

REVIEW ARTICLE



Standard survey methods for estimating colony losses and explanatory risk factors in *Apis mellifera*

Romée van der Zee^{1*}, Alison Gray², Céline Holzmann³, Lennard Pisa¹, Robert Brodschneider⁴, Róbert Chlebo⁵, Mary F Coffey⁶, Aykut Kence⁷, Preben Kristiansen⁸, Franco Mutinelli⁹, Bach Kim Nguyen¹⁰, Adjlane Noureddine¹¹, Magnus Peterson¹², Victoria Soroker¹³, Grażyna Topolska¹⁴, Flemming Vejsnæs¹⁵ and Selwyn Wilkins¹⁶

¹Netherlands Centre for Bee Research, Durk Dijkstrastr. 10, 9014 CC, Tersoal, Netherlands.

²University of Strathclyde, Department of Mathematics and Statistics, Glasgow, G1 1XH, UK.

³Formerly at Institut Technique et Scientifique de l' Apiculture et de la Pollinisation, Institut de l' abeille, 75595 Paris, France.

⁴Department of Zoology, Karl-Franzens University Graz, Universitätsplatz 2, A-8010 Graz, Austria.

⁵Slovak University of Agriculture, Department of Poultry Science and Small Animals Husbandry, Tr. A. Hlinku 2, 94976 Nitra, Slovakia.

⁶Department of Life Sciences, University of Limerick, Limerick, Ireland.

⁷Middle East Technical University, Department of Biology, 06800, Ankara, Turkey.

⁸Swedish Beekeepers Association, Trumpetarev 5, SE-59019 Mantorp, Sweden.

⁹Istituto Zooprofilattico Sperimentale delle Venezie, National Reference Laboratory for Beekeeping, Viale dell'Università 10, 35020 Legnaro (PD), Italy.

¹⁰University of Liege, Gembloux Agro-Bio Tech, Department of Functional and Evolutionary Entomology, B-5030 Gembloux, Belgium.

¹¹Departement de Biologie, University of Boumerdes, Algeria.

¹²Scottish Beekeepers Association, 20, Lennox Road, Edinburgh, EH5 3JW, UK.

¹³Agricultural Research Organization The Volcani Center, 50250 PO Box 6, Bet Dagan, Israel.

¹⁴Warsaw University of Life Sciences, Faculty of Veterinary Medicine, Ciszewskiego 8, 02-786 Warsaw, Poland.

¹⁵Danish Beekeepers Association, Fulbyvej, DK-4140 Sorø, Denmark.

¹⁶National Bee Unit, Food and Environment Research Agency, Sand Hutton, York, YO41 1LZ, UK.

Received 15th June 2012, accepted subject to revision 18 December 2012, accepted for publication 3 June 2013.

*Corresponding author: Email: romee.van.der.zee@beemonitoring.org

Summary

This chapter addresses survey methodology and questionnaire design for the collection of data pertaining to estimation of honey bee colony loss rates and identification of risk factors for colony loss. Sources of error in surveys are described. Advantages and disadvantages of different random and non-random sampling strategies and different modes of data collection are presented to enable the researcher to make an informed choice. We discuss survey and questionnaire methodology in some detail, for the purpose of raising awareness of issues to be considered during the survey design stage in order to minimise error and bias in the results. Aspects of survey design are illustrated using surveys in Scotland. Part of a standardized questionnaire is given as a further example, developed by the COLOSS working group for Monitoring and Diagnosis. Approaches to data analysis are described, focussing on estimation of loss rates. Dutch monitoring data from 2012 were used for an example of a statistical analysis with the public domain R software. We demonstrate the estimation of the overall proportion of losses and corresponding confidence interval using a quasi-binomial model to account for extra-binomial variation. We also illustrate generalized linear model fitting when incorporating a single risk factor, and derivation of relevant confidence intervals.

Métodos estándar de encuestas para la estimación de la pérdida de colonias y los factores de riesgo que los explican en *Apis mellifera*

Resumen

Este capítulo trata sobre la metodología de encuestas y el diseño del cuestionario para la recogida de datos relativos a la estimación de las tasas de pérdida de colonias de abejas de la miel y la identificación de los factores de riesgo de la pérdida de colonias. Se describen las fuentes de error en las encuestas. Se presentan las ventajas y desventajas de las diferentes estrategias de muestreo aleatorio y no aleatorio y diferentes modos de recogida de datos que permitan al investigador tomar una decisión informada. Discutimos sobre la metodología de las encuestas y los cuestionarios con cierto detalle, con el propósito de dar a conocer las cuestiones a tener en cuenta durante la fase de diseño de la encuesta con el fin de minimizar el error y el sesgo en los resultados. Se ilustran aspectos de la encuesta que a través de encuestas realizadas en Escocia. Se da como ejemplo parte de un cuestionario estandarizado, desarrollado por el grupo de trabajo COLOSS de Monitoreo y Diagnóstico. Se describen enfoques para el análisis de datos, centrándose en la estimación de las tasas de pérdida. Se utilizaron datos de un monitoreo holandés de 2012 como ejemplo de análisis estadístico con el software de dominio público R. Demostramos la estimación de la proporción total de las pérdidas y el intervalo de confianza correspondiente usando un modelo cuasi-binomial para dar cuenta de la variación extra-binomial. También ilustramos ajustes del modelo lineal generalizado al incorporar un solo factor de riesgo, y la derivación de los intervalos de confianza correspondientes.

西方蜜蜂估测蜂群损失和风险因子的标准调查方法

摘要

本章针对为估测蜂群损失率和鉴定蜂群损失相关风险因子采集数据时所需的调查方法和问卷设计。描述了调查中的误差来源，介绍了不同的随机或非随机取样方法和不同数据采集模式的利弊，以供研究者做出合理选择。为提高对于在调查设计阶段需要考虑的问题的认识，使结果中的误差和偏差最小化，我们在一些细节上进一步讨论了调查和问卷的方法学。调查设计部分以在苏格兰的调查为例说明，并以由 COLOSS 工作组开发的用于监测和诊断的标准化问卷的一部分作为另一个例子进一步说明。描述了针对估测损失率的数据分析方法。以荷兰 2012 年的监测数据为例，使用公共领域 R 软件，介绍统计分析方法。我们使用类似二项式模型解释额外二项式变化，示范了总体损失比例及相应置信区间的估算方法。同时，我们举例说明了当合并单一风险因素时广义线性模型的拟合，以及相关置信区间的推导

Keywords: COLOSS *BEEBOOK*, estimating colony losses, *Apis mellifera*, surveys, questionnaire, random / randomized sampling, non-random sampling, generalized linear models (GZLMs)

Table of Contents

| | | Page No. | | Page No. | |
|-----------|---|-----------|------------|--|-----------|
| 1. | Introduction | 3 | 6.1.4. | Non-random methods | 14 |
| 2. | Objectives and case definitions | 5 | 6.2. | Need for and use of a sampling frame in random sampling | 15 |
| 2.1. | Objective of epidemiological studies on honey bee colony losses | 5 | 6.3. | Availability of a sampling frame | 15 |
| 2.2. | Application of definitions associated with honey bee colony losses | 6 | 6.4. | Sources of sampling frames appropriate for different target populations | 15 |
| 2.3. | Case definitions | 6 | 7. | Questionnaire design | 16 |
| 3. | Data collection methods | 6 | 7.1. | Completeness of the questionnaire | 16 |
| 3.1. | Choosing the method of data collection | 6 | 7.2. | Appropriate designs for different sampling methods | 16 |
| 3.2. | Available data collection methods with advantages and disadvantages | 7 | 7.3. | Common problems to avoid in questionnaire design | 16 |
| 3.2.1. | Surveyor administered questionnaires | 7 | 7.3.1. | Ambiguity of interpretation | 17 |
| 3.2.1.1. | Face-to-face interviews | 7 | 7.3.2. | Loaded questions | 17 |
| 3.2.1.2. | Telephone interviews | 7 | 7.3.3. | Questions on sensitive issues | 17 |
| 3.2.2. | Self-Administered questionnaires | 7 | 7.4. | Questionnaire design for minimisation of measurement error and ease of analysis. | 17 |
| 3.2.2.1. | Postal or email survey | 7 | 7.5. | Need to limit data sought, for a high response rate and accurate measurement | 18 |
| 3.2.2.2. | Internet survey | 7 | 7.6. | Problems of multi-lingual/multi-cultural questionnaires | 18 |
| 3.2.2.3. | Questionnaire published in beekeeping journals | 8 | 7.7. | Testing survey questions: importance of pilot studies | 19 |
| 3.2.2.4. | Questionnaires disseminated during meetings | 8 | 7.8. | Example of a standardized questionnaire on colony losses | 19 |
| 3.3. | Data validity and accuracy | 8 | 8. | Response rates | 19 |
| 4. | Quality issues in surveys | 8 | 8.1. | Use of incentives and reminders to improve response rates | 19 |
| 4.1. | Errors | 8 | 9. | Choice of sample size | 20 |
| 4.1.1. | Coverage error | 8 | 10. | Analysis of survey data | 22 |
| 4.1.2. | Sampling error | 9 | 10.1. | Assessing data quality | 22 |
| 4.1.3. | Measurement error | 9 | 10.1.1. | Dealing with missing data | 22 |
| 4.1.4. | Non-response error | 9 | 10.2. | The use of weighting in statistical analysis | 23 |
| 4.1.5. | Errors caused by selection bias | 9 | 10.3. | Elementary analysis | 23 |
| 4.1.6. | Processing errors | 9 | 10.3.1. | Descriptive analysis | 23 |
| 4.2. | Effort and costs in data accumulation | 10 | 10.3.2. | Loss calculations and confidence intervals | 24 |
| 4.3. | Issues of anonymity and ethical approval | 10 | 10.3.3. | Loss rate per factor including stratification on the operation size | 25 |
| 5. | Coverage | 11 | 10.4. | Advanced analysis, identification of risk factors by logistic regression | 25 |
| 5.1. | Effective coverage | 11 | 10.4.1. | Logistic regression | 25 |
| 5.2. | Potential bias of the survey sample | 11 | 10.4.2. | Dispersion in statistical models | 26 |
| 5.3. | Identifying the target population | 12 | 10.4.3. | Multilevel analysis | 26 |
| 6. | Sampling | 12 | 10.4.4. | Software for logistic models | 26 |
| 6.1. | Random and non-random sample selection methods | 12 | 10.4.5. | Example of advanced analysis | 27 |
| 6.1.1. | A census | 13 | 11. | Conclusions | 33 |
| 6.1.2. | Random sampling | 13 | 12. | Acknowledgements | 33 |
| 6.1.3. | Recommended approach for random sampling | 14 | 13. | References | 33 |
| | | | | Online Supplementary Material. Useful resources for survey design | 35 |

1. Introduction

Surveys on honey bee colony losses have been conducted by many researchers over the years to understand the factors that contribute to colony losses. Recognizing the importance of standard

questionnaires for use in surveys, a network of honey bee specialists preceding the establishment of the COLOSS Action network, initiated by Cost Action FA0803, established at its first meeting a working group (Working Group 1 - WG1) whose aim was to develop and implement research surveys for the purpose of identifying such

factors. The working group currently represents a global network of scientists who monitor colony losses. This group was conscious of the fragility of many survey results and addressed crucial issues to obtain a valid research framework (Van der Zee *et al.*, 2012). Using other literature sources, the group developed and/or recognized appropriate case definitions, statistics and relevant factors associated with honey bee colony losses. The present manuscript aims to make the results of these efforts available to all researchers working in this field and to provide guidelines for conducting effective surveys.

