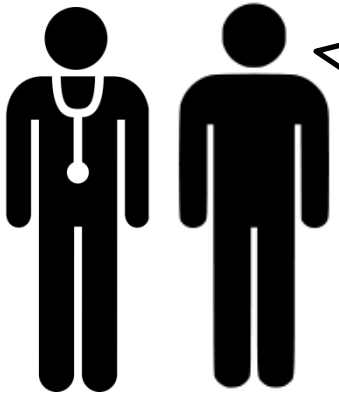
Harnessing Next-Generation Informatics for Personalizing Medicine: Personalization of care and research

Jesse Berlin, ScD
VP and Global Head of Epidemiology, Johnson & Johnson
With thanks to Patrick Ryan, PhD and Martijn Schuemie, PhD
Janssen Research and Development

BASS: 4 November 2014 (Election Day)



Patient's concerns:

- What's wrong with me?
- What should I do about it?
- What's going to happen to me?

Subjective:

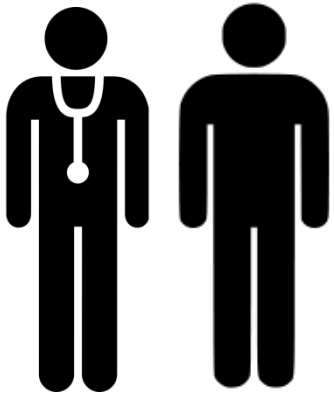
- Chief complaint and other symptoms
- Medical history – what has happened in the past?

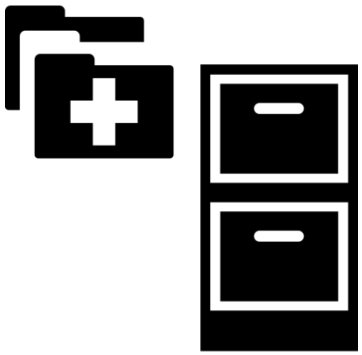
Objective:

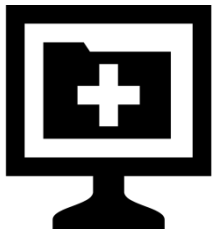
- What can we observe that is happening right now?
- Measurements – vital signs, laboratory tests, radiology/ pathology findings

Assessment and Plan:

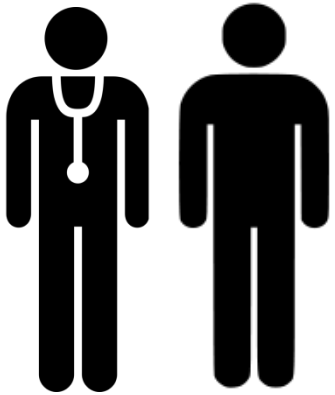
- Diagnosis, based on available information
- Treatment, based on evaluating prognosis for alternative options





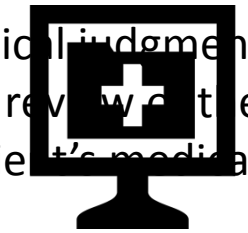


- The subjective, objective, assessment, and plan are all captured in a SOAP (subjective, objective, assessment, and plan) note and recorded in the medical record.
- Medical records are increasingly being captured electronically in electronic health record (EHR) systems
- Data within the EHR are becoming increasingly structured, opening their use in analysis



- Treatment is inherently personalized
- Incomplete external information for decision-making
- Current evidence based on average treatment effects

Information guiding treatment choices



Clinical judgment in the review of the patient's medical data

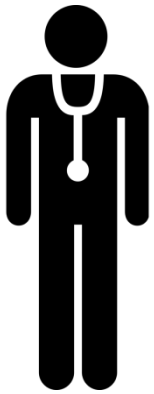
Medical education and treatment guidelines

Evidence about treatment effects from prior clinical research in published literature

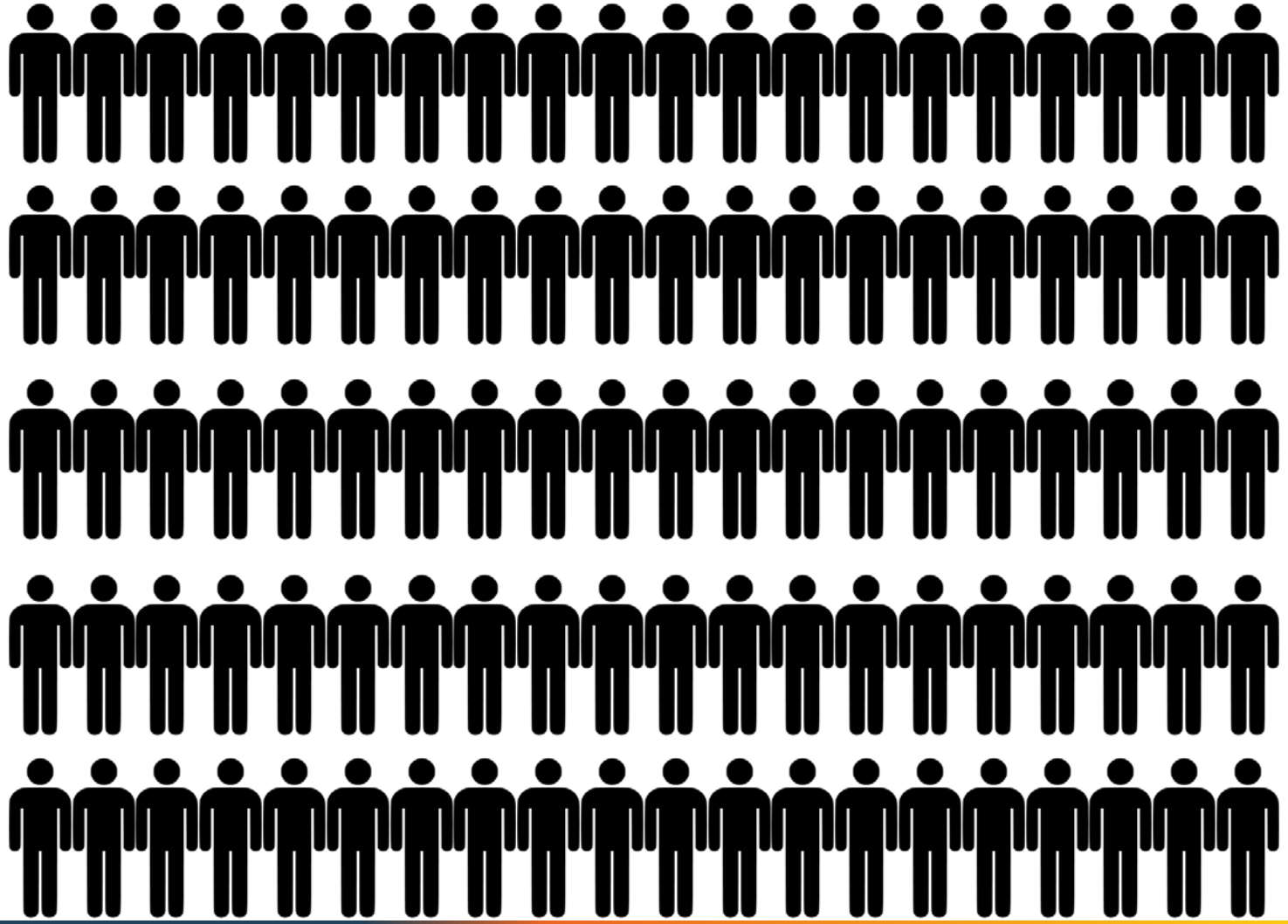
Patient preferences on benefit-risk trade-offs of alternatives

