Commentary

Endocrine physiology and the value of case studies

B. G. Charlton

Case studies have a bad reputation. Group studies are seen as preferable, and the bigger the better. In medical circles, individual case studies are widely regarded as the least satisfactory kind of clinical research; inferior to the case series, case-control study, cohort study and (as a gold standard) the prospective study of a randomly selected population. Such group techniques, appropriately modified, are also used as the method of choice for investigation of endocrine disease (in man and other animals). Nevertheless, in the study of mechanisms (i.e. endocrine physiology) there is an essential and complementary role for intensive study of individual subjects.

There are problems with group studies, and especially with the statistical manipulations of their data. The point was made many years ago by the founding father of modern physiology, Claude Bernard, who warned that using group averages 'leads, so to speak, necessarily to error ... The greatest obstacle to applying calculation to physiological phenomena is still, at bottom, the excessive complexity which prevents their being definite and comparable one with another ... Averages must therefore be rejected, because they confuse while aiming to unify, and distort while aiming to simplify. Averages are applicable only to reducing very slightly varying numerical data about clearly defined and absolutely simple cases' (Bernard, 1865).

Bernard's argument remains valid today, and is particularly applicable to endocrinology. The essence of endocrine physiology is variability—which is an unsurprising characteristic for a control mechanism. For example, hormone release may be prone to alteration by circadian rhythms, external stimuli (pain, food, sleep etc.), interactions with other hormones and neurotransmitters; and the mode of release is characteristically pulsatile. Blood concentrations may be further modified by changes in rate of elimination, protein binding and a host of other factors.

The net effect of such a multiplicity of influences is that any longitudinal pattern of hormonal change over a given period of time in an individual subject is likely to be unique and unrepeatable ('excessive complexity'): hence the temptation to smooth things out by the use of statistics. The impossibility of controlling for all interfering factors is one common justification for averaging groups results—although it carries the direct penalty that because the individuals making-up the group are never absolutely identical (especially if they are human) therefore much of the most interesting information is lost: buried in the statistical 'noise'.

However, as Bernard indicates, averaging is not always legitimate. Imagine a profile of multiple measurements of a hormone's blood concentration over 24 h. I contend that these data should be presented primarily as the detailed results of a representative individual ('present our most perfect experiment as a type, which, however, still stands for true facts', to quote Bernard). If there is no such thing as a 'representative' individual then it is not legitimate to use averages to invent one. We must not 'distort while aiming to simplify'. Either the average of the group is essentially the same as the individual, in which case statistics simply serve to confirm the results of the individual; or else the average is different from any individual, in which case it does not have any validity for physiological purposes. In considering such data, group statistics suffer the same absurdity as the average married couple with 2-4 children: no such situation exists!

This is important because the fine individual detail, lost in the statistics, is itself the 'biological character of phenomena': if the hormone profile is not highly variable, then it's not endocrinology. It is exactly this fine detail which is necessary for understanding mechanisms, for constructing models of function. I am not suggesting anything outrageous here: there are many branches of (respectable) science which routinely present representative individual data as their primary results; for example electrophysiology, electron microscopy or cognitive neuropsychology. Indeed the latter group of scientists might be said to have 'rediscovered' the case study for contemporary clinical science (see Shallice (1989) for a sophisticated discussion of the case versus group controversy).
I am proposing that in testing endocrine hypotheses, whether of normal function or pathophysiology, it is vital to ensure that any theory is confirmed by the detailed longitudinal study of individuals. Unless the ‘best’ experiments with individuals are able to confirm the predictions of a proposed mechanism, then the hypothesis could be an artefact of averages. This does not mean that we are lost in a mass of unanalysable data, instead that we are testing a model of function (derived from consideration of the best individual experiments—which may include pathological conditions); and testing that model against a number of individual cases. Large group size does not compensate for a poorly designed or executed experiment. Numbers merely serve to confirm hypotheses generated on individuals.

The mania for large groups in all situations is related to a misunderstanding of the function of statistics. Some biologists would say (privately) that if you need statistics to prove something, then it’s not significant!: meaning that unless an effect is big enough to hit you between the eyes, then it is not important when compared with the uncertainties of biological variability. This is an exaggeration, and only applies to simple situations, but carries a certain validity.

Of course, available technology may not be adequate for the longitudinal study of individuals—especially human individuals. Repeated sampling of peripheral blood can be tricky (the reliability of intravenous cannulae still leaves a lot to be desired) and the cumulative quantity of blood required for assay may be excessive. In man this may be due to the ethical dimension: it is one matter to take a few blood samples over a few days, but quite another to keep pestering and puncturing more often and over much longer periods. Some vessels may be all-but inaccessible, requiring surgery, cumbersome apparatus, or even ‘sacrifice’ of an animal: which certainly puts a stop to any longitudinal study.

So, in endocrinology the longitudinal study of individuals is not always possible; although it is always scientifically desirable. This is not intended as a green-light for an explosion in submissions of papers of the genre characterized as ‘an interesting case of . . .’. On the other hand, really detailed scrutiny of unusual (or unique) pathological conditions can be extraordinarily valuable in generating or testing a clearly defined functional model which can then be tested more widely. Examples from medical history include Alexis St Martin with his gunshot-induced gastric fistula and Phineas Gage whose prefrontal cerebral cortex was amputated by an unfortunate work accident.

Be that as it may, I suggest that any paper addressing endocrine physiology—especially involving hormone interactions or correlations with other factors—should display detailed longitudinally sampled ‘raw’ data from single subjects, as well as any necessary statistical digest of the groups involved. By such means we might hope to prevent Claude Bernard spinning in his grave.

REFERENCES


Department of Anatomy,
University of Glasgow,
Glasgow G12 8QQ.