Investigating the Moderating Effect of Respiratory Masks on the Association Between H2S Exposure Levels and Cognitive Performance

I. Introduction

The current analysis investigates the moderating role of respiratory mask type on the relationship between hydrogen sulfide exposure levels and cognitive performance post-exposure. The subset of data analyzed here includes 200 employees working in industrial environments with exposure to hydrogen sulfide, a toxic gas known to affect attention, reaction time, and working memory. Participants completed regular work tasks for 60 minutes wearing either a new respiratory mask or a standard model, while average airborne concentration of hydrogen sulfide was measured in parts per million. Afterwards, participants completed a cognitive performance test, scored on a 0-100 scale where higher scores reflect higher cognitive performance. Using Null Hypothesis Significance Testing (NHST) and Bayesian hierarchical linear regression, I estimate and compare the following two models:

Model 1 (M1): The first model includes the main effects of mask type and exposure level, assuming that the fixed effects are additive and linear.

Model 2 (M2): The second model expands upon M1 by including an interaction term between exposure level and mask type, allowing the effect of exposure on performance to vary by mask type.

II. Background

I provide descriptive statistics for mask type in Table 1 and exposure levels in Table 2. Exposure level is roughly uniformly distributed while performance seems to be bimodal, with peaks at 65 and 85; I hypothesize that these two peaks correspond to the two mask types based on the means provided in Table 1. Linearity was visually verified with scatterplots and boxplots, and independence was similarly checked and satisfied, though there also appeared to be a funnel-shaped pattern, where cognitive performance is more variable at higher levels of exposure. This might indicate a ceiling effect for cognitive performance at lower exposure levels representing a maximal performance level; conversely, the higher variability observed at higher exposure levels could indicate individual sensitivities to H2S exposure. Other assumptions were checked and reported with method of statistical inference. Exposure level was grand-mean centered. Notably, fewer individuals were given the new mask compared to the standard, which may influence later standard errors.

Table 1.

Descriptive Statistics for Cognitive Performance by Mask Type (n = 200)

Mask Type	*n*	*M*	*SD*	*Med*	*Q1*	*Q3*	*Min*	*Max*
New	84	84.26	5.70	83.70	79.90	88.05	69.57	95.28

Standard	116	70.03	12.41	67.55	60.75	79.35	43.80	96.45

Table 2.
Descriptive Statistics for Exposure Levels to Hydrogen Sulfide in ppm (n = 200)

n	*M*	*SD*	*Med*	*Q1*	*Q3*	*Min*	*Max*
200	26.11	14.60	27.77	12.92	38.04	0.012	49.44

NHST Analysis

Residuals were normal and centered at zero for both models. Model 1 satisfied homoscedasticity by mask type and for exposure level, but overall residuals were homoscedastic and showed a mountain-shaped pattern, which could mean the model is missing a part of the relationship of mask type with performance. For Model 2, the previous mountain-shaped residual pattern is less pronounced; instead, residuals are more spread out at high and low predicted values of cognitive performance.

Alpha was set to 0.05 for both models, reflecting a 5% tolerance rate for false positives. The null hypothesis was that the slope of each predictor was zero (that is, that the empty linear model performs better than M1 or M2) while the alternative hypothesis was that the predictor slopes were non-zero. I report 95% confidence intervals for all results, which may be interpreted as follows: if the population was sampled repeatedly and a linear model with 95% confidence intervals for each slope was constructed for each sample, 95% of constructed confidence intervals would contain the true parameter value.

Results for both models are reported in Table 3. Model 1 [$F(2,197) = 293.3$, $p < 0.01$] indicated that exposure levels to H2S and mask type as predictors of cognitive performance were significant and explained about 75% of the total variation [$R^2 = 0.746$]. The main effect of exposure level was a significant predictor of performance [$p < 0.01$] and uniquely explained 42% of variance in cognitive performance [$\eta^2 = 0.423$]; for every 1ppm increase in H2S concentration, post-exposure cognitive performance was expected to decline by 0.55 points, holding mask type constant. Mask type was also found to be a significant predictor of performance [$p < 0.001$] with a large effect size [*Cohen's d* = 2.43], as workers wearing the standard respiratory mask performed 15.20 points worse than workers wearing the new mask, on average.

