UNIVERSITY
OF
CALIFORNIA

Precisely Practicing Medicine from 700 Trillion Points of University of California Health Data

Atul Butte, MD, PhD

Chief Data Scientist, University of California Health (UC Health)

Priscilla Chan and Mark Zuckerberg Distinguished Professor

Director, Bakar Computational Health Sciences Institute, UCSF

atul.butte@ucsf.edu ▪ @atulbutte

Conflicts of Interest

- Scientific founder and advisory board membership
  - Genstruct
  - NuMedii
  - Personalis
  - Carmenta
- Honoraria for talks
  - Lilly
  - Pfizer
  - Siemens
  - Bristol Myers Squibb
  - AstraZeneca
  - Roche
  - Genentech
  - Warburg Pincus
  - CRG
  - AbbVie
  - Westat
- Past or present consultancy
  - Lilly
  - Johnson and Johnson
  - Roche
  - NuMedii
  - Genstruct
  - Tercica
  - Ecoeos
  - Helix
  - Ansh Labs
  - uBiome
  - Prevendia
  - Samsung
  - Assay Depot
  - Regeneron
  - Verinata
  - Pathway Diagnostics
  - Geisinger Health
  - Covance
  - Wilson Sonsini Goodrich & Rosati
  - Orrick
  - 10X Genomics
  - GNS Healthcare
  - Gerson Lehman Group
  - Coatue Management
- Corporate Relationships
  - Northrop Grumman
  - Genentech
  - Optum
  - Aptalis
  - Allergan
  - Astellas
  - Thomson Reuters
  - Intel
  - SAP
  - SV Angel
  - Progenity
  - Illumina
- Speakers' bureau
  - None
- Companies started by students
  - Carmenta
  - Serendipity
  - Stimulomics
  - NunaHealth
  - Praedicat
  - MyTime
  - Flipora
  - Tumbl.in
  - Polyglot
  - Iota Health
  - Ongevity Health

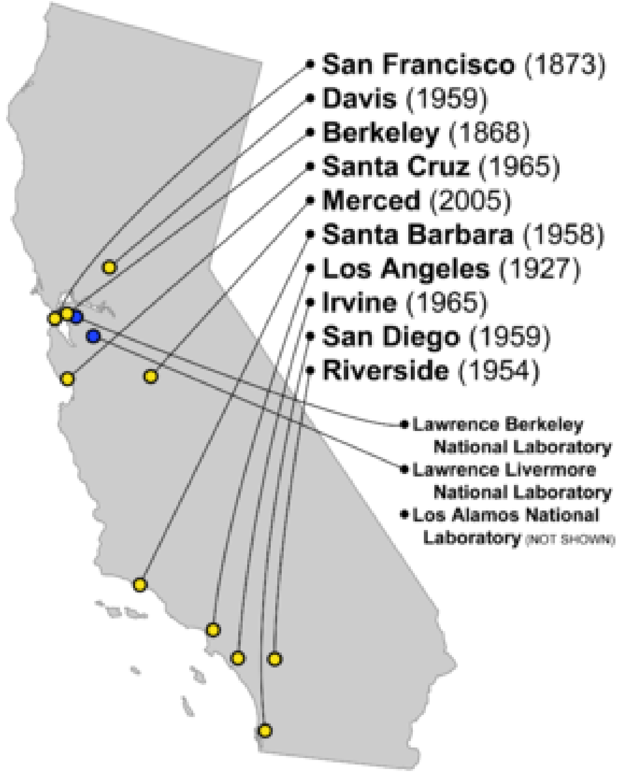
University of California

- 10 campuses and 3 national labs
- ~200,000 employees, ~250,000 students/yr

UC Health

UC Health

- 18 health professional schools (6 med schools)
- Train half the medical students and residents in California
- ~$2 billion NIH funding
- $13+ billion clinical operating revenue
- 5000 faculty physicians, 12000 nurses
- UCSF and UCLA are in US News top 10
- 5 NCI Comprehensive Cancer Centers, 5 NIH CTSA
- IRB reliance, centralized contracting



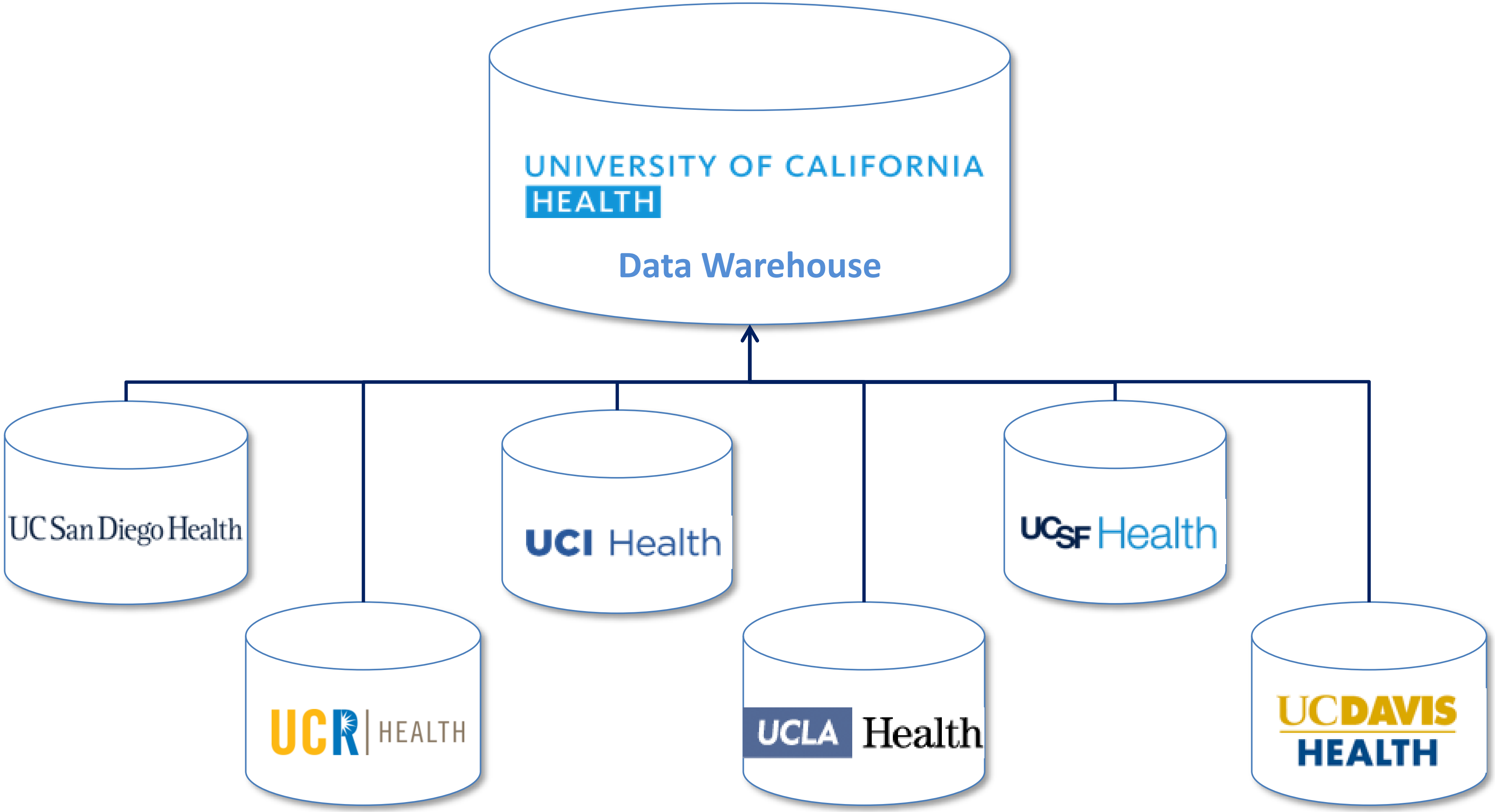
UC Health, United Healthcare Form New ACO & Clinically Integrated Networks

by Staff Writer 10/03/2016 0 Comments

UNIVERSITY OF CALIFORNIA UC Health

The University of California and UnitedHealth Group are teaming up to form a new accountable care organization (ACO) and clinically integrated network. As part of the 10-year strategic relationship, UC Health's five academic medical centers will expand use of Optum's clinically integrated network services and advanced data analytics services.

Combining healthcare data from across the six University of California medical schools and systems



UC Health Data Analytics Platform

The University of California has an incredible view of the medical system

- Combined EHR data from UCSF, UCLA, UC Irvine, UC Davis, UC San Diego, and UC Riverside
  - 15 million patients treated over the past 15 years
- Central database built using OMOP (not Epic) as a data backend
  - Structured data from 2012 to the present day
  - **5.7 million patients with “modern” data**
  - 228M encounters, 96M procedures, 593M med orders, 640M diagnosis codes, 2.3B lab tests and vital signs
  - “Everything from Tylenol to CAR-T cells…”
  - OSHPD data, pathology and radiology text elements, death index
  - Claims data from our self-funded plans now included
  - Continually harmonizing elements
- Quality and performance dashboards

Many operational teams within UC Health now using and benefitting from the UC Health Data Warehouse, saving $millions

