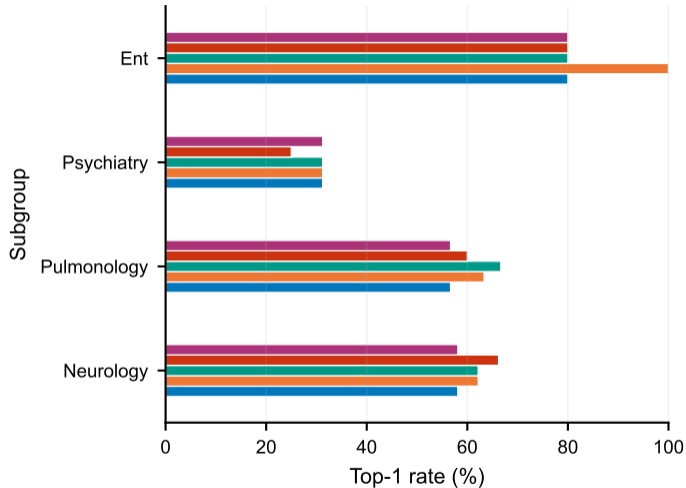
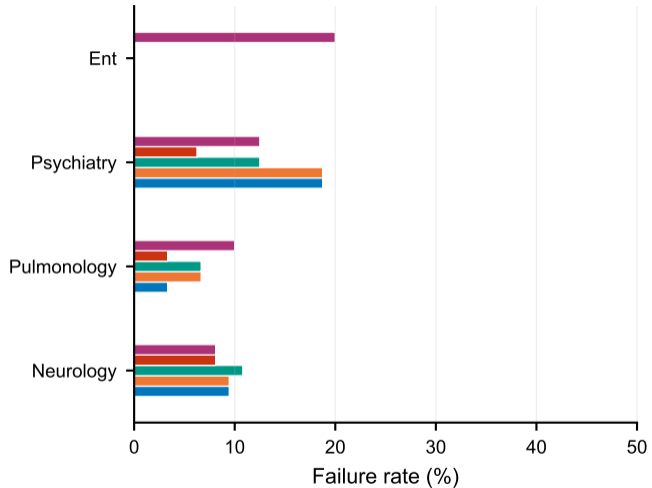


a) Top-1 rate**b) Failure rate**

Grok Claude DeepSeek GPT Gemini