**Editorial comments:**  
  
***Editor’s Note: Please note that multiple reviewer # 1 has raised some concerns about aspects of your manuscript. Please thoroughly address or rebut each individual comment below to further strengthen and clarify your submission.***   
  
  
**Reviewers' comments:**  
  
**Reviewer #1:**   
*Manuscript Summary:*   
The manuscript by Feather-Schussler and Ferguson presents a battery of functional tests adapted to neonatal mice, and their application to the evaluation of deficits in a neonate mouse model of cerebral palsy.

This summary is accurate, however, misses some key details. This battery of tests was not ‘*adapted’* for neonatal mice, but was compiled from already published neonatal behavior tests, with appropriate references cited within the manuscript. This manuscript is not claiming to have invented these tests (with the exception of the foot-angle hindlimb ambulation test), but was designed to demonstrate the proper technique for performing a battery of neonatal motor tests using our qualitative scoring sheet, and using JoVE as a platform to visually demonstrate how each of these tests are to be performed. One of the major problems with behavioral tests, especially in neonatal rodents, is that the text descriptions on how to perform these tests and score them are not clear. The current literature does not have good supportive figures or finely detailed descriptions, causing each research group to “modify” the tests and their reporting method. By publishing in JoVE, the video format will show researchers how to perform these tests and report their results using our quantitative scoring sheet so that data can be compared between research groups reliably and reproducibly.

Each test was specifically selected based upon the fact that neonatal mice are able to perform the behavior with reliable and repeatable results. These specific motor tests were chosen to emphasize motor deficits and place less emphasis on sensory-based behaviors. Furthermore, these tests can be used on a wide variety of injury paradigms, whether in our model of cerebral palsy, used as an example model here, or in other neonatal neuromuscular or muscular disorders where motor function is affected.

*Major Concerns:*  
The first part of the paper describes a battery of nine tests based on a subset of tests previously developed by Fox and widely used in a number of studies. The tests are presented with too much approximate age delimitations to be applicable in developmental studies. Thorough analyses of neuro-behavioral developmental testing were already published and are more interesting for developmental studies (see Leroy et al 2001 for instance).

All reported developmental ranges were taken from previously published papers and were cited, as appropriate and are also listed in the Le Roy 2001 paper, as mentioned by the reviewer (age range from 0-17 days). The tests are to determine whether a motor deficit persists in a neonatal injury model, either over time or following therapeutic intervention (not shown). Additionally, a comprehensive battery of neonatal motor tests is not found elsewhere, including the Le Roy 2001 paper, which is a statistical analysis of Fox’s battery of tests with no description regarding scoring or how to perform each test. Our paper highlights motor functions that may be impaired after injury, rather than changes during normal development. Fox (1965) and Wahlsten (1974) test for a combination of behaviors, including the development of motor and sensory functions. Their tests compare the onset of a behavior in the neonate to a mature or adult phenotype with a 0 or 1 scoring system (present or not present). Our paper compares neonatal behavior within the context of an injury paradigm with a more sensitive scoring system and a visual description of how to perform each test. Thus, our paper is unique in this field.

The authors propose to use parametric tests (Student t-test) to analyze the data. This is a misleading procedures in some cases, because several tests use qualitative scoring (ambulation, HL suspension, grasping reflex). Student t-test should not be used with categorical variables: use Chi2, or non-parametric tests at least.

Non-parametric tests are used when data is nominal (yes/no) or ordinal (ranking). These data are interpreted using subjective findings. Non-parametric tests are incorrect and less powerful than parametric tests for our studies. Because the three tests listed above are scored using interval (scale of 1-5 with equidistant points) or ratio (having a true zero point) scales, parametric tests are appropriate; although non-parametric tests can be used with interval or ratio data. However, this statistical analysis would be less powerful. With that being said, a non-parametric Mann-Whitney test for ambulation, hindlimb suspension, and right paw preference in the grasping reflex was determined for comparison sake. The non-parametric p values were extremely similar to the parametric p values and all values were still significant. Additionally, a statistician reviewed the data prior to paper submission and found that all statistical analyses were appropriate. The paper will remain as is.

