REPRESENTATION OF DIVERSE GROUPS IN TEST SETS
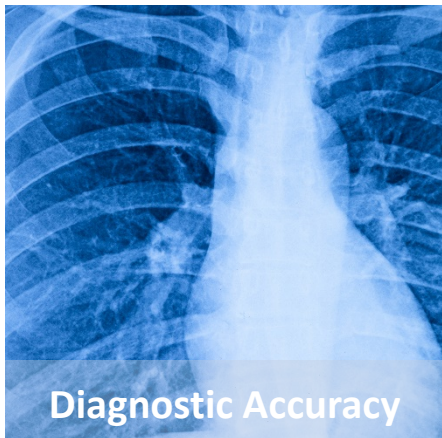
Terri L. Cornelison, M.D., Ph.D.
Director and Chief Medical Officer
Health of Women Program
Center for Devices and Radiological Health

Data are the most important ingredient for training AI/ML algorithms

"Worldview" could be narrow in focus if data do not represent a diverse set of patients

Advances in AI/ML Transforming
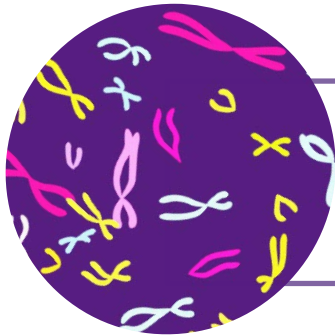## *Health Systems and Daily Lives*

Diagnostic Accuracy

Intervention Response

Predict Prognosis

Advances in AI/ML Transforming
Health Systems and Daily Lives

Diagnostic Accuracy

Intervention Response

Predict Prognosis

Sex
Gender
Age
Race
Ethnicity

Attributes

Sex, Age, Race

Alter physiology at the molecular, cellular, and macro-organism level

Attributes

Gender, Ethnicity — **Affect behavior, perception, health**

Sex, Age, Race — **Alter physiology at the molecular, cellular, and macro-organism level**

Attributes

Constitute relevant sources of variation in a number of clinical and subclinical conditions

Affect risk factors, prevalence, age of onset, symptomatology, prognosis, biomarkers, treatment effectiveness
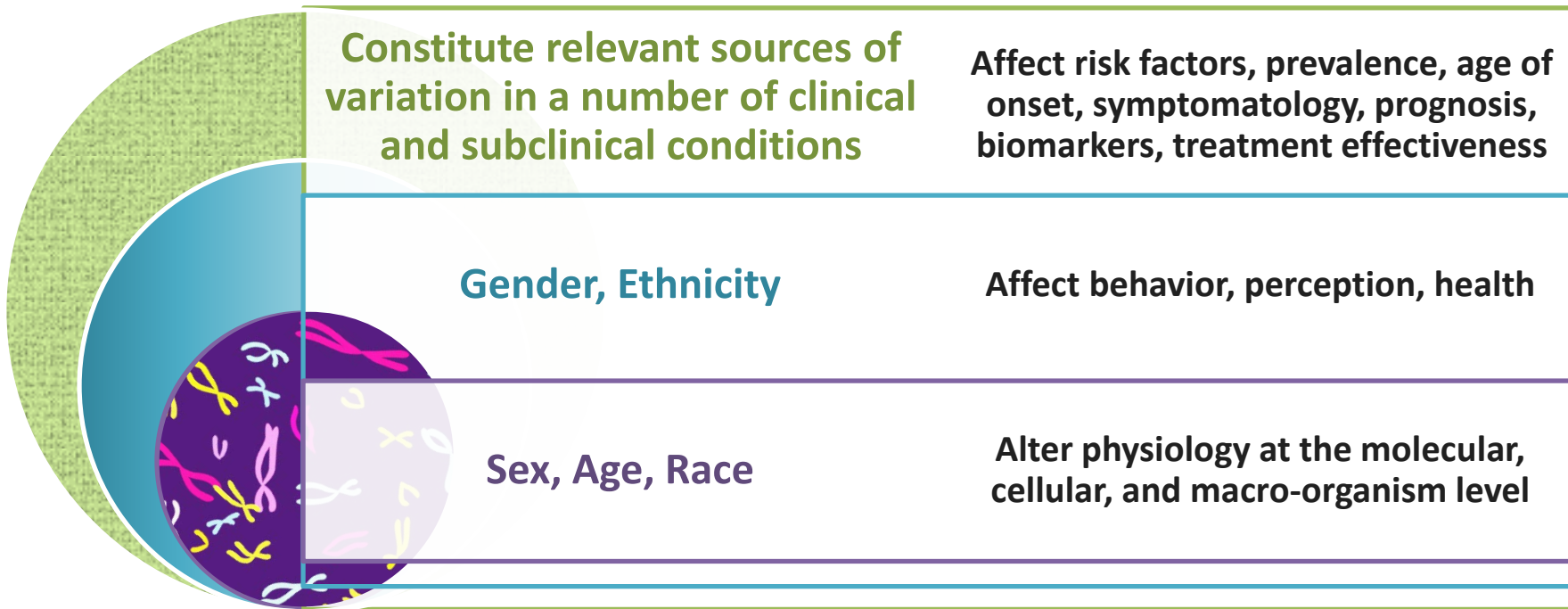
Gender, Ethnicity

