

A Guide to Good Practice for Quantitative Veterinary Epidemiology

Report by Veterinary Training in Research Initiative project
VT0101 on quantitative epidemiology



AUTHORS

Mark Woolhouse

Centre for Infectious Diseases and School of Biological Sciences, University of Edinburgh, Ashworth Bldg, West Mains Rd, Edinburgh EH9 3JT

Eric Fèvre

Centre for Infectious Diseases and School of Biological Sciences, University of Edinburgh, Ashworth Bldg, West Mains Rd, Edinburgh EH9 3JT

Ian Handel

Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Roslin, Midlothian, EH25 9PS

Jane Heller

School of Animal & Veterinary Sciences, Charles Sturt University, Locked Bag 588, Wagga Wagga, NSW 2678, Australia

Tim Parkin

Boyd Orr Centre for Population and Ecosystem Health, Faculty of Veterinary Medicine, University of Glasgow, Glasgow G61 1QH

Mike Tildesley

Centre for Immunity, Infection and Evolution, University of Edinburgh, Ashworth Bldg, West Mains Rd, Edinburgh EH9 3JT

Stuart Reid

Faculty of Veterinary Medicine, University of Glasgow, Glasgow G61 1QH

ACKNOWLEDGEMENTS

Katie Atkins Yale University

Mark Bronsvoort Roslin Institute, Edinburgh

Margo Chase-Topping University of Edinburgh

Rob Christley University of Liverpool

Alex Cook Veterinary Laboratories Agency

Matthew Denwood University of Glasgow

George Gunn Scottish Agricultural College

Giles Innocent Biomathematics and Statistics Scotland

Mike Lamont Scottish Government

Alison Mather University of Glasgow

Louise Matthews University of Glasgow

George Milne University of Western Australia

Fletcher Morgan DEFRA

Catherine O'Connor University of Glasgow

Michael Pearce Pfizer

Jay Santhanam Warwick University

Jim Scudamore University of Liverpool

Barbara Weiland Royal Veterinary College

Julia Yates University of Glasgow

and all participants in the Quantitative Veterinary Epidemiology in the 21st century workshop held at Hinxton Cambridge in November 2009.

FUNDING

Department for the Environment, Food and Rural Affairs (DEFRA) and the Scottish Funding Council through Veterinary Training in Research Initiative (VTRI) grant VT0101 awarded to the Universities of Edinburgh and Glasgow.

Research and Policy for Infectious Disease Dynamics (RAPIDD) programme through the Fogarty International Center for Advanced Study in the Health Sciences.

CITATION

This document should be cited as:

Woolhouse, M.E.J., Fèvre, E.M., Handel, I., Heller, J., Tildesley, M.J., Parkin, T. and Reid, S.W.J. (2011). Guide to Good Practice for Quantitative Veterinary Epidemiology. <http://www.qve-goodpracticeguide.org.uk/>.

Version 1 was released on 31st January 2011.

All authors contributed equally to this work.

PREFACE

1. INPUT DATA

- 1.1 Background
- 1.2 Preparation
- 1.3 Data collection
- 1.4 Data management
- 1.5 Data sharing

2. STATISTICAL ANALYSIS

- 2.1 Background
- 2.2 Aims of analysis and choice of approach
- 2.3 Implementation
- 2.4 Inputs
- 2.5 Outputs

3. RISK ANALYSIS

- 3.1 Background
- 3.2 Aims of analysis and choice of approach
- 3.3 Implementation
- 3.4 Inputs
- 3.5 Outputs

4. DYNAMICAL MODELLING

- 4.1 Background
- 4.2 Aims of analysis and choice of approach
- 4.3 Implementation
- 4.4 Inputs
- 4.5 Outputs

5. COMMUNICATION WITH POLICY MAKERS

- 5.1 Background
- 5.2 Relationships
- 5.3 General principles
- 5.4 Informing policy

BOX I. EXPERT OPINION

BOX II. SOFTWARE DEVELOPMENT

BIBLIOGRAPHY

There are no generally accepted standards for ‘good practice’ in quantitative veterinary epidemiology. In this document we address a variety of issues to do with data collection, ways of analysing and modelling data and communication of the results. These issues are relevant in a wide variety of contexts, from ongoing attempts to analyse the impact of badger culling on the incidence of bovine tuberculosis through to the use of mathematical models to inform policy during the epidemics of foot-and-mouth disease in 2001 or of ‘mad cow’ disease (bovine spongiform encephalopathy) in the 1990s. An underlying aim here is to move away from the current position, where quantitative methods are all too often viewed by policy makers and other stakeholders as a mysterious ‘black box’, to their becoming familiar and valued tools.

Many of the issues we address are not just relevant to veterinary epidemiology. Much of the material presented here is equally relevant to medical epidemiology and, beyond that, many of the general principles are relevant to any discipline that utilises sophisticated tools for quantitative analysis, such as climate change studies, economic forecasting or environmental risk assessments. A common feature of these disciplines, in addition to their quantitative nature, is that they increasingly involve ‘big science’ with large teams of researchers and collaborators, direct communication with policy makers and other stakeholders and, in some cases, durations longer than the typical research career. The corollary of this is that many different individuals may be involved in data collection, analysis, communication or implementation of the findings. Therefore it is essential that such studies follow strict guidelines so that all those involved understand and have confidence in what has been done in other places or at other times.

Having identified an absence of good practice guidelines for veterinary epidemiology, we also recognise that good practice guidelines are already available that cover some of the activities of veterinary epidemiologists, starting with the BBSRC/DEFRA/FSA/ NERC Joint Code of Practice (for a full reference to this and other relevant material see the Bibliography). Wherever possible, we refer the reader to existing material. That said, there is inevitably some repetition and we have made little attempt to avoid this on the grounds that it is better to make a valid point twice than not at all. Indeed, we would be delighted if diligent practitioners of veterinary epidemiology found nothing new in this entire document and regarded the contents as no more than common sense and common practice. However, we believe that there is intrinsic merit in developing codes of practice and we hope that this document will be useful for those new to the discipline, for those seeking to review their working practices and for reviewers and auditors of research output. At the very least, putting these guidelines in a single document provides a useful point of reference.

These guidelines are written by practitioners for practitioners, whether they are data managers, statisticians, risk analysts, simulation modellers, or scientists advising policy makers. They are not written explicitly for the users of the work undertaken and, although users may have an interest in the contents (if only to see the view from the other side of the fence), the language is unashamedly technical where appropriate.

The guidelines are organised as five chapters which cover the following topics:

- Input data (Chapter 1);
- Statistical analysis (Chapter 2);
- Risk analysis (Chapter 3);
- Dynamical modelling (Chapter 4);
- Communication with policy makers (Chapter 5).

Each of these chapters is intended to stand alone but nevertheless there are a number of common themes (and hence some repetition). Of these we would highlight eight overarching principles that pervade this entire document:

- 1) clarity of objectives;
- 2) transparency;
- 3) documentation and record keeping;

- 4) verification¹;
- 5) validation²;
- 6) peer review and audit;
- 7) reproducibility;
- 8) communication.

We are well aware that to implement these guidelines in full would be expensive, in terms of time, money and the demand on skills. All concerned need to realise this and should be prepared to ask for, or make available, the necessary resources. At the planning stage for any individual project it may be appropriate to consider what may reasonably be achieved in terms of conforming to good practice guidelines, though we hope that many studies (including our own) will aspire to meet them all.

Finally, we consider this to be a living document. Over time, standards evolve, technologies improve and expectations may grow. Nor do we claim that it is complete: the contents reflect the experiences of the authors – others may wish to add material based on their own experience. For these reasons, comments and suggestions for improving these guidelines can be submitted through our web site:

<http://www.qve-goodpracticeguide.org.uk/>

¹ Defined as the process of determining that a model or simulation implementation accurately represents the developer's conceptual description and specifications.

² Meaning that a model is acceptable for its intended use because it meets specified performance requirements.

1.1 Background

1.1.1 This chapter provides good practice guidelines for the provision of data for quantitative epidemiological analysis and modelling. We provide a checklist of issues that should be considered when planning and implementing an epidemiological data collection exercise. These are organised under three headings: preparation, data quality and data sharing.

1.1.2 Before embarking on an epidemiological study, whether this involves de novo data collection or the analysis of existing data, established references for good practice in study reporting should be consulted. Key sources (see Bibliography) are the STARD statement for studies reporting diagnostic accuracy, the STROBE statement for observational studies and the CONSORT guidelines for clinical trials. These documents give both general and specific advice regarding standards that must be met before such studies can be published, and should therefore be consulted at the initial design stage. Epidemiological studies also have many parallels with the more general topic of data management in other sectors, including the business sector. As such, valuable additional detail is available from the Data Management Book of Knowledge - DAMA DMBOK (see Bibliography).

1.2 Preparation

1.2.1 Research may be researcher-led, or commissioned. When commissioned, the context established by the commissioner's expectations is crucial. Resource and time constraints may impact on the extent to which the study can adhere to good practice. Before agreeing to the work, the research team should report back to the commissioners to ascertain whether, given the known constraints, they can deliver outputs to an acceptable standard. This is a general principle of good project management.

1.2.2 Any research activity must clearly identify the question being addressed, i.e. what is the objective of the work (which may be separated into primary and secondary objectives)? Most data collection exercises address one of three kinds of question: i) hypothesis generation; ii) hypothesis testing; and iii) estimation. At the outset, the question being asked may help to define whether the research will have an immediate bearing on policy or whether it could or should lead to further research.

1.2.3 The case definition and/or the criteria that allow entry of a participant into the study protocol must be clearly set out with the study objectives in mind.

1.2.4 The variable(s) to be studied must be clearly defined, their method of measurement ascertained and sample sizes determined from the outset. This will help to identify when data collection can be completed and whether the study has been a success.