Conditions in which to perform surveys on honey bee colony losses and achieve results which meet methodological standards are very different between and within countries. The present chapter offers guidelines to attain good quality surveys, even under unfavourable conditions. The main objective of these surveys (section 2.1.) is the estimation of winter colony loss rates, identification of specific areas with a higher or lower risk of honey bee mortality and information on possible determinants such as the control of *Varroa destructor*. This will enable the provision of advice on loss prevention and control.

We are conscious that the case definitions we present (section 2.3.) may be refined or changed in the following years because not enough knowledge is yet available to resolve many important issues. However what we present here does, in our view, give a good set of standards to which all researchers in this field should aim to conform in order to produce robust and reliable results.

The target population of the surveys is usually the set of active beekeepers in a country or specific area. The possibilities for reaching the target population vary between and within countries. Sometimes registers of beekeepers can be used for collecting data; but more often, cooperation with beekeepers' associations is necessary (section 5.). In some situations, both are absent and the investigator has to develop other survey strategies. Suggestions are given for sampling

frames in situations where cooperation with a beekeeper association is not possible or if a beekeeper infrastructure is absent (section 6.4.).

The sample selection method used is one of the main issues for obtaining reliable survey outcomes. Selecting a random sample of beekeepers gives results whose accuracy can be quantified (section 6.1.), if, as is usually the case, studying the whole target population instead of a sample is not feasible. At present most monitoring surveys with questionnaires will gain in quality if the shift is made from the present common practices of self-selected samples (samples in which participants volunteer to take part) towards at least simple random sampling. The same or better survey results may be achieved by using other more sophisticated forms of probability sampling, and relatively small sample sizes might then be sufficient.

Detailed consideration will also be given to the various sources of bias which may affect survey outcomes (De Leeuw *et al.*, 2008), whose effect will usually be to introduce errors into results whose effects are difficult or impossible to assess. In particular, it is good practice to strive for high response rates, although there is no empirical support for the notion that low response rates necessarily produce estimates with high nonresponse bias (Groves, 2006). However that risk is inevitably present if response rates are low.

Attention is given here to a variety of methods of statistical data analysis. These range from simple analyses to examine the effects of different *Varroa* controls or other individual risk factors on mortality, to more advanced methods involving the use of Generalized Linear Models (GZLMs) to investigate simultaneously the possible effects of multiple different factors on colony loss rates. We use the statistical program R to illustrate an analysis in section 10, using data from the Netherlands. Survey design and sampling methodology is illustrated throughout the manuscript using Scotland as a case study. An introduction to this is provided in Box 1.

Box 1. Introduction to the example of surveys of beekeepers in Scotland.

Throughout this chapter we illustrate some of the methodology and concepts using as an example surveys of beekeepers in Scotland. These surveys are not perfect in the way they have been conducted, but most of them have used random sampling and we describe how this was achieved. They provide a case study, for those who may be interested, of how random sampling can be done in a situation where reasonably good records of beekeepers are available for use as a sampling frame. It is recognised that this approach is not possible in all situations. We describe the sampling design, the use of records for sample selection, obtaining permission for use of records, maintaining anonymity etc.

These surveys have run since 2006. They have all used as the survey population members of the Scottish Beekeepers' Association (SBA), the national body for beekeepers in Scotland. Persons keeping any number of honey bee colonies, or with no bees but having an interest in bees, can belong to this organisation. Affiliated with the SBA are a large number of local associations for beekeepers. It is possible to belong to one or more local associations as well as, or instead of, the SBA.

The first survey used a quota sampling approach, as permission had not at that time been sought for using the membership records of the SBA for sampling purposes (see section 6. and box 5), and made use of SBA representatives to identify beekeepers to include in the sample. Subsequent surveys have used a stratified random sampling approach, by dividing the membership into those belonging to each of several large administrative areas and taking a separate random sample from each of those identified groups of SBA members using the membership records (see section 6.).

As these administrative areas are geographical, this approach was chosen to try to ensure coverage of the different geographical areas, since conditions for beekeeping vary across the country. In particular, the more remote parts of the north and west of Scotland are thought to be free of *Varroa destructor*. Weather patterns also vary geographically.

All of the surveys from 2006 to 2012 have been conducted using a postal questionnaire.

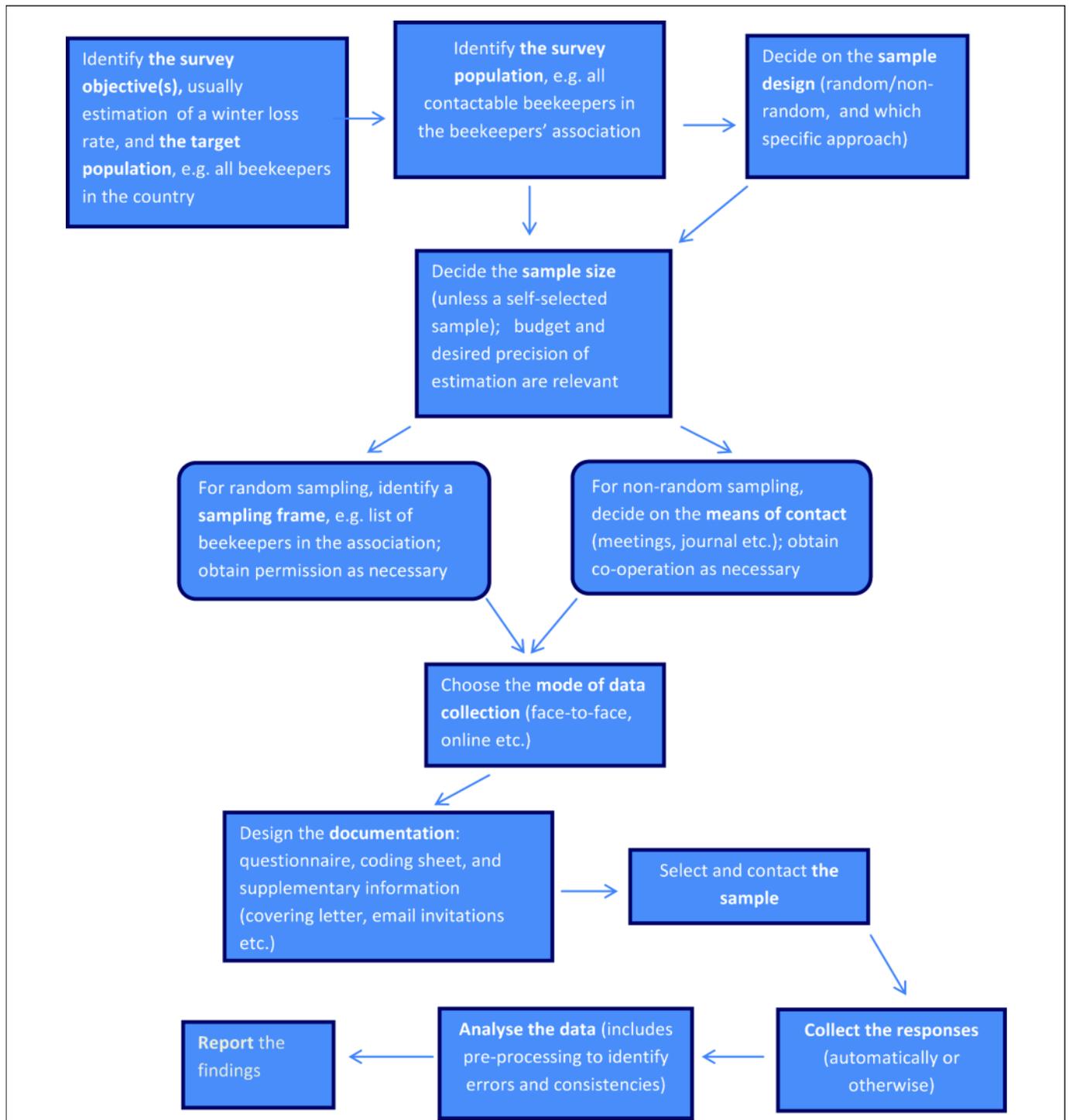


Fig. 1. A basic flowchart of the key steps in carrying out a survey.

2. Objectives and case definitions

Fig. 1 shows the steps to be addressed in designing a survey. We address each of these in the sections below.

2.1. Objective of epidemiological studies on honey bee colony losses

The epidemiological study of honey bee colony losses aims to determine explanatory factors for, and to monitor the magnitude and

the spatial distribution of honey bee colony losses at the operation, apiary or colony level. This enables the formulation of good advice on prevention and control of colony stressors. In this section, we propose standardized case definitions to facilitate various objectives as follows:

1. Reporting and classification of cases of honey bee colony losses by national and international honey bee experts.
2. Standardization of language for communication purposes.
3. Comparability of data across time and geographical areas.

2.2. Application of definitions associated with honey bee colony loss

1. The case definitions for use in surveillance are based on available epidemiological data summarizing what is currently known about the magnitude and spatial distribution of honey bee colony losses. Countries may need to adapt case definitions depending on their own situation.
2. The case definitions have been developed to help national authorities classify and track cases.
3. The case definitions are not intended to provide complete descriptions of the symptomatology of lost colonies but rather to standardize reporting of these losses.
4. The case definitions will describe the symptoms of dwindling and lost colonies and the timeframe of observation during which honey bee colony losses occurred.
 - b. a noticeable lack of dead worker bees both within and surrounding the hive (vanEngelsdorp *et al.*, 2009);
 - c. the delayed invasion of hive pests (e.g., small hive beetles (Neumann *et al.*, 2013) and wax moths (Ellis *et al.*, 2013)) into affected colonies and kleptoparasitism of affected colonies by neighbouring colonies (Cox-Foster *et al.*, 2007);
 - d. the absence of *Varroa* and *Nosema* at levels thought to cause economic damage (vanEngelsdorp *et al.*, 2009).
5. Time frames during which honey bee colony losses occurred can be distinguished as:
 - a. Time frames related to seasonal characteristics:
For example winter: the period between the moment that a beekeeper finished pre-winter preparations for his/her honey bee colonies and the start of the new foraging season.
 - b. Fixed time frames:
For example: observations every half year.

2.3. Case definitions

1. Lost honey bee colony is a honey bee colony that:
 - a. is reduced to such a small number of bees that it cannot perform the normal biological activities needed for survival (brood rearing, resource gathering) or
 - b. has queen problems, such as drone laying queens or drone laying worker bees in absence of a queen, which could not be solved or
 - c. was missing due to burglary, or didn't survive fire, inundation, desert storms or similar causes unrelated to health problems.
 - d. no longer has any living bees present.
2. Weak honey bee colony:

A honey bee colony that is not considered as lost, but in which the number of bees is less than would be expected from the colony size observed at an earlier inspection.