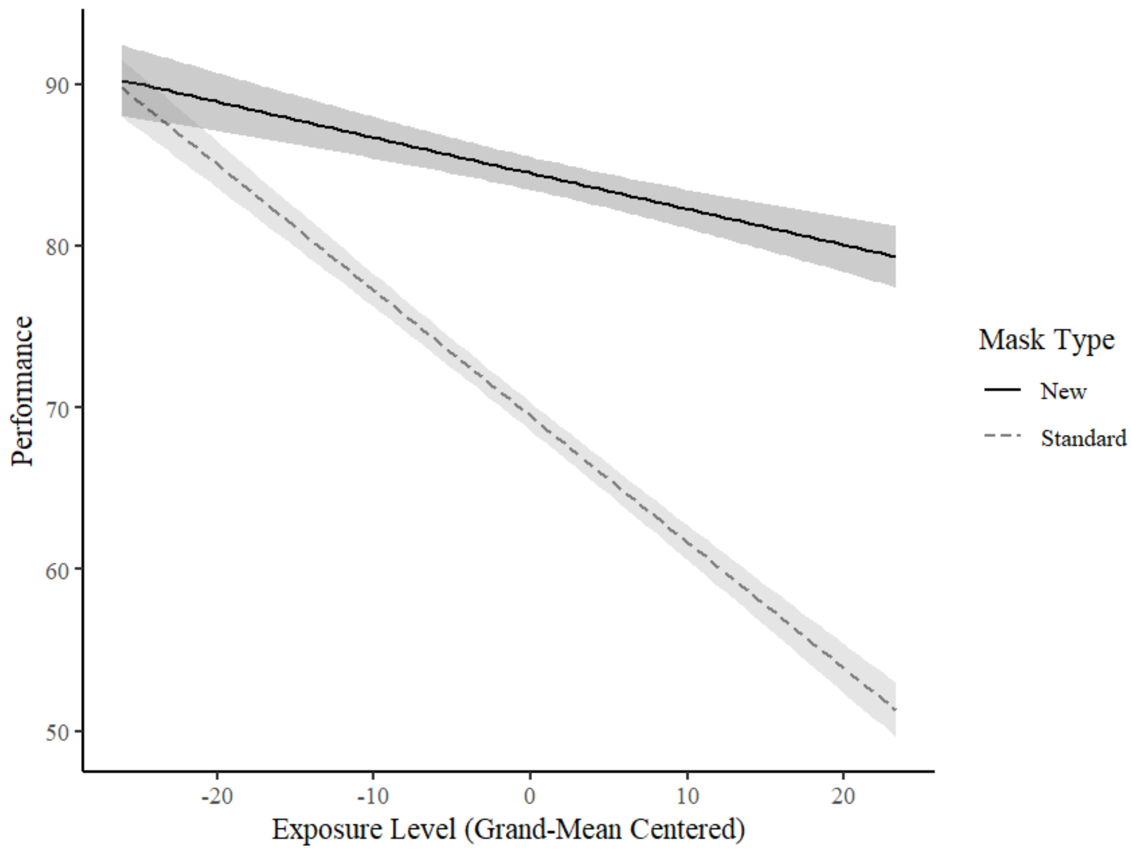
The interaction model suggested that the inclusion of the interaction effect explained significantly more variance in cognitive performance than the main effect model [$F(1,196) = 140.94$, $p < 0.01$]. The interaction effect between mask type and exposure level was significant, such that the negative association between exposure level and cognitive performance was weaker for individuals wearing the new mask [$b = -0.558$, $SE = 0.047$, $CI(-0.650, -0.465)$, $p < 0.01$]. Figure 1 shows a visual representation of the interaction.

