**Response to Reviewer #1:**

*Major Concerns:*  
1. The authors present a battery of tests. It would be beneficial in the text and subsequent video to mention other behaviors that can be substituted into this battery. Fujimoto, 2004 (Neurosci Biobehav Rev) and/or Gold, 2013 (Regen Med) provide a nice breakdown of various tasks for different aspects of function (e.g. Fine motor, Gross motor, somatosensory, etc).

The last paragraph of the Discussion has been expanded to include mention of other possible tests that can be substituted into the battery presented in this paper and the two references mentioned above have been cited. There is some concern that this is beyond the scope of this paper, as there is only room to mention each alternative example by name and not detailed description as this is not meant to be a comprehensive review.

2. It would be good to note the behaviors that this battery will not address: depression/emotionality-related deficits, social interactions, decision making, impulsivity. This is hinted at in the discussion, but not particularly emphasized.

A statement to this effect has been added to the discussion.

3. TBI models have many different manifestations. The current paper utilizes FPI, which is typically regarded as a moderate to severe injury. Would this battery be as effective in milder or concussive TBI?

Considering this is a methods paper describing the method that we use, it is beyond the scope of this paper to speculate on whether or not this battery of tests would be as effective in other forms of TBI. These are experiments that could be done in the future.

4. The authors should clarify in the neuroscore section where repeating by side adds to the score (e.g. Repeat on opposite side and score (0) or (1)). A quick read-through suggests 5 total points.

Possible scores have been added to each section of the neuroscore for clarification.

5. The description of the Beam Walk apparatus is not clear. The video may clarify this, but I am not clear on what purpose the 'pegs' serve.

The pegs are inserted at equally spaced intervals along the beam to create obstacles for the rats to navigate as they walk the length of the beam. The purpose of the pegs is to increase the difficulty of the Beam-Walk task.

6. The figures do not match the methods of analysis. Since continuous data (Group x Day) is being presented, they should be line graphs. When a bar graph is presented, it implies that the data are independent between days. This is especially egregious for the MWM data which appears to be completely collapsed across multiple days of testing.

All the figures and associated statistical analysis has been revised. Please see Section 6.0 Data Analysis and the Representative Results section.

7. For the MWM analysis, please clarify if you are using a 3x5x2 (Injury x Day x Trial) ANOVA. A simpler approach might be to do an ANOVA on the difference scores, which would also make the graphing clearer by providing a single data point.

We have revised our analysis. We now use a two-way repeated measures ANOVA with Injury as a between group factor and Day as a within group factor. When a significant difference was found, a Tukey’s post-hoc test was used to find which groups were different.

A one-way ANOVA was used to determine if there were differences between NAÏVE, SHAM and TBI on each day. If a significant difference was found, a Tukey’s post-hoc test was used to determine where the differences lay.

Figure 5 now shows the line graph time course of each Trial of the working memory water maze test as well as the summarized data I a box and whisker plot.

8. The posthoc analyses suggested would be inappropriate for anything beyond 3 groups, and should be noted as researchers who are unfamiliar are likely to follow the protocol very closely.

Please see Section 6, Data Analysis for our revised statistical analyses. For post-hoc testing we used either the Tukey’s post-hoc test or pairwise t-test comparisons as necessary.

9. Figures 3 & 4 are missing a y-axis label. The y-axis of Figure 5 is confusing - it indicates that the authors have summed the average latencies, when it appears to more likely just be the average.

Thank you for pointing out the missing labels. Figures 3 and 4 have been corrected. The label for the Y-axis of Figure 5 has been changed to say “overall average”.

10. Some discussion of the inherent problems of using ordinal scored tests (NeuroScore, Beam balance), particularly as it applies to domain-specific impairments as opposed to general impairments as well as statistical challenges would improve the discussion.

A paragraph has been added to the discussion to address these issues.

