VFM Technical Guides

Collecting, analysing, and presenting data

How software can help

Why use this guide?

A key feature of good quality value for money reports is a comprehensive analysis of data. Study teams already use a range of quantitative techniques to collect, analyse, and present information from various sources.

Software can assist in a number of ways and the purpose of this guide is to:

- outline the software packages and what they can do to help;
- demonstrate how software can be used for:
- data collection;
- analysis and evaluation;
- presentation;
- illustrate with examples, how software has been used on a number of studies.

Quantitative techniques such as sampling, sensitivity analysis, and modelling can often help teams estimate the financial savings which might be achieved if audited bodies were to implement our recommendations. This is very important in quantifying and demonstrating the impact of the NAO's vfm work.

Used appropriately, software can also contribute to improvements in the cost-effectiveness of studies. However, careful thought needs to be given to how software is applied and, in particular, the type of analysis that is undertaken.

This guide provides practical advice and points to watch for in using software. It is not intended to make vfm staff experts, and often they will need to consult specialists either in-house - such as members of the Technical Advisory Group (TAG) - or externally.

However, it is important that staff have sufficient knowledge to:

- know when a technique or software is appropriate;
- consult specialists as required; and

- act as intelligent customers in appointing consultants and reviewing and acting on their advice.

Contents

Part 1 **Introduction**

Introduction

A significant factor in the quality of the NAO's value for money (vfm) work is the soundness of the analysis and interpretation of data supporting our conclusions and recommendations.

Data analysed in NAO reports are very varied. Typically they will include expenditure trends; staff statistics; indicators of resource utilisation; throughputs, for example, hospital bed occupancy; number of claims processed; geographic distribution of resources; measures of client satisfaction; and a range of other performance measures.

Increasingly, our studies are making use of information from a wider range of sources. Audited bodies are producing more sophisticated data and this should continue as they develop new performance measures. We now draw more on information published by academic institutions and other centres of excellence. And by commissioning surveys we include totally new and independent information in our reports.

Data analysis requires a structured approach involving a number of key stages:

Availability of data. In proposing subjects for potential vfm studies we need to be very clear on the type, extent, and quality of the data that will be available, its source, and ease of access, in order to carry out sufficient reliable analysis.

Seek expert help. Consider the need for expert help in determining the analysis required and in carrying it out either internally - for instance, members of the Technical Advisory Group (TAG) - or externally. Members of TAG and their fields of expertise can be found in a leaflet which is distributed by the central vfm team. The leaflet will be updated regularly.

Relevance to study objectives. The way in which the proposed analysis will help achieve the study objectives should be well thought out and demonstrated at the study design stage.

Data collection. Methods of collecting or obtaining data should be considered very early in the study, together with the potential technical difficulties and resource implications of the proposed approach.

Cost and resource implications. Consider the cost and resource implications of the proposed approach and whether it is affordable.

Reliability. Obviously, we should have sufficient confidence that the data on which the analysis is based are reliable, and this may require some validation checks which can be carried out by a TAG expert such as Janet Snelling (ext 7149) or Fahera Sindhu (ext 7850).

Type of analysis. A range of analytical tools and techniques is available. In selecting the ones to use it is important that they are the most appropriate for the analysis required.

Interpretation. The interpretation of the results of analysis should avoid ambiguity; any assumptions made should be stated and the way in which conclusions have been reached should be explained.

Presentation. Careful thought should be given to the best way to present the results of analysis in order to get messages across simply and effectively.

Software packages can make an important contribution to these various stages, particularly in collecting and sorting data, analysing and interpreting it, and in presenting the results. This guide illustrates how a number of different software packages now available in the Office can be used on vfm work.

The guide assumes that the reader has no prior statistical or computing knowledge. It explains how the different software packages can be applied and the circumstances when study teams may need more specialist assistance.

The table below shows who to contact for advice and help in using software on vfm work.

Key stages **in analysing data**

1 ***Assessing what data are available***

- Type of data, eg quantitative or qualitative
- Extent, quantity, and relevance
- Quality
- Source (manual or electronic)
- Ease of access

2 ***Seek expert help***

- Consider the need for expert help either internally - for example, members of the Technical Advisory Group (TAG) - or externally in determining the analysis required and in carrying it out.

3 ***Relevance to study objectives***

- How will data and the analysis of data meet study objectives?

4 **Type of analysis**

- What type of analysis is appropriate?
- Will it achieve the study objectives?
- Will raw data be in a form suitable for analysis or will they need to be converted to a different format?

5 ***Convince the audited body***

- Demonstrate to the audited body that the approach and analysis is relevant and will give added value.

6 Data collection

- How are the data to be collected eg file examination, survey, electronic transfer to NAO software?

7 ***Cost and resource implications***

- Consider the cost and resource implications of the proposed approach and whether it is affordable.

8 Reliability

- What checks are needed to ensure that the data is reliable?
- How will data be validated?

9 Interpretation

- Have all assumptions been stated?
- Have the results been fully reviewed to validate them, using expert opinion where appropriate?
- Has ambiguity been avoided?

10 Presentation

- Does the presentation of the results of the analysis communicate the key messages clearly?
- Seek expert advice if necessary to advise on the best presentation.

Part 2 **How software packages can assist in vfm work**

How software **packages can assist in vfm work**

Microsoft Office (MS Office) includes the word processing application Word; the spreadsheet Excel; the database system Access; and PowerPoint for graphics presentation.

Excel is suitable for much of the data analysis and presentation required in vfm studies. Access is better for handling large volumes of data and text data.

Three other packages are also available with a number of features which make them particularly useful for vfm work. They are:

• **IDEA**. This can handle much larger data files and facilitate a range of different interrogations.

• **SNAP**. This is useful for designing questionnaires and analysing survey results.

• **SPSS**. This can carry out a range of statistical analyses.

Several of these packages can be used to undertake the same analysis but often one is best for a particular task. The choice of software in each case will be influenced by the type and volume of data and the analytical approach required.

Part 2 gives brief summaries of the features of each package and how they can contribute to vfm work. Parts 3, 4, and 5 then focus on their use in collecting, analysing, and presenting data.

The chart ‘Selecting the most appropriate software’ provides an overview to the most suitable software depending on the type of vfm data collection, analysis, or presentation required.

Selecting the **most appropriate software**

Software	**Excel**	**Access**	**Power-Point**	**IDEA**	**SNAP**	**SPSS**
Business analysis (eg financial	**					

calculations and summary tables)							
Transferring large data files (eg over 10,000 records) from audited bodies' databases.		*		**		*	
Checking, summarising, and sorting large data files.		*		**		*	
Transferring smaller data files (under 10,000 records) from audited bodies' computer systems.	**	**		*		*	
Checking, summarising, and sorting smaller data files.	**	**		*	*	*	
Selecting random samples for vfm analysis.	*	*		**	*		
Designing survey questionnaires, establishing survey databases, and analysing the responses.	*	*			**	*	
Entering and analysing narrative text (over 20 words).	*	**					
Entering and analysing narrative text (under 20 words).	*	**					
Preparing frequency and cross tabulation tables.	*	*			**	**	
Statistical analysis.	*					**	

* appropriate

** most appropriate

Software	Excel	Access	Power-Point	IDEA	SNAP	SPSS	Other*
Analysis using:							
• Index numbers and performance measures	**					*	
• Association between two variables	*				*	**	
• Time series and trend analysis	*					**	**
• Linear and multiple regression	*					**	**
• Multivariate analysis						**	**
• Decision theory							**

• Network analysis						**
• Simulation	*					**
• Linear programming						**
• Queuing theory						**
Presentation						
• Graphs	**	*	*	*	*	
• Statistical graphs (eg box plot)					**	
• Drawing diagrams, objects, organisation charts			**			*
• Visual presentations as screen shows			**			

* appropriate

** most appropriate

* More advanced software required. Consult Technical Advisory Group (extension 7150) or Specialist Software Resource (extension 7167)

Excel

Excel is the MS Office spreadsheet application. It is easy to use and has many powerful features.

What Excel can be used for

- Creating and maintaining simple databases.
- Manipulating, calculating, sorting, and summarising data.
- Carrying out an extensive range of financial calculations, for example converting figures into real prices, calculating net present values, and extrapolating figures.
- Facilitating 'what if' analysis, for example calculating the effect of changing a value or percentage on a given figure.
- Producing a range of graphical presentations including pie diagrams, bar charts, and trend lines; also editing and drawing facilities to enhance graphical presentations.

An advantage of Excel is that the spreadsheets are held in workbooks. Each workbook can hold one or many worksheets so that all related data including graphics can easily be organised in one file.

Ease of use

Simple to advanced skills are required depending on the type of analysis, but generally the software is easy to use.

Examples of how Excel can assist in vfm work

• In *Planning for Change: NAO Examinations of vfm at Grant-Maintained Schools (HC 768 1994-95)* Excel was used to analyse data collected from a sample of schools. The data were sorted to produce bar charts showing performance by type and size of school.

• In a study of Health and Safety in the NHS, Excel was used to examine data on accidents occurring in a sample of NHS Trusts. The type and rate of incidence was analysed and the sample results extrapolated to estimate likely levels of accidents for all Trusts.

• Excel has also been used to analyse information on the type of work carried out by Inland Revenue offices, filtering and sorting it by type, region, and size of office to select a sample for more in-depth examination.

Points to watch for

Excel is easy to operate but users should bear in mind that:

• The maximum number of rows in a spreadsheet is 16,000, so files with more than 16,000 records have to be analysed using another package such as Access or IDEA.

• All spreadsheet work must be clearly documented and easy to follow and review.

• Calculations and totals should be checked to ensure that they include the correct cells of data.

• Building in checks, and checking formulae through the auditing function and the formula view should all be used to validate spreadsheet work.

• One workbook should not hold too many large worksheets as this can slow the work down and the risk of losing data from corrupt files can increase (it is usually best to keep up to six worksheets).

Contact point

For further advice contact IT Resource Centre (extension 7272).

Access

Access can be used for a range of vfm work, from administrative tasks such as tracking the examination of a large number of files and recording the results, to interrogation and detailed analysis of large data files. Access is particularly suitable for studies requiring large volumes of data to be collected, interrogated, and summarised in a standard and secure manner.