Affect behavior, perception, health

Sex, Age, Race

Alter physiology at the molecular, cellular, and macro-organism level

Attributes

Sex, Gender

Differences reported in cardiovascular disease, pulmonary dysfunction, neurological debility, irritable bowel syndrome, endocrine and autoimmune disorders, mental illness

May influence disease course and outcome in all organ systems of the body

Age

Older patients and pediatric patients with age-specific co-morbidities, concomitant therapies, or development considerations that impact health

Race, Ethnicity

Racial and ethnic groups experience different mortality rates for many health conditions

Global Patterns of Gene Expression

Differ in Males and Females

FDA

Although all sex differences stem from X and Y genes

Sex hormones cause sexual bias in gene expression by acting directly on genes *throughout the genome*

50-75% of active genes are sex-biased
72% liver, 68% fat, 55% muscle
Mean difference of 8-9%

Yang et al. *Genome Res* 2006
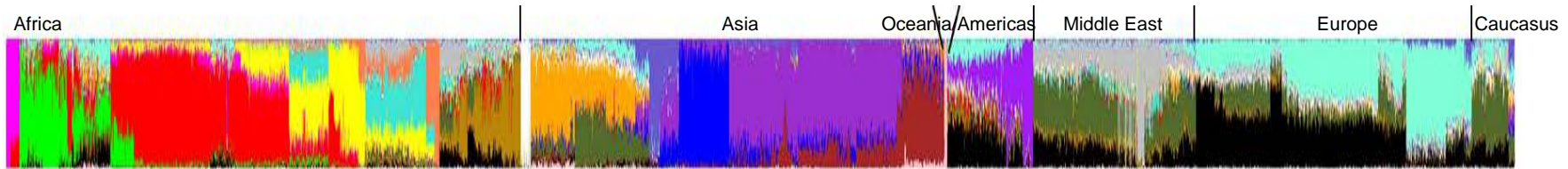
Age

- Reparative capacity

- Physiology

- Disease manifestations

- Disease prevalence

- Developmental stages
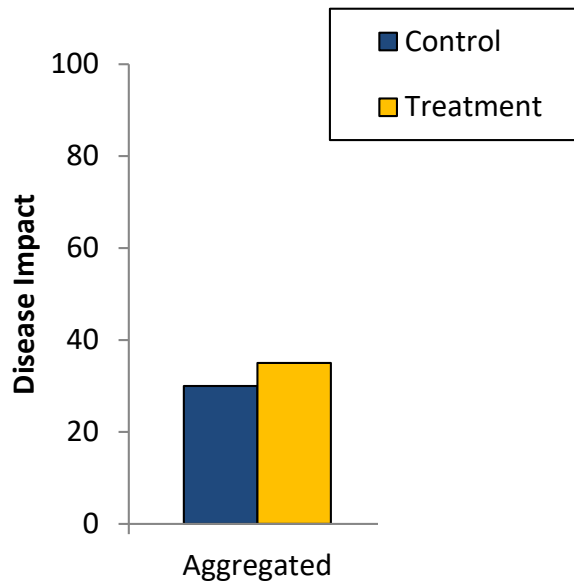
- Behaviors and preferences

