Data, Information, and Indicators*

Marvin L. Birnbaum, MD, PhD

*Expanded from Sundnes KO, Birnbaum ML, and the Steering Committee of the Task Force for Quality Control of Disaster Medicine: "Information" in Volume 2 of the Health Disaster Management: Guidelines for Evaluations and Research in the Utstein Style, to be published as a Supplement to *Prehospital and Disaster Medicine* in 2008 (in press)

It is better to understand a little than misunderstand a lot.

Anatole France
Revolt of the Angels, Ch 1

In 2001, I wrote an editorial on research data and information and how each should be incorporated into a research paper.[1] In that editorial, I did not clearly define the differences between data and information nor the appropriate uses of the terms reliability, reproducibility, or validity.

Confusion about this issue arose for the first time for me during a meeting with the Steering Committee of the Task Force for Quality Control of Disaster Medicine in which we were trying to distinguish between data and information. I explored the uses of these terms further with Dr. Knut Ole Sundnes during our work on a chapter on Information that is part of Volume 2 of the Guidelines[2], and it became apparent that many of us did not understand these terms and their appropriate use. Furthermore, many authors who have published in PDM and other scientific publications also have had difficulty with the appropriate use of these terms. So, let's take another look.

The term *data* is defined as known facts or things (or observations) used as a basis for inference to produce the needed information.[3] Data is the plural form of "datum". Data can be collected using quantitative techniques, qualitative techniques, or a combination of both. Data are collected to help to answer a question; collection of data that are without a clear relationship to the question being posed, not only is irrelevant, but is frivolous and expensive and has high opportunity costs.

Important characteristics of data include their reliability and reproducibility. Reliable means to be capable of being relied on; of sound and consistent character or quality.[4] The reliability of the data means that exactly the same data will be obtained regardless of the processes used, the tools used, and/or who is collecting the data. To reproduce means to produce a copy or representation of.[5] Reproducibility in the process/method is used to denote that when using the same method in the same circumstances, in the same population, the same data will be obtained. If the process is reproduced, but different data are obtained for that indicator, a change has occurred. Thus, if repeated assessments using a specific indicator are conducted during a control period (pre-event) and yield the same data, but following an intervention or event, yield different data, then a change in the data for the chosen indicator has occurred, indicating that an intervention or event has produced an effect. The significance of such changes is interpreted through processing of the data.

The use of the term "validity" when referring to the data collected is a misnomer and is not related directly to the data. To be valid is to be sound or defensible; well-grounded.[6] Thus, validity relates to the conclusions derived, inferred, or interpreted from the data—giving the data meaning. It is this process of converting the data to information and its interpretation that must be validated. The data collected may be unreliable, non-reproducible, inappropriate, or inaccurate, but the data cannot be invalid.

Data by themselves do not have meaning. In order for data to become meaningful, they must be processed. Data processing includes interpretation or making inferences. Inference and interpretation require human input, and therefore, may be accurate or inaccurate and may or may not be valid. Consequently, the inferences and information obtained from the data have a degree of internal and external validity. If an inference leads to proving a cause:effect relationship, it is considered to have a high degree of internal validity. When the inference derived from the data can be generalized to other situations, the inference has a high degree of external validity. The process of deriving the information (e.g., statistical processing) may have errors embedded in it due to the methods used.

Since raw data are not meaningful, the term "data analysis" also is a misnomer. To analyze means to separate into its component parts.[7] For example, a substance may be analyzed to identify its chemical components. A situation may be analyzed to see if and where there are any problems. A problem may be analyzed in an attempt to identify the factors responsible for the problem. Information is analyzed to find out whether the conclusions derived from the data are accurate, reproducible, reliable, and valid or whether the inferential process was appropriate and adequate. Thus, information and conclusions, and not the data, can be analyzed. The data are processed into information.

An indicator is a thing that indicates; a sign or marker that defines the status of a specific component.[8] Assessments using indicators produce data relevant to the indicator. Indicators are chosen because, in the judgment of the persons who require the information that can be derived from them, they are likely to provide answers to the question being posed. The ability of indicators to reflect what is being sought is called construct validity. For example, data collected relative to the bacterial content of water has good construct validity to inform whether the water is contaminated. The collection of data that do not have accurate construct validity is frivolous and expensive and upon interpretation may lead to an incorrect conclusion. The problem, then, lies in the choice of appropriate indicators that, when processed, will lead to an appropriate and valid conclusion.

Thus, it is essential that appropriate indicators are defined prior to any attempts to collect the data. Inferencing is the process of deducing new information from information you already know.

Information is knowledge derived from study, experience, or instruction; a collection of facts or data e.g., statistical information.[9] Decisions are made based on information that is derived from the data available. Information gives meaning to the data and is transmitted among people. The information used for decision-making depends on the expertise of those who select the indicators and those who interpret the data.

The same terminology applies to the data and information regardless of whether the indicators chosen are quantitative or qualitative. Research studies attempt to answer one or more question or propose a hypothesis. Selecting the best possible indicators will improve the ability to answer the question posed, and will direct how the data are acquired and processed to yield valid interpretations and conclusions. The better and more clearly defined the indicators selected and processes used, the better will be the science upon which we base our practice.

Comprehension must be the soil in which grow all the fruits of friendship.

Woodrow Wilson
Address
Mobile, Alabama, 1913

The improvement of understanding is for two ends: first, for our own increase of knowledge; second, to enable us to deliver and make out that knowledge to others.

John Locke
Some Thoughts Concerning Reading and Study, Appendix B

References

1. Birnbaum ML: Data, information, results, discussion, hypotheses, and theory. *Prehospital Disast Med* 2001(1);1–2.
2. Arnold JL, Levine BN, Manmatha R, Lee F, Shenoy P,Tsai MC, Ibrahim TK, O'Brien DJ,Walsh DA: Information-sharing in out-of-hospital disaster response:The future role of information technology. *Prehospital Disast Med* 2004;19(2):201–207.
3. Sundnes KO, Birnbaum ML (eds) and the Task Force for Quality Control of Disaster Management: Health disaster management: Guidelines for evaluation and research in the Utstein style. *Prehospital Disast Med* 2003;17(Suppl 2):148.
4. Thompson D (ed): *The Concise Oxford Dictionary of Current English.* 9th ed, Clarendon Press: Oxford, 1995, p 1161.
5. *Ibid,* p 1168.
6. *Ibid,* p 1548.
7. Thompson, p 44.
8. Sundnes, 154.
9. Kleinelder SR (senior ed): *American Heritage College Dictionary,* 4th ed, New York: Houghton Mifflin Co., 2002, p 712.