1.2.5 Confounding variables - variables that influence the values of other variables - should be identified and accounted for. There are many analytical methods to deal with confounding, but its existence must be explicitly acknowledged in both the design and analysis stages. This is of particular importance when assessing the utility of existing datasets for novel analyses.

1.2.6 The population under study should be clearly defined, and it should be borne in mind that the results obtained apply to that population and others of which it is representative, but should not be widely extrapolated. The study population is defined as the population of individuals selected to participate in the study, whether or not they do in fact participate. The sample frame is then the list of all sampling units including in the target/study population.

1.2.7 Most studies involve sampling the population, rather than census sampling which would involve inclusion of all elements of the sample frame. Based on the analysis being carried out, the sample size needs to be calculated, for a given power and confidence in a study. This must be done at the outset to determine, e.g. in an observational study, how many individual units from the sample frame need to be selected in order to be confident that the results are representative of the whole of the target/study population.

1.2.8 Often sampling will be clustered, e.g. individuals will be selected for sampling from within a list of homes in the study area where the home is a level of clustering and is referred to as the primary sampling unit. This must be recorded as it affects the analysis of the resultant data.

1.2.9 All procedures for sample collection in the study should be piloted prior to starting data collection to ensure that logistics and data collection systems (data forms, questionnaires, databases, personnel, etc) are working as expected.

1.2.10 Resources must be available to complete the project. Resources should be tracked through the project cycle and objectives adjusted as necessary.

1.2.11 Careful documentation should be kept regarding all the design of the project in the form of Standard Operating Procedures (SOPs)¹. Establishing an SOP, or protocol, for the process of data collection and recording is thus required for effective research and audit. Any changes in SOPs must be recorded and dated (see Bibliography).

1.2.12 Before any data are collected appropriate permissions must be sought from Ethical Review Boards of the institutions to which the study is associated, and countries/ institutions in which the study will be carried out. This applies equally to human studies and to work on animals. All elements of the study (including data management) should be presented to these committees.

1.2.13 Clinical trials must be appropriately registered (see Bibliography) with centralised clinical trial databases (doing so is a requirement of the CONSORT guidelines in human studies). Work on experimental or other animals must be approved institutionally and by national legislative bodies such as the UK Home Office.

1.2.14 Informed, preferably written, consent should be sought when collecting samples or personal data, through the use of an ICD - Informed Consent Document (for adults) or IAD (Informed Assent Document) for adults giving consent on behalf of minors. This will often be a requirement of the ethical review process. The consent process should include details about the study, the risks to the participants, benefits participants will receive and measures for data security.

1.2.15 Consideration may be given to the use of pre-existing data to address the research question. In such cases the objectives of the original data collection exercise must be clearly understood, the original SOPs inspected and ethical requirements considered to help establish whether the data are suitable for current purposes.

1.2.16 The planning elements discussed above are an integral part of the project: they should not be considered as separate or of marginal relevance, and it lies with all levels of the project management team to consider the planning elements before any work is conducted. Planning may also require a budget, and this budget should be included explicitly in the overall project budget.

1.3 Data collection

1.3.1 Data collection is the process of measuring and recording an item of data. Some of the information that accompanies the data itself (“meta-data”) should be stored together with the data (see 1.3.4).

1.3.2 Data collection should be contemporaneous. An observation, when made, should be recorded in some physical or digital permanent form rather than in an ad-doc basis for later transcription. Transcription at a later date to a different medium (e.g. paper to digital) is acceptable given appropriate quality controls, but the original format should be retained as a permanent record.

¹ SOPs are defined as detailed, written instructions to achieve uniformity of the performance of a specific function.

1.3.3 Careful consideration should be given to the possibility of bias in data collection. This could include measurement bias resulting from issues to do with diagnostic sensitivity and specificity or other forms of differential misclassification, operator bias, systematic differences between readings from different pieces of equipment, or the outputs from different laboratories involved in the same study. Every effort should be made to eliminate bias and, where this is not possible, to quantify it.

1.3.4 Metadata recorded will include at a minimum:

- time and date of the observation for each item of data if being collected serially or for a group of observations if they are made as a group;
- geographical position of an observation (e.g. location name, map reference, Global Positioning System reading) noting the units of the positioning data;
- operational details of items of equipment used in the data collecting process – this facilitates auditing and is especially important if faults or mis-calibrations are subsequently identified;
- any issues relating to possible measurement and/or selection bias;
- full name and affiliation of the individual collecting the data, full name and affiliation of the individual recording the data and full name and affiliation of the individual responsible for transcribing the data between formats (e.g. from paper to digital form) – the latter is to ensure that the data recording process is attributable and facilitates auditing.

1.3.5 Data collection and recording should be ‘executed with due care’. This includes the need to calibrate all equipment used to make observations according to manufacturers’ instructions and using recognised standards and working according to clearly set out SOPs.

1.3.6 A decision must be made at the data collection stage regarding the precision of recording of data: is it to match a particular item of equipment or to fulfil a study objective? This level of precision must be maintained consistently throughout, and borne in mind at the data analysis stage.

1.3.7 As far as possible, there should be no missing data, and every effort should be made in data collection and recording to ensure that missing data are minimised. Where data are missing, a consistent method of recording the missing status (bearing in mind that missing \neq 0) should be used. Missing data should be flagged so that it can be appropriately dealt with at the analysis stage (see 2.4.4).

1.3.8 To ensure accuracy and completeness, data should be recorded in a format defined in advance (a ‘proforma’), either in digital or hardcopy form. Substantial effort should go into the data collection SOPs to ensure that all variables are recorded, and external opinion should be sought where possible to ensure the format is clear and unambiguous.

1.3.9 A negative response, a non-response/missing data and a zero entry are separate entities. Blank entries should be properly defined (e.g. in the meta-data) and such values should be consistently recorded in the dataset. Particular regard should be given to the way in which the chosen data management and analysis software handles these values.

1.3.10 The research question must be kept in mind when considering other operational issues such as deviations from designed data collection process and premature stopping of sequential data collection. For example, ethical or cost factors may require a study to be ceased early or to be modified in the light of preliminary results; however, this is likely to have an impact on subsequent statistical analysis.

1.3.11 It may be useful to record an activity log along with the data; this log is a qualitative account of the data gathering exercise, and may help to clarify uncertainties later in the process.

1.3.12 If pre-existing data are to be used to address the study objectives then procedures for the collection of the original data should be reviewed with the question “are these data fit for purpose?” in mind.

1.3.13 It may be appropriate to formally review existing, published data (indeed this may be the objective of the study). This should be undertaken as a systematic review, which involves adhering to certain rules regarding the acceptability of data being used (depending on reporting standards etc in the original studies). The methods for systematic reviews have been outlined by the Cochrane Collaboration (see Bibliography).

1.4 Data Management

1.4.1 Data management covers entering, cleaning, managing, securing and disseminating data.

1.4.2 Ideally, epidemiological data should be stored in an electronic form, which makes using data for modelling or statistical analysis more straightforward and makes data sharing easier. However, electronic storage brings with it some special considerations on security which should be adhered to.

1.4.3 Data transcribed from hardcopy to digital format should be double entered by different operators and each set should be compared for discrepancies. In the case of direct digital data entry, a proforma (see above) should be used to reduce the error rate in data entry, and checks should be put in place to ensure that errors have not occurred.

1.4.4 Before analysis, data held in electronic format should be validated - using validation rules or check routines to ensure correctness, meaningfulness, and security of data that are input to the system. The rules may be implemented through the automated facilities of a data dictionary, or manually by the researcher.

1.4.5 The choice of data management software should be made with the data in mind. For example, some software packages are unable to handle large data sets or certain data values (e.g. blank entries). This must be considered prior to electronic data storage.

1.4.6 Data should ideally be stored in a transferable format (eg .csv file) that is usable across packages and between operating systems.

1.4.7 Epidemiological data may be sensitive. It may contain personal details of study participants, for example, and there is an ethical obligation on the researcher to protect these details. Data should be stored in a secure fashion, while allowing access to those involved in data processing both in the present and in the future.

1.4.8 Special attention should be paid to data on portable media. Such data should be synchronized with a master version where possible, and preferably changes should be linked to the master version with appropriate versioning (see 1.4.9).

1.4.9 A detailed record of versions and edits to versions should be kept. All previous versions should be stored with filenames reflecting versioning, and with metadata descriptions of changes between versions.

1.4.10 Thought should be given as to who on a project team is given permission to READ, WRITE and DELETE data.

1.4.11 Ideally, data should be stored centrally with all versions held on a single (and regularly backed up) server.

1.4.12 A data dictionary is a “centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format.” All data storage should be accompanied by such a dictionary, which identifies the meaning of headings, links between data sets etc. This is an imperative for long-term utility of the database.

1.4.13 Documentation should be in place detailing the security in place to protect a dataset (including back-up protocols, access, user lists etc).

1.4.14 In the same way that SOPs are an integral part of laboratory activity, so they should be for data collection and management. All protocols for data management (as well as collection) should be defined in writing and agreed by all parties.

1.4.15 There will be many instances in which existing datasets or samples can be used to answer research questions. Indeed, making existing data more widely available is an aim of funding bodies and may become a condition of funding in the future. Data may be stored by the original project, or in a data repository. It is essential, however, to ensure that these existing data are appropriate for their intended use and of appropriate quality. The criteria set out for the collection of new data sets in this section can be applied to existing data to help to assess their quality.

1.4.16 In some cases expert opinion may be used to provide inputs into statistical and modelling analyses. This topic is covered in more detail in Box I. Here, we emphasize that the collection and management of data from exercises to solicit expert opinion should conform to the same standards as for any other kind of input data.

1.4.17 Steps should be taken to ensure the security and integrity of the data. This may take the form of a risk assessment and mitigation strategies, for example considering catastrophic data loss, both inadvertent and intentional tampering, and security of sensitive data. This may include use of an off-site backup resource, multiple data redundancies where appropriate, and use of manual or automated access control and usage logging systems protecting the data resource. There are several commercial companies offering these services at different levels of security and redundancy.

