Validation of a Large Osteoarthritis Cohort

Michael A. Grasso, MD, PhD[1], Yelena Yesha, PhD[2], Naphtali D. Rishe, PhD[3], Virginia B. Kraus, MD[4], Amanda Niskar, DrPH[5]

[1]University of Maryland School of Medicine, Baltimore, Maryland, [2]University of Maryland Baltimore County, Baltimore, Maryland, [3]Florida International University, Miami, Florida, [4]Duke University School of Medicine, Durham, North Carolina, [5]Arthritis Foundation, Atlanta, Georgia

INTRODUCTION

We identified a large cohort of osteoarthritis (OA) patients to demonstrate the feasibility of a big data approach to identify patients for clinical studies. The data were obtained from the national clinical repository from the Veterans Health Administration (VA).

This project was funded in full by the Arthritis Foundation. Additional resources and facilities were provided by the Department of Veterans Affairs and the Arthritis Foundation. Regulatory approval was obtained from the University of Maryland and the Baltimore Veterans Affairs Medical Center.

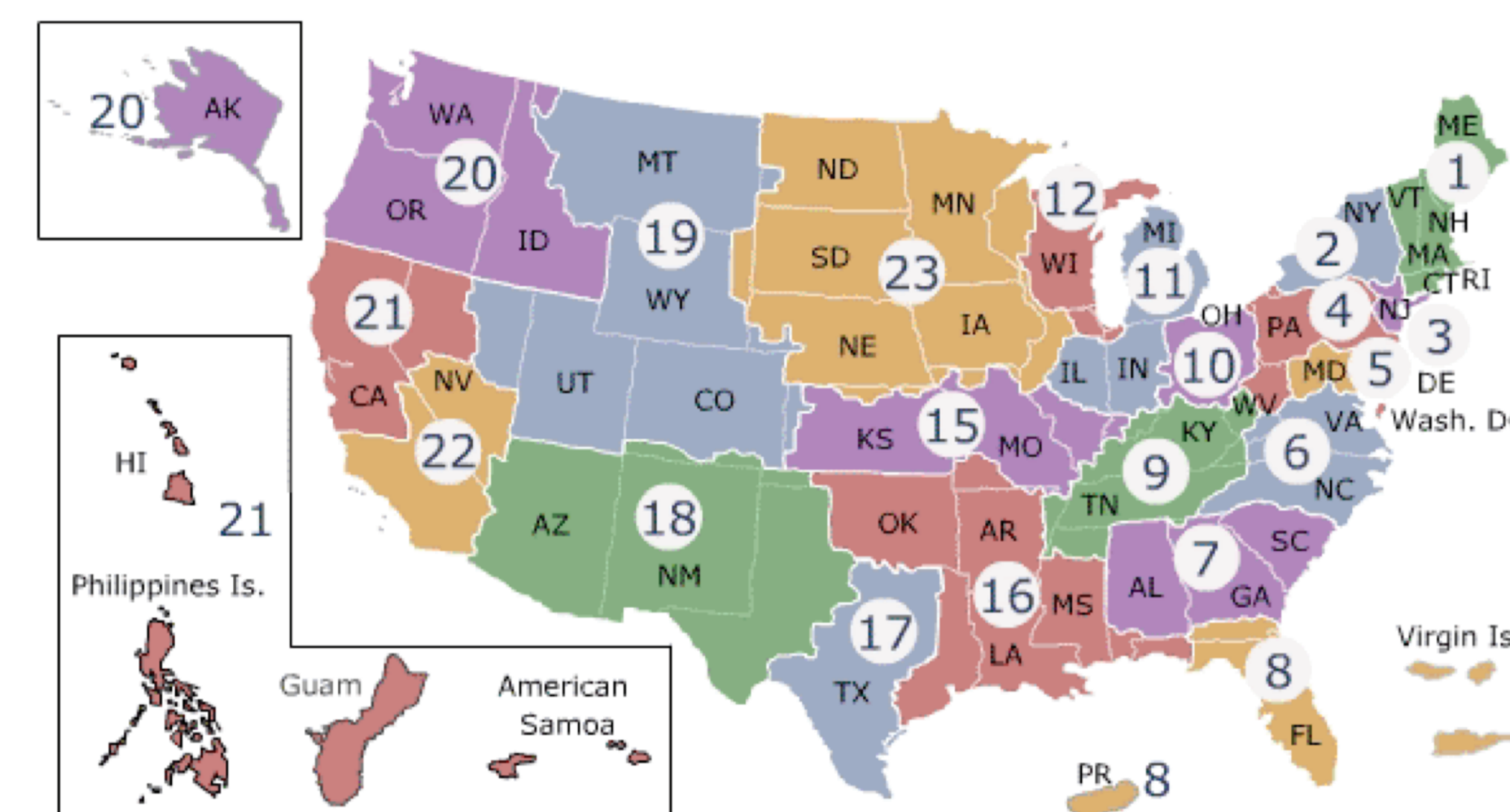### Figure 1. Veterans Affairs (VA) Clinical Locations
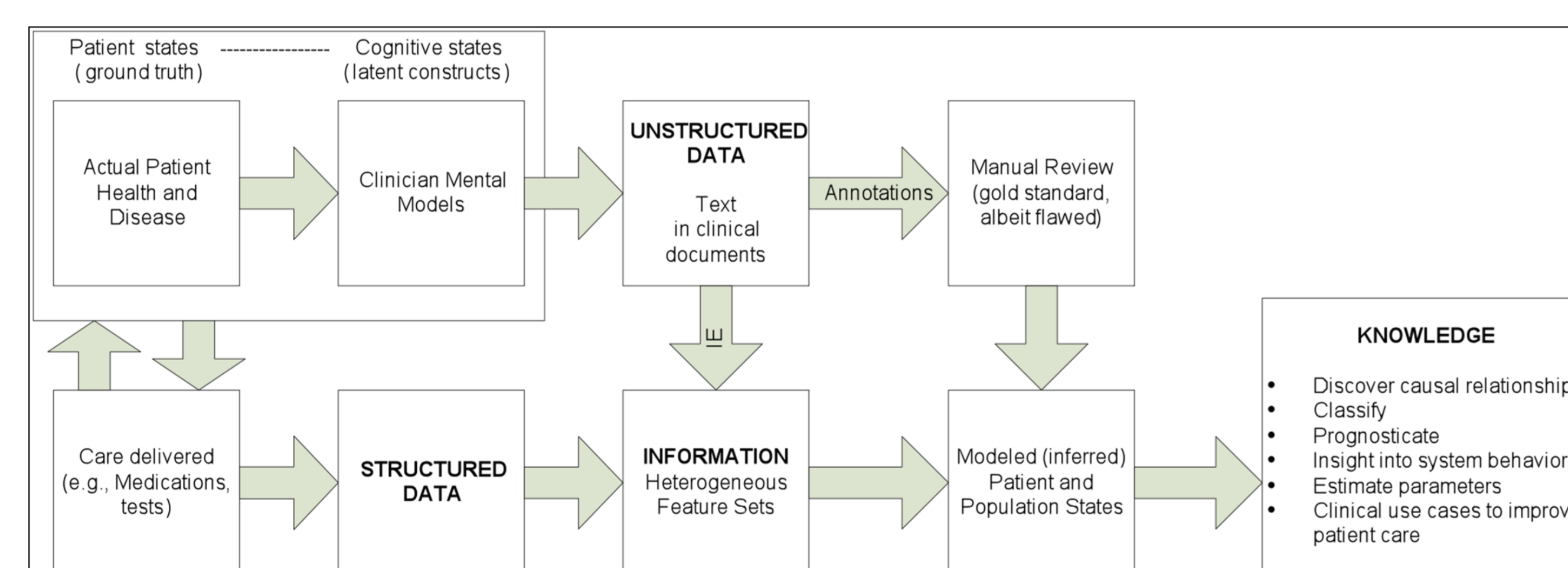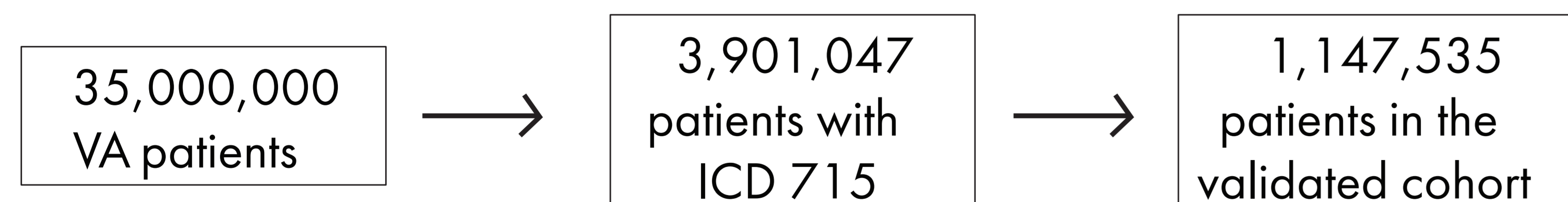