The second part reports representative results based on the functional evaluation of a mouse model of cerebral palsy carried out on a single measure at P8. The results are limited to differences in the hindlimbs (HIL angles during crawling, HIL Strength and grasping) without allowing to decide if they results from leg weakness or a delay of maturation.

We did perform all tests at a variety of time-points, ranging from 24 hours after surgery (P7) to 1 week following surgery (P13). We reported results 48 hours after surgery (P8), as this time point showed the largest deficits, as part of our “representative results”. Our primary goal was to show there are motor deficits in our model, not the developmental time course of these deficits, nor whether they are due to delayed maturation versus weakness. We are in the process of writing up our full data, including therapeutic intervention, but that was not the purpose of this particular manuscript. As previously mentioned, the purpose of this manuscript was to properly demonstrate how to perform neonatal motor tests complied from a variety of cited sources using the JoVE video format to assure reproducibility and accuracy, using our quantitative scoring methods rather than subjective reporting methods.

A single stage measure does not allow to detect correctly a developmental delay. Neurobehavioral developmental analyses need a longitudinal analysis of the appearance of critical functions. In addition several of these results are based on inappropriate statistical testing.

While the reviewer is correct in that a longitudinal analysis would have to be performed to detect a developmental delay, this was not the purpose of our study. Our intentions were to determine the presence or absence of motor impairment in our cerebral palsy model, which we feel was accurately shown. Additionally, the statistical analysis has been addressed above.

*Minor Concerns:*  
Minor points:  
In the Statistical significance chapter the authors suggest to "express data as mean +- SEM". They did not applied it to their own data that are apparently expressed as mean +- S.D. (or the tests are not

Although we cannot assume what the reviewer is attempting to say with the rest of this statement as it is not provided, we can guarantee that the data shown is expressed +/- SEM.

3.2.1. HL foot angle:The measures on the feet should be done better by filming from below on a transparent surface.

We agree with this statement and we have changed the wording to indicate foot angle can be viewed from above or below, as the figure showed. We thank the reviewer for catching this wording omission.

3.8.3 the scoring is not clearly stated.

We have re-worded the grasping scoring and paw preference.

3.7 four limb grip strength. The description of the test is confusing.

We appreciate this comment, but since no specific points of confusion were mentioned, we are unsure what needs editing.

Li323. P8+7days, is probably to late. So you cannot conclude to a delay in the development. It would have been more pertinent do determine the age of acquisition of quadruped locomotion.

We are unclear to what this statement refers. Lines 322-323 are as follows, “2.9.3) Once both the snout and paws have been removed from the edge, stop the timer and record time.” Unfortunately, even trying to determine where this comment came from in the text was impossible, unlike the comments below. If the reviewer would care to clarify, we can appropriately address this comment.

Li328: values for hil and sham are inverted.

Thank you for finding this flaw (line 349-350 in our version)

Li330. This remark is an over-interpretation.

Line 330 does not make any assumption; it is in the methods section. However, we believe the reviewer is concerned about our interpretation of what a wide-based gait indicates. From many human studies, as well as various rodent injury studies, a wide-stance during gait activities is a sign of imbalance or ataxia. We do not feel this remark is an over-interpretation.

Li 345 what is the logic to test at P8 a function that develop after P10 ?

We tested forelimb suspension from P7 to P13 knowing that the selected time point we used for demonstration (P8) was too young for this test. We have clarified the time points we actually examined at the beginning of our representative results section. We still included this test result due to JoVE’s policy that ALL results, positive or negative, are presented. We do have a clear statement within the paragraph that reads, “This test is better for mice older than 10 days”. If the reviewer feels strongly about this result and if it is okay with JoVE, we will remove this result -or- report a different time-point that is within the appropriate age range.

Li 348: decrease in suspension time (data not shown ).