1.5 Data sharing

1.5.1 Particularly in multi-user projects, data will need to be shared. Indeed, wherever possible the sharing of data more widely should be the norm, not the exception. Data sharing increases transparency, allows independent reproduction of results, and promotes further analysis.

1.5.2 Appropriate authorisation should be in place for sharing of data in consultation with institutional requirements and with the principal investigator (and with the institution or funder for the period beyond the project itself).

1.5.3 Security of data should be borne in mind during sharing – e.g. use of encrypted connections (i.e. not using open FTP servers). Portable media should be secure during transport.

1.5.4 Elements of the data may be sensitive and/or subject to specific agreements (e.g. confidentiality and data protection). Appropriate measures should be taken before data are shared to ensure that these requirements are followed. This should include awareness of any relevant legislation such as the UK Data Protection Act.

1.5.5 It will often be appropriate, and sometimes a requirement, to report back the results of a data collection exercise to various stakeholders, including the study participants. This aspect of good practice should be respected.

1.5.6 Rules for data sharing and other communications should be set out at the start of the project and included in project documentation.

2.1 Background

2.1.1 This chapter provides good practice guidelines on all aspects of statistical analysis with a scope that includes defining the aims of the analysis, implementing the analysis, management of input data and interpretation and communication of outputs. Due to the wide variety and rapid development of analytical techniques it is not possible or advisable to provide here detailed prescriptive methodologies. The primary focus of these guidelines is to outline the principles of good statistical practice, leaving the choice of specific methodology and tools to the individual analyst. Issues related to quality assurance are also discussed where relevant.

2.1.2 General, discipline-based and organisational guidelines exist to inform good statistical practice. Examples include Preece (2002) for a general discussion of approach to statistical analysis and the Reading University Guides to Good Statistical Practice (see Bibliography).

2.2 Aims of analysis and choice of approach

2.2.1 Identification of the research questions should inform the analysis process. Most statistical analyses address one of four kinds of question: i) hypothesis generation; ii) hypothesis testing; iii) estimation; and iv) prediction.

2.2.2 The collection of data, particularly high volume, high dimensional data can be a legitimate approach for subsequent analysis to generate hypotheses for future investigation. However, this approach requires an explicit a priori statement that hypothesis generation is the aim of the study and analysis. Subsequent analysis of the same data to test these hypotheses may be statistically invalid.

2.2.3 Hypotheses to be tested must be set out in advance of data collection and statistical analysis with rigorous sample size calculations used to inform the experimental procedure. Negative results should be reported to reduce publication bias. Hypothesis testing should be used with caution and only when explicit consideration of effect sizes to be identified and appropriate power calculations have been performed.

2.2.4 Epidemiological research often focuses on estimation of effects rather than hypothesis testing. All estimates should include a measure of uncertainty such as a confidence interval or credible interval, but p-values would not normally be appropriate. The magnitude of the effect and the degree of uncertainty should be considered as separate issues.

2.2.5 Statistical analysis may be used for extrapolation, interpolation and prediction. Such analyses require careful assessment of validity, clear descriptions of methodology, and explicitly defined underlying assumptions.

2.2.6 The analytical approach should be selected and justified in advance of data collection and exploratory data analysis (see 2.3.2). A change of statistical approach between the experimental design and data analysis should not normally be required, and must be adequately justified.

2.3 Implementation

2.3.1 It is essential that a clear, transparent and defensible route from data cleaning to communication of results be documented and adhered to. However the choice of analytical methods will likely vary between statisticians for any given dataset. It is not intended to be prescriptive about this choice here, so long as the methods were chosen and justified in advance of the project, the analyst is competent in the chosen methods and the analyses are conducted with an appropriate question and outcomes in mind.

2.3.2 Exploratory data analysis is a useful component of a thorough statistical analysis, and should be considered obligatory unless analytical or philosophical approaches require statistical models to be estimated without examination of the data (e.g. fully Bayesian approaches). The exploratory methodology should be fully recorded, documented and thus be repeatable.

2.3.3 Wherever possible the chosen statistical analyses should be implemented using validated software. Local validation may involve the comparison of results from an installed package with results using a standard data set with another software system. More generally, software validation, integrity and testing are covered in EU good manufacturing practice guidelines for medicinal products (see Bibliography). Software version numbers and the software routines utilised should be documented against all analyses.

2.3.4 Programmes written for statistical software packages should be carefully checked and tested against data sets that provide readily verifiable results.

2.3.5 It is strongly advised to assess the sensitivity of the results to all assumptions made by at every step of the analysis.

2.3.6 Consideration should be given, preferably at the funding stage, to collection of sufficient data to allow both model fitting and model validation using different data subsets. Results and predictions using multiple subsets of data may be more robust and form better basis for policy decisions.

2.3.7 Meta-analytical methodologies may collate the results from multiple studies to attempt to generalise estimates and conclusions and/or to increase the precision of estimates by effectively increasing the size of the meta-study population. Paragraphs 2.3.1 to 2.3.6 apply equally to meta-analyses, together with considerations of utilising pre-existing data (see 1.4.15).

2.4 Inputs

2.4.1 There must be a reproducible and transparent description of the process by which raw data obtained from data generators is transformed into a cleaned, re-coded and locked format for statistical analysis. The original data should also be maintained, in exactly the format in which it was obtained, in case it needs to be referred to at any point. The format in which the transformed data is kept should be non-proprietary and open, in order that data may be easily accessed and used by other parties both at the time and in the future.

2.4.2 Consistency of file formats between datasets within a collaborating group of analysts is strongly recommended. To ensure consistency of data used for analysis, the data should be held statically in a centralised resource with electronic identity and version checks such as MD5 check-summing. Any changes to the data or requests for and dissemination of the data should be recorded in an electronic lab book.

2.4.3 To ensure transparency and reproducibility, data cleaning and extraction should, where possible, be performed using a script-based approach to extract the data required from a centralised repository. Where it is not possible to use scripts, a detailed log of the process undertaken to retrieve data should be kept. It should always be possible for third parties to reproduce the data used in analysis if given access to the original data source.

2.4.4 Missing or incomplete observations, co-variates or outcome variables require explicit statistical management. An a priori approach to missing data should always be defined during the data collection phase. Management of missing data should be recorded explicitly and should use techniques that will minimise bias and correctly estimate uncertainty in estimators rather than simple omission of incomplete results or carrying forward of previous values. The missing data approach should be clearly documented and be repeatable.

2.4.5 Evidence from previous studies addressing the same research questions may be included intuitively in non-Bayesian analyses or explicitly in Bayesian approaches, so long as the methodology is sound and information from previous studies is used only once in the analysis.

2.5 Outputs

2.5.1 Good communication is essential for an effective, appropriate and valid statistical analysis. This includes careful discussion with the team sourcing the data so the data set subject to statistical analysis preserves the information content of the field data. Good communication with the ultimate customer for the analysis will help the statistical team to provide estimates, inference and predictions appropriate to the particular decision problem.

2.5.2 The statistician must ensure that all records, tables and figures correctly communicate the methodology and findings of the analysis without mis-representation or selection.

2.5.3 Uncertainty in parameter estimates and effect sizes must be clearly communicated, e.g. in the form of confidence intervals or posterior distributions.

2.5.4 Significance test values should be reported with great care. Differences between the results of statistical tests on different data sets should be formally examined, not inferred from differences in levels of significance.

2.5.5 A formal assessment of model fit to the observed data should be provided alongside parameter estimates and significance tests.

2.5.6 Active consideration should be given to the possibility of bias in the interpretation of the results (for examples, see Bibliography).

2.5.7 Major policy changes should not be made on the basis of single studies unless an issue is particularly urgent and important. In such cases, the study should be designed to increase the precision of parameter estimates and the robustness of the findings (see 2.3.6).

3.1 Background

3.1.1 Risk analysis may be defined as the identification and assessment of risk, where risk refers to the unique combination of probability and consequence. In general, risk modelling aims to estimate, with uncertainty, the probability of occurrence and severity of known adverse effects from defined hazards. These hazards may take the form of diseases, infections or events, depending on the studied scenario and discipline.

3.1.2 The implementation of good practice within the process of risk analysis requires careful consideration of the broad areas of model selection and implementation, quality of input data (including expert opinion) and interpretation and communication of outputs. To this end, it is imperative that the risk analyst has, or has access to, a thorough and deep understanding of their discipline, along with adequate knowledge of the system to be modelled.

3.1.3 Published guidelines exist in the areas of good risk assessment practice (e.g. by the European Food Safety Authority (EFSA)), import risk assessment and animal health (e.g. by the World Organisation for Animal Health (OIE)), and microbial risk analysis (e.g. by the Codex Alimentarius Commission¹ (CAC)) – see Bibliography. These pre-existing guidelines are thorough and encompassing, but may not be appropriate for all risk analyses or assessments and are, to some degree, constrained within their specified fields.

3.2 Aims of analysis and choice of approach

3.2.1 The first step is identification of the question to be addressed within the risk analysis, with consideration of the time horizon and units to be modelled. This step should be supported with feedback from stakeholders and experts within the field.

3.2.2 A thorough literature search should be conducted in the field to be analysed at the outset of the project. The analyst should be aware of and have access to published data and possess knowledge of key research groups and stakeholders.

3.2.3 A conceptual model, representing all potential pathways that exist between the hazard and the putative outcome, should then be defined and documented and this conceptual model should form the basis for any (qualitative or quantitative) assessment.

3.2.4 Measurement units (e.g. colony forming units [of bacteria] per g meat) should be defined and their relevance to the question to be addressed justified.