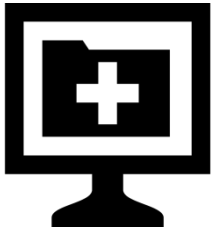
The average primary care physician sees 20 different patients a day, each seeking personalized care

…that’s 100 visits in a week…



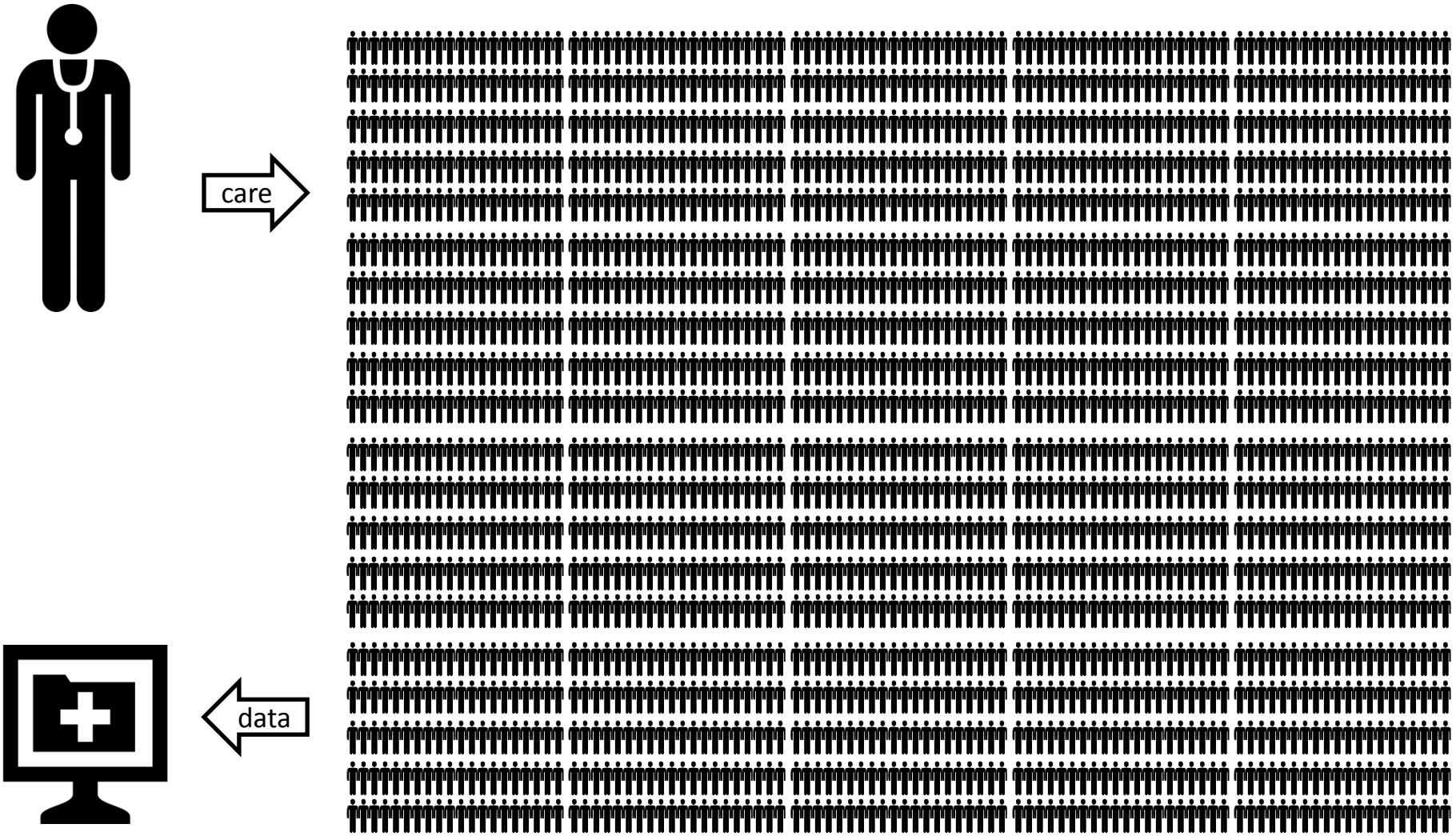
care





data

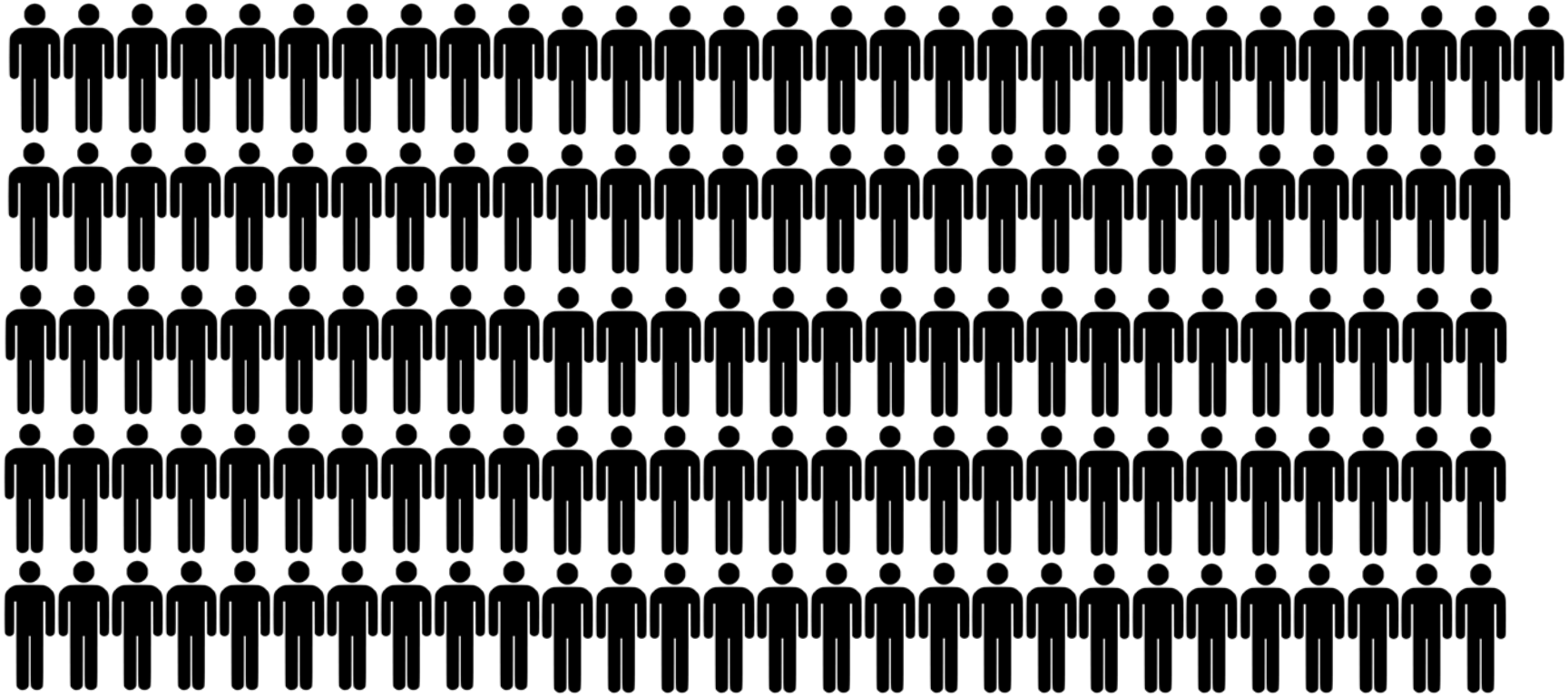
…which can reflect a panel of over 2,000 patients in a year