Probing of simple slopes revealed the following: exposure level was significantly associated with lower cognitive performance for the standard mask type [$b = -0.779$, $SE = 0.030$, $p < 0.001$]. In particular, for workers wearing the standard mask, every 1ppm increase in H2S was associated with a 0.78 point decrease in cognitive performance. While exposure level was similarly associated with lower cognitive performance for the new mask type, the association was weaker [$b = -0.221$, $SE = 0.036$, $p < 0.001$], indicating that the newer masks reduce the impact of exposure to H2S on cognitive performance; specifically, every 1ppm increase in H2S exposure was associated with a 0.22 point decrease in cognitive performance for those wearing the new mask.

To further compare the models, 10-fold cross validation was performed using the *caret* package to estimate the root mean squared error (RMSE) and mean absolute error (MAE) for each model. The mean absolute error provides the average magnitude of prediction error, while the root mean squared error provides more weight to larger errors, indicating the likelihood of the model to make large mistakes. Model 1 had an RMSE of 6.20 ($SE = 0.93$) and an MAE of 5.13 ($SE = 0.94$) , while Model 2 had an RMSE of 4.78 ($SE = 0.64$) and an MAE of 3.82 ($SE = 0.62$), showing a moderate improvement in predictive accuracy. Additionally, the decrease in standard errors shows that the interaction model is both more accurate on average and more consistent across folds. In practice, a MAE of 3.82 is around 7% of the observed range of performance scores (42.8 - 96.5) which suggests a moderate level of predictive accuracy.

Figure 1.

Interaction between Exposure Level and Mask Type (n = 200).



Note: Shaded region reflects 95% Confidence Interval. Exposure level is grand-mean centered.

Table 3.
NHST Hierarchical Regression Results predicting Cognitive Performance (n = 200 individuals)

	Dependent variable:	
	Cognitive Performance	
	(1) Main Effects Model	(2) Interaction Model
Exposure Levels (ppm)	-0.550^{***}	-0.221^{***}
	(0.030)	(0.036)
Mask Type (Standard)	-15.205^{***}	-0.472
	(0.892)	(1.416)

Exposure Level * Mask Type (Standard)		-0.558^{***}
		(0.047)
Intercept	99.186^{***}	90.253^{***}
	(1.064)	(1.108)
Observations	200	200
R^2	0.749	0.854
Adjusted R^2	0.746	0.852
Residual Std. Error	6.213 (df = 197)	4.751 (df = 196)
F Statistic	293.344^{***} (df = 2; 197)	381.462^{***} (df = 3; 196)
Note: Standard errors reported in parentheses.		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Bayesian Analysis

Using the default priors in the *brms* package (v 2.20.4) for all parameters, I fit both models with the default settings (4 chains of 2,000 iterations each, with 1,000 dedicated to warming up). I selected a skewed normal distribution for performance scores based on its distribution, showed in Figure 2. Although the distribution of cognitive performance appears bimodal, I hypothesized that this was due to the difference in mean cognitive performance for each mask type; after controlling for mask type with the moderating term, both types showed a slightly skewed distribution.

Convergence was checked and satisfied using R-hats, effective sample sizes, and visual inspection of trace plots. Model residuals were checked for both models and found to be homoscedastic and nearly normal. Notably, Model 1 did not sufficiently satisfy predictive accuracy, whereas Model 2 did; see Figure 3 for a visual representation. This indicates that the skew-normal likelihood was better able to capture the bimodal trend of cognitive performance with the inclusion of the moderating term.

I report Bayesian 95% high density intervals (*HDI*) for all results, which may be interpreted as a 95% probability that the true parameter falls within the given range. I also report probability of direction (*pd*) which is the probability that the effect is non-zero, and regions of practical equivalence (*ROPE*) which indicate the proportion of the posterior distribution that falls within a range of values considered practically equivalent to zero.