3. Colony Depopulation Syndrome (CDS):

This is observed if a honey bee colony shows the following conditions within a certain time-frame:

- a. reduced to no, or only a few remaining, living bees in the hive and
- b. no, or only a few dead bees in or in front of the hive or at the apiary while
- c. food is present in the hive (Van der Zee *et al.*, 2012).
4. Colony Collapse Disorder (CCD)

This is observed if the following conditions are present:

- a. a rapid loss of adult worker bees from affected honey bee colonies, as evidenced by weak or dead colonies with excess brood populations present relative to adult bee populations (vanEngelsdorp *et al.*, 2009);

It is difficult to come to conclusions on losses with a fixed timeframe approach since the outcome depends on beekeeper practices such as merging, splitting, buying and selling of colonies. Recalling the numbers of colonies involved in these practices later when a questionnaire is disseminated may easily lead to errors in the data (Van der Zee *et al.*, 2012). Another problem is that no information is collected on when these increases/reductions were made within the timeframe, with the effect that colonies bought at the start of a time frame have the same weight in the risk estimation as colonies bought later on. Not recognising these problems may severely bias the outcome.

3. Data collection methods

3.1. Choosing the method of data collection

When choosing the mode of data collection, a number of issues must be considered: effective coverage of target populations, data accuracy, potential bias of the survey sample, the survey mode(s), and the effort and cost of data collection (De Leeuw, 2008; Charrière and Neumann, 2010; Dahle, 2010; Hatjina *et al.*, 2010; Mutinelli *et al.*, 2010; Nguyen *et al.*, 2010). In a survey, the data can be collected by direct contact of an interviewer with the respondent (face-to-face interviews, telephone interviews) or the questions can be administered and answered by beekeepers without the assistance of an interviewer by means of a self-administered questionnaire (postal surveys, email surveys, internet surveys). The advantages and disadvantages of each mode are described in section 3.2. Further references are given in the online supplementary material.

3.2. Available data collection methods with advantages and disadvantages

3.2.1. Surveyor administered questionnaires

3.2.1.1. Face-to-face interviews

Pros:

- The interviewer can explain the importance of a survey and clarify questions if needed.
- The data can often be entered directly into the computer database and checked for a valid data entry, question by question.
- Answers can easily be corrected *in situ*. It may become apparent immediately that some answers are inconsistent or wrong, especially if suitable data checks are built into a computer questionnaire programme.

Cons:

- A representative list of beekeepers is needed; there are few countries in which it would be available through beekeeping associations, census registers or the veterinary services.
- The presence of an interviewer may influence the answers, unless the interviewer is well-trained.
- This method is time-consuming and costly because of: travel costs (unless some form of cluster sampling is used), the need for many highly trained interviewers, and the need for multiple call-backs to ensure a high response rate.

3.2.1.2. Telephone interviews

Pros:

- The interviewer can explain the importance of a survey and clarify the questions if needed.
- Beekeepers may feel obliged to participate in the survey, though others may simply say they do not have time to participate. The data can often be entered directly into the computer database, as above.
- Lower costs than face-to-face interviews and call-backs are much faster and easier.
- Many interviews can be completed in a relatively short time.

Cons:

- A representative list of telephone numbers of the target population of beekeepers is needed; there are very few countries where it would be available.
- Beekeepers under pressure may give the answer without careful consideration, and it is difficult to correct such answers later.
- Time necessary for some interviews may be longer than is necessary to answer the questions because some people tend to be garrulous, although other people will become impatient if the time required is long.

3.2.2. Self-Administered questionnaires

3.2.2.1. Postal or email survey

Beekeepers receive a questionnaire by mail (or email), answer the questions and return the questionnaire.

Pros:

- Beekeepers have time to check their apiary notes and to answer the questions fully.
- Quick distribution of questionnaires for survey organisers and quick return (if emailed).

Cons:

- A good (complete) list of addresses (email addresses) of the target population of beekeepers is needed. Limited access of beekeepers to the internet can very badly influence the survey coverage if only an email survey is performed.
- Especially clear questions and instructions are necessary.
- Beekeepers are not always actively involved and very often they do not respond. Free return postage raises the costs for postal surveys but also beekeepers' participation in the survey.
- Reminders are likely to be necessary for a good response rate.
- The questionnaires may be filled in carelessly with answers missing.
- The time for and cost of data entry can be high, unless the sample is very small.

3.2.2.2. Internet survey

Beekeepers complete a questionnaire hosted on the internet. The same questionnaire could be hosted or linked on different websites, e.g. research institutes, reference laboratories, beekeepers' associations or beekeeping journals. There are several approaches to the calculation of response rates for internet surveys (AAPOR, 2011). The response rate for an online survey is often comparable with the response to a questionnaire published in a beekeeper journal, if the invitation to participate in an online survey is published on the website of a beekeepers' association.

Pros:

- Large numbers of completed questionnaires can be collected in a very short time.
- There is no need for transferring data from paper questionnaires to an e-database.
- Low cost of data collection and reduced cost of data analysis.
- Beekeepers' associations or beekeeping journals could contribute to increasing the number of filled in questionnaires by advertising the survey.
- Beekeepers have enough time to check their apiary notes and to answer the questions fully.

Cons:

- A list of email addresses of the target population of beekeepers is needed to advise them of the survey or issue reminders.
- Beekeepers with ready computer access may not be wholly representative of the general beekeeping population of interest, which can adversely affect the survey coverage.
- In principle, a beekeeper could complete the survey more than once; however, there is a choice of a suitable survey software that makes it possible to prevent duplicate submissions.

3.2.2.3. Questionnaire published in beekeeping journals

Completed questionnaires are usually posted, faxed or emailed by beekeepers.

Pros:

- If a journal has a large circulation, the dissemination of a questionnaire is widespread.
- Cost of questionnaire dissemination is low.
- Cost of data collection is low. Free return shipment raises the costs but also beekeepers' participation in the survey.
- Beekeepers have enough time to check their apiary notes and to answer the questions fully.
- Non-response can be estimated, if all beekeepers receive the journal for free because they are a member of a beekeeper association.

Cons:

- Compared to disseminating copies of a questionnaire at beekeepers' association meetings and encouragement to respond by leaders of those associations, participation rates are lower.
- The survey will not cover non-readers of these journals, so its representativeness may be limited to the readers of the journal.
- The questionnaires may be filled in carelessly with answers missing.
- The time for and cost of data entry can be high.

3.2.2.4. Questionnaires disseminated during meetings

Completed questionnaires can be collected immediately by the survey organiser or an association representative, or posted, faxed or emailed by beekeepers.

Pros:

- An easy way to disseminate the questionnaires if co-operation with the beekeeping association hosting the meeting is good.
- The survey organiser can explain the importance of the survey and clarify the questions if needed.
- If beekeepers post the filled-in questionnaires by themselves they have enough time to check their apiary notes and to answer the questions fully.

Cons:

- Not all beekeepers attend beekeepers' association meetings, leading to coverage problems.
- The survey will cover beekeepers only from specific regions or associations.
- The questionnaires are often filled in carelessly with many answers missing.
- The time for and cost of data entry can be high.

3.3. Data validity and accuracy

The use of unambiguous questions is critical. However, the clarity of any international questionnaire may well be culturally dependent. In the absence of a face-to-face or telephone interviewer to conduct the survey, the questions could be misunderstood by a respondent though this may not be immediately obvious. This would reduce the validity or accuracy of the response to the question asked. In self-administered surveys, the respondent is the locus of control and can spend as much time as s/he wants to consult records to answer detailed questions. Especially in telephone interviews, but also face-to-face, one may feel pressured to answer and not let the interviewer wait and give an estimate instead of looking up the correct answer. Multi-stage manual data entry, such as is often involved in the collection of data in electronic form from paper questionnaires, completed by individual beekeepers, and then read and entered by another individual later, is error-prone and needs careful checking.

4. Quality issues in surveys**4.1. Errors**

Any survey will be vulnerable to errors which may invalidate the extrapolation of the sample results to the target population. The most important sources of error are discussed in detail below. The aim in each investigation must be to minimise the non-sampling errors, and to quantify as far as possible the (unavoidable) sampling error. Therefore we describe these various sources of error in detail, as they should be borne in mind at the planning stage of the survey.

4.1.1. Coverage error

Coverage error arises when the survey population listed in the "sampling frame" – the list from which the sample selection is made (see section 6.2., usually of beekeepers in a particular country or region) – does not match well with the target population (the population about whom an inference is to be made; usually the set of all beekeepers in that country or region). The results of the survey will be seriously affected if those omitted from the sampling frame differ in some respect relevant to the aims of the investigation (in size of enterprise, for example, so that perhaps large commercial beekeepers are not represented).

4.1.2. Sampling error

Sampling error is the error that occurs because a sample is taken instead of examining the whole population (Lohr, 2010). A survey allows estimation of characteristics of variables (typically a population mean, sum, or proportion) concerning a whole population, on the basis of a sample. Such an estimate is inevitably not the exact value of the population quantity. The only way to obtain the exact value is to calculate it from the whole population, but this is rarely possible.

If the sample has been randomly selected from the population, this error can be quantified by calculating the standard error (standard deviation) of the estimate. This is an estimate of the variation of the estimator used between different samples of the same size selected from the same population. When a non-random sample is used, there is no appropriate analytical form for the standard error (see any elementary textbook on survey sampling, e.g. Schaeffer *et al.*, 1990) and therefore the results of any such calculation should be viewed with caution.

This variation from sample to sample is usually presented by quoting a confidence interval for the estimate obtained from the sample. A confidence interval can only be reported if the sample is representative for the population of interest. An example of the calculation is given in Box 2.

The sampling error can be reduced by increasing the size of the sample. It is possible to calculate the optimum size of the sample for a satisfactory estimate at a given cost, depending on the chosen confidence level or required precision or margin of error. Having some approximate knowledge of the population quantities is needed for such a calculation of the sample size (see section 9.).

4.1.3. Measurement errors

Measurement error occurs when the recorded answer to a question deviates from the true answer. The risk of such error is increased by an imprecise or wrongly formulated question. These errors can also occur in face to face surveys, if the interviewer influences the respondent (interviewer bias). Moreover, a survey is always declarative, so the respondent can voluntarily give a wrong answer if the question is in some way sensitive.

4.1.4. Non-Response errors

Non-response can be complete non-response (if a sampling unit, i.e. a beekeeper, did not answer at all) or partial non-response (some questions are not answered or only partially answered).

Keeping questions clear and simple can help to reduce the chance of missing data. Emphasising the importance of the survey and explaining how the results will be used may increase the response rate. If the participant can appreciate that there is some benefit to completing the survey, they will be more likely to take part. Use of rewards and incentives can be useful to increase participation of people selected already to take part in the survey.

Non-response bias occurs when there is something systematically different about participants who do not respond from those who do. Trying to minimise non-response is therefore very important. Reminders are useful in this regard. Non-response can in principle be estimated by randomly sampling some sample units after termination of the survey, approaching the non-respondents with another survey mode and comparing the main outcome with the survey response on these sample units. However such an effort is rarely regarded as a good use of scarce resources of time and money.

4.1.5. Errors caused by selection bias

Selection bias should also be considered as a source of error. Selecting an unrepresentative sample will lead to bias in the results. Actively selecting a random sample rather than a self-selecting sample who may well differ from those not included in the survey (volunteer bias), is the best way to avoid this. Using well-trained personnel will avoid any haphazard substitution of properly selected participants with more conveniently available participants who had not been selected, which is another possible source of selection bias.

4.1.6. Processing errors

These errors affect the data set. They can arise by errors caused by the person who records the data (mistyping, copy/paste errors, stretching cells in an Excel spread sheet etc.). If different people capture the data, harmonisation of notation and careful procedures for recording data should be in place, checks should be made, and personnel should be well-trained and informed to avoid introducing errors and biases (Schaeffer *et al.*, 1990).