What Access can be used for

- More structured (organising data in a more formal and logical framework) and secure databases.
- Handling non numeric data (Access is probably the best package for dealing with narratival information).
- Collecting standard information during fieldwork and surveys, particularly if it is to be combined with information from another source.
- Linking and analysing data held on a number of different files - for example staffing statistics held on one file and workload indicators on another; there needs, however, to be a common identity code.
- Screen interrogations and printed reports.
- Access databases can be used with Excel through an add-on called Data Query - for example, data can be initially collected in Access and then analysed in Excel.

Ease of use

Advanced skills are usually required so seek specialist help.

Examples of how Access can assist in vfm work

In a study of Health and Safety in the NHS Access was used to compile a database of accidents at a number of NHS trusts. Data on the type of accident, its cause and location etc were input to Access, sorted, and initially analysed.

Points to watch for

• Access is excellent for holding large volumes of data, but it is less appropriate for one-off ad-hoc vfm analysis because setting up an Access database can be time consuming. Careful advanced planning and preparation are required. Technical advice and assistance is usually needed to create, manage, and interrogate Access databases reliably.

• Always consider the analysis required and ways to build in checks on the data, for example checks for any duplicate entries.

Contact point

For further advice contact IT Resource Centre (extension 7272).

PowerPoint

PowerPoint is a useful tool for producing graphics for reports and presentations.

What PowerPoint can be used for

- Diagrams, organisation charts and a range of other presentations.

- PowerPoint's library of symbols holds additional graphics such as maps, shapes, signs etc which can be used to enhance standard charts and diagrams.
- Presentations can be run as screen shows from the computer or printed for overhead projector slides and handouts.
- Sets of slides for each presentation are held in one file where they can be created, edited, re-ordered, and re-styled easily and quickly.
- A symbol can be taken from PowerPoint into a chart produced by Excel, for example to convert a bar chart into a pictogram.
- Each PowerPoint screen has an optional speaker's notes facility which can be used when preparing presentations and then printed with the slides.
- Up to six slides can be printed on one page reducing the number of handouts required to print a presentation.

Ease of use

Simple skills are required.

Examples of how PowerPoint can assist in vfm work

• Screen shows on an audio visual unit (AVU) are an effective way of demonstrating the justification for carrying out a study or presenting findings to a range of audiences - including client departments, members of the Committee of Public Accounts, senior management, and other interested parties. PowerPoint can assist in enhancing visual displays as well as producing high quality graphics for reports. Unit E used Powerpoint to give a presentation to the Society of British Aerospace Companies on the NAO report on Initiatives to Reduce Technical Risks.

Points to watch for

Although graphs can be produced by PowerPoint they tend not to be as good as those produced by Excel. Bar graphs, line, and pie charts can be produced in Excel and copied into PowerPoint.

Contact point

For further advice contact IT Resource Centre (extension 7272).

For presentations to smaller audiences the computer can be linked to a projector panel which is placed on an overhead projector. For larger audiences the computer can be linked to an audio visual unit. To use projection equipment contact the IT Resource Centre to book a projector panel, or Accommodation Services to book the audio visual unit.

IDEA

IDEA is a software package developed for auditors by the Canadian Institute of Chartered Accountants. Although it is mainly used for financial audit, IDEA has features that are useful for vfm work including specifically sample selection. The NAO currently uses the DOS version of IDEA which is available throughout the Office.

What IDEA can be used for

- Transferring data held on departments' etc computer files to IDEA for interrogation and analysis by NAO staff.
- The package is capable of handling very high volume data files.
- It is very useful for selecting samples for detailed examination.
- IDEA can cope with data produced in a number of different formats.

Ease of use

Intermediate to advanced skills are required.

Examples of how IDEA can assist in vfm work

• *Metropolitan Police Service: Responding to Calls from the Public (HC 753, 1994-95)* used IDEA to interrogate a large data file on telephone calls made to the police. Approximately one million records holding information on the source of the call, type of call, and response time were interrogated to extract information by types of call.

• *Contracting for Acute Health Care in England (HC 261 1994-95)* IDEA was used to select a sample of general practices from six health authorities for more detailed examination. A sample was drawn based on the number of general practitioners working in each practice.

Points to watch for

• Because of the high volume of data that is often analysed using IDEA, large files can quickly take up computer space. Good housekeeping is needed to ensure that only essential files are retained.

• Training in IDEA is available for vfm staff, covering the software's basic use and an awareness of the data formats. IDEA can be self-taught but it is advisable to attend the vfm IDEA training course if you intend to use the application or need to understand its capabilities. All course members are provided with a manual (IDEA: Beginner's Guide to Auditing with IDEA version 5) produced by Clark Whitehill.

The CAATs team can assist staff with downloading data into IDEA and any complex data interrogation.

Contact point

For further advice contact Ron Cook (extension 7167).

SNAP

SNAP is a package developed by a company called Mercator, for carrying out surveys including questionnaire design, entering data, and analysing survey results. In 1994 SNAP changed from DOS to WINDOWS which produced an easier to use, more powerful package. SNAP is particularly suited for surveys carried out in-house by NAO staff. Alternatively, data collected through surveys can be processed externally, either by Mercator or an agency that uses SNAP, and transferred in SNAP format for in-house analysis.

What SNAP can be used for

All survey work can be processed in one package with the following features:

- Easy design of questionnaires on the computer screen.
- Printing high quality, well-formatted and presented questionnaires.
- Easy and quick inputting of written responses to questionnaires or during face-to-face or telephone interviews.
- A range of data summaries is available, for example percentage responses to questionnaires, analysed by different type and group of respondents.
- Easy interactive analysis on the survey data.
- Batch methods of analysis are available whereby routine tables can be saved and automatically re-run as new data is entered.

Ease of use

The package is easy to learn and use.

Examples of how SNAP can assist in vfm work

NAO have used SNAP since 1992 for a number of surveys, for example:

• *University Purchasing in England (HC 635, 1992-93)* included a postal survey and around 40 completed questionnaires were analysed.

• *Payment of Social Security Benefits to Overseas Customers (HC 407 1993-94)* included a postal survey and around 4000 responses were analysed.

• *NHS Outpatient Service in England and Wales (HC 359 1994-95)*. Some 750 responses from a postal survey of General Practitioners were analysed.

• *Handling Small Claims in the County Courts (HC 271 1995-96)*. A survey was used to assess the quality of service provided by County courts in handling small claims. Some 700 questionnaires completed by plaintiffs and 250 by defendants were analysed.

In many of the surveys the data were processed externally and then analysed in-house using SNAP.

Points to watch for

• It is usually more cost-effective for larger surveys, involving 100 or more respondents, to be processed externally or carried out by market research agencies.

• SNAP can handle only short textual responses to questionnaires etc; for surveys involving larger textual responses another package such as Access should be used.

• Training is available on how to use SNAP. Course members are provided with more detailed SNAP course notes and the SNAP User Manual.

Contact point

Staff wanting to use SNAP should contact Janet Snelling (extension 7149).

SPSS

SPSS is used mainly for statistical analysis. SPSS is a WINDOWS package and includes a large number of statistical analytical techniques.

What SPSS can be used for

- Producing descriptive and summary statistics, frequency tables and cross tabulations, for example, a table counting the number of visits made to an exhibition by the age group of the visitors.
- Modelling different sets of data, for example, regression analysis to deter-mine the strength of the relationship between two or more variables such as staffing levels and workload; this is a useful technique for testing efficiency.
- Calculating different averages, for example weighted averages; geometric means used to average proportions, percentages, and index numbers; and trimmed means used to exclude extreme values eg 5 per cent trim excludes the top 5 per cent and bottom 5 per cent of cases.
- Producing bar charts, histograms, and other presentation formats.

Ease of use

Advanced skills are required.

Examples of how SPSS can assist in vfm work

• *HM Customs and Excise: Writing Off VAT Arrears (HC 209 1994-95)* SPSS was used to analyse the distribution of VAT write-offs; a sample of 4700 traders with VAT debts written off was analysed by type and size.

• *Contracting for Acute Health Care in England (HC 261 1994-95)* A survey of general practices from six regions was selected on the basis of a sample proportional to the number of GPs in each practice. To do this the team required data on both the number of GPs and patients at each practice. However, one region only provided data on patient numbers. Regression analysis was used to determine the strength of the relationship between the number of GPs and patients in five regions. Because there was a direct correlation, patient numbers were used in place of the likely number of GPs in the sixth region. The survey data were processed externally and transferred from a data processing agency in SPSS format for further examination in-house and were analysed mainly using frequency tables and cross-tabulations.

Points to watch for

While SPSS is relatively easy to use a good knowledge of statistics is still needed to analyse data correctly and interpret the results. Vfm training is available for basic data analysis but more sophisticated analysis will usually require specialist help.

Contact point

Staff wanting to use SPSS should contact Janet Snelling (extension 7149).

Data transfer

Data can often be transferred from one software package to another. This may be necessary to facilitate more detailed analysis, for example down-loading a department's data file into IDEA to select a sample and then trans-ferring the sample into Excel for further work. Or the analysis might be partly done in Excel and partly in SPSS. Transfers can be made in a variety of ways and staff should seek advice from Ron Cooke (IDEA) (extension 7167).

Other specialist **software**

The software covered by this guide is capable of carrying out most types of analysis required for the NAO's vfm work. But there may be circumstances when staff need to undertake more advanced analyses requiring other software. For example, as part of the preliminary study on the Mothballing of Defence Equipment, the team used the London School of Economics' decision analysis program, HIVIEW. This enabled them to evaluate the potential impact of a range of options including leasing or selling equipment surplus to current operational requirements.

A number of more advanced packages are held by Unit A's Specialist Software Resource Centre (extension 7167).

Training

Training is available in the use of all these packages on vfm work. Unit C's IT Resource Centre provides introductory and advanced training in Microsoft Office applications. The vfm central team offers courses on the use of Excel and PowerPoint, IDEA, SNAP, and SPSS for data collection, analysis and presentation in vfm studies. Staff interested in attending these courses should consult the Training Programme.

Part 3 **Data collection techniques**

Data collection **techniques**

Software packages can assist data collection in three main ways:

• Downloading data from audited bodies' databases.