Central tools to improve quality of care

Decreasing expenses in our self-funded health plans

Managing costs in our self-funded health plans

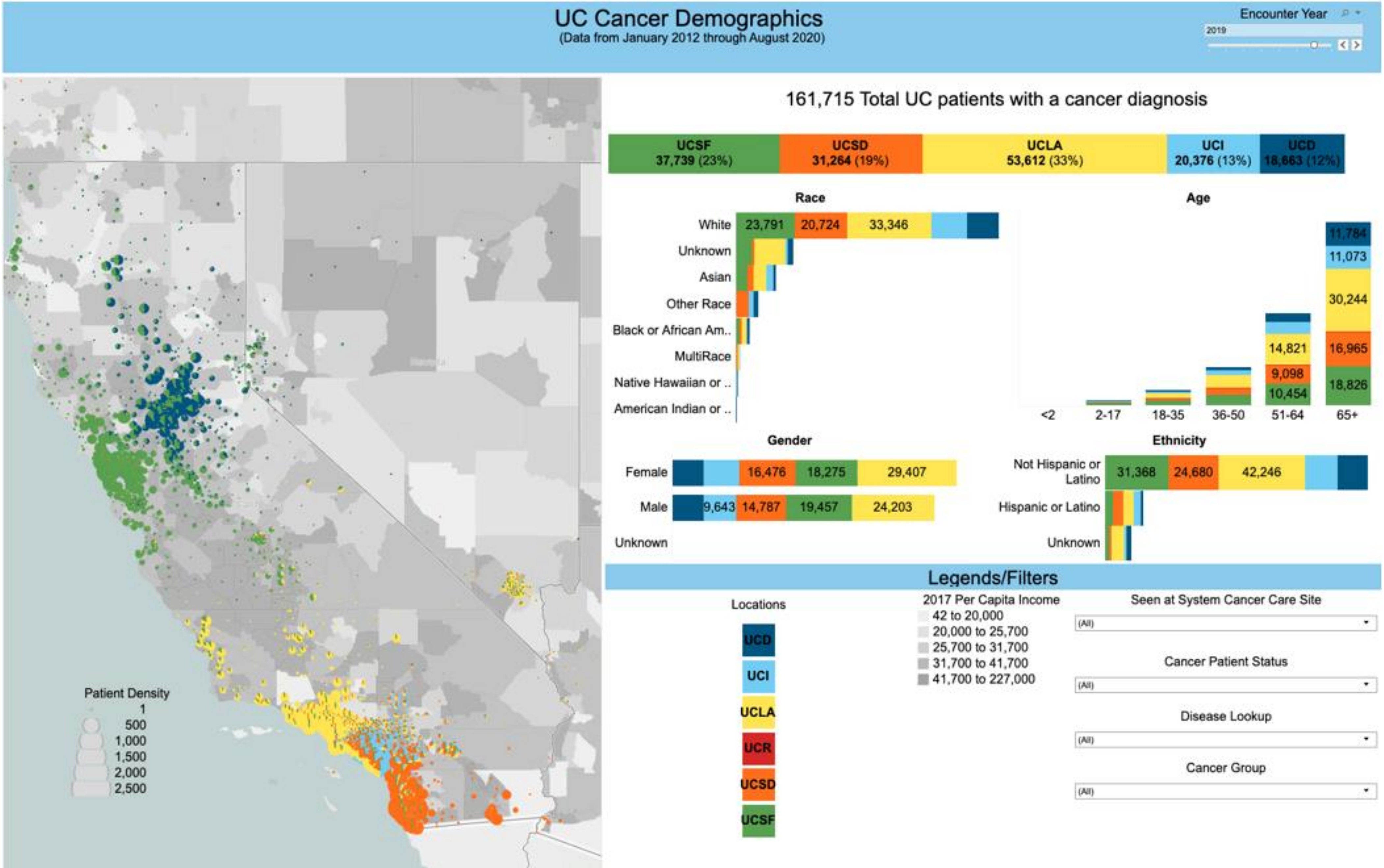
Central management of primary care patients

University of California Cancer Consortium Takes on California's $14 Billion Killer

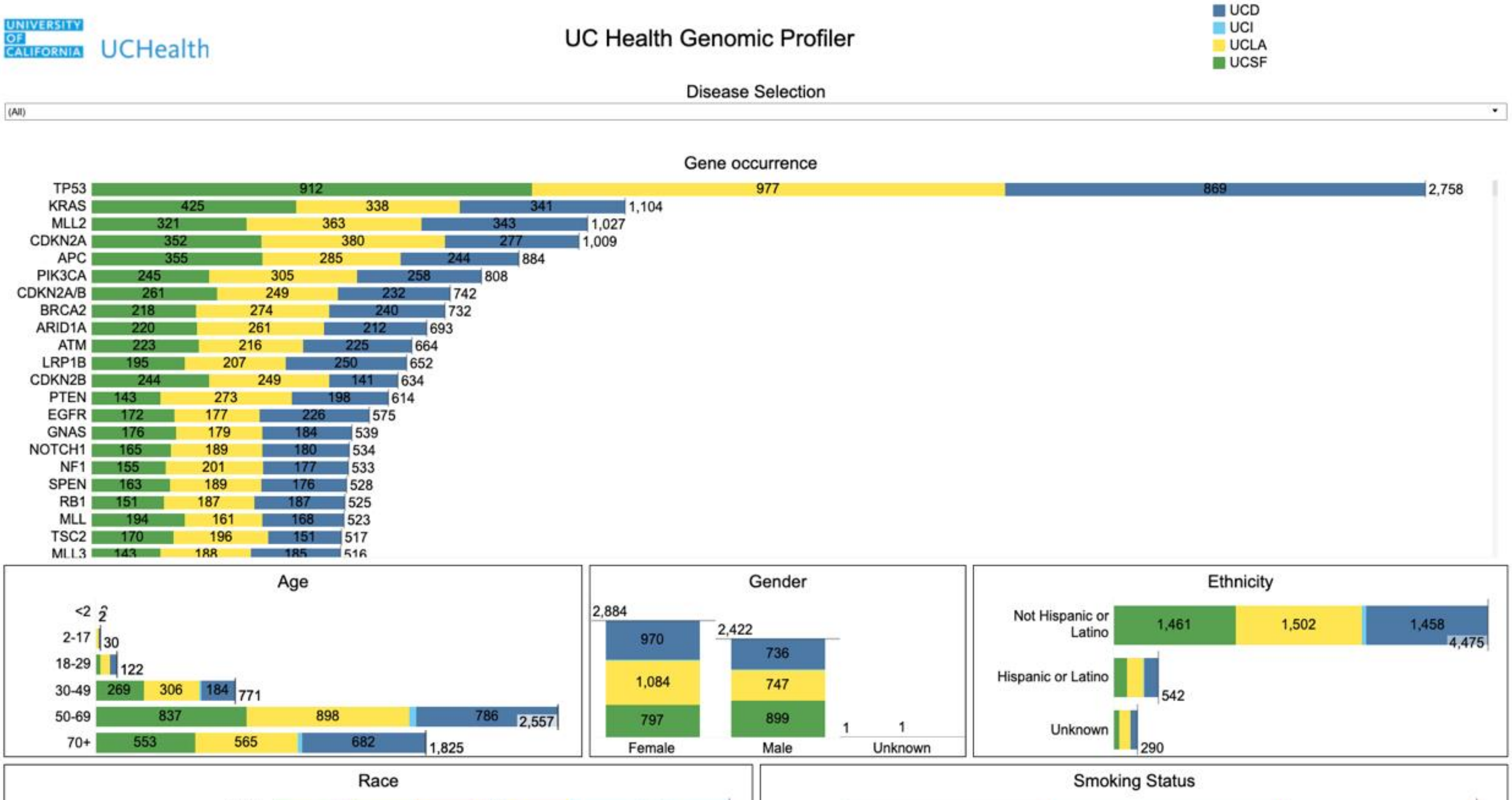
By Elizabeth Fernandez on September 11, 2017

University of California President Janet Napolitano announced the formation of a UC Cancer Consortium to include five of the nation's leading academic cancer centers during a press conference in Genentech Hall at Mission Bay. *Photo by Susan Merrell*

161,715 cancer patients seen in 2019 across UC Health

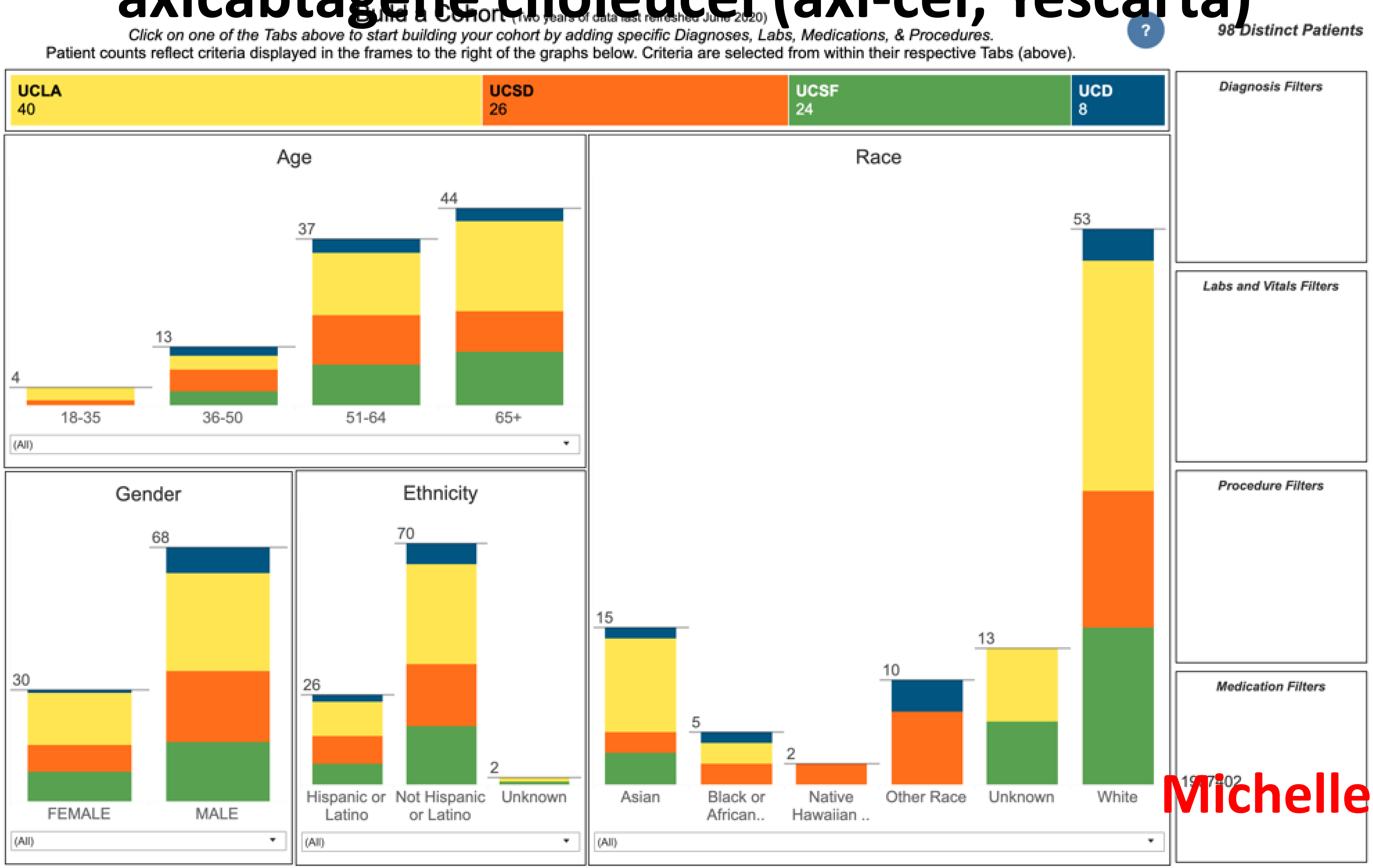


Foundation Medicine cancer genomic reports centralized and parsed



Top 10 Drug Charges across UC Health

Nearly 100 patients treated so far with axicabtagene ciloleucel (axi-cel; Yescarta)



