Treasure our rarities

Paul van Helden

At some point in his or her life, every scientist will encounter the Gaussian distribution. It is such a useful statistical tool that it is used regularly, for example, to compare a set of experimental data with controls, or to define the 'normal range' of a wide range of observational data. Generally, the smaller the difference between the measurements of samples compared with controls, the narrower the Gaussian-distribution curve, and the more samples are needed to detect statistically significant differences; that is, to increase the 'power' of the study. Many scientists in biology and medicine rely on the Gaussian distribution when they compare data or individuals, in order to define those who fall within the normal range and those who are outside this norm or within a defined deviation from the mean. In the case of clinical medicine, individuals who fall outside the norm might have a disease or require attention to prevent disease development. Such a deviation from the norm might be temporary (acute) or permanent (chronic).

It is these outliers that can provide us with unique insights into the biology and physiological mechanisms of disease. They often represent a genetic defect or polymorphisms that cause deviation from the norm, such as those seen in metabolic diseases including phenylketonuria, the performance of outstanding athletes, or people who experience adverse effects from drugs, or do not respond to drugs at all. One example of a truly surprising finding is that individuals with mutations or polymorphisms in the interferon- γ pathway are highly susceptible to mycobacterial infection, which demonstrates the importance of interferon- γ in defending against mycobacteria. Each example of an individual, family or species with deficiencies or abilities that differ from the norm can be informative for research. In general, the further an individual sample deviates from the norm, the smaller the sample size that is needed to show statistical significance; a single individual with a natural knockout gene in a crucial

pathway might be sufficient to identify and prove the crucial function of the gene.

Where are these rarities, and do we pay them enough attention? The natural world offers many examples of 'oddities' that range from individual humans with rare conditions or abilities to organisms with peculiar abilities or habits. The resurrection plant, *Myrothamnus flabellifolius*, which is found in the Namib desert, can probably teach us much about survival under water stress—potentially vital information for many countries with insufficient water supplies for agriculture. Similarly, understanding the physiology of extremophiles, which flourish at high salinity or temperature extremes, can provide us with many potential benefits, such as Taq polymerases and PCR.

The finding that most genomes are extremely complex and have only a small proportion of coding regions has led to the concept of the 'minimal genome' for comparative studies. Examples include the use of *Mycobacterium leprae* to study mycobacterioses generally, or the use of the puffer fish to study vertebral genomes.

Luckily, taxonomists and field-workers such as zoologists or clinicians carefully document rare cases and store materials, and basic researchers study things such as the puffer-fish genome, even though the relevance of doing so is not obvious. These researchers face a tough game of survival, since their research is not necessarily a top priority for funding agencies. More importantly, though, journal editors need to appreciate the importance of documenting deviations from the norm. Rejecting clinical case-reports or the basic description of a new organism with the argument that such a finding is 'not novel or of interest to our readers' is counter-productive. It is a waste of time and effort to ask patients with a rare condition to volunteer for research and to be documented if we do not complete the circle and publish the information; not least because the insight gained could benefit many individuals.

Over time, it is possible to imagine the way in which seemingly disparate information could be combined to create and improve complex models, so that the crucial information from many individuals could enable the prediction of the outcome of various perturbations including drug treatment or chronic health conditions.

It is also for this reason that we must preserve biodiversity. The extinction of species has a direct impact on our own survival in ways that we cannot imagine and will never know. The loss of biodiversity directly affects resilience, and increases the vulnerability of ecosystems and the species within them, including ourselves.

What role, then, should researchers, reviewers and publishers have in treasuring our rarities? I might not be the only one to see the trend, dictated by the ubiquitous impact factor, that journals increasingly reject papers that are 'mere' descriptions of new organisms or case reports, unless they contain a mechanistic insight and comprehensive functional explanation. We might have all received stock rejection letters stating that 'the journal no longer accepts genome papers' or 'the journal does not accept case reports'. Instead, reviewers and editors must remind themselves that an individual case or observation might be the first step towards an important discovery; its publication will capture the attention of the wider community to further investigate these interesting phenomena. Unless we continue to document interesting cases, meta-analysis and evidence-based science or medicine will not be possible. Evidence, however insignificant, is the basis for everything we do in science, and we need to make sure that all of it is both captured and treasured.

Paul van Helden is a professor and head of the Department of Biomedical Science at Stellenbosch University, Cape Town, South Africa.

E-mail: pvh@sun.ac.za

EMBO reports (2011) 12, 611. doi:10.1038/embor.2011.107