• Sampling.

• Carrying out surveys.

A summary of the key points to bear in mind when using these data collection techniques is provided in the following pages.

Downloading data **from audited bodies' systems**

Audited bodies generally hold considerable amounts of information (expenditure trends, staffing statistics, records of claims processed etc) on their computer systems which study teams analyse as part of their vfm work. Data like this can often be 'downloaded' to NAO software, such as **IDEA**, to facilitate in-house interrogation and analysis.

This has the advantage that the study team has access to the data at any time, allowing quicker and potentially less time-consuming examination. The audited body is also less inconvenienced as NAO staff do not have to access their systems directly or request a large number of reports or interrogations.

Key stages to follow

• Acquire a reliable and comprehensive understanding of the audited body's databases and supporting systems, and determine what information is needed.

• Secure agreement of the audited body to downloading of their files to NAO software.

• Relevant data files, records, and fields to be downloaded should be specified and agreed with the audited body.

• The data transfer method and format should be precisely defined and agreed with the audited body.

• Once data has been transferred to NAO software it should be verified to ensure that the correct data was sent and was not corrupted during the transfer.

Transfer medium

The method of data transfer depends on the size of the computer file(s). Small files can be transferred by diskette. For larger files:

• The IT Help Desk provides specialist support for data transfer using backup tapes or direct PC to PC links. Other transfer methods using modems and networks may be possible.

• Large mainframe files can be transferred on magnetic tape. Tape downloading is a specialist task and should be undertaken by the NAO CAATs team. Always discuss any plans for tape transfer with the team, who will advise on the specification required or liaise directly with the audited body's computer department.

Transfer format

Standard formats such as **Excel, Access files, Lotus 123**, or **Dbase** are the best and easiest computer transfer formats.

The best format for transferring large files is **ASCII** (ASCII stands for American Standard Code for Information, a standard text file which is easy to transfer and read). The data file can be in fixed length or variable length format. Fixed length records are easier to transfer but they take up more disk space. With variable length records the fields are separated by delimiters, usually commas. Variable length records are sometimes more difficult to transfer but they are often preferred as they make more efficient use of disk space.

Data verification

All data transferred needs checking. Key points to consider are:

• The confidentiality of data should, in the first instance, be cleared with the audited body. Further concerns over the sensitivity of data, particularly under the Data Protection Act, should be cleared with Ian Petticrew, the NAO Computer Security Officer (extension 7267).

• Any diskette should be checked by the IT Resource Centre for any potential viruses.

• Placing verification checks on fields, looking for outliers and missing results in the data; running check totals and reconciliations to test the consistency of data and to identify errors or anomalies.

Points to watch for

• Data transfer can be time-consuming if the process goes wrong. Incorrect transfer can occur for a number of reasons such as poor specification, out-of-date information, and file corruption during transfer. These risks can be minimised by careful planning, clear specifications, and seeking appropriate technical advice.

• Data must be 'cleaned up' at data transfer and data collection stages. Checks for completeness and reliability of the data must be made.

Contact point

Advice and specialist help are available from:

• IT Resource Centre on data transfer methods (extension 7272)

• Specialist Software Resource Centre (CAATs) on tape transfer, transfer formats, record definitions, and the use of IDEA and other specialist software (extension 7167).

Sampling

Vfm studies often use sampling. It is not practicable or necessary to examine every transaction, for example, in an expenditure programme, or every recipient of a benefit. Conclusions drawn from sample examinations can, with various degrees of confidence, be applicable to a larger population. The software Excel, IDEA, and SPSS are particularly suited for selecting and analysing samples. The NAO publication Use of Sampling - Value for Money Studies 1992 provides more detailed advice, but a number of key points are summarised below.

Getting sampling right is crucial. Savings identified in a sample can often be extrapolated to indicate the potential financial implications for the whole population. This can be very important in demonstrating the financial savings which might be achieved by implementing the study's recommendations, and demonstrating the impact of the NAO's work.

Planning to sample

When planning to sample it is important to be clear on the evidence required, what it will be used for, and the degree of reliability to be placed on the results of the sample.

• **Seek specialist advice**. When using sampling you should seek statistical assistance to check that your approach is valid and to advise on interpreting the result. In more complex cases, it may be appropriate for specialists to undertake the sampling.

• **Determining the population**. The population can be a set of offices, institutions, employees, or recipients of a grant or benefit etc, from which samples are selected. If the population is held on a computer database, the sample can be selected from this. Sometimes a larger population is analysed initially, to identify a special or minority population which is then sampled.

• **Choosing the sampling method**. A number of statistical sampling designs are available: simple random sampling, interval, cluster, and probability proportional to size. The central vfm team can advise on the most appropriate type of sampling to use. In some cases, for example when small sample sizes are needed or random sampling is not possible, non-statistical methods of selection may be necessary.

However, subjective samples may prevent the results from being extrapolated to form judgements about the whole population.

• **Use of software programs**. IDEA can be used to select samples using any of these designs. SPSS is better for simple random or interval sampling, Excel has the flexibility to design any random or interval sample.

• **Interpreting sample results**. When analysing sample results you will need to allow for some uncertainty. The results from a sample examination can be used to calculate a best estimate for the whole population, but the result may be subject to sampling error. The confidence interval allows us to say with a specific degree of confidence that the true figure lies within a range of values.

• **Determining the accuracy required**. Choice of sample size will be influenced by various factors: the precision required; the variability of the components, eg range in the size of capital projects or benefits; and the number of sub-groups in the data being examined. The greater the precision required and the more varied the population, the larger the sample that will be needed. It is important to determine the accuracy required at a very early stage.

• **Sample size**. When determining the sample size, as a general rule you should aim for a statistical sample large enough so there are at least 30 cases in each sub-group you wish to analyse.

Examples of NAO studies using sampling

• *Contracting for Acute Health Care in England (HC 261 1994-95)*. Probability proportional to size design was used to select a sample of general practices for more detailed examination. This method was adopted to ensure that the sample was biased more towards larger practices with a higher number of doctors.

• *Payments of Social Security Benefits to Overseas Customers (HC 407 1993-94)*. This included a survey of recipients: to demonstrate the degree of reliability that could be placed on the survey results, the confidence intervals adopted were published in the report.

• *Department of Trade and Industry's Support for Innovation (HC 715 1994-95)*. This report included a survey to gauge industry's awareness of, and views on, the changes made to the Department's support for innovation. To demonstrate the degree of reliability that could be placed on the results the sampling errors for the overall survey were explained in the report as + or - 3 to 4 per cent at 95 per cent confidence.

Points to watch for

• Estimates and confidence interval calculations are influenced by the method of sampling. Estimates and sampling errors from simple random sampling are fairly straightforward and can be produced in **Excel, SPSS, or SNAP**. However, estimates from other sample designs are usually more complex and expert assistance may be needed.

• Samples designed out of proportion to the population will need re-weighting to calculate population estimates. Survey findings based on a sample should be re-weighted to represent the population, not the respondents to the survey.

Contact point

For further advice contact initially:

• Central vfm team (Janet Snelling, extension 7149)

• Technical Advisory Group (Robin Swan, extension 7150, or Michael Whitehouse, extension 7078).

• Consultants and other external experts appointed to the study.

Surveys

Surveys are now widely used in NAO studies to collect both quantitative and qualitative data. Study teams usually commission market research agencies to carry out large surveys, but some smaller surveys are managed in-house.

The software package **SNAP** can be used to help in designing surveys and analysing the results. Survey data collected by agencies using SNAP can be transferred to SNAP for in-house analysis.

The NAO publications *Designing and Carrying out a Survey 1991* and *Designing, Commissioning and Analysing Surveys 1995* provide more detailed advice but a number of key points are summarised below.

Planning to carry out a survey

• **Seek specialist help**. Even if the study team, and not an agency, is to carry out the survey it is important to seek specialist help.

• **Determining the type of survey**. The main quantitative survey methods are (self-completion) questionnaires, personal interviewing, and telephone interviewing. Computer-assisted methods to support these techniques are: computer-assisted data input (CADI); computer-assisted personal interviewing (CAPI); and computer-assisted telephone interviewing (CATI).

• - ***Postal/self completion surveys***. These are a relatively inexpensive way to reach a widely dispersed population. The potential drawback is that response rates can be low: some NAO reports have recorded response rates as low as 30 per cent. To promote a good response rate, questionnaires should be clearly written and attractively presented. **SNAP** can assist in designing questionnaires and in establishing a database of survey results for analysis.

• - ***Personal interview surveys***. This approach is better for covering complex issues, because interviewees can be guided through difficult questions and their responses observed. Software programs can assist, because questions can be read from a portable computer screen and responses typed directly into the computer. This type of survey can be expensive if a large number of people have to be interviewed to obtain a representative result. But the response rate is usually good.

• - ***Telephone interviews***. These are an inexpensive means of reaching a large audience very quickly. They are particularly good for business surveys. Again software such as **SNAP** can help in formulating the questions and ensuring a logical sequence. Telephone interviews can be difficult to control and validate, and only simple questions can be asked. However these potential difficulties can be overcome through good management.

• **Preparation**. Qualitative research is often necessary before carrying out a survey. In-depth interviews and focus groups may be required to obtain information on which to base the design of the survey and to help prepare questionnaires.

• **Validating results**. In some cases, after completion of the survey, it may be worth holding a series of discussions with some respondents to validate the results of the survey and to obtain additional perspectives.

Recent examples of NAO studies using surveys

Study teams have used a number of market research firms to carry out surveys. These firms often specialise in a particular field such as health. The Market Research Society Yearbook provides a complete list of firms who have experience in undertaking surveys and a copy is held in the NAO library.

Examples of market research firms used to carry out NAO surveys

The Rutherford Appleton Laboratory (HC 138 1993-94). British Market Research Bureau (BMRB) surveyed a sample of users of the laboratory to assess their views on its facilities.

Improving Social Security Services in London: the Provision of Service to Customers (HC 525 1993-94). Research Services Ltd (RSL) based at (Harrow) interviewed face-to-face a sample of 1400 benefit recipients in Greater London.