Michelle Wang

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma

S.S. Neelapu, F.L. Locke, N.L. Bartlett, L.J. Lekakis, D.B. Miklos, C.A. Jacobson, I. Braunschweig, O.O. Oluwole, T. Siddiqi, Y. Lin, J.M. Timmerman, P.J. Stiff, J.W. Friedberg, I.W. Flinn, A. Goy, B.T. Hill, M.R. Smith, A. Deol, U. Farooq, P. McSweeney, J. Munoz, I. Avivi, J.E. Castro, J.R. Westin, J.C. Chavez, A. Ghobadi, K.V. Komanduri, R. Levy, E.D. Jacobsen, T.E. Witzig, P. Reagan, A. Bot, J. Rossi, L. Navale, Y. Jiang, J. Aycock, M. Elias, D. Chang, J. Wiezorek, and W.Y. Go

ABSTRACT

BACKGROUND

In a phase 1 trial, axicabtagene ciloleucel (axi-cel), an autologous anti-CD19 chimeric antigen receptor (CAR) T-cell therapy, showed efficacy in patients with refractory large B-cell lymphoma after the failure of conventional therapy.

METHODS

In this multicenter, phase 2 trial, we enrolled 111 patients with diffuse large B-cell lymphoma, primary mediastinal B-cell lymphoma, or transformed follicular lymphoma who had refractory disease despite undergoing recommended prior therapy. Patients received a target dose of 2×10^6 anti-CD19 CAR T cells per kilogram of body weight after receiving a conditioning regimen of low-dose cyclophosphamide and fludarabine. The primary end point was the rate of objective response (calculated as the combined rates of complete response and partial response). Secondary end points included overall survival, safety, and biomarker assessments.

RESULTS

Among the 111 patients who were enrolled, axi-cel was successfully manufactured for 110 (99%) and administered to 101 (91%). The objective response rate was 82%, and the complete response rate was 54%.With a median follow-up of 15.4 months, 42% of the patients continued to have a response, with 40% continuing to have a complete response. The overall rate of survival at 18 months was 52%. The most common adverse events of grade 3 or higher during treatment were neutropenia (in 78% of the patients), anemia (in 43%), and thrombocytopenia (in 38%). Grade 3 or higher cytokine release syndrome and neurologic events occurred in 13% and 28% of the patients, respectively. Three of the patients died during treatment. Higher CAR T-cell levels in blood were associated with response.

CONCLUSIONS

In this multicenter study, patients with refractory large B-cell lymphoma who received CAR T-cell therapy with axi-cel had high levels of durable response, with a safety profile that included myelosuppression, the cytokine release syndrome, and neurologic events. (Funded by Kite Pharma and the Leukemia and Lymphoma Society Therapy Acceleration Program; ZUMA-1 ClinicalTrials.gov number, NCT02348216.)

The autho grees, and pendix. A Neelapu a Anderson Blvd., Hou mdanders Drs. Neela ly to this a

This article 2017, at N

N Engl J M DOI: 10.10

Supplementary Appendix). An additional 9 patients were enrolled and awaiting treatment at the time that the 92nd patient received the axi-cel infusion. Among the 101 patients who received axi-cel, the objective response rate was 82% (95% CI, 73 to 89), with a 54% complete response rate (Fig. 1A, and Fig. S2 in the Supplementary Appendix).

The median time to response was rapid (1.0 month; range, 0.8 to 6.0). The median duration of response was 8.1 months (95% CI, 3.3 to could not be estimated). Response rates were consistent across key covariates, including age, disease stage, International Prognostic Index score at enrollment, presence or absence of bulky disease, cell-of-origin subtype, and use of tocilizumab or gluco-

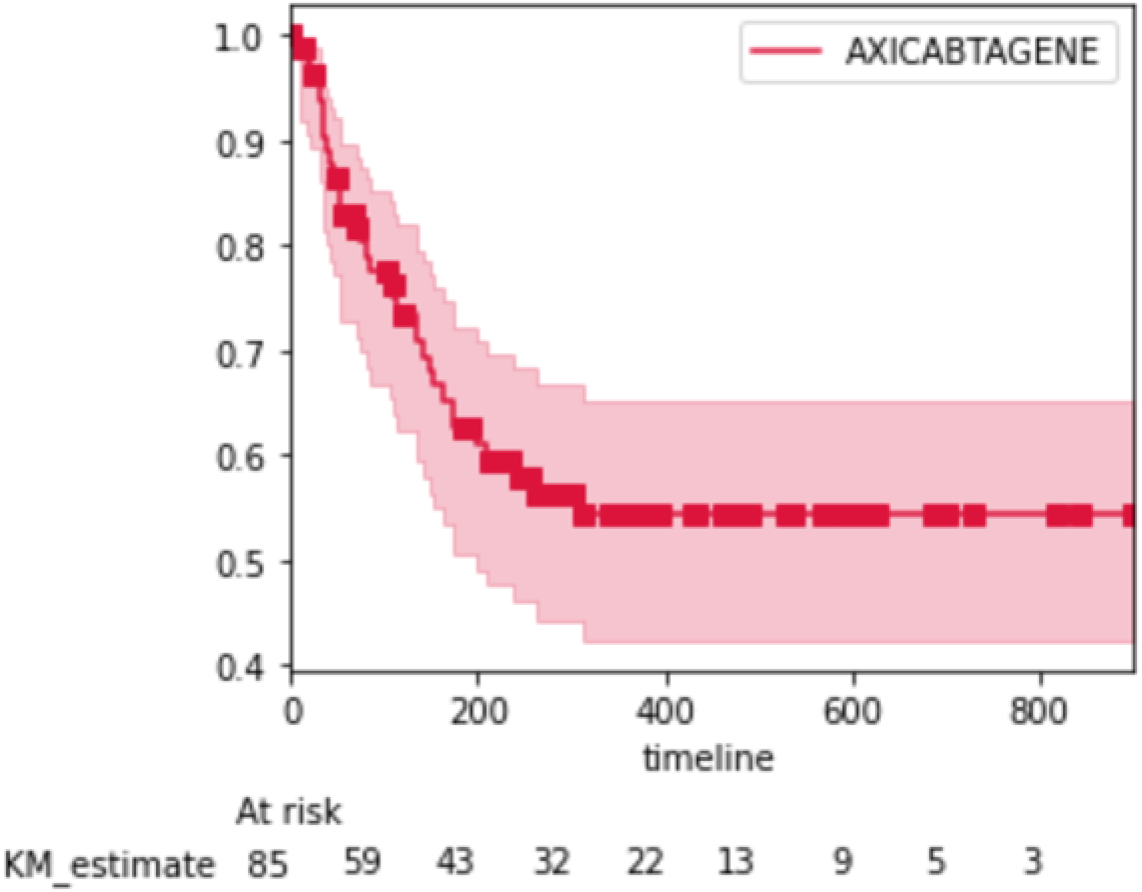
Table 1. Treatment Disposition and Baseline Characteristics of the Patients.*

Variable	Patients with DLBCL	Patients with PMBCL or TFL	All Patients
Treatment disposition			
No. of patients enrolled	81	30	111
Treatment with axi-cel — no. (%)			
Yes	77 (95)	24 (80)	101 (91)
No	4 (5)	6 (20)	10 (9)
Death before treatment†	1 (1)	2 (7)	3 (3)