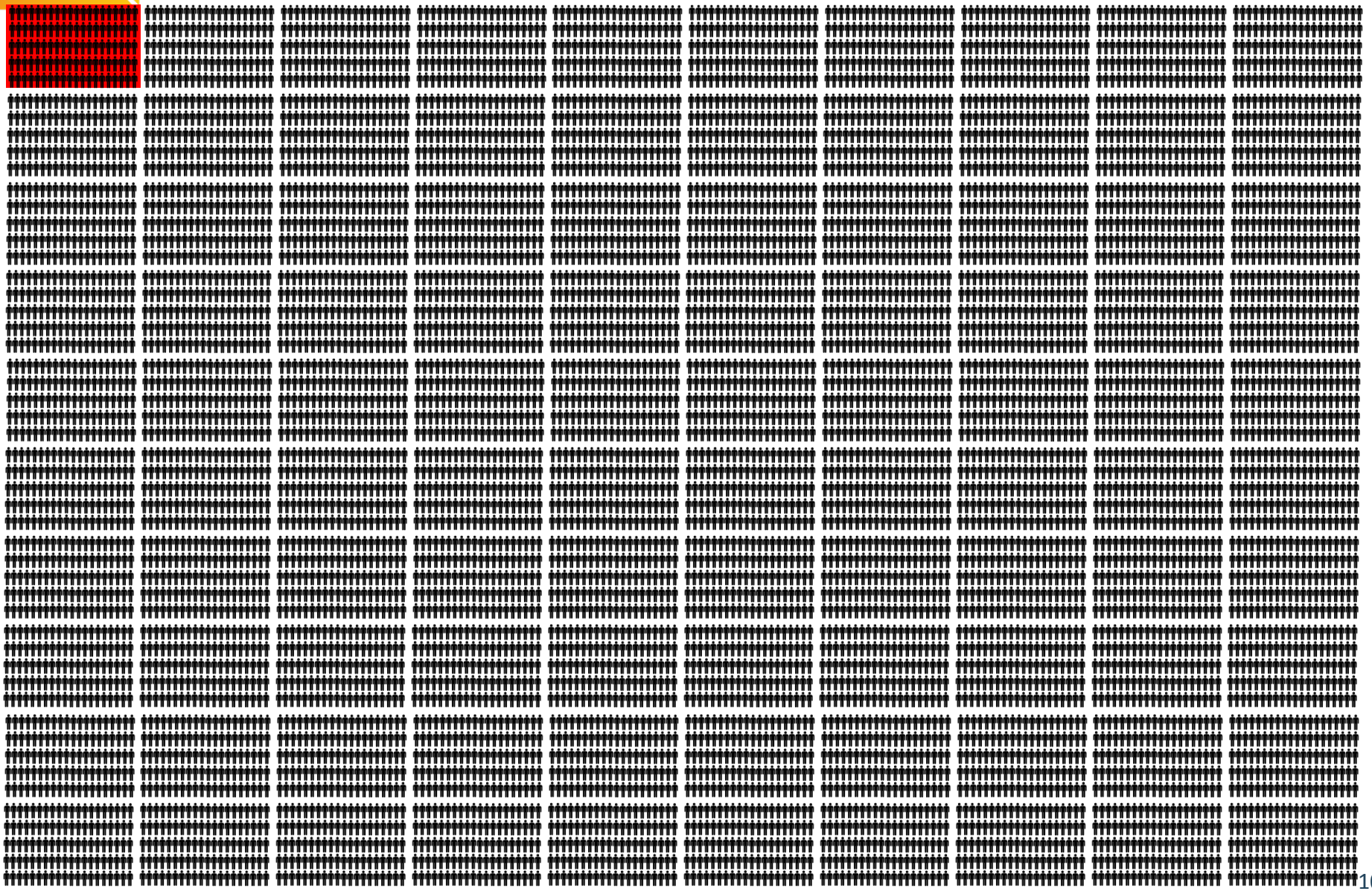
We’re used to the research context

- Larger numbers (depending on the question)
- Multiple doctors and centers
- Goal: Inferences about populations