Unlike the sampling error, the non-sampling errors are very difficult, if not impossible, to measure, and cannot be reduced by increasing the sample size. In a large sample, non-sampling errors are the more important source of errors, as the sampling error is reduced. The only way of controlling non-sampling error is to know what sorts of errors are possible, and to be very careful to avoid these as much as possible in the conduct of the survey.

Box 2. Example of confidence interval.

The overall proportion of colonies lost from those at risk is estimated as 19.5%, with a standard error (s.e.) of 1.5%. The corresponding 95% confidence interval is obtained in the usual way as the estimate ± 1.96 (s.e.) giving 19.5% ± 1.96 (1.5)%, which yields 16.6% to 22.4%. This means that in about 95% of cases when such a calculation is made, using a sample of the same size from the same population, the interval quoted will contain the true value of the overall proportion of colonies lost. This calculation assumes that the sample estimate is approximately normally distributed.

4.2. Effort and costs in data accumulation

Manual data entry is costly and time consuming. Although Optical Character Recognition systems for automatic data entry are available, the fastest mode of data collection and accumulation is when beekeepers directly answer the questionnaire using an internet database. Furthermore, after the end of the data collecting period, the respondent can receive feedback and evaluate his/her losses or other aspects of beekeeping experience relative to data accumulated from other participants in the survey. Such systems can encourage participation by other respondents and so achieve a higher response, and are available nowadays for affordable prices. However this approach will fail to achieve a representative sample and may be a source of selection bias if the availability of internet access is associated with questions of underlying interest.

As none of the discussed survey methods is flawless, nowadays survey organisers often use a combination of data collection modes (a mixed-mode survey) to offset the weaknesses of one mode with the strength of another (Brodtschneider *et al.*, 2010; Topolska *et al.*, 2010; Soroker *et al.*, 2010, van der Zee *et al.*, 2012). Data validity and accuracy can be improved by interviewer-administered questionnaires of a selected group of beekeepers, ideally randomly selected, by following up a postal or email survey by a telephone interview or offering the opportunity to clarify any points of difficulty. Such a limited follow-up can also sometimes reveal the kinds of bias incurred by the more extensive survey. However, unless follow-up is so limited as to produce little information, it is a very costly option.

Repeating annual surveys among the same group of beekeepers will provide information on time trends, either by simply sampling the same population or possibly by following the same sample of beekeepers through time (i.e. using a panel design), when this is feasible. Use of panel surveys does require some replacement of panel of members who are no longer available to participate in the

survey, while trying to keep the sample representative. Comparing the results of self-administered surveys which are widely distributed and interviewer-administered surveys of a selected group of beekeepers will enable evaluation of the extent of colony losses area/countrywide and indicate the reliability or otherwise of a non-randomly selected sample. It may also identify any special cases that require further study, such as extreme losses in specific geographical areas that were overlooked by random sampling and in the event that identification of such areas is the purpose of the study.

4.3. Issues of anonymity and ethical approval

Considering the issue of anonymity versus confidentiality is important before a questionnaire is disseminated. In an 'anonymous' type questionnaire, the subject is totally unknown to the survey organiser, while in a 'confidential questionnaire' all the data is known to the survey organiser, but kept confidential. Box 3 gives an example of how anonymity can be preserved when postal surveying is used, using experience from surveys in Scotland.

With the increase in use of email / webmail, using these means of communication make it virtually impossible to guarantee total anonymity, as the respondent's name – or at least the email address – is automatically included in their reply, although satisfactory survey software packages include the option of suppressing from the recorded responses all means of identifying individual respondents. While the perceived possibility of lack of anonymity may raise levels of non-response or compromise the validity of responses to any sensitive questions in an email questionnaire, the ease of access to a worldwide population of beekeepers, the low administration costs and its unobtrusiveness to respondents generally outweigh this negative effect. It is also a simple matter to issue reminders by email.

However, it is important that the level of confidentiality of the questionnaire is clearly outlined to participants. Hence, the covering

Box 3. Example: Preserving anonymity in a postal survey in Scotland.

The membership records of the Scottish Beekeepers' Association (SBA) provide a well-organised sampling frame of the target population. The help of the SBA Membership Convener was obtained. He is the only person with full access to those records. He was asked not to supply the survey organisers with the full records (which he would not have been permitted to do in any case), but only to supply the list of "Short Reference Numbers", each of which uniquely identifies one of the members, along with the associated postal code. Before supplying that list, he was asked to remove from the list those ineligible to participate in the survey, including for example members not resident in Scotland, institutional members (such as libraries), and those who had declared themselves unwilling to participate in surveys. (The opportunity to opt out of surveys is available to new members when joining the SBA, and the opportunity was given to all existing members of the SBA to opt out via a short article published in the SBA's regular publication for members, prior to the first survey using the SBA records for sample selection).

The postal codes were abbreviated in the list supplied, so that while preserving the broad geographical location of each potential survey participant, it was not possible to identify any particular address. The postal codes were used to assign each member on the list to a particular geographical region in Scotland, so that the sample selected could be stratified on a geographical basis. This was done by dividing the country into a number of regions related to the administrative areas used by the SBA, in order both to give greater precision in estimation and to ensure greater geographical coverage in sampling and therefore hopefully a more representative sample.

Then a stratified sample of the agreed size was selected from the list, using a sampling function available in the R software for objective random selection. Each questionnaire sent out was put into an envelope on which was written a questionnaire number provided by the survey organisers, also written on the questionnaire in the envelope. These envelopes were then sent to the Membership Convener along with a key file linking the questionnaire numbers to the corresponding "Short Reference Numbers". This enabled the Membership Convener to print the appropriate address on each envelope and to mail out the questionnaires, without the organisers knowing the identity of the selected members.

In fact, the majority of participants in the Scottish surveys have willingly provided personal contact details as part of their questionnaire return. (For example, in the survey in 2011, 85% out of 94 respondents did so; there was a 47% response rate).

letter with the original questionnaire should clearly state that a reminder will be forthcoming if no response is received. The availability of this option to issue reminders is important since research indicates that reminders increase the response rate (Campbell and Waters, 1990; see also the Scottish example below on reminders and incentives). Furthermore, email software allows the dispatcher of the questionnaire the option of notification when the recipient has opened the message.

The use of questionnaires raises the question of personal liberty and ethics and, because of this, many research institutes/universities require the survey organiser to acquire ethical approval prior to disseminating questionnaires. Since questionnaires related to honey bee research are primarily concerned with generic rather than personal information, it may be possible to acquire multi-annual ethical approval in advance, thus allowing the annual dissemination of the questionnaire. However, this will be specific to different institutes and thus clarification on the ethical requirements should be sought locally before a questionnaire is disseminated.

5. Coverage

5.1. Effective coverage

Collecting representative data on the extent of colony losses in any area or country depends on one's ability to identify and reach the target population (beekeepers in the country, commercial and/or non-commercial). This ability is affected by factors such as:

- the size of the country and beekeeper community
- the means of contacting the beekeeper community
- the degree of affiliation of beekeepers with beekeeping associations
- availability of professional magazines (including the possibility of publishing the questionnaire in a beekeeping journal)
- the holding of regular meetings of beekeepers
- the extent and accessibility of internet and telephone networks
- availability of addresses, e-mail addresses or telephone numbers
- willingness and ability of beekeepers' associations to cooperate in providing information
- the possibility of cooperation with beekeeping inspectors and veterinary services who may hold registers of beekeepers
- the number of staff engaged or available to conduct the survey and analyse the data
- and, of course, available time and funds.

Legal issues in relation to the preservation of the confidentiality of data also arise in some countries – for example, all organisations in the UK are constrained in what data they make available outside their own membership by the requirements of the Data Protection Act 1998.

If any method of probability sampling is to be used for a survey, access is also required to a sampling frame (see section 6.2.) which gives good coverage of the target population.

5.2. Potential bias of the survey sample

The responsiveness of the target population can be biased. For example, beekeepers suffering higher colony losses might be more likely to respond to surveys than those suffering fewer losses, although other biases are also possible. For example, Brodschneider *et al.* (2010) found a potential bias in reported colony losses from the same region collected by different media. Respondents who returned a postal questionnaire from a beekeeping journal reported higher proportions of losses than those responding online or at a convention in this particular region. This suggests that responses of some groups (such as from visitors of a convention) may not constitute a statistically representative sample, being from a different population than the target population. In such situations a mixed sampling approach must be considered. This enables comparison of the outcome of the different sampling methods, which should be reported in the final report. This could be due to different experiences of the target groups and hence a different level of motivation of the beekeeper to respond. A randomized sample may suffer from the same non-response problems, but maybe to a lesser extent because the approach is more focussed on the individual beekeeper. It may be possible to overcome this problem by increasing the response rate via encouragement of broad participation in the survey and the use of reminders. Using mixed media surveys or surveys using random surveying of the population of beekeepers to achieve a more representative sample may help, but the underlying problem remains that of the association of the response rate with the underlying questions of interest to the survey organisers. Response rates can be rather low. Dahle (2010) quotes a 15% response rate from surveys sent out in a beekeeping journal. Van der Zee (2010) quotes a 7.5% response rate from beekeepers who were invited in a national beekeeper journal to participate in an internet survey as well and they found differences between the results of these surveys and those from random surveys. In Denmark, response rates of up to 33% have been achieved (Vejsnæs *et al.*, 2010). In the Netherlands, an average of 22% of the beekeepers surveyed from 2006-2012 (van der Zee *et al.*, 2012) responded to a mixed mode approach of a questionnaire included in the 2 national beekeeper journals. The letters could be

Table 1. Some sampling methods with advantages and disadvantages.

| Method groups | Method | Explanation | Advantages | Disadvantages |
|---------------|--------------------------|--|---|--|
| Complete | Census | Whole population selected. | With good response rate should give excellent information. | Potentially expensive and often infeasible. If attempted without a sampling frame, may be misleading. |
| Random | Stratified | Population divided into groups thought to be relevant to the survey objective, and each group sampled independently at random. | Gives well-targeted information where size of errors can be properly estimated if response rate is good. Has the potential for lower sampling error than simple random sampling. | Infeasible without a sampling frame. |
| | Simple | One random sample chosen from whole population. | Size of errors can be properly estimated if response rate is good. | Infeasible without a sampling frame. Less well-targeted than a well-stratified sample. |
| | Cluster | Population split into clusters (convenient subgroups), and a census is conducted within each of a randomly selected set of clusters. | Size of errors can be properly estimated if response rate is good and clusters are well chosen. Only requires good sampling frames within clusters selected. Potentially cheaper than stratified or simple random sample. | Requires sampling frame of clusters, and of individuals within selected clusters only. If clusters selected are not representative then may introduce bias to results. Sampling error is higher than with simple or stratified sampling. |
| Non-random | Quota | "Quotas" are set for researchers to fill of respondents matching given criteria. | Requires no sampling frame. Can work well, if quotas are based on sound underlying information. | Can fail very badly in achieving representative results. It is sometimes hard to fill quotas, and non-response is disguised. |
| | Purposive or judgemental | Researchers aim to select "representative" respondents. | Requires no sampling frame, and aims to achieve a good mix of respondents. | Personal selection is notoriously bad at choosing a genuinely representative sample. |
| | Convenience | Respondents are chosen because they provide a conveniently available sample. | Cheap and easy to implement. Useful for pilot studies. Requires no sampling frame. | Extremely unreliable and impossible to assess accuracy of results. |

sent back without charges or through an email with a personalized link to the questionnaire on the internet.

