

Machine Learning in Population and Public Health: Challenges and Opportunities

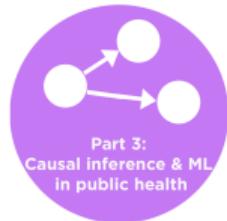
Rumi Chunara, Yuan Zhao, Vishwali Mhasawade

ACM CHIL 2020 Tutorial



Tutorial Goals

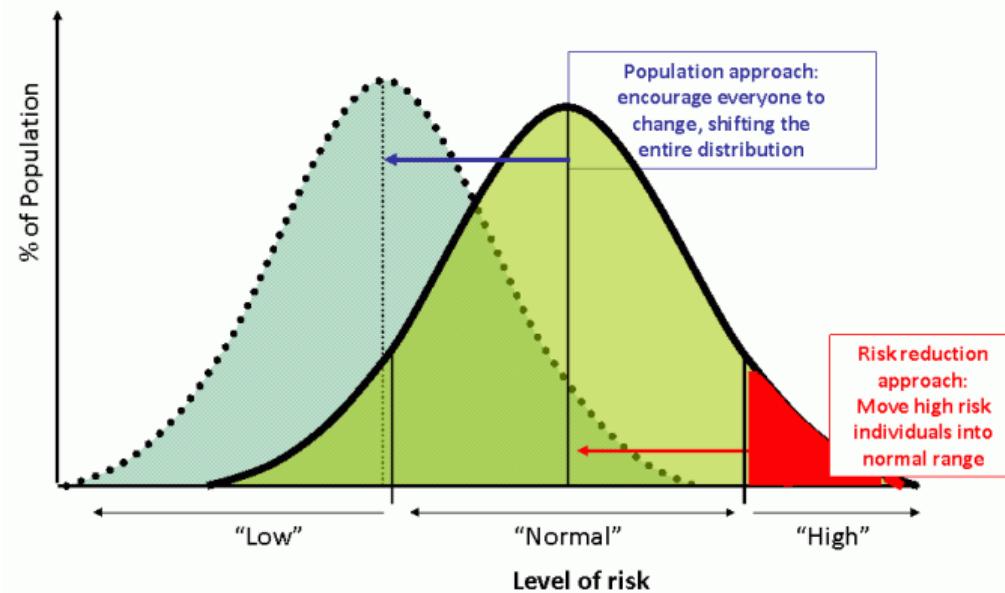
*Familiarize with research in public and population health
Identify open areas related to health equity
Activate the machine learning community*



- Introduce Public and Population Health
- Theory and framework of social determinants of health (SoDH)
- Measurement of SoDH
- SoDH interventions
- Integration of SoDH in machine learning models
- Taxonomy of health tasks
- Causal inference in public health
- Challenges with using proxies
- Algorithmic fairness and health disparities

Population Health

Focus is on what makes and keeps people healthy (holistically). Aim is to improve the health of the entire population and to reduce health inequities among population groups [Rose, 1985].



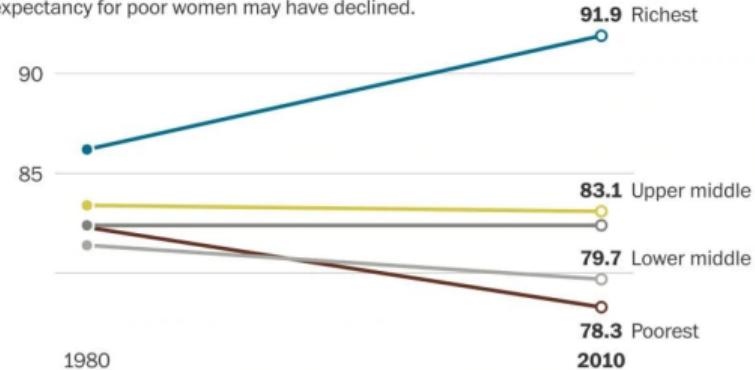
What is Health Equity?

Health equity

“Minimizing avoidable disparities in health and its determinants—including but not limited to health care—between groups of people who have different levels of underlying social advantage or privilege, i.e., different levels of power, wealth, or prestige due to their positions in society relative to other groups.”
[Braveman, 2006]

Inequality in life expectancy widens for women

Wealthier women can expect to live longer than their parents did, while life expectancy for poor women may have declined.



Life expectancy for 50-year-olds in a given year, by quintile of income over the previous 10 years

Source: National Academies of Science, Engineering and Medicine

Figure 1: Growing inequalities for women in the United States [National Academies, 2015].

Public Health



Figure 2: Socio-ecological model of health [Bronfenbrenner, 1977].

Impact



Figure 3: [CDC, 2014]

United States Social factors account for 25–60 percent of deaths in any given year according to results from various meta-analyses. [Heiman and Artiga, 2015]

Worldwide: Eighty per cent of noncommunicable diseases could be prevented through primary prevention – through modifying behaviours such as reducing tobacco consumption and fat, alcohol and salt intake, preventing obesity, and promoting physical activity, and improving environmental conditions such as air quality and urban planning [WHO, 2009].

Importance of considering all determinants of health - COVID-19 case study

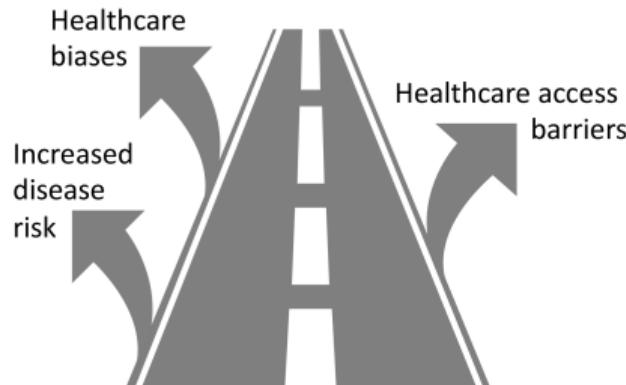


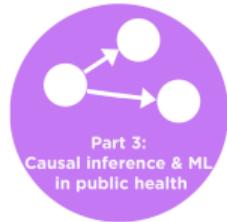
Figure 4: "Highway to health".

Determinant (COVID relevance [Rollston and Galea, 2020])

- Housing conditions (crowding, poor sanitation)
- Healthy food access (increased comorbidities)
- Education access and quality (health literacy, future socioeconomic status)
- Socio-economic status (occupation/essential worker, ability to self-isolate)
- Healthcare trust (access treatment)
- Racism/discrimination (socioeconomic status, segregation, housing quality, health care access, quality, etc..)

Tutorial Goals

*Familiarize with research in public and population health
Identify open areas related to health equity
Activate the machine learning community*



- Introduce Public and Population Health
- Theory and framework of social determinants of health (SoDH)
- Measurement of SoDH
- SoDH interventions
- Integration of SoDH in machine learning models
- Taxonomy of health tasks
- Causal inference in public health
- Challenges with using proxies
- Algorithmic fairness and health disparities

Social Determinants of Health



Figure 5: Five key areas of SoDH

“Conditions in which people are born, grow, live, work and age. These circumstances are shaped by the distribution of money, power and resources at global, national and local levels.” [WHO, 2008]

Purposes of conceptual frameworks [WHO, 2010]:

- Guide empirical work to enhance understanding of determinants and mechanisms
- Guide policy-making to illuminate entry points for interventions and policies

Two Theories:

