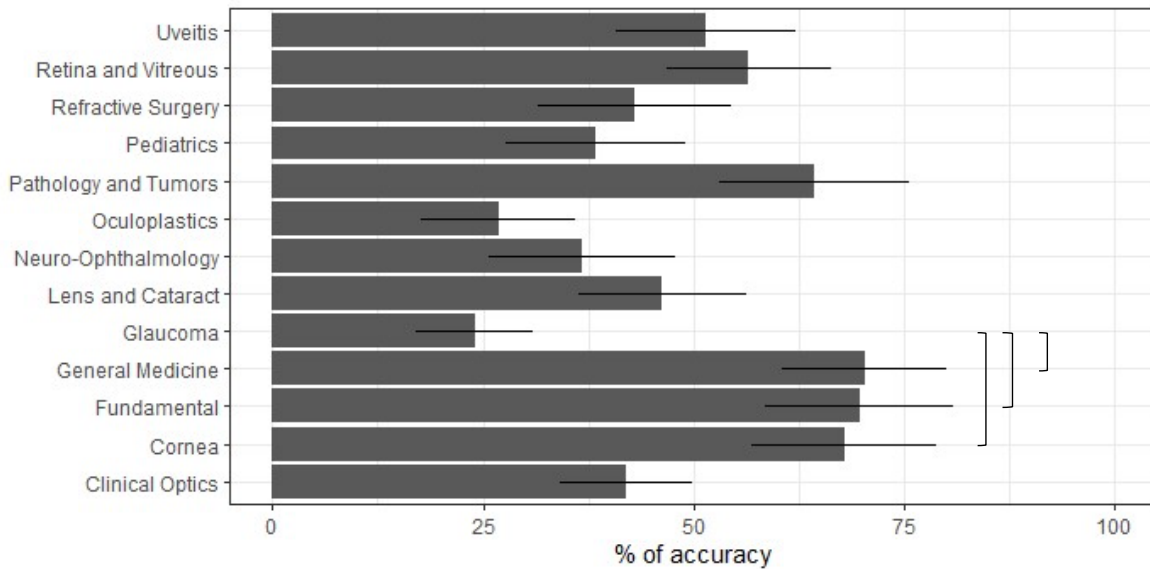


Supplemental Figure 8: Post hoc analysis using Tukey's test to isolate the effect of exam section and difficulty index using the OphthoQuestions questions for ChatGPT Plus. (A)

Bar plot of the percentage of accuracy by exam section. The grey square brackets identify the significant differences. The significant contrasts were the following: Cornea – Glaucoma (estimate 0.438, $p=0.046$), Fundamentals – Glaucoma (estimate 0.457, $p=0.034$) and General Medicine – Glaucoma (estimate 0.463, $p=0.009$) (B) Predicted percentage of accuracy by difficulty index. Accuracy increased with increasing difficulty index (easier questions).

(A)



(B)