Race

Human Biogenetic Variation

Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry

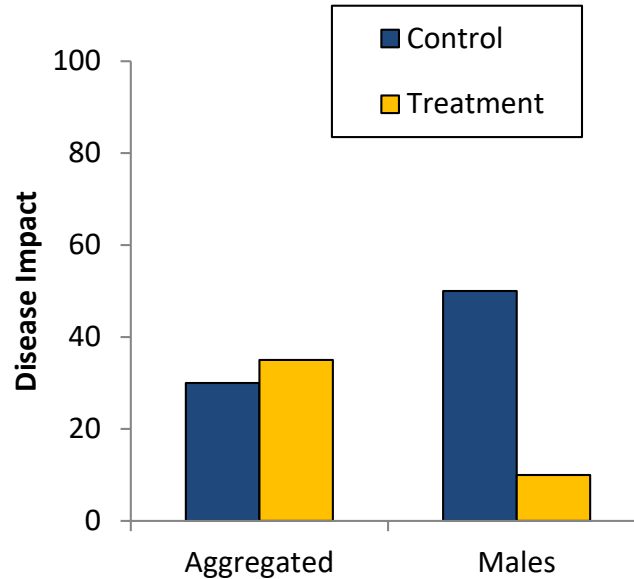| Africa | Asia | Oceania/Americas | Middle East | Europe | Caucasus |

- Investigated ancestry of 3,528 modern humans
- (163 ethno-linguistic groups); 19 ancestral components
- **94.4% of individuals showed mixed ancestry**
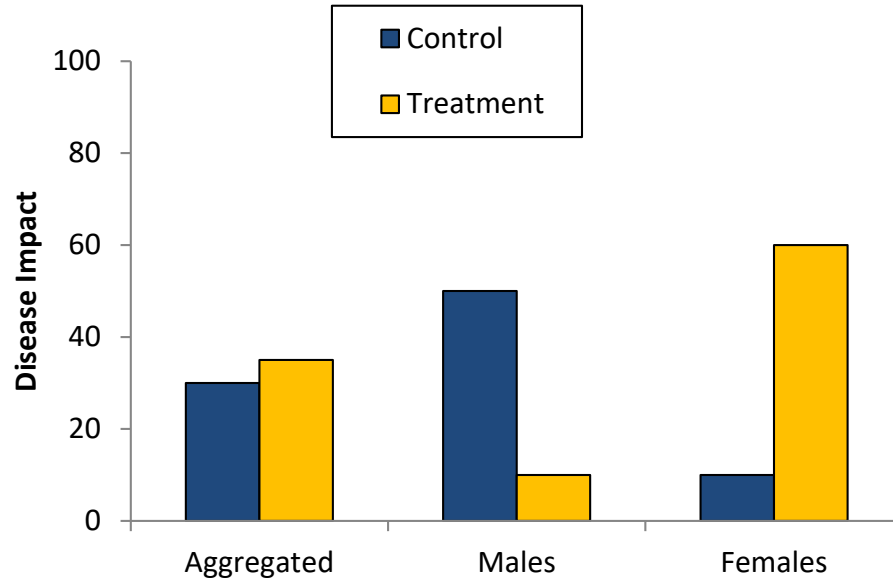- Emphasizes importance of accounting for ancestry in history, forensics, health