5.3. Identifying the target population

A survey of beekeepers can be used to collect reliable data on beekeepers and beekeeping activity and/or practice in a certain area or territory of a country. The target population of such a survey should be defined according to the data that one aims to collect. It could be targeted to beekeepers' associations (local to national in scope) or the individual beekeeper. However, the target population is usually the set of all active beekeepers whose colonies are kept in the area of interest during the time period concerned. Consequently, coverage of individual beekeepers' operations via associations might be incomplete and variable. How easy it is to access such a population depends on whether beekeepers voluntarily or under legal compulsion are registered with some record-keeping organisation. Examples might be:

- A legally required register of all beekeepers within a given country;
- A voluntary register of beekeepers within a region to which most beekeepers subscribe;

- One or more regional or national beekeeping associations to which most beekeepers belong.

The coverage provided by such potential sources of data is clearly very variable, and the reliability of survey efforts will depend heavily on the adequacy of this coverage.

6. Sampling

6.1. Random and non-random sample selection methods

There are numerous sample selection methods for drawing the sample from the population, broadly classified into random or probability-based sampling schemes or survey design methods, and non-random or non-probability based sampling. An overview may be found for example in Schaeffer *et al.* (1990) or any such survey sampling textbook. A number of terms relating to surveys are introduced in the following sections, and may all also be found in such standard texts. The main methods which might be used for sampling beekeepers are summarised in Table 1. Some more detail is given about each method below. In general those methods higher up in the table will cost more, but will

give more reliable results, provided that a good response rate can be achieved.

6.1.1. A census

It is possible to approach most of the beekeeper population in smaller countries such as the Netherlands, where questionnaires are included in both of the two beekeeper journals, which are sent to all Dutch beekeepers who are by their membership of a local organization also a member of a national association (>90%). The questionnaire could be returned without postal costs. Beekeepers who provided an email address (>85%) in the past received a personalised link to the online questionnaire. In 2012, about 70% of the data was submitted this way and processed immediately, which reduced costs substantially. This approach is in fact not sampling, but addressing the total population of organized beekeepers.

6.1.2. Random sampling

Survey designs based on random sampling are designed to select sampling units from the population with known probabilities. This means that the sampling properties of estimators of population quantities can be determined, such as whether or not the estimator is unbiased (i.e., does it on average give the right answer?) and what is its precision (i.e., how do we calculate its variance or its standard error).

This is the objective scientific approach to sampling and the only one for which sampling properties of estimators are known. Other methods may provide good information but there is no guarantee that they will, and their sampling properties are unknown. However even with random sampling, if response rates are poor then the possibility of non-response bias will compromise the estimation.

Implementation of random sampling methods requires a mechanism for random selection, usually accomplished by use of random number generators in computer software, e.g. the "sample" function in the public domain software R (downloadable from <http://www.r-project.org/>). It also usually requires a sampling frame, or list of sampling units in the population (section 6.2).

The simplest scheme is *simple random sampling*, which samples randomly without replacement from the sampling frame so that at every stage every sampling unit not already selected from the sampling frame is equally likely to be chosen. This results in all samples of a given size being equally likely to be selected.

Systematic sampling is sometimes used as a simple alternative to simple random sampling and works at least as well in situations where the population sampled from is "randomly ordered" with respect to the value of a quantity being measured or recorded, or is ordered in order of size of such a quantity. It does not always require a sampling frame. For example, if 1000 beekeepers attend a convention, to achieve a 10% sample of those attending, a participant may be selected at random from the first 10 beekeepers to arrive or register, and then every 10th person after that also selected.

Stratified sampling splits the population into subgroups or *strata*, using stratification factors such as geographical area or degree of experience of the beekeeper, or beekeepers/bee farmers, which are judged to be important in terms of coverage of the population and which are likely to be related to the response variable(s) or interest. Then a random sample, in the simplest case a *simple random sample*, is selected separately from each stratum, using predetermined sample sizes. This ensures representation of all these important groups in the sample (which might not be achieved by a single simple random sample), and the random sampling should compensate for any other relevant stratification factors which may have been overlooked in the survey design. It also allows comparison of the responses from each stratum, provided enough responses are achieved in each stratum.

If the average responses do differ between the strata, and/or the variation in recorded responses differs between strata, stratified sampling should provide estimates with a lower variance than simple random sampling. The lower variance is achieved because separate samples have been taken from populations with smaller variation within them compared to the population as a whole (Schaeffer *et al.*, 1990).

One basis for stratified sampling is operation size. The scale of beekeeping operations and management practices are very different for hobbyist beekeepers and professional/commercial operators (bee farmers). Due to the potential for different numbers of lost colonies and consequences of losses among these two groups, both should be included whenever possible in a survey. This allows the colony loss rates experienced by both groups to be compared and it is more representative of overall levels of loss. Box 4 gives an example.

The migration of colonies (the movement of colonies to/from nectar flows or for purposes of crop pollination) differs widely between beekeeping operations. Therefore, it is also desirable to consider different classes of migratory practice where possible when designing and analysing the survey. As migration may be a factor in loss rates (although see VanEngelsdorp *et al.*, 2010), comparing migratory and non-migratory beekeepers is important, if the sample

Box 4. Example: Case study of stratified random sample selection.

In Scotland, any beekeeper (or other person interested in bees) can choose to become a member of the Scottish Beekeepers' Association, while there is a separate Bee Farmers' Association for the UK, the qualification for the latter being that the beekeeper should keep at least 40 colonies of honey bees within the UK. There is known to be some overlap between the two membership lists, and care needs to be taken not to request survey participation of the same person twice for the same survey. Despite the fact that there are far fewer bee farmers than hobbyist beekeepers in Scotland, it is clear that they manage more than half the managed colonies, so that their contribution to the overall bee population is far greater than their numbers would suggest. Therefore in a recent survey it was decided to sample all of the bee farmers who could be identified, while selecting a random sample of non-commercial beekeepers (Gray and Peterson, 2012).

sizes permit valid comparisons. In places where there is widespread practice of migration on a large scale, this comparison becomes much more important. Identifying beekeepers practising migration of bees in advance of drawing a sample may be difficult, unless auxiliary sources like membership records include this information. If this information is available, then a stratified approach may be adopted to ensure coverage of both migratory and non-migratory beekeepers.

Geographical stratification may also be important, especially if different regions are subject to different weather conditions and differing exposure to bee diseases. However, combining multiple stratification factors with lower than ideal response rates can make the desired comparisons statistically invalid or impossible due to small samples.

Cluster sampling is the other main method of probability sampling. If the population can be divided into convenient groups of population elements rather than strata thought to differ in ways relevant to the response(s) of interest, then randomly selecting a few of the groups and including everyone in those selected groups as part of the sample will provide a representative sample from the whole population if the groups or *clusters* are representative of the population. For example, these clusters might consist of local beekeeping associations, which would be viewed as groups of beekeepers. This is a one-stage cluster sample design. There are other variants of this method, but they are unlikely to be of practical importance in this field of application.

6.1.3. Recommended approach for random sampling

In view of the discussion above, stratified sampling is recommended to achieve good spatial coverage of all main geographical areas in a country or region whose beekeepers are to be surveyed, and to give a more representative sample and more precise estimation. Using proportional allocation (see section 9.) is the simplest way to implement this. By ensuring that the chosen strata are represented in the sample, comparisons can be made, and these stratification factors can also be used as risk factors in modelling the risk of colony loss, for example. Achieving good spatial coverage of the population is also essential for spatial or spatial-temporal model fitting, which requires a high degree of data resolution.

6.1.4. Non-random methods

Non-random sampling, is any other kind of sampling. Such methods are often used for speed and convenience, and also they do not require a sampling frame. Their big disadvantage is that sampling error cannot reliably be quantified, as the sampling properties of any estimators used are not known (since the probability of choosing any one individual or sample cannot be determined).

Convenience or accessibility sampling involves asking a sample of people to respond to a survey. An example is distributing survey questionnaires at a meeting of a local beekeeping association or at a beekeepers' convention. However these people may not be representative of the whole target population of beekeepers, for

example due to local weather conditions in the first case, or the fact that attendees at a convention may be real enthusiasts whose bee husbandry practices are not typical of the general beekeeping population. A small convenience sample may be very useful for a pilot survey (see section 7.7) but is not recommended more generally.

An invitation to respond to a survey available on a web-site for example, is an example of taking a *self-selected sample* unless the people invited to respond to the survey have been selected already (as in Charrière and Neumann, 2010).

In some countries, such as Algeria, the most effective method in terms of response rates is a face-to-face survey in the beekeeper's home or at meetings of beekeepers' associations or co-operatives, as using mailed surveys produces an extremely low response. In Slovakia, it is also reported that the only method which works well is to disseminate questionnaires at meetings, as data collection via emails, web pages and journals has very low rates of return. For example, only 5 questionnaires were returned from a beekeeping journal with a circulation of 8 thousand copies (Chlebo, 2012; Pers. Comm.).

Given access to a population to be sampled, a survey organiser could try to take a "representative" sample, which is called *judgemental* or *purposive sampling*, to select what they think is a suitable mix of people to participate in the survey. The difficulty is that some important factors which have a bearing on the responses made to the questions may have been overlooked. Using judgemental sampling leads to a serious risk of badly biased samples.

Quota sampling is like stratified sampling in that stratification factors are identified which are thought to be relevant to the survey, but instead of sampling randomly the participants to come from each stratum, the survey samplers themselves choose the people subjectively from each stratum until sufficient people have been chosen and have responded. The main difficulty with this is the subjective choice of participants. Use of quota sampling also disguises non-response, as invited participants may decline to take part but the sampling will continue until the quotas are achieved. Quota sampling can work well, but can also fail spectacularly badly (as seen most notably in pre-election polling; see Schaeffer *et al.* (1990) for an overview of this and other methods). An example of using non-randomized quota sampling to survey American beekeepers is described in VanEngelsdorp *et al.* (2010), who recognised the dangers of using this approach but judged that it had given results consistent with the pattern of US beekeeping. Box 5 gives an example comparing quota and stratified random sampling.

A fundamental guiding principle in survey sampling is to use randomisation wherever possible in sample selection, to avoid subjective selection bias affecting survey results. Genuinely random samples are well-known to have the best chance of being representative of the survey population and should therefore be used unless it really is not possible (Schaeffer *et al.*, 1990) or will lead to such a low response rate that the results are of little use.

Box 5. Example: A case study: comparing quota and stratified random sampling.

In Scotland, the first survey conducted in 2006 for the Scottish Beekeepers' Association (SBA) used a form of quota sampling. This survey used strata which were broadly geographical, as has also been done subsequently. After the organisers decided on the split of the sample size between strata, they contacted the SBA Area Representatives, in order for them to choose the required number of participants from those known to them personally and known by the Secretaries of the Local Associations of beekeepers in that area. This allowed a known quota to be obtained from each geographical area. This was done purely because permission had not been gained at that stage to use the SBA membership records for sampling purposes and there was no other means of obtaining a list of beekeepers. The results (Peterson *et al.*, 2009) suggested to the organisers that the participants ran larger beekeeping operations than were typical of beekeepers in Scotland as a whole, and also that they were more conscientious and organised beekeepers than was typical. This is not entirely surprising, as the Area Representatives and Local Association Secretaries probably would have chosen people they thought were more organised and more likely to complete and return their questionnaire.