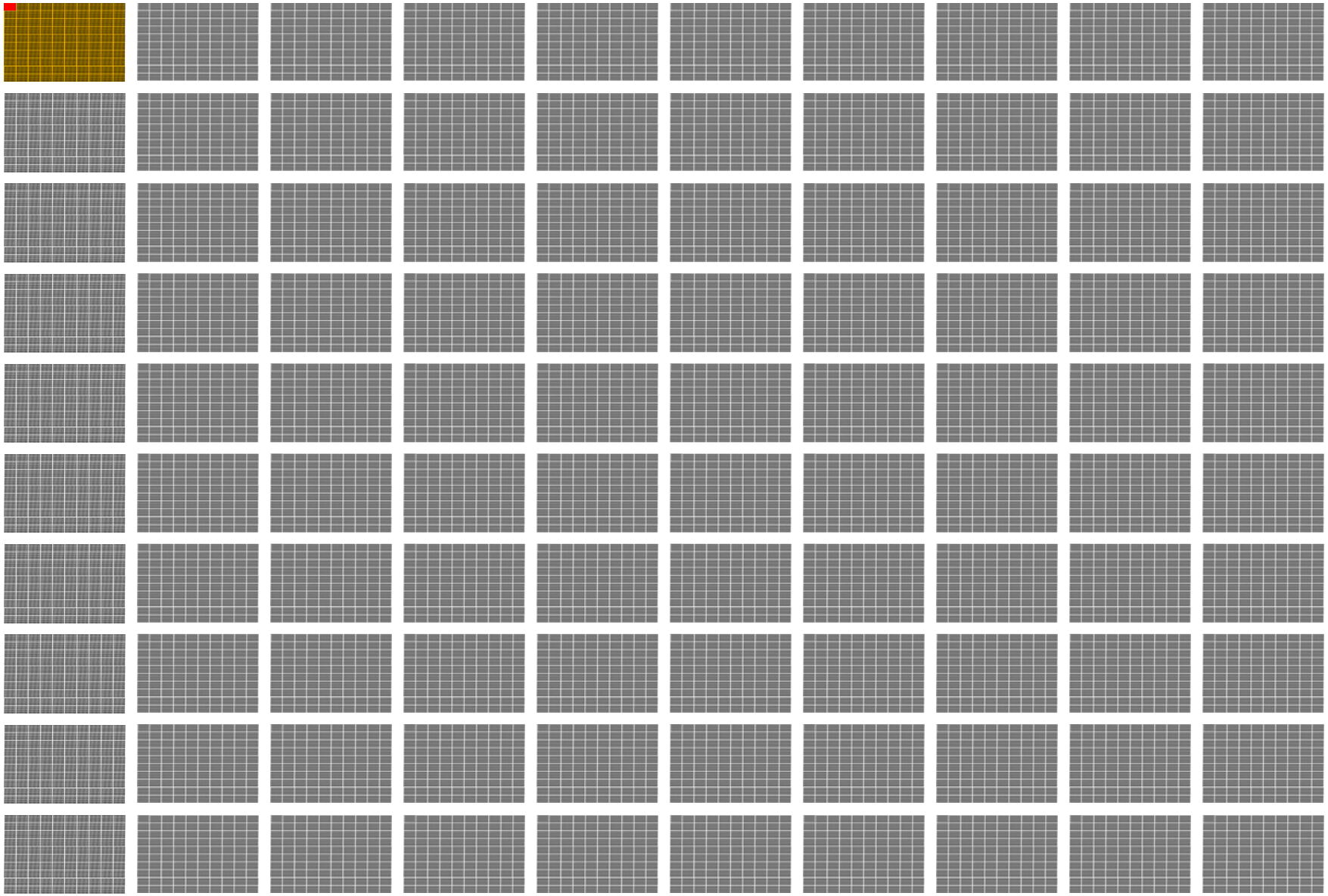
141 patients exposed in THE pivotal study for metformin



>10,000 patients exposed across canagliflozin clinical development program



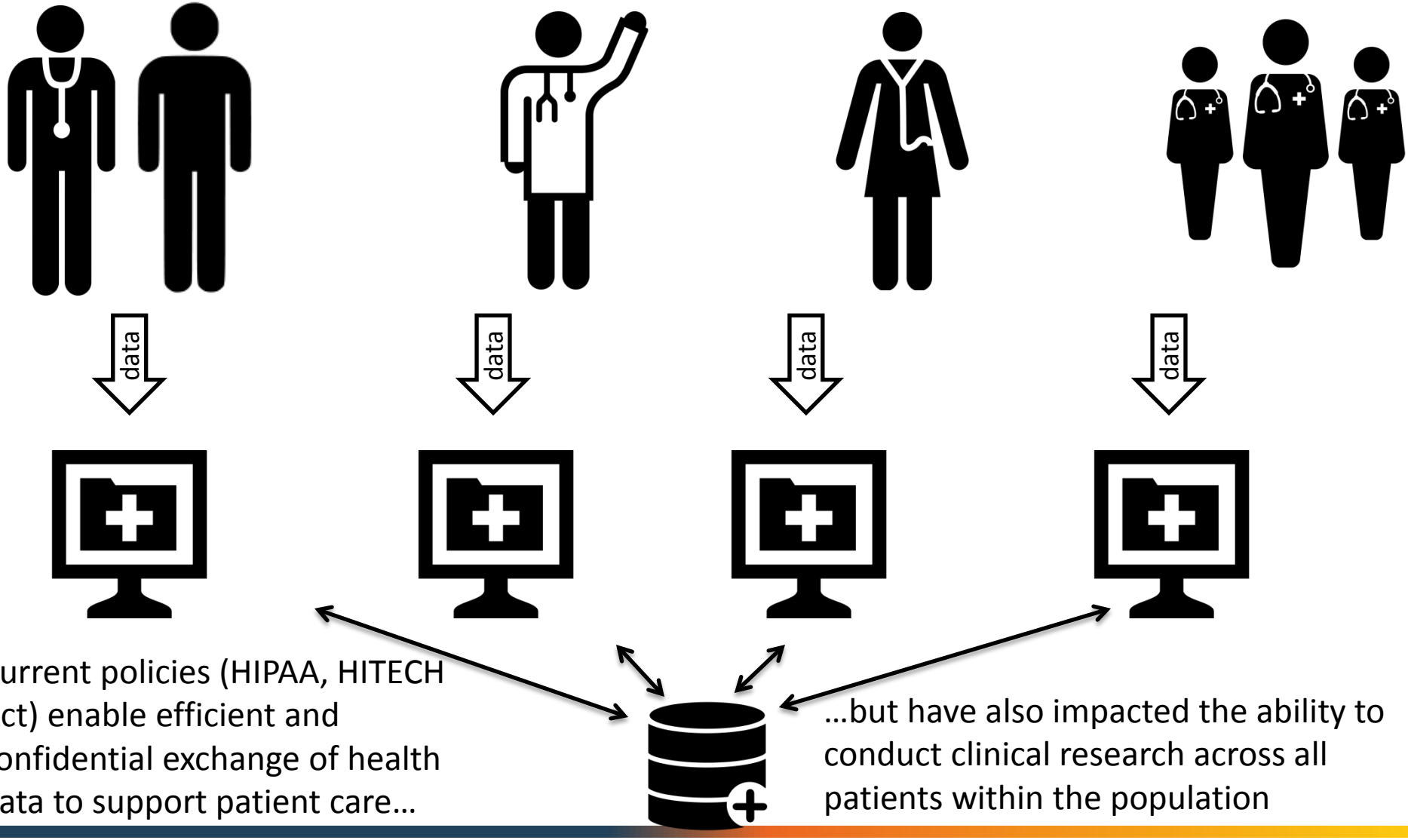
>1,000,000 new users of metformin in one administrative claims database



Let’s get back to the patient

- Different perspective
- What’s going to happen to **me**?