3.2.5 An iterative approach should be adopted from the outset and maintained throughout the entire analysis, whereby communication between analyst, stakeholders, risk managers and field experts, is fluid, dynamic and readily accessible. Communication is an encompassing aspect of risk analysis and should occur in an iterative manner from inception through completion and beyond. Previously defined best practices in risk communication should be adhered to (see Bibliography).

3.2.6 All communication should be transparent, enabling access to and understanding of assumptions, methods and results for all likely audiences. Lexicon should be appropriate for the audience to which communication is aimed and psychological aspects of risk framing and communication should also be considered to maximise comprehension.

3.2.7 At this point, the next step is the identification of the overall risk analysis framework to be used. While previously published model frameworks should be considered, it is possible that these will not be applicable to the process under investigation. In this case, it may be desirable to specify and utilise a modified or novel framework.

¹ Currently under revision and extension to include Risk Assessment Guidance regarding food-borne antimicrobial resistant micro-organisms.

3.2.8 The chosen framework should cover the entire process of risk analysis. There is usually a common structure consisting of the following steps:

- i) identifying the risk (see 3.2.1);
- ii) undertaking a qualitative assessment of the risk (see 3.3.1);
- iii) quantitative or semi-quantitative analysis of the risk and associated management options (see 3.3.4);
- iv) communication of the result and its basis to risk managers and other stakeholders (see Section 3.5);
- v) implementation of an approved risk management strategy (not addressed here).

3.2.9 Each of the steps in 3.2.8 must be represented within the framework that is specified by the analyst. All frameworks (pre-existing, modified or novel) should be accompanied by full documentation and justification for use.

3.2.10 Once a risk analysis framework has been specified, the roles of risk analyst, risk manager, field experts and stakeholders have been defined and lines of communication between these participants secured, specification of the risk assessment, or modelling, component of the analysis should be commenced.

3.3 Implementation

3.3.1 All risk assessments should be undertaken qualitatively in the first instance and a review of the need for, scope of and ability to undertake a quantitative risk assessment should be made based on the outcome of this qualitative assessment. While qualitative assessments may be insufficient or inappropriate for the question posed, they should document areas of lack of knowledge, identify the available literature and define important dependencies or correlations that may require specification in future quantitative models.

3.3.2 The nature of the descriptors used in the risk analysis should be clearly defined. If categorical descriptors are to be used for qualitative assessments, these should be transparent, defined early in the assessment and appropriate to the system under analysis.

3.3.3 All outcomes of qualitative assessments should be defined and interpreted with respect to these initial descriptors. The mechanism of combining categorical estimates in qualitative assessments should be transparent, reproducible, biologically appropriate and defensible. Subjectivity and ambiguous definitions should be avoided. Use of previously published categorical descriptors should be justified with respect to the system now being studied. The analyst should possess an understanding of the limitations of categorical descriptors and matrix-based combinations of these (Cox, 2008).

3.3.4 Specification of a quantitative model requires additional considerations (see also Chapter 4). The model should be of a justifiable level of complexity. Where possible, software that is validated and appropriate to the skill base of the analyst should be used.

3.3.5 The risk analyst should understand the difference between uncertainty and variability, and these should be accounted for where possible in both qualitative and quantitative assessments.

3.3.6 All available data should be considered to establish whether a first (considering uncertainty) or second (considering uncertainty and variability) order model can be parameterised.

3.3.7 A convergence analysis should be undertaken to define the minimum iterations required to reach convergence prior to running any simulation model.

3.3.8 Where possible, uncertainty analyses should be specified and interpreted in light of the outcome of the model and its representative input structure.

3.3.9 All models should be presented with an accompanying sensitivity analysis and this analysis should be appropriate for the model that is specified. Local and global sensitivity analysis techniques should be considered and where possible, global techniques applied.

3.3.10 The influence of the model structure for the risk analysis that has been specified should be considered and any hierarchies, dependencies or co-variables accounted for within the sensitivity analysis. In line with global analysis technique, all epidemiologically plausible interactions should be considered within a sensitivity analysis.

3.3.11 The sensitivity analysis should be broad enough to allow consideration of some input outliers or extreme, but plausible, values. Exploration of the potential effect of outliers should be achievable through a thorough literature review, use of expert opinion, careful consideration of unexpected results and outputs and appropriate and wide-ranging sensitivity analysis.

3.3.12 Model validity should be maximised by the appropriate training of analysts, use of appropriate techniques and software, model testing against scenarios with a known outcomes and iterative updating of the model through consistent communication and feedback from experts and stakeholders.

3.3.13 Transparency and ease of updating should be ensured in both model structure and code. All models should be open to scrutiny and peer review, should be replicable through a defined structure, explicit assumptions and documented input distributions and all code should be published or, at the very least, be available from the author on request.

3.3.14 The duration of validity of the model inputs and structure should be considered and defined within the model and the literature should be frequently scrutinised for new data that presents the need for model update.

3.3.15 External validation of the model should be attempted and this may be undertaken by repeating the model with a separate data set, using stakeholder or expert opinion input, or comparison of model outputs with other statistical, mathematical or risk-based models. Internal model validation may be undertaken through identification of influential parts of the model. The use of appropriate uncertainty analysis, sensitivity analysis, expert opinion and stakeholder feedback may again be used for this internal validation.

3.3.16 Careful consideration of extreme outputs should be undertaken as should the use of restricted models to assess output variation.

3.3.17 Discrepancies between the results of quantitative and qualitative assessments should be explored.

3.4 Inputs

3.4.1 Appropriate and sufficient data is vital to populate a valid risk analysis. Sufficient data may be obtained by implementing an independent data collection process, consideration of 'grey' literature (with acknowledgement of the source, type and limitations posed) and the use of expert opinion.

3.4.2 All data should be obtained from a reliable source and scrutinised carefully for omissions and errors where possible. Where outcomes of previous statistical analyses are used as inputs for the risk assessment, the analyst should possess a thorough understanding of the statistical methods used and an appreciation for potential confounders, which may need to be accounted for within the risk assessment.

3.4.3 As with most areas of risk analysis, networking, communicative feedback and a thorough knowledge of the key research groups and stakeholders are imperative for data identification and acquisition.

3.4.4 It may be appropriate to utilise data that represents a different yet similar system. If this occurs, the applicability of the data to the system under consideration should be outlined and any relevant transformations should be justified and applied.

3.4.5 In some situations, data may be unattainable from any source, in which case, after explicit discussion of the data deficiency, the structure of the model may be altered to accommodate the missing data, or expert opinion may be used (see Box I). These decisions and the reasons for them should be properly documented. In addition, an appropriately encompassing sensitivity analysis (see 3.3.8) should be undertaken for all data-deficient input factors to ascertain their potential relevance for the overall model and to justify omission, or the need for future data collection exercises.

3.5 Outputs

3.5.1 From the outset the outputs of the risk analysis should be determined with a view to effective communication both to risk managers and the wider community.

3.5.2 Separation should exist between the risk analyst and risk manager with respect to decision-making and clear communication between these roles is essential to ensure appropriate dissemination of study findings.

3.5.3 Scenarios that may be appropriate for risk management should be identified through communication between all members of the risk analysis team; risk analyst, risk manager and stakeholders, along with consideration of expert opinion.

3.5.4 All risk analyses should be subjected to peer review where possible and all efforts should be made to ensure that the analysis is published in a journal appropriate to stakeholders and other intended recipients, given the findings of the study.

3.5.5 Communication of findings should not be restricted to model outputs: publication of the model framework and structure should be ensured, as should communication of the results of the sensitivity analysis, along with data gaps that have been identified and any unexpected or counterintuitive results.

4.1 Background

4.1.1 The increased use of dynamical mathematical modelling of infectious diseases for both research and policy development purposes has created a need to consider good practice for the development, implementation and validation of the models. We focus here on simulation models, which are becoming both increasingly complex and widely used as computing power accessible to modellers continues to grow. Simulation models are “virtual worlds” which aim to mimic the real world but by necessity they are approximations of the real world. The challenge facing model designers is to make “good” approximations, ones which capture the key features of the real world. The intention is to obtain results that we can be confident are not only correct in terms of the model but are also valid for the real world too. This chapter summarises the key issues surrounding potential hazards in model selection and implementation, obtaining input data and interpreting and communicating the outputs. We will also provide a guide to ways of minimising the risk of each of these hazards, both when developing a model from scratch and when selecting a pre-existing simulator to model a particular scenario.

4.1.2 There are no published good practice guidelines for dynamical mathematical models of the kind of interest here. However, good practice has been examined for decision analytic healthcare modelling and many of the same guidelines developed in that context (see Bibliography) can be applied to simulation models for infectious diseases.

4.2 Aims of analysis and choice of approach

4.2.1 The first step is to identify clearly the scientific question that the modelling exercise is designed to answer. A well defined question will aid in identification of the modelling approach (or approaches) most appropriate to answer it. Therefore, this step must be completed before any model development is undertaken.

4.2.2 Once the scientific question has been identified, the objectives of the modelling process must be stated and documented such that they are reproducible at the time of publication. This will prove necessary to aid in the reviewing process and is core to stating the requirements which the model should meet. It will also determine whether to use or adapt an existing model or designing and develop a new one.

4.2.3 The next step is to select the appropriate mathematical modelling framework, that is, the appropriate conceptual model. We may investigate existing approaches (see 4.2.2) as to their appropriateness to address the questions to be posed by the model to determine whether to select an existing model or to develop a new one. Clearly, the next step depends on this choice. A particular model design may be appropriate for a model written to provide a qualitative understanding the epidemiological processes whilst an alternate model design may be appropriate for a predictive model used to identify, for instance, optimal control policies to limit disease spread.

4.2.4 Model selection must take into account the nature of the data available. Model selection in infectious disease modelling is often determined by the availability of both demography data and disease specific data. A lack of data may influence a modeller to adopt a simpler model, whilst extensive data may allow a modeller to develop a more complex model. With this in mind, prior to any software development, the data availability must be determined and fully documented. It is important to determine whether the data are ‘fit for purpose’ (see Section 4.4).