Scotland's National Museums and Galleries: Quality of Service and Safeguarding the Collections (HC 14 1995-96). System Three Scotland carried out a postal survey of primary schools and personal interviews of Edinburgh residents.

The Protection of Scottish Fisheries (HC 28 1995-96). System Three Scotland surveyed a sample of Scottish fishermen to assess their awareness of the role and function of the Scottish Fisheries Protection Agency.

Overseas Trade Service: Assistance to Exporters (HC 293 1995-96). BJM Research and Consultancy Limited undertook a telephone survey of existing and potential UK exporters to four South-East Asia countries.

The central vfm team maintain a database of recent NAO surveys. All study teams are asked to register surveys they carry out with the Central vfm team and also with the Government's Survey Control Unit (SCU) using NAO Form 163. The database includes information on who conducted the survey, its cost, estimates of interview times involved where appropriate, and the response rate achieved, and can be accessed by all staff. Information provided to the Survey Control Unit informs departments about the NAO's survey work. In return the Unit provides information on surveys carried out by departments as well as advice on best practice.

Points to watch for

Smaller surveys, mainly involving self-completion questionnaires sent and returned by post, are often carried out in-house. It is important to validate data, for example to identify inconsistent responses and to check data entry to avoid input errors. Sufficient time should be allowed for data entry validation, for example double entry of data to ensure accurate input and data checks.

For larger surveys carried out by market research agencies it is important that the study team obtain assurance that the agency has formal data processing procedures and quality control standards.

The survey method, the size of the sample and the purpose of the survey should be fully explained and discussed with the audited body.

Contact point

For further advice contact initially Janet Snelling (extension 7149).

Part 4 **Analysing data**

Analysing data

Software can help in analysing data, using a range of quantitative techniques. This section briefly summarises the following typical techniques which may be appropriate:

• Initial data analysis

• Index numbers and performance measures

• Association between two variables

• Time series

• Modelling

• Financial analysis

Initial data analysis

All analysis begins with an exploratory stage to check the data for reasonableness, bias, abnormalities, outliers, and missing information. The next stage should focus on the profile of the data to explore patterns, trends, and relationships in the information.

Simple data analysis techniques

• **Collating data into groups** so that the magnitude of the numbers can be understood, and arranging the data in numerical order so that the lowest and highest values can be seen at a glance.

• **Determining the spread of data** – the difference between the maximum and minimum values and the distribution of the data.

• **Ranking data** by value, geographic distribution or other relevant criteria.

• **Determining the frequency** with which events, incidences or a particular value occurs.

• **Cross tabulating data**. There may be a multitude of measurements or counts recording different aspects or characteristics of the data. Cross tabulation is appropriate to show the results of two or more methods of classification.

• **Plotting simple time series**, for example to show the costs of a service over a number of years.

• **Calculating simple averages** such as **arithmetic means**, medians (the value of the middle item of an ordered list of data), and modes (the most frequently occurring value or item in a set of data).

• **Measuring potential variations in data**, particularly differences about the mean by calculating the standard deviation. For example, if you intend to describe a set of data and consider that the mean is the most appropriate average you should use the standard deviation to describe the likely variability in the data.

Software

Excel and SPSS can undertake all of these simple techniques, and SNAP can be used for survey analysis. In addition, SPSS can be used for more complex analyses.

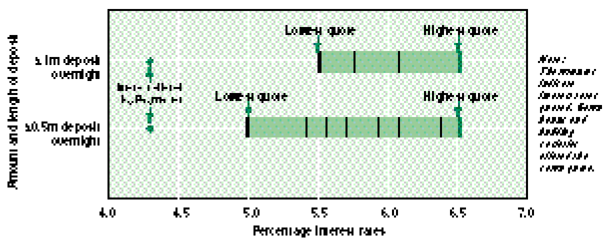
Examples

Highways Agency: The Bridge Programme (HC 282, 1995-96). Often the results of initial data analysis can be presented in reports as descriptive background to put the study into context and set the scene for subsequent more detailed analysis. Here is an example:

See hard copy for example.

Treasury Management in NHS Trusts in England (HC 7 1994-95). This example illustrates one way of showing a range and highlighting maximum and minimum values.

Interest rates quoted by eight banks and building societies to United Bristol Healthcare NHS trust on a day in December 1993



Points to watch for

• Initial data analysis should start with simple techniques, to avoid wasting time in carrying out more sophisticated analysis which may not be appropriate or be based on inaccurate data. Often the most simple analyses are the most telling.

• Outliers should always be investigated to check that they are true values and not errors. The reason for this is that outliers have a very strong effect on many derived statistics such as the mean, standard deviation, and correlation coefficient.

• When we talk of an average, or mean, we are usually referring to the arithmetic mean and so reference to any other form of average such as median should be clearly defined, and if necessary explained.

• The main drawback of using the mean is that it is affected by extreme values, whether outliers or errors, which can distort the message conveyed.

• Some terms, such as ‘standard deviation’, are not widely understood other than by statisticians, and therefore need to be carefully explained in layman terms.

Index numbers **and performance measures**

In presenting data in reports it is often important to show the change in the value or price of a commodity over time and also to compare variables of different units or size. A base period is selected, usually a year, and all subsequent values, commodities, or other units of measurement etc are scaled against the base value to produce index numbers.

The main advantage of index numbers is that percentage increases from the base year can be seen at a glance, and that the numbers shown are of a manageable and understandable size, making it easier for the lay reader to understand the trend over several years.

There are various methods for compiling index numbers. The most common is a simple price index showing the change in prices or costs over time. There are a number of well known indices which are used in vfm studies: the Retail Price Index, the Gross Domestic Product (GDP) deflator, and the Financial Times Share Index. Audited bodies' performance measures, used for internal management or published in annual reports, are sometimes based on index numbers.

Weighting

Composite index numbers cover a larger number of items, which can be weighted so that the indices take account of their relative importance. For example, in a typical basket of goods prices can be weighted by the quantities in the basket to produce indices such as the Retail Price Index (RPI).

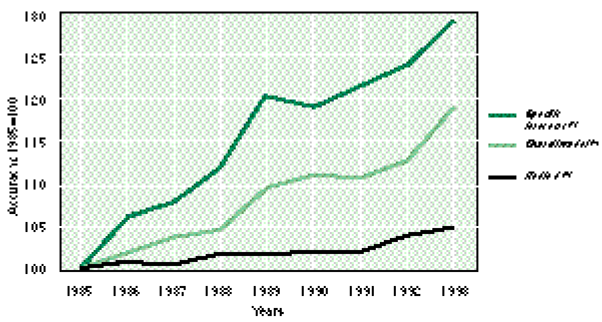
Software

Constructing index numbers involves multiplying one column of numbers by another. Obviously, this sort of calculation can be put into a spreadsheet, and either Excel or SPSS can be used.

Example

The Meteorological Office Executive Agency: Evaluation of Performance (HC 693 1995-96). Taking 1985 as the base year (=100) the team constructed an aggregate index, derived from more than 100 data sets, to measure improvements in the accuracy of the Met. Office's forecasting model.

Forecast accuracy



Notes: (1) Average for specific forecasts derived from the model

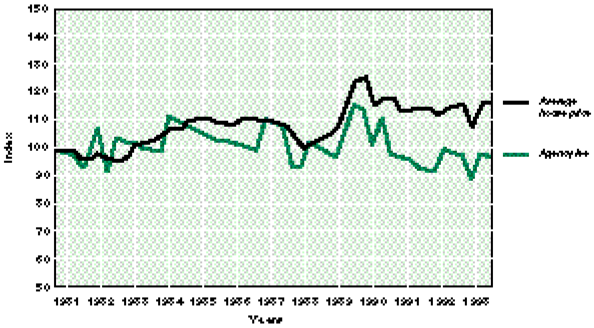
(2) Aggregate index for the whole model

(3) 17:55 Radio 4 forecast

The figure shows the improvements in accuracy of the Meteorological Office's forecasts from their specific targets, the overall model, and the Radio 4 forecast.

The Registers of Scotland: Service to the Public (HC 330 1993-94). This is an example of index numbers being used to show the movement in land registration fees in real terms over twelve years.

Changes in house prices and fees. Average priced house 1981-1993



This figure shows that the average house price in Scotland increased by 18 per cent in real terms from 1981 to 1993, but that the corresponding land registration fee fell by one per cent in real terms.

Source: Nationwide Building Society and gross domestic product deflator.

Points to watch for

• While index numbers are a useful tool they can be misleading if care is not taken.

• Any year can be used as the base but the choice should be carefully considered as it can make a difference to the way the trend is perceived, for example if the base year was in any way atypical.

• In constructing a composite index, the choice of items, the weights applied and the method of calculation can all affect the trend.

• When items are excluded or new items included in an index, there may be drastic movements in the series which do not reflect major changes in prices or quantities, but merely the changed composition of the index.

Association between two variables

In vfm studies we are often interested in looking at the association between two variables. This might be between staffing levels and achieved workload at local offices or the take up of benefits or grants and the number of eligible recipients.

A key consideration is to determine the strength of any association. This approach is useful for vfm work because it can help assess the efficiency with which resources are used as well as giving some indication of their effectiveness. For example, if the association between the number of police officers allocated to solve crimes and the clear up rate was low, this might suggest that more police were needed or that they could be better deployed.

Techniques

The most common techniques for examining association between two variables are:

• **Scatter diagrams**. These show two variables together, one on each axis. The values of the variables are plotted as a point on the graph, for example claims processed against staff resources by location. If the plot forms a straight or a curved line it indicates a clear relationship and can be measured using a correlation coefficient.

• **Chi-squared test**. This is a useful tool for analysing the results from opinion surveys. It can be used to determine whether there is any association between the answers given to a pair of questions. For example, two questions might be asked: what age group are you in (16-24, 25-34 etc) and how many times have you visited a museum in the last three years (0, 1, 2-3, 4 or more). A chi-squared test could determine if there was a statistical association between the two sets of answers, and hence whether age group had any possible influence on visits made.