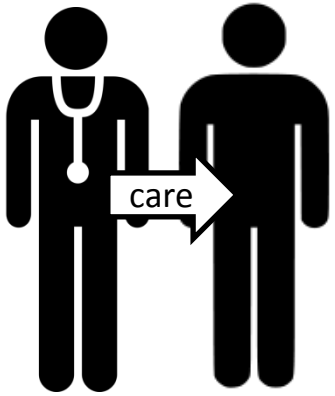
Every patient encounter with every provider generates valuable health data that could improve the quality of patient care

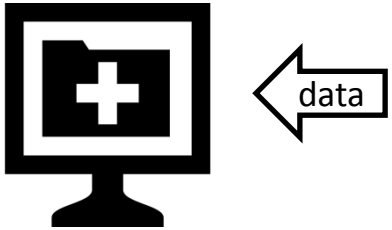


Current policies (HIPAA, HITECH Act) enable efficient and confidential exchange of health data to support patient care…

…but have also impacted the ability to conduct clinical research across all patients within the population

What if real-world evidence could be generated in real-time from the patient-level data across the population to support individual patient care?





- How many other similar patients do we know?
- What treatments were used?
- What outcomes did they experience?

The NEW ENGLAND JOURNAL of MEDICINE

N Engl J Med 2011; 365:1758-1759
November 10, 2011
DOI: 10.1056/NEJMp1108726

Evidence-Based Medicine in the EMR Era

Jennifer Frankovich, M.D., Christopher A. Longhurst, M.D., and Scott M. Sutherland, M.D.

Results of Electronic Search of Patient Medical Records (for a Cohort of 98 Pediatric Patients with Lupus) Focused on Risk Factors for Thrombosis Relevant to Our 13-Year-Old Patient with Systemic Lupus Erythematosus.*

Outcome or Risk Factor	Keywords Used to Conduct Expedited Electronic Search	Prevalence of Thrombosis	Relative Risk (95% CI)
		no./total no (%)	
Outcome — thrombosis	"Thrombus," "Thrombosis," "Blood clot"	10/98 (10)	Not applicable
Thrombosis risk factor			
Heavy proteinuria (>2.5 g per deciliter)			
Present at any time	"Nephrosis," "Nephrotic," "Proteinuria"	8/36 (22)	7.8 (1.7–50)
Present >60 days	"Urine protein"	7/23 (30)	14.7 (3.3–96)
Pancreatitis	"Pancreatitis," "Lipase"	5/8 (63)	11.8 (3.8–27)
Antiphospholipid antibodies	"Aspirin"	6/51 (12)	1.0 (0.3–3.7)

* In all cases, the sentences surrounding the keywords were manually reviewed to determine their relevance to our patient. Pancreatitis was defined as an elevated lipase level (twice the upper limit of normal) coexisting with abdominal pain. We used the word "aspirin" as a proxy for antiphospholipid antibodies, since it is standard practice at our institution to give all patients with these antibodies aspirin; if "aspirin" was found in the chart, than antiphospholipid-antibody status was confirmed by investigating the laboratory results.

enabled data analysis, we made the decision to give our patient anticoagulants within 24 hours after admission.

Our case is but one example of a situation in which the existing literature is insufficient to guide the clinical care of a patient. But it illustrates a novel process that is likely to become

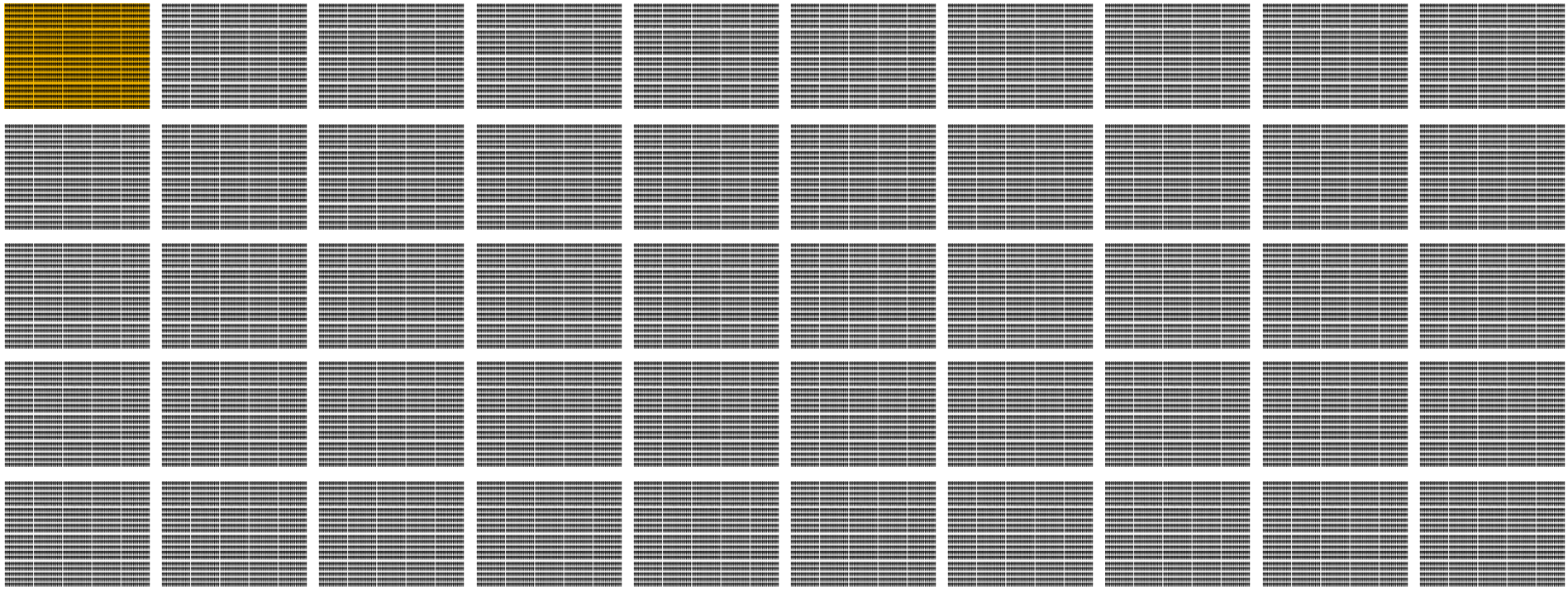
sion making has already transformed other industries,[4] and the growing prevalence of EMRs along with the development of sophisticated tools for real-time analysis of deidentified data sets will no doubt advance the use of this data-driven approach to health care delivery. We look forward to a future in which health information

by intelligence."[5] In the practice of medicine, one can't do better than that.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the Division of Rheumatology (J.F.), the Division of Systems Medicine (C.A.L.), and the Division of Nephrology (S.M.S.), Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA.