4.2.5 Based upon the scientific question and data availability, an extensive literature search should be carried out to determine prior use of equivalent models. Each of these models should be considered under the following criteria:

- Does the model provide a good representation of the biological processes in the system?
- Is the method computationally expensive in terms of run time and software development?
- Whilst the previously used model may have been the “best” approach at the time of development, have there been significant advances which might suggest an alternate approach is more appropriate?

4.2.6 Once the particular mathematical model to be coded by the software has been determined, the choice of approach should be documented and justified. If a novel approach is to be used the new method should be justified with reference to previous methods.

4.2.7 It is good modelling practice to develop simpler models with fewer parameters first, adding complexity should the simpler method be unable to capture the real world process successfully or should more data become available.

4.2.8 Designing, building and parameterising a simulation model for application to a real world situation is often a very complex, time-consuming and expensive procedure. In reacting to an immediate threat, it may be necessary for policy makers to utilise pre-existing models (see 4.2.2).

4.3 Implementation

4.3.1 If we chose to develop a new model, the next step is to design a framework for the model (often a flow diagram) and to present this as a detailed design specification document. A highly structured, detailed, modular and hierarchical design specification will significantly aid the subsequent steps where the simulator software is developed.

4.3.2 The individual modules which have been specified are then implemented as software modules, blocks of program code which may be tested in isolation and then in conjunction with the other modules comprising the simulator. The simulator is thus the software embodiment of the conceptual model whose design is represented in the design document. Some aspects of good practice for software development are listed in Box II.

4.3.3 A structured approach to simulator development is helpful for validating the model and for verification of the simulator. Following good software engineering practice, this design-then-develop process should be implemented with all modelling decisions being made prior to any construction of the software.

4.3.4 A feature of good software development is to start with a simpler, “bare bones” model, develop it and test it before moving on to add more detail (i.e., represent more features of the real world, and then add these features step-by-step, testing after each development step). This procedure is carried out to determine correctness of both modelling the methods and the software. If testing realises anomalous results then the modeller needs to check for conceptual errors and, separately, for software bugs.

4.3.5 The level of detail in the design document should be such that it can be used to audit the model design while the testing process may use test data from past events, or expert knowledge, to determine whether results are plausible.

4.3.6 Where possible, more than one model should be used for analysis to reduce the possibility of erroneous results. This is especially important where the models are intended to inform policy.

4.3.7 A series of test simulations should be carried out for all the models under consideration, to ensure that the outputs are in agreement. This requires determination of a series of scenarios under which to test the simulation software. Such scenarios require care in selection such that they give good coverage of an appropriate range of demographic and movement characteristics and disease biology parameter settings, which range from probable to possible in a real world event.

4.3.8 Once the parameter estimation procedure has been carried out (see Section 4.4), the entire model can be tested to assess its validity. Attempts at model validation are typically of three kinds:

- i) At a minimum, the model should recreate key features of the input data. In itself, this is not true validation. However, it is sometimes possible to derive model expectations independent of inputs, e.g. the date a control policy was changed where this was not input during model development.
- ii) Another validation step is comparison of outputs with previously developed models.
- iii) Ideally, model validation should involve comparison of model predictions with independent data.

4.3.9 Criteria for model validation should be determined in advance of the validation exercise. These should set out which model outputs need to be validated, how this is to be done, and what level of agreement constitutes an adequate fit. These criteria will reflect the aims of the modelling exercise.

4.4 Inputs

4.4.1 Data inputs are usually of three kinds: i) structural data, e.g. demographic data that describes the size and make-up of the population of interest; ii) parameter values, sometimes available from existing literature; and iii) initial conditions. Data inputs are determined by the structure of the model. These inputs must be fully described in model documentation. Primary data should conform to the good practice guidelines set out in Chapter 1.

4.4.2 Original data sources must also be fully referenced. Secondary referencing, e.g. to other modelling studies, is not adequate.

4.4.3 For many applications not all data inputs will be available ‘off-the-shelf’ and so some kinds of estimation, approximation or assumption must be undertaken. There are two common problems with data inputs, as described in 4.4.4 and 4.4.5.

4.4.4 The data may be incomplete (i.e. there is missing information or missing values). If assumptions are made to bridge the data gap, these need to be fully documented so that a reviewer could establish whether the assumptions are appropriate. An alternative approach to missing data is to implement the model for a range of possible values for the “missing data” gap. Separate simulations are then required for each assumed value or set of values. The results are then constrained to be applicable for these values, but the intent is to cover enough representative values that the results are applicable to most if not all probable scenarios.

4.4.5 The data may be obtained from different contexts. The available data may have been obtained from a different population and at an earlier time. Often a judgement has to be made about the suitability of the data for the exercise at hand.

4.4.6 Model development can in turn be used to guide future data collection. A well designed modelling exercise can provide detailed information to data collectors to eventually feed back into future modelling exercises.

4.4.7 Parameter estimation is a key step in model implementation. It is generally a sophisticated procedure and care is needed before embarking upon the process. Various techniques are available. Which is most appropriate depends on both the model and the data available.

4.4.8 One commonly used technique is Maximum Likelihood. This method is used extensively to obtain parameter estimates for infectious disease models, particularly when extensive epidemic data is available.

4.4.9 In addition, Markov chain Monte Carlo (MCMC) approaches are increasingly widely used within a Bayesian framework. This approach is used to match multiple parameters simultaneously and involves running multiple simulations with parameters re-selected from a pre-determined distribution for each simulation. This method is computationally expensive but can often provide parameter estimates that best capture the epidemiological process.

4.4.10 Selection of the best statistical approach to estimate model parameters is often best done by a trained statistician.

4.4.11 In many cases, a lack of data or use of large numbers of parameters make the parameter estimation procedure intractable. Should this be the case, alternate methods should be pursued to obtain some parameter estimates (see 4.4.12, 4.4.13 and 4.4.14).

4.4.12 The literature may provide estimates of some parameters. Sources must be clearly documented. In the event that only point estimates of parameters are available (i.e. with no associated measure of uncertainty), this should also be recorded. It may be possible to use existing parameter estimates as priors in Bayesian approaches to parameter estimation.

4.4.13 Expert opinion can be used (subject to good practice guidance – see Box I) to provide parameter estimates, or to provide point estimates for plausible ranges of values that a given parameter could take. However, if expert opinion is to be used it is vital to identify the appropriate experts and ensure that they are fully informed as to the nature of the parameter(s) of interest.

4.4.14 In some cases, some parameters are simply unknowable in advance; for example, parameters representing the “human response” to an epidemic (e.g. behavioural changes such as avoiding risk of infection). In these cases, estimates of upper and lower bounds for these parameters must be assumed (usually based on expert opinion) and steps taken to explore the effects of values within this range on model outputs.

4.4.15 Once parameter estimates for the model have been obtained it is vital to perform a sensitivity analysis to investigate the dependency of the model to all the parameter value choices made. This is carried out by allowing parameters to vary within set ranges and investigating the relationship between parameter values and model outputs.

4.4.16 Where more than one model is to be used for analysis the models may be structured in very different ways, which makes ensuring that all models are parameterised correctly difficult. It should not simply be assumed that a parameter value used in one model is appropriate for use in a different model.

4.4.17 Ideally, even for a single model, software should be written and parameter estimation procedures be carried out independently by more than one individual or, failing that, the process should be repeated in its entirety by the same individual. This provides internal validation of the software and the mathematical model. In practice, cost and time constraints often preclude this step, but it should be attempted whenever possible.

4.5 Outputs

4.5.1 From the outset of the modelling exercise the outputs of the model should be considered, especially how these outputs should be formatted and communicated and how the outputs relate to the aims of exercise.

4.5.2 The purpose of the model will influence the type of output necessary in the model. It is important that the outputs are written such that they are useful and understood by the appropriate audience.

4.5.3 Carrying out multiple model runs for a range of parameter settings can be computationally expensive. It is important for the modeller to identify the level of detail required in model output, which will be dependent upon the scientific or policy question posed.

4.5.4 The output can provide another internal validation of the model choice and the software. The modeller needs to justify whether the output makes sense in an epidemiological context. If this is not the case, further software or model validation is needed. The fault may lie in the conceptual model (i.e. wrong assumptions, methodology or parameter settings) or could be due to software errors (either intrinsic mistakes in coding or in the mapping from the conceptual model into software).

4.5.5 Model outputs may need to be communicated to policy makers or other non-modelling stakeholders. It is important to be able to provide a full account of the model outputs for a non-mathematical audience. Good practice in output dissemination must include the points covered in

4.5.6 to 4.5.11 (see also Chapter 5).

4.5.6 Results must be carefully explained as there is often scope for misunderstanding model outputs. For example, presentation of mean values of epidemic simulations without confidence intervals presents an inaccurate picture of epidemic risk or the effect of a particular control strategy.

4.5.7 In particular, great care must be taken to communicate clearly “low probability, high risk” scenarios, as these may greatly influence a policy maker’s decision making process.

4.5.8 In instances where good communication with a non-mathematical audience is required, it may be necessary to employ specialist communicators such that the key results are reported in an understandable manner.

4.5.9 The assumptions inherent in the particular modelling approach used need to be identified and set out in the documentation in a very clear way such that they are transparent to non-experts. In a case when an assumption simplifies or approximates the underlying epidemiology, it should be clearly stated why this assumption has been made and how this may influence the results.

4.5.10 The governing equations for the model should be stated clearly. The information provided in papers should be sufficient for the reader to write an equivalent code.

4.5.11 It is essential that all data inputs – structural data, specific parameter settings, initial conditions – are clearly stated, indicating that the outputs are valid only for the combinations of inputs used.

4.5.12 Where more than one model has been used it is possible that outputs from the different models may disagree. In such cases, each model should be deconstructed and tested in order to identify the cause of the discrepancy.