- Life course approach to inter-generational effects
- Intersectionality

Life Course Approach to Inter-generational Effects

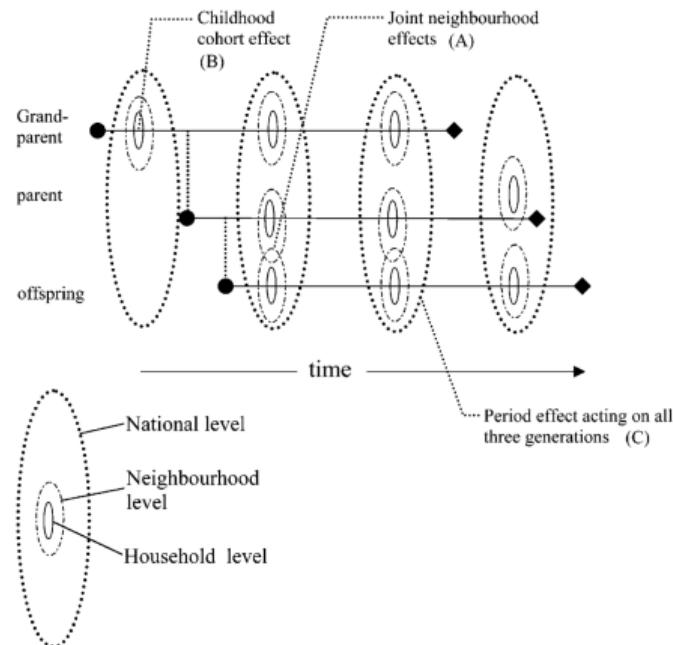
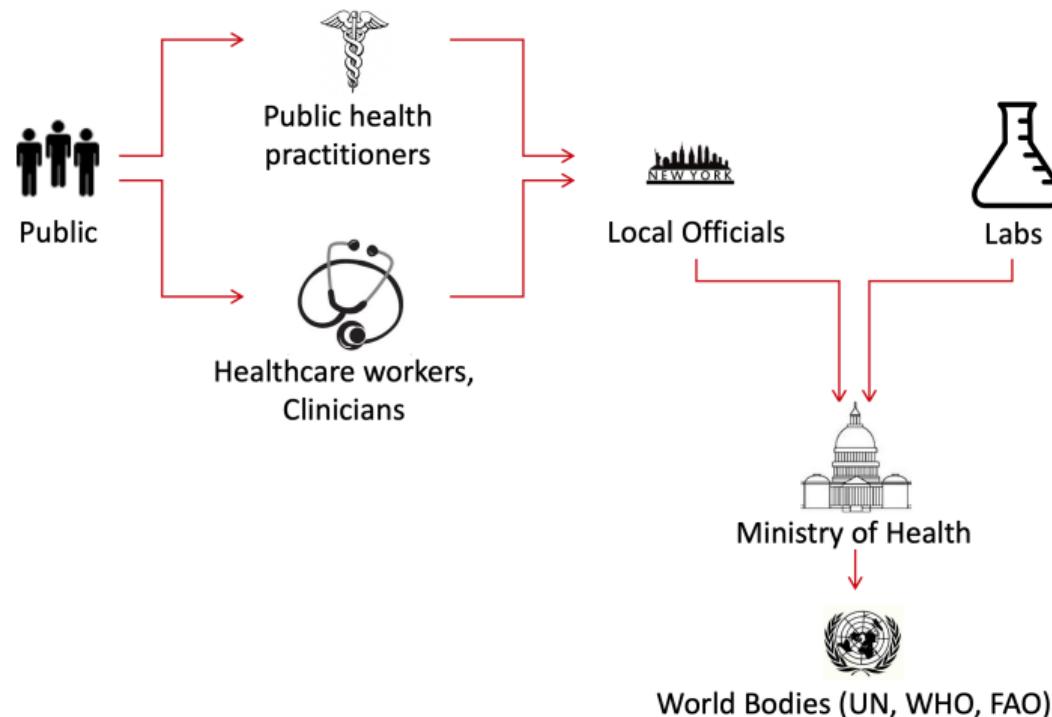


Figure 6: Hierarchical and life course exposures on disease risk across generations

- Longitudinal effects of socioeconomic adversities to which people are exposed to at various developmental stages on health. [Cable, 2014]
- Individual life-course perspective to inter-generational association between social determinants and diseases (e.g. parent adverse childhood experience was associated with higher odds of poor child overall health status [Lê-Scherban et al., 2018])

Traditional Data Collection in Public Health



Measurement of Social Determinants of Health

- Data sources at individual and population level: Electronic Medical Record (SoDH screening), American Community Survey (ACS), U.S. Census Bureau, Nationally representative surveys (e.g. NHANES, BRFSS)
- Each construct is measured by multiple indicators (e.g. Housing: has housing, rental housing, sanitation status, crowding, indoor air quality)
[Kusnoor et al., 2018]

Internet and Mobile Data - examples



Health outcome (measured) examples

- Parkinsons (symptoms)
[Zhan et al., 2018]
- Diabetes (mood, behaviors, diet)
[Akbari and Chunara, 2019]
- Mental health (activity)
[Quisel et al., 2016]

Figure 7: Person-generated data sources used today in public health

Data vs. Models

... while the fair ML literature has largely focused on “de-biasing” methods and viewed the training data as fixed, most of our interviews report that their teams consider data collection, rather than model development, as the most important place to intervene [Holstein et al., 2019].

Challenge in Measuring Social Determinants

- Data sharing; calls for centralized repository and open source assessment tools [National Academies of Sciences et al., 2019]
- Denominator challenges need to be addressed in order to understand and reach populations at risk [Chunara et al., 2017]
- **Internal validity:** the degree of confidence that the causal relationship being tested is trustworthy and not influenced by other factors or variables.
External validity: the extent to which results from a study can be applied (generalized) to other situations, groups or events.
[Mitchell and Jolley, 2004]

Challenge in Measuring Social Determinants: Collection medium moderates person-generated data

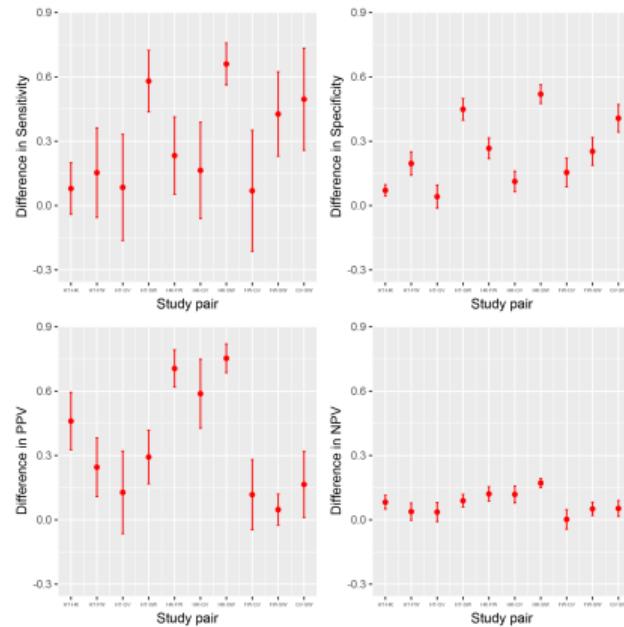


Figure 8: Varying specificity, sensitivity, NPV and PPV in multiple influenza syndromic surveillance systems.

Characteristics of the surveillance approach (data collection method, survey questions, the places/groups and time periods data are collected from) for which predictive performance of the same case definition can vary [Chunara et al., 2020].

