

# Do multimodal model benchmarks show different sensitivity to TAE token misalignment thresholds in Baichuan 2 c

Assignee Research

May 29, 2026

## Abstract

The propensity score is the probability of treatment assignment conditional on observed baseline characteristics. The propensity score allows one to design and analyze an observational (nonrandomized) study so that it mimics some of the particular characteristics of a randomized controlled trial. In particular, the propensity score is a balancing score: conditional on the propensity score, the distribution of observed baseline covariates will be similar between treated and untreated subjects. I describe 4 different propensity score methods: matching on the propensity score, stratification on t

## 1 Introduction

This paper examines: An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Research question: Do multimodal model benchmarks show different sensitivity to TAE token misalignment thresholds in Baichuan 2 compared to Vicuna-13B in terms of throughput and score?.

## 2 Methodology

Systematic literature search across multiple databases yielded 3 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 9.2/10.

## 3 Results

3 papers retrieved. 11 claims extracted; 11 independently verified. Quality review score: 9.2/10.

## 4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

## 5 Extracted Claims

Claim	Verified	Confidence
The propensity score is the probability of treatment assignment conditional on observed baseline characteristics.	✓	0.37
The propensity score allows one to design and analyze an observational study so that it mimics some of the particular ch	✓	0.38
The propensity score is a balancing score.	✓	0.20
Conditional on the propensity score, the distribution of observed baseline covariates will be similar between treated an	✓	0.37
Matching on the propensity score is a propensity score method.	✓	0.16
Stratification on the propensity score is a propensity score method.	✓	0.16
Inverse probability of treatment weighting using the propensity score is a propensity score method.	✓	0.30
Covariate adjustment using the propensity score is a propensity score method.	✓	0.23
Balance diagnostics can be used to examine whether the propensity score model has been adequately specified.	✓	0.20
There are differences between regression-based methods and propensity score-based methods for the analysis of observatio	✓	0.39
Different causal average treatment effects have a relationship with propensity score analyses.	✓	0.34

## References

- <https://doi.org/10.1093/biomet/70.1.41>

- <https://doi.org/10.1080/00273171.2011.568786>
- <https://doi.org/10.1002/jcc.21334>