Subsequent surveys from 2008 onwards have used stratified random sampling. In the 2008 survey a modified Neyman allocation method (Schaeffer *et al.*, 1990; Särndal *et al.*, 1992; section 9.) was used to split the sample between the main SBA areas, and subdivided proportionally within these large areas to smaller geographical areas according to the number of SBA members (Gray *et al.*, 2010). In 2010 the simpler proportional allocation was used, as there was insufficient data from the 2008 survey on which to base Neyman allocation and the 2006 data was felt to be out of date. In 2011, Neyman allocation was used again, based on winter loss rates. The results were more in accord with what was expected, and therefore are probably more representative samples than the earlier one. The response rates however have been lower, and the higher response rate in the 2006 survey (of 77%, compared to 42% in the 2008 survey) almost certainly resulted from the element of personal contact.

Finally, it is essential in any reporting of survey results that the survey methodology and response rate should be clearly stated. This enables assessment of the reliability of the results, based on how representative the sample is likely to be. One way to assess whether or not the survey has been successful in achieving a representative sample is to check the responses to a standard question, to which the responses are not expected to change much from survey to survey, if past surveys have been carried out on the same population. If the results of this are different from what is expected this may indicate that the sample is not a representative one. The breakdown of the participants by key indicators such as geographical area or class of operation size can also be examined, although some of these factors will ideally have been controlled for in the sample design, by use of stratification.

6.2. Need for and use of a sampling frame in random sampling

Implementation of random sampling generally requires a list of sampling units, the *sampling frame*, from which to select the sample by random means, although systematic sampling can in some cases be carried out without one. This operates using a numbering system for each person listed. Random selection without replacement of the required sample size from the list of numbers of those having given permission for their records to be used for the purposes of sampling identifies the selected numbers and hence the sample selected, where a simple random sample is required.

For a stratified design, sub-samples are selected in the same way from each stratum, by selecting each sub-sample separately from the list of identifying numbers for those belonging to the relevant stratum. A cluster sample might select a few local beekeeping associations randomly from a centrally held list of associations, as for a simple random sample, and either target everyone in the selected associations as a survey participant, or take a further random sample from the membership list of each local association again using random selection of numbers as the means of sampling.

6.3. Availability of a sampling frame

In some countries, a substantial number of hobbyist beekeepers may choose not to belong to any kind of association of beekeepers or to be registered on an official list of beekeepers, meaning that there can never be 100% coverage in any list used as a sampling frame.

Personal knowledge of some of these beekeepers may enable survey organisers to extend their sampling frame, however the possibilities for this are likely to be very limited. If such independent, unregistered beekeepers form a significant proportion of the beekeepers in a country, then it will be virtually impossible to obtain for that country a truly representative sample of beekeepers. To make matters worse, it is difficult to determine how many such unregistered beekeepers there are. This may be a cause of biased survey results, if the beekeeping experience and loss rates of non-registered beekeepers are likely to be systematically different from those that are registered.

6.4. Sources of sampling frames appropriate for different target populations

The ideal situation is one in which all beekeepers of a country (or other geographical unit) are equally represented in a sample. In some countries, beekeepers, usually with a minimum number of colonies, may be required to register on an official list, in which case gaining access to that list enables access to a selected part of the beekeeping population. In practice, not all beekeepers will register even if this is legally required. The level of compliance with registration requirements may vary greatly from one country to another. In the absence of a satisfactory list of registered beekeepers, other sources of sampling frames may be membership lists of beekeepers' associations or records held by veterinary services. In some countries (e.g. Norway, Sweden, Denmark and the Netherlands), beekeepers' associations may represent up to 90% of the beekeepers, so use of these association records seems to be the best approach currently.

Use of any of the above sampling frames for random sampling does require prior consent, by some means, of the beekeepers on the

list sampled from, for their record to be used in the selection of a survey sample. Those who would not wish this must have the opportunity to opt out and, having done so, should be omitted from the list before random selection takes place. Ethical approval may also be required (section 4.3.). If cooperation with beekeeper associations which represent the majority of beekeepers in a country is not possible, or complicated because there are many small ones all with a limited number of members, another approach is advised.

Firstly, generally most of the bee stocks in the country are managed by large scale commercial beekeepers (even though there are also large numbers of small scale beekeepers). Often the commercial beekeepers have their own trade organisation which will list them all, as well as the approximate sizes of their operations, and if access can be obtained to them, estimation is possible.

If cooperation with a commercial beekeeper association is not possible, an approach using the fragmented smaller organisations may result at least in some kind of sampling frame from which a sample can be drawn with some hope of being representative of beekeepers who belong to these associations, at least in some local areas. How representative these associations are of all beekeepers is of course unknown.

If no beekeeper infra-structure is available, even if, say, in some parts of the world the post office is the only main central information hub, it may be possible to find for each post office district a nucleus of beekeepers. A representative sample of post offices with respect to climate and suitability of area for beekeeping could be drawn up. If such a sample were not too large, then putting out an enumerator for the survey into each of those post office areas might enable that enumerator to find within that area a fairly complete list of the beekeepers in the area. Then the cluster sampling approach would be sensible, where the post offices sampled were regarded as the clusters. A return would be made for each sampled post office area, and the usual techniques for cluster sampling could be used to analyse the results. These approaches may provide a way forward in situations where there is very little by way of an existing sampling frame and limited resources are available or only small scale surveys are possible. Even an imperfect investigation will yield some information. The important thing in considering the results of such work is to be open about the shortcomings of the results, and not to claim more for them than is justified.

7. Questionnaire design

7.1. Completeness of the questionnaire

In the framing of survey questions, the aims and objectives of the desired analysis, and the methods to be used in the analysis, should be borne in mind, in order to ask all of the questions needed to enable collection of the appropriate data. For example, for modelling

the odds of colony loss through CDS, questions should be asked relating to any suspected risk factors as well as collecting data on numbers of colonies lost and what number of losses are attributed to each of a list of possible causes.

In a survey concerned with honey bee colony losses, migratory habits need to be stated as clearly as possible by the respondents in order to avoid misunderstanding about the place(s) where losses were/were not recorded, as well as possible causes of losses, since migratory habit could be one such cause or a contributing factor in honey bee loss.

It is also important to collect and record auxiliary information for statistical purposes such as weighting and multivariate analysis.

7.2. Appropriate designs for different sampling methods

A questionnaire for use by a trained interviewer conducting face-to-face or telephone interviews can be much more elaborately constructed than one used for general distribution (say at a conference) or for a postal or email-based survey or a web-based survey.

In the former situation, the time and cost of obtaining the interviews at all is so large that the additional expenditure for training specialist interviewers and possibly providing them with aids such as laptop computers with purpose-designed questionnaire software is often considered worthwhile. Many large scale government-funded surveys seeking national statistical data are conducted in this way. Questionnaires may then have complicated question routing for various alternative pathways through the questionnaire, as indicated by answers to key initial questions. Sophisticated questionnaire software will have these paths encoded within it. In other cases the interviewer will be familiar with the routes to take depending on the responses given. This is the more common situation in surveys with visiting inspectors of the extension service, or in general in projects which require advanced diagnosis of disease and/or sampling of colonies.

In the second situation, where respondents to the survey have full control of the progress of the response and interpretation of the questions, it is vital that shorter and simpler questionnaires are used, with clear instructions and clear questions, to avoid low response rates and inaccurate responses. This is almost always the situation relevant to surveys of beekeepers.

7.3. Common problems to avoid in questionnaire design

Some common problems to avoid in questionnaire design include *ambiguity of interpretation*, *loaded questions* and *questions on sensitive issues*.

7.3.1. Ambiguity of interpretation

If respondents can interpret a question in various ways, the returns made will not be easy to interpret and the analysis can become difficult or impossible to conduct. Box 6 gives an example. The way to minimise ambiguity is, first of all, to ensure that the early drafts of a questionnaire are always criticised by an independent evaluator before they are used, and once all obvious ambiguities have been removed, to pilot the questionnaire (see section 7.7.) in order to try to detect any remaining problems with the questions.

Box 6. Example: Colony management in Canada.

An example is provided by recent COLOSS surveys in which beekeepers were asked about increases and decreases during a certain timeframe. All Canadian respondents who reported increases or decreases during the defined wintering period were contacted to verify whether such changes truly reflected the dynamics of the wintering population. Invariably, these changes reflected spring-time activities (typically splitting colonies), where these activities could occur in warmer areas of the country prior to the defined end date of the wintering period. Moreover, these changes were not reflected in total colony counts at the end of the wintering period. The question was clear about the timeframe, but a substantial number of beekeepers ignored this information (van der Zee *et al.*, 2012).

7.3.2. Loaded questions

Questions can often be framed in such a way that the respondent is guided towards selecting a particular response, even when that response does not reflect the true state of affairs of interest to the investigator. Box 7 gives an example.

Critical analysis of the original questions for possible loading, and careful analysis of pilot survey results, with subsequent revision of the questionnaire where necessary, are essential.

Box 7. Example: Case study: Experience in a Scottish survey.

An example is provided by a question used in a recent Scottish survey in which respondents were asked in what year they had first become aware of varroa infestation of their colonies. The question was intended to discover how far in the past it was when this parasite had first been detected in that area of the country, since there are still remote areas of Scotland where it has not yet been found. However some newly established beekeepers interpreted this as meaning that they were expected to have personally observed the parasite, and so were inclined to respond that the parasite had "not yet been found" - a biased answer leading to an over-optimistic interpretation of the extent of the parts of the country which were still free of varroa.

7.3.3. Questions on sensitive issues

Even in surveys of beekeepers, some issues can be sensitive. Matters such as financial returns, incidence of disease, location of beehives, and methods for treating diseases and parasites may be sensitive topics for some beekeepers. Taxation, personal and commercial confidentiality issues may be important in financial questions. Beekeepers may feel sensitive about exposure to criticism for poor management if they report disease. If unorthodox treatments for disease have been used, then exposure to the risk of prosecution may make respondents reluctant to respond. Concern about safety of

beehives, or any of these other issues may mean that some beekeepers will not answer those questions at all, will provide incomplete information, or will supply wrong information. There is no doubt that seeking too much sensitive information will seriously reduce response rates and also lead in many cases to incomplete survey returns, thus defeating the object of asking the questions.

It is hard to know how best to address this problem. Firstly, investigators should be aware of what may be sensitive points in their target population, and if necessary, avoid directly asking about them. Sometimes a suitable non-sensitive substitute question may be available for some of these issues, but that too can be problematic. Perhaps seeking information about the mean honey yield per production colony may be felt to be less threatening to a respondent than asking about the financial return for their honey harvest in a particular season, for example. Participants should be assured that all information held will be treated in confidence and only used for the stated purposes of the survey, and that permission would be sought for any subsequent use of the data for other purposes. Any wider data sharing should only be undertaken with extreme caution and great care taken to remove any information which could lead to the identification of the individual beekeeper. There are limits on how successful this can be, in a small area, for example, where beekeepers may be well-known. Information on exact hive location is probably best not shared at all.

