QUALITY AND RELIABILITY ENGINEERING INTERNATIONAL
Qual. Reliab. Engng. Int. 2003; **19**:265–272 (DOI: 10.1002/qre.585)

Special Issue

Statistical Efficiency: The Practical Perspective[‡]

Ron S. Kenett[1], Shirley Coleman[2,*,†] and Dave Stewardson[2]
[1] *KPA Ltd, Management Consulting, 25 Hataasia Street, P.O. Box 2525, Raanana 43100, Israel*
[2] *ISRU, Stephenson Building, Newcastle University, Newcastle NE1 7RU, U.K.*

The idea of adding a practical perspective to the mathematical definition of statistical efficiency is based on a suggestion by Churchill Eisenhart who, years ago gave, in an informal 'Beer and Statistics' seminar, a new definition of statistical efficiency. Later Bruce Hoadley from Bell Laboratories picked up where Eisenhart left off and added his version nicknamed 'Vador'. Blan Godfrey, former CEO of the Juran Institute, more or less used Hoadley's idea during his Youden Address at the Fall Technical Conference of the American Society for Quality Control. We expand on this idea adding an additional component, the value of the data actually collected, which we believe is critical to the overall idea. The concept of Practical Statistical Efficiency (PSE) derived from these developments is introduced and demonstrated using five case studies. We suggest that PSE be considered before, during and after undertaking any quality improvement projects. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: practical statistical efficiency; consultancy tools; Six Sigma; quality improvement

1. INTRODUCTION

Statistical efficiency is usually defined in mathematical terms. In the context of hypothesis testing, the relative efficiency of Test A versus Test B is the ratio of sample size for test A to sample size for test B, where sample sizes are determined so that both tests reach a certain power against the same alternative hypothesis[1]. The idea of adding a practical perspective to the mathematical definition of statistical efficiency is based on a suggestion by Churchill Eisenhart who, years ago, gave in an informal 'Beer and Statistics' seminar, a new definition of statistical efficiency. Later Bruce Hoadley from Bell Laboratories picked up where Eisenhart left off and added his version nicknamed 'Vador'[2]. Blan Godfrey, former CEO of the Juran Institute, used Hoadley's idea during his Youden Address at the *Fall Technical Conference of the American Society for Quality Control*[3,4]. We expand on this idea adding an additional component, the value of the data actually collected, which we believe is critical to the overall idea. The concept of Practical Statistical Efficiency (PSE) derived from these developments is introduced and demonstrated using five case studies. It is presented as an important addition to the set of statistical consultancy tools, thus enhancing the ability of practicing statisticians to show the extent of their contribution to the resolution of real life problems. The original motivation of the authors comes from the pervasive observation that the application of statistical methods is in many cases an exercise in using statistical tools rather than a focused contribution to specific problems. Section 2 defines PSE, Section 3 describes five case studies that are evaluated in Section 4 with PSE. We conclude with some remarks and recommendations.

*Correspondence to: S. Y. Coleman, ISRU, Stephenson Building, Newcastle University, Newcastle NE1 7RU, U.K.
†E-mail: shirley.coleman@ncl.ac.uk
‡This paper is based on a presentation given at the second ENBIS Conference, Rimini, September 2002.

Received 17 January 2003
Revised 28 March 2003

2. PRACTICAL STATISTICAL EFFICIENCY

We define PSE as:

$$\mathrm{PSE} = E\{\mathrm{R}\} \times T\{\mathrm{I}\} \times P\{\mathrm{I}\} \times P\{\mathrm{S}\} \times V\{\mathrm{PS}\} \times V\{\mathrm{P}\} \times V\{\mathrm{M}\} \times V\{\mathrm{D}\}$$

where $V\{\mathrm{D}\}$ is the value of the data actually collected; $V\{\mathrm{M}\}$ is the value of the statistical method employed; $V\{\mathrm{P}\}$ is the value of the problem to be solved; $V\{\mathrm{PS}\}$ is the value of the problem actually solved; $P\{\mathrm{S}\}$ is the probability that the problem actually gets solved; $P\{\mathrm{I}\}$ is the probability the solution is actually implemented; $T\{\mathrm{I}\}$ is the time the solution stays implemented; and $E\{\mathrm{R}\}$ is the expected number of replications.

The eight terms in PSE are defined in detail below. Clearly PSE is a conceptual assessment of efficiency rather than the original idea of mathematical statistical efficiency which is encompassed in the definition of $V\{\mathrm{M}\}$ below.