Challenge in Measuring Social Determinants: Understanding the data generating process

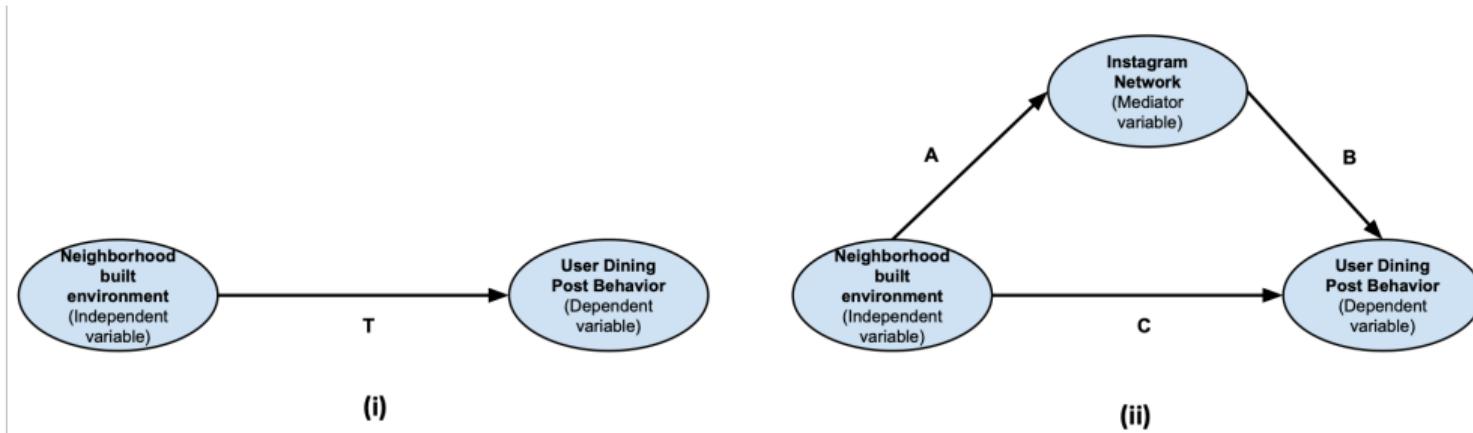


Figure 9: Role of the built and online social environments on expression of dining on Instagram [Mhasawade et al., 2020a]

Challenge in Measuring Social Determinants: Variable forms and pathways

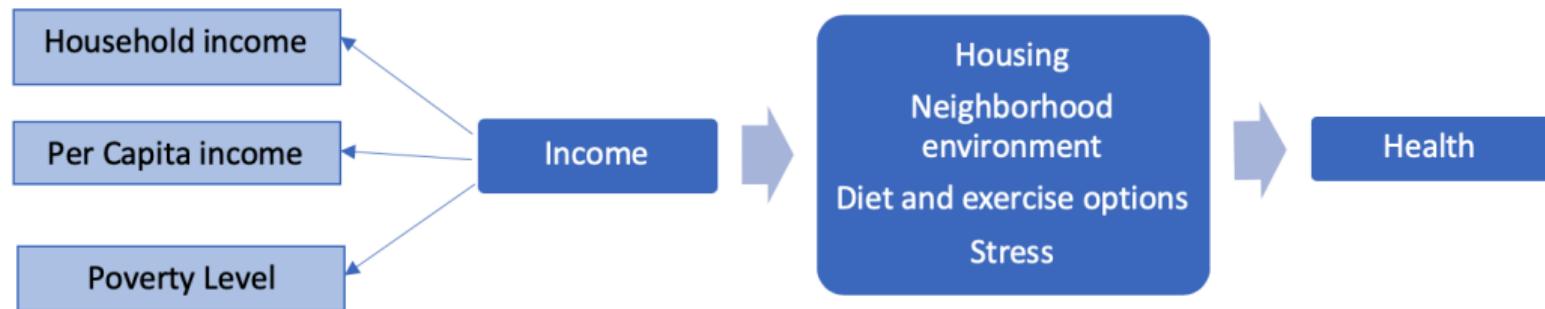


Figure 10: The Influence of Income on Health [Marmot, 2002]

Challenges in Measuring Social Determinants

- Mapping from construct space to observed space through imperfect proxy features and associated metrics [Friedler et al., 2016]

Construct space	Observed space
Intelligence	IQ
Success in High School	GPA
Propensity to commit crime	Family history of crime
Risk-averseness	Age
Knowledge of job	Number of Years of Experience

Definition (Observed space (OS)). The observed space (with respect to T) is a metric space $OS = (\hat{P}, \hat{d})$. We assume an observation process $g : P \rightarrow \hat{P}$ that generates an entity $\hat{p} = g(p)$ from a person $p \in CS$.

Challenge in Measuring Social Determinants: Biases in measurement

“There is a need for better understanding—both among data analysts and among consumers of data—of the weighty implications of analytic choices in measurement of health disparities, health inequities, and social determinants of health”

“Bias can stem not only from the value preferences or habits that inform choice of measurement practice(s) but also from the effect that different data presentation approaches have on audience perceptions or judgments of the resulting meaning of the data or analysis”

[Penman-Aguilar et al., 2016]

Are Social Determinants Intervenable? [Schwartz et al., 2016]

- Violating Stable Unit Treatment Value (SUTVA) assumption in causal inference
- Downstream manipulable mediators of social constructs as the exposures of interest (e.g. encouraging reading to children to improve cognitive development)
- Consider structural change for upstream social determinants (e.g. improving childhood SES)



Figure 11: From upstream to individual interventions [Lehman, 2019]

Simulation Study of Social Determinant Interventions on Chronic Illness

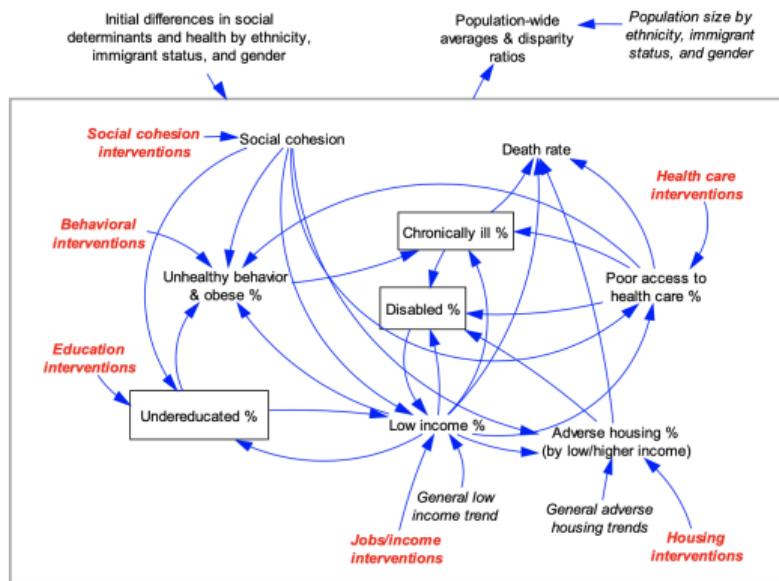


Figure 12: Hypothesized causal pathways of social and biological determinants of health [Mahamoud et al., 2013].

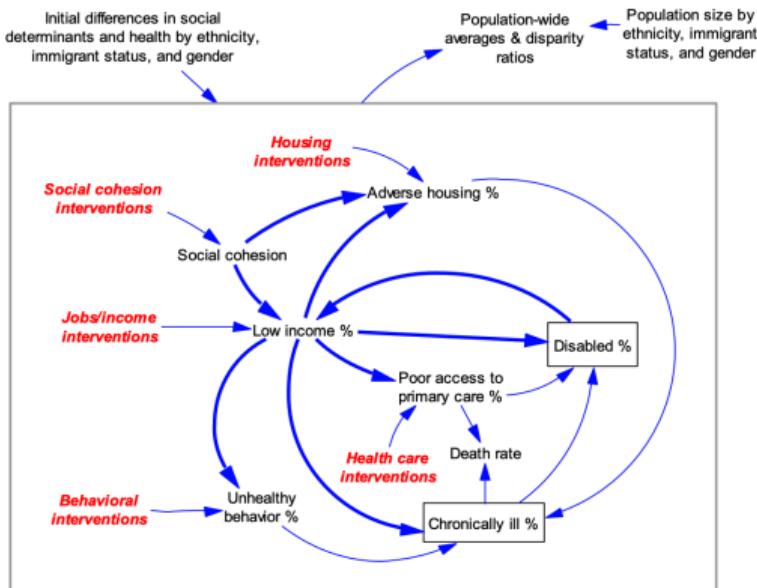


Figure 13: Modified causal graph based on relations in data [Mahamoud et al., 2013].