Patient-level predictions for personalized evidence requires big data



Aggregated data from a large medical group of 50 providers may contain 100,000 patients

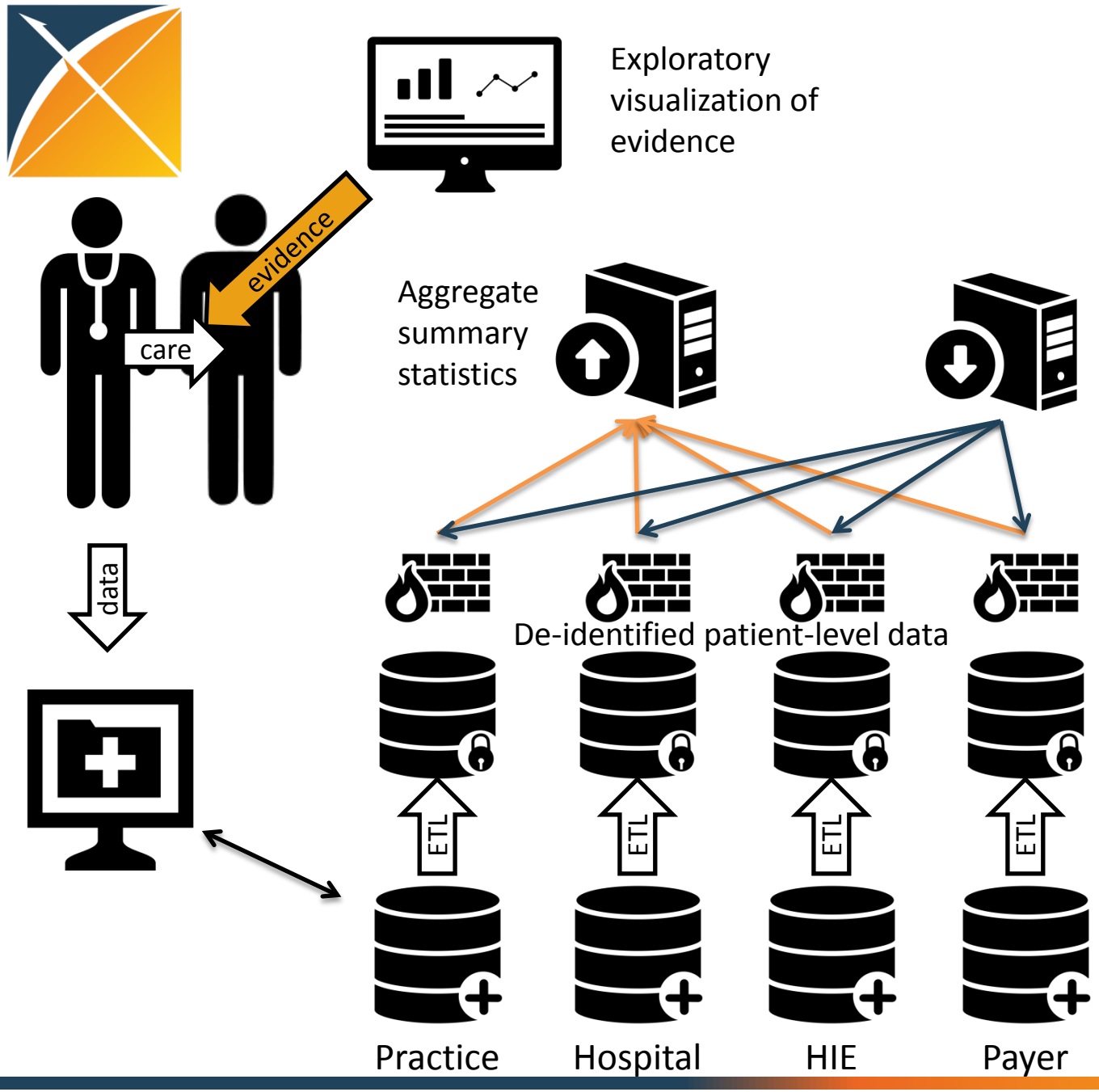
Patient-level predictions for personalized evidence requires big data

2 million patients seem excessive or unnecessary?

- Imagine a provider wants to compare her patient with other patients with the same gender (50%), in the same 10-year age group (10%), and with the same comorbidity of Type 2 diabetes (5%)
- Imagine the patient is concerned about the risk of ketoacidosis (0.5%) associated with two alternative treatments they are considering
- With 2 million patients, you'd only expect to observe 25 similar patients with the event, and would only be powered to observe a relative risk > 2.0

Aggregated data across a health system of 1,000 providers may contain 2,000,000 patients

One informatics approach, being developed by OHDSI



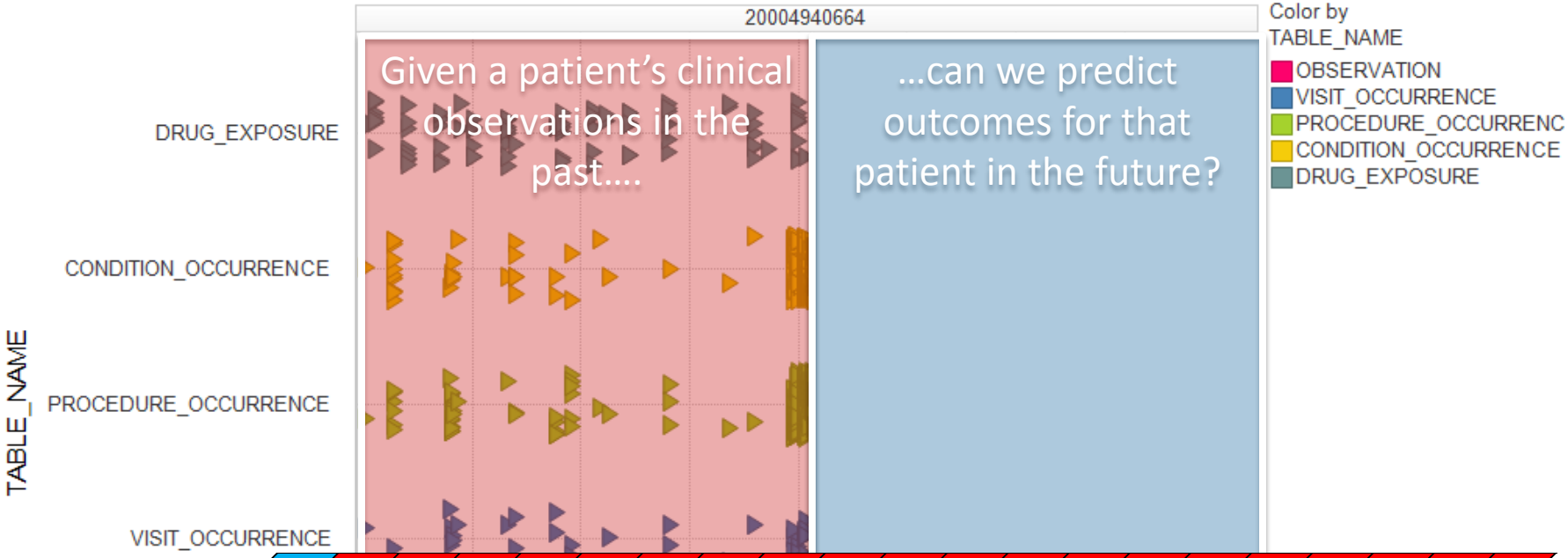
Standardized Analytics:
- Characterization
- Estimation
- Prediction

Standardized Structure:
OMOP
Common Data Model

Standardized Content:
Meaningful Use
vocabularies (SNOMED, RxNorm, LOINC)