A straightforward approach to evaluating PSE is to use a scale of '1' for not very good to '5' for excellent. This method of scoring can be applied uniformly for all PSE components. Some of the PSE components can be also assessed quantitatively. $P\{\mathrm{S}\}$ and $P\{\mathrm{I}\}$ are probabilities, $T\{\mathrm{I}\}$ can be measured in months, $V\{\mathrm{P}\}$ and $V\{\mathrm{PS}\}$ can be evaluated in euros or dollars. If such values are available we suggest defining five categories for these values and, thereby, achieve the classification outlined above. If no quantification is possible we can rely on an intuitive qualitative evaluation, for example of experts using the Delphi method, in order to estimate the PSE components. In all cases we obtain an evaluation of each component on a 1–5 scale. This reflects the fact that PSE is a conceptual measure rather than a numerical one.

2.1. $V\{\mathrm{D}\}$ = *value of the data actually collected*

The application of statistical methods is necessarily data driven. In many cases using the data which are readily accessible is like taking observations below the lamppost where there is light—not necessarily where you lost your key or where the answer to your problem lies. Collecting the right data is a critical step in any statistical investigation. Data collection is a translation process where issues raised at the management or technical level are formulated at a quantitative level. Random variables are a misnomer since they correspond to functions from the event space to the numeric dimension—they are not 'variables', merely 'functions' which need to be well specified. This can be done in a proper or improper manner. Mistranslating a given problem into the wrong data definition will lead to irrelevant statistical answers. For example, non-responses to a sample survey can significantly bias the results if the units missing from the survey have a unique characteristic that will therefore be completely ignored in the analysis. Non-responses could all belong to a certain geographical area where something different occurs or have cultural issues[5]. Without proper follow up the survey will not be able to uncover these phenomena. The value of the data actually collected reflects how well the data used in the statistical analysis reflect the original problem raised by the content expert. A $V\{\mathrm{D}\} = 5$ corresponds to data being most relevant to the problem, clearly understood by relevant stakeholders and collected comprehensively without bias.

2.2. $V\{\mathrm{M}\}$ = *value of the statistical method employed*

This concept is closest to the original idea of mathematical statistical efficiency and includes the idea that the method should be as efficient as possible[1]. Take, for example, the case of a client who launched an effort to reduce billing errors and has asked for a method that estimates the number of errors accurately and efficiently. Let's say that we are considering two alternative methods: A and B. Method A is said to be more efficient than method B if method A requires a smaller sample to provide an estimate of equal accuracy and precision than method B. A $V\{\mathrm{M}\} = 5$ corresponds to the method with proven mathematical optimality properties.

2.3. $V\{\mathrm{P}\}$ = *value of the problem to be solved*

Statisticians too often forget this part of the equation. We frequently choose problems to be solved on the basis of their statistical interest rather than the value of solving them. In the billing example mentioned in $V\{\mathrm{M}\}$ above, a calculation showed the billing errors are worth somewhat over \$700 000. This information is crucial to management. The perceived impact of a problem being addressed by statistical methods has direct bearing on the ability of the organization to deal with it seriously in order to achieve actual results—not only 'recommendations'. Working on a problem with high $V\{\mathrm{P}\}$ is a necessary, but not sufficient, condition to achieve significant results for the client. A value of $V\{\mathrm{P}\} = 5$ corresponds to a problem with strategic impact on the organization, i.e. with long-term implications.

2.4. $V\{\mathrm{PS}\}$ = *value of the problem actually solved*

Usually no one method or attempt actually solves the entire problem, only part of it. So this part of the equation could be expressed as a fraction of $V\{\mathrm{P}\}$. In the case of the billing example, the client expected to reduce the billings errors from 24 000 to 3000 per invoice. In this case the fraction of the problem solved is large at 87.5%. Problems with high $V\{\mathrm{P}\}$ that are fully solved get a high $V\{\mathrm{PS}\}$.

2.5. $P\{\mathrm{S}\}$ = *probability that the problem actually gets solved*

This is both a statistical question and a management question. Did the method work and lead to a solution that worked and were the data, information and resources available to solve the problem? Part of this PSE component is related to the 'buy in' of management and technical personnel in meeting the challenge of facing the problem tackled. This is achieved by getting this group of stakeholders to play an active role in specifying the problem and interpreting the results. A high value of $P\{\mathrm{S}\}$ implies that proper planning and effective execution has been carried out.

