CHATGPT Thoughts on my power analysis

Your power analysis appears to be thorough, and you've considered different slopes (differences in probability between dominant and submissive conditions) for the effect of the manipulation. Here's a breakdown of key points and some suggestions:

**Strengths:**

1. **Multiple Effect Sizes**: You've tested power estimates for various manipulation effects (6%, 10%, 16.7%), which is a great way to assess how sensitive your study design is to different effect sizes.
2. **Use of Mixed Models**: You’re appropriately using a GLMM with random intercepts for participants, which accounts for the variability in participant responses and is more realistic than simpler models.
3. **Power Simulations**: The powerSim() and powerCurve() functions are used well, providing power estimates across different sample sizes, ensuring a good understanding of how many subjects you need to reach a desired power level.

**Suggestions for Improvement:**

1. **Variance in Manipulation Slopes**: From the pilot data, it seems that the variability in slopes across participants is quite small. If you expect more variability in a larger sample, you might want to simulate data with a larger variance in slopes to reflect this. This would likely lower the power estimates since higher variability means the effect will be harder to detect.
2. **Sample Size Considerations**:
   * **With a 6% difference in probability** between conditions, the power is quite low, even at larger sample sizes (e.g., 62% at 1400 rows or 2240 subjects). If you're targeting a small effect, this suggests you'd need a very large sample to achieve 80% power. Given the large numbers needed, consider whether such small effect sizes are practically meaningful in your context.
   * **With a 10% difference** in probability, you achieve around 77-96% power at sample sizes of 50-100 participants. This seems like a reasonable sample size, and if this effect size is realistic, it may offer a more achievable target.
3. **Model Specification**: In your simulation code, you're using only random intercepts ((1 | player)), but if there's enough variability in individual slopes (as your variance-covariance matrix suggests), you might also want to include random slopes for manipulation ((manipulation | player)). This would model individual differences in response to the manipulation and give you a more accurate estimate of power, especially if you expect participants to vary widely in their responses.
4. **Confidence Intervals**: The confidence intervals (CIs) provided by powerSim() are a good way to assess the precision of your power estimates. For the smallest effect size (6%), the CIs are wide, suggesting that there's quite a bit of uncertainty around those estimates. This further underscores the need for larger samples when detecting smaller effects.
5. **Pilot Data**: Your pilot data suggests a Cohen's d of 0.33, which is a small-to-moderate effect. If your final study effect sizes are similar to this, aiming for a larger sample size (e.g., around 100-150 participants) seems reasonable to ensure sufficient power, especially if you’re targeting a power level of 80%.

**Final Thoughts:**

* **Large Samples for Small Effects**: If you’re primarily concerned with small effect sizes (e.g., a 6% probability increase), then be prepared for large sample sizes, as you've noted. However, for more moderate effect sizes (e.g., 10-16% probability increase), your sample size estimates (50-100 participants) look more achievable.
* **Modeling Variability**: If you believe that participant variability in response to the manipulation could be significant, ensure that the variance of random slopes in your simulation reflects this. This would provide more realistic power estimates.

Your current approach is solid, and with adjustments to the variability of slopes and a focus on effect sizes with practical significance, you’re in a good position to make well-informed decisions about your study design.