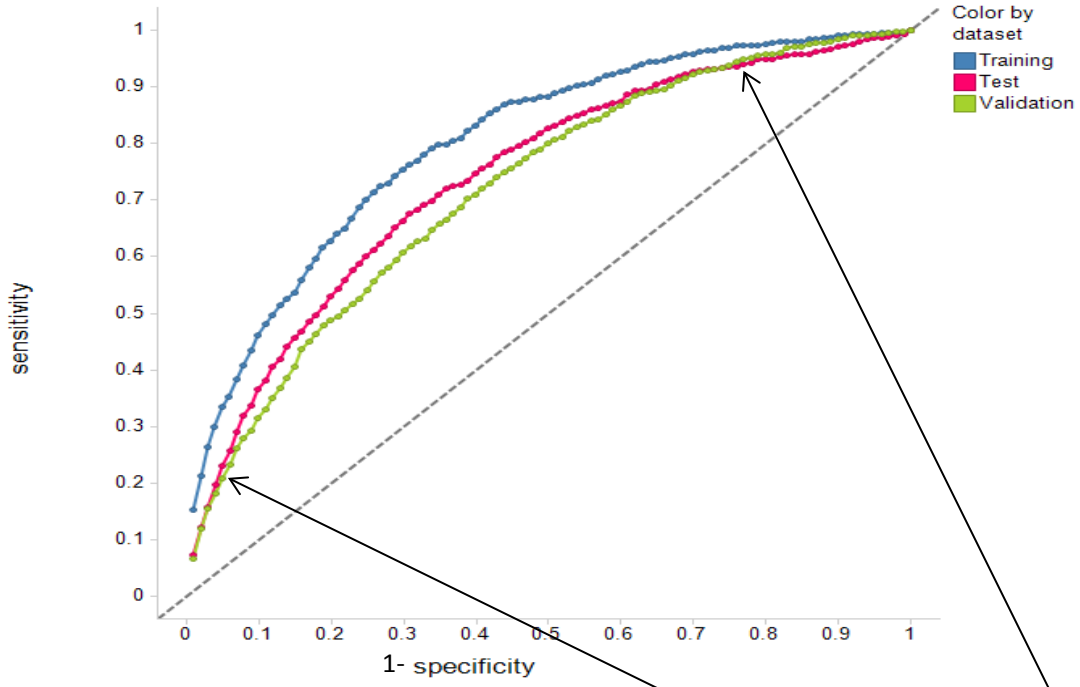
Identifiable patient-level data

Large-scale analytics can help reframe the patient-level prediction problem



Outcome	Age	Gender	Race	Location	Drug 1	Drug 2	…	Drug n	Condition 1	Condition 2	…	Condition n	Procedure 1	Procedure 2	…	Procedure n	Lab 1	Lab 2	…	Lab n
0	76	M	B	441	0	0	1	1	1	1	1	1	0	1	0	1	1	0	1	1
1	77	F	W	521	1	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0
1	96	F	B	215	1	1	1	0	1	0	1	1	1	0	0	0	1	1	0	1
1	76	F	B	646	0	1	0	0	1	0	1	1	0	0	0	0	1	0	1	0
0	64	M	B	379	0	0	1	1	1	1	0	1	1	1	0	1	0	0	0	0
1	74	M	W	627	0	1	1	1	0	0	1	1	0	1	1	1	1	0	0	1
1	68	M	B	348	0	0	0	1	0	0	1	0	1	1	1	0	1	1	0	1
	Demographics				All drugs				All conditions				All procedures				All lab values			

Example: Among patients with diabetes, can we predict who will have short-term complications?

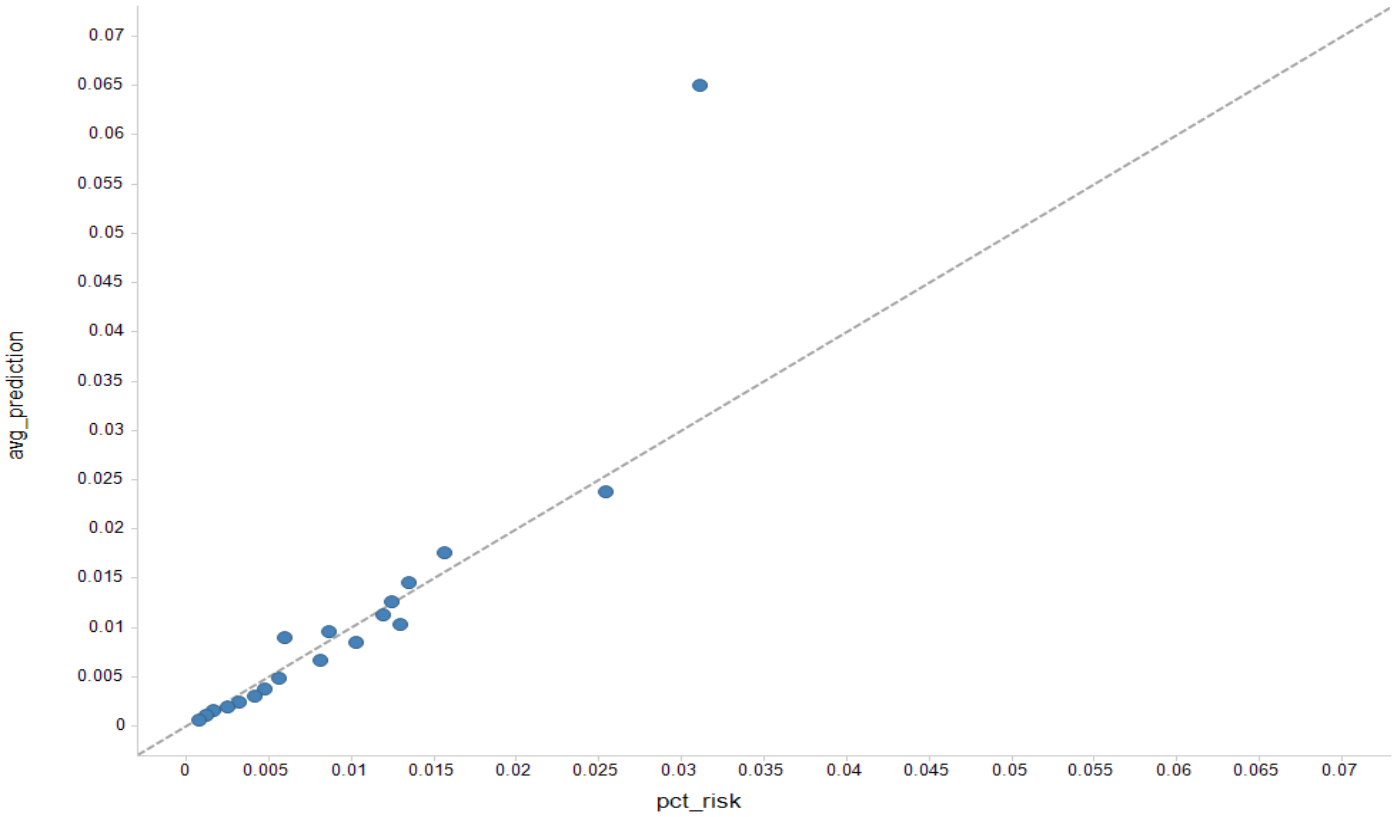


Dataset	AUC	Sensitivity at p05	Prediction at p05	False negative rate at p80	Prediction at p80
Training	**0.802**	0.34	0.013	0.02	0.001
Test	**0.741**	0.23	0.014	0.05	0.001
Validation	**0.720**	0.21	0.013	0.04	0.001