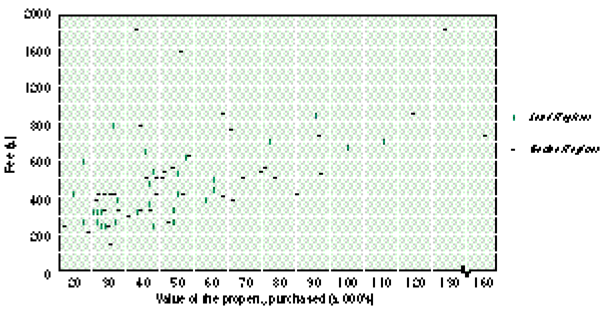
Software

Excel and SPSS can both produce scatter diagrams. SPSS, Excel, and SNAP can calculate chi-squared tests.

Example

The Registers of Scotland: Service to the Public (HC 330 1993-94). This is an example of a study using a scatter diagram to examine the relationship between fees charged by solicitors and the value of houses purchased. While fees and property values are shown to increase together, there is no discernible difference in the fees between registers.

Solicitors fees



The chi-squared test is applied to cross-tabulation tables and is more likely to be used only as part of the analysis in a vfm study. Where appropriate, the use of the technique should be clearly described in the report.

Points to watch for

• A scatter diagram may not present a sufficiently clear picture to include in a published report, particularly if there is a large number of data points or no obvious message.

• The chi-squared test should only be used for the frequency of occurrences (eg responses to a question) and never for percentages.

• Applying the chi-squared test and interpreting its results may require some specialist assistance. Contact Janet Snelling (extension 7149) for advice.

Time series

Trend analysis is useful for investigating the significance of changes in indicators and measures of performance over time.

Approaches

There are two broad approaches:

• **Simple trend analysis** looks at an observation or set of observations at regular intervals over a period of time for clear trends, patterns, or changes – for example, peaks and troughs in activity, seasonality, and cycles. Simple trend analysis assists in measuring and interpreting progress and can be the basis for more sophisticated modelling and forecasting. The first step in analysing a time series is to construct a graph of the data.

• **Time series models**. A wide variety of factors may influence values over a period of time. Some will not be continuous or particularly significant but create ‘noise’ in the picture which makes it difficult to see underlying patterns and trends. Seasonal, cyclical, and random factors are examples of this. Smoothing can be applied to remove the impact of such transient events so that the underlying trend can be seen more clearly. Time series models, based on simple moving averages and other methods, can be used to achieve smoothing.

A simple moving average involves looking at relatively small sections of the data, finding an average, and then moving on to another section. The series of averages is then plotted.

Time series models can also be used to make predictions by extending the trend line into the future. Predictive models vary in their complexity and usually require some weighting to reflect that recent data may have more influence than older data. Applying this approach usually requires specialist advice but it can be used in studies to determine the likely financial impact of a procedure or programme over time.

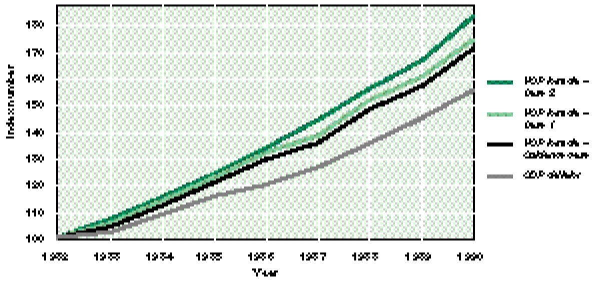
Software

Excel and SPSS can be used to carry out time series analysis and construct more complicated models.

Examples

Accounting for Inflation in Defence Procurement (HC 54 1993-94). This is an example of analysing time series data. The study team used the analysis to demonstrate that a strict application of the Ministry of Defence variation of price guidance would increase the aggregate price of contracts by £50 million a year more than if these contract prices were enhanced in line with the GDP deflator.

Variation of price formulae and the GDP deflator



The graph shows that typical VOPformulae currently in use escalate faster than the GDPdeflator.

Points to watch for

• Over a number of years the definitions of information and the methods of collecting data can change, so it is important to ensure that like is compared with like.

• Where such changes have occurred they should either be acknowledged or, if possible, adjusted for.

• If the time series covers more than two years, all monetary data should be shown in real terms, that is after adjusting for inflation (see ‘Financial reporting’).

• Analysis of time series data is based on the assumption that the behaviour of data in the past will be a good guide to its behaviour in the future. This may be valid, but there will be cases where data are susceptible to change. In these circumstances time series analysis on its own will not be able to predict what the new figure is likely to be.

Modelling techniques

Modelling is used to understand organisational processes, to predict results, to test the effect of a change, and to identify unusual occurrences which might require more detailed investigation. The basis of any model is the assumption that there is some relationship between the items being analysed. A model must be able to cope with a range of conditions and outcomes. Software is usually essential for modelling techniques.

Vfm teams should have some basic awareness of the different types of models:

• to consider whether they might make a valuable contribution to vfm analysis;

• be aware of the circumstances when audited bodies should be using them to assist in the management of their operations; and

• where such techniques are used by audited bodies, to have sufficient knowledge to appraise whether the techniques have been used appropriately and are likely to give reliable results.

Types of modelling techniques

• **Regression**. This is the most common modelling technique, which ranges from simple linear regression to more complex multiple regression models. It is used to examine the relationship between two or more variables and, where a strong relationship exists, to predict possible outcomes.

Practical application for vfm. The technique can be particularly useful in vfm studies. Multiple regression can assist in assessing the strength of any association, and hence the influence of, for example, the size of classroom, number of teachers, level of resources invested in education etc on schools' performance in terms of examination pass rates. Regression can also be used to assess the influence of various inputs on different outputs, for example the number of staff and the number of claims or grants processed or reductions in sick absences in local offices.

• **Multivariate analysis**. This draws on a range of statistical techniques, including principle component analysis, factor analysis, cluster analysis, log-linear models, and discriminant analysis, to examine the relationship between a number of variables. Multivariate analysis is often used by major retailers to determine the optimum location for their outlets taking account of age distribution, socio economic mix of the local population, and optimum travel distances.

Practical application for vfm. One relevance of this type of modelling to vfm work is in its appropriate use by audited bodies in determining the most cost-effective location, for example of ambulance centres, local offices etc.

• **Decision theory**. This can be used to determine the best choice from a number of options. Various techniques are available but essentially all options are assessed using the same set of criteria. Relative costs and probabilities are attached to the criteria for each option. These can be determined subjectively, based on expectations and past knowledge, or objectively using empirical data from past performance. The 'best choice' is then calculated. This approach is applicable for making decisions on which equipment to purchase or capital project or schemes to select.

Practical application for vfm. Decision theory can be used to re-assess a range of projects managed by audited bodies to determine whether the most cost-effective options were selected. Also, in designing vfm studies it can assist teams in determining which cases or projects to examine.

• **Network analysis**. This breaks down the timing and resourcing of a project, or area of work, into smaller tasks. The most efficient order and sequence of tasks is then determined to make the best use of time and resources. Network analysis can be used to help manage large equipment procurement projects, research and development, and construction projects.

Practical application for vfm. Network analysis could be used in vfm studies to assess how well a department had managed a major project and whether they had made optimum use of time and resources. For projects experiencing major cost and time overruns the technique could assist in identifying the causes of the poor performance. Network analysis can also be used to plan more complex vfm studies.

• **Simulation**. This type of modelling can be used to solve 'what if' problems. A model can be constructed consisting of a range of components, some of which will be constant and others variable. Various different assumptions can be applied and a simulation program can be used to show a range of possible outcomes.

Practical application for vfm. Simulation models can be used to show the potential impact of changes in key factors on a programme or project. This would be important if initial evidence suggested that major components such as sub-contractors' work were likely to be delayed or significant design changes were necessary with significant cost implications. Conversely, where a study had identified significant management weaknesses which had resulted in cost overruns, simulation models could be used to indicate a 'what if' the project had been a better managed scenario.

• **Linear programming**. This uses a mathematical or graphical model to find the optimum solution for allocating resources. This could be, for example, the best mix of staff resources to achieve a desired level of output or to determine the output that can be achieved with a given level of resources.

Practical application for vfm. Study teams may face circumstances when they have to evaluate the reliability with which audited bodies have allocated resources. Teams may also need to examine the potential implications of different resource allocation assumptions; linear programming could assist in this.

• **Queuing theory**. This can be used to model, for example, waiting times in processing tax returns or response times in dealing with letters from the public. Models can be constructed to reflect expected demands, the time needed to service demands and the resources available to do so.

Practical application for vfm. Study teams might use queuing theory to model the impact of different efficiency assumptions or improved productivity on the resources needed to process claims or to treat patients etc.

Software

Simple models can be set up on spreadsheets using Excel. Regression models can be built using SPSS. More complex models require advanced programs. In most cases specialist advice is usually needed.

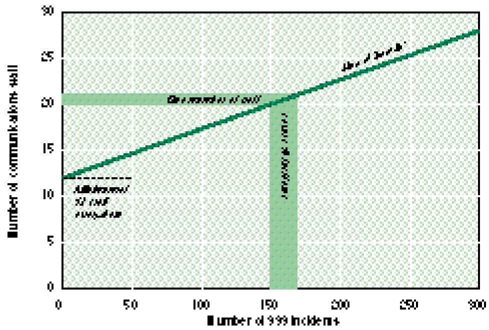
Examples

A number of NAO studies have used modelling techniques as part of their analysis:

Foreign and Commonwealth Office Manpower (HC 358 1989-90). This used multiple regression to identify Foreign and Commonwealth Office overseas posts where staffing appeared high in comparison to reported workload. The Department were asked to give possible explanations why staffing identified by the multiple regression model was higher or lower than expectations.

