Big data, social determinants of health, and health inequities

Sandro Galea

Boston University School of Public Health
1. The trouble with the population’s health
Life expectancy vs. health expenditure over time (1970-2014)

Health spending measures the consumption of health care goods and services, including personal health care (curative care, rehabilitative care, long-term care, ancillary services and medical goods) and collective services (prevention and public health services as well as health administration), but excluding spending on investments. Shown is total health expenditure (financed by public and private sources).

Our World in Data. “The link between health spending and life expectancy: The US is an outlier.”
Andrasfay T, Goldman N. Reductions in 2020 US life expectancy due to COVID-19 and the disproportionate impact on the Black and Latino populations. PNAS 2021 Vol. 118 No. 5 e2014746118
2. Understanding population health
The health outcomes of a group of individuals, including the distribution of such outcomes within the group
“The health outcomes of a group of individuals, including the distribution of such outcomes within the group”

Why? So that we may intervene

3. Four principles that can help
a. Population health data as continuous, not binary
Figure 1. Distribution of BMI in two populations illustrating health as a continuum in the population.

Panel A

BMI = 30
Mean = 23
Figure 3 Percentage distribution of serum cholesterol levels (mg/dl) in men aged 50–62 who did or did not subsequently develop coronary heart disease (Framingham Study^5^)
Figure 3  Percentage distribution of serum cholesterol levels (mg/dl) in men aged 50–62 who did or did not subsequently develop coronary heart disease (Framingham Study).

**FIGURE 2.** Probability distributions of a marker, $X$, in cases (solid curves) and controls (dashed curves) consistent with the logistic model $\log \frac{P(D=1|X)}{1-P(D=1|X)} = \alpha + \beta X$. It has been assumed that $X$ has a mean of 0 and a standard deviation of 0.5 in controls so that a unit increase represents the difference between the 84th and 16th percentiles of $X$ in controls. The marker is normally distributed, with the same variance in cases. The odds ratio (OR) per unit increase in $X$ is shown.
b. Illuminating ubiquitous causes
The goldfish are surrounded by water and everything they do is influenced by the quality of the water in which they live; therefore, water is a ubiquitous factor influencing the fish and needs to be taken into consideration every time we may want to improve the lives of the fish.
Crack Babies: The Worst Threat Is Mom Herself

By Douglas J. Besharov

LAST WEEK in this city, Greater Southeast Community Hospital released a 7-week-old baby to her homeless, drug-addicted mother even though the child was at severe risk of pulmonary arrest. The hospital's explanation: "Because [the mother] demanded that the baby be released."

The hospital provided the mother with an apnea monitor to warn her if the baby stopped breathing while asleep, and trained her in CPR. But on the very first night, the mother went out drinking and left the child at a friend's house—without the monitor. Within seven hours, the baby was dead. Like Dooney Waters, the 6-year-old living in his mother's drug den, whose shocking story was reported in The Washington Post last week, this child was all but abandoned by his parents.
<table>
<thead>
<tr>
<th>Predictor for Peabody Picture Vocabulary Test score</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gestational cocaine exposure</td>
<td>-2.89</td>
<td>0.26</td>
</tr>
<tr>
<td>Assessment no.</td>
<td>2.72</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Gestational cocaine exposure x assessment no.</td>
<td>0.58</td>
<td>0.51</td>
</tr>
<tr>
<td>Age at 1st assessment</td>
<td>-0.36</td>
<td>0.76</td>
</tr>
<tr>
<td>Female gender</td>
<td>-4.93</td>
<td>0.058</td>
</tr>
<tr>
<td>Parental nurturance</td>
<td>-0.31</td>
<td>0.89</td>
</tr>
<tr>
<td>Environmental stimulation</td>
<td>5.91</td>
<td>0.039</td>
</tr>
<tr>
<td>Caregiver BDI-II depression score</td>
<td>0.03</td>
<td>0.84</td>
</tr>
</tbody>
</table>

STOP DRINKING ALCOHOL
By Craig Beck
<table>
<thead>
<tr>
<th>Food</th>
<th>20 Years Ago</th>
<th>Today</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bagel</td>
<td>140 calories (3” diameter)</td>
<td>350 calories (6” diameter)</td>
</tr>
<tr>
<td>Muffin</td>
<td>210 calories (1.5 oz)</td>
<td>500 calories (4 oz)</td>
</tr>
<tr>
<td>Cheeseburger</td>
<td>333 calories</td>
<td>590 calories</td>
</tr>
<tr>
<td>Pasta (Spaghetti &amp; Meatballs)</td>
<td>500 calories</td>
<td>1025 calories</td>
</tr>
<tr>
<td>French Fries</td>
<td>210 calories (2.4 oz)</td>
<td>610 calories (6.9 oz)</td>
</tr>
<tr>
<td>Soda</td>
<td>85 calories (6.5 oz)</td>
<td>250 calories (20 oz)</td>
</tr>
<tr>
<td>Theater Popcorn</td>
<td>270 calories (5 cups)</td>
<td>630 calories (1 tub)</td>
</tr>
<tr>
<td>Turkey Sandwich</td>
<td>320 calories</td>
<td>820 calories</td>
</tr>
<tr>
<td>Pizza</td>
<td>500 calories (2 slices)</td>
<td>850 calories (2 calories)</td>
</tr>
</tbody>
</table>
Poor food environment in New York City

c. The role of co-occurring causes
How much of our obesity risk is determined by our genes?
= GE+
= ENV+
Scenario 1