Social Determinants in Cardiovascular Disease Prediction Machine Learning Models

How are researchers incorporating social determinants in cardiovascular disease prediction models? [Zhao et al., 2020]

- 2728 publications were identified, and 120 publications were included
- Commonly measured social determinants: age, gender, ethnicity/race, education and income
- Most popular machine learning algorithms: random forest, SVM, decision tree, neural nets
- Data sources: cohort/observational studies, clinical trials, hospital electronic medical records

Reflections from CVD & Social Determinants Case Study

- Lack of comprehensive social determinants data (e.g. social environment, contextual information) from cohort studies or clinical datasets
- Most measurement at individual instead of community/population level
- Outcomes vary, including risk scores, prognosis of CVD and readmission to hospitals
- How to apply models and results to actionable interventions to address health disparity and improve health equity

Recap

*Familiarize with research in public and population health
Identify open areas related to health equity
Activate the machine learning community*



- Introduce Public and Population Health
- Theory and framework of social determinants of health (SoDH)
- Measurement of SoDH
- SoDH interventions
- Integration of SoDH in machine learning models
- Taxonomy of health tasks
- Causal inference in public health
- Challenges with using proxies
- Algorithmic fairness and health disparities

What is the Role of Machine Learning in Public and Population Health?

What are the challenges with *health* tasks? Are the challenges any different from healthcare or other domains?



Taxonomy of ML in Population and Public Health

Identify ML opportunities for health, comprehensively

Public health and healthcare tasks can be grouped as follows:

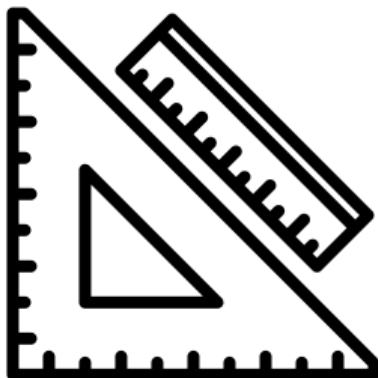
- (1) Identification
- (2) Design
- (3) Prediction
- (4) Allocation

ML in Health: (1) Identification



- Disease incidence at population level [Bhatt et al., 2013]
- Disease detection including behavior and markers [Gulshan et al., 2016]
- Multi-level factors related to health outcomes [Weichenthal et al., 2020]
- External validity of policies [Hermanspann et al., 2017]
- Fairness of policies [Obermeyer and Mullainathan, 2019a]
- Causal treatment effects [Lodi et al., 2019]
- Individuals/subpopulations to target interventions [McGuire, 2016]
- Individuals/populations to minimize healthcare costs [Rose et al., 2017]

ML in Health: (2) Design



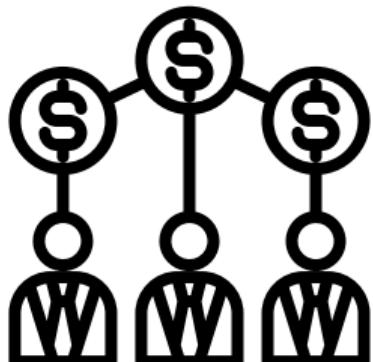
- Individual-level interventions [Rahmattalabi et al., 2018]
- Community/group- level interventions [Ahsan et al., 2013]
- Public policy [Braveman et al., 2004]

ML in Health: (3) Prediction



- Risk score (clinical algorithms) [Vyas et al., 2020]
- Disease prognosis [Dugan et al., 2015]
- Treatment/procedure effectiveness [Kreif et al., 2015]
- Chance of mortality [Rajkomar et al., 2018]
- Treatment adherence [Franklin et al., 2016]
- Hospital readmission [Galiatsatos et al., 2019]

ML in Health: (4) Allocation



- Resources to individuals [Snyder et al., 2018]
- Resources to populations [Lord et al., 2015]
- Resources within hospitals [da Silveira Grüber et al., 2018]
- Care management (at both individual and population level) [Osborn et al., 2017]

Causality and Health Tasks

Introduction to Causality

Causal methods have been used in epidemiology for representing domain knowledge via causal graphs. This helps capture epistemic uncertainty as well as incorporate prior knowledge into methods.

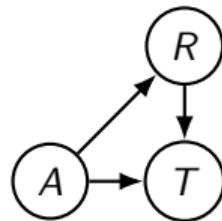


Figure 14: Example causal graph where A represents age, R represents risk factor, T represents treatment outcomes.

Nodes represent quantities of interest. Edges represent the relationship between different quantities. For example, a directed arrow from A to T represents that age affects the treatment outcome.

A Natural Experiment: Estimating Population Level Education Effect on CVD

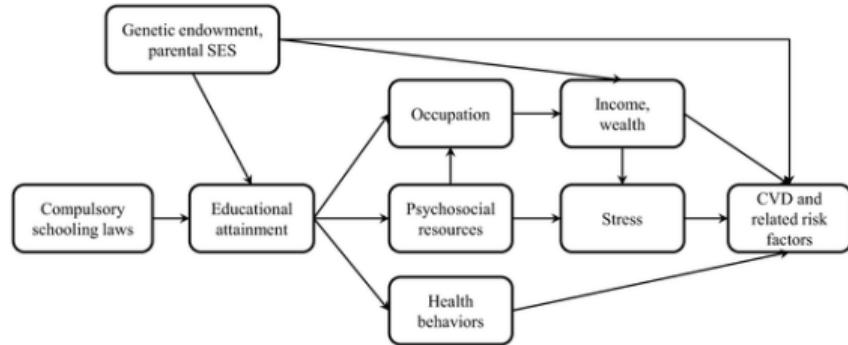


Figure 15: Conceptual model linking educational attainment with CVD [Hamad et al., 2019]. Numerous pathways linking education with cardiovascular disease (CVD).

- United States state-level compulsory education laws provide grounds for a natural experiment.
- Multiple risk factors like smoking, depression, cholesterol levels, and BMI.
- OLS - improvements in all risk factors with increased education level.
- IV - improvement in only some, worsening few risk factors (cholesterol and BMI).

Should We Control For Smoking?

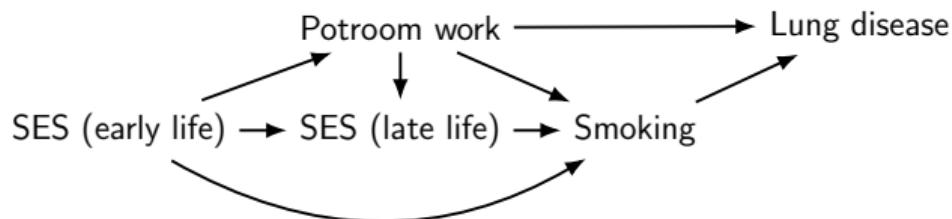


Figure 16: Controlling for smoking blocks the effect of early life SES on disease risk but smoking representative of downstream effect of job [Mcclure, 2018].



Figure 17: Hazardous conditions in an aluminium pot room^a.

Control downstream of the exposure when estimating causal effect?

a. Source: <http://www.foilvedanta.org/articles/a-nice-place-to-work-in-experiences-of-icelandic-aluminium-smelter-employees/>

How to Model Social Variables?

Using DAGs (Directed Acyclic Graphs) is *not* the only solution [Krieger and Davey Smith, 2016].

Challenges

- Complex social-health relations question validity of postulated causal graphs.
- Current approaches consider DAGs as drivers of the study.
- Can a *single* study support the hypothesis?

Triangulation to the rescue.

- Integrate results from several approaches for reliable conclusions.
- Inference to best explanation (IBE):
 - ➊ Generate candidate hypothesis
 - ➋ Select from them based on explainability.

Can we just incorporate social variables into analyses like any other?
Are there any pitfalls to this?