Are you referring to the Front-limb suspension test results? If so, does the reviewer suggest that we add the wording, “decrease in suspension time”? However, for forelimb suspension, there is no significant difference in the suspension time. We have adjusted the manuscript to attempt to clarify this section.

Li 356 : Test values for suspension time ?

We misworded this section and thank the reviewer for the comment. We have fixed this statement.

Li 357 P>0.005 ? (idem li 365, 429, 433, 443,452)

We thank the reviewer for catching this “copy and paste” error! The multiple errors have been corrected in the document.  
  
**Reviewer #2:**   
*Manuscript Summary:*   
In this paper, authors created a neonatal mouse model for cerebral palsy, incorporating the major phenotypes associated with the disorder, including hypoxia, ischemia and inflammation. One of the main neonatal behavioral readouts of cerebral palsy is motor dysfunction, which persists into adulthood. Therefore, in addition to creating a neonatal mouse model for cerebral palsy, authors have also attempted to establish a series of repeatable motor tests, designed specifically to test motor function in neonatal mice up to two weeks of age. The surgical protocol and subsequent techniques used to effectively induce cerebral palsy in the neonatal mouse appear to be appropriate. Although the established model is appropriate to test deficits associated with cerebral palsy, I find a few potential problems with some of the behavioral assays testing motor outcomes of the disorder, as I think a few of the assays may cause stress and anxiety to the neonate, which may potentially interfere with the proper testing and outcome of motor function.

The stated summary is very accurate; however, the behavioral tests are well established and thus, the stress and anxiety inherent in these behavioral tests are minimal and do not interfere with the results. All of these tests are well-established and referenced in this and other papers, with the exception of the gait angle test, which is a novel gait analysis test for neonatal mice.   
  
*Major Concerns:*  
For examples of potential methodological confounds:  
a) Authors suggest "gentle prodding by touching the pup's tail" (line 182) be used in order to motivate the pup to walk in the Hindlimb foot angle assay. 1) I question the consistency and repeatability of the interaction with the animal across all cases. 2) Instead of motivation, could the prodding cause stress/anxiety in the neonate leading to more freezing behavior?

Because angles of the hind limbs as well as ambulation scoring are not taken when the mouse first begins walking, when the mouse stops waking, nor when the mouse is turning around while walking, the repeatability is high and interaction between observer and mouse does not pose a problem in obtaining data. Furthermore, gentle prodding of the mouse does not cause freezing behavior, increased anxiety or stress, as has been shown previously. Many papers with mouse behavior cite gentle prodding or pinching of the mouse tail to coax the mouse to continue forward movement, including (at random): a 2011 JoVE paper titled “Assessment of Motor Balance and Coordination in Mice using the Balance Beam” by Luong et al.; a 2001 paper published in human molecular genetics titled “The HD mutation causes progressive lethal neurological disease in mice expressing reduced levels of huntingtin” where gentle prodding of the tail was used to cause adult mice to walk across a surface so hindlimb gait could be determined; and in 1998, the Journal of Cell Biology published a paper titled “AnkyrinG Is Required for Clustering of Voltage-gated Na Channels at Axon Initial Segments and for Normal Action Potential Firing” where authors prodded the tails of mice to observe gait. This is not an exhaustive list of publications in which prodding of the tail has been used to observe gait, therefore, we do not believe gentle prodding of the tail is negatively or inconsistently affecting the results of gait analysis.

b) Surface Righting (line 193). Authors suggest putting the pups on their backs and holding them down, "in position" for five seconds then releasing the pup and recording the time it takes for the pup to return to prone position. Wouldn't forcefully holding the animal down be a confound in the measurement of time it takes for the animal to change position, as it can induce stress/anxiety?

While your concern is reasonable, again, restraining the mouse is not long enough to cause stress and is gentle enough that the pup lays still in this position. While adult mice undergo the “fight or flight” response when turned over, the pups are too young to undergo this fear response. Righting reflex is a common neonatal motor test and the practice of holding the mice on their back is well-established.

c) Cliff Aversion (line 307). Olfactory cues could serve as potential confounds. How do authors plan on removing olfactory cues from the box? Also, if the cliff aversion is used to test for inherent fear, how's that related to motor impairment or vestibular difficulties. It is not clear how an aversion test is measuring motor control without also measuring fear response (line 410).