4.5.13 There may be a tendency to prefer the model which agrees with current preconceptions of the epidemiological scenario under investigation. Nevertheless, discrepancies still need to be investigated formally.

4.5.14 At the point that outputs are being communicated to both the scientific community and to stakeholders, it will be necessary to make the model and software documentation available for inspection and challenge.

4.5.15 Peer review prior to publication in scientific journals is an important step. However, in some cases this process may not provide a sufficiently in depth critique of the modelling exercise. In such cases, it may be necessary to appoint external, independent auditors who will examine not just a draft scientific paper but also the full model documentation and the software code. This is especially important when the model is being used to inform policy making.

4.5.16 There is often a pressure to publish in the scientific community, particularly in situations where more than one research group are working on aspects of the same problem. Whilst this may be a key consideration for a mathematical modeller, care must still be taken to allow time for model validation and auditing to be carried out.

4.5.17 When publishing papers in scientific journals, the mathematical modeller should provide details about the modelling approaches used, giving justification for model selection, assumptions and approximations made, details of all model inputs and records of the model testing regime used. This is analogous to the methods section in an experimental paper and will aid in the peer review process – a reviewer will be able to provide a much more informed opinion on the validity of the work.

4.5.18 There are advantages to using journals which allow supplementary material to be submitted, allowing more detail of the modelling exercise to be made widely available.

4.5.19 It is important to clearly define which community the results are aimed at. Is the model aimed at the scientific community and hence designed to handle specific scientific questions, or will it be used to answer policy questions? A model designed to answer policy questions may be scrutinised heavily by a non-mathematical audience and hence it is vital that the whole process is transparent.

4.5.20 Compiled code could be supplied on request to funding bodies or scientists at recognised academic institutions. This would allow external data-driven testing to be carried out to validate the model, following appropriate agreements to protect IP of authors/developers and their institutions.

4.5.21 Software code is often an important piece of intellectual property and for this reason it may not always be appropriate to make it freely available: whether or not this is to be done should be agreed in advance by the modellers and any funding body involved. However, this consideration does not override the need for comprehensive peer review and/or independent auditing.

4.5.22 If software is to be made widely available care must be taken to ensure that the software is easy to use correctly (informally referred to as “idiot-proofing”). Misuse of software may result in incorrect outputs or false error reports, either of which could undermine confidence in the outputs of the original modelling exercise. The release of software for general use will typically require additional investment of time and effort.

5.1 Background

5.1.1 Effective communication or knowledge transfer is the key to a successful outcome of veterinary epidemiological research. Indeed it is a crucial aspect of most applied research. In this context, an outcome can be defined as the impact that research output has on policy, a target population or the wider community. A useful and effective outcome is only achievable if the research output is scientifically robust. The previous themes in this guide to good epidemiological practice have highlighted how different types of epidemiological work can be optimised to best achieve scientifically robust outputs. This final chapter identifies some of the key requirements to enable effective translation of scientific output into successful outcomes. The chapter is organised as three sections: relationships; general principles of effective communication; and informing policy development.

5.1.2 Guidelines for the communication of science to the media have been produced by the Royal Society and as part of the EU-funded MESSENGER project. There are also a number of examples of guides specifically designed to aid the communication of risk in different fields of research (see Bibliography). The UK government's "Guidelines for Scientific Analysis in Policy Making" emphasises how evidence should be sought by policy makers. Guidelines for the research community on communication with policy makers are limited and it was only in October 2009 that The National Environmental Research Council published their very useful guidelines on "Making Science Count in Policy" (see Bibliography).

5.2 Relationships

5.2.1 The establishment and continued development of working relationships and open lines of communication are vital for successful science-policy communication. The research scientist must build relationships with key policymakers and potential stakeholders (i.e. those likely to be effected by or potentially implementing changes in policy or management practice). The lines of communication should be defined at the outset of new working relationships or new projects, within existing relationships. In some circumstances it may not be appropriate for the scientist to have direct contact with stakeholders. It may be more appropriate for all communication to be conducted through the policy maker.

5.2.2 The best working relationships are based on a mutual trust that will develop over time. However, the research scientist can help to engender trust by being transparent in the processes used to achieve the scientific output. If the policy maker can understand the output and also understand how that output was achieved, mutual respect and trust is more likely to develop.

5.2.3 Communication must be frequent and multi-directional. This is essential to provide early warning of unexpected events or results. The policy maker will be able to prepare for and react to unexpected scientific results before wider publication. The scientist may be able to tailor research to fit with new events or policy directions that arise during on going work, helping to maintain the relevance of the scientific output.

5.2.4 Scientific researchers must make themselves aware of the topical issues that are of concern to policy makers and stakeholders. Obviously their research must be relevant to the current specific areas of concern. But, in addition to this they should be aware of funding cycles and delivery expectations. In doing so it is much more likely that their work will be received and accepted into current policy decisions.

5.2.5 The research scientist should be aware at which stage in the funding cycle a particular topic has reached. For example, has work already progressed around the cycle or are the policy makers still setting the agenda? Ideally those with interest in specific topics should be aiming to help define the issues immediately after or while policy makers are setting the agenda.

5.3 Principles

- 5.3.1 The principles behind effective communication strategies have been well documented (see Bibliography). To summarise, the following guidelines are intended to make communication with policy makers more effective.
- 5.3.2 Ensure you are speaking to the correct person or people. It is essential to understand the lines of communication within the department/industry you are communicating with. If you cannot speak to the decision maker, ensure you are at least speaking to an influential advisor to that decision maker.
- 5.3.3 Emphasise what you can do for the policy maker as well as making it clear they can help you. Ensure that the policy maker understands that what you are proposing is part of a wider partnership and that both the scientist and policy maker can drive forward initiatives together.
- 5.3.4 Prepare bullet-point summaries and leave the policy maker with some of the crucial succinct messages to digest after you have left the meeting. This may also be reinforced by a short letter of thanks following a meeting which should include the main points from the meeting as well as creating the opportunity for further contact.
- 5.3.5 Recognise and identify uncertainties in your advice. If you do not do this another advisor to the policy maker may identify flaws in your advice that will ultimately discredit your work.
- 5.3.6 Provide policy makers with options. It is better to provide several options for change, highlighting the pros and cons (including costs if appropriate) for each. It is never appropriate to say “you must do this”, but you can express a preference for a particular option as long as you provide robust reasoning for that choice.
- 5.3.7 The scientist must be proactive in developing contacts. The policy maker will often simply contact the person they know well and whose advice they trust. It is useful to offer your services as a contact for policy makers when they need to seek scientific advice.
- 5.3.8 Avoid the use of jargon and abbreviations and do not get immersed in the details of methodology when trying to get a specific message across.
- 5.3.9 Allow time for absorption and questions. It is likely that what the policy maker/advisor is being asked to consider is, at least in some part, “new” to them. It is essential that they are given time to understand and explore the science and the implications of the advice. It is crucial the policy maker/advisor feels comfortable when “passing on the message” in order to justify his or her decisions.
- 5.3.10 Check that the policy maker fully understands the advice you have provided. If necessary be prepared to present the advice in a different manner or from a different point of view.
- 5.3.11 Keep a record of what was communicated, to whom and when. This will avoid both repetition and gaps in advice to the same or different policy makers at future meetings.
- 5.3.12 Communication to wider audiences may sometimes be appropriate. Use the media (in its many forms, including web based media) where necessary to emphasise a message. But always advise the policy maker who is most likely to be affected by media coverage that you are doing so well in advance.
- 5.3.13 When putting together proposals for work include some budget for costs associated with communication (including travel, time and subsistence). This aspect is often overlooked in project proposals but given the likely small proportion of total costs it is unlikely to prevent work being funded.

5.3.14 Research scientists should react to and learn from past experiences when interacting with policy makers. When developing new proposals and presenting results of current work it is essential that the scientist has an understanding of the potential benefits of a study in terms of the wider policy context. Ensuring that the potential benefits are actually realised is important, but often overlooked. This usually requires active “benefit management” including identifying performance metrics, the measurement of which may well be beyond the scope of your study. However, you can introduce these concepts in the formulation of your study and it may be possible to persuade others to take the measurements for you.

5.3.15 Establishing that your contribution was of real benefit, and being able to quantify the degree of benefit, is an important criteria that policy makers and potential funding bodies will recognise when making future funding decisions.

5.3.16 Effective communication should be: accessible; complete; accurate; relevant; include information on likely cost; secure; reliable; verifiable; timely; flexible; and simple and concise.

5.3.17 The nature of the output or method of reporting will determine how many and which of the characteristics listed in 5.3.16 be considered to be important. For example, lay reports will not include all relevant methodological material but must be accessible, simple and concise. Whereas papers in the scientific literature should include as much detail as necessary to allow replication and as such should be complete, verifiable and accurate.

5.4 Informing policy

5.4.1 Communications strategies will vary dependent on the nature of the policy being advised upon. For example, scientific policy advice to deal with very specific problems will often seek multiple, independent sources of evidence. Individual researchers should aim to collate all possible evidence in order to best appraise the policy maker of likely outcomes of suggested interventions. This will include their work and that of others working in the same field. However, when attempting to drive or influence specific science policy, such as future research priorities, it may be more important to highlight the long-term potential of different research, rather than the immediate impact of potential interventions. Nevertheless, in all scenarios the crucial aspect to remember is that advice and communication should be outcome focused. A thorough examination of the societal outcome or impact of research output should be the ultimate aim of most communication or knowledge transfer.

5.4.2 Policy is developed in response to a strategic aim or agenda. Underpinning policy development will usually be one or more intended outcomes (“real world impacts”) that the policy is aiming to help bring about. These outcomes act as a reference point throughout the policy development cycle to help maintain its focus.