Example: ‘Race’ as a Proxy in Clinical Algorithms

<i>Clinical Utility</i>	<i>Effect of race on risk score of black patients in comparison with non black patients</i>	<i>Equity concern</i>
Cardiology	Lower mortality risk	Higher threshold for allocating clinical resources.
Cardiac Surgery	Higher risk of operative mortality	Lower allocation of procedures to black patients.
Nephrology	Higher eGFR	Delay in specialized care assuming better kidney functions.
Obstetrics	Lower estimated success rate	Reduced quality of clinical care.
Urology	Lower risk of a uretral stone	Reduced quality of clinical care.
Oncology	Higher risk for cancer survival	Fewer interventions.
Endocrinology	Lower risk of osteoporosis	Delayed diagnosis and intervention.

Table 1: Equity concern of clinical algorithms using ‘race’ [Vyas et al., 2020].

- Should sensitive attributes be considered as proxies?
- Do social variables play the same role across all tasks, merely being considered as protected attributes?

Health vs Other Domains

Sensitive attributes like race, gender are often used as proxies for different social interactions across various domains including health.

However, especially in health, they are not just mere proxies but are 'social determinants' of health.

Modeling sensitive attributes in health requires understanding the complex mechanism between them and health outcomes.

Example: Assessing Kidney Function

Assessing kidney function is essential to recognize kidney injury which is evaluated using glomerular filtration rate (GFR) that accounts for the serum creatinine level.

What is *serum creatinine*?

A waste product in blood from muscle activity. With impaired kidney function, the amount of serum creatinine increases in the blood which would normally be removed from the blood by the kidneys.

CKD-EPI Equation

Developed in 2009 using a diverse population estimate GFR from serum creatinine, age, sex and race.

$$\text{GFR} = 1.41 \times \min(\text{Scr}/\mathcal{K}, 1)^\alpha \times \max(\text{Scr}/\mathcal{K}, 1)^{-1.209} \times 0.993^{\text{Age}} \times 1.019[\text{if female}] \\ \times 1.159[\text{if black}]$$

where Scr is standardized serum creatinine in mg/dL, \mathcal{K} is 0.7 for females and 0.9 for males, α is -0.329 for females and -0.411 for males.

Case for ‘Race Correction’ in eFGR Function

- Sensitive attributes like gender, race, and class may be intended as proxies for the interactions of systems of oppression (sexism, racism, classism) and other social processes in producing population-level incidence [Bauer, 2014].
- Estimate kidney function value (eFGR) without race correction. *Researchers have been actively demonstrating that use of race multipliers can lead to important care delays. Recently, such efforts have led to elimination of the race multiplier at multiple places including MGH/Brigham.* ¹.
- “The challenge that scientists must address is how to report genomic variation without inappropriately describing racial and ethnic groups as discrete population groups?” [Bonham et al., 2018].

¹<https://twitter.com/LashNolen/status/1276181898394558467/photo/1>

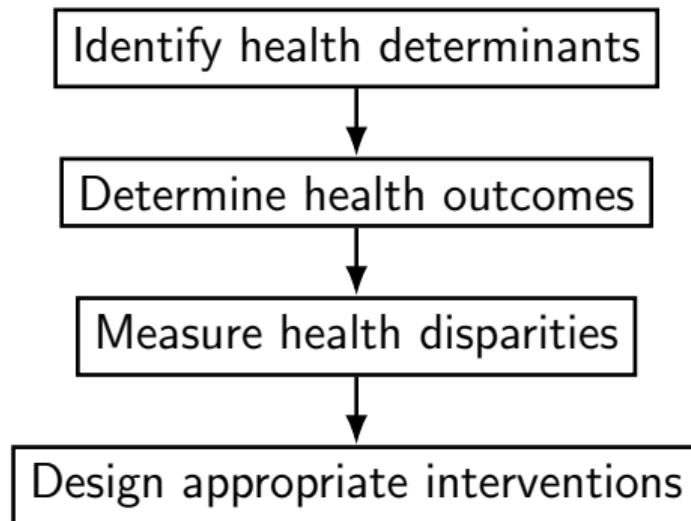
Algorithmic fairness and health disparities

What are Health Disparities? And Why is it Important to be Aware of Them?

A health disparity/inequality is a particular type of difference in health (or in the most important influences on health that could potentially be shaped by policies); it is a difference in which disadvantaged social groups—such as the poor, racial/ethnic minorities, women, or other groups who have persistently experienced social disadvantage or discrimination—systematically experience worse health or greater health risks than more advantaged social groups.

Identifying health disparities is essential to understand the dynamics of social, economic, cultural environments and their effect on health outcomes that is related to social disadvantage.

How to Recognize Health Disparities?



Be *aware* of the pipeline.
Be *fair* to public health.

- Ensuring *health equity* involves swimming in a *complex* public health ecosystem!
[Braveman, 2006]
- Touching surface of just one facet can drown down the effort!

Figure 18: Pipeline for detecting health disparities

Are Average Rates Representative of the Actual Health Disparities?

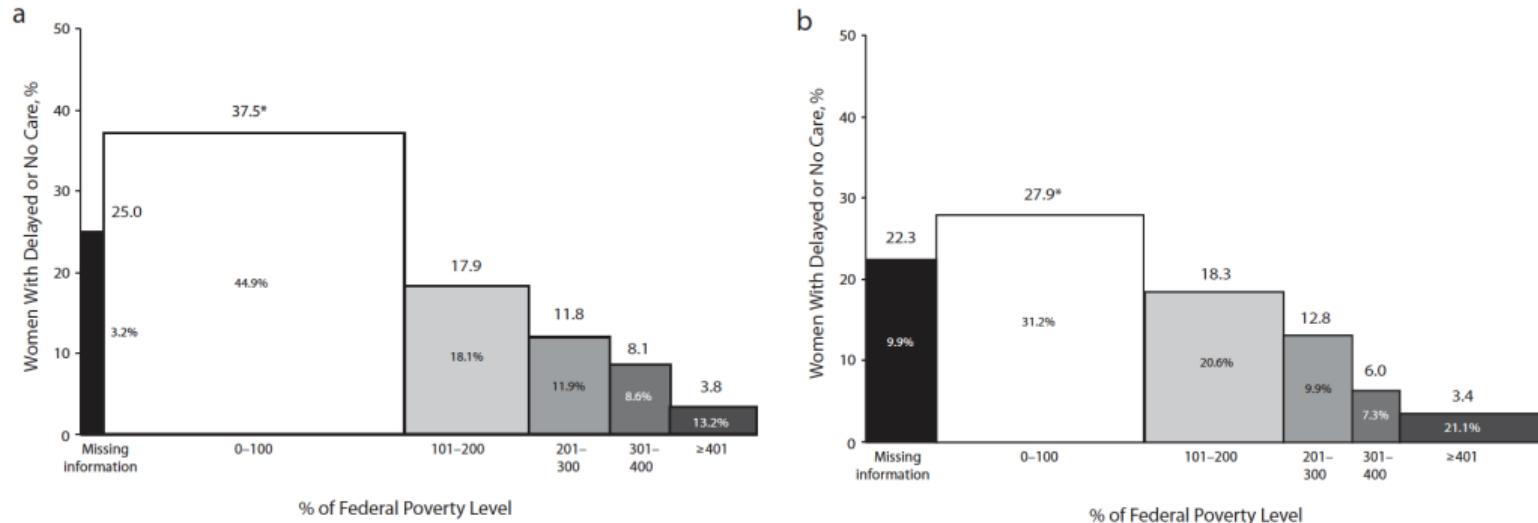


Figure 19: Proportion of women in California with delayed or no prenatal care by income in a) 1994-1995 and b) 1999-2001. Overall improvements in prenatal care rates among childbearing women in California but disparities by income persisted [Braveman et al., 2004].

