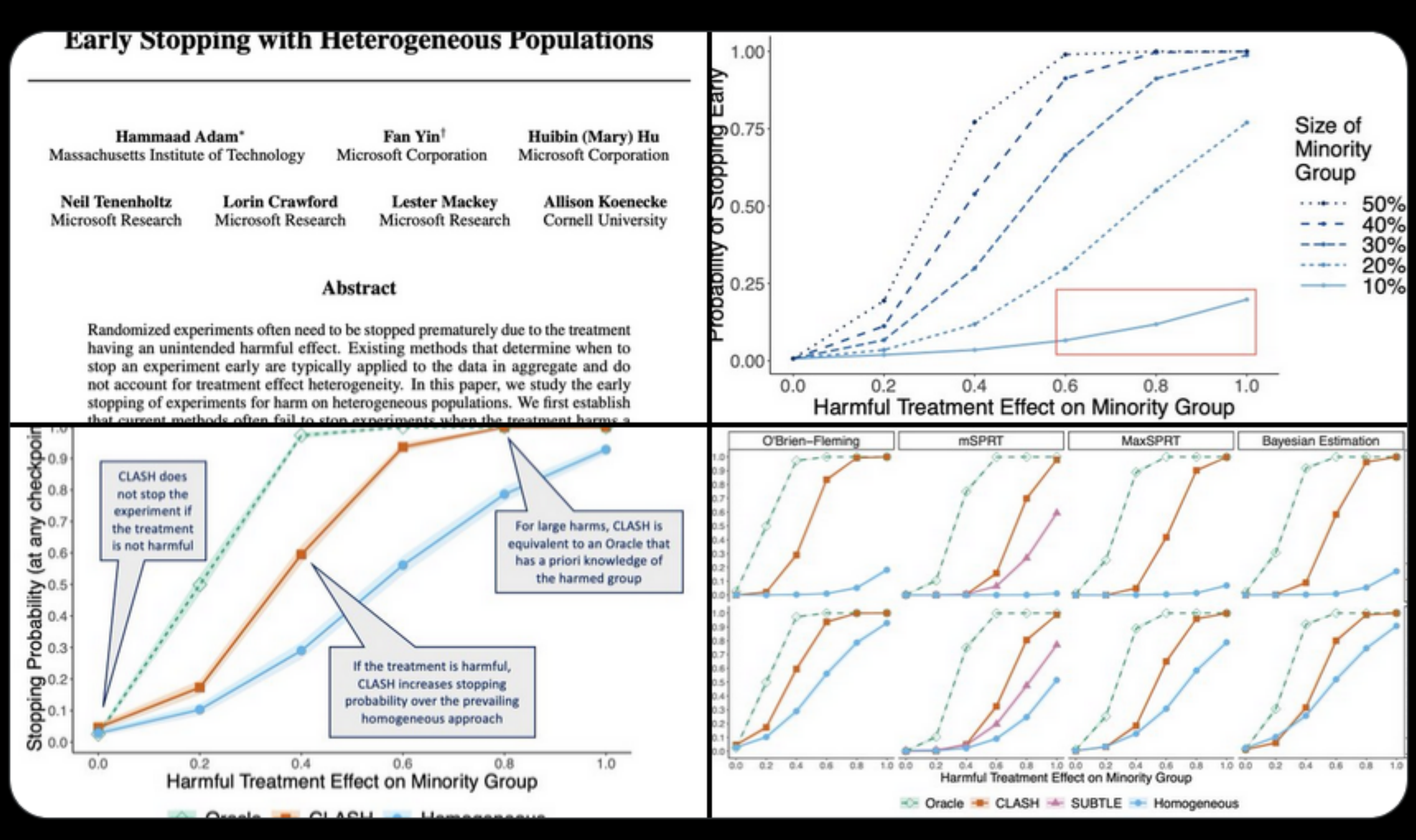
Hammaad Adam
@hammaadadam1

Clinical trials and A/B tests often need to be stopped early if the treatment shows signs of harm. Can we detect harm if it only affects a minority group of participants?

Our #NeurIPS2023 (spotlight) paper investigates!

