Uncertainty and quality in science for policy

S.O. Funtowicz and J.R. Ravetz

We all know how statistics can be, and are, manipulated for ulterior ends, so that numbers can be made to convey an impression of what we know is not the case. How can this abuse be counteracted? The time-honoured way is to expose each individual distortion, and also to write reviews with a number of examples that illustrate an unannounced policy of the degradation of statistics for political ends. Necessary as this approach is, it suffers from two deficiencies. One is the implicit assumption that there is a true and objective measure for each aspect of social reality, which all honest and competent statisticians should recognize. The other is that each case of abuse must be exposed on its own demerits; there is no standard procedure whereby the quality of statistics can be characterized.

We have devised a notational system for quantitative information by which these difficulties can, to some extent at least, be overcome. It is based in large part on the experience of workers in the matured natural sciences. Contrary to the impression conveyed by popularisers and philosophers, the success of these sciences is not so much due to their being 'exact' or certain, but rather to their effective control of the inherent uncertainties (of which inexactness is one sort) in their data and theories. Those social sciences which most exhibit 'physics-envy', such as economics, are just those which are most crippled by an anachronistic and fantasised vision of the methods and results of physical science.

Of course, the myth of unlimited accuracy in quantitative sciences has many uses; only with the debate over 'Star Wars' did the public learn that computer programming is a highly uncertain and inexact art. With the proliferation of 'models' of all sorts in natural and social science, we must recognize a new sort of pseudo-science, which we may call GIGO (Garbage In, Garbage Out). This can be defined as one where the uncertainties in the inputs must be suppressed, lest the outputs become indeterminate. Hence the crisis in statistics is only one manifestation of a general problem of quantitative information, where incompetence and malevolence interact to produce meaningless numbers, deceiving the experts themselves along with the public. The fate of the State-socialist economies is a reminder of how even the largest and most comprehensive bureaucratic systems can drift out of touch with reality. While defective social statistics were not a simple, unique cause of this collapse, they were certainly an essential part of a syndrome that inhibited any attempts at reform.

Our contribution is at the technical level, a notational system whereby the different sorts of uncertainty in quantitative information are displayed in a standardized and perspicuous way. By itself, it enables providers and users of such information to be clear about its uncertainties. Since the management of uncertainty

is at the core of quality of quantitative information, the system 'NUSAP' also fosters an enhanced appreciation of the issue of quality in information, and thereby enables a more effective criticism of quantitative information by users of all sorts, expert and lay. We also believe that with the better diffusion of self-conscious skills in the production of quantitative information, critical analysis will not be drawn into the philosophical trap of commitment to some essentially true Platonic statistic for each aspect of the world.

The system is based on five categories, which generally reflect the standard practice of the matured experimental sciences. By providing a separate box, or 'field', for each aspect of the information, it enables a great flexibility in their expression. By its means, nuances of meaning can be conveyed concisely and clearly, to a degree that is quite impossible otherwise. The name 'NUSAP' is an acronym for the categories. The first is Numeral; this will usually be an ordinary number; but the choice of number (say, between '2' and '2.13') will be assisted by the entries in the successive boxes. Second comes Unit, which may be of the conventional sort, but which may also contain extra information, as the date at which the unit is evaluated (most commonly with money). The middle category is Spread, which generalizes from the 'random error' of experiments or the 'variance' of statistics. Although Spread is usually conveyed by a number (either =/-, % or 'factor of', it is not an ordinary quantity, for its own inexactness is not of the same sort as that of measurements.

This brings us to the more qualitative side of the NUSAP expression. The next category is Assessment; this provides a place for a concise expression of the salient qualitative judgements about the information. In the case of statistical tests, this might be the significance-level; in the case of numerical estimates for policy purposes, it might be the qualifier 'optimistic' or 'pessimistic'. In some experimental fields, information is given with two +/- terms, of which the first is the spread, or random error, and the second is the 'systematic error' which must estimated on the basis of the history of the measurement, and which corresponds to our Assessment.