*Minor Concerns:*  
1. It should be noted that while the beam balance is sensitive to vestibular dysfunction, gross motor dyscoordination will also generate deficits. A similar statement could be mentioned about cerebellar deficits on the beam walk task. This is brought up again in the discussion - do the authors think it is more likely to be gross motor deficits that mimic vestibular dysfunction, or actual vestibular dysfunction?

This is a good question and possibly not one that we can prove one way or the other with these data. However, these rats do not look grossly impaired and perform quite well on all these tasks with only small differences detected compared to sham or naïve rats. Thus we speculate that the small deficits we detect are due to vestibular dysfunction more than gross motor dysfunction.

2. A comment on counterbalancing design such that injured and sham animals are interspersed in the testing period would be beneficial. Sadly, some researchers do not consider the effects of variables such as these.

A statement mentioning our use of a balanced design has been added as a note in the surgical procedures section.

3. The authors mention the MWM-WM can be used to test reference memory. Please clarify what is meant by this as there are usually separate reference and working paradigms for this task.

It is true that there are separate tests designed to evaluate reference memory or working memory. We suggest that the Trial 1 latencies are representative of reference memory. Considering that there are four possible platform locations, even though these locations change randomly over the course of the 5 day testing period, the rats demonstrate that they learn the four platform locations. This is shown by the decrease in the latency of Trial 1 over the 5 day period. This indicates that the rats retain the platform locations over the 5 days, indicating reference memory.

The Trial 2 measurement is a measure of working memory because the location of the platform for trial 2 is only relevant for that one trial and then it is no longer useful in the next session.

4. The original reference you provide for the FPI model describes an air blast, which eliminates the fluid component. A qualification ("a similar model was first published…") or an update to the 1976 paper describing the modern iteration of the device would be preferred.

I apologize if there has been a misunderstanding. We are not aware of the 1976 paper that is mentioned here. The articles we referenced describing the FPI model were:

1. (reference #10 in the manuscript) Dixon, et al., J Neurosurg 67:110-119, 1987 entitled, “A fluid percussion model of experimental brain injury in the rat”, in which Figure 1 is a diagram of the exact device we currently employ in our laboratory.

2. (reference #11 in the manuscript) DeWitt, et al., J Neurotrauma 14(4):223-233, 1997 entitled, “L-Arginine and superoxide dismutase prevent or reverse cerebral hypoperfusion after fluid-percussion traumatic brain injury”. This paper is an early paper from our laboratory.

Both of these papers describe the fluid-percussion injury model. Additionally, we have added our recent paper Sell et al., J Neurotrauma 33:1-11, 2016, which also describes the injury procedure.

5. The authors make the claim that FPI is the most common method of injury. CCI is arguably as common or more common at 1200 references [PubMed search string: ("controlled cortical impact" OR "cortical contusion injury")]. The referenced search does not link properly in the references. Further, the search string should be [("fluid percussion injury" rat)] rather than terms without the quotes. When searching for the term without the quotes, it returns any abstract that contains all of the words, while the quoted term returns all abstracts with the words in that specific order. This search returns 427, a number which is much lower.

This statement has been altered to reflect this point and now reads, “…the fluid-percussion injury model is one of the most widely used…”

6. A note should be added to 1.2.4 that this takes considerable practice to reliably elicit a vibrissae reflex, and that researchers should be well trained on this aspect.

A note to this effect has been added to what is now 2.2.4.

7. Spelling error in 3.2.13 'noice'

Corrected – thank you

8. Section 4.1 should note the size of tank necessary for rat water maze

The dimensions of the tank are in the Materials Table as specified.

9. In your MWM materials, the authors describe everything as conducted by hand, but then reference AnyMaze in the analysis.

Section 5.1.8 has been added to mention the use of a video tracking system.

**Reviewer #2:**  
*Major Concerns:*  
1. While the tasks chosen for motor and cognitive function are commonly used among traumatic brain injury researchers, it is a limited list. More complex exams are being used in conjunction as MWM along is not a complete story. Novel object task, rotorod, gait, social interaction, Y-maze, Barnes maze are all techniques that are often preferred by reviewers and researchers in the field. There should be mention of the other exams.