Metropolitan Police Service: Responding to Calls from the Public (HC 753 1994-95). The study used linear regression to determine optimum staffing levels based on a line of 'best fit' between hourly average scheduled staff and 999 incident numbers. Optimum staffing was compared to actual staffing to identify significant differences. The regression model used is illustrated below

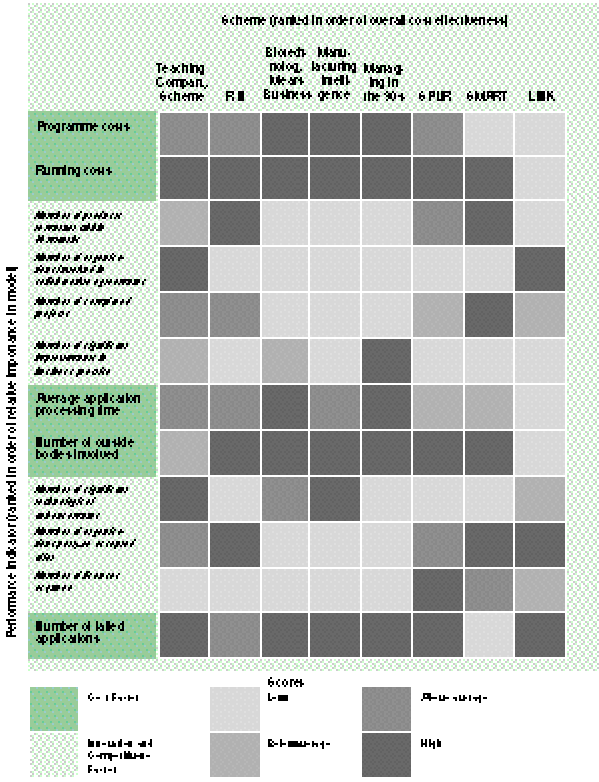
Number of 999 communications room staff against number of incidents recorded for Force Z



The line of 'best fit' gives an optimum staffing level of one for every 19 incidents recorded each hour with a minimum of 12 staff each hour.

Department of Trade and Industry's Support for Innovation (HC 715 1994-95). Assisted by consultants the team developed a model based on decision theory to compare the cost effectiveness of eight schemes to support innovation. The model consisted of seven criteria and 12 performance indicators,which were used to rank schemes as illustrated by the table.

Principal factors contributing to the ranking of the scheme



The table shows the scores achieved by each scheme on each of the 12 most important performance indicators. In particular it shows the relatively poor scores achieved by LINK compared with the relatively high scores achieved by the Teaching Company Scheme.

Points to watch for

• Modelling techniques are useful tools but their appropriateness for vfm studies needs to be carefully thought out and justified. They should only be used when they are likely to provide added value and not for their own sake.

• Specialist advice is nearly always needed.

• Modelling techniques require reliable data, a comprehensive understanding of the material to which the techniques is being applied, and sufficient time to test the model.

• The assumptions on which the model is based need to be clearly understood and communicated.

• The conclusions drawn from the model need to be carefully explained and any caveats highlighted.

Financial reporting

All NAO reports should show financial values in real terms when comparing more than two years of data. Actual values are deflated to show the underlying trend in real terms, after adjusting for the effects of inflation. As a general rule, the GDP (Gross Domestic Product) deflator is used to adjust cash prices to real prices.

Other financial calculations typically included in studies are net present value, internal rate of return, simple and compound interest, depreciation, and percentage growth rates.

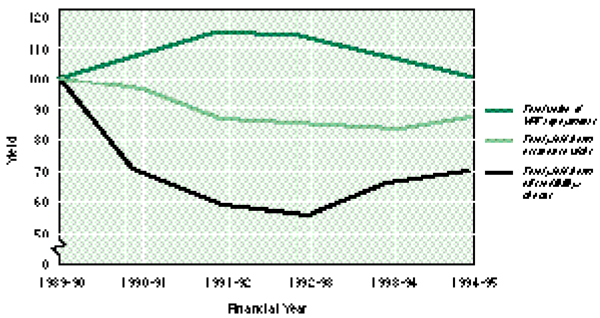
Software

Excel's financial function can be used for the financial calculations.

Examples

HM Customs and Excise: Checking Claims for Repayment of VAT (HC 703 1994-95). The report examines claims over the period 1989-90 to 1994-95. All monetary values are adjusted to 1989-90 prices the first year in the time series, using the GDP deflator.

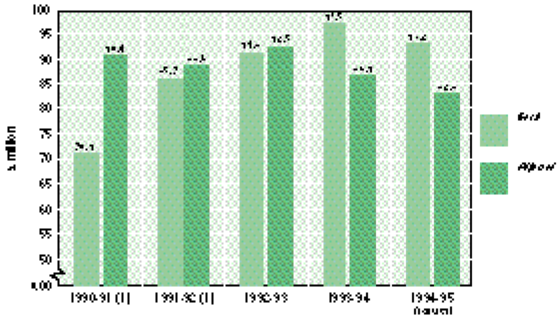
Comparison of the yields from credibility work and assurance visits, 1989-90 to 1994-95 at 1994-95 prices (1989-90 = 100)



After 1989-90, the yield from credibility work showed an initial sharp decline, followed by an upturn in 1993-94 and 1994-95. By comparison the yield from assurance visits has declined each year between 1989-90 and 1993-94, before rising in 1994-95.

The Meteorological Office Executive Agency: Quality of Service to Customers (HC 693 1994-95). A group bar chart is used to show cash and adjusted prices over five years.

Net expenditure



The figure shows that net expenditure in cash terms has increased in each year since 1990-91 but is targeted to reduce in 1994-95. After adjustments net expenditure in 1993-94 was the lowest in this four year period.

PSA Services: The transfer of PSA Building Management to the Private Sector (HC 130, 1995-96). This is an example of calculating net present values.

Payment arrangements for the five PSA Building Management businesses

	1993-94 £m	**1994-95 £m**	**1995-96 £m**	**1996-97 £m**	**1997-98 £m**	**Total £m**
BM Scotland	3.5	0.0	0.0	0.0	0.0	3.5
BM Manchester	(11.5)	0.0	0.0	0.0	0.0	(11.5)
BM Noreast	0.0	2.0	3.0	3.0	0.0	8.0
BM South East and BM South and West	(14.0)	8.0	5.5	5.5	5.4	10.4
Total Cash Proceeds	(22.0)	10.0	8.5	8.5	5.4	10.4
Net present value [1]	(22.0)	9.2	7.2	6.5	3.8	4.7

[1] calculated using a commercial loan rate of 9 per cent – the base rate at 1 October 1993 plus 3 per cent.

Points to watch for

• There will be exceptions where it may not be possible or appropriate to deflate the prices.

• The retail price index is more appropriate than the GDP deflator when deflating prices related to individuals, such as employees' income.

• Deflation indices are sometimes compiled for specialist subjects; for example civil construction indices may be more appropriate for some studies than the broad GDP deflator or Retail Price Index.

Part 5 **Presenting data**

Presenting data

For reports to communicate clearly their conclusions and recommendations, and the analysis supporting them, it is very important that the data they contain are well presented. There are many ways of presenting data, including tables, graphs, charts, and a range of diagrams. The choice depends on the type of data and the messages the illustration is intended to convey.

Most of the software discussed in this guide can present data in several different formats. Data can in some cases be analysed in one package and then transferred to another to present the results in the most appropriate format, whether a graph, a chart, or a diagram.

Excel has a range of charts and a flexible editing facility and is usually the best package for producing graphs for final reports.

PowerPoint is better for diagrams, symbols and non-standard charts, and can also be used to prepare handouts, slides, and screen shows. To include a diagram in a draft report it is simply copied ('cut and paste') from the software which produced it into the correct position in the word processing document.

The following pages show the type of presentations which the software can produce. The *NAO Guide Presenting Data in Reports (November 1991)* provides more detailed advice on design and format. For further advice and specialist assistance contact the DTP Section (extention 7553).

Pie charts

A pie chart is simply a circle divided into segments to provide a clear visual summary of the relative sizes of component parts to the whole. Each segment represents the percentage or quantity of the total. A pie chart should have only six, or at most eight, segments before looking too crowded. Two or three pie charts can be used together to show changes over time or to illustrate differences between groups.

The segments of a pie chart should be placed in logical order, ideally by size or order of importance.

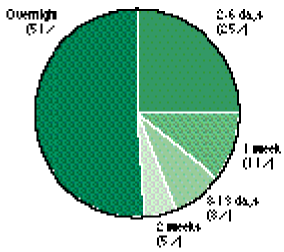
Software

Excel and PowerPoint can both produce pie charts.

Example

Treasury Management in NHS Trusts in England (HC 7 1994-95). This is a good example as a pie chart can only show a small amount of information and should, therefore, only take a small space in a report. The segments are clearly labelled.

Length of investments made by the six trusts



Source: National Audit Office

The figure shows that most investments made by the six trusts were for periods of one week or less.

Points to watch for

• Three dimensional pie charts should be avoided as they can sometimes distort the presentation and provide no extra information.

• Pie charts should be shown as a circle: no more than one segment should be exploded and then only to emphasise its difference.

• Where a series of pie charts is used their size should be kept constant and the segments of different charts kept in the same order.

Simple bar charts

Bar charts consist of a series of vertical solid bars. If more space for labelling is needed, horizontal bars may be used. The length of the bars is proportional to the measurement they represent - the width is common. A simple bar chart is used to compare totals, groups, changes over time, and the component parts of a group (each part presented as a single bar). A bar chart should be used in preference to a pie chart when there are more than six components to display.

Histograms

This is a form of simple bar chart used to display frequencies. It is the graphical equivalent of a frequency table. A histogram should be used when the shape of the frequency distribution is important, for example to illustrate a skewed distribution.

In a histogram, the area of each bar represents the number of observations in each class interval. The bars of a histogram are drawn adjacent to each other, not at the discrete intervals seen in most bar charts.

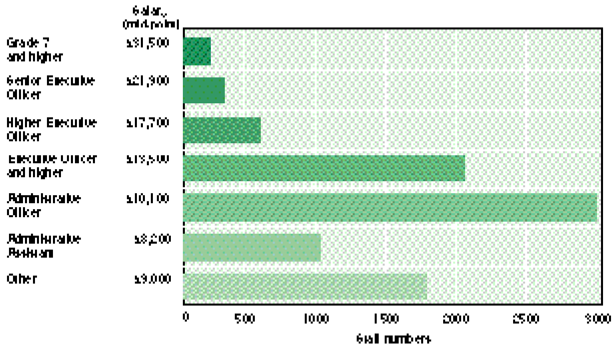
Software

Excel and PowerPoint can both produce bar charts. Excel and SPSS can both produce histograms.