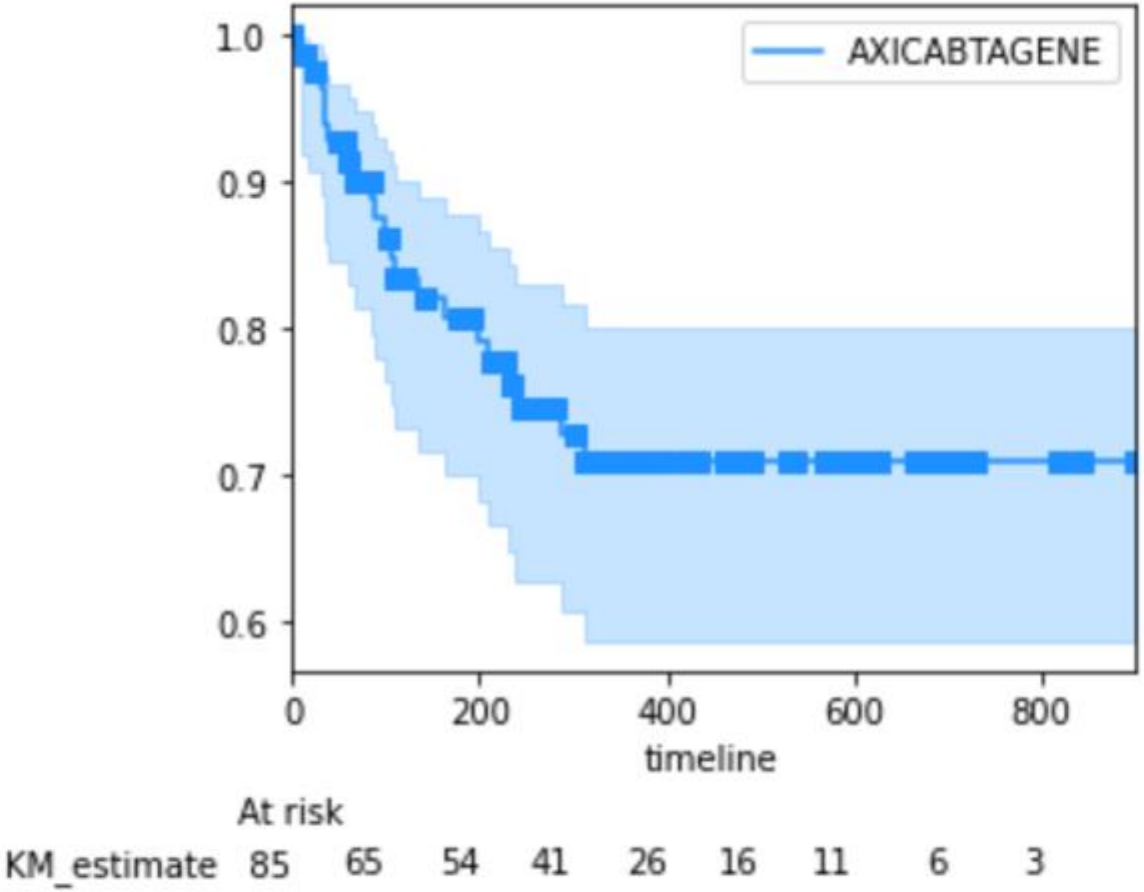
Michelle Wang

PFS and OS at 12 month

Progression Free Survival

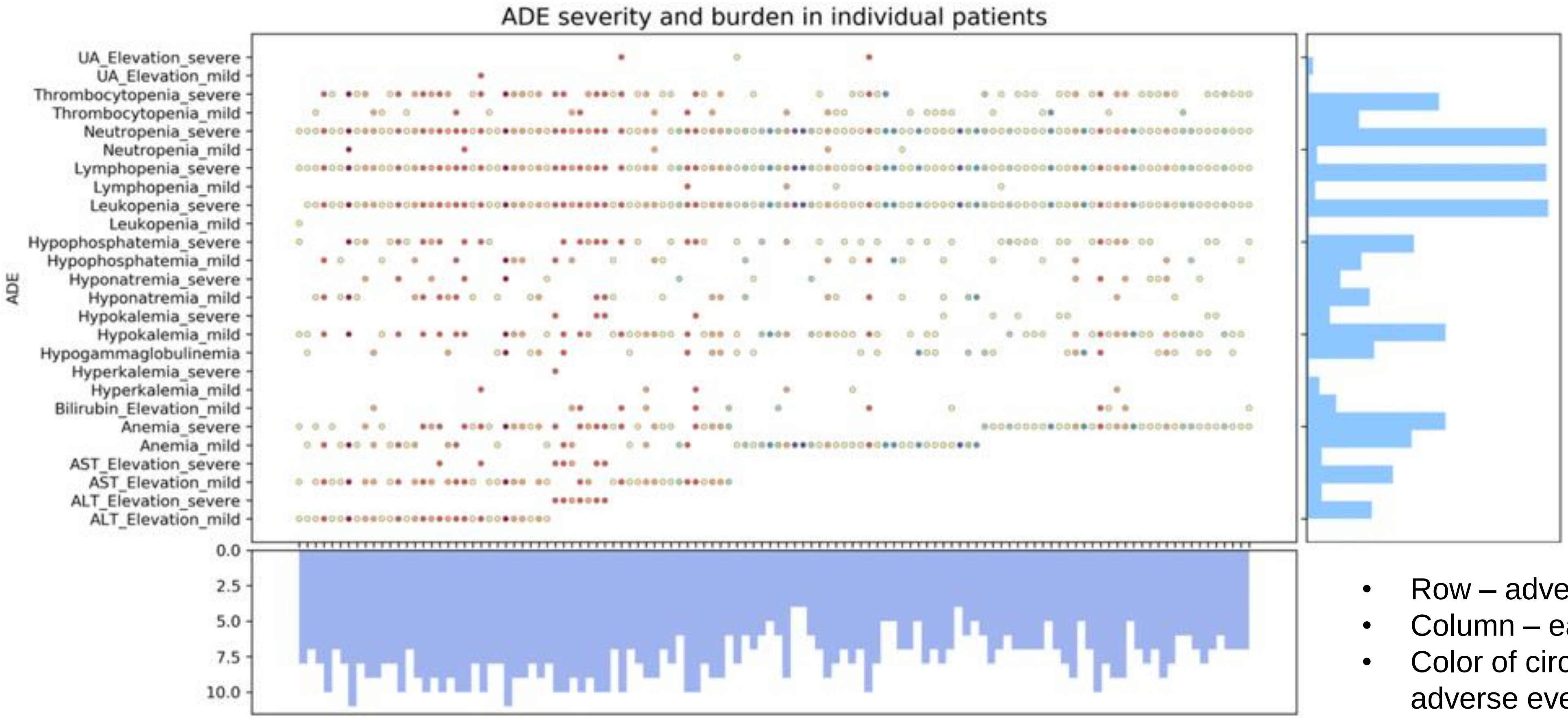


Overall Survival



Michelle Wang

Overview of ADE burden in individual patients

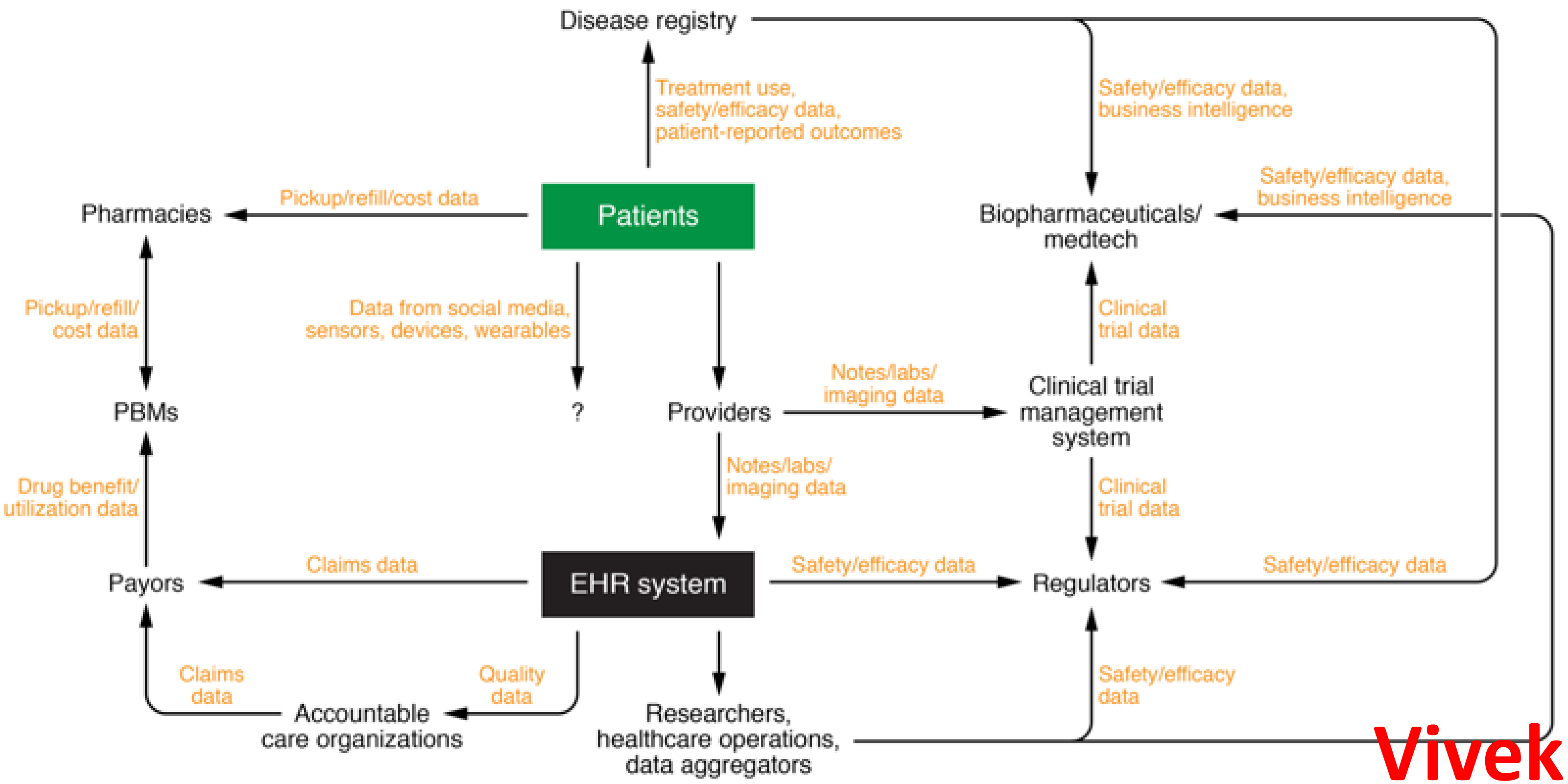


- Row – adverse event
- Column – each individual patient
- Color of circle represents number of adverse events the patient experienced following axi-cel infusion (dark red: highest number of ADE, dark blue: lowest number of ADE)

Michelle Wang

Announcing a new publication on Real World Data

- Journal of Clinical Investigation (JCI) February 2020
  - Opportunities and Challenges in Using Real-World Data for Healthcare
  - Special issue on Big Data's Future in Medicine bit.ly/JCIbigdata



Vivek Rudrapatna

bit.ly/JCIbigdata

Twenty-one uses for Real World Data

Post-approval safety

Updating side effect rates

Discovering novel side effects

Supporting regulatory approval

Single-arm experimental trials

”Digital approvals”

Biosimilar development

Informing clinical trials design

Better patient selection

Trimming the trials: more efficient data collection

Continually establishing efficacy

Assessing the efficacy-effectiveness gap

Searching for efficacy in specific populations

Effect modifiers and precision medicine

Long-term, post trial outcomes

Comparative effectiveness

Integrating costs with comparative effectiveness

Understanding effects of pharmacy practices on healthcare utilization

Studying novel on-label pharmaceuticals versus older off-label drugs

Studying the practice of medicine

Quality of practice, medical errors

Standardizing care and care delivery

The effect of payors on medical care

Are new-generation diagnostics improving outcomes?