Intersectionality

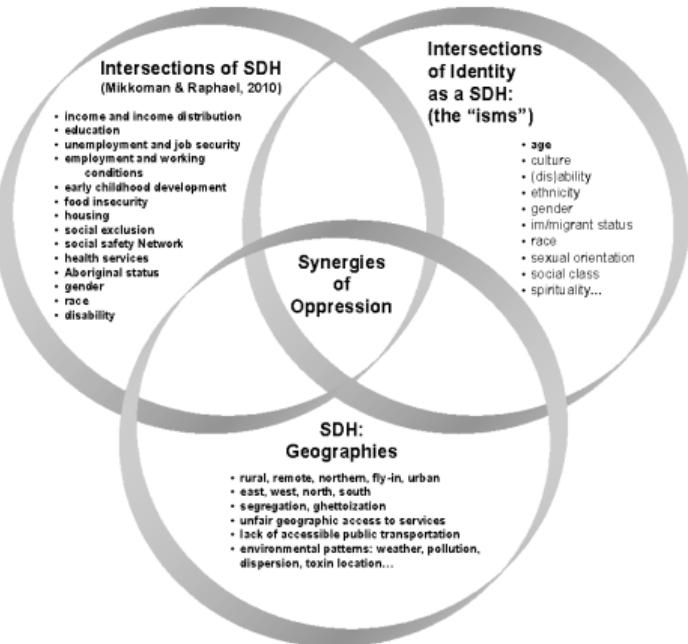
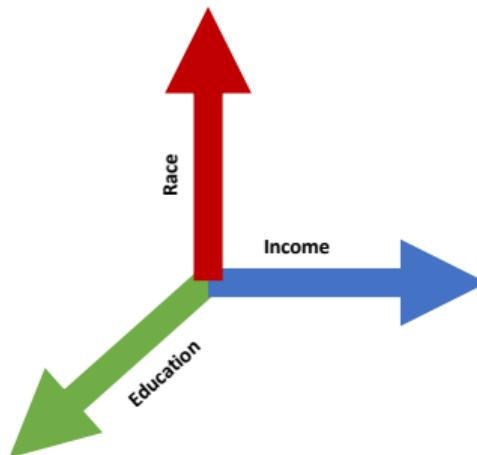


Figure 20: Intersectionality in health
[McGibbon and McPherson, 2011]

Intersectionality is an approach or lens that recognizes that health is shaped by a multi-dimensional overlapping of factors such as race, class, income, education, age, ability, sexual orientation, immigration status, ethnicity, indigeneity, and geography.



Eco-epidemiology and Interaction terms for Social Variables

- It is common practice to evaluate an interaction between race and an exposure of interest as evidence (or lack thereof) that an exposure contributes to a racial health disparity. However, when using this method, researchers may attribute too much authority to the significance of this interaction term [Ward et al., 2019]
- “Tyranny of means”: the average causal effect of a treatment is not the same as an individual causal effect [Merlo and Wagner, 2013]
- In a multilevel framework, the “effect” of being influenced by a higher level like, the family, neighborhood, or school can be considered as a general contextual effect. This general influence is not properly operationalized by measuring differences between average risks. Rather, the general influence of the context is better quantified by measuring the share of the total interindividual heterogeneity that appears at that specific level [Merlo et al., 2009]

MAHIDA: Multilevel Analysis of Individual Heterogeneity and Discrimination Accuracy

$$\text{Total variance} = (\text{Within-strata variance}) + (\text{between-strata variance})$$

“Including interaction terms encourages us to only study the intersectionality of marginalization.” [Evans et al., 2018]

<i>Comparison criteria with additional sensitive attributes</i>	<i>Fixed effect model (interaction terms for intersectionality)</i>	<i>Multilevel model</i>
Increase in fixed effect parameters Estimates adjusted for sample size in each strata?	Geometrically No	Linearly Yes

Table 2: Comparison between fixed effect and multilevel approaches

Multilevel Models

- Consistent with eco-epidemiology approach to situate individuals within intersectional social strata instead of individual level variables
- Intersectionality situates the problems of disparities in the structural power hierarchies, social processes, social determinants that shape the the social experiences of individuals with the specific intersectional identities.
- It is important to examine the magnitude and direction of the intersectional interaction effect to recognize disparities and privileges at the intersection of social experiences.
- Example work harnessing age/gender in a multilevel model helps to capture invariant information in population attributes for a flu prediction task, to improve prediction in datasets where groups may be under-represented [Mhasawade et al., 2020b].

Can Intersectionality Theory Inform the Way Forward?

- Numerous interlocking systems of privilege and oppression such as racism, classism, sexism, and ageism push back against the “additive approach,” which treats the advantages or disadvantages conferred through simultaneous occupation of multiple social positions as simply accumulated [Collins, 2002, Crenshaw, 1989, McCall, 2005].
- To describe joint effects of these systems, need a meaningful reference point: one choice is a world where effects of all power hierarchies are independent and additive.
- In real situations different intersectional groups have radically different sizes and levels of social power and position, thus the average of stratum-level means would not be a meaningful quantity.

Algorithmic fairness

“Fairness” of decisions quantitatively defined based on statistical and machine learning predictions [Mitchell et al., 2018].

Absence of discrimination of individuals with the same “merit”
[Kasy and Abebe, 2020].

Several definitions based on maximizing utility, ensuring equal prediction, equal decision across advantaged and disadvantaged groups²

²For a complete summary we refer to [Mitchell et al., 2018].

Algorithmic Fairness in Public Health

Algorithmic fairness has not accounted for complex causal relationships between biological, environmental and social factors that give rise to differences in medical conditions across protected identities [McCraadden et al., 2020].

Social and structural factors affect health across multiple intersecting identities, but the mechanism(s) by which social determinants affect health outcomes is not always well understood.

Questioning the Questions we Ask!

“Don’t just ask how the algorithm treats different people differently, but also who gets to do the treating” [Kasy and Abebe, 2020].

- Assess the causal impact of introducing the algorithm on inequality [Kasy and Abebe, 2020].
- Consider inclusion decisions [Yang et al., 2020, Nishtala et al., 2020].
- Is the objective of eliminating disparities in line with health equity [Obermeyer and Mullainathan, 2019b]?
- Improve methods for understanding the relation between observed space and decision space especially when the construction is complex across different social variables [Friedler et al., 2016].
- Assess disparities with direct and indirect path-specific causal effects [Wu et al., 2019].

Advancing Health Disparities with 'Fair ML': Can Fair Algorithms be Inequitable?

Unexplained variance: sensitive attributes as 'proxy'.

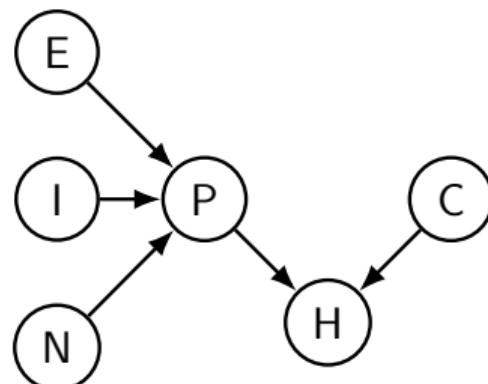


Figure 21: If the *perceived protected attribute* P is not distilled into components like education E , income I , neighborhood SES N but its effect on the health outcomes H along with the clinical variables C is assessed; then the variance between the intersectional groups will not be identified leading to inequity across intersectional strata.

Advancing Health Disparities with 'Fair ML': Can Fair Algorithms be Inequitable?

Should you only treat highly insured patients? Can it lead to inequity?

