
CHEAT SHEET

Disparity in Data

Why health care data is not as objective as we expect

Published - June 03, 2020 • 5-min read

Key takeaways

- Health care data reflects underlying disparities. Populations that face socioeconomic barriers to accessing care are underrepresented or misrepresented in our systems.
- The “digital divide” in health care contributes to disparities in data. Digital technologies provide a new source of patient-generated health data, but only for patients who can afford and access them.
- Data enables analytics and artificial intelligence (AI), but we must understand the shortcomings of our data and the social implications of using biased data to drive care decisions. Marginalized populations that are left out of data will not reap the benefits of AI.

What is it?

Data is often considered an objective source of truth. But there are underlying issues in health care data that can lead to skewed inferences and decisions.

- 1. Incomplete data:** Gaps in data prevent us from having a holistic view of a patient or population. Marginalized groups may experience more fractured care and less documentation of conditions and outcomes. Demographics and social determinants of health (SDOH) influence outcomes but aren't routinely captured in systems. Only one-third of commercial plans reported having complete or partially complete data on race, a pattern that is likely reflected in electronic health records (EHRs).
- 2. Small sample size:** Marginalized populations are not adequately represented in health care data. More data is available for those that are able to access care and treatments, and "data deserts" exist for groups that experience systemic barriers to accessing care. Underrepresentation in data can lead to less informed care decisions or flawed inferences about a population.
- 3. Historical inequities embedded in data:** Health care outcomes are not the same across populations. For example, black women are 42% more likely to die from breast cancer. This can be at least partially attributed to factors like a higher burden of comorbidities and barriers to accessing care that stem from the enduring legacies of structural racism and intergenerational poverty. Black women are also more likely to be diagnosed at later stages of the disease and experience delays in treatment of two or more months. These types of inequitable outcomes are baked into health care data.

Why does it matter?

Data is viewed by many leaders as a critical asset that will lead to better, more efficient health care. Advanced analytics and artificial intelligence can turn complex EHR data into actionable insights that improve decision-making and care delivery. But if EHRs are to become the new textbooks for health care, we must consider the disparities that live in their data.

The IT adage “garbage in, garbage out” suggests that flawed inputs will produce poor outputs. The same is true for bias—bias in, bias out. Algorithms learn from historical patterns to make predictions and decisions, but if they learn from biased data they will produce biased outputs. By using these insights to inform care decisions, systems may unintentionally create or perpetuate inequities.

For example, one emerging application of AI is predicting intensive care unit (ICU) demand. Algorithms can be used to identify which inpatients are at risk for clinical deterioration and will require a transfer to an ICU. A model could be built using historical health records of patients who were transferred to ICUs. But if the training data contains more white than black patients, the model will make better predictions for white patients. Deterioration might be underestimated for black patients, leading to fewer transfers and worse outcomes.

Structural inequities in health care prevent marginalized groups from accessing timely and quality care. These inequities are embedded in health care data. If we aren't careful, analytics and AI can automate and scale health disparities.



It's not ideal to make broad-sweeping suggestions in communities where we have data deserts.

Robert Winn

Physician and director, Virginia Commonwealth University (VCU) Massey Cancer Center

How does it happen?

Data disparities, while unintentional, can be introduced by a number of factors.

- **Data gaps or lags occur when disadvantaged groups experience barriers in access to care.** Data collection is skewed toward those who can afford and access care. For example, a diagnosis documented in the EHR signifies when a patient accessed care and a provider made a diagnosis, but not when a patient first got sick. Populations with fragmented or limited care are underrepresented in health care systems. When data is available for these groups, it often includes the more severe cases and can misrepresent the overall health of the population.
- **Socioeconomics status influences where patients access care, and data collection varies across care sites.** Health care facilities don't all collect the same information in the same way. Patients of low socioeconomic status are more likely to seek care in teaching clinics or community hospitals, where data input may be less consistent due to constrained time and resources.
- **Providers' subjectivity can influence what data is captured during a visit.** Unstructured data like chart notes can reflect the implicit biases of providers. When it comes to demographic data, staff may be hesitant to have potentially sensitive conversations and may make assumptions about gender or race.
- **Health record systems and processes are not built to collect holistic patient data.** EHRs were built to be billing tools first and clinical tools second. They are often missing data and are not set up to easily capture demographic and SDOH data. Although there are SDOH codes and Z codes, providers have few incentives to track this information because it's not reimbursable.
- **Remote patient monitoring (RPM) devices produce additional data, but exacerbate the "digital divide."** RPM devices provide detailed data about a patient's health condition when they are not in a facility, but data is collected only for patients who can access and use devices. Populations who are less tech-savvy or don't have consistent internet access will be left out of these data sets.

1. Z codes are a subset of ICD-10-CM codes, used as reason codes to capture factors that influence health status and contact with health services.

Source: Gianfrancesco M, et al., "Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data," *JAMA*, Jan. 2019; [Z Codes Utilization among Medicare Fee-for-Service \(FFS\) Beneficiaries in 2017](#), CMS, Jan 2020, Advisory Board interviews and analysis.