= GE+  = OB+  = ENV+
Scenario 1

= GE+  = OB+  = ENV+
Scenario 1

RR (OB+IGE+) = 3.99
PARP (OB+IGE+) = 1

= GE+      = CA+      = ENV+
Scenario 2

= GE+  = OB+  = ENV+
Scenario 2

= GE+  = OB+  = ENV+
Scenario 2

RR (OB+|GE+) = 1.7
PARP (OB+|GE+) = 0.4

= GE+  = OB+  = ENV+
Therefore under a very plausible assumption of co-occurring causes, the gene-obesity association can only be understood if we understand the urban factors that create the conditions for disease
Figure 3. Predicted body mass index (BMI), calculated as weight in kilograms divided by height in meters squared, as a function of residualized age- and sex-specific ln-transformed physical activity accelerometer counts according to FTO rs1861868 genotypes. On the left side of the plot (low physical activity), BMI levels are strikingly dissimilar between rs1861868 genotypes. In contrast, on the right side of the plot, similar BMI levels can be seen across genotypes, particularly in subjects with very high levels of physical activity.
d. The centrality of health equity
Figure 5: Widening income-related inequalities in survival across birth cohorts
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. Gaining overall population health while increasing health inequity

- **No intervention**
  - DALY = 50
  - DALY = 25
  - Inequality of 25 DALY

- **Intervention adding 1 DALY**
  - DALY = 51
  - DALY = 25.5
  - Inequality of 25.5 DALY

- **Intervention adding 10 DALY**
  - DALY = 60
  - DALY = 30
  - Inequality of 30 DALY
Figure 2. Gaining overall population health while creating health inequalities

<table>
<thead>
<tr>
<th>Before intervention</th>
<th>After intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High SES</td>
</tr>
<tr>
<td>Daly = 50</td>
<td>Daly = 60</td>
</tr>
<tr>
<td>Inequality of 5 Daly</td>
<td></td>
</tr>
<tr>
<td>Daly = 50</td>
<td>Daly = 55</td>
</tr>
<tr>
<td>No inequality</td>
<td></td>
</tr>
</tbody>
</table>
4. Methods for population health
Populations are

1. Heterogeneous, ie have diversity of agents
2. Characterized by nonlinear dynamics
3. Characterized by contact structure, networks, organization
4. Have feedback, adaptation, learning, evolution
5. Stochastic with important tails
6. Display emergent properties
Populations are

1. Heterogeneous, ie have diversity of agents
2. Characterized by nonlinear dynamics
3. Characterized by contact structure, networks, organization
4. Have feedback, adaptation, learning, evolution
5. Stochastic with important tails
6. Display emergent properties

Complex systems
Figure 5.2: The full obesity system map with thematic clusters (see main text 5.1.2 for discussion). Variables are represented by boxes, positive causal relationships are represented by solid arrows and negative relationships by dotted lines. The central engine is highlighted in orange at the centre of the map.
Causes and counterfactuals

Observed

Counterfactual (parallel universe)
Causes and counterfactuals

Observed

Counterfactual (parallel universe)
Causes and counterfactuals

Observed

Counterfactual (parallel universe)
<table>
<thead>
<tr>
<th>Study Design</th>
<th>Idealized Randomized Controlled Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population of Interest</td>
<td></td>
</tr>
<tr>
<td>Manipulation</td>
<td></td>
</tr>
<tr>
<td>Exposure or Treatment Distribution</td>
<td></td>
</tr>
<tr>
<td>Disease Distribution</td>
<td></td>
</tr>
<tr>
<td>Causal Effect?</td>
<td>Yes</td>
</tr>
</tbody>
</table>

“Everything should be made as simple as possible, but not simpler”

Attributed to Albert Einstein.
Simple approaches, a foundational myth
The effectiveness of simple approaches?

5. The big picture
Figure 2.1 Levels of influence on the health of populations. Source: Modified from Kaplan, G. What’s wrong with social epidemiology, and how can we make it better? Epid Rev. 2004; 26:124–135.
The global decision-maker survey

Learning directly from stakeholders about the best ways to make data on determinants relevant to decision-making is critical to the 3-D Commission's approach. If you identify as a decision-maker, please share your thoughts by taking our short, anonymous survey.

Take the survey