Regularized logistic regression, n=185k, p=10k

Wang et al., ADA, 2014

Predicted probability demonstrate strong calibration in validation set

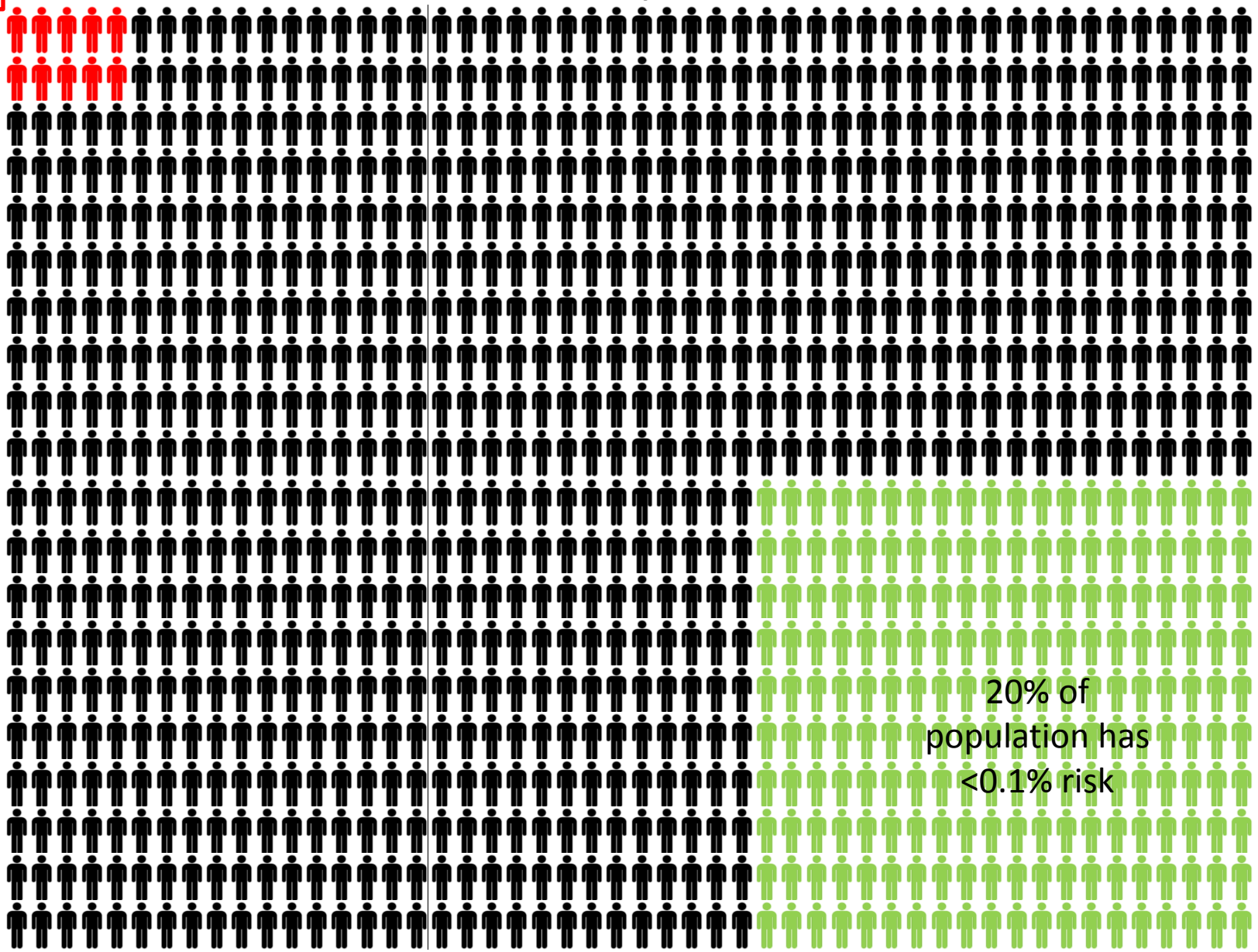


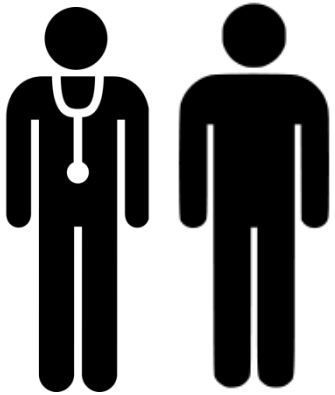
- Model is well-calibrated with predicted probability near true risk in all but top 1%, where model overestimates risk

Wang et al., ADA, 2014

Translating predictive model into public health impact

1% of population has >3% risk

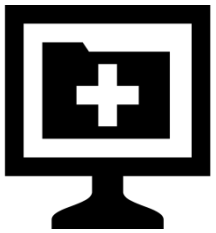




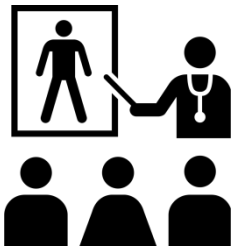
Next-generation informatics is already here

OHDSI: Real-world evidence from data characterization, population-level estimation, and patient-level prediction

Information guiding treatment choices



Clinical judgment in the review of the patient's medical data



Medical education and treatment guidelines

Evidence about treatment effects from prior clinical research in published literature

Patient preferences on benefit-risk trade-offs of alternatives

Concluding thoughts

- Patients deserve personalized evidence to improve the quality of their care
  - Personalized evidence for one patient requires use of data from all patients
  - Current policies speak to access for the patient under care, but preclude access to other patient's data
- Patient-level predictions for personalized evidence requires big data
  - Accuracy and precision will be impacted by size of available population
  - Policies need to support infrastructure to enable aggregation of patient-level data across providers, health systems, and payers
- Personalized evidence doesn't (necessarily) require exposing patient-level data
  - Some patient-level predictive models can be trained on patient-level data, but applied using only aggregate statistics
  - Policies should recognize trade-off between improving quality of evidence vs. protecting patient privacy
- Establishing the reliability of the real-world evidence is a necessary pre-requisite for a learning health system
  - Real-world evidence should complement, not replace, current sources of information in supporting medical decision-making and should only be used when shown to be appropriate
  - Policy needs to encourage more methodological research to establish appropriate statistical techniques to address sources of bias that plague observational analyses