Rapid Implementation of Test Design Using Python

David H. Oh

Economist

Office of Compensation and Working Conditions

GASP Workshop

September 17, 2019

Background

- Machine learning (ML) prototype
  - Developed in early 2019
  - Used Occupational Requirements Survey (ORS) and other supplemental datasets to train the model
  - Predicts the top-five most likely Standard Occupational Classification (SOC) codes

Pilot Test

■ Computer-Assisted Review (CAR) Pilot

  ▶ A small team formed in May, 2019

  ▶ Test the feasibility of implementing the ML algorithm into the production cycle, specifically in the review process

  ▶ Three weeks of testing in the actual ORS production environment

    – From late July to early August

Team kickoff

Begin pilot test

End pilot test

May   June   July   August   September
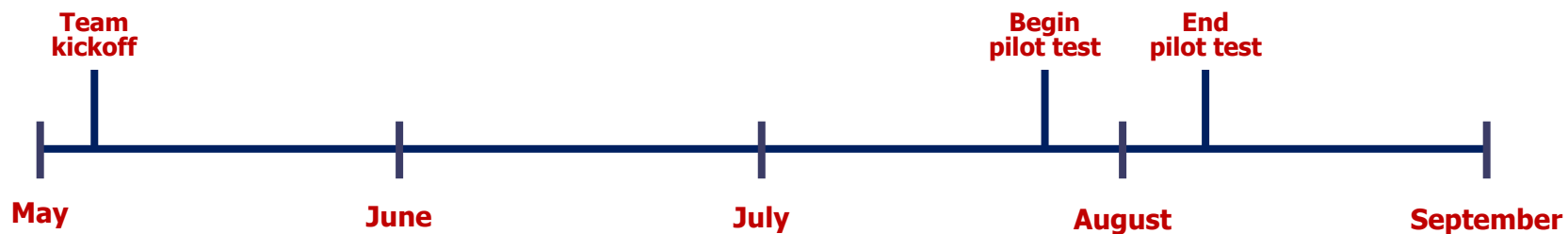
Questions to Answer

■ The effects of CAR on SOC code review

▶ Its effects on the time spent reviewing the SOC code

▶ Its effects on the number of questions being sent out

▶ Its effects on the review resulting in a positive change

■ The effects of CAR on reviewer bias

▶ Do exposures to ML algorithm's outputs result in reviewers favoring (or not favoring) specific SOC codes?

Test Design

- Randomized, controlled crossover trial
  - ▶ Eight participants from the microdata review staff
  - ▶ Each participant expected to review approximately 150 incoming data
  - ▶ Every incoming data reviewed by a participant gets randomly assigned to a control/treatment group