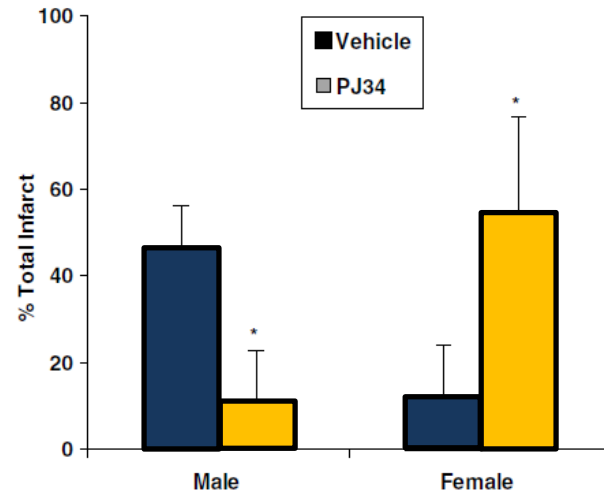*Shriner D, Tekola-Ayele F, Adeyemo A, Rotimi C. Scientific Reports, 2014.*

11

Impact of Experimental Design

Tabak L, *Workshop on Methods and Techniques for Integrating the Biological Variable "Sex" in Preclinical Research*, NIH, 2014

Impact of Experimental Design

13

Impact of Experimental Design

Tabak L, *Workshop on Methods and Techniques for Integrating the Biological Variable "Sex" in Preclinical Research*, NIH, 2014

14

Real Life

The effects of the selective poly-ADP ribose polymerase **(PARP-1) inhibitor** PJ-34 in wild-type (WT) mice of both genders. Treatment with PJ-34 at ischemic onset **reduced total infarction in male mice** compared with saline-treated controls (* $P<0.001$). A significant **increase in ischemic damage was seen in PJ-34-treated females** compared with control (* $P<0.001$).

McCullough LD, et al. *Journal of Cerebral Blood Flow & Metabolism* 2005

Keystone

Include all populations, including underrepresented populations, in our clinical trials and our data sets

Analyze the data disaggregated by subgroup(s) to better inform the science and refine artificial intelligence algorithms to perform best in all populations for which intervention is intended
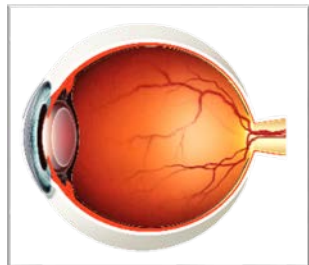
Improving data quality, strengthening the science, enriching patient information
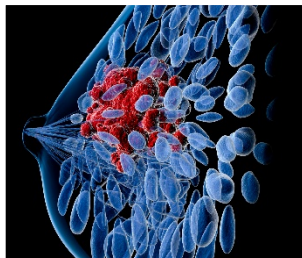
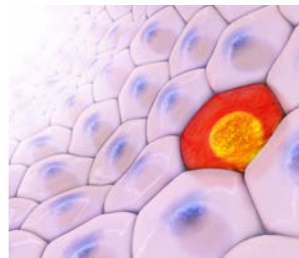# Diagnostic Tools
Visual Input Data

| Diabetic Retinopathy	Mammography	Melanoma	Pulse Oximetry
Retinal pigment may influence ability to discern abnormal capillaries or micro-lesions	Breast density is influenced by **age, genetics, body habitus, parity, estrogen use, and menstrual cycle phase**	**Lack of the full spectrum of skin phenotypes and lesions** in the training datasets is a limitation of these techniques	Undesired bias occurred with the device itself which showed errors associated with **sex and skin color**

Diagnostic Tools
ECG and **Cardiovascular Disease**

Age: False positives more common in young or thin individuals, whose voltage may exceed conventional thresholds

Co-morbidities: False negatives with right bundle branch block (RBBB), obesity, or chronic obstructive pulmonary disease