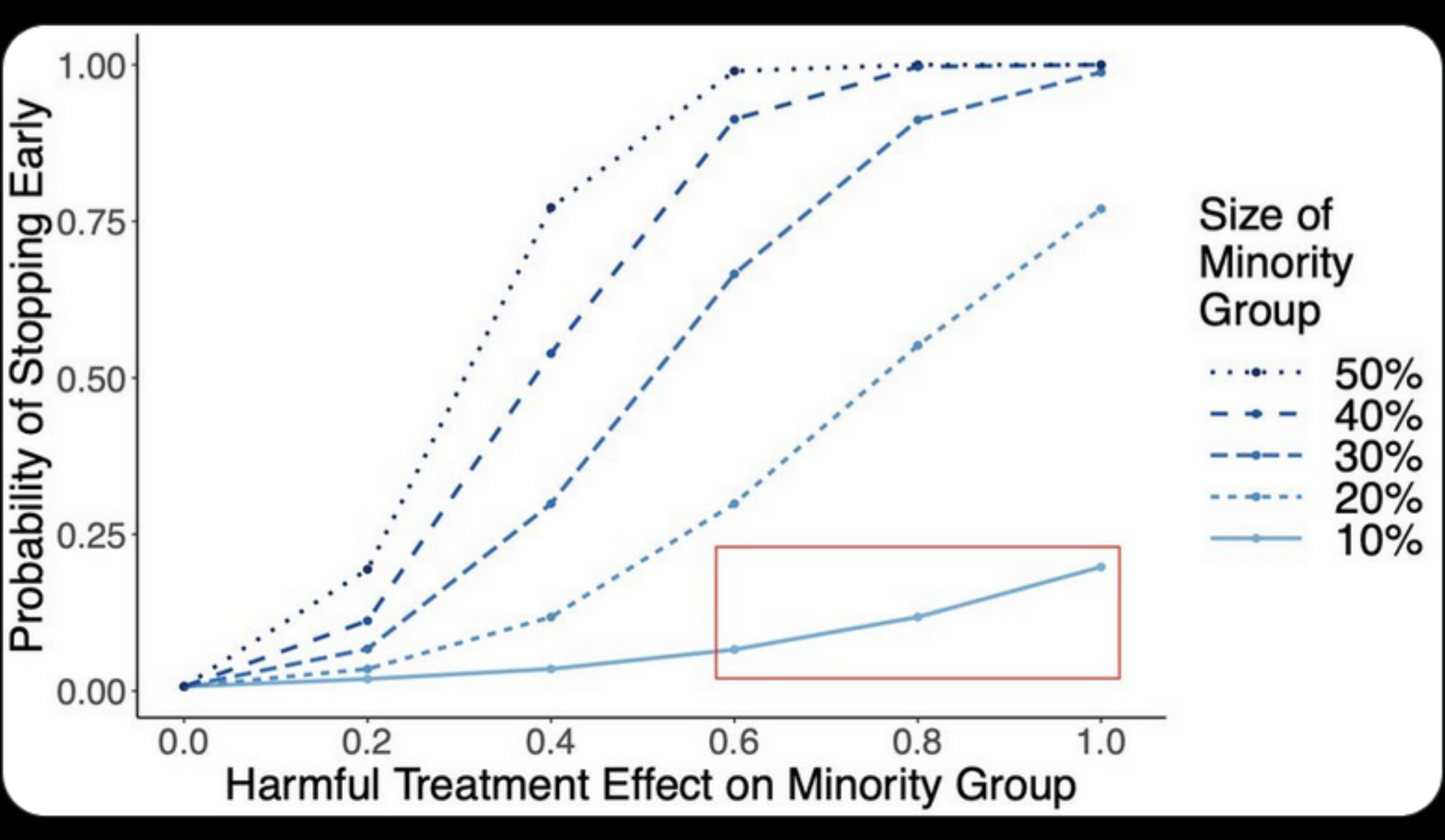
arxiv.org/pdf/2306.11839…; 🧵👇 (1/11)



Hammaad Adam @hammaadadam1 · Nov 10, 2023
Randomized experiments are often stopped prematurely if the treatment shows evidence of harm. For example, if early data from an RCT suggests that a drug increases mortality risk, the trial is immediately stopped to protect participants. (2/11)

Hammaad Adam @hammaadadam1 · Nov 10, 2023
Harm is often heterogeneous: for example, a drug (e.g. warfarin) may increase the rate of adverse events only in elderly patients. However, "stopping tests"-- the existing methods used to detect harm--are typically applied in aggregate and do not allow for heterogeneity. (3/11)