Test Design
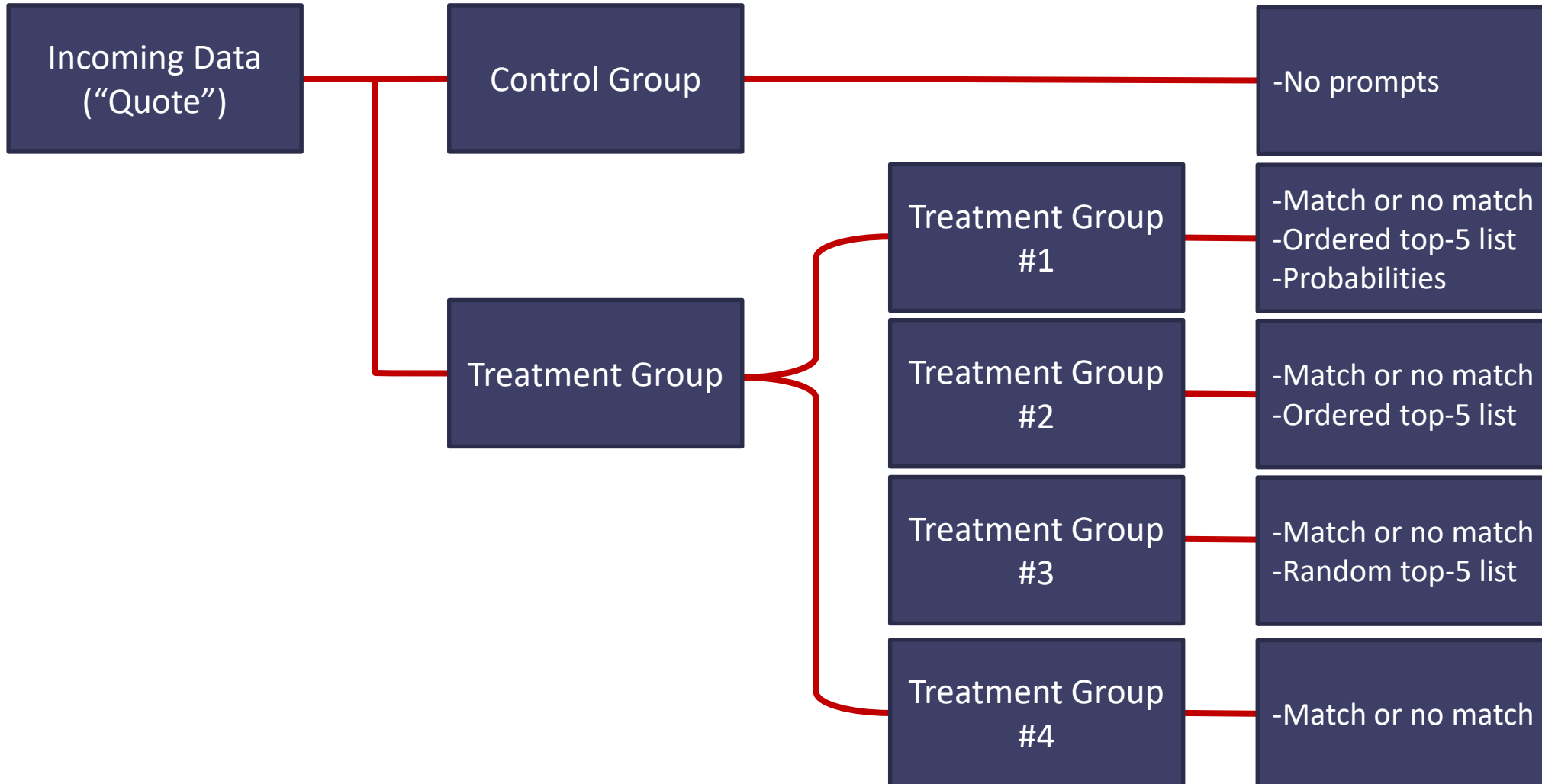
Challenges and Constraints

- Some information are readily captured by the existing production/review system
  - SOC codes
  - Questions sent
- Other information are not available
  - Time spent on reviewing a SOC code
  - Reviewer's expected SOC code
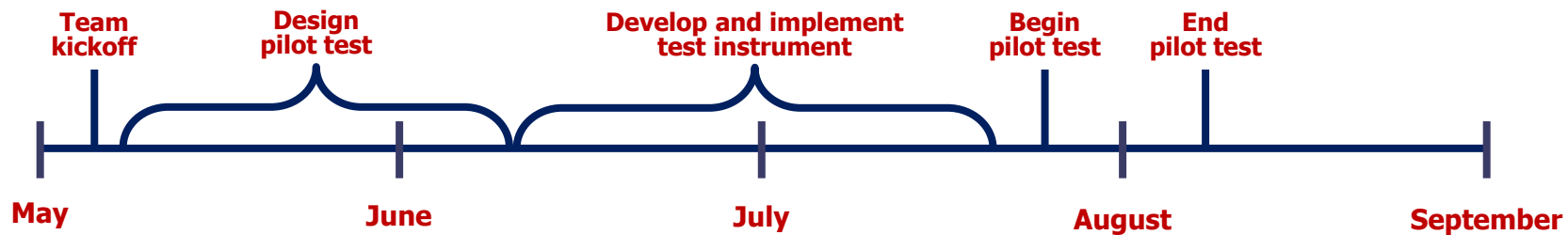
BLS

Challenges and Constraints

- Random assignment
  - Test instrument must be able to perform random assignments
- Resource constraint
  - Minimal disturbance on the actual production
- Time constraint
  - A little over a month to develop and implement the test instrument

Team kickoff   **Design pilot test**   **Develop and implement test instrument**   **Begin pilot test**   **End pilot test**