Sex: Sensitivity reduced in women, sex-based difference in pathophysiology
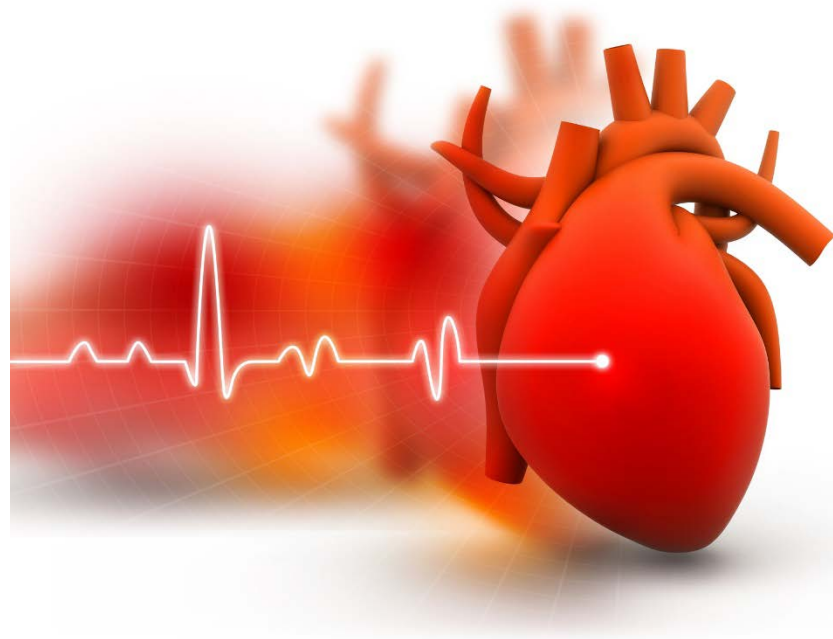
Diagnostic Tools
MRI in **Parkinsonian Syndromes**

Sex-based differences in neural anatomy

- **Broad regions of the cerebral cortex are thicker in women than in men and ratios of grey to white matter also differ**

- Some structures (hippocampus) are larger in the female brain, others (amygdala, hypothalamus) are larger in the male brain relative to cerebral size

- Sex-based differences in cognitive loci such as the amygdala, hippocampus, neocortex; affecting brain morphology and neurocognitive function

Digital Biomarkers

Early Detection of **Alzheimer's** and **Parkinson's**

Physiological, psychological and behavioral indicators based on data collected by portable, wearable, implantable, ingestible devices – Facilitate diagnosis, assessment of treatment and predicted prognosis

Sex and gender differences in these indices of health and disease have not yet been examined, and not fully examined for age, race, ethnicity

- *Significant **sex differences** on neurodegenerative, physiological and cognitive aspects during the preclinical stages of **Alzheimer's***

- *Study assessing digital biomarkers for **Parkinson's**, 18.6% were women; **skews diagnosis** towards symptoms found more in males (rigidity and rapid-eye movement) than in females (dyskinesias and depression)*

Prognostic Tools

Useful for estimating disease severity and survival
Serve as helpful medical decision-making tools for guiding patient care

Survival estimates in advanced terminal cancer	**Help guide conversations and decisions about treatment and end of life preferences**
Predictive scoring systems in the intensive care unit	**Identify patterns that trigger early intervention in critically ill patients and guide care**
Prognostic models for patients with cirrhosis	**For prioritization of patients awaiting liver transplantation**

Although machine learning algorithms hold promise, **algorithms are best when derived from all populations, and when they account for variables such as sex, gender, age, race and ethnicity**.

Bias and Trust

The use of historical data to train algorithms may lead to inaccurate conclusions in different racial and ethnic groups

Human biases in decision making may be inadvertently introduced into the algorithms

Medical records of vulnerable groups might be poorly collected or digitized resulting in sample size disparity

Validating algorithms in diverse data sets promotes trust

Closing

AI and ML statistical methods provide extraordinary opportunity for improving the efficiency, effectiveness and efficacy of health care delivery

Help ensure that diverse patient demographics and the full spectrum of disease are adequately represented in AI algorithms and ML paradigms for the population on which the device will be used

"Worldview" includes you because the data represent you