Yes, there are many other useful behavioral measures of brain function. Describing these is beyond the scope of this paper. The discussion has been expanded to include mention of other paradigms.

2. The surgical procedure for FP involves intubation, which is not always used by researchers who utilize the injury model. How might this affect behavior? It is also important to mention whether or not a stereotaxic frame was used and the type of ear bars. Puncturing of the ear drums will directly affect animal behavior. This is not mentioned at all. There is no description as to the type of TBI pathology produced with the FP in this study. Is there cell loss? Contusion? The severity of pathology will dictate the duration of detectable impairments. This must be clarified.

Intubation and a stereotaxic frame have been the standard procedure for our laboratory since at least 1997 (see DeWitt et al., 1997 mentioned above and referenced in the manuscript). We have no evidence that intubation alters the results of behavioral testing, but the reason that SHAM animals are included is to take into consideration all of the effects of the surgical preparation which includes intubation and stereotaxic mounting, minus the injury. Our stereotaxic technique does not puncture the ear drums. Since this manuscript is focused on behavioral methods, the readers are referred to two previous publications for details on the FPI surgical technique.

Several previous publications have addressed the type of brain pathology that is produced by FPI. We expanded the introduction to address the issue of neuronal injury and added some references.

3. The beam balance and beam walking beams are wood, which is great for gripping, but they also absorb urine…there is no mention of wood treatment. This can especially affect the walking.

The wood is sealed with polyurethane varnish; this has been specified in the Materials Table.

4. There is no mention of a video camera for beam walking. This is often used for analysis of the limb slips as they go across the beam. A mirror can be used for the contralateral limb. Analysis of only time is limited.

We do not use a video camera for our beam walking task as we do not measure limb slips.

5. In the MWM task description, there is no mention of extramaze cues (type, size or location relative to the maze). This varies a lot between labs and can be a major issue. Also any intramaze cues can alter results. The use of a plexiglass platform can also be problematic since it is really slippery. Others have made multiple holes in the platform for the animals to grip. There is also no mention of the color of the tank, which is essential for video tracking.

Thank you for bringing up this important point. The walls surrounding the pool have large shapes including three dimensional objects hung on them and one side has a curtain that the computer and researcher stay behind while the rat is swimming. There are no intra-maze cues. The plexiglass platform has silicone applied to the surface that has hardened and provides a surface that the rats can grip. These details have been specified in the Materials Table.

The tank is white, but with the proper settings on the video-tracking system this is not a problem.

6. There is also no mention of the temperature maintained in the tank. The water temperature will affect performance.

This is a good point. We maintain the water temperature at 26+/- 1o C. A statement to this effect has been added to the protocol.

7. The duration of time allowed to find the platform is also something that varies between groups. 120s is a long time and can add to fatigue and poor performance. No mention of this.

It is true that the time allowed to the rats to locate the platform varies between laboratories and experimental designs. In fact, one of the early Morris papers (Morris, 1981) allows 180 seconds. However, in our search of the literature 120 seconds seems to be quite common and we have not had any problems with fatigue in our rats.

8. The analysis for MWM is also rather simplistic. Most researchers now utilize tracking programs for latency, distance, pathway pattern analysis. There is much more data that can be acquired than latency alone. Also the 2 day testing is not always used. Many people use the MWM for working memory without pre training. This is not mentioned.

The explanation of the use of AnyMaze has been addressed previously. (see item #5). Item 6.4.2 in the protocol mentions that other outcomes can be measured using tracking software:

“6.4.2. Many possible outcomes are available for analysis from computer tracking programs so outcomes to be analyzed must be selected. Examples of outcomes selected for analysis may include: latency, path length, thigmotaxia and swim speed.”

We choose latency to use as an example because it has proven to be the most reliable and significant outcome measure for this version of the water maze.