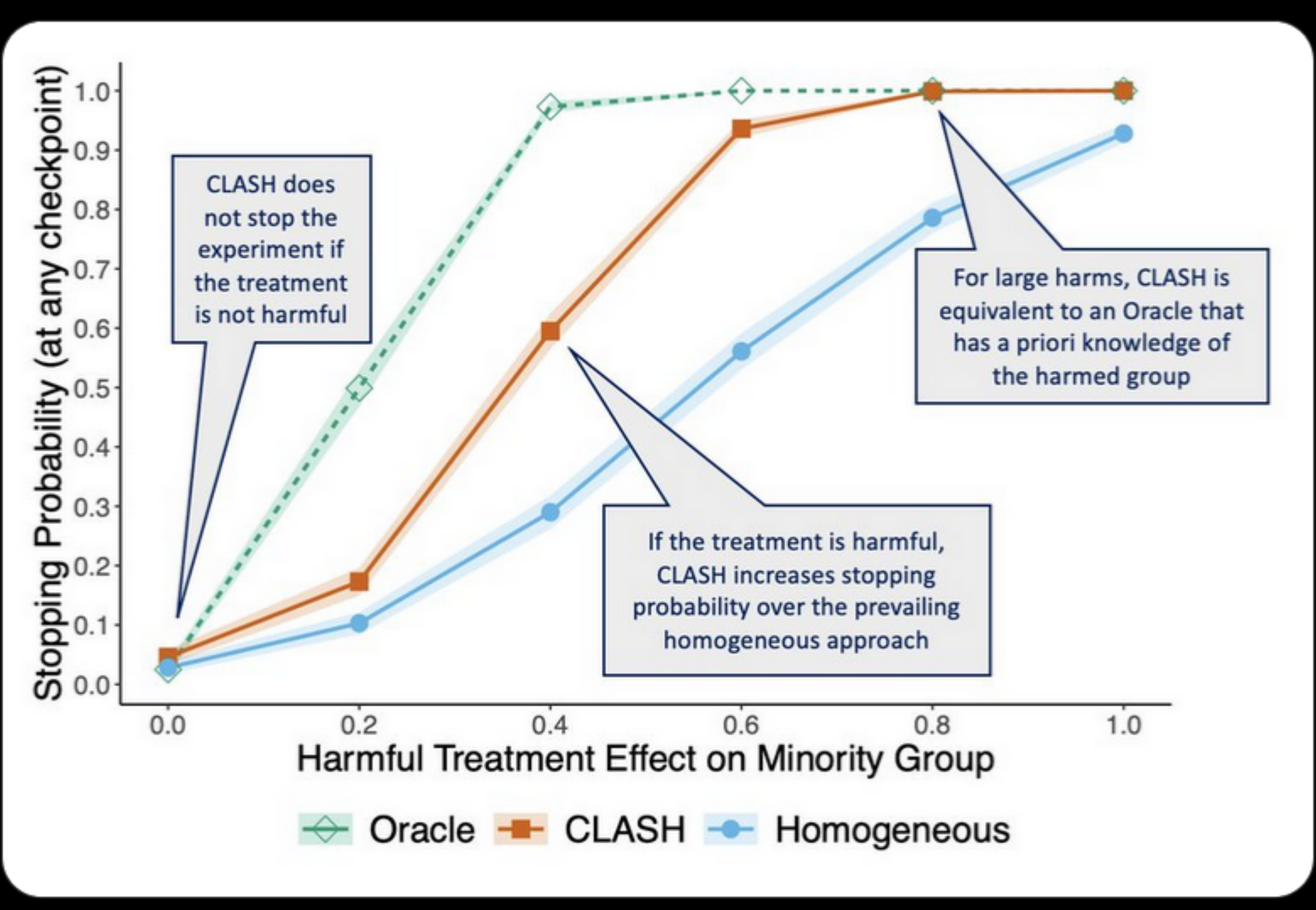
Hammaad Adam @hammaadadam1 · Nov 10, 2023
We find that existing stopping tests (e.g. O'Brien Fleming, mSPRT) often fail to stop experiments if the treatment harms only a minority group of participants. In fact, if the treatment harms a 10% minority group, the trial is almost never stopped (red box in plot below). (4/11)



