

DIVERSIFYING SYNTHETIC MEDICATION DATA USING OPEN-SOURCE TOOLS AND PUBLIC DATA SOURCES



Joseph LeGrand, PharmD, MS^{1,2}; Kent Bridgeman, PharmD, MHI^{1,3}; Kristen Tokunaga, PharmD, BCGP^{1,4}; Yevgeny Bulochnik, PharmD, ACE^{1,5}, CACP; Robert Hodges, PharmD, MBA, MSDS^{1,6}; Dalton Fabian, PharmD^{1,7}

1. CodeRx, 2. Vanderbilt University Medical Center, 3. Allina Health, 4. Komodo Health, 5. HealthPartners, 6. McKesson, 7. UnityPoint Health

WHAT IS SYNTHETIC HEALTH DATA?

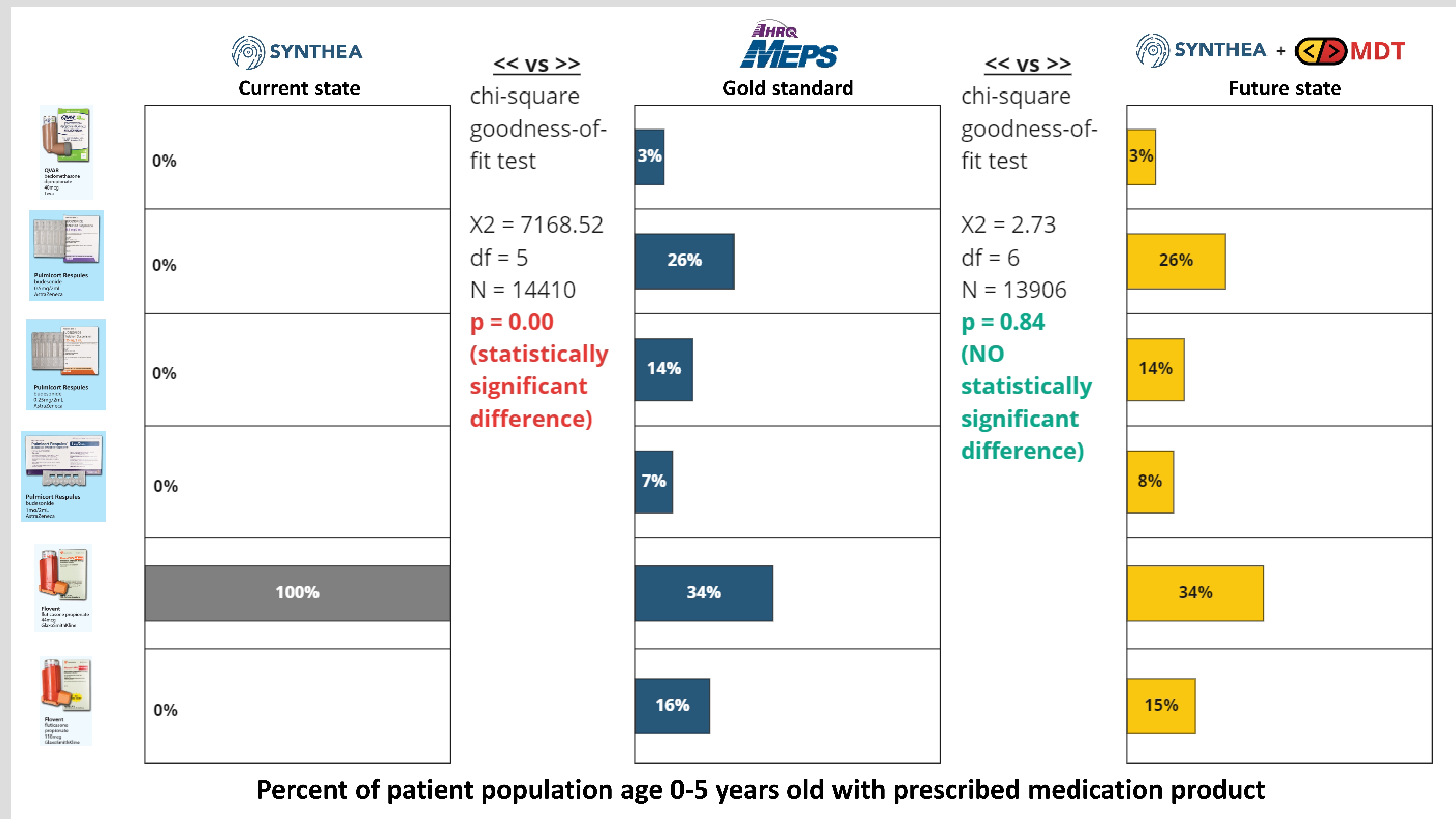
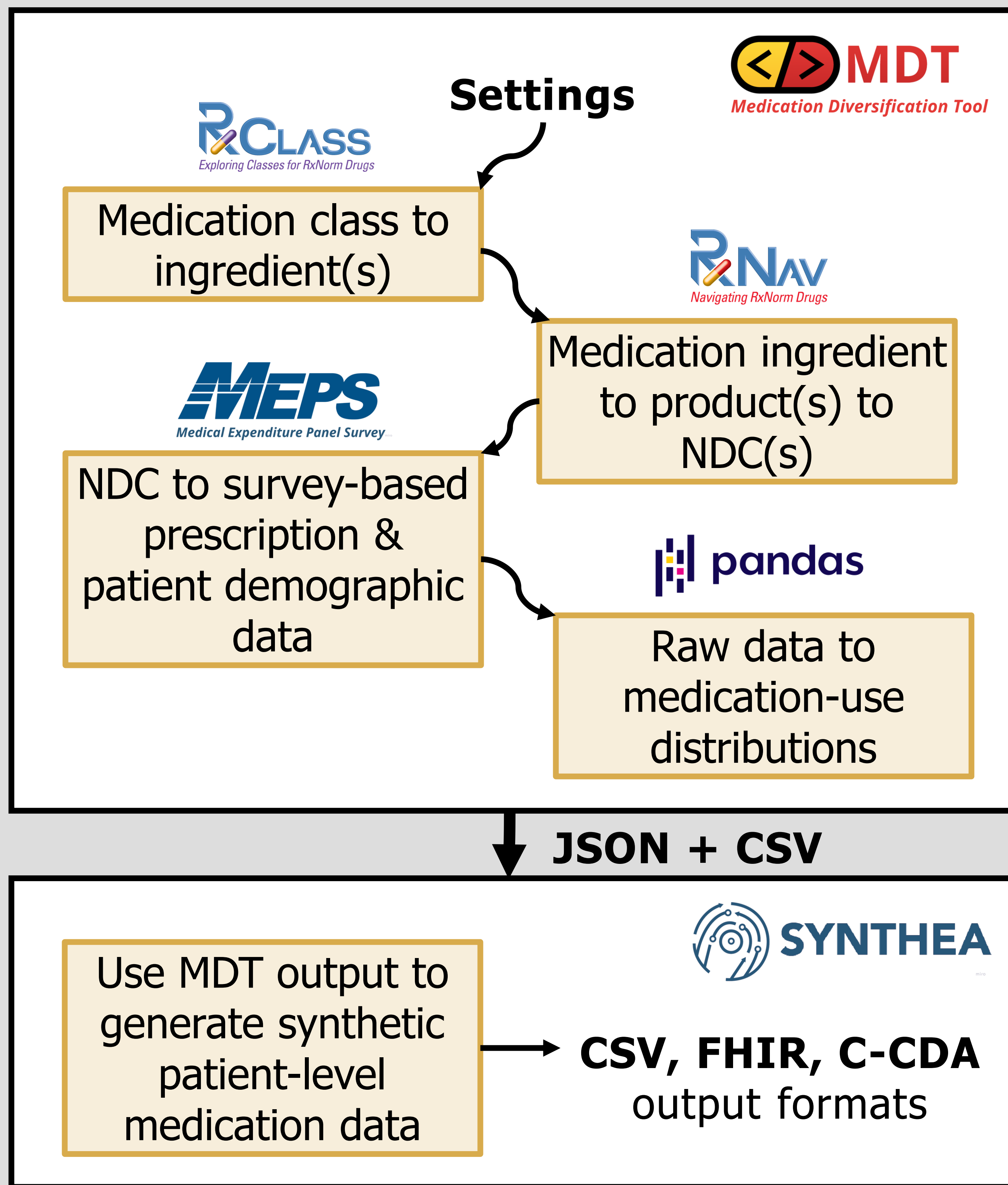
Synthetic data is realistic, but not real, patient data and associated health records. Synthea (created by MITRE) is an open-source tool that generates synthetic patient-level medical records free of PHI, cost, and legal risk - in formats required for research and development purposes.

PROBLEM STATEMENT

The current Synthea tool is creating records with very few medications for 100% of the patients for a given disease state. For example, 100% of asthma patients are getting the same asthma inhaler regardless of age—which is not consistent with real world clinical practice.

OUR SOLUTION

We created a Python module called Medication Diversification Tool (MDT). Our open-source tool accepts user input and uses public data sources to generate a Synthea submodule representative of a US population distribution of medication-use patterns which can be integrated into Synthea.



This work was supported by CodeRx... Please contact the corresponding author at joseph.legrand@vumc.org