2.6. $P\{\mathrm{I}\}$ = *probability the solution is actually implemented*

Here is the non-statistical part of the equation that is often the most difficult to evaluate. Implementing the solution may be far harder than just coming up with the solution. Here we deal with resistance to change and all the other reasons why implementation is by far the hardest part of statistical consulting. Part of this issue is related to properly matching the management approach with the statistical tools used in the analysis. A proposed approach to achieving this match has been suggested in the first chapter of Kenett and Zocks[6] in the form of a Quality Ladder. Typical approaches to the management of industrial organizations can be summarized and classified as (1) fire fighting, (2) inspection, (3) process control and (4) quality by design. Each management approach requires a matching set of statistical methods. The Quality Ladder provides such a match. Managers involved in reactive fire fighting need to be exposed to basic statistical thinking. Managers attempting to contain quality and inefficiency problems through inspection and 100% control can have their tasks alleviated by implementing sampling techniques. More proactive managers investing in process control and process improvement are well aware of the advantages of control chart and process control procedures. At the top of the quality ladder is the quality by design approach where upfront investments are secured in order to run experiments designed to impact product and process specifications.

Efficient implementation of statistical methods requires a proper match between management approach and statistical tools. For example, if a fire fighting manager tries to implement control charts he might have the workforce maintain charts covering the walls but the early warning signals provided by the charts will go unnoticed, grossly reducing the effectiveness of control charts or any other process control procedures. In order to take full advantage of the benefits of proactive process control and improvement, fire fighting management habits have to change. The implication of this concept is that $P\{\mathrm{I}\}$ is determined by the management style and the match achieved between management style and statistical approach. A high value of $P\{\mathrm{I}\}$ implies that a proper match between management approach and statistical methods has been established.

2.7. $T\{\mathrm{I}\}$ = *time the solution stays implemented*

Problems have the tendency not to stay solved. This is why we need to put much emphasis on holding the gains in any process improvement. Take the billing example. Suppose that in the first year the client saves \$700 000 minus the cost of solving the problem and implementing the solution. Even if for some reason new types of errors are introduced into the billing process, if there is tight control so that the original problem stays solved for three years, the client saves almost \$2 100 000. An extensive application of the solution, indicating sustainable implementation, results in $T\{\mathrm{I}\} = 5$.

2.8. $E\{\mathrm{R}\}$ = *expected number of replications*

This is the part most often missed in companies. If the basic idea of the solution could be replicated in other areas of the company, the savings could be enormous. A value of $E\{\mathrm{R}\} = 5$ represents a very large number of potential replications of the solution.

3. *THE CASE STUDIES*

In order to demonstrate the concept of PSE we will review five case studies, four of which were first presented in Coleman and Stewardson[7] the other in Linsley *et al*.[8].

3.1. *Tea packing*

An old family owned company decided, in 1998, to modernize their management of the company. Key performance indicators (KPIs) including efficiency, staff absence and waste were adopted and the concept proved so useful that the original 39 KPIs were increased to 50. There were several early wins, for example a 90% reduction in outloading errors to a major customer and an improvement in performance of two parallel shifts from 54% and 68% to 80% each. A questionnaire was given to staff to find out how they felt about the KPIs. It was found that staff did not like the staff absence KPIs because they were so negative and they felt powerless to influence them but did relate to waste KPIs because they are traditionally given in real terms rather than percentage. So staff absence was changed to staff attendance and KPIs focussed on output using both quantities and value made and lost. Reporting changed from displaying the daily performance of 10 machines on a page to displaying weekly performance of one machine per page. This meant that charts could be placed near to machines and operators could observe their performance. It also led to a clearer understanding of causes of downtime. It was then decided to take the plunge and change from reporting efficiency to reporting utilization as a much more realistic and meaningful way to assess performance. Performance appeared to drop from 80 to 70% but is was felt that staff were ready for the psychological blow and now performance on this measure is gradually rising.

The tea packers embraced the quality improvement project wholeheartedly and initially had a specific problem to solve. This was that they had two apparently identical packing lines and yet one produced 54% efficiency and the other produced 68% with the same staff. The problem was eventually located to the availability of the engineers to mend the line when problems occurred.