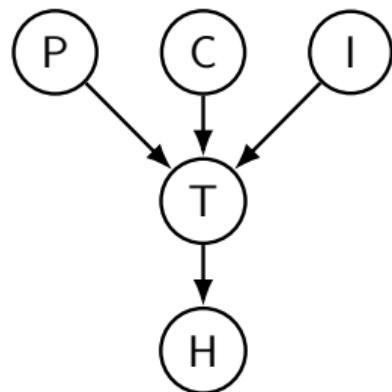
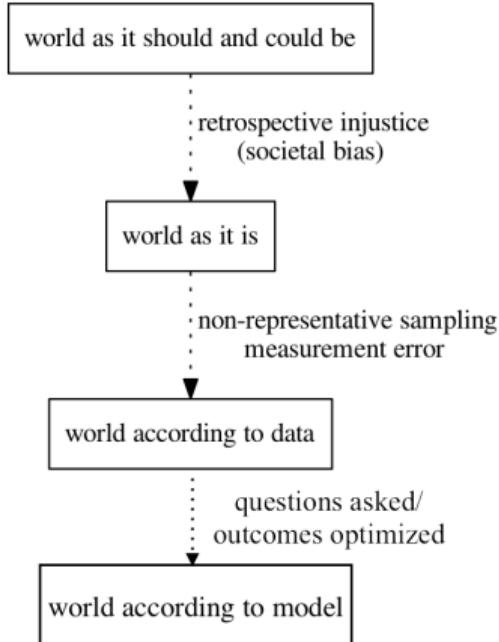


Figure 22: Is it fair to treat (T) highly insured patients (I) considering their *perceived protected attribute* P and clinical variables C ? The resulting health outcomes H may be approximately equal across advantaged and disadvantaged groups with respect to P but social health disparity still persists for lower insurance patients.

Take-aways



- Social determinants are paramount for attaining health equity.
- There are many Machine Learning opportunities for better measuring, understanding and incorporating social determinants across health tasks.
- Health equity can be prioritized in Machine Learning models via types of questions asked, how data is represented, etc.

Figure 23: Perception vs. reality (modified from [Mitchell et al., 2018]).

ML in Population and Public Health



Part 1:
Intro to Population
and Public Health



Part 2:
Social determinants,
data challenges



Part 3:
Causal inference & ML
in public health

For a complete list of related articles, more information and to give feedback:

<https://ChunaraLab.github.io/MLPH/>

Acknowledgements

Vishwali Mhasawade
Yuan Zhao
Harvineet Singh
Yijun Tian
Alene Rhea
Allan Porter



References I

- Ahsan, G. T., Addo, I. D., Ahamed, S. I., Petereit, D., Kanekar, S., Burhansstipanov, L., and Krebs, L. U. (2013). Toward an mhealth intervention for smoking cessation. In *2013 IEEE 37th Annual Computer Software and Applications Conference Workshops*, pages 345–350. IEEE.
- Akbari, M. and Chunara, R. (2019). Using contextual information to improve blood glucose prediction. *Machine Learning for Healthcare*, arXiv preprint arXiv:1909.01735.
- Bauer, G. R. (2014). Incorporating intersectionality theory into population health research methodology: challenges and the potential to advance health equity. *Social science & medicine*, 110:10–17.

References II

- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., et al. (2013). The global distribution and burden of dengue. *Nature*, 496(7446):504–507.
- Bonham, V. L., Green, E. D., and Pérez-Stable, E. J. (2018). Examining how race, ethnicity, and ancestry data are used in biomedical research. *Jama*, 320(15):1533–1534.
- Braveman, P. (2006). Health disparities and health equity: concepts and measurement. *Annu. Rev. Public Health*, 27:167–194.
- Braveman, P. A., Egerter, S. A., Cubbin, C., and Marchi, K. S. (2004). An approach to studying social disparities in health and health care. *American Journal of Public Health*, 94(12):2139–2148.

References III

- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American psychologist*, 32(7):513.
- Cable, N. (2014). Life course approach in social epidemiology: an overview, application and future implications. *Journal of epidemiology*, page JE20140045.
- CDC (2014). Up to 40 percent of annual deaths from each of five leading us causes are preventable. *Atlanta, GA: Centers for Disease Control and Prevention*.
- Chunara, R., Plymoth, A., and Martin, L. (2020). Diversity in surveillance data: implications for infectious disease forecasting models.

References IV

- Chunara, R., Wisk, L. E., and Weitzman, E. R. (2017). Denominator issues for personally generated data in population health monitoring. *American journal of preventive medicine*, 52(4):549–553.
- Collins, P. H. (2002). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. routledge.
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *U. Chi. Legal f.*, page 139.
- da Silveira Grüber, M., da Costa, C. A., da Rosa Righi, R., Rigo, S. J., and Chiwiacowsky, L. D. (2018). A hospital bed allocation hybrid model based on situation awareness. *CIN: Computers, Informatics, Nursing*, 36(5):249–255.

References V

- Dugan, T. M., Mukhopadhyay, S., Carroll, A., and Downs, S. (2015). Machine learning techniques for prediction of early childhood obesity. *Applied clinical informatics*, 6(03):506–520.
- Evans, C. R., Williams, D. R., Onnela, J.-P., and Subramanian, S. (2018). A multilevel approach to modeling health inequalities at the intersection of multiple social identities. *Social Science & Medicine*, 203:64–73.
- Franklin, J. M., Shrank, W. H., Lii, J., Krumme, A. K., Matlin, O. S., Brennan, T. A., and Choudhry, N. K. (2016). Observing versus predicting: initial patterns of filling predict long-term adherence more accurately than high-dimensional modeling techniques. *Health services research*, 51(1):220–239.

References VI

- Friedler, S. A., Scheidegger, C., and Venkatasubramanian, S. (2016). On the (im) possibility of fairness. *arXiv preprint arXiv:1609.07236*.
- Galiatsatos, P., Follin, A., Uradu, N., Alghanim, F., Daniel, Y., Saria, S., Townsend, J., Sylvester, C., Chanmugam, A., and Chen, E. (2019). The association between neighborhood socioeconomic disadvantage and readmissions for patients hospitalized with sepsis. In *C94. The Impact of Social Determinants in Pulmonary and Critical Care*, pages A5569–A5569. American Thoracic Society.

References VII

- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410.
- Hamad, R., Nguyen, T. T., Bhattacharya, J., Glymour, M. M., and Rehkopf, D. H. (2019). Educational attainment and cardiovascular disease in the united states: A quasi-experimental instrumental variables analysis. *PLoS medicine*, 16(6):e1002834.
- Heiman, H. J. and Artiga, S. (2015). Beyond health care: the role of social determinants in promoting health and health equity. *Health*, 20(10):1–10.

References VIII

- Hermanspann, T., Schoberer, M., Robel-Tillig, E., Härtel, C., Goelz, R., Orlikowsky, T., and Eisert, A. (2017). Incidence and severity of prescribing errors in parenteral nutrition for pediatric inpatients at a neonatal and pediatric intensive care unit. *Frontiers in pediatrics*, 5:149.
- Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., and Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–16.
- Kasy, M. and Abebe, R. (2020). Fairness, equality, and power in algorithmic decision making. Technical report, Working paper.

References IX

- Kreif, N., Grieve, R., Díaz, I., and Harrison, D. (2015). Evaluation of the effect of a continuous treatment: a machine learning approach with an application to treatment for traumatic brain injury. *Health economics*, 24(9):1213–1228.
- Krieger, N. and Davey Smith, G. (2016). The tale wagged by the dag: broadening the scope of causal inference and explanation for epidemiology. *International journal of epidemiology*, 45(6):1787–1808.
- Kusnoor, S. V., Koonce, T. Y., Hurley, S. T., McClellan, K. M., Blasingame, M. N., Frakes, E. T., Huang, L.-C., Epelbaum, M. I., and Giuse, N. B. (2018). Collection of social determinants of health in the community clinic setting: A cross-sectional study. *BMC Public Health*, 18(1):550.

References X

- Lê-Scherban, F., Wang, X., Boyle-Steed, K. H., and Pachter, L. M. (2018). Intergenerational associations of parent adverse childhood experiences and child health outcomes. *Pediatrics*, 141(6):e20174274.
- Lehman, C. (2019). Addressing social determinants of health. *Physical Therapy in Motion*.
- Lodi, S., Phillips, A., Lundgren, J., Logan, R., Sharma, S., Cole, S. R., Babiker, A., Law, M., Chu, H., Byrne, D., et al. (2019). Effect estimates in randomized trials and observational studies: comparing apples with apples. *American journal of epidemiology*, 188(8):1569–1577.

