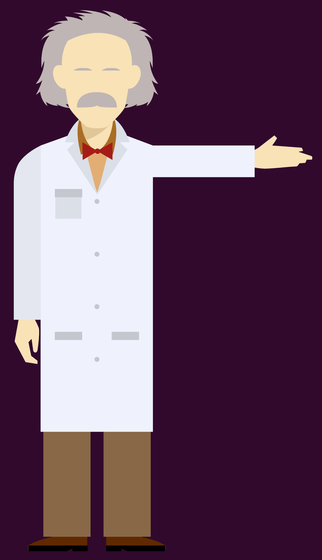
PRESENTER
Richard Evans



LLMs outperform human experts generating priors for Bayesian Glioblastoma Survival Analyses

Background

• Small, underpowered trials make frequentist inference unreliable.
• Bayesian analyses incorporate *prior information* which improves inference but sometimes uses subject matter **expert** opinion.

• **Experts** are often overconfident, overemphasize memorable research or patients, are anchored in their initial training, and find it hard to translate their knowledge into density parameters.
• It is exceptionally time consuming to elicit information from medical experts.

Solution

• LLM priors replace expert opinion priors

Methods

• Bayesian Cox proportional-hazards model
• Six priors generated by LLMs and validated by a Radiation Oncologist

Prompt

I described the disease and the patient characteristics and stated the two treatments and the variable types (e.g., censoring, time-to-event)

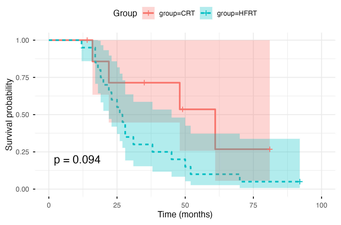
I stepped the LLMs through their responses:

1. A review of the information on HFRT and CRT trials in the glioblastoma literature.

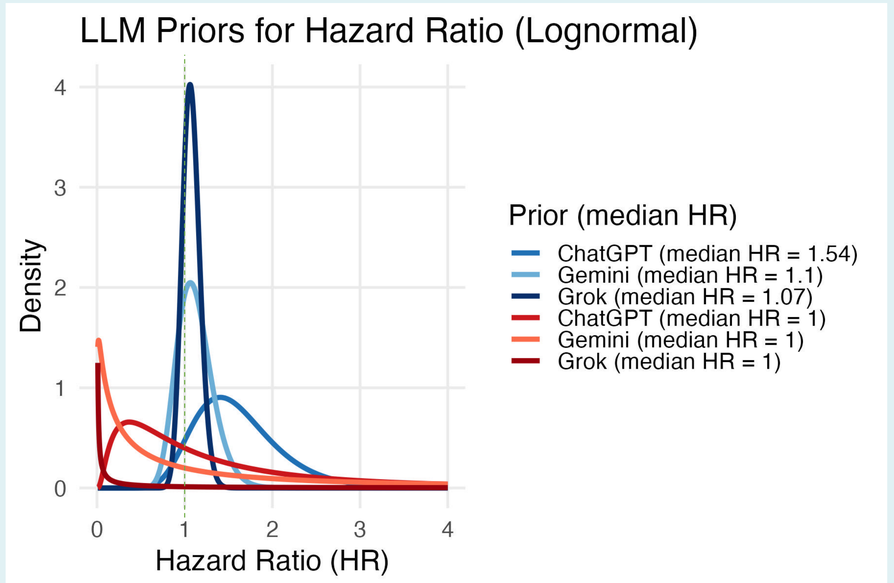
2. An informative log-normal prior for the hazard ratio

3. A justification of the informative prior using the information on HFRT and CRT trials in the glioblastoma literature.

4. A non-informative log-normal prior for the hazard ratio



• 28 Adult GBM patients: 8 CRT, 20 HFRT
• Two variables: censoring and time-to-death in months



Prior	Pr(HR > 1)	Median HR	2.5% CI	97.5% CI
ChatGPT (noninf)	0.973	2.748	0.987	9.477
ChatGPT	0.976	2.712	1.006	9.555
Gemini	0.974	2.713	0.995	10.036
Grok	0.975	2.713	1.007	9.547
None (Classic KM)		2.457	0.828	7.291

PrePrint

MASONIC CANCER CENTER
UNIVERSITY OF MINNESOTA
MEDICAL COLLEGE OF WISCONSIN

Rich Evans, Max Felland, Susanna Evans, Joseph Moore, Lindsey Sloan