Data driven decision support

Clinical-Decision Support: the Provider Perspective

Clinical-Decision Support: the Patient Perspective

Clinical-Decision Support: the Community Perspective

Vivek Rudrapatna

bit.ly/JCIbigdata

The Team

UCLA Health

Mohammed Mahbouba
Albert Duntugan
Jay Shah
Vajra Kasturi
Danielle Belmontez
Michael Swinford (Azure)
Scott Bailey (Azure)
Andrew Weaver (Azure)

UCI Health

Lisa Dahm
Ayan Patel
Charles Wilson
Aiden Barin
Kathy Pickell
David Gonzalez
Lattice Armstead
Tim Hayes

UC DAVIS HEALTH

Kent Anderson
Doug Berman
Steve Covington
Jeff Evoy
Raj Sankala
Hemanth Tatiparthi
Brian Paciotti

UCSF Health

David Dobbs
Rick Larsen
Nelson Lee
Brian Chan
Oksana Gologorskaya
Rao Venigalla

UC San Diego Health

Josh Glandorf
Jennifer Holland
Eugene Lee
Travis Mitchelar
Peter Ryan

The CIO Team

Mike Pfeffer* – UCLA
Chris Longhurst – UCSD
Joe Bengfort – UCSF
Chuck Podesta – UCI
John Cook - UCD

*UCHDW Sponsoring CIO

UC HEALTH

Atul Butte
Tom Andriola
Liz Engel
Jack Stobo
Leslie Yuan

Enabling UC researchers and patients to go beyond… machine learning in a safe, respectful, fair, equitable way in medicine

npj | Digital Me

ARTICLE OPEN

Scalable and
records

Alvin Rajkomar[1,2], Eyal Oren[1], Kai Chen[1], Andrew M. Dai[1], Nissan Hajaj[1], Michaela Hardt[1], Peter J. Liu[1], Xiaobing Liu[1], Jake Marcus[1], Mimi Sun[1], Patrik Sundberg[1], Hector Yee[1], Kun Zhang[1], Yi Zhang[1], Gerardo Flores[1], Gavin E. Duggan[1], Jamie Irvine[1], Quoc Le[1], Kurt Litsch[1], Alexander Mossin[1], Justin Tansuwan[1], De Wang[1], James Wexler[1], Jimbo Wilson[1], Dana Ludwig[2], Samuel L. Volchenboum, Katherine Chou[1], Michael Pearson[1], Srinivasan Madabushi[1], Nigam H. Shah[4], Atul J. Butte[2], Michael D. Howell[1], Claire Cui[1], Greg S. Corrado[1] and Jeffrey Dean[1]

Predictive modeling with electronic health record (EHR) data is anticipated to drive personalized medicine and improve healthcare quality. Constructing predictive statistical models typically requires extraction of curated predictor variables from normalized EHR data, a labor-intensive process that discards the vast majority of information in each patient's record. We propose a representation of patients' entire raw EHR records based on the Fast Healthcare Interoperability Resources (FHIR) format. We demonstrate that deep learning methods using this representation are capable of accurately predicting multiple medical events from multiple centers without site-specific data harmonization. We validated our approach using de-identified EHR data from two US academic medical centers with 216,221 adult patients hospitalized for at least 24 h. In the sequential format we propose, this volume of EHR data unrolled into a total of 46,864,534,945 data points, including clinical notes. Deep learning models achieved high accuracy for tasks such as predicting: in-hospital mortality (area under the receiver operator curve [AUROC] across sites 0.93–0.94), 30-day unplanned readmission (AUROC 0.75–0.76), prolonged length of stay (AUROC 0.85–0.86), and all of a patient's final discharge diagnoses (frequency-weighted AUROC 0.90). These models outperformed traditional, clinically-used predictive models in all cases. We believe that this approach can be used to create accurate and scalable predictions for a variety of clinical scenarios. In a case study of a particular prediction, we demonstrate that neural networks can be used to identify relevant information from the patient's chart.

npj Digital Medicine (2018)1:18 ; doi:10.1038/s41746-018-0029-1

INTRODUCTION

The promise of digital medicine stems in part from the hope that, by digitizing health data, we might more easily leverage computer information systems to understand and improve care. In fact, routinely collected patient healthcare data are now approaching the genomic scale in volume and complexity.[1] Unfortunately, most of this information is not yet used in the sorts of predictive statistical models clinicians might use to improve care delivery. It is widely suspected that use of such efforts, if successful, could provide major benefits not only for patient safety and quality but also in reducing healthcare costs.[2–6]

In spite of the richness and potential of available data, scaling the development of predictive models is difficult because, for traditional predictive modeling techniques, each outcome to be predicted requires the creation of a custom dataset with specific variables.[7] It is widely held that 80% of the effort in an analytic model is preprocessing, merging, customizing, and cleaning datasets,[8,9] not analyzing them for insights. This profoundly limits the scalability of predictive models.

Another challenge is that the number of potential predictor variables in the electronic health record (EHR) may easily number in the thousands, particularly if free-text notes from doctors, nurses, and other providers are included. Traditional modeling approaches have dealt with this complexity simply by choosing a very limited number of commonly collected variables to conside This is problematic because the resulting models may produ imprecise predictions: false-positive predictions can overwhel physicians, nurses, and other providers with false alarms an concomitant alert fatigue,[10] which the Joint Commission identifie as a national patient safety priority in 2014.[11] False-negativ predictions can miss significant numbers of clinically importa events, leading to poor clinical outcomes.[11,12] Incorporating th entire EHR, including clinicians' free-text notes, offers some hop of overcoming these shortcomings but is unwieldy for mo predictive modeling techniques.

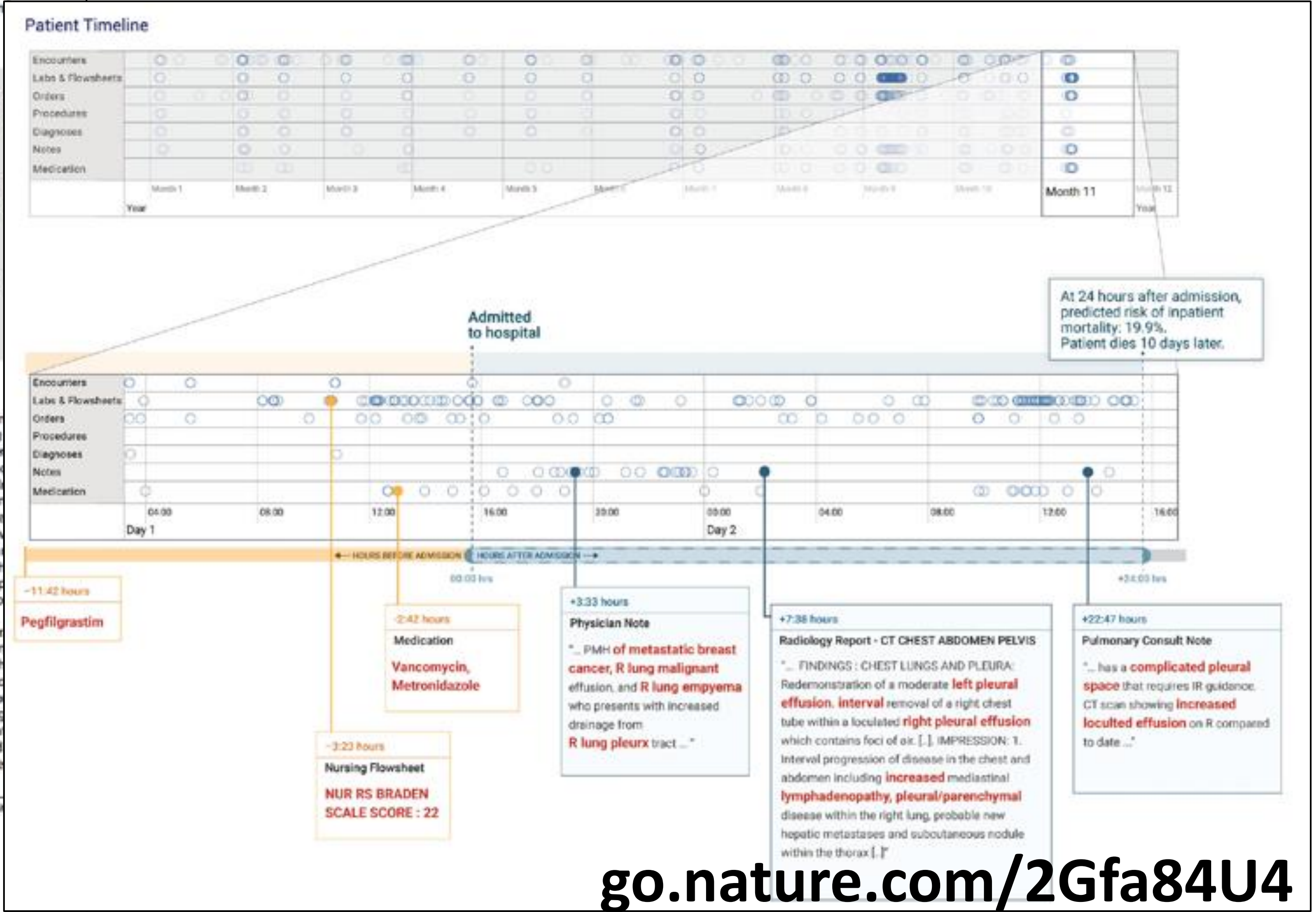
Recent developments in deep learning and artificial neur networks may allow us to address many of these challenges an unlock the information in the EHR. Deep learning emerged as th preferred machine learning approach in machine percepti problems ranging from computer vision to speech recognitio but has more recently proven useful in natural languag processing, sequence prediction, and mixed modality da settings.[13–17] These systems are known for their ability to hand large volumes of relatively messy data, including errors in labe

