Data

Ah Young (Amy) Kim

12/09/2018

Data Module: Table of Contents

- What is Bias?
- Types of Biases
- Effects of Biases in Data
- Causes of Biases from the Use of CDSS
- How to Avoid Biases?

What is Bias?

Bias

- "Statistical bias is a feature of a statistical technique or of its results whereby the expected value of the results **differs** from the true underlying quantitative parameter being estimate" (Wikipedia)
- "Bias refers to the tendency of a measurement process to **over** or **underestimate** the value of a population parameter" (Stat Trek)
- "Bias is systematic favoritism that is present in the data collection process, resulting in **lopsided**, **misleading** results" (Dummies)

Any biases in data can misrepresent the actual phenomenon at hand!

Types of Biases

Major types of biases that are more relevant to the CDSS are the followings:

- ► Selection bias: occurs when particular subsets of the entire population are sampled accidentally
- ► Exclusion bias: occurs when certain individuals are excluded from a study
- ► Automation bias: occurs when people favor any suggestion from automated decision-making systems even though their decision without those systems are correct

Example: Cabiza et al. (2012) Pneumonia Case Study

An ML-DSS(Machine Learning based Decision Support Systems) suggested that patients with pneumonia and asthma had a lower risk of death from pneumonia than those with pneumonia but without asthma

- Patient with a history of asthma who presented with pneumonia were admitted directly to intensive care units and had better outcomes than patients diagnosed with pneumonia and without a history of asthma
- This information could not be included in the ML-DSS

Example: Melanoma

Melanoma is a type of skin cancer, and it is known to be more common these days for in **women under 40** with **fairer skins**. Consider a registry data which collects incidence of melanoma in the U.S., and imagine a ML-DSS based on the registry data is used to predict the risk of melanoma. It is dangerous to rely on the ML-DSS suggestions because

- ▶ men over 40 may have lower risk of melanoma based on the ML-DSS suggestion because most incidences are women under 40
- ▶ the ML-DSS may suggest that people with darker skins have lower risk of melanoma

It is important to understand the data used to have better clinical outcomes!

Effects of Biases in Data: Char et al. (2018)

"Algorithms introduced in non-medical fields have already been shown to make problematic decisions that reflect biases inherent in the data used to train them. . . It's possible that similar racial biases could inadvertently be built into health care algorithms. Health care delivery already varies by race. An algorithm designed to predict outcome from genetic findings will be biased if there have been few (or no) genetic studies in certain populations. For example, attempts to use data from the Framingham Heart Study to predict the risk of cardiovascular events in nonwhite populations have led to biased results, with both overestimation and underestimations of risk."

Why Should We Be Aware of Biases?

- Friedman et al. (1999) found that in 6% of cases, clinicians over-rode their own correct decisions in favor of incorrect suggestions from the DSS
- Berner et al. (2003) noticed tat in 21 cases of 272 the correct decisions made without DSS suggestions was changed to incorrect decisions after DSS use
- Westbrook et al.(2006) also found that 7% of correct unaided answers were changed incorrectly after system use

These are called automation biases that arise from the use of DSS.

Causes of Biases from the Use of CDSS

- Experience: user experience with the DSS may cause them to be more/less reliant on the DSS
- Confidence: Dreiseitl and Binder (2005) showed that physicians were more likely to be biased by automation when they were less confident of their own decisions
- Individual differences: users' personality and cognitive characteristics may affect users making errors based on automation bias
- Task type: more complex tasks, higher workloads, high time pressure on tasks may affect reliance on the DSS and errors caused by automation bias

How to Avoid Biases

- Make users aware of the CDSS reasoning process
- Provide training on appropriate CDSS use to make users to be able to recognize errors and reduce automation bias
- Present additional information of the CDSS for users to prevent users from being over reliant

References (1/2)

[1] Berner, E. S., Maisiak, R. S., Heudebert, G. R., & Young Jr, K. R. (2003). Clinician performance and prominence of diagnoses displayed by a clinical diagnostic decision support system. In Amia annual symposium proceedings (Vol. 2003, p. 76). American Medical Informatics Association.

[2] Cabitza, F., R. Rasoini, and G.F. Gensini, Unintended consequences of machine learning in medicine. Jama, 2017. 318(6): p. 517-518.

[3] Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. The New England journal of medicine, 378(11), 981.

[4] Dreiseitl, S., & Binder, M. (2005). Do physicians value decision support? A look at the effect of decision support systems on physician opinion. Artificial intelligence in medicine, 33(1), 25-30.

References (2/2)

[5] Friedman, C. P., Elstein, A. S., Wolf, F. M., Murphy, G. C., Franz, T. M., Heckerling, P. S., ... & Abraham, V. (1999). Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. Jama, 282(19), 1851-1856.

[6] Westbrook, J. I., Coiera, E. W., & Gosling, A. S. (2005). Do online information retrieval systems help experienced clinicians answer clinical questions?. Journal of the American Medical Informatics Association, 12(3), 315-321.