We are confused about the mention of “2 day testing” and “pre training”. Our paradigm includes 5 days of testing and no pre-training. On each day there are 4 sessions which are distinguished by 4 different platform locations. The rats are given two trials per platform location. The first trial is a learning trial and the second trial is the memory trial. Please see section 5.2 of the protocol.

9. These tasks can be used with other age groups and gender, for which accommodations need to be made in the tasks. Including pretraining duration, beam size and length, durations of swimming in the maze. These needs to be mentioned.

A statement mentioning these issues has been included in the discussion.  
  
**Reviewer #3:**  
*Major Concerns:*  
1. Authors should re-write the whole section of discussion to describe more clearly about the interactions between the results from different neuromotor tests after brain injury. For example, why the score is increased in naïve group but decreased in TBI group in beam walk and beam balance tests after 6 months later? Why the behavioral situation is getting worse in the control group but getting better in the brain injury rats?

A section has been added to the discussion to explain the results of the 6 month beam balance and beam walk test:

*“The results of the Beam-Balance and Beam-Walk tasks shown here demonstrate that these tests are useful early after injury to detect deficits in vestibulomotor function. These deficits typically resolve over time. In this model, by six months after injury, the injury-induced deficits have resolved (Sell et al., 2016; Dixon et al., 1987) The results of the six month time point indicate that there are no differences between naïve, sham or injured rats. However, all the rats have been relaxing in their home cages for 6 months, aging and gaining weight. Thus, by the time they are re-tested at six months post-surgery (or equivalent in the case of NAÏVE), they are essentially becoming old and fat and therefore all the groups do not perform as well as they did compared to their baseline Day 0 results.”*

*Minor Concerns:*  
1. Authors should explain the meaning of score of beam walk and beam balance tests in the section of protocol.

The Beam-Balance scores are defined in the protocol section 3.2 as follows:

**1** = Shows stable balance (grooms, walks, attempts to climb barrier)

**2** = Shows shaky balance (grasps sides of beam and/or has unsteady movements)

**3** = Tries to balance but slips or spins on the beam, hangs on by hugging the beam

**4** = Tries to balance, but falls after 10 s

**5** = Hangs over or from the beam and falls off in under 10 s

**6** = Falls off, making no effort to balance or hang onto the beam

The Beam-Walk is a timed assessment, so there is no score as such. The time it takes the rat to traverse the beam and enter the escape box is recorded in seconds. Three timed trials are conducted and the average of these trials is the outcome for that rat for that day. These processes are described in the protocol Section 4,

2. Some citations are inappropriate; please check them clearly.

The citations have been carefully checked. It would have been helpful if the reviewer would specify which citations s/he felt were inappropriate. We did find that the wrong Morris paper (Morris et al., 1982, *Nature*) was listed, so this has been replaced with Morris & Hagan, 1986 *Q J of Exp Psych*). We hope this is satisfactory.  
  
**Reviewer #4:**  
*Major Concerns:*  
-The limitations of the FPI model should be discussed. The variability in this model is high for example. Further I believe the CCI model is more common in the present times.

We have altered the wording to reflect that the FPI model is one of the popular models, not the most popular model.

-A description of the location of the injury and how that affects behavioral outcomes should also be included.

The details on the site of the injury are available in the references cited. The focus of this paper is how to do behavior testing. A statement mentioning that the site of the injury could alter the outcome of the tests has been added.

-A description of pre training should be included.

The pre-training is described in the protocol section for each of the behaviors that requires pretraining.

-How do you get a total score for each round of 7? Each test (5 in total) only with 0 and 1 as possible scores suggests the formula should be 5\*3=15 as the high score.

Possible scores have been added to each section of the neuroscore for clarification.

-For the Beam-balancing test you should specify how long to wait between each trial during training and during experimentation.

The words “allow the rat to rest for 15 s in the home cage between trials” has been added to the Beam Balance section.

*Minor Concerns:*  
There are some typos in the article that should be corrected.  
Thank you for pointing these out. We have attempted to rid our manuscript of all typos.