While it is true that mice respond to olfactory cues, we remove cues by over-scenting the box with many mice. We neglected to add that to our description and thank the reviewer for noticing this. We have now added the box pre-scenting to the description. Pre-scenting the box makes it so the mice do not follow olfactory cues because there are too many cues to follow. In terms of the fear response, the test is examining whether the mice recognize they are at an edge, and then we observe their motor response, which include a horizontal vestibular movement. Motor impairment (not moving or moving with very poor limb coordination) or side preference (turning) is determined when the mouse moves (either to the right or the left) from the edge of the box. The response to turn from the edge of the box is inherent in mice and this test is well established.  
  
*Minor Concerns:*  
Line 78: Inflammation of what region of the brain?

Chronic inflammation is from an unknown source, which may include chorioamnionitis, maternal illness or fetal infection. In this model, inflammation is due to a single injection intraperitoneally of lipopolysaccharide, a well-characterized model of inflammation. As to exactly which areas of the brain react to this global inflammation is unknown, but the animal undergoes a general increase in inflammatory cytokines.

Line 329: this section: 3. Statistical Significance  
328 3.1 Using a statistical software analyze the results. Express data as mean ± standard error of the mean (SEM). Tests are parametric so analyzed them by a t-test.   
  
Should be re-written as: (NAME OF SOFTWARE) was used to analyze the results. Data was expressed ad mean +/- standard error of the mean (SEM). Tests were parametric and thus, the data was examined using t-test analyses.

Thank you; we have changed the paper to reflect this suggestion.

Line 371: Authors state that this test is more appropriate for mice older than 10 days, but conduct testing at PND 8. Please provide rationale for this statement.

We tested forelimb suspension from P7 to P13 and have clarified the age ranges tested in our introduction to representative results. We only report one representative time-point (P8) for all tests, knowing that the selected time point was too young for this test. We still included this test in the results as it is a test that could be used for neonatal motor disorders at P10 or later, and JoVE asks for “representative results” for all sections. We can change this result from P8 to another time-point, but it will be different than the other representative time-points reported. We are in the process of publishing a paper with all results, including therapeutic intervention. This paper reports methods of testing neonatal motor deficits using JoVE’s video format to demonstrate proper technique using our quantitative scoring sheet to allow for reliable and reproducible testing between research groups.

*Additional Comments to Authors:*  
Overall this is an important and interesting method paper that highlights fairly novel ways to test very young mice for motor control.  
  
We thank the reviewer for his/her kind words.

**Reviewer #3:**   
*Manuscript Summary:*   
This manuscripts describes a battery of tests to detect some forms of cerebral palsy in newborn mice, illustrated by the authors' own results employing the Rice-Vanucci model of stroke caused by unilateral carotid artery cauterization followed by a period of systemic hypoxia.   
  
*Major Concerns:*  
The major strength of the manuscript is the detailed description of the experiments, so investigators considering to repeat this type of experiments might find valuable Information.   
  
*Minor Concerns:*  
The introduction could be shortened significantly, while a section on pitfalls might be helpful

Thank you for your review. If the reviewer has specific points or parts of the introduction they feel are superfluous, we will be happy to edit them out.

Pitfalls are inherent when working with animals. Although there is not a ‘pitfalls’ section in the paper, the major pitfalls for some of the tests have been outlined within each test. For example, we suggest the examiner determine the scoring for pups that fall in negative geotaxis, that there can be learning in the absence of negative reinforcement for falling in the hindlimb and front limb suspension tests, and that grasping reflex must be performed prior to 15 days, at which time this reflex disappears (Crawley, 2007: What’s wrong with my mouse? Behavioral Phenotyping of Transgenic and Knockout Mice, 2nd Ed).