

Collecting Big Data

Have you ever hit the “accept” button without carefully reading the terms or conditions for a program on your computer or an App on your cell phone or tablet? If so, you may already be a participant in a big data study that collects current data and future data related to your life and your health.

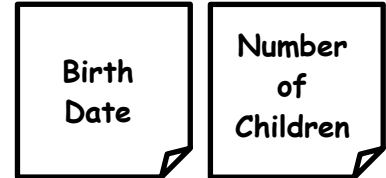
You may have granted big data researchers access to **electronic data sources** such as:

- A. Health information from doctors, hospitals, and insurance records
- B. Computer files, search engines used, and Apps on digital devices
- C. Online purchases, store reward cards, credit card payments
- D. Wearable devices, home security systems, or home control devices

1. Your teacher provided a sheet of poster paper with the **electronic data sources** assigned to your team.

2. **What kinds of data (information about you or other people) might be stored in these electronic data sources?** You will have 10 minutes to brainstorm with your team to make a set of sticky notes. Each sticky note should list one possible type of data that could be obtained from the electronic data sources assigned to your team.

Use the colored marker and write neatly.



3. Attach your sticky notes to your team’s poster.

4. You will have 5 minutes to observe and add sticky notes to one of the posters made by another team. Read the title of the poster and read what previous teams have written on the sticky notes. Use your marker and sticky notes to add **new** ideas for types of data to the poster.

5. Repeat observing the posters until you have read and added **new** ideas for types of data to each of the posters.

Big data health research involves using electronic sources of information and new technologies to create huge data sets containing many different types of data. The data collected includes many potential **variables** without identifying which might be used as dependent or independent variables.

Not all data collected will be relevant to health issues. However, it is possible that big data analysis may result in the discovery of unanticipated connections, patterns or trends. Sometimes new variables are discovered that show trends or patterns related to health issues.

A **variable** is any factor, trait, or condition that can exist in differing amounts or types.

6. Give three examples of variables from the class sticky notes that are likely to be linked to human disease.

- _____
- _____
- _____

7. Give three examples of variables from the class sticky notes that are unlikely to be related to human diseases.

- _____
- _____
- _____

8. Should big data scientists limit the data they collect to variables (factors) that are clearly identified as related to health or disease? Explain why or why not.

The **right to privacy** refers to the concept that one's personal information (including health information) may be kept confidential and may be protected from becoming public knowledge. Some experts are concerned that big data research may allow unregulated access to personal information, and this can threaten individuals' rights to privacy. Other experts think that the potential benefits of big data research far outweigh the risks to individuals' privacy.

9. With your team, look at the posters and select and list four types of data that people might want to keep private.

- _____
- _____
- _____
- _____

10. Do you think the benefits from big data research outweigh the risks to individuals' privacy?

11. What actions might be taken to reduce the risks that big data research will interfere with individuals' rights to privacy?
