48. Abundance—Big Data

All the grains of sand
On all existing beaches
Just the beginning

Over time, data sets have grown from hand-gathered, penciled-in numbers to vast arrays of cross-linked information. From marketing to weather to less transparent government interests, data on individuals, events, and domains of many descriptions are being gathered and mined in unprecedented quantities. Due to issues such as statistical power, even trivially small effect sizes can easily become highly statistically significant. In short, the entry of big data heralds mandates for effect sizes and the exit of *p* values. When *p* values no longer serve their intended purpose, they are no longer of value to the statistical process, although they will likely need to be mentioned for the comfort of those whose pervasive conditioning requires the existence of *p* values for work to be interpretable.

The amount of data that can be used to model any given topic can simply be too large for many researchers to handle with all but the most basic designs. Data acquisition, storage, and retrieval are now outpacing the ability to process the data in a reasonable amount of time. As new and better processors become available, more of the available data will become available for use. The extent to which those data will be put to personally intrusive purposes remains to be seen.

The high school principal knows about big data, but that situation does not exist for him. He can easily process all the data he has for any project that might come his way.

The director of public health again finds herself in the opposite position. This past year, her data contained more than 800 federally collected measures for use in national public health activities. Her statistics courses did not include what to do with more than 800 potential dependent variables—or even with that many independent variables. More and more each year, she finds herself swimming in deeper and deeper water with an ever shrinking life jacket with regard to data handling issues. Even her statistician is telling her that there are others who specialize in this field and would make a recommendation, if asked.