7.4. Questionnaire design for minimisation of measurement error and ease of analysis

It is important to word questions carefully, using neutral language, clearly and unambiguously to avoid misunderstanding and consequent errors in data supplied. Carrying out a pilot survey of a questionnaire (see section 7.7), or part of it, involving any new questions is essential in order to check that the questions are appropriately worded. Modifications of existing questions are best tested in the same way, unless the change is minor.

It is worth considering the order of questions asked. The usual guideline is to ask more general questions before more specific ones. This avoids attitudes/responses from becoming fixed early on and may encourage a more flexible way of thinking. It also allows for appropriate question routing, i.e. based on their response to certain questions the respondent is then directed to go to the next appropriate question for them to answer. For example, the questionnaire can state something like "If you responded "yes" to this question, please go next to question X". This is important in surveys drawn randomly from membership lists of beekeeping associations, where not all members may be active beekeepers at the time of the survey and it cannot be determined prior to carrying out the survey which members are active beekeepers. Asking whether or not the respondent is an active beekeeper early on in the survey allows asking such respondents any questions directed at them specifically,

while directing the active beekeepers to the start of the main part of the questionnaire and the questions intended for them.

For ease of data coding and analysis it is best to use closed format questions where possible, with a fixed number of response options of most interest and/or thought to be the most common, and to provide an "other" or "not applicable" category to cater for responses that cannot be fitted into the supplied list of responses, with the means to provide further details of the answer to the question. Closed questions with a given format make it possible to compare responses from surveys of different populations, for example in different countries, or of the same or a similar population at different times. Completely open questions inviting a written response are much less easy to deal with in data coding and analysis and are best avoided. The number of response options is best not to be too long, to avoid confusion or error.

Questions asking for a numerical response are best worded and set out to allow the respondent to supply the exact number, of colonies managed, for example, as the answers can be categorised later if required but having the exact numbers provides more information for analysis.

It is well-known that asking sensitive questions in surveys (see section 7.3.3.) is less likely to elicit an accurate or complete response than less emotive questions. Some survey methods are more successful in this matter than others (Schaeffer *et al.*, 1990). Questions requiring more knowledge than a participant has are likely to be answered inaccurately. Either some background information should be provided, or a screening question(s) should be asked first to determine whether or not it is appropriate for the participant to be asked the particular question of interest.

While constructing the questionnaire and accompanying documentation, including a covering letter/invitation to participate and instructions to the survey participant, a coding sheet should also be prepared. In online surveys, responses will automatically be entered into a database, and the coding of them is part of the questionnaire design (0 = no, 1 = yes, N/A for missing, for example). In surveys requiring manual data entry, a coding sheet is important for consistent translation of survey responses into data entered in a spreadsheet. This is especially important in situations where there is more than one person involved in data entry.

7.5. Need to limit data sought, for a high response rate and accurate measurement

Constructing a long and detailed questionnaire offers the survey organisers the opportunity to collect a great deal of useful information from those survey participants who return their questionnaire, but is likely to result in a rather lower response rate than would be desirable, owing to the time and possible difficulty involved in completing it. Few respondents may return the questionnaire and it is

likely that amongst those who do, some of the information will be missing. Asking very detailed questions is also likely to result in information being less reliable, as not all beekeepers will recall the details or will not have kept sufficiently detailed records to be able to provide correct information, or be unwilling to take time to find the information requested. Longer, more detailed questionnaires and more complex response options are more likely to be successful in face-to-face surveys, but less so in more modern forms of survey. (In telephone surveys, participants may become impatient with long and complicated surveys with many response options, and are likely to terminate the interview prematurely, so shorter and simpler is best.) Postal and self-administered surveys in general require especially clear questions and response options and should be kept to a manageable length (Schaeffer *et al.*, 1990; Brodschneider *et al.*, 2010). Balancing the desire for more information with simpler questions and a shorter questionnaire is likely to produce a higher response rate and more accurate data.

7.6. Problems of multi-lingual/multi-cultural questionnaires

Care should be taken in constructing a multi-lingual or multi-cultural questionnaire, to ensure that the questions and response options are relevant to those receiving them, to avoid needless complication and needless irritation of survey participants, with a view to securing the goodwill and co-operation of the questionnaire recipients and hopefully therefore a high response rate. Accurate translation of specialised concepts requires translation by those familiar with specialist terms in both languages involved, which can be hard to achieve.

Local modifications may be necessary, for example in specifying in relevant questions the month of the start of the winter/summer season for beekeeping, however care should be taken to preserve the meaning intended by the original question. Similarly, differing response options may be appropriate in different countries. For example, in a question about possible disturbances to bee colonies, bears are a possible hazard in some countries, but not in others. Bee races kept will also vary from country to country. Providing "Other" as a response option allows for any more unusual responses, while keeping the specific response options relevant to the participants. Even some questions may not be felt to be relevant ones for some countries. These local variations have implications for the return of the data for central processing and also for its interpretation. Data coding needs to allow for the different response options and care is required in returning accurate data to avoid introducing errors.

One difficulty when colony losses are being recorded is the time period of observation. Lost colonies are common within a period when colonies are not foraging. Depending on the climatic zone this may be winter or other periods. Such periods also differ in duration between years and areas. Using seasonal characteristics allows for comparing

effects on honey bee survival of the length of the non-foraging periods between climate zones. However, in some parts of the world like the USA, Southern Europe and Asia, migration of colonies for pollination purposes to warmer zones during winter can be substantial. This suggests the use of fixed timeframes and determining how many colonies are present at some fixed moments in time.

7.7. Testing survey questions: importance of pilot studies

It is always sensible to test a new or modified questionnaire in a small scale pilot survey before circulating it more widely to a larger group of survey participants. Inevitably in the answering and reviewing of the pilot questionnaires, some unanticipated problems will be highlighted, from minor issues such as duplication of question numbers, to misinterpretation of question wording and issues requiring modification of question wording, new response options and/or additional questions. Box 8 gives an example.

If re-using a well-tried and tested questionnaire, clearly there is less need for a pilot run. However, if new questions are added a small pilot run is still advisable. Almost always some small point has been overlooked or can be improved upon, despite the most careful survey design. Even with an old questionnaire, piloting is often advisable to ensure that questions are still comprehensible and relevant.

Box 8. Case study: Pilot surveys in Scotland.

In recent surveys in Scotland, for example, about 6 people known to one of the survey organisers through his local beekeeping association were identified as suitable candidates who were readily contactable, covering a wide span of years of beekeeping experience from the beginner to the much more experienced. The questionnaire was delivered to them personally at a time when they able to deal with it immediately or an arrangement made to collect it shortly thereafter, so that no responses went missing. In the face to face situation, any immediate difficulties in understanding the questions are easily dealt with and explained, and a note made that these questions need to be re-worded. In all cases this exercise has suggested some points to be changed in the survey questionnaire, if only minor ones, and has been felt to be very useful.

7.8. Example of a standardized questionnaire on colony losses

An example of a standardized questionnaire, produced by the monitoring working group of the COLOSS network (section 1), is provided in Fig. 2. The questionnaire can be split into essential questions which should be implemented in all participating countries and optional questions which are left to the national survey organisers to use or not. Optional questions ask more information about the operation such as postal code and location of the apiary, migration, bee race, increases and decreases made by the beekeeper during winter, origin of queens, queen replacement, pollination services, honey and pollen sources, comb replacement and winter feeding. The survey organisers may replace the concept of winter by another seasonal concept suitable for the local situation.

8. Response rates

There are different ways to calculate response rates. The examples given here simply use the number of usable responses (complete or otherwise) divided by the size of the sample selected or number of participants approached. Variations on this as well as several other measures of outcome rate are discussed fully in the reference AAPOR (2011).

8.1. Use of incentives and reminders to improve response rates

As mentioned above, reminders are an important means of improving response rates in self-administered surveys. A personal reminder is likely to be more effective than a more general public one. Providing an incentive to participate in the survey to those already selected to participate can also encourage return of a questionnaire and hence may have some beneficial effect on response rates.

In telephone surveys call-backs are easy to arrange. Sending repeat emails is also straightforward. In an online web-based survey, as in self-selected survey samples generally, it is the more motivated who will respond to a general call for participation and these may well coincide with those who have more extreme opinions or experiences to report. Therefore reminders are important to try to overcome the bias which this creates, by involving some of those who are less inclined to participate but who may be more representative of the population as a whole. Box 9 gives an example.

Box 9. The Scottish surveys: use of reminders and incentives.

In the 2008 survey in Scotland, a short public reminder was published in "The Scottish Beekeeper", but the final response rate was only 42%. In that survey no personal reminder was possible as anonymity was built into the survey and questionnaire numbering was not used. Numbering of the questionnaires allows identification of the selected survey participants who have not responded. In recent Scottish surveys the numbers of questionnaires returned by the deadline were removed from the list of numbers of all the questionnaires sent out, the remaining numbers matched to the reference number of the person concerned and this list of numbers sorted into order and sent to the membership convenor for identification of the people in order for him to send a short reminder letter. The first time this was done, in 2010, the response rate was considerably improved, to 69%, although in 2011 it had little effect (response improved from 45% to 49%) and there was barely any effect in 2012. Nonetheless reminders are recommended.

In the last few annual surveys of beekeepers, a well-known commercial supplier of beekeeping equipment has willingly provided a generous voucher to be awarded to the winner of a prize draw at the end of the deadline specified for return of the questionnaire. The winner was randomly selected from the list of questionnaire numbers returned by that deadline. The winning number was matched to the identifying short reference number for that participant, and the details were sent to the SBA membership convenor. The convenor identified and contacted the winner, and contacted the commercial company to arrange for the sending of the prize to the winner. The winner was asked what details they would be willing to have published in the SBA's monthly publication for members, for example, information such as "The winner of the £50 voucher kindly offered by Company A as a prize to the successful participant in the SBA 2010 survey lives in Argyll)", hence giving some publicity to Company A.



Survey on Colony Losses 2012

In this questionnaire we try to gather information about production colonies. Please consider colonies which are queen-right and strong enough to provide a honey harvest as production colonies.

1 Country _____

2 Region/Province _____

In the next questions you are asked for numbers of colonies lost. Please consider a colony as lost if it is dead, or reduced to a few hundred bees, or alive but with queen problems, like drone laying queens or no queen at all, which you couldn't solve.

Please consider winter as the period between the moment that you finished the pre-winter preparations for your colonies and the start of the new foraging season.

3 How many production colonies did you have before **winter 2011-2012**?

4 How many of these colonies were lost during **winter 2011-2012**?

5 How many of the lost colonies did not have dead bees in the hive or in the apiary?

6 How many of the lost colonies had dead workers in cells;
and no food present in the hive?
and food present in the hive?

7 How many of the lost colonies had queen problems, like drone laying queens or no queen at all?

8 How many production colonies did you have after;
winter 2010-2011
winter 2011-2012

9 Have you treated your colonies against Varroa during the period November 2010 - January 2012?

| | |
|-----|----|
| yes | no |
|-----|----|

To learn about the number of Varroa treatments during the period November 2010-January 2012 we would like to get information on the months you started a Varroa treatment with a product.