[1]Google Inc, Mountain View, CA, USA; [2]University of California, San Francisco, San Francisco, CA, USA; [3]University of Chicago Medicine, Chicago, IL, USA and [4]Stanford Univers Stanford, CA, USA
Correspondence: Alvin Rajkomar (alvinrajkomar@google.com)
These authors contributed equally: Alvin Rajkomar, Eyal Oren

Received: 26 January 2018 Revised: 14 March 2018 Accepted: 26 March 2018
Published online: 08 May 2018

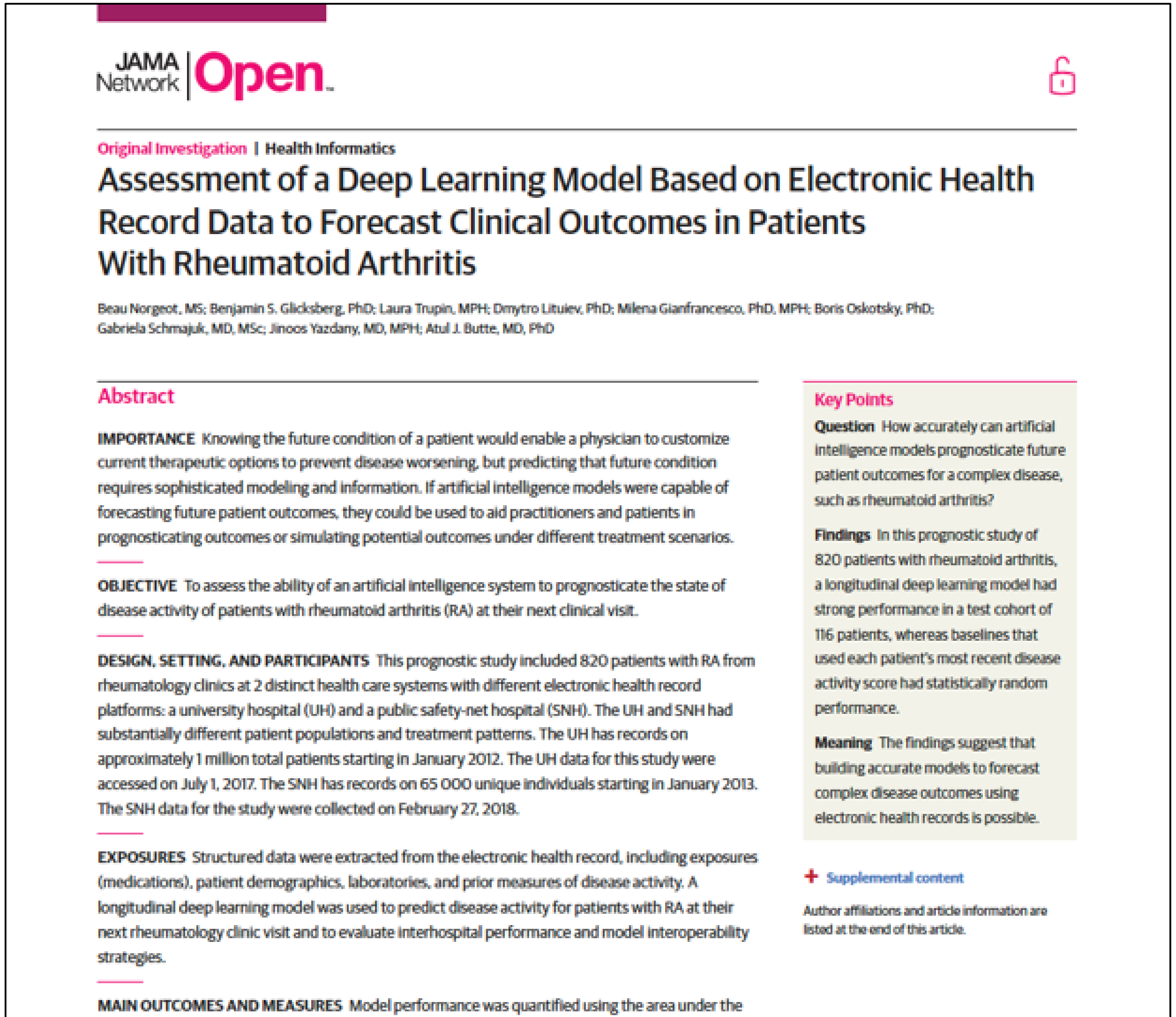
Published in partnership with the Scripps Translational Science Institute

npj nature partner journals



go.nature.com/2Gfa84U4

Predicting the future state of a patient with Rheumatoid Arthritis


JAMA Network | Open

Original Investigation | Health Informatics

Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis

Beau Norgeot, MS; Benjamin S. Glicksberg, PhD; Laura Trupin, MPH; Dmytro Lituiev, PhD; Milena Gianfrancesco, PhD, MPH; Boris Oskotsky, PhD; Gabriela Schmajuk, MD, MSc; Jinoos Yazdany, MD, MPH; Atul J. Butte, MD, PhD

Abstract

IMPORTANCE Knowing the future condition of a patient would enable a physician to customize current therapeutic options to prevent disease worsening, but predicting that future condition requires sophisticated modeling and information. If artificial intelligence models were capable of forecasting future patient outcomes, they could be used to aid practitioners and patients in prognosticating outcomes or simulating potential outcomes under different treatment scenarios.

OBJECTIVE To assess the ability of an artificial intelligence system to prognosticate the state of disease activity of patients with rheumatoid arthritis (RA) at their next clinical visit.

DESIGN, SETTING, AND PARTICIPANTS This prognostic study included 820 patients with RA from rheumatology clinics at 2 distinct health care systems with different electronic health record platforms: a university hospital (UH) and a public safety-net hospital (SNH). The UH and SNH had substantially different patient populations and treatment patterns. The UH has records on approximately 1 million total patients starting in January 2012. The UH data for this study were accessed on July 1, 2017. The SNH has records on 65 000 unique individuals starting in January 2013. The SNH data for the study were collected on February 27, 2018.

EXPOSURES Structured data were extracted from the electronic health record, including exposures (medications), patient demographics, laboratories, and prior measures of disease activity. A longitudinal deep learning model was used to predict disease activity for patients with RA at their next rheumatology clinic visit and to evaluate interhospital performance and model interoperability strategies.

MAIN OUTCOMES AND MEASURES Model performance was quantified using the area under the

Key Points

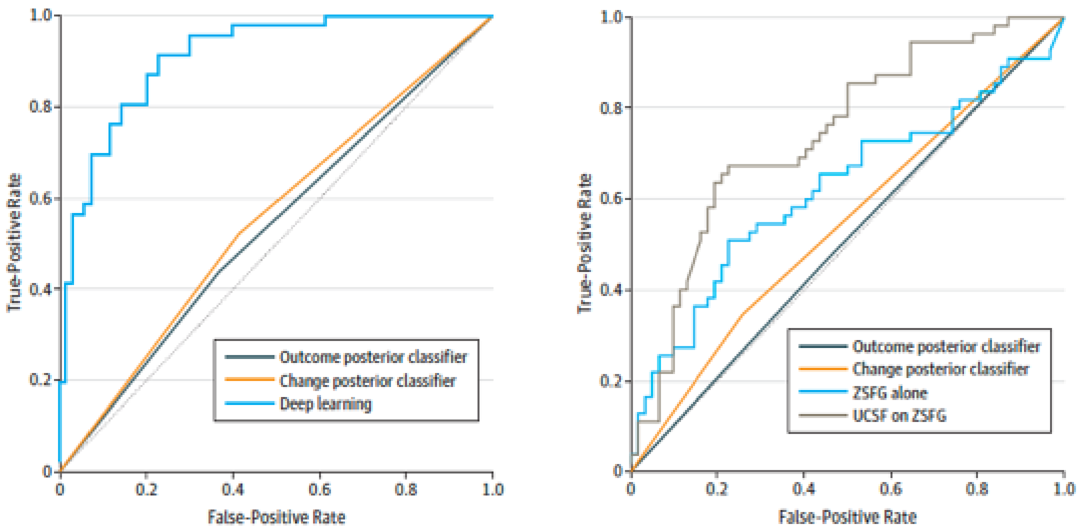
Question How accurately can artificial intelligence models prognosticate future patient outcomes for a complex disease, such as rheumatoid arthritis?

Findings In this prognostic study of 820 patients with rheumatoid arthritis, a longitudinal deep learning model had strong performance in a test cohort of 116 patients, whereas baselines that used each patient's most recent disease activity score had statistically random performance.

Meaning The findings suggest that building accurate models to forecast complex disease outcomes using electronic health records is possible.