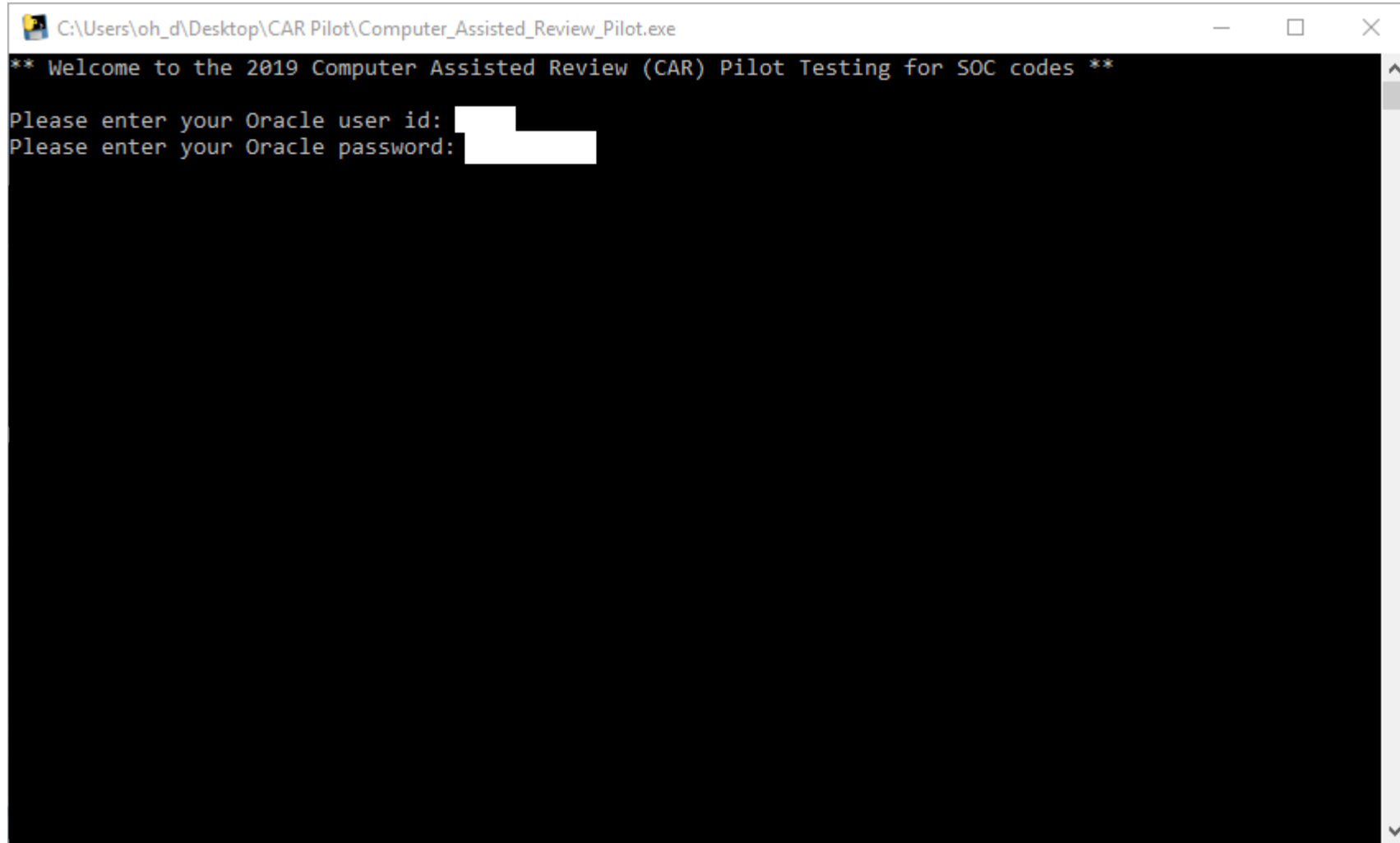**May**   **June**   **July**   **August**   **September**

Benefits of Using Python

- Same language as the one used to develop the ML algorithm
  - ▶ Can easily import the ML algorithm that has been *Pickle*-ed
- Can create a standalone application that is easily distributable
- Can take user inputs
- Can access database
- Can output varying prompts based on random assignments
- Can write to a centralized dataset

Test Instrument

Retrieve information from the database ➡

Show varying prompts based on random assignment ➡

Test Instrument – Treatment #1 Prompt

Test Instrument – Treatment #2 Prompt

Test Instrument – Control Prompt


```
C:\Users\oh_d\Desktop\CAR Pilot\Computer_Assisted_Review_Pilot.exe

** Welcome to the 2019 Computer Assisted Review (CAR) Pilot Testing for SOC codes **

Please enter your Oracle user id:
Please enter your Oracle password:

Please wait while we verify your access and import supporting modules.

Your access has been verified.
Module import process is now complete.
*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*.*

Please enter the quote information:
Schedule:
Hit number:

Job title: Truck Driver
Current 6-digit SOC code: 533032 (Heavy and Tractor-Trailer Truck Drivers)

Do you suspect the current SOC code might be incorrect? (Y)/(N):
```

Test Instrument – Follow-up Questions

- Whether the participant suspects the entered SOC code to be incorrect
  - If yes, a follow-up question on what the correct SOC code would be
- Participant's familiarity with the entered SOC code
  - On a scale from 1 to 5
- Duration (in seconds) collected in the background
  - From the time the random assignment was made to the time participant moved on to the next quote

Lessons for Future Iterations

- Create a more robust centralized database structure for collecting information
  - ▶ Few instances of application crashing on the users due to multiple users writing to the central dataset at the same time
- Develop a web-based application to improve user experience

Conclusion

- Pilot test ended in early August with almost 1,500 quotes reviewed using the test instrument

- The flexibility of Python language and its various applicability enabled the rapid implementation of a randomized, controlled crossover trial

- Results from the pilot test?
  - ▶ Currently being analyzed

Contact Information

David H. Oh
Economist
Office of Compensation and Working Conditions
www.bls.gov/ect
202-691-5985
oh.david@bls.gov

BLS