Assessing the Quality of Patient-Facing Dietary Plans Post-Bariatric Surgery Recommended by Large Language Models

Melanie Beceiro, BS; Andres Wong, BS; Georgeta Vaidean, MD MPH PhD

Herbert Wertheim College of Medicine, Florida International University, Miami, FL, USA,

INTRODUCTION

- Bariatric surgery patients experience changes in gastric capacity, nutrient absorption, and digestive enzyme secretion, requiring optimal lifestyle adaptations.
- Increasing number of patients use LLMs for health information

AIM

- To assess the quality of dietary plans created by LLM

METHODS

- Prompt development
- ChatGPT-4o → 25 Clinical vignettes
- Gemini Basic & ChatGPT-4o→ Dietary plans creation
- Human/ researchers & ChatGPT o3 → Evaluation rubric
- Rubrics 1 (min) - 5 (max)
- ChatGPT o3 → blinded assessor of dietary plans quality
- Claude Sonnet 4→ statistical analysis
- SAS 9.4 → human verification of statistical analysis
- Statistical analysis: Mann-Whitney U test (Wilcoxon rank-sum test) and chi square test for proportions

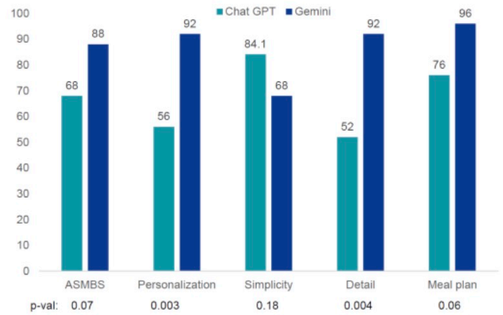
RESULTS

Quality measures for multiple domains by LLM

	ChatGPT			Gemini			P-val*
	Mean (SD)	Median	Mode	Mean (SD)	Median	Mode	
ASMBS Guides	4.68 (0.48)	5	5	4.88 (0.33)	5	5	0.07
Personalization	4.64 (0.49)	5	5	4.92 (0.28)	5	5	0.005
Simplicity	4.84 (0.37)	5	5	4.80 (0.41)	5	5	0.18
Detail	4.36 (0.70)	4	4	4.92 (0.28)	5	5	0.002
Meals+Portions	4.76 (0.44)	5	5	4.96 (0.20)	5	5	0.004
Total Score	23.2 (1.78)	24	25	24.48 (0.71)	25	25	0.09

* p-values for the Mann-Whitney U test (Wilcoxon rank-sum test)

Percent vignettes with maximum score



* p-values for the chi square test

STRENGTHS:

- 1st study comparing LLM's diet plans
- Realistic and diverse clinical scenarios

LIMITATIONS:

- Rubric still in development phase
- Limited statistical power/sample size

CONCLUSIONS and FUTURE studies

- LLMs generated personalized, guideline-informed dietary plans for post-bariatric patients
- Gemini outperformed ChatGPT for personalization and detail
- Future studies should refine the rubrics and explore implementation in real life settings