3.2. *Soft drinks manufacturer*

A major quality improvement drive in a soft drinks manufacturer was lead from the top and saw the introduction of quality improvement groups, KPIs in production, distribution and resources, quantitative evaluation of new equipment and downtime reduction. The main charts showing percentage of efficiency, however, highlighted the problem of trying to encapsulate the whole business performance in one chart. Changes in percentage of efficiency were found to be seasonal and highly dependent on product changes and product mix.

It was decided to investigate producing an interactive, visual display of the knock on effects of management decisions. This required the construction of process maps using an appropriate team of people Process mapping simultaneously makes people talk to each other, records the current process flow, challenges the way things are done and indicates where measurements need to be made, all in terms familiar to the company.

The soft drinks company were enthusiastic but had a very vague problem area, defined as general improvement, and the company were more or less taking part in a quality improvement programme to please the programme organizers rather than definitely expecting some benefit. This is reflected in their PSE score.

3.3. Heavy metal company

In a heavy metal company the level of effluent was a major area of concern. Initial data gathering showed large fluctuations, charts were informative but the data were too unstable for SPC techniques to be applied immediately. The current level and variability were calculated to provide a comparison before and after investigation. Extreme values, both large and small, were analysed to help interpret the data but yielded very little insight into the underlying causes of variation. The next step was a process map drawn up by a team including charge hands, laboratory staff and management. It was apparent that routine maintenance could be very useful and a plan was drawn up. The lack of guidelines gave clues to other causes of variation and new drainage was introduced. Lack of communication between all players was also important and a new post was created to coordinate things.

The effluent levels improved. Control charts were put in place. However, process improvement is a complex affair and attention then went to looking at the use of services including electricity and water. Higher management decided to halve the use of water. The effluent levels began to rise. Communication between management and operators was non-existent which reinforced the feeling of non-empowerment. The relative costs of water, prosecution, loss of customers and environmental considerations may have been evaluated but were not shared.

The measurement and charts provided all the necessary information, process mapping and brainstorming were effective and yet still the good ideas were not adopted. What was missing was the inclination within the company to share information and ideas. Probably what was also missing was confidence, courage, belief in the power of quantitative analysis and leadership from the consultant

The heavy metal company were very keen but only to reduce short-term costs and did not give management support to the technicians involved in the project. Again there was no specific problem rather there was an overall interest to see where statistics could help. This lack of focus is reflected in the lower PSE.

3.4. Major utility

Quality is very much a part of the gas transportation system. The management of a major utility company decided to see what benefits could be gained from SPC. Intern students managed to bring together large numbers of disparate experts to discuss fundamentals, such as what makes an operator feel that the day has gone well. From these discussions suitable measures were determined and SPC was implemented. Reports have been written about the management issues, the use of steering groups, importance of feedback and review[9], and the statistical issues encountered[10].

There were some early wins where control charts highlighted problem areas, which although known to the people concerned, were not widely recognized. SPC has since been developed to provide a safety measurement framework. About 30 metrics are analysed periodically and presented in charts in a quarterly report. The report provides an informative and powerful analysis of the current safety situation. Feedback from readers helps to ensure that the report maintains vitality and relevance[11].

The utility company had a specific project and was focused on solving the problem, retaining the gains and cascading the methodology throughout the company. This is reflected in the high PSE.

Table I. Scores for components of PSE (1 is poor, 5 is excellent)

Case	$E\{R\}$	$T\{I\}$	$P\{I\}$	$P\{S\}$	$V\{PS\}$	$V\{P\}$	$V\{M\}$	$V\{D\}$	PSE
Tea	4	3	5	5	4	4	3	4	57 600
Drink	4	4	3	3	4	3	2	1	3456
Metal	4	3	3	4	5	1	1	1	720
Utility	5	5	4	3	4	5	4	4	96 000
Filter	4	5	5	5	4	5	5	5	250 000

3.5. *Filter manufacture*

This example represents a particularly good result for the company involved and is discussed more fully in Linsley *et al*.[8]. Variation in quality from the main production line was causing concern and the consequent need to reduce rejection rates helped to focus this company's management on solving the problem quickly. An earlier project had already demonstrated the value of a systematic team-based approach, combined with the use of designed experiments, and there was little need for the consultant to 'sell' these ideas.