5.4.3 Epidemiological studies and their outputs will contribute to these outcomes. Knowing what these are is an essential part of understanding the purpose of the study. For example, the study may contribute to the scientific evidence base that helps to:

- i) more clearly understand the issues that help formulate the policy;
- ii) develop and appraise implementation options for the policy;
- iii) monitor the implementation, evaluation and any future adaptations of the policy.

5.4.4 When considering the merit of a proposed intervention the policy maker is likely to consider the value, reliability, acceptability, affordability, feasibility and accountability of the proposal.

5.4.5 Many outcomes are interrelated and some are only achievable indirectly (i.e. via intermediate outcomes). Knowing how your study is likely to contribute, can give you scope to adjust your study’s characteristics, and sometimes suggest changes to the form of the output(s).

5.4.6 The information being presented to policy makers should be open to quality assurance (e.g. in a form suitable for peer review). Independent validation is often sought by policy makers and this can only be done effectively if the work and its communication are entirely transparent.

5.4.7 The type of work that has been commissioned should guide expectations. Reactive work is often commissioned by policy makers in response to particular scenarios. This type of work is often conducted over short time scales using seconded staff and so may be sub-optimal, incomplete or not fully robust. On the other hand, proactive work may be conducted over a much longer time scale with dedicated researchers funded to conduct the work. In this scenario it is entirely reasonable to expect the highest level of science. Policy makers and independent reviewers must be made aware of the circumstances surrounding the inception of different pieces of work and must bear these in mind when reviewing work or interpreting research outputs.

5.4.8 Communication of scientific research is more likely to lead to a successful outcome and influence policy if the policy maker knows that the researcher has adhered to good practice guidelines in their work.

- I.1 The use of formalised expert opinion requires adherence to guidelines to ensure that opinions truly represent expert belief and the estimates provided represent those requested and required by the analyst.
- I.2 If it is necessary to rely on expert opinion, careful consideration should be given to the exact approach to be taken. These may range from the informal examination of model output to the elicitation of expert inputs through a structured, formalised approach.
- I.2 The technique used to elicit expert opinion should be transparent, well described, defensible and follow previously published techniques.
- I.3 The analyst should consider whether convergence of opinion is required or desired and if it is, an appropriate technique to ensure convergence is reached should be followed (e.g. Delphi technique).
- I.4 It is advisable that the overall framework from the analysis or modelling exercise is constructed prior to eliciting expert opinion to ensure that all areas requiring expert opinion are encompassed in as few elicitation exercises as practicable.
- I.5 If a workshop is to be run for expert opinion elicitation, this should ideally be run by a person not directly involved in the project to avoid introduction of bias in the discussion and guidance of opinion convergence.
- I.6 Prior to implementation, the constructed questionnaire, or other elicitation technique, should be pre-tested amongst experts and non-experts and refined appropriately. Written documentation should be available for experts that describes the study, introduces the concept of expert opinion elicitation technique and discusses the intended use of all collected information.
- I.7 Communication between experts and the analyst should be encouraged and readily available. Over-reliance on a small number of experts should be avoided.
- I.8 A number of measures may be undertaken to ensure that the opinion obtained is as representative of the truth as possible. Firstly, identification and selection of true experts in the field in which they are to be questioned should be undertaken through appropriate networking, prior knowledge of the field and use of rigorous, well-defined selection criteria. Experts should be selected from as wide a catchment as possible. Non-response or lack of willingness for expert participation may be reduced by explicit discussion and documentation of the requirements of the expert, the manner in which the provided opinion will be used and provision for contact, discussion and feedback from the risk analyst. Potential bias in non-responders should be explored.
- I.9 The questions should be framed to ensure that the information that is obtained is what was intended and this should be undertaken with the consideration of the required input for the model, along with explicit consideration of timescale and units to be modelled.
- I.10 ‘Anchoring’ should be avoided by requesting the ‘most likely’ value prior to asking about minimum and maximum. Experts’ responses should not be limited to a pre-determined range.
- I.11 Estimates should be simplified as much as possible and reliance on experts’ prior knowledge of probability, distributions or density should be avoided.
- I.12 Interpretation of the results of expert opinion exercises should commence with an overview of responses. A thorough understanding of the pertinent literature should allow the analyst to discern the effect of influential published data on the responses.

I.13 Inconsistency between experts may be appropriate or inappropriate depending on the technique used and the presence of geographical, professional or other differences that may legitimately account for divergent estimates. If inconsistency is present and unaccountable then incorrect question framing should be considered. If legitimate divergence exists then the use of subsets of experts to populate certain sections of the analysis should be considered and defended where used. The use of expert based ‘confidence’ ratings for each question may help to account for divergent opinion in some cases.

I.14 Combination of expert opinion should be undertaken using pre-defined, transparent and defensible methods and these may be based on weightings relative to analyst, expert or colleague’s estimates of ‘expertness’.

II.1 Software development is subject to its own good practice guidelines (see Bibliography). A key aspect of good practice in this area is thorough testing at every stage; testing should be regarded as an inherent part of the software development process.

II.2 Code must be well written. Transparency is necessary should internal auditing be required. It is also possible that a personnel change may result in the code being handed over to be utilised by another individual. In this eventuality, transparency in the code is vital to aid in the transition.

II.3 Code should be modular. This will aid in the testing process and assist in clarity for other users. Construction should be carried out stage by stage, with simple simulators being built into more complex simulators. Each routine within the code should be accompanied by an extensive comment section explaining how the routine should work.

II.4 Code should be accompanied by a document with diagrams explaining the model structure, the inputs and outputs of the simulator. This will provide an easy-to-use guide for both future software users and for other scientists and policy makers at the time that results are disseminated.

II.5 During simulator construction, testing should be carried out to ensure that bugs are eliminated. Code also needs to be highly documented, for transparency to aid modification by others and to identify software modifications and rational for modifications.

II.6 When “black box” subroutines are used from previously written software or from other literature (e.g. Fortran Numerical Recipes packages) it is vital that the validity of the subroutine is tested in the context of the model selected. For example, some random number generators are inadequate for large scale simulations – it is important that the software developer investigates all options for black box subroutines before progressing.

II.7 As each module is added to the software, the simulator should be run to test the validity of each update. The results of these tests should be carefully documented so that, in the event of a bug being identified at a later stage, it will be much simpler for the modeller to track down the error.

II.8 Once the software has been fully developed the entire code should be tested for validity. Prior to model parameter estimation (see Sections 3.4 and 4.4) the only testing that can be carried out are testing of the numerical routines. Further testing of the mathematical model will be carried out once parameters have been determined.

PREFACE

- Department for Environment Food and Rural Affairs (DEFRA) (2004) *Joint Code of Practice for Research*. Available from: (<http://www.defra.gov.uk/evidence/science/how/documents/JCoPRGv02-04.pdf>). Accessed 21/09/2010.
- Oreskes, N., Shrader-Frechette, K., Belitz, K. (1994) *Verification, validation, and confirmation of numerical models in the earth sciences*. *Science* 263:641-646. Available from: (<http://courses.washington.edu/ess408/OreskesetalModels.pdf>). Accessed 21/09/2010.
- Rykiel, E.J. (1996) *Testing ecological models: the meaning of validation*. *Ecological Modelling* 90(3):229-244. Available from: (<http://www.sciencedirect.com/science/article/B6VBS-3VWK7V1-7/2/6a5c8fc2b57707e1f028e00432baa7b2>). Accessed 22/09/2010.
- The Royal Society (2005) *The roles of codes of conduct in preventing the misuse of scientific research*. Available from: (<http://royalsociety.org/The-roles-of-codes-of-conduct-in-preventing-the-misuse-of-scientific-research-/>). Accessed 22/09/2010.

CHAPTER 1

- BioPAX (2006) Website: *BioPAX - Biological Pathway Exchange*. Available from: (<http://www.biopax.org/>). Accessed 22/09/2010.
- Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., Moher, D., Rennie, D., de Vet, H.C.W., Lijmer, J.G. (2003) *The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration*. *Annals of Internal Medicine* 138(1):W1-W12. Available from: (<http://www.annals.org/content/138/1/W1.full>). Accessed 22/09/2010.
- Council for International Organizations of Medical Sciences (CIOMS) and World Health Organization (WHO) (2002) *International Ethical Guidelines for Biomedical Research Involving Human Subjects*. Geneva. 62 pp. Available from: (http://www.fhi.org/training/fr/retc/pdf_files/cioms.pdf). Accessed 22/09/2010.
- Current Controlled Trials (2010) Website: *Current Controlled Trials*. Available from (<http://www.controlled-trials.com/>). Accessed 22/10/2010.
- DAMA International, Henderson, D., Mosley, M. (Updated 2010) Website: *Data Management International, Data Management Body of Knowledge (DMBOK) Functional Framework*. Available from: (<http://www.dama.org/i4a/pages/index.cfm?pageid=3552>). Accessed 22/09/2010.
- Dohoo, I., Martin, W., Stryhn, H. (2003) *Veterinary Epidemiologic Research*. Charlottetown, Canada: Prince Edward Island: AVC Inc.
- Dublin Core (2010) Website: *Dublin Core Metadata Initiative (DCMI)*. Available from: (<http://dublincore.org/>). Accessed 22/09/2010.
- Information Commissioner's Office (ICO) - U.K. Government (1998) *Data Protection Act (DPA)*. Available from: (http://www.ico.gov.uk/what_we_cover/data_protection.aspx). Accessed 22/09/2010.
- Moher, D., Schulz, K.F., Altman, D.G. (2001) *The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials*. *The Lancet*, 357(9263):1191-1194.
- The Cochrane Collaboration (2010) Website: *The Cochrane Collaboration*. Available from: (<http://www.cochrane.org/>). Accessed 22/09/2010.