Finally there is P for Pedigree. It might be surprising to imagine numbers as having pedigrees, as if they were showdogs or racehorses. But where quality is crucial, a pedigree is essential. In our case, the pedigree does not show ancestry, but is an evaluative description of the mode of production (and where relevant, of anticipated use) of the information. Each special sort of information has its own pedigree; and we have found that research workers can quickly learn to formulate the distinctions around which a special pedigree is constructed, and also thereby learn much about the characteristic uncertainties of their own field. Although we have not yet had the opportunity to test this with non-expert groups, we are equally sure with some preliminary training, any concerned citizen could learn how to elicit the pedigree of information being provided by an expert.

The pedigree is expressed by means of a matrix; the columns represent the various phases of production or use of the information, and within each column there are modes, normatively ranked descriptions. These can be numerically graded, so that with a coarse arithmetic, a 'quality index' can be calculated for use

in Assessment if desired. For general statistical information, the pedigree is laid out as in the Table, where the top row has grade 4 and the bottom two, 0. For a numerical evaluation, average scores of 4 downwards are rated as High, Good, Medium, Low and Poor.

Definitions & standards	Data-collection & analysis	Institutional culture	Review
Negotiation	Task-force	Dialogue	External
Science	Direct survey	Accommodation	Independent
Convenience	Indirect estimate	Obedience	Regular
Symbolism	Educated guess	Evasion	Occasional
Inertia	Fiat	No-contact	None
Unknown	Unknown	Unknown	Unknown

Pedigree Matrix for Statistical Information

The first column describes how the job is defined, for any competent statistician knows that 'just collecting numbers' leads to nonsense. The whole pedigree matrix is conditioned by the principle that statistical work is (unlike some traditional lab research) a highly articulated social activity). So in 'Definition and Standards' we put 'negotiation' as superior to 'science', since those on the job will know of special features and problems of which an expert with only a general training might miss. It is important to be able to describe low-quality work; and so 'symbolism' in statistics is something which any comprehensive scheme must allow for. Similarly, a 'Task Force' gets a higher rating than a 'Direct Survey', for the latter (like a census) may produce information that is not tailored to the problem at hand. The other two columns relate to the more directly social aspects of the work. 'Institutional Culture' describes the relations between the various levels in hierarchy or command and control; and we allow for the phenomena variously described by 'Clochemerle' or 'Schweik'. Since quality-assurance is an essential part of any productive process, including that for information, we have a column for 'Review'. This needs no explanation.

Thus the pedigree matrix, with its multiplicity of categories, enables a considerable variety of evaluative descriptions to be simply scored and coded. In our book Uncertainty and Quality in Science for Policy (Kluwer Academic, 1990), we

illustrate the NUSAP system with a somewhat imaginary example of the history of the statistics on hand-pumps for drinking water in a Third-World country. The earlier efforts had a distinctly low Pedigree profile; inertia, symbolism and fiat were prominent, along with the absence of effective review; but by the end, with the lessons of experience, improvements could be recorded. The use of NUSAP could also highlight crucial features of the process. For example, the problems of definition can be explored; is a 'pump' one that is listed in an old census, one that is ordered from abroad, one that is registered as delivered, one that is installed, or one that is in full and satisfactory use? At the other end of the process, NUSAP alerts us to the meaningfulness or otherwise of numerical expressions. If we see a number like 15,432 coming out of an unsophisticated statistical exercise, we should be able to recognize hyper-precision. On the other hand, a correctly framed estimate like 11,300:pumps: +/- 5%: has certain uses, but an equivalent form, >10 1/2 :K pumps: may be more appropriate for policy purposes. Thus with NUSAP we should be able to provide numerical statements with considerable nuance of expression.

We know enough about the use of numbers to be aware that no single system will prevent abuse and corruption of statistics. But we believe that with the improvement of competence all around, and especially arming non-expert users with an instrument of analysis, NUSAP will at least make it possible for the debate to be conducted at a higher level. Beyond that, we can imagine that with a conception of mathematics that is freed from Platonic perfectionism, there could eventually arise a public that is genuinely numerate as well as literate. The implications of that for the conduct of public policy do not need elaboration at this point.

S.O. Funtowicz and J.R. Ravetz are Directors of the Research Methods Consultancy, 13 Temple Gardens, London NW11, OLP; Tel. 081 455 2770; Fax/answerphone 081 458 7945; e-mail j.ravetz@vaxa.cc.ic.uk.ac. They specialise in quality assurance of scientific information, and the social practice of science on problems of risk and the environment.