

Evaluating Agentic AI for Autonomous Hypothesis Generation in Clinical Datasets

AUTHORS & INSTITUTION

Abdullah Hamdan

School of Medicine, The University of Jordan

Ola Hamdan

Jordan Center for Disease Control (JCDC)

An average hospital produces about 50 petabytes of data per year, and 97% of that data goes unused.

But still unexpectedly, in clinical research, data collection and patient recruitment remain among the biggest causes of research delays.



INTRODUCTION

In this study, we evaluated whether an autonomous AI research agent could...

Although healthcare systems possess substantial data.

Researchers often have to start from the beginning.

This without mentioning the time needed to clean this data

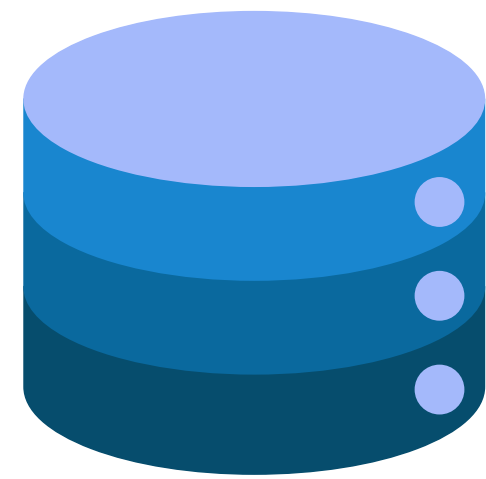
At the same time, researchers may still struggle to identify meaningful research ideas or generate testable hypotheses.



1
Independently examine a deidentified clinical dataset

METHODOLOGY

We used the Brain Mets Lung MRI Path Segs dataset from The Cancer Imaging Archive.



We then submitted these data to ChatGPT in agent mode.



2
Identify a significant association or generate testable hypothesis

To automatically identify patterns in the data and formulate clinically relevant, statistically significant hypotheses.

Utilizing Structured Prompt Engineering and Guardrails

3
Choose a suitable statistical test and carry out the analysis.

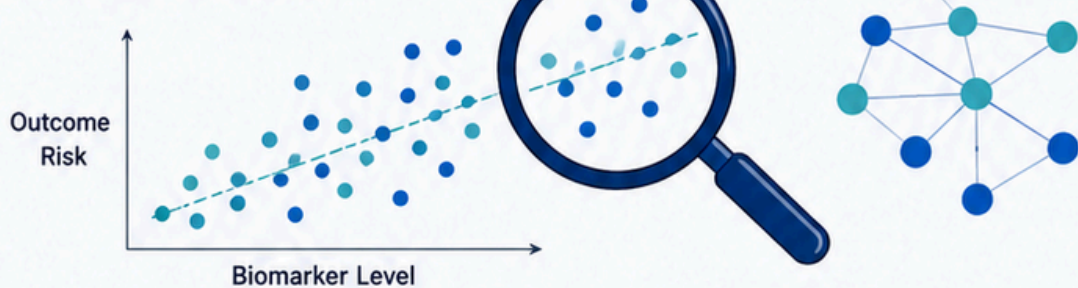
AI Research Agent Workflow

1 Data cleaned & structured

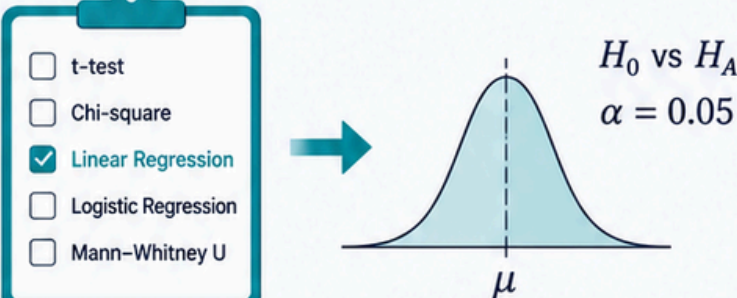
Consistent rows and columns

ID	Age	Sex	Lab A	Lab B	Outcome	...
1001	62	F	5.2	120	0	...
1002	55	M	6.1	98	1	...
1003	70	F	4.8	110	0	...
1004	47	M	5.9	130	1	...
...

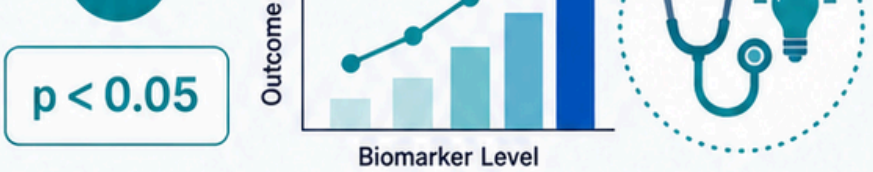
2 Association identified



3 Appropriate statistical test chosen



4 Accurate, clinically significant findings



From raw clinical data to validated hypothesis

RESULTS

As an example of successful autonomous hypothesis generation, the workflow identified an association between the dataset field labeled PD L1 status and dominant lesion size, yielding a statistically significant result in the initial exploratory analysis.

It discovered PD-L1 positivity was significantly associated with smaller lesion size, not higher lesion size

✓ We manually reviewed the agent's findings and confirmed the association was statistically significant.

What is new about this?

This hypothesis arose from the AI system's ability to autonomously detect patterns in the dataset, marking its first study validating its capability to generate hypotheses independently.