Following sessions with the teams of engineers and process staff, during which the involvement of people from all areas of the company was obtained, process maps and basic quality tools were used in an initial analysis of the problem. The measurement system was investigated first, a strategy similar to that adopted in Six Sigma programmes, and then a designed experiment conducted over a number of important factors in the early stages of the process quickly established the main causes of faults. A second round of trials in the later process stages established that one complete stage, rather than adding value, was actually having no beneficial effect at all but was damaging the product. This stage was subsequently dropped from the process entirely, saving much time and resources, in addition to improving product quality.

The company ended up with a better controlled, faster process, better product quality and tighter process capability, with lower costs all round. The involvement of staff throughout the process in designing the study was a major factor behind the quick success of the project. Based on this and other successes they have spun out many of the statistical ideas used and now employ a full time engineer to coordinate the implementation of all-round process improvements. They have reported savings from this one project of around £500 000 per annum, a major result for a company with fewer than 100 employees.

The filter manufacturer had already seen the value of designed experiments in an earlier project and had established the need to improve the process, hence the commitment to this project. A key success factor here was the involvement of process staff in the main discussions and decisions about the investigation. The major benefits achieved and the spin-out via the subsequent programme is reflected in the very high PSE.

4. *PSE EVALUATION OF THE FIVE CASE STUDIES*

Referring to the five case studies, the scores we derived subjectively for the eight components of PSE are as given in Table I. A radar plot of these scores is presented as Figure 1. It can be seen from the table and figure that the scores are generally high for the longevity ($E\{R\}$ and $T\{I\}$) and probability components ($P\{I\}$ and $P\{S\}$) for all five case studies and the difference comes in the values. In particular $V\{D\}$, the value of the data. This value reflects the effort put into the project. A similar comment applies to $V\{M\}$, the value of the method used. When the problem is of low priority, resources are not made available for methodology to be as good as possible. For example, in the heavy metal company, the time available to the operative involved and his expertise limited the methodology to the most basic charts.

In all of these case studies $V\{PS\}$, the value of the problem actually solved, is high. In the less successful case studies $V\{P\}$, the value of the problem to be solved, was not appreciated and the resources required to deal with the side effects were not available.

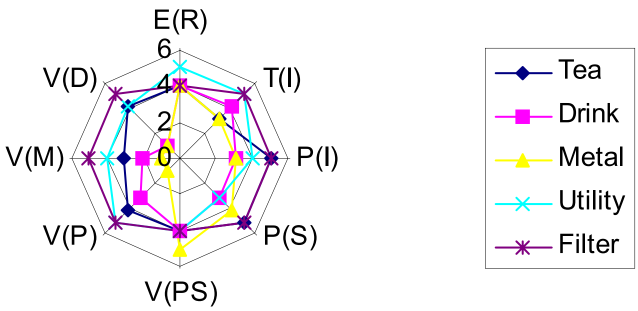

Figure 1. PSE evaluation of the five case studies

5. CONCLUSION

Considering PSE via its various components provides a check on the consultancy service offered and can usefully be done before and after the intervention. It is easier to look at PSE for specific projects rather than for the overall more general improvement programmes.

It should be noted that not all statistical consultancy is aimed at solving major problems, often it is part of a general pro-active quest for quality improvement. The PSE scores above highlight the importance of the Six Sigma approach, which is very problem orientated. As soon as a general statistical improvement project is started a number of possible areas for statistical investigation are identified. Rather than work on them in an unstructured way, it is more practically efficient from a statistical viewpoint to work on separate agreed problem areas.

In all companies, but especially in SMEs, the company situation can change very fast in reaction to orders or to management restructures. For example in the heavy metal company there was a desperate reactive effort to produce a better quality product for a particular customer, but as soon as that order was completed resources were shifted to the next problem. All the discoveries and improvements made to the process were lost. There is a move particularly amongst SMEs to cast off old information very quickly which results in a total lack of knowledge about anything in the past.

PSE provides an efficient approach to compare various statistical projects and, by its composition, is forcing practitioners to take a broad view of their work. Readers of this article may like to experiment with finding the PSE of all their next projects to see if it helps clarify them at the outset. The authors would welcome any feedback or comments as to the benefits of this approach.

Acknowledgements

This paper was supported by funding from the 'Growth' programme of the European Community and was prepared in collaboration by member organizations of the Thematic Network, Pro-ENBIS-EC contract number G6RT-CT-2001-05059.