Figure 2. Using All Types of Data

Figure 3. Cohort Identification

35,000,000 VA patients → 3,901,047 patients with ICD 715 → 1,147,535 patients in the validated cohort

METHODS

CLINICAL DATA

The data for our cohort were obtained from the VA Informatics and Computing Infrastructure (VINCI). This dataset is one of the largest clinical repositories available.

- Data on 35 million veterans who received care between the years 2000 and 2014
- More than 160 hospitals and 800 clinics in the VA (Figure 1)
- Comprehensive repository of physician notes, imaging reports, pharmacy data, vitals, laboratory data, admissions, clinic visits (Figure 2)

APPROACH

Our big data approach included the following steps.

- Cohort identification and validation
- Feature extraction
- Data analysis

CLINICAL COHORT

IDENTIFICATION AND VALIDATION

Our inclusion criteria were based on guidelines from the American College of Rheumatology (ACR).

- Screened the national repository of 35 million veterans
- We started with an initial cohort of veterans with a history of OA (ICD-9 715)
- For Knee OA, we included those who were at least 50 years old, and who had at least one clinic visit for knee pain (ICD-9 719.46)
- For Hip OA, we included those who never had an erythrocyte sedimentation rate above 20, and who had at least one clinic visit for hip pain (ICD-9 719.45)
- We validated the cohort through a manual review of 40 charts, half selected from the validated cohort, half selected from among those excluded

We identified 3,901,047 patients with a diagnosis of osteoarthrosis (Figure 3). The validated cohort included 1,147,535 patients, of which 1,073,169 (94%) were male, and 755,010 (66%) were Caucasian.

RESULTS

Our manual review of the cohort included 20 true positives, 13 true negatives, 7 false negatives, and no false positives. For this analysis, sensitivity was 74% and specificity was 100%.

Clinic visits by the OA cohort.

- 5,980,233 primary care visits
- 884,882 rheumatology clinic visits
- 4,649,522 emergency medical visits
- The top five ICD-9 codes for these visits were hypertension (719,558 visits), hyperlipidemia (372,900 visits), diabetes (348,787 visits), lower back pain (222,397 visits), and joint pain (211,739 visits)

Medications prescribed to members of the cohort (over-the-counter medications were not included).

- NSAIDs (919,469, 79%)
- Acetaminophen (874,721, 76%)
- Opiates (829,063, 72%)
- Neuropathic pain medication (478,462, 42%)
- Tramadol (431,338, 38%)
- Muscle relaxers (355,325, 31%)

CONCLUSIONS

OA COHORT

- Primarily Caucasian males, driven by the demographics of the VA
- Clinic visits by the cohort were similar to others within the VA system
- Majority of patients in the OA cohort have used NSAIDs, acetaminophen, and opiate pain medication
- The sensitivity of the cohort could be increased by extracting additional ACR criteria from the VA repository

BIG DATA

This project is a first step in demonstrating the power of big data to improve our understanding of OA. These analytics can be used for a variety of OA scientific discovery topics such as:

- Phenotyping and disease classification to improve clinical study recruitment
- Understanding which treatment is the best fit for an individual
- Detecting differences in care patterns by location and other characteristics