Both the main effect only model [*BF* = 6.69*e55 to 1] and the interaction model [*BF* = 2.26*e77 to 1] were highly favored over the null. For the main effects model, the estimated performance score for a worker with average exposure levels wearing the new mask type was 98.67 [*HDI*

(96.79, 100.48), *SD* = 0.94], with performance credibly decreasing with higher exposure to H2S [*b* = -0.54, *HDI*(-0.60, -0.48), *SD* = 0.03], holding mask type constant. Additionally, workers wearing standard masks scored credibly lower on the cognitive exam, on average [*b* = -14.71, *SE* = 0.91, *CI*(-16.47, -12.91)], holding exposure levels constant; mask type was found to have a large effect [*Cohen's d* = 2.32] on cognitive performance, and to be practically credible [0% of *HDI* outside of *ROPE* range, *ROPE* range = -1.23, + 1.23, performance SD = 12.33], whereas exposure level was not practically significant [100% of *HDI* in *ROPE* range]. These results were using the default ROPE range, which may be too wide in this case as a variation of even 1ppm in H2S concentration can be significant – see conclusion for further analysis.

The regression model incorporating the moderation term suggested that the moderating model was highly favored over the main effect only model [*BF* = 3.67*e21 to 1]. The interaction between exposure levels and mask type was credible, such that the association between exposure levels and cognitive performance was credibly more negative for the standard mask than the newer model [*b* = -0.55, *SE* = 0.05, *HDI*(-0.64, -0.46)]. See Figure 4 for a visual representation of the interaction. Probing of simple slopes revealed the following: while exposure levels were credibly associated with cognitive performance for both standard [*b* = -0.78, *SE* = 0.030, *HDI*(-0.84, -0.72)] and new masks [*b* = -0.23, *SE* = 0.04, *HDI*(-0.30, -0.16)], those wearing standard masks were generally more sensitive to changes in H2S exposure. Specifically, for every 1ppm increase in H2S exposure, cognitive performance was expected to decrease linearly by 0.78 points for workers wearing the standard masks and by 0.23 points for workers wearing the newer masks. Notably, 100% of the HDI for the moderation term fell within the ROPE range, meaning that these predictors may not be practically significant.

Although the effect of exposure level on cognitive performance was smaller when participants wore new masks (*b = -0.22*) than when they wore standard masks (*b = -0.78*), both effects were credibly negative (*pd = 100%*). This comparison suggests that the new mask substantially attenuates the harmful cognitive impact of H_2S exposure. Even though the smaller effect falls within the default ROPE, the contrast between slopes illustrates a meaningful protective benefit of the new mask, reducing the performance drop by approximately two-thirds.

While the main effects model explained 74% of total variance [R^2 = 0.738], the moderation model explained significantly more variance [R^2 = 0.852], suggesting that the moderation term improves overall predictive accuracy for this sample. To further assess model performance, leave-one-out (LOO) cross validation was performed using the *brms* package. All Pareto k-values were less than 0.6, indicating that no individual observations greatly impacted results. The moderation model was moderately favored over the main effects only model, with an expected log predictive difference of 50.8 (SE = 8.7). The ELPD represents the model's predictive accuracy for out-of-sample data, which is high relative to its standard error. This suggests that the model is expected to perform strongly for new data should the study be repeated.

Figure 2.

Distribution of Cognitive Exam Performance (n = 200)