REFERENCES

1. Bickel P, Doksum K. *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day: San Francisco, CA, 1977.
2. Hoadley B. A zero defect paradigm. *ASQC Quality Congress Transaction*, Anaheim, 1986.
3. Blanton GA. Statistics, quality and the bottom line, part 1. *ASQC Statistics Division Newsletter* 1988; **9**(2):211–213.
4. Blanton GA. Statistics, quality and the bottom line, part 2. *ASQC Statistics Division Newsletter* 1989; **10**(1):14–17.
5. Kenett RS, Albert D. The international quality manager: Translating quality concepts into different cultures requires knowing what to do, why it should be done and how to do it. *Quality Progress* 2001; **34**(7):45–48.

6. Kenett RS, Zacks S. *Modern Industrial Statistics: Design and Control of Quality and Reliability*. Duxbury Press: San Francisco, CA, 1988.
7. Coleman SY, Stewardson DJ. Use of measurement and charts to inform management decisions. *Managerial Auditing Journal* 2002; **17**(1/2):16–19.
8. Linsley MJ, Pearce KF. Stewardson DJ, Coleman SY. The introduction of quality tools to reduce product reject rates in an SME. *Proceedings of the 5th SMESME International Conference*, East Anglia, U.K., 2002; 302–307.
9. Chambers PGR, Piggott JL, Coleman SY. SPC—a team effort for process improvement across four area control centers. *Journal of Applied Statistics* 2001; **28**(3):307–324.
10. Coleman SY, Gordon A, Chambers PR. SPC—making it work for the gas transportation industry. *Journal of Applied Statistics* 2001; **28**(3):343–351.
11. Coleman SY, Arunakumar G, Foldvary F, Feltham R. SPC as a tool for creating a successful business measurement framework. *Journal of Applied Statistics* 2001; **28**(3):325–334.

Authors' biographies

Ron Kenett, PhD, FRSS is CEO and Senior Partner of KPA Ltd, an international management consulting firm and Visiting Professor at the University of Torino, Italy. He was Professor of Management at the State University of New York (1987–1992), where he was awarded the General Electric Quality Management Fellowship, Director of Statistical Methods for Tadiran Telecommunications Corporation (1981–1990) and Member of the Technical Staff at Bell Laboratories in New Jersey (1979–1981). He obtained his PhD in Mathematics (1978) from the Weizmann Institute of Science, Rehovot, Israel, BSc in Mathematics (1974) from Imperial College of Science and Technology, with first class honors, London University, London, U.K. He has over 25 years of experience in restructuring and improving the competitive position of organizations by integrating statistical methods, process analysis, supporting technologies and modern human resource management systems. He is the author of 90 papers and four books.

Shirley Coleman, PhD, C.Stat, FRSS is principal statistician at the Industrial Statistics Research Unit, Newcastle University and Associate Professor of the Technical University of Wroclaw, Poland. Previously she was a Statistics Lecturer at Hong Kong University, Visiting Lecturer at Hong Kong Polytechnic and Senior Quantitative Analyst at Barings Securities Hong Kong (1986–1990). She obtained her PhD in Morphometry, Computer Graphics and Statistics (1986), her MSc in Statistics (1981) and her BSc in Mathematical Sciences (1975) with first class honours from Newcastle University. She was senior engineer at On Line Inspection Centre of British Gas (1979–1980). She has many years of experience of practical application of statistical methods to investigate processes and unravel facts from information. She has carried out extensive work with manufacturing companies, research and training in applied statistical methods for quality improvement and Six Sigma achievement. She is the author of 75 papers and chapters in books.

Dave Stewardson BSc (1st), C.Stat, FRSS is Director of the Industrial Statistics Research Unit at the University of Newcastle, the largest group of its type attached to a University in Europe, and Lecturer at the Technical University of Wroclaw, Poland. He is former President of ENBIS (European Network for Business and Industrial Statistics) and former Director of a number of small businesses in the U.K. over the previous 25 years. He came to Academia in his late 30s and gained a first class degree in Applied Statistics for Business and Industry from Northumbria University (1995). Now a Chartered Statistician, he is currently coordinating two European-funded thematic networks, including Pro-ENBIS, which is helping to underpin the new Pan-European Society of statistical practitioners that is ENBIS. He has many years of experience in business and industry and of applying statistical and operational research methods within the manufacturing, service and public sectors. He is particularly involved in assisting small and medium-sized companies within the north-east of England, and in developing European manufacturing standards, both supported by European Funding. He is the author of 80 papers and articles.

