Morgan et al propose a registered report to assess the role of reward on sleep-dependent memory consolidation. Several studies have been conducted in this area with mixed results. They propose to conduct a large-scale online study to potentially reconcile these discrepant results, given several previous studies have had a relatively low N and therefore may have been underpowered. Overall, I am impressed with the level of care and detail that has gone into planning and presenting this Stage 1 RR and think the study could make a valuable contribution to the literature.

**1A. The scientific validity of the research question(s).**

The research question is scientifically valid. There is a lot of evidence in favour of sleep-dependent consolidation (i.e., an active consolidation process during sleep). Given the number of memory traces encoded during the day, it is likely that some process of selectivity is needed to consolidation specific traces. One possible driver of selective consolidation is reward at the point of encoding, and there is some understanding of the neurobiological basis for this relationship, involving dopamine and the VTA and hippocampus.

**1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.**

**The logic and rationale of the proposed hypotheses are appropriate. Given past research, both hypotheses are plausible. The authors present two clear hypotheses: (1) memory performance will be greater following a period of sleep than wake (i.e., a standard “sleep effect”), and (2) that the sleep effect will be greater for high than low reward stimuli. This latter hypothesis is the critical hypothesis related to reward and sleep-dependent consolidation. The former isn’t presented as an control analysis to ensure the sleep manipulation has worked, but could be interpreted as such (see comments to criteria 1D).**

1. **Is a significant sleep effect in H1 critical to assessing H2? How will H2 be interpreted in the absence of an effect for H1?**
2. **In relation to H2, you appear to be predicting no sleep effect for low reward items (and an effect for high-reward). Given the ample evidence for sleep effects in studies that don’t include a reward I wasn’t sure about this. Is it not more likely that a sleep effect will be seen for both low and high reward, but that it is greater in the high reward condition?**

**1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable).**

**The methodology and analysis pipeline are appropriate, however I have several comments/questions in relation to this:**

1. **Overall the methodology and analysis pipeline are clear and explained in great detail. However, that detail came at the expense of clarity to me. I don’t want to increase the length of the manuscript more than is necessary, or remove any of the detail (which is needed), however I wonder whether a summary/overview is needed at the beginning of the methods, or a reordering of the methods might help. For instance, I wanted more info about the actual experimental task (Figure 3 and 4 could have been presented earlier) and overall experimental design before I then tackled the detail.**
2. **Data collection and demographically diverse sampling is ambitious (high N, 8 month data collection window, wide range of individuals). I wonder whether the authors have any backup plans in terms of ensuring they reach their target N (e.g., if they can’t collection enough data in one demographic area, will they sacrifice this aspect of the study to ensure they reach the target N, or will they sacrifice N to ensure a representative sample)?**
3. **Participants will be told not to take a nap in the wake condition. Is this potentially problematic to individuals who do typically nap during the day (e.g., older populations)?**

**1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses.**

**There is appropriate methodological detail and the analysis pipeline is appropriately explained. Again, I have a few comments/questions though:**

**Methods:**

1. **The N reported in the abstract is different from that reported in the main text, so please correct this.**
2. **The two sessions will be conducted a maximum of 4 weeks apart. What is the minimum gap between sessions?**
3. **On p. 18 you describe the total number of stimuli etc., but I couldn’t see information on the number of encoding trials per item. I presume it is one?**
4. **On p. 28 you explain the 4 validation questions. It isn’t clear to me whether inclusion only occurs if all 4 questions are answered correctly, or exclusion occurs if all 4 questions are answered incorrectly.**
5. **On p. 29 you explain the “seriousness check”. I presume during this you still make clear that their answers to these questions will not affect their inclusion in the prize draw, otherwise they have a stake in saying they were being serious.**

**Analysis:**

1. **On p. 29 you explain exclusion criteria in relation to d’. I think this is d’ collapsed across all conditions (which you do mention later), but I think this should be made explicit here.**
2. **In relation to both H1 and H2 the hypothesis is directional, but the interaction analysis isn’t. You state you will follow up significant interactions with posthoc tests to appropriately characterise the interactions, however there is a slight gap in relation to the interpretations. For example, in Table 1 you state “If there is no difference in memory performance between the sleep and wake groups…”. What if a difference is seen but it is the opposite to that predicted for H1? If this occurred, I don’t think you have made clear what the conclusion would be. This probably just requires a slight tightening of the wording in relation to the last two columns of the table to ensure you have covered every statistical eventuality.**
3. **On p. 31 you describe the GLMMs, and that they will include “all interactions, main effects and random slopes for each participant for all parameters”. What about intercepts in this model?**
4. **On p. 31 you say “reward will be a mean-centred continuous predictor”. I presume this is a linear predictor, based on the (mean-centred) raw reward values, but it should perhaps be made clear.**
5. **On p. 32 & 38 you discuss adding covariates to the analysis (e.g., memory performance at immediate test). It isn’t clear whether these will be added to the primary analysis, or whether the primary analysis will be conducted without these covariates and you will then run a second analysis including the covariates.**
6. **Why is the p-value set at p<.02?**

**1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).**

**I do not think the authors have presented a positive control to ensure the data are of sufficient quality to assess their main hypothesis. There are multiple checks of data quality, which will help in relation to this. One potential possibility is to only test H2 conditional on H1 being true (i.e., a significant sleep effect, regardless of reward, is present). However, this might be too limiting. Another possibility is to assess a reward effect at immediate test. There might be other possibilities that I have missed.**

**Further comments:**

1. The first sentence of the abstract is overly complex and could be simplified for clarity.
2. The mention of psychiatric disorders in the abstract does not seem appropriate to me.
3. In the introduction there is quite a bit of discussion about dopamine and the VTA. Although interesting and relevant I wonder whether this could be reduce somewhat. There could also be better signposting on which studies in the introduction are human vs non-human research.