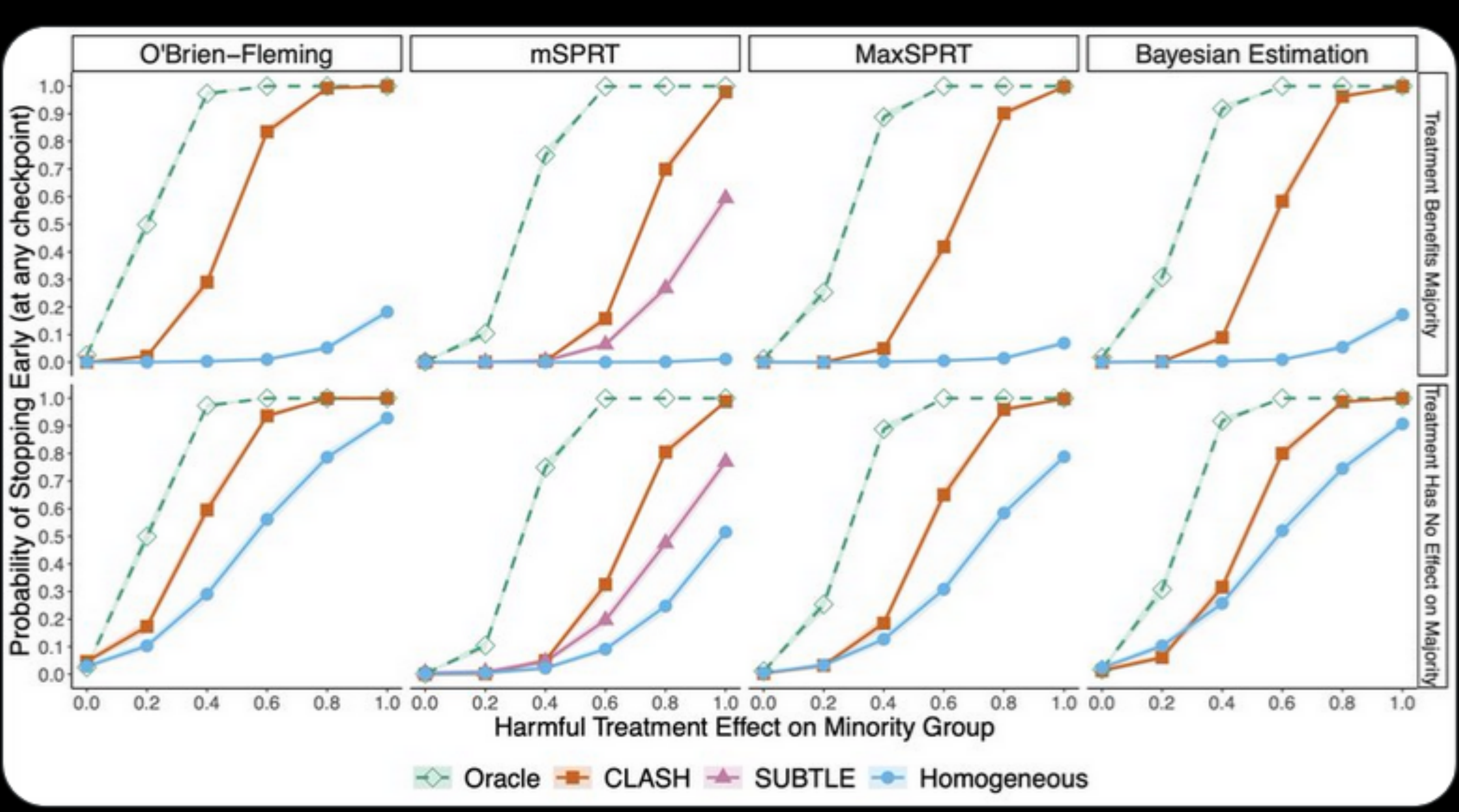
Hammaad Adam @hammaadadam1 · Nov 10, 2023
This is a major issue: not stopping an experiment in which participants are harmed can have serious ethical, legal, and financial repercussions! It is also a key equity concern, since experimental treatments often display heterogeneous harms by age, race, and gender. (5/11)

Hammaad Adam @hammaadadam1 · Nov 10, 2023
In this paper, we develop CLASH, the first broadly-applicable method for heterogeneous early stopping. CLASH uses causal machine learning to infer the probability that a participant is harmed by the treatment, then adapts existing stopping tests to better detect this harm. (6/11)

Hammaad Adam @hammaadadam1 · Nov 10, 2023
We establish (theoretically + empirically) that CLASH is more likely to stop an experiment than existing approaches if the treatment harms a minority group. Crucially, CLASH does not stop unnecessarily: it only stops the experiment if a group of participants is harmed. (7/11)



Hammaad Adam @hammaadadam1 · Nov 10, 2023
CLASH is extremely flexible: it can be used with any existing stopping test (e.g. O'Brien Fleming, mSPRT) and data distribution (e.g. continuous, binary, time-to-event). It is also easy to use and can be implemented in ~25 lines of code (see: github.com/hammaadadam1/c…). (8/11)



Hammaad Adam @hammaadadam1 · Nov 10, 2023
Takeaway: if you are a practitioner running clinical trials or A/B tests and suspect that your treatment may affect participants heterogeneously, CLASH offers an easy way to improve the equity and effectiveness of your early stopping procedure. (9/11)

Hammaad Adam @hammaadadam1 · Nov 10, 2023
Excited to present our paper today @CODEConference and @NeurIPSConf in December! Really enjoyed working on this project with a wonderful team of co-authors: @LesterMackey @allisonkoe @ntenenz @lorin_crawford, Fan Yin, and Mary Hu. (10/11)

Hammaad Adam @hammaadadam1 · Nov 10, 2023
Finally, thank you to all our research inspirations for this work! @rameshjohari @ronnyk @deaneckles @david_sontag @darbour26 @Susan_Athey @oziadias @rina_friedberg @goodmanmetrics @emrek @irenetrampoline @amt_shrma @MarzyehGhassemi @spschmit (11/11)