

REPLICATION PACKAGE FOR:

Does LLM Assistance Improve Healthcare Delivery? Evaluation Using On-site Physicians and Laboratory Tests (Abaluck et al. 2025)

Please use Stata version 19.5 with the estout package installed to run the replication .do files.

The replication package zip file is set up with three folders.

1. The “analysis” > “code” folder contains all do files generating the tables in the paper using our datasets.
2. The “build” > “code” folder contains all do files generating our analysis datasets from the raw data.
3. The “questionnaires” folder contains a printable version of the SurveyCTO questionnaires used to collect data from providers.

For complete replication

1. Ensure that all raw data files (available to World Bank staff on the World Bank Microdata Library (MDL) [here](#)) are placed in the “build” > “data” folder. There should be a total of 16 raw data files:
 - *4 .csv files:*
 - chew_exit_feedback.csv
 - md_evaluation_wide.csv
 - mo_evaluation_wide.csv
 - patient_flow_wide.csv
 - *8 Excel .xlsx files*
 - LLM_follow_through_evaluation.xlsx
 - LLM_mo_evaluation_soap1.csv
 - LLM_mo_evaluation_soap2.csv
 - LLM_similarity_evaluation_regular_franken.xlsx
 - objective_clinical_codes_groupings.xlsx
 - objective_correct_drug_dosage.xlsx
 - objective_rx_groupings.xlsx
 - patient_flow_recovered_wide.xlsx
 - *4 Stata .dta files*
 - EMR_Labtests.dta
 - EMR_Main.dta
 - EMR_Prescriptions.dta
 - mo_evaluation_recovered_wide.dta

2. Run “build/code/build_master.do” to construct all analysis datasets from raw data
3. Run “analysis/code/analysis_master.do” to produce all tables in the paper

Data Sources:

All raw data files should be placed in “build/data/”. The data come from the following sources:

- **Survey data – de-identified**

Conducted via Survey CTO for secure data collection. Survey questions available in “questionnaire” folder

- **mo_evaluation_wide.csv:** On-site physician’s quality of care submissions. The evaluation was done via a detailed survey in conjunction with a custom review mask implemented directly in the EMR. Deployed between Jan - May 2025.
- **mo_evaluation_recovered_wide.csv:** On-site physician’s quality of care submissions recovered from server shutdown. The evaluation was done via a detailed survey in conjunction with a custom review mask implemented directly in the EMR. Deployed between Jan - May 2025.
- **md_evaluation_wide.csv:** Medical Doctor’s quality of care submissions. The sample was selected based on either a high DALY loss rating from the on-site physician in the unassisted note or a substantial change in the on-site physician’s loss rating between unassisted and assisted notes. Collected between Sep - Oct 2025.
- **Chew_exit_feedback.csv:** Exit survey submissions collected from health worker when they rotated out of the study. A follow-up survey with the same questions were shared to health-workers who participated but missed the exit survey. Collected between Feb - Oct 2025.

- **Electronic Medical Records (EMR) – medical records, direct identifiers removed**

- **EMR_Labtest.dta:** Laboratory test records from EMR
- **EMR_Main.dta:** Primary patient encounter data from EMR
- **EMR_Prescriptions.dta:** Prescription records from EMR

- **Patient Flow Data – de-identified**

- **Patient_flow_wide.csv:** records of patient test eligibility, consent status, and study flow
- **Patient_flow_recovered_wide.xlsx:** patient flow data recovered from server shutdown

- **Clinical Reference Data – de-identified**

The following reference files were created by research team members and physicians Aaron Schwartz, MD PhD and Nirmal Ravi, MD PhD:

- **objective_clinical_codes_groupings.xlsx**: Categorized clinical diagnostic codes used by physicians and health workers to check for objective discordance
- **objective_correct_drug_dosage.xlsx**: Reference for correct medication dosages for testing conditions malaria, UTI, and anemia
- **objective_rx_groupings.xlsx**: Categorized prescriptions used by physicians and health workers to check for objective discordance

- **LLM-Generated Evaluations**

The following files were generated by Robert Pless using LLM prompts for specific evaluation tasks:

- **LLM_follow_through_evaluation.xlsx**: LLM assessment of health worker follow-through and similarity to physicians on LLM suggestions
- **LLM_mo_evaluation_soap1.xlsx**: LLM evaluation of health worker's unassisted SOAP note
- **LLM_mo_evaluation_soap2.xlsx**: LLM evaluation of health worker's assisted SOAP note
- **LLM_similarity_evaluation_regular_franken.xlsx**: LLM evaluation on similarity to physician comparing LLM note from being provided the on-site physician's subjective-objective (S-O) compared to providing only the health worker's subjective-objective (S-O)

Notes on the data:

- The server used to collect survey data was disrupted between April 10-May 10. Partial data was recovered.
- For de-identification purposes, exact dates were removed or masked.
- Some data files contain personal medical information. Users of the data should never attempt to use this information to re-identify individuals in the study.

Additional information on the link between replication outputs and paper exhibits below.

Code-Exhibit Linkages

In Manuscript

Code File	Code Output	Exhibit	Medical data
"analysis/code/sample_construction.do"	"sample_construction.tex"	Table 1	Yes
"analysis/code/chew_assessment.do"	"chew_combined_feedback_table.tex"	Table 2	Yes
"analysis/code/mo_evaluation_plan_elements.do"	"mo_plan_element_evaluation.tex"	Table 3	Yes
"analysis/code/care_quality_assessment.do"	"mo_ratings_general.tex"	Table 4	Yes
"analysis/code/care_quality_assessment.do"	"md_ratings_combined.tex"	Table 5	Yes
"analysis/code/under_over_testing.do"	"combined_testing_objective_imp.tex"	Table 6	Yes
"analysis/code/misallocate_treat.do"	"misallocated_treatment.tex"	Table 7	Yes
"analysis/code/objective_disconcordance.do"	"objective_rx_diag_disconcordance_any.tex"	Table 8	Yes
"analysis/code/chew_follow_through_summary.do"	"followed_summary_table_count.tex"	Table 9	No
"analysis/code/synthetic_concordance.do"	"llm_franken_analysis_count.tex"	Table 10	No

In Appendix

Code File	Code Output	Exhibit	Medical data
"analysis/code/care_quality_assessment.do"	"alt_mo_ratings.tex"	Table B.1	Yes
"analysis/code/care_quality_assessment.do"	"llm_ratings.tex"	Table B.2	Yes
"analysis/code/blinding_validation.do"	"blinding_combined.tex"	Table B.3	Yes
"analysis/code/care_quality_assessment.do"	"ratings_w_mo_assist_guess.tex"	Table B.4	Yes
"analysis/code/md_mo_concordance.do"	"mo_md_concordance.tex"	Table B.5	Yes
"analysis/code/under_over_testing.do"	"testing_by_result_imputed"	Table B.6	Yes
"analysis/code/under_over_testing.do"	"combined_testing_objective_actual.tex"	Table B.7	Yes
"analysis/code/under_over_testing.do"	"testing_concordance_cond_MO_decision.tex"	Table B.8	Yes
"analysis/code/under_over_testing.do"	"combined_testing_objective_imp_rst.tex"	Table B.9	Yes
"analysis/code/misallocate_treat.do"	"misallocated_treatment_rst.tex"	Table B.10	Yes
"analysis/code/objective_disconcordance.do"	"objective_rx_diag_disconcordance.tex"	Table B.11	Yes
"analysis/code/objective_disconcordance.do"	"objective_rx_diag_disconcordance_any_rst"	Table B.12	Yes