+ Supplemental content

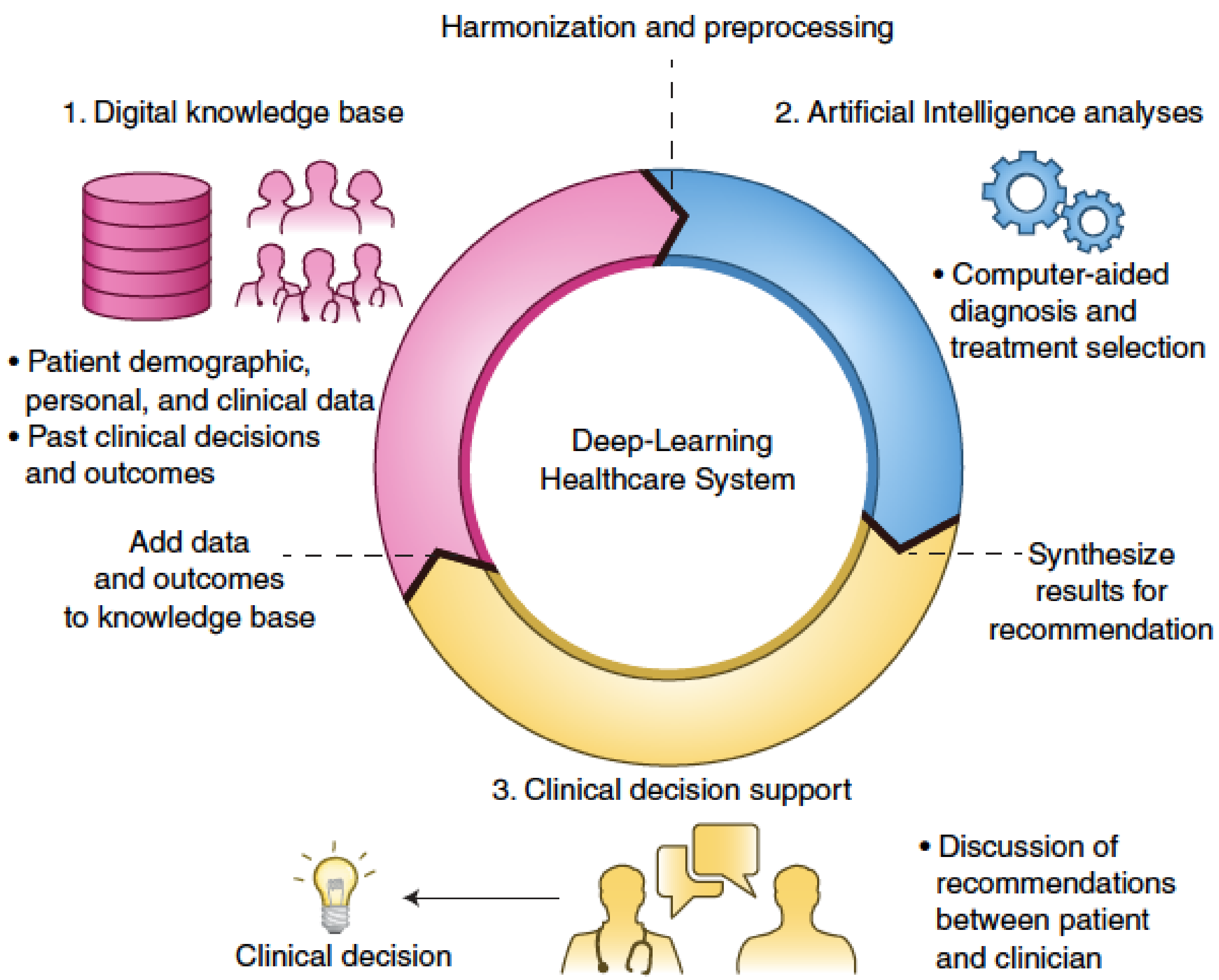
Author affiliations and article information are listed at the end of this article.



Beau Norgeot

bit.ly/jamaRA

A New Deep Learning Healthcare System



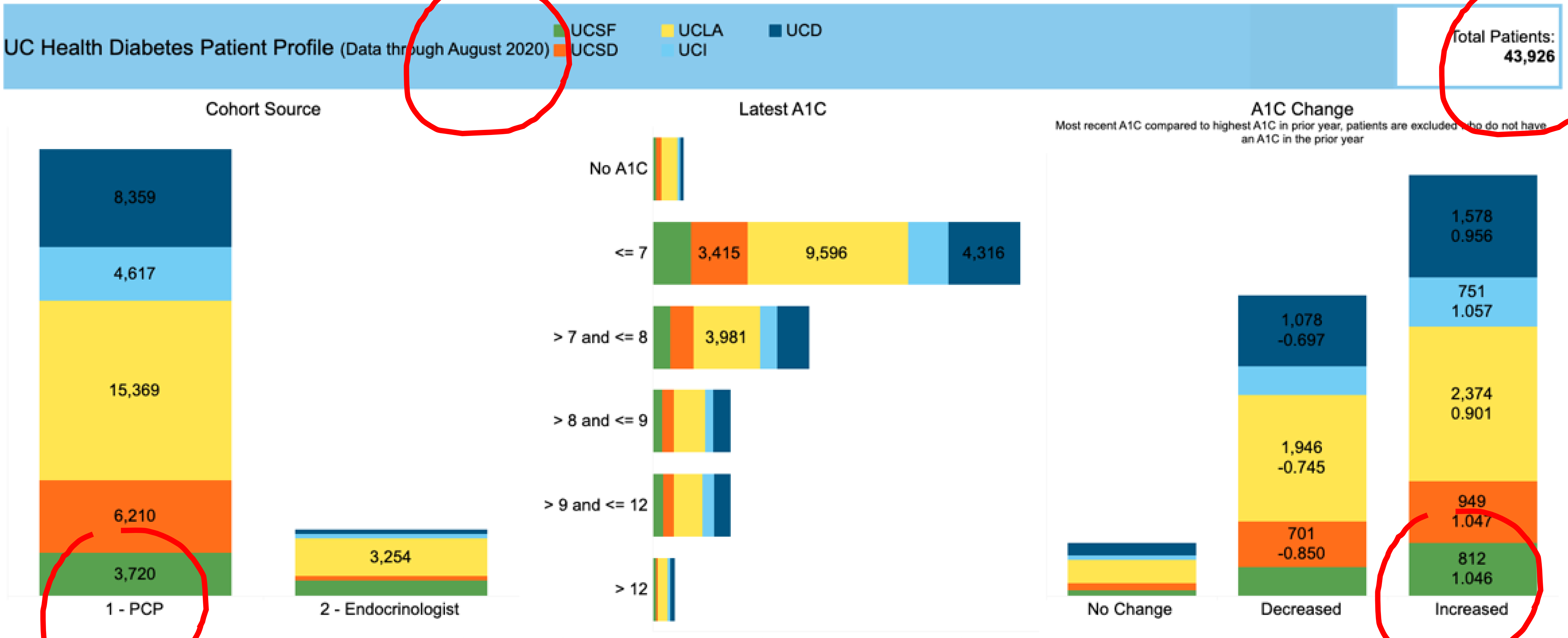
Beau Norgeot

bit.ly/DLHCare

What could we do with clinical data?

- Clinical researcher at UCLA could run a genome wide association study across UC Health
- Mobile health researcher at UCSD can enable patients to contribute data for research
- Community activist and researcher UC Merced can study environmental factors contributing to health and disease
- Transplant patient at UC Irvine can download all their data across UC Health
- Data scientist at UC Santa Barbara can model development of Alzheimer's disease and build a multi-modal predictor
- App designer at UC Riverside can show patients their choices with chronic disease
- CMO at UCSF can build predictive models for readmission, test, share across UC Health
- AI researcher at UC Berkeley can build deep-learning models for image-based diagnostics
- Health services researcher at UC Davis can build predictive models for drug efficacy, and maybe enable pay-for-performance
- Cancer genomics researcher at UCSC can study all our clinical cancer genomes

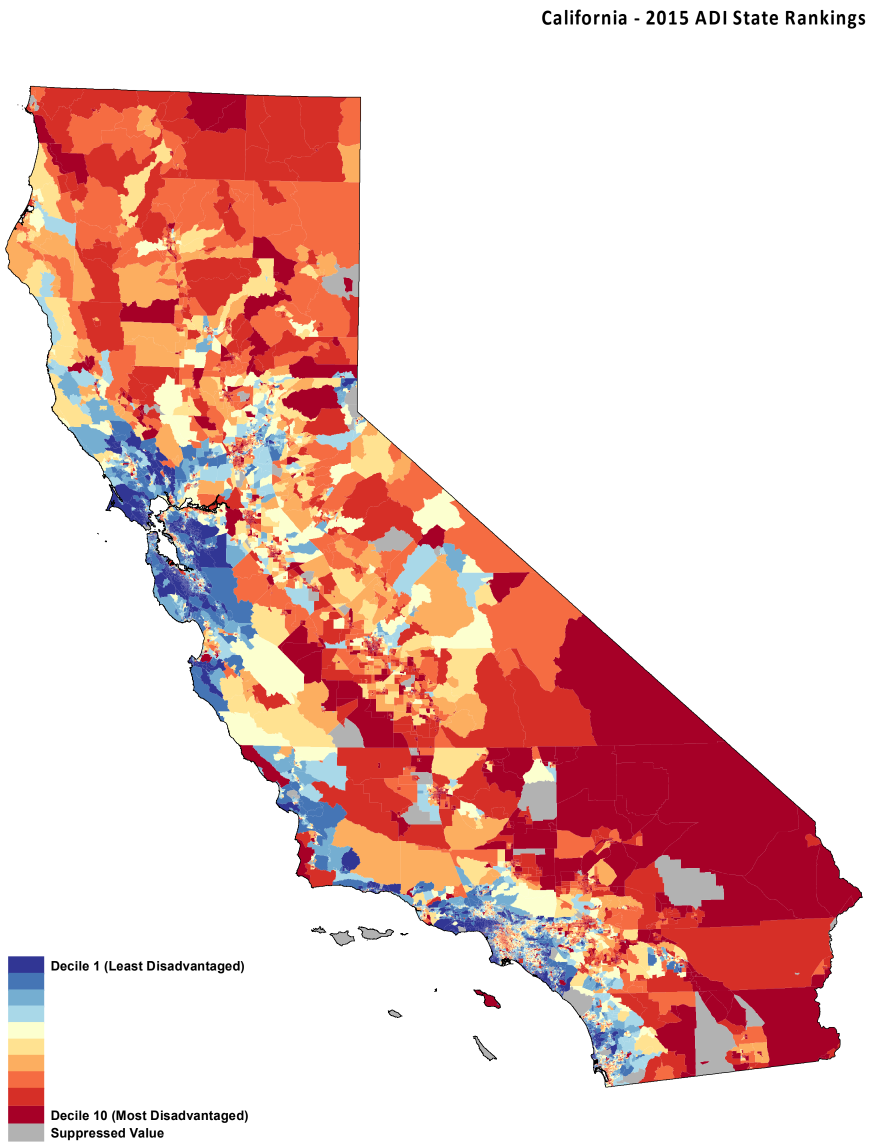
One dashboard for primary care and specialists covers all 44 thousand UC Health patients with type 2 diabetes