References XI

- Lord, A. S., Carman, H. M., Roberts, E. T., Torrico, V., Goldmann, E., Ishida, K., Tuhrim, S., Stillman, J., Quarles, L. W., and Boden-Albala, B. (2015). Discharge educational strategies for reduction of vascular events (deserve): design and methods. *International journal of stroke*, 10(SA100):151–154.
- Mahamoud, A., Roche, B., and Homer, J. (2013). Modelling the social determinants of health and simulating short-term and long-term intervention impacts for the city of toronto, canada. *Social science & medicine*, 93:247–255.
- Marmot, M. (2002). The influence of income on health: views of an epidemiologist. *Health affairs*, 21(2):31–46.

References XII

- McCall, L. (2005). The complexity of intersectionality. *Signs: Journal of women in culture and society*, 30(3):1771–1800.
- McClure, L. (2018). *Controlling for smoking?*
- McCradden, M. D., Joshi, S., Mazwi, M., and Anderson, J. A. (2020). Ethical limitations of algorithmic fairness solutions in health care machine learning. *The Lancet Digital Health*, 2(5):e221–e223.
- McGibbon, E. and McPherson, C. (2011). Applying intersectionality & complexity theory to address the social determinants of women's health.
- McGuire, T. G. (2016). Achieving mental health care parity might require changes in payments and competition. *Health Affairs*, 35(6):1029–1035.

References XIII

Merlo, J., Ohlsson, H., Lynch, K. F., Chaix, B., and Subramanian, S. (2009). Individual and collective bodies: using measures of variance and association in contextual epidemiology. *Journal of Epidemiology & Community Health*, 63(12):1043–1048.

Merlo, J. and Wagner, P. (2013). The tyranny of the averages and the indiscriminate use of risk factors in public health: a call for revolution. *Eur J Epidemiol*, 28(Suppl 1):148.

Mhasawade, V., Elghafari, A., Duncan, D. T., and Chunara, R. (2020a). Role of the built and online social environments on expression of dining on instagram. *International journal of environmental research and public health*, 17(3):735.

- Mhasawade, V., Rehman, N. A., and Chunara, R. (2020b). Population-aware hierarchical bayesian domain adaptation via multi-component invariant learning. In *Proceedings of the ACM Conference on Health, Inference, and Learning*, pages 182–192.
- Mitchell, M. and Jolley, J. (2004). Research design explained 5 th ed. Victoria: Wadsworth Publisher.
- Moebert, J. & Tydecks, P.(2007). *Power and Ownership Structures among German Companies. A Network Analysis of Financial Linkages*.
- Mitchell, S., Potash, E., Barocas, S., D'Amour, A., and Lum, K. (2018). Prediction-based decisions and fairness: A catalogue of choices, assumptions, and definitions. *arXiv preprint arXiv:1811.07867*.

- National Academies (2015). *The growing gap in life expectancy by income: Implications for federal programs and policy responses*. National Academies Press.
- National Academies of Sciences, E., Medicine, et al. (2019). *Integrating social care into the delivery of health care: moving upstream to improve the nation's health*. National Academies Press.
- Nishtala, S., Kamarthi, H., Thakkar, D., Narayanan, D., Grama, A., Padmanabhan, R., Madhiwalla, N., Chaudhary, S., Ravindra, B., and Tambe, M. (2020). Missed calls, automated calls and health support: Using ai to improve maternal health outcomes by increasing program engagement. *arXiv preprint arXiv:2006.07590*.

References XVI

- Obermeyer, Z. and Mullainathan, S. (2019a). Dissecting racial bias in an algorithm that guides health decisions for 70 million people. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 89–89.
- Obermeyer, Z. and Mullainathan, S. (2019b). Dissecting racial bias in an algorithm that guides health decisions for 70 million people. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 89, New York, NY, USA. Association for Computing Machinery.
- Osborn, S. R., Yu, J., Williams, B., Vasilyadis, M., and Blackmore, C. C. (2017). Changes in provider prescribing patterns after implementation of an emergency department prescription opioid policy. *The Journal of emergency medicine*, 52(4):538–546.

References XVII

- Penman-Aguilar, A., Talih, M., Huang, D., Moonesinghe, R., Bouye, K., and Beckles, G. (2016). Measurement of health disparities, health inequities, and social determinants of health to support the advancement of health equity. *Journal of public health management and practice: JPHMP*, 22(Suppl 1):S33.
- Quisel, T., Kale, D. C., and Foschini, L. (2016). Intra-day activity better predicts chronic conditions. *arXiv preprint arXiv:1612.01200*.
- Rahmattalabi, A., Adhikari, A. B., Vayanos, P., Tambe, M., Rice, E., and Baker, R. (2018). Influence maximization for social network based substance abuse prevention. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

References XVIII

- Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., et al. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1(1):18.
- Rollston, R. and Galea, S. (2020). Covid-19 and the social determinants of health. *American Journal of Health Promotion*, 34(6):687–689. PMID: 32551932.
- Rose, G. (1985). Sick individuals and sick populations, int. journ. *Of Epidemiology*, 14(1).
- Rose, S., Bergquist, S. L., and Layton, T. J. (2017). Computational health economics for identification of unprofitable health care enrollees. *Biostatistics*, 18(4):682–694.

References XIX

- Schwartz, S., Prins, S. J., Campbell, U. B., and Gatto, N. M. (2016). Is the “well-defined intervention assumption” politically conservative? *Social science & medicine* (1982), 166:254.
- Snyder, J. J., Salkowski, N., Wey, A., Pyke, J., Israni, A. K., and Kasiske, B. L. (2018). Organ distribution without geographic boundaries: a possible framework for organ allocation. *American Journal of Transplantation*, 18(11):2635–2640.
- Vyas, D. A., Eisenstein, L. G., and Jones, D. S. (2020). Hidden in plain sight—reconsidering the use of race correction in clinical algorithms.

- Ward, J. B., Gartner, D. R., Keyes, K. M., Fliss, M. D., McClure, E. S., and Robinson, W. R. (2019). How do we assess a racial disparity in health? distribution, interaction, and interpretation in epidemiological studies. *Annals of epidemiology*, 29:1–7.
- Weichenthal, S., Olaniyan, T., Christidis, T., Lavigne, E., Hatzopoulou, M., Van Ryswyk, K., Tjepkema, M., and Burnett, R. (2020). Within-city spatial variations in ambient ultrafine particle concentrations and incident brain tumors in adults. *Epidemiology (Cambridge, Mass.)*, 31(2):177.
- WHO (2008). *Closing the gap in a generation: Health equity through action on the social determinants of health: Commission on Social Determinants of Health final report*. World Health Organization.

- WHO (2009). 2008-2013 action plan for the global strategy for the prevention and control of noncommunicable diseases: prevent and control cardiovascular diseases, cancers, chronic respiratory diseases and diabetes.
- WHO (2010). A conceptual framework for action on the social determinants of health.
- Wu, Y., Zhang, L., Wu, X., and Tong, H. (2019). P_c-fairness: A unified framework for measuring causality-based fairness. In *Advances in Neural Information Processing Systems*, pages 3404–3414.
- Yang, K., Loftus, J. R., and Stoyanovich, J. (2020). Causal intersectionality for fair ranking. *arXiv preprint arXiv:2006.08688*.

References XXII

- Zhan, A., Mohan, S., Tarolli, C., Schneider, R. B., Adams, J. L., Sharma, S., Elson, M. J., Spear, K. L., Glidden, A. M., Little, M. A., et al. (2018). Using smartphones and machine learning to quantify parkinson disease severity: the mobile parkinson disease score. *JAMA neurology*, 75(7):876–880.
- Zhao, Y., Mirin, N., Wood, E., Dorice, V., Rajesh, V., Cook, S., and Chunara, R. (2020). Machine learning for integrating social determinants in cardiovascular disease prediction models: A systematic review.