Conversations to have

01

Understand how your data may reflect health disparities.

- What populations are missing or underrepresented in our data? Are our sample sizes representative of our community?
 - What types of information are not captured in our data? Consider clinical, demographic, community, and SDOH data.
 - How do health outcomes differ across population subgroups such as race, ethnicity, gender, and socioeconomic class?
-

02


Discuss ways to account for gaps in data.

- How can we capture more holistic data on our patients? Do our providers screen for SDOH when interacting with patients? Have we considered partnering with third parties?
 - How can we regularly engage community leaders to better inform our understanding of the patients we serve?
-


03


Determine what factors might be causing disparities in data and how you can address these issues.


- Do implicit biases of providers influence data collection and data entry? Do we provide system-wide training on bias?
 - What barriers to accessing care exist in our communities? How can we ease access to care for marginalized groups?
-


To identify the social implications of using analytics and AI, health care leaders must first understand the quality and completeness of their data. Only then can the data be useful for the entire patient population. 

Related content

 **OUR TAKE**
Addressing Discrimination in Artificial
Intelligence
[Read now](#)

 **INFOGRAPHIC**
How to Combat AI Bias
[Read now](#)

 **EXECUTIVE BRIEFING**
The 7 Simple Practices of Health Care Data
Governance
[Read now](#)

 **CHEAT SHEET**
Social Determinants of Health Data
[Read now](#)

 **RESEARCH REPORT**
Using IT to Help Address the Social
Determinants of Health
[Read now](#)

Project director

Mackenzie Rech

Research team

Ty Aderhold

Benjamin Wheeler

Madeleine Chaillet

Program leadership

Micha'le Simmons

John League

Darby Sullivan

LEGAL CAVEAT

Advisory Board has made efforts to verify the accuracy of the information it provides to members. This report relies on data obtained from many sources, however, and Advisory Board cannot guarantee the accuracy of the information provided or any analysis based thereon. In addition, Advisory Board is not in the business of giving legal, medical, accounting, or other professional advice, and its reports should not be construed as professional advice. In particular, members should not rely on any legal commentary in this report as a basis for action, or assume that any tactics described herein would be permitted by applicable law or appropriate for a given member's situation. Members are advised to consult with appropriate professionals concerning legal, medical, tax, or accounting issues, before implementing any of these tactics. Neither Advisory Board nor its officers, directors, trustees, employees, and agents shall be liable for any claims, liabilities, or expenses relating to (a) any errors or omissions in this report, whether caused by Advisory Board or any of its employees or agents, or sources or other third parties, (b) any recommendation or graded ranking by Advisory Board, or (c) failure of member and its employees and agents to abide by the terms set forth herein.

Advisory Board and the "A" logo are registered trademarks of The Advisory Board Company in the United States and other countries. Members are not permitted to use these trademarks, or any other trademark, product name, service name, trade name, and logo of Advisory Board without prior written consent of Advisory Board. All other trademarks, product names, service names, trade names, and logos used within these pages are the property of their respective holders. Use of other company trademarks, product names, service names, trade names, and logos or images of the same does not necessarily constitute (a) an endorsement by such company of Advisory Board and its products and services, or (b) an endorsement of the company or its products or services by Advisory Board. Advisory Board is not affiliated with any such company.

IMPORTANT: Please read the following.

Advisory Board has prepared this report for the exclusive use of its members. Each member acknowledges and agrees that this report and the information contained herein (collectively, the "Report") are confidential and proprietary to Advisory Board. By accepting delivery of this Report, each member agrees to abide by the terms as stated herein, including the following:

1. Advisory Board owns all right, title, and interest in and to this Report. Except as stated herein, no right, license, permission, or interest of any kind in this Report is intended to be given, transferred to, or acquired by a member. Each member is authorized to use this Report only to the extent expressly authorized herein.
2. Each member shall not sell, license, republish, or post online or otherwise this Report, in part or in whole. Each member shall not disseminate or permit the use of, and shall take reasonable precautions to prevent such dissemination or use of, this Report by (a) any of its employees and agents (except as stated below), or (b) any third party.
3. Each member may make this Report available solely to those of its employees and agents who (a) are registered for the workshop or membership program of which this Report is a part, (b) require access to this Report in order to learn from the information described herein, and (c) agree not to disclose this Report to other employees or agents or any third party. Each member shall use, and shall ensure that its employees and agents use, this Report for its internal use only. Each member may make a limited number of copies, solely as adequate for use by its employees and agents in accordance with the terms herein.
4. Each member shall not remove from this Report any confidential markings, copyright notices, and/or other similar indicia herein.
5. Each member is responsible for any breach of its obligations as stated herein by any of its employees or agents.
6. If a member is unwilling to abide by any of the foregoing obligations, then such member shall promptly return this Report and all copies thereof to Advisory Board.



655 New York Avenue NW, Washington DC 20001
202-266-5600 | [advisory.com](https://www.advisory.com)