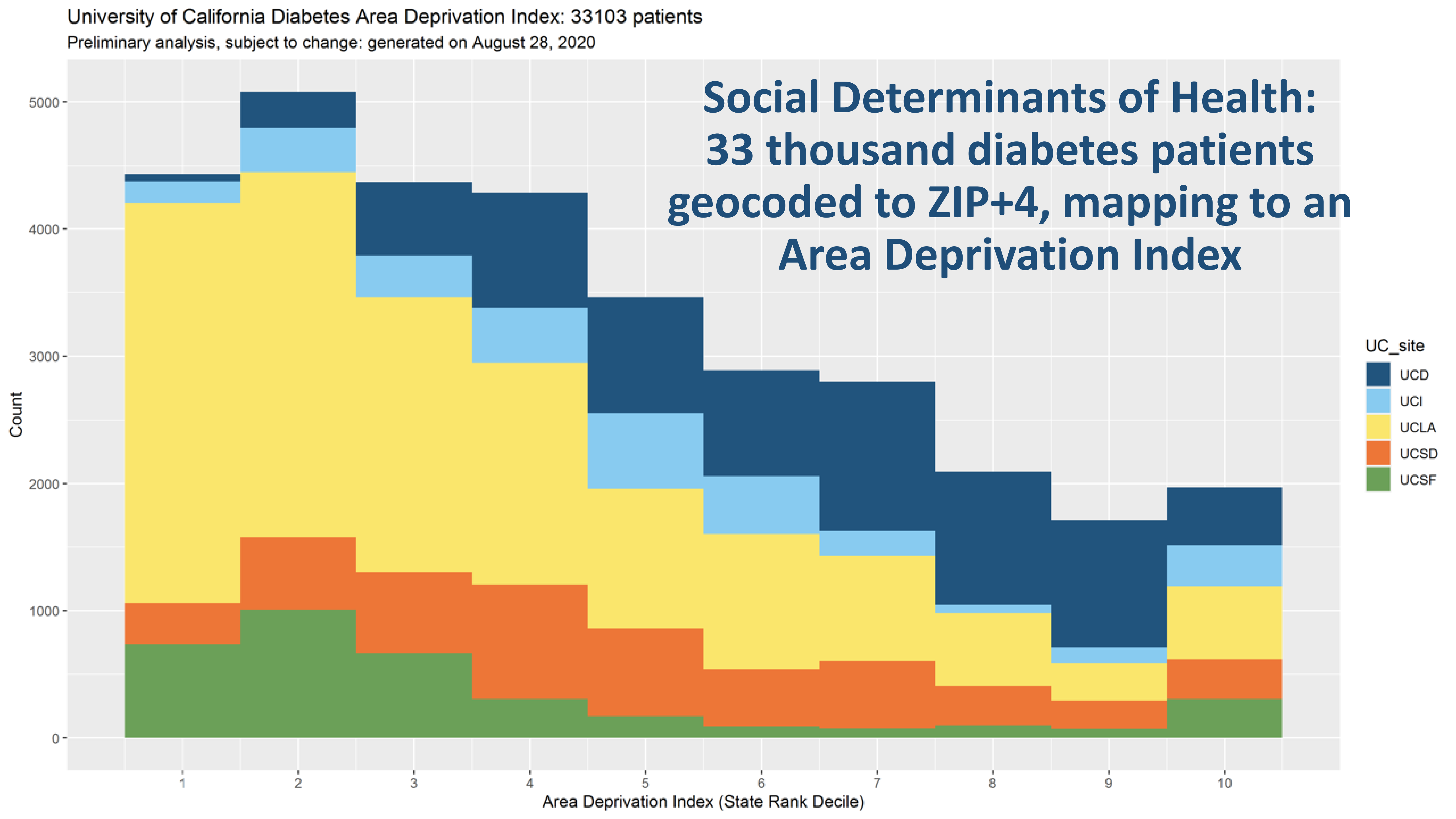


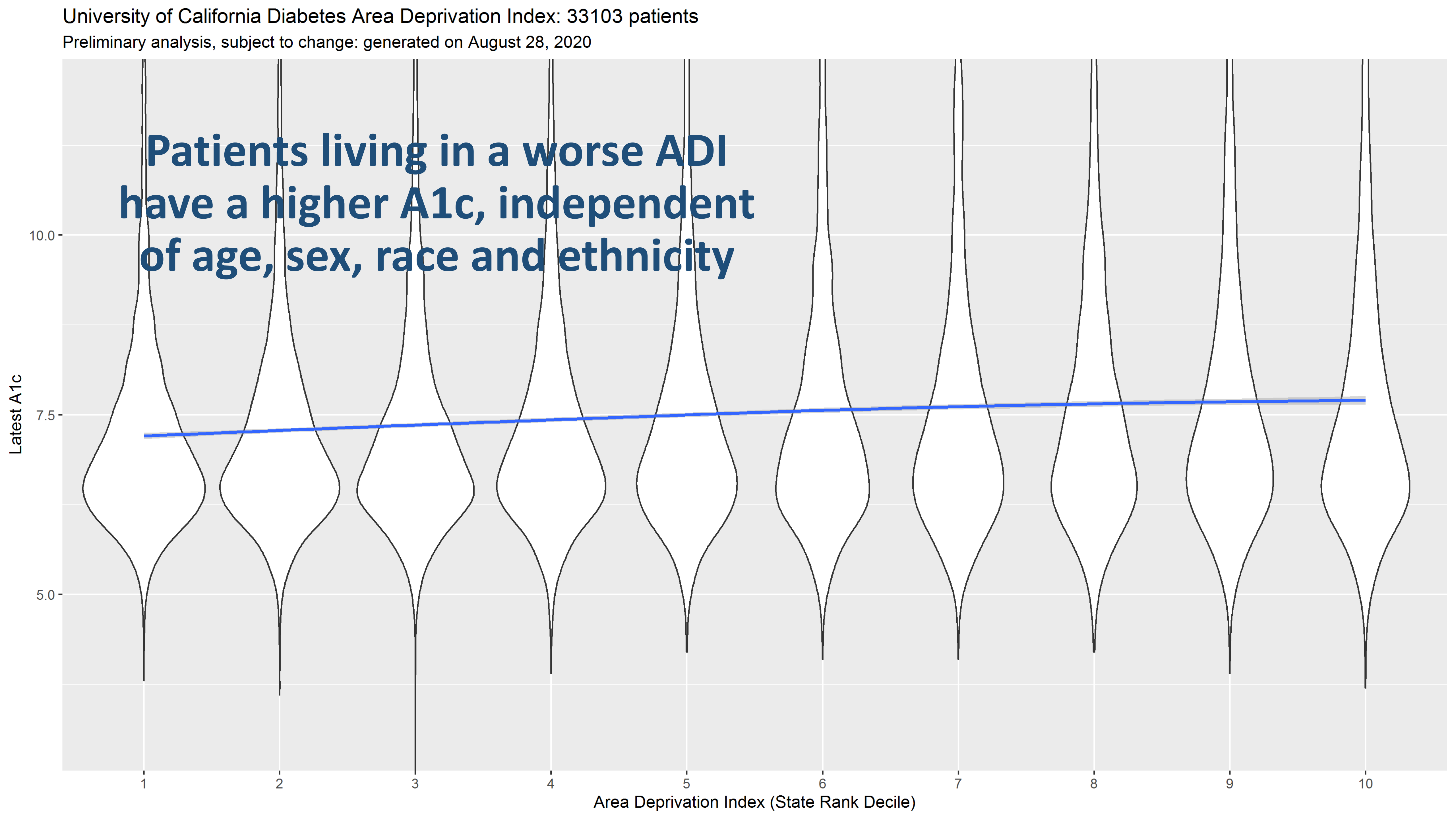
Measure Name	50% IHA Benchmark	UCLA 1 - PCP	UCLA 2 - Endocri..	UCSF 1 - PCP	UCSF 2 - Endocri..	UCSD 1 - PCP	UCSD 2 - Endocri..	UCD 1 - PCP	UCD 2 - Endocri..	UCI 1 - PCP	UCI 2 - Endocri..
Optimal Diabetes Care (A1C control, BP control, Eye Exam, Nephropathy Attention)	18.02%	14.15%	5.32%	21.67%	5.32%	25.71%	5.95%	35.71%	11.04%	21.55%	15.41%
A1C Control (A1C < 8.0%)	61.07%	57.11%	52.23%	57.47%	46.17%	57.38%	35.04%	61.32%	32.79%	55.14%	67.55%

Social Determinants of Health: Area Deprivation Index (ADI)

- Estimate socioeconomic status based on income, education, employment, housing quality at the neighborhood level
- Use with race, ethnicity, gender, age to identify health disparities
- Central geocoding capability to link 9-digit zip code for addresses
- We have now geocoded our UC-wide primary care population
- ADI is significantly associated with adverse health outcomes in our patients, in addition to race, ethnicity, and age




University of California Diabetes Area Deprivation Index: 33103 patients
Preliminary analysis, subject to change: generated on August 28, 2020
Social Determinants of Health:
33 thousand diabetes patients
geocoded to ZIP+4, mapping to an
Area Deprivation Index
Count
0
1000
2000
3000
4000
5000
1
2
3
4
5
6
7
8
9
10
Area Deprivation Index (State Rank Decile)
UC_site
UCD
UCI
UCLA
UCSD
UCSF


University of California Diabetes Area Deprivation Index: 33103 patients
Preliminary analysis, subject to change: generated on August 28, 2020
Patients living in a worse ADI have a higher A1c, independent of age, sex, race and ethnicity
Latest A1c
10.0
7.5
5.0
1
2
3
4
5
6
7
8
9
10
Area Deprivation Index (State Rank Decile)

Working closely with our pharmacist teams, inappropriate use of IV acetaminophen has significantly dropped over the past 2.5 years