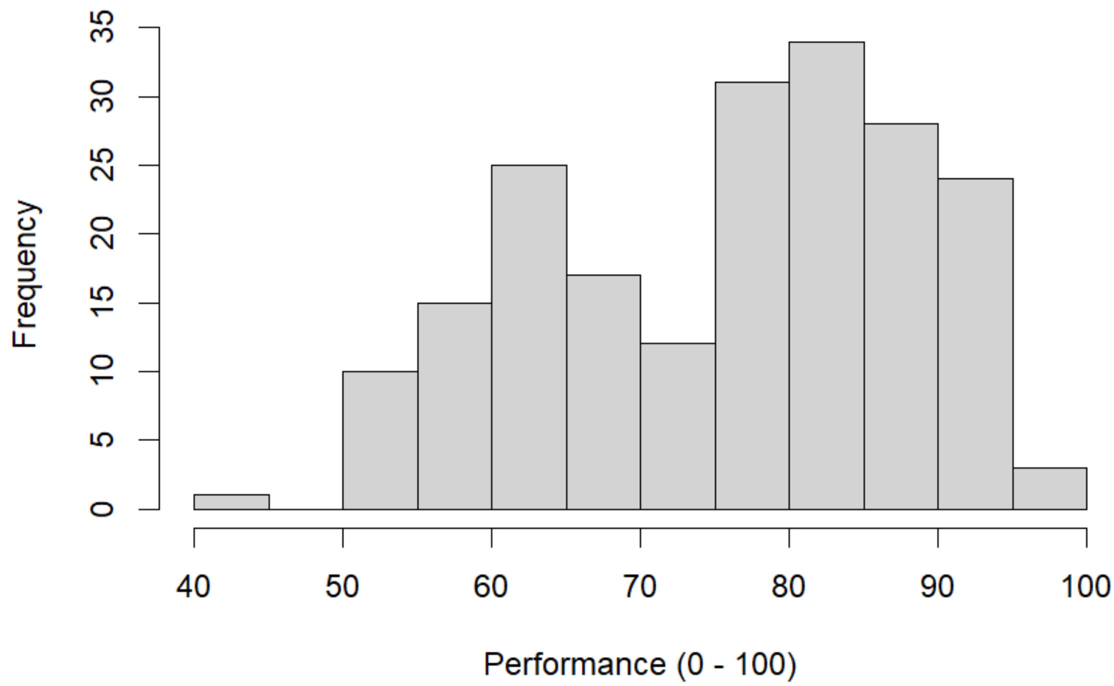
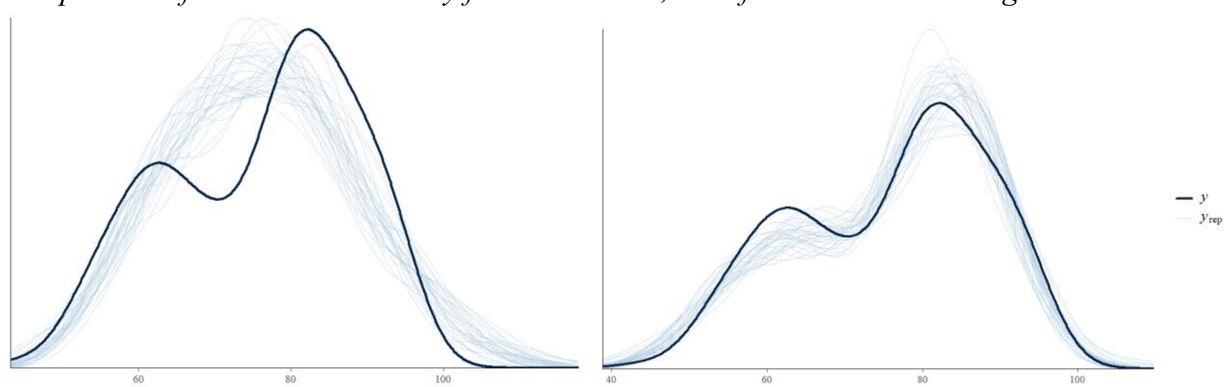