Examples

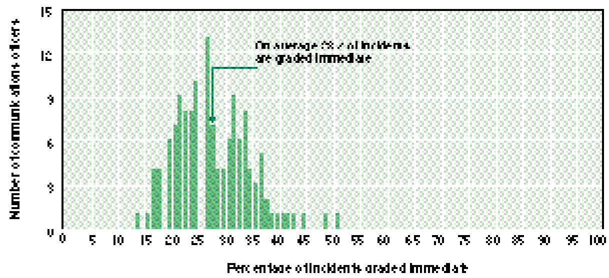
The Management of Sickness Absence in Her Majesty's Land Registry (HC 94 1995-96). This is a simple bar chart comparing staff numbers by grades.

Land Registry staff resources by grade



Metropolitan Police Service: Responding to Calls from the Public (HC 753 1994-95). This is a good example of a histogram. The chart shows the distribution of incidents graded 'immediate'.

Analysis of the grading consistency of communication officers at the Central Communications Complex



Points to watch for

• Bar charts using shading and depth can distort the presentation and should be avoided.

• Three-dimensional charts (that is, with three axes) should be used with care to ensure presentation is not distorted.

• Bars should be measured against a scaled axis. A visual comparison of the bars against each other and the axis should provide sufficient inform-ation, but sometimes it may be necessary to show the actual numbers on top of the bars. The key message is keep the presentation simple.

Component bar charts

A more complex bar chart divides the bars into groups or component parts. Component bar charts should be used to show changes in the components of a group over time, or to compare related groups. Percentage component bar charts should be used to show proportions of a total.

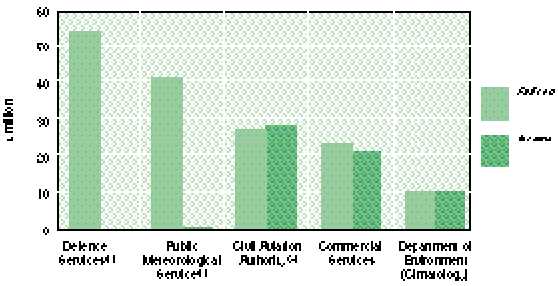
Software

Excel, PowerPoint, and SPSS can produce component and group bar charts.

Examples

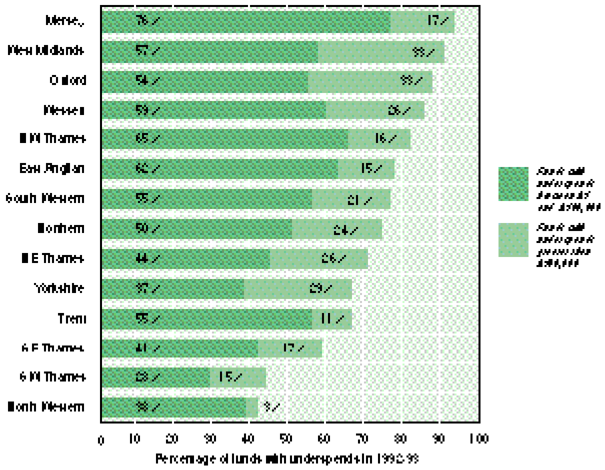
The Meteorological Office Executive Agency: Evaluation of Performance (HC 693 1994-95). This is a good example of a group bar chart. There are two groups: budgeted full costs and income from services to customers. The bars represent categories of business areas and are ordered by costs.

1994-95 budgeted full costs and income of services



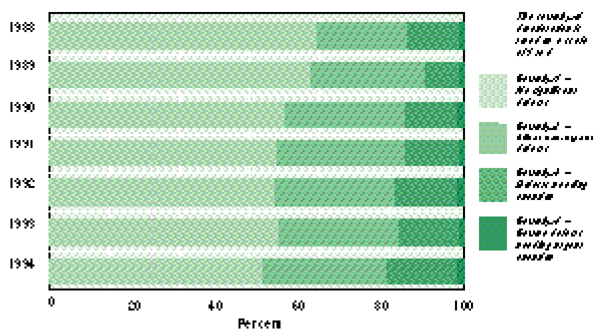
General Practitioner Fundholding in England (HC 51 1994-95). This is a good example of a component bar chart. There are just two components: underspends of up to £100,000 and over £100,000. The bars, shown horizontally, represent regions and are ordered by the percentage of funds underspent.

Percentage of funds with underspends in 1992-93



Highways Agency: The Bridge Programme (HC 282 1995-96). This is a good example of a percentage component bar chart. There are four groups, representing four levels of severity. The bars represent the years 1988 to 1994.

The average severity ratings for key bridge elements, 1988 to 1994



Source: National Audit Office analysis of Highways Agency data

The Highways Agency has kept the proportion of key bridge elements with severe defects to less than two per cent between 1988 and 1994.

Points to watch for

• Component bar charts become less clear as the number of components increase; ideally no more than four components should be used.

• Only use percentage bar charts when the actual values are not important.

• Think carefully about the correct order for the component parts or groups. The components should follow a natural sequence or be placed in order of size following the sequence of the first bar.

• If there is no required order, make the component bar chart easier to follow by placing the most variable component on the top layer of the bars and the least changeable component on the bottom layer.

High-low charts

This type of chart shows the spread of groups of data being compared. It charts the range of data within each group, with the minimum, maximum, and average points plotted. High-low charts can typically be used to show the longest, average, and shortest time to process, for example, claims, grants, or passports over different periods. Their primary function is to analyse data, but they can also be used to present information in reports.

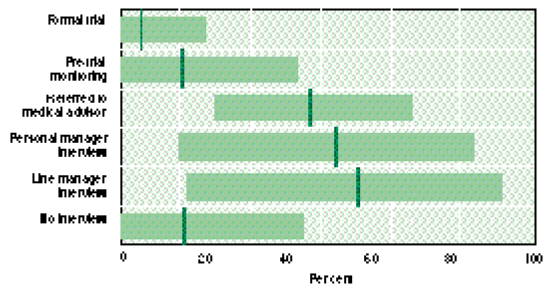
Software

SPSS can produce a variety of high-low charts.

Example

The Management of Sickness Absence in Her Majesty's Land Registry (HC 94 1995-96). This is an example of a high-low chart. It shows the lowest, average and highest percentage of management involvement in cases at district registries.

The range of management involvement in district land registries



Source: National Audit Office analysis of data provided by district registries

Note: The horizontal boxes show the extent of variations between district registries, while the vertical line indicates the average level for all registries.

There were large variations between district registries in the extent of local management involvement, and the percentage of cases which proceed to formal monitoring and trials is low.

Points to watch for

• Plotting data in a high-low chart requires clear labelling and supporting explanation if the message is to be understood.

• For presentation in reports, high-low charts are only useful if there are clear extremes in the distribution of the data.

Graphs showing trends

The line graph is the most appropriate method for illustrating changes over a period of time for one or more series of data. This could be, for example, the level of stocks held by an organisation over a number of years or the change in workload over a number of months. In comparing the trend in related data, mixed graphs can be effective with one series shown as a line and the other as bars.

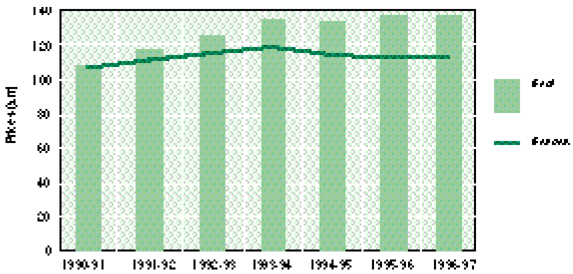
Software

Excel, PowerPoint, and SPSS can produce line graphs. Excel can produce mixed charts.

Examples

BBC World Service (HC 811 1994-95). This is an example of a mixed chart that uses only one scale. The bars represent cash prices and the trend line shows prices in real terms based on 1990-91 prices.

Revenue Grant-in-Aid in cash and constant prices

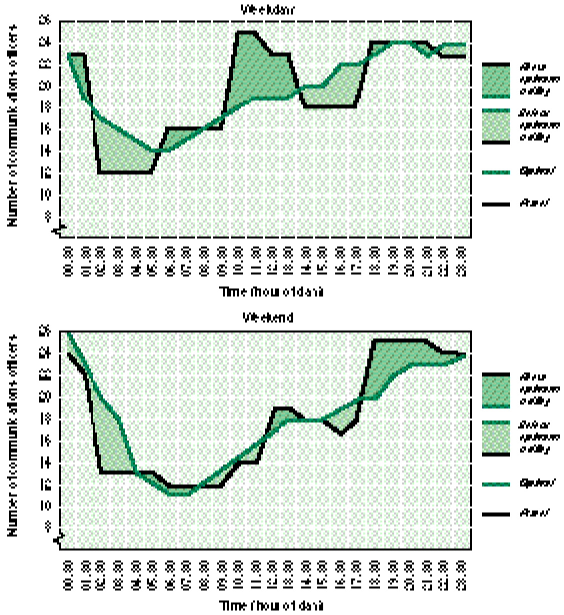


Source: National Audit Office

In the five years 1990-91 to 1994-95 the value of the grant-in-aid has risen by 7.5 per cent in real terms; by the end of the current triennium the increase since 1990-91 will have dropped to 2.8 per cent.

Metropolitan Police Service: Responding to Calls from the Public (HC 753 1994-95). This is a good example of a line graph; having just two lines results in a very clear presentation.

A comparison of actual with optimum levels for the New Scotland Yard central communications during weekdays and at the weekends between 6 June 1994 and 3 July 1994.



The figure shows that whilst staffing broadly matches the optimum levels calculated, there remains some under and over staffing.

Points to watch for

• Five lines are the maximum that should be used before the graph becomes too cluttered.

• Lines should be clearly distinguished by colour or style and clearly labelled.

• Avoid using two different scales for two lines on one graph as they can show misleading comparisons. If two scales have to be used ensure that each axis includes zero and finishes at the maximum point in the series and the related line is clearly labelled. Alternatively, convert both sets of data to index numbers so that they can be compared on the same scale.

X-Y (scatter) charts

X-Y charts are used to show the relationship (if any) between two measures - such as performance achieved against number of staff - by plotting the figures for a number of cases.