- The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1996) *ICH Harmonised Tripartite Guideline for Good Clinical Practice E6(R1)*, Current Step 4 version (including the Post Step 4 corrections). International Conference on Harmonisation. Conference Report. Available from: (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002874.pdf). Accessed 22/09/2010.
- U.S. National Institutes of Health (Updated 2010) *Website: ClinicalTrials.gov*. Available from: (<http://clinicaltrials.gov/ct2/home>). Accessed 22/09/2010.
- Uhrowczik, P.P. (1973) *Data dictionary/directories*. IBM Systems Journal, 12 (12):332-350.
- von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gotsche, P.C., Vandenbroucke, J.P. (2007) *Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies*. BMJ 335(7624):806-808. Available from: (<http://www.bmj.com/content/335/7624/806.full.pdf>). Accessed 22/09/2010.
- Wellcome Trust (2010) *Policy on data management and sharing*. Available from: (<http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>). Accessed 22/09/2010.
- CHAPTER 2**
- Boden, L.A., Parkin, T.D. (2008) *Current guidelines on good reporting of analytical observational studies in epidemiology*. Equine Vet J. 40(1):84-6. Available from: (<http://onlinelibrary.wiley.com/doi/10.2746/042516408X255927/pdf>). Accessed 22/09/2010.
- Carpenter, J., Bartlett, J., Kenward, M. (Updated 2010) *Website: Missing Data*. Available at: (<http://www.missingdata.org.uk/>). Accessed 22/09/2010.
- European Union - Enterprise and Industry Directorate-General (2008) *The Rules Governing Medicinal Products in the European Union: Volume 4 EU Guidelines to Good Manufacturing Practice Medicinal Products for Human and Veterinary Use, Draft Annex 11 Computerised Systems*. Available from: (http://ec.europa.eu/health/files/eudralex/vol-4/pdfs-en/2008_02_12_introduction_gmp_en.pdf). Accessed 22/09/2010.
- Golbeck, A.L. (1986) *Evaluating statistical validity of research reports: a guide for managers, planners, and researchers*. Gen. Tech. Rep. PSW-87. Berkeley, California: U.S. Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Exp. Stn. 22 p. Available from: (http://www.fs.fed.us/psw/publications/documents/psw_gtr087/psw_gtr087.pdf). Accessed 22/09/2010.
- Kaptchuk, T.J. (2003) *Effect of interpretive bias on research evidence*. BMJ 326(7404):1453-1455. Available from: (<http://www.bmj.com/content/326/7404/1453.full.pdf>). Accessed 22/09/2010.
- Preece, D.A. (1987) *Good Statistical Practice*. Journal of the Royal Statistical Society: Series D (The Statistician) 36(4):397-408.
- Royal Statistical Society and the UK Data Archive (2002) *Preserving and sharing data. Working Group on the Preservation and Sharing of Statistical Material: Information for Data Producers*. Available from: (<http://www.rss.org.uk/PDF/Archiving.pdf>). Accessed 23/09/2010.
- University of Reading (2004) *Guides to Good Statistical Practice*, Statistical Services Centre, University of Reading. Available from: (<http://www.reading.ac.uk/ssc/publications/guides.html>). Accessed 23/09/2010.

CHAPTER 3

- Codex Alimentarius Commission (CAC) and Food and Agriculture Organization of the United Nations (1999) Conference report: Host - Rome, Italy. *Principles and guidelines for the conduct of a microbiological risk assessment. CAC/GL-30*. Available from: (www.codexalimentarius.net/download/standards/357/CXG_030e.pdf). Accessed 23/09/2010.
- Codex Alimentarius Commission (CAC) ad hoc Intergovernmental Task Force on Antimicrobial Resistance (2009) Conference report: Host Government - Republic of Korea. Available from: (http://www.codexalimentarius.net/web/comm_info.jsp?id_1st=40). Accessed 23/09/2010.
- Cox, L.A.T. (2008) *What's Wrong with Risk Matrices?* Risk Analysis 28(2):497-512. Available from: (<http://www.risksociety.org.nz/sites/default/files/uploads/oldsite/risk%20matrix.pdf>). Accessed 23/09/2010.
- European Food Safety Authority (EFSA) (2006) *Transparency in risk assessment carried out by EFSA: Guidance Document on procedural aspects*. EFSA Journal 353:1-16. Available from: (<http://www.efsa.europa.eu/en/scdocs/doc/353.pdf>). Accessed 23/09/2010.
- European Food Safety Authority (EFSA) (2009) *Guidance of the scientific committee on transparency in the scientific aspects of risk assessments carried out by EFSA. Part 2: General Principles*. EFSA Journal 1051:1-22. Available from: (<http://www.efsa.europa.eu/en/scdocs/doc/1051.pdf>). Accessed 23/09/2010.
- International Union of Food Science and Technology (IUFoST) (2007) *Best practices in risk and crisis communication. Advice for food scientists and technologists*. IUFoST Scientific Information Bulletin. Available from: (<http://www.iufost.org/sites/default/files/docs/IUF.SIB.Risk.Communication.pdf>). Accessed 23/09/2010.
- MacGillivray, B.H., Hamilton, P.D., Hrudey, S.E., Reekie, L., Pollard, S.J.T. (2006) *Benchmarking risk analysis practice in the international water sector*. Water Practice and Technology 1(2):1-7.
- Murray, N., World Organisation for Animal Health (OIE) (2006) *Handbook on import risk analysis for animals and animal products*, OIE, Paris.
- National Center for Food Protection and Defense - University of Minnesota (2010), *Best practices for effective risk communication*. Available from: (<http://www.ncfpd.umn.edu/Ncfpd/assets/File/pdf/NCFPDRiskCommBestPractices.pdf>). Accessed 23/09/2010.
- Vose, D. (2000) *Risk analysis: a quantitative guide*. 1st edition. Chichester: John Wiley and sons Ltd.
- World Organisation for Animal Health (OIE) (2009) *Aquatic Animal Health Code - 2008*. Available from: (http://www.oie.int/eng/normes/fcode/en_sommaire.htm). Accessed 23/09/2010.
- World Organisation for Animal Health (OIE) (2009) *Terrestrial Animal Health Code - 2008*. Available from: (http://www.oie.int/eng/normes/Mcode/en_sommaire.htm). Accessed 23/09/2010.

CHAPTER 4

- Abran, A., Moore, J.W., Bourque, P., Dupuis, R. (2004) *SWEBOK guide - Guide to the Software Engineering Body of Knowledge: 2004 Version*. Available from: (<http://www.computer.org/portal/web/swebok>). Accessed 23/09/2010.

- Citro, C.F., Hanushek, E.A., National Research Council (U.S.) (1991) *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling - Reiview and Recommendations*. Washington: National Academy Press. Available from: (http://books.google.co.uk/books?id=9bRcVKbjDEoC&printsec=frontcover&source=gbg_summary_r&cad=0#v=onepage&q&f=false). Accessed 23/09/2010.
- Models of Infectious Disease Agent Study (MIDAS) (2006) *Consultation on Modeling and Public Policy Meeting Report*. National Insititue of General Medical Sciences, Virginia Bioinformatics Institute, Blacksburg, VA, USA. Available from: (http://www.nigms.nih.gov/News/Reports/modelling_policy_060925.htm). Accessed 23/09/2010.
- Weinstein, M.C., Toy, E.L., Sandberg, E.A., Neumann, P.J., Evans, J.S., Kuntz, K.M., Graham, J.D., Hammitt, J.K. (2001) *Modeling for Health Care and Other Policy Decisions: Uses, Roles, and Validity*. Value in Health 4(5):348-361. Available from: (<http://onlinelibrary.wiley.com/doi/10.1046/j.1524-4733.2001.45061.x/pdf>). Accessed 23/09/2010.

CHAPTER 5

- Brownson, R.C., Royer, C., Ewing, R., McBride, T.D. (2006) *Researchers and Policymakers: Travelers in Parallel Universes*. American Journal of Preventive Medicine 30(2):164-172. Available from: (<http://www.sciencedirect.com/science/article/B6VHT-4J5Y9JG-B/2/fbea81c9a62e260c3731178deabcafb8>). Accessed 23/09/2010.
- Catford, J. (2006) *Creating political will: moving from the science to the art of health promotion*. Health Promotion International 21(1):1-4. Available from: (<http://heapro.oxfordjournals.org/content/21/1/1.short>). Accessed 23/09/2010.
- Clayton, H., Culshaw, F., National Environmental Research Council (2009) *Making Science Count in Policy*. (2nd edition). Available from: (<http://www.nerc.ac.uk/publications/corporate/documents/science-into-policy.pdf>). Accessed 23/09/2010.
- Department of Health U.K (1997) *Communicating about risks to public health: pointers to good practice*. Available from: (http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4039670.pdf). Accessed 23/09/2010.
- HM Government U.K (2005) *Guidelines for Scientific Analysis in Policy Making*. Available from: (<http://webarchive.nationalarchives.gov.uk/+http://www.berr.gov.uk/files/file9767.pdf>). Accessed 23/09.2010.
- Keeney, R.L., von Winterfeldt, D. (1986) *Improving Risk Communication*. Risk Analysis 6(4):417-424. Available from: (<http://onlinelibrary.wiley.com/doi/10.1111/j.1539-6924.1986.tb00954.x/pdf>). Accessed 23/09/2010.
- Social Issues Research Centre, Royal Society and Royal Institution of Great Britain (2001) *Guidelines on science and health communication*. Available from: (http://www.sirc.org/publik/revised_guidelines.pdf). Accessed 23/09/2010.
- Stair, R.M., Reynolds, G.W. (2008) *Fundamentals of Information Systems, A Managerial Approach*. Canada: Thomson Course Technology.
- Social Issues Research Centre (SIRC) and Amsterdam School of Communications Research (ASCoR) *Guidelines for scientists on communicating with the media*, EU-funded MESSENGER project. Available from: (http://www.sirc.org/messenger/messenger_guidelines.pdf). Accessed 23/09/2010.