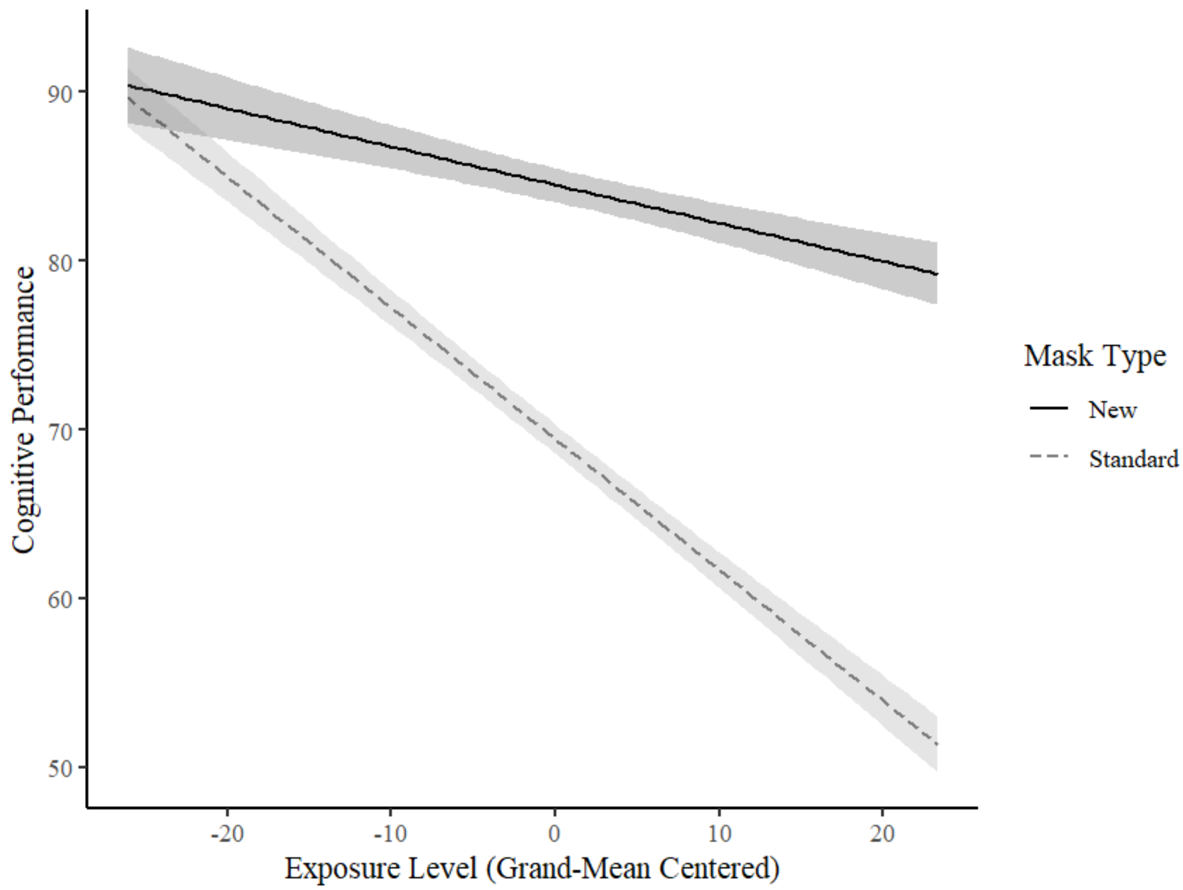


