

Health Data Science: Homework #1
Karis Van Noord

1. What is the primary goal of data anonymization?
 - a. The primary goal of data anonymization is to protect the identities of those whose data is being collected and stored. The process of anonymizing data erases or encrypts the specific connections between an individual and their data that is being stored. The goal is to make it nearly impossible to identify the individual, while simultaneously still allowing their information to be useful for data analysis.
2. What is the Havasupai Tribe case about?
 - a. In the 1990s a study was conducted to research the high prevalence of Type 2 diabetes in the Havasupai Tribe in Arizona. With the tribe's consent, the researchers from Arizona State University collected blood samples and investigated the predisposition to Type 2 diabetes. However, without the tribe's consent, the researchers retained these blood samples, and studied genetic markers for schizophrenia, metabolic disorders, and the tribe's geographical origins. The tribe discovered how their blood samples had been utilized for further research beyond the tests that had been consented to, and felt taken advantage of, as these tests went against some of the tribe's religious beliefs.
3. What was the largest data breach in history?
 - a. The largest data breach in history took place in early 2015. Anthem Inc. was the victim of a cyber attack that leaked the personal information of nearly 78.8 million individuals. This cyber attack exposed many individuals to identity theft and fraud, and it exposed the vulnerabilities of health data security systems and the dangers of future cyber leaks.
4. What is the pulse oximeter controversy about?
 - a. The pulse oximeter controversy is an issue regarding bias in health data. This device is commonly used worldwide to measure levels of oxygen in the blood. These devices are minimally invasive (placed on a patient's finger), yet they provide vital information for doctors and surgeons. In recent years, pulse oximeters have been found to be three times more likely to miss hypoxemia or low blood oxygen levels in Black patients compared to White patients. Inaccurate results reported by pulse oximeters can have far-reaching implications, including inappropriate treatment plans and the inability to identify critical health crises in many individuals. These devices are biased because the algorithms used to develop them were based largely on data that was collected from light-skinned individuals. This situation demonstrates the prevalence of hidden bias in healthcare and is a reminder to actively seek out and overcome these biases.
5. What is the role of an Institutional Review Board (IRB) in research?

- a. Institutional review boards are committees of independent members including scientists, non-scientists, and community members who review and ensure ethical standards of research studies that involve humans. Before such studies can commence, research protocols must be reviewed and approved by an IRB. These committees assess the study design, methodology, and procedures. The IRB ensures that consent is given and that there are measures in place to protect participant's privacy and confidentiality.
6. What is the difference between spreadsheet software and database software for health data management?
 - a. In spreadsheet software management, data is stored in a single cell and can be edited, formatted, and changed within that cell. In a database, cells containing specific information cannot be manipulated once they are in the database, due to these records coming from external sources. There are advantages and disadvantages when it comes to both methods of data entry. The only criteria for health data management is that the correct individuals have access to the information.
7. What is a conflict of interest in public health research?
 - a. There are many different kinds of conflicts of interest in public health research, including financial conflicts, personal conflicts, and academic conflicts. Conflicts of interest in public health research occur when a secondary gain is thrown into the mix, clouding the validity of findings and/or the integrity of research. These secondary gains can include financial gain, personal advancement, or institutional benefit. For example, if a smoking company conducts research into the effects of smoking, these findings will often be skewed, resulting in the downplaying of the potentially harmful effects of smoking.
8. What is informed consent in the context of health data collection?
 - a. In the context of health data collection this ensures that individuals are provided with sufficient information to make informed decisions about the care that they receive. This informed consent is important because it not only respects personal autonomy and decision-making but it protects the privacy and confidentiality of each patient's unique information.
9. What is data masking in the context of de-identification of health data?
 - a. The process of de-identification of health data includes anonymizing each patient's data to remove identifiable factors that could connect someone's data to their person. Data masking replaces a patient's identifiable information with randomized characters to protect their privacy.
10. What is the principle of k-anonymity in the context of de-identification of health data?
 - a. K-anonymity is the practice of grouping data together if each group contains at least k records. This strategy of data anonymization makes it difficult to single out and identify the patients/data in each group.

11. What is the main concern about conflicts of interest in public health research?
 - a. The main concerns about conflicts of interest in public health research include bias in research, loss of trust in public health institutions/researchers, and damage to research integrity. All of these potential concerns can decrease public engagement with research and alter the results of research.
12. What is the first step in managing conflicts of interest in public health research?
 - a. There are several steps that can be taken in order to manage the prevalence of conflicts of interest in public health research. These steps include (but are not limited to) avoidance, divestment, management plans, public disclosure, institutional policies, and education and training. These steps decrease the likelihood and in some cases avoid altogether the occurrence of conflicts of interest in public health research.
13. What is the role of independent oversight in managing conflicts of interest in public health research?
 - a. Independent oversight is imperative in public health research. Similar to IRB's, these oversight committees review research proposals for potential conflicts of interest, monitor ongoing research, and review findings before publication. These committees also ensure that conflicts of interest do not take place.
14. What is one strategy for mitigating conflicts of interest in public health research?
 - a. Some of these strategies can include divestment, meaning if there is a financial conflict of interest, a researcher or institution must divest (remove their stake) from the conflicting financial interest.
15. What is the potential impact of not properly managing conflicts of interest in public health research?
 - a. Not properly managing conflicts of interest can result in skewed results, loss of professional judgment, and a public mistrust in public health institutions and research.