Comparison of AI-Assisted and Traditional Analytic Workflows in Health Economics and Outcomes Research Using Health Survey Data

MSR25

Alysha M. McGovern, MBA[1], Joseph Yeb, BS[1], Harshini Mashruwala, MS[1], Praveen Kumar Potukuchi, PhD[1], **Amy Bolton, MPH, MS[1]**, Hamid Zarei, PhD[2]
[1]Boston Scientific, Marlborough, MA, USA; [2]University of Louisville, Louisville, KY, USA

BACKGROUND

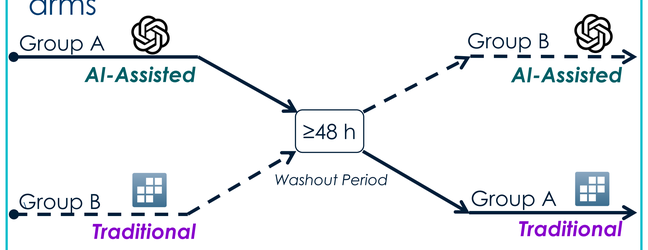
Artificial Intelligence (AI) is increasingly used in analytic workflows; however, direct comparisons with human-only analyses are limited.

OBJECTIVE

To compare the efficiency and accuracy of AI-assisted and traditional analytic workflows using publicly available health survey data.

METHODS

- **Design**: Two-sequence crossover design with ≥48-hour washout period between arms



- **Data Source**: 5% sample of the 2024 Behavioral Risk Factor Surveillance System survey data
- **Comparison**: Traditional analytic workflow (Stata 19; StataCorp) vs. AI-assisted analytic workflow (ChatGPT GPT-5; OpenAI)
- **Outcomes**:
  - *Efficiency*: Completion time in minutes
  - *Accuracy*: Concordance with blinded validator results (n=2 validators)
- **Workflow Standardization:**
  - *Traditional*: Limited modification Stata do file template
  - *AI-Assisted*: Standard prompting template (TRACI [Task-Role-Audience-Create-Intent] framework)
- **Analysts**: n=4

RESULTS

AI-assisted analyses demonstrated comparable speed and similar accuracy to traditional workflows.

Differences primarily reflected specification rules rather than computational errors.

Efficiency

Traditional	AI-Assisted
Mean Completion Time	
47.8 minutes	**50.3** minutes
Analyst-Level Comparison	
Faster for **2 analysts**	Faster for **2 analysts**

Accuracy

Traditional	AI-Assisted	
Descriptive Output		
99.3%	Raw **75.0%**	Adjusted* **99.2%**
Chi-Square Results		
87.5%	**93.8%**	

* Adjusted for small percentage denominator-related discrepancies due to non-exclusion of refused/unknown/missing responses (median absolute difference = 0.2%).

LIMITATIONS

- The small sample size (n=4 analysts) precludes formal statistical inference.
- Findings reflect performance under structured study conditions and may not generalize to less structured real-world analytic environments.
- Accuracy was defined by agreement with pre-generated reference outputs, which may not capture all analytical limitations.
- Completion times may be confounded by differences in analyst experience with Stata and ChatGPT.

CONCLUSIONS

- In this pilot comparison, AI-assisted analyses showed similar completion times and broadly comparable accuracy to traditional workflows.
- Most discrepancies reflected denominator-handling rules versus computational errors.
- Explicit analytic specifications and independent validation remain important for AI-assisted workflows.

DISCLOSURES

This study was funded by Boston Scientific. Alysha M. McGovern, Harshini Mashruwala, Praveen Kumar Potukuchi, and Amy Bolton are full-time employees of, and shareholders in, Boston Scientific. Joseph Yeb was a paid intern with Boston Scientific at the time of this research. Hamid Zarei has no conflicts of interest to disclose.

Poster presented by **Amy Bolton (amy.bolton@bsci.com)** at ISPOR 2026, May 17-20 in Philadelphia, PA