Figure 3.
Comparison of Predictive Accuracy for M1 and M2, both fitted with a skewed gaussian.



Note: The navy line shows the distribution of cognitive performance scores, while the light blue lines are random draws from the predicted posterior distribution, performed with posterior predictive checks in R. The left panel shows the fit for the main effect only model while the right panel shows the fit for the interaction model. Both were fitted with a skewed gaussian distribution.

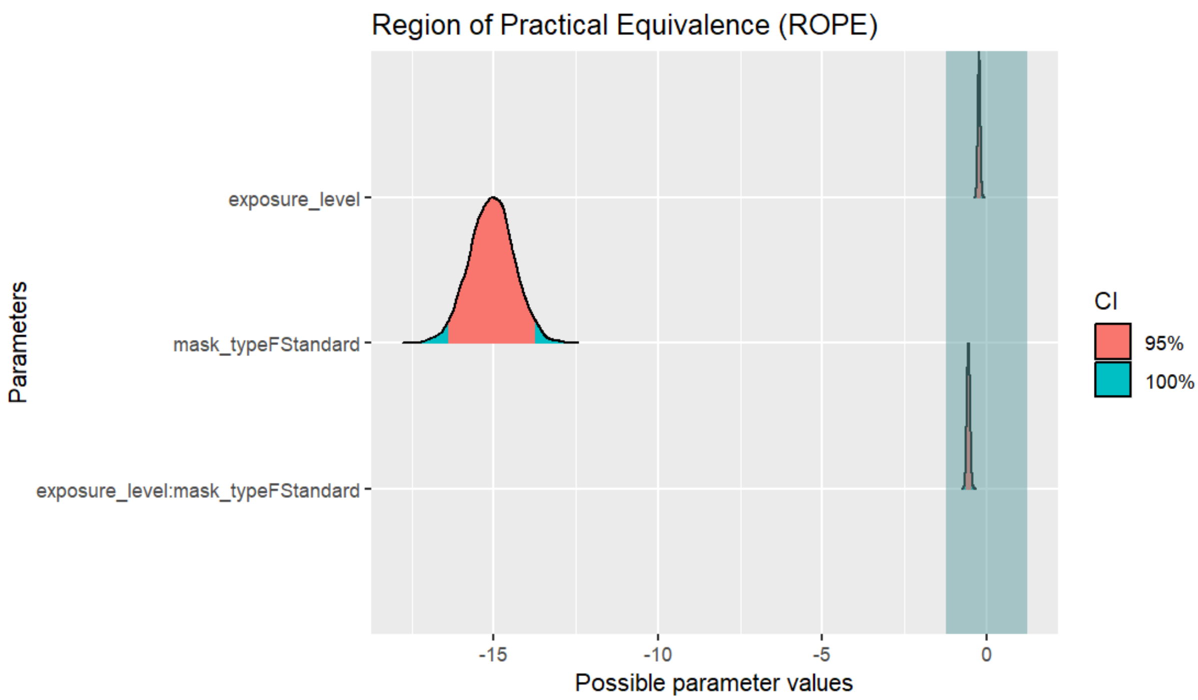
Figure 4.
Interaction between Exposure Level and Mask Type (n = 200)



Note: Shaded region reflects 95% High Density Interval. Exposure level was grand-mean centered.

Figure 5.

Regions of Practical Equivalence for Interaction Model.



Concluding Remarks

I used NHST and Bayesian analytic methods to investigate the moderating effect of mask type on the relationship between exposure levels and cognitive performance. Although exposure levels and cognitive performance had a negative relationship across both mask types, the relationship was stronger for the standard mask type, suggesting that the newer masks reduce sensitivity to H2S exposure. Practically, these newer masks were more effective at protecting workers from the harmful effects of H2S exposure across the board.

For both the NHST and Bayesian models, the interaction model explained more overall variance in cognitive performance. In the NHST analysis, the interaction model showed a moderate improvement from 10-fold cross validation that was found to be practically minimal. Similarly for the Bayesian analysis, the interaction model was greatly favored over the main effect only model using Bayes Factors and leave-one-out cross validation. However, the interaction term was fully within the ROPE range, meaning that the effect was not practically significant.

Regarding the ROPE range, the span of about 2.5 points is feasibly small since the outcome ranges from 0 to 100 points on the cognitive exam; even when considering that all participants received scores between 40 and 97, a 1ppm increase in H2S exposure causing a change of performance of less than 1 point is not practically credible. However, the short-term limits for H2S exposure are around 50ppm, meaning that even a 1ppm change might have significant effects. Regarding the NHST analysis, although the exam has a range of 100 points, a difference of one point may reflect significant changes in memory and attention that should be addressed. Without the background knowledge of the cognitive exam content and safe levels of H2S, determining what effects are practically significant is difficult; while I use the provided results for this analysis, future work should assess what is practically significant in this field.

Overall, while the interaction model is statistically favored over the main effect only model, it seems to have a relatively small effect. Even without the moderating effect, newer masks were associated with higher performance on the cognitive exam, with an average cognitive improvement of 14.71 points (NHST) and 15.21 points (Bayesian), holding exposure levels constant. Considering that this improvement is around 15% of the total range of the outcome, its significance should be noted. In particular, mask type was the only predictor that was fully outside of the ROPE range for both the main effect only model and the interaction model, solidifying its practical and statistical significance. Hence newer masks should be worn in high-exposure work environments to minimize impacts on cognitive ability.