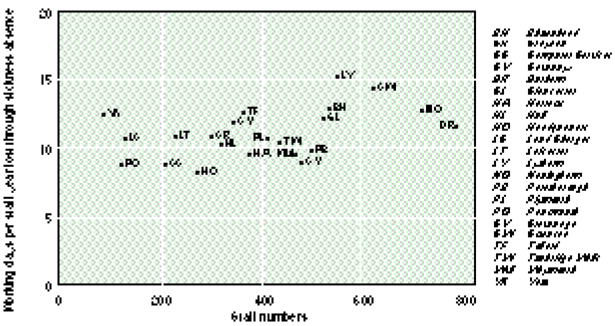
Software

Excel and SPSS can produce X-Y charts.

Example

The Management of Sickness Absence in Her Majesty's Land Registry (HC 94, 1995-96)

Size of office and sickness absence levels



Points to watch for

• When there are only a small number of plotted points, these should be labelled directly on the graph.

• X-Y charts are better for the analysis of data during fieldwork and are only shown in reports if the plotted points are clearly presented and illustrate an important message.

Tables

Tables are used when it is important to provide the actual figures and to highlight small variations. Tables can show summary statistics, time series, and frequencies.

Software

Excel, SNAP, SPSS, and Access can produce tables, Excel can prepare any form of table for presentation in reports.

Example

HM Customs and Excise: Checking Claims for Repayment of VAT (HC 703 1994-95)

Comparison of the yields from credibility work and assurance visits, 1989-90 to 1994-95 (at 1994-95 prices)

Financial year	Total Value of VAT repayments £m	Total yield from all credibility checks on VATrepayments £m	Total yield from assurance visits to all traders (excluding credibility visits) £m
1989-90	24,000	349	947
1990-91	25,800	252	912
1991-92	27,500	221	833
1992-93	27,300	212	813
1993-94	25,800	241	804

1994-95	24,300	251	839
Percentage change 1989-90 to 1994-95	+1%	-28%	-11%

Points to watch for

• The figures should be rounded wherever appropriate: usually two effective digits is sufficient.

• Lists of figures should be organised in a logical order, either following a natural order such as geographical sequence or by ranking by size of the most important item.

Diagrams and organisation charts

Diagrams can often portray important messages in a simple and eye-catching way. Charts illustrating organisations and systems can show basic or complex structures and procedures efficiently and clearly.

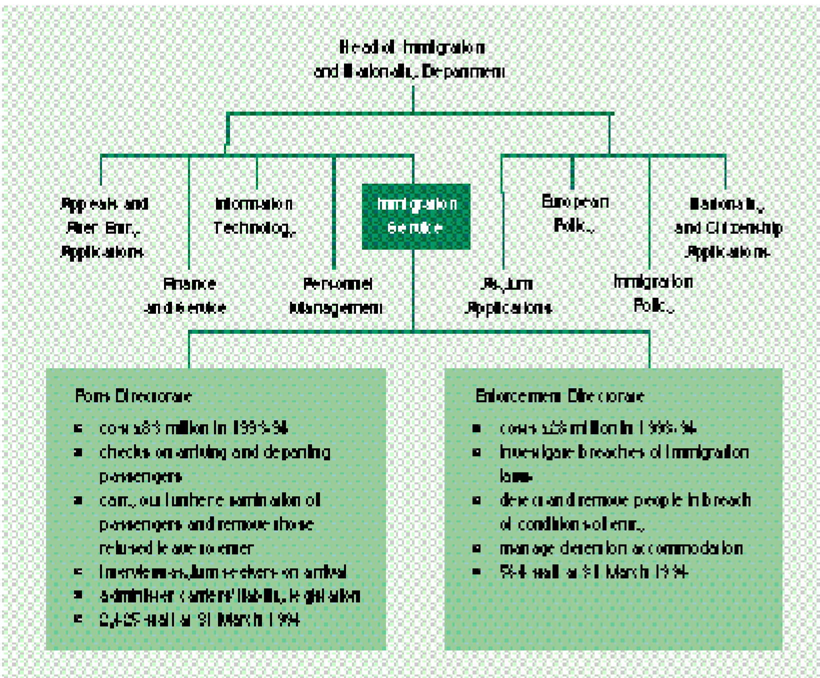
Software

PowerPoint can produce diagrams and organisation charts.

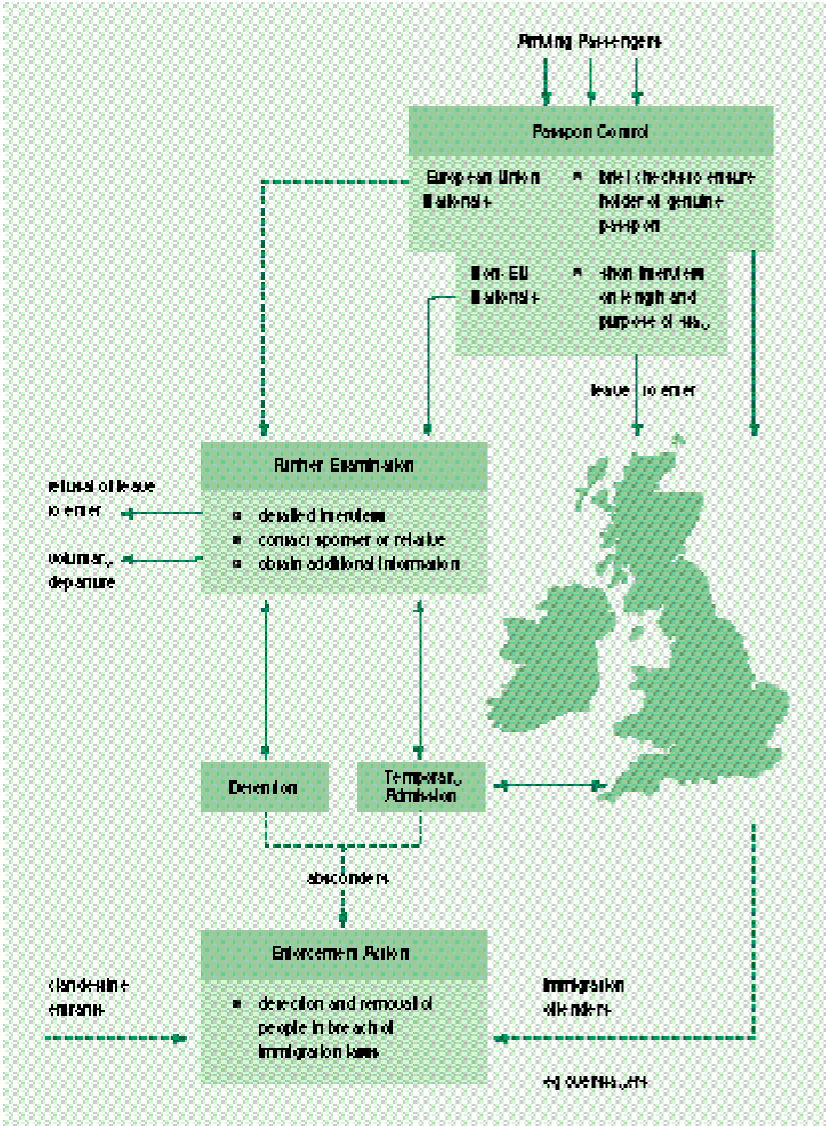
Examples

Entry into the United Kingdom (HC 204 1994-95). The examples show an organisation chart and a diagram illustrating the various stages of immigration control in the United Kingdom.

Structure of the Home Office Immigration and Nationality Department



Stages of immigration controls on passengers arriving in the United Kingdom



Points to watch for

• The diagram should portray a clear and important message.

• The chart should be logical and easily understood.

Maps

Maps are often included in NAO reports to show the geographic distribution of offices, assets, or the recipients of economic assistance or overseas aid.

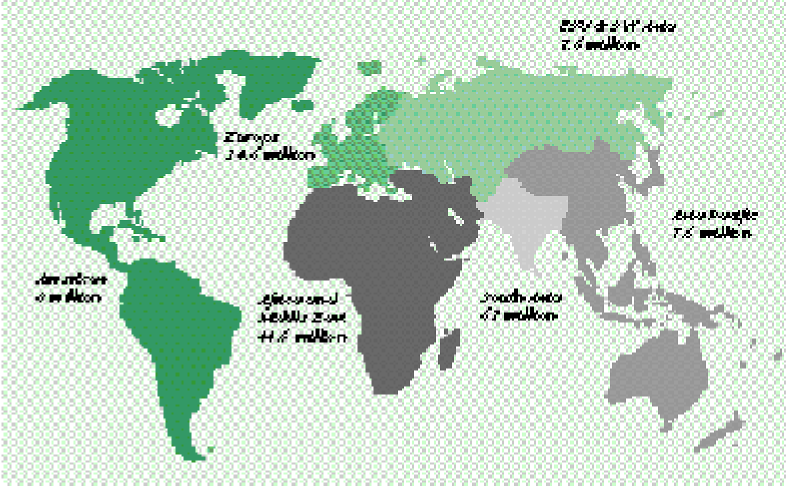
Software

PowerPoint can produce simple maps.

Example

BBC World Service (HC 811 1994-95)

World Service regions and global regular audience



Points to watch for

• Presentation should be clear and simple.

• Careful attention should be given to colour, shading, and labelling.

• Be clear on the point(s) and message which the map is intended to communicate

References

	Author	*Areas covered*
SNAP user guide	Mercator	Use of SNAPsurvey software and surveys
SPSS user guide	SPSS	Use of SPSSstatistical software and statistical analysis
Use of Sampling – Value for Money Studies, NAO, 1992	NAO	Reference guide to sample designs including calculations for estimation and precision
Designing and Carrying out a Survey, NAO, 1991	NAO	Basic guide to surveys and sampling
A Guide to Designing, Commissioning, and Analysing Surveys, NAO, 1995	SCPR	Technical guide to survey work
Market Research Society Yearbook	MRS	List of survey agencies
Presenting Data in Reports, 1991	NAO	Presenting charts and tables
Problem Solving – A Statistician's Guide	Chatfield	Analysing data
Quantitative Methods for	Curwin &	Analysing data

Business Decisions	Slater	
Statistics for Business: Data	Cryer & Miller	Analysing data