Letters

Qualitative, Quantitative, and Mixed-Methods Research

I followed with great interest the discussion about descriptive versus hypothesis-driven research (A. Casadevall and F. C. Fang, Microbe, December 2008, p. 552–553, and May 2009, p. 207; J. L. Slonczewski and E. S. Kaneshiro, Microbe, February 2009, p. 50). The authors talked about various ways of defining “descriptive research” and evaluated its value to science.

My understanding is that quantitative research is typically associated with hypothesis-driven, experimental approaches. An example is bench-level research in microbiology which is characterized by structured observation, hard data, control group, deductive approach, and use of statistical tools, among others. Another form of research is purely descriptive. The description of an observed phenomenon (e.g., the discovery of a microorganism in a particular environmental setting) without having an immediate explanation for its existence might qualify as an example.

In epidemiology, descriptive (or observational) research relates to different types of study designs (e.g., case reports and case series, ecologic and cross-sectional studies, as well as cohort and case-control studies), which are considered nonexperimental and noninterventional. This is in contrast to experimental study designs (e.g., clinical trials and time series). Another way to describe these study designs is by dividing them into descriptive studies (e.g., case reports, case series, and ecologic studies) and analytical studies (e.g., cross-sectional and cohort studies, case-control studies, and clinical trials). A common sequence in epidemiology research may include clinical observations (i.e., recognizing a new or increased pattern of disease), descriptive studies (describing the distribution of disease and identifying clues for further investigation), and analytical studies (testing a specific hypothesis that Exposure A leads to Disease B) [J. L. Kelsey, A. S. Whittemore, A. S. Evans, and W. D. Thompson, Methods in observational epidemiology (2nd ed.), Oxford University Press, New York, NY, 1996; K. E. Nelson and C. M. Williams (eds.), Infections disease epidemiology: theory and practice (2nd ed.), Jones and Bartlett Publishers, Boston, Mass., 2007].

In the social sciences (e.g., anthropology, sociology, and psychology), the umbrella term “qualitative” approach is used to refer to a research design that is evolving and flexible; it creates descriptive (soft and rich) data through observation (e.g., people’s own words, field notes, and official documents/artifacts) and analysis is ongoing and inductive, leading to the description of multiple realities, to understanding and meaning, and the development of sensitizing concepts. The qualitative research approach is aimed at examining the world with the assumption that “nothing is trivial, that everything has the potential of being a clue that might unlock a more comprehensive understanding of what is being studied” [R. C. Bogdan and S. K. Biklen, Qualitative research for education: an introduction to theories and methods (5th ed.), Pearson Education, Boston, Mass., 2007].

Finally, there is mixed-methods research. This approach uses both qualitative and quantitative methods and blends the resulting data in certain ways. An increasing number of scientists considers this type of research the most powerful of all. Investigators in various fields (including public health and education) and across professional disciplines have begun to use it in order to get a more holistic view on topics under study [e.g., L. Curry, R. Shield, and T. Wele (eds.), Improving aging and public health research: qualitative and mixed methods, American Public Health Association, Washington, D.C., 2006; J. W. Creswell and V. L. P. Clark, Designing and conducting mixed methods research, Sage Publications, Thousand Oaks, Calif., 2007].

In conclusion, I believe there is not one single method that is inferior to another. In this sense, I agree with Casadevall and Fang as well as with Slonczewski and Kaneshiro that “descriptive research can illuminate novel phenomena or give rise to novel hypotheses” and that we should “avoid privileging one mode of discovery over another,” respectively. I believe that the choice of research design solely depends on the kind of question an investigator wants to answer. On a final note: My presentation here of information about different types of research approaches is essentially descriptive—nevertheless, I hope it will shed some more light on the terminology of various research designs used to advance our understanding of the world in which we live.

Christian T. K.-H. Stadtländer
University of St. Thomas
Minneapolis-St. Paul, Minn.

The First Paper in Bioinformatics?

In the early years of molecular biology, once it became known that DNA was the hereditary material and that DNA encoded information exactly specifying protein sequences, a number of coding schemes were proposed to explain how various combinations of the four unique bases of DNA might specify particular amino acids. It is obvious that at least three bases would be required to form “words” (codons in contemporary parlance) corresponding to the 20 unique amino acids. A doublet code could specify at most 16 (4²) amino acids while a triplet code, the most parsimonious option, could easily specify 20 amino acids, with some degeneracy, because it allows for 64 (4³) unique codons. A more vexing and less tractable problem was related to the actual
reading of the code: was the code an overlapping one? That is, once the (unknown) reading machinery had sensed a triplet, did it advance by three bases to the next triplet, or did it advance by only one base, sensing a triplet that had two letters in common with the preceding one?

In his 1957 paper addressing this problem (S. Brenner, Proc. Natl. Acad. Sci. USA 43:687–694, 1957), Sydney Brenner noted that, in the event of the code being an overlapping one, the sharing of two bases among adjacent codons automatically imposed constraints on the identity of neighboring amino acids. Thus, the identity of each successive amino acid residue in the polypeptide chain is constrained by the identity of the preceding one, and four bases would be required to specify any given dipeptide. This would imply that only certain amino acid pairs could occur as neighbors, and other pairs would be forbidden. Therefore, there could not be more than 256 (i.e., $4^4$) unique dipeptides, however large the set of protein sequences. But, in 1957, fewer than 256 dipeptides were known and therefore, a verdict could not be given based on the number of unique dipeptides observed in nature. On the other hand, if the code were nonoverlapping, 400 ($20 \times 20$) unique dipeptides would be found to occur if a sufficiently large set of protein sequences were available.

Brenner’s elegant solution to this problem was to use the limited protein data to draw inferences about the genetic code by counting the observed number of amino acid neighbors (N- or C-terminal) for each of the 20 amino acids. As noted earlier, an overlapping code implies the sharing of two bases between successive triplets. Therefore, any triplet can be followed by (or preceded by) only four unique triplets, i.e., the two shared bases plus any one of the four bases. Thus, for any given amino acid $x$, one unique base triplet must be assigned for every four unique N-terminal (or C-terminal) neighbors observed. Now, the number of unique N- and C-terminal neighbors observed for each of the 20 amino acids can be compiled from the list of known dipeptides. The greater of the numbers of unique N-terminal and C-terminal neighbors for $x$ indicates the minimum number of unique triplets required to encode $x$. Taking the specific case of serine, 17 unique N-terminal neighbors and 13 unique C-terminal neighbors are observed from the list of dipeptides. The greater of these numbers is 17, and corresponds to 4 sets of 4 unique amino acid neighbors plus one lone amino acid (forming the fifth set). The minimum number of unique triplets required to encode serine is therefore at least five. This number can be determined for each of the 20 amino acids in a like manner from the list of dipeptides. The total of this quantity for all 20 amino acids is the minimum number of triplets required to encode all amino acids, assuming an overlapping code.

However, upon actual enumeration, Brenner found that the minimum number of unique triplets required to encode the 20 amino acids came to 70, 6 more than the theoretically possible 64. This summarily ruled out the possibility of an overlapping triplet code.


Ramakrishnan Sitaraman
TERI University
New Delhi, India