10 Could you please indicate in what month and year you started every Varroa treatment of your colonies with a product during the period November 2010 - January 2012?

| 2010 | | 2011 | | | | | | | | | | | | 2012 | | |
|----------|----------|---------|----------|-------|-------|-----|------|------|--------|-----------|---------|----------|----------|---------|----------|-------|
| November | December | January | February | March | April | May | June | July | August | September | October | November | December | January | February | March |

Fig. 2. An example of a standardized questionnaire, produced by the monitoring working group of the COLOSS network.

9. Choice of sample size

In a probability-based sample, the sample size can be calculated statistically in order to achieve a required level of precision of

estimates from the data collected, where these estimates have been identified in advance as being of interest. The formulae required depend on the sampling scheme to be used. Schaeffer *et al.* (1990) give details.

For example in a simple random sample, to estimate a mean, e.g. average number of colonies kept per beekeeper, to within a distance or error bound B of the correct value with approximately 95% confidence, the formula for the sample size is $n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$ where $D = \frac{B^2}{4}$ and σ^2 is the variance in the population of the quantity of interest, e.g. the number of colonies kept, and N is the population size. In the case of a very large population of beekeepers, where N is not known exactly, an approximation to this sample size is given by $n = \frac{\sigma^2}{D}$. The population variance may be estimated from the variance calculated from data in a previous survey of the same population, or from a pilot survey. To estimate a total (by the population size N times the sample average) with the same precision uses this same formula but with $D = \frac{B^2}{4N^2}$. Box 10 provides an example of the calculations.

Box 10. Sample size calculation for a survey to estimate a mean or a total.

For example, using a simple random sampling approach, to estimate the average number of colonies kept to within a margin of error of 10% ($B = 0.10$) of the true value with an approximate confidence level of 95%, the sample size is calculated as follows. We use the formula $n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$ where $D = \frac{B^2}{4} = \frac{0.10^2}{4} = 0.0025$. Assuming that the total number of beekeepers in the population is 1500, and if we have recent information from a previous survey that the variance σ^2 of the number of colonies per beekeeper is about 4, then we should sample $n = \frac{1500(4)}{(1499)(0.0025) + 4} = 775$ beekeepers, rounding up to the nearest integer. If we wished to estimate the total number of colonies kept, say to within 200 of the actual total with the same level of confidence, then making use of the same information, we calculate instead $D = \frac{B^2}{4N^2} = \frac{200^2}{4(1500^2)} = 0.00444$, which now gives $n = \frac{1500(4)}{(1499)0.00444 + 4} = 563$ beekeepers to be sampled.

To estimate a proportion p to within an error bound B of the true value with approximately 95% confidence, the same exact and approximate formulae are used as for estimating a mean, but with $\sigma^2 = p(1-p)$, so in the large population case $n = \frac{p(1-p)}{D}$, $D = \frac{B^2}{4}$. These formulae require an approximate value for p based on prior experience, or else substitution of a conservative value of $p = 0.5$ to maximise the required sample size. Box 11 shows the calculations.

Box 11. Sample size calculation for a survey to estimate a proportion.

For example, using a simple random sampling approach, to estimate an overall proportion of losses which was 20% last year (so $p = 0.20$ approximately), to within a margin of error of 5% ($B = 0.05$) of the true value with an approximate confidence level of 95%, the sample size is calculated as follows. The population size is assumed large, but is unknown. So we use the large population version of the sample size formula for estimation of a proportion given by $n = \frac{p(1-p)}{D}$, $D = \frac{B^2}{4}$. Here this gives $n = \frac{0.20(0.80)}{D}$, $D = \frac{0.05^2}{4}$, giving $n = 256$ exactly. So the sample should be composed of at least 256 individuals to achieve the required level of precision.

If there is more than one quantity to be estimated, as there will be in surveys of beekeepers, the larger of the relevant calculated sample sizes can be used, where this is feasible, or it can be decided to focus on one more important estimator, e.g. the proportion of beekeepers experiencing winter colony loss or the proportion experiencing CDS losses. It is then accepted that any other estimates requiring a larger sample size will be estimated with lower precision than is desirable.

For a stratified sample, which takes simple random samples from each stratum, similar calculations may be done to obtain the overall sample size required to estimate the mean or total or proportion to within an error bound B of the true value with approximately 95% confidence. See Schaeffer *et al.* (1990), for example, for details.

Various approaches are possible to divide the chosen sample size between the strata, including the *proportional* method which takes the sample size n_i in the i th stratum proportional to N_i/N , where N_i is the size of the i th stratum and N is the population size. This means taking $n_i = \left(\frac{N_i}{N}\right)n = W_i n$, where W_i is the i th stratum weight or the proportion of the population belonging to stratum i .

Neyman allocation is a more complex method which splits the sample between strata in order to minimise the variance of the unbiased estimator of the population mean (given by $\sum_{i=1}^k w_i \bar{x}_i$, where $k = N_i/N$, where k is the number of strata, $W_i = N_i/N$ and \bar{x}_i is the mean of the sample from stratum i) or of the total (taken as N times the estimator for the mean) by taking the i th stratum sample size proportional to $N_i\sigma_i / N$ or $W_i\sigma_i$, where σ_i^2 is the variance within stratum i and σ_i is the standard deviation the variance within stratum i . So

$$n_i = \left(\frac{N_i\sigma_i}{\sum_{i=1}^k N_i\sigma_i} \right) n = \left(\frac{W_i\sigma_i}{\sum_{i=1}^k W_i\sigma_i} \right) n.$$

The within stratum variances may be estimated from previous experience or a pilot survey.

To estimate a proportion (by $\sum_{i=1}^k w_i \hat{p}_i$, where \hat{p}_i is the sample proportion in stratum i), the same formula can be used for allocation as for estimating a mean, but σ_i is replaced by $\sqrt{p_i(1-p_i)}$ where p_i is the value of the population proportion in stratum i (and in practice an estimate of this is used).

The Neyman approach can also be modified, if required, to incorporate different sampling costs for each stratum. More complex modified Neyman allocation schemes are also possible (Särndal *et al.*, 1992).

More generally it may be decided, in order to achieve a suitable coverage of the population, that a fixed percentage of the population should be sampled. For some of the COLOSS surveys, a guideline for acceptable coverage has been that, where possible, at least 5% of beekeepers should be surveyed. This is a simple way to choose sample size, especially in a non-probability sample for which sample size calculations are not valid.

Another concern in a smaller population which may be surveyed repeatedly is not to overburden individuals, but to maintain goodwill.

This may mean taking a smaller sample than is ideal. Data processing concerns may also limit the sample size.

If the level of non-response can be anticipated, for example, from recent experience, the calculated or chosen sample size can be increased accordingly, in order still to give a sample of the required size, as $n_2 = n_1 / (1 - r)$, where n_1 is the original sample size, n_2 is the new size, and r is the expected non-response rate as a proportion, e.g., $r = 0.25$.

Obtaining standard errors of estimates, or confidence intervals, as part of the data analysis indicates how precisely the various quantities of interest have been estimated (see sections 4.1.2. and 10.).

10. Analysis of survey data

10.1 Assessing data quality

Prior to the analysis, some assessment and possible improvement of the quality of the data is essential. This is of utmost importance when these data are to be used in statistical models. Errors of different kinds can easily result in false inferences of general patterns, meaning that effort expended in complex modelling may be largely wasted if the data are unreliable.

As numerical data is not directly measured but derived from surveys, the means of data collection used in the surveys has to be designed in such a way that respondents have limited opportunity to generate extreme or erroneous responses. Thorough data validation must precede modelling procedures, i.e. checking for out-of-range data (invalid responses), and inconsistent responses. The proportion of missing values is also an indication of data quality. See De Leeuw *et al.* (2008), chapters 17-22, for an overview of quality control and data validation for survey data.

If the results of data checking suggest that the data are unreliable, then it may be sensible to limit analysis to simple procedures, or else interpret the results of model fitting with some caution. This is also true for small data sets where complex model fitting may not be feasible.

If the selected sample size is known, as for example in a randomized sample, the overall non-response rate can be calculated as a first indicator of quality, as a survey with a high non-response rate (a low achieved sample size) may be unrepresentative of the population of interest. Assessing non-response involves comparison of the actual sample size and the planned sample size (see § 9 on choice of sample size).

Examining responses to individual questions is also necessary. For each question, several simple quality measures may be calculated:

1. The missing data rate can be checked (for partial non-response).

A high proportion of missing data may indicate inappropriate or sensitive questions, for example those which will be important to reconsider for the question design of future surveys. Missing data may be left as missing for the purposes of the analysis, or a data imputation method may be used to replace the missing data with a plausible value (De Leeuw *et al.*, 2008, chapter 19).

2. The proportion of invalid values can be checked. The size of any deviations from what is a valid response is also of interest.

The response may be a value outside the valid range of responses for that variable, such as a percentage above 100 or a negative number of colonies lost, or it may be a suspiciously extreme value. This problem occurs when the question was not correctly answered by the respondent, or when the data was not correctly captured at the point of data collection or data entry. A question with many invalid answers should probably be reformulated. If there is no way of checking what is the correct answer, the response should be considered as missing data and should be omitted from the analysis.

3. The proportion of inconsistent values must be checked.

It may be clear from examination of the data that the responses to some questions are inconsistent with the responses to some other questions. For example, the calculation of the number of colonies lost in periods when bee management is practised may give a different answer from the number of colonies stated as having been lost. A variable recording the difference in these two quantities may be used as a filter to remove cases with inconsistent data from analysis.

These data quality descriptors can be obtained from descriptive analysis, for example using summary statistics including the range of a variable, tabulations, cross-tabulations and histograms.

10.1.1. Dealing with missing data

The treatment of missing data is a rather specialised statistical topic. Missing data is difficult to deal with adequately in the analysis of questionnaire data, especially if it is not "missing at random". Data which is missing at random is such that the responses that would have been given are not related to the probability of non-response. If data is missing at random, then the data that is available can be analysed and the results should still be representative of the population, provided that the selected sample was representative. If it is not missing at random, then the results of analysing the available data are likely to be badly biased. Missing data reduces the number of responses available to analyse and hence reduces the precision of any estimates made. The best approach therefore is to try to minimise the chances of data being missing, by careful questionnaire design and by choosing a survey mode which gives respondents time to complete all the questions and secures their co-operation to do so.

10.2. The use of weighting in statistical analysis

In an analysis of a survey based on random selection, if the survey does not use simple random sampling or sampling with replacement, then participants are not all equally likely to be selected from the population. In this case, to achieve unbiased estimation, a case weight should be assigned to each of the participants returning data. These sampling weights should be inversely proportional to the probability of selection of each participant and should sum, over all the participants, to the sample size. Sampling weights can only be calculated if probability sampling is used. The software package SPSS, for example, allows a weighted analysis to be carried out.

For example, if a stratified sample is used, based on geographical area, for a case sampled from stratum i the sampling case weight is given by $(N_i/N)/(n_i/n) = (N/n)/(n_i N)$ where N is the population size, N_i is the size of stratum i in the population, n_i is the number of people sampled from stratum i and n is the total sample size. This requires knowledge of which area or stratum a respondent comes from. Numbering questionnaires and recording in the data spreadsheet the area in a column beside the questionnaire number is probably the best way to ensure that the required information is available. Inclusion of appropriate questions can make it feasible to set up weights to be used in a weighted analysis of the data, however not all participants may respond to these questions, so it is safer to record the information in advance.

Weights can also be used to allow for *unit non-response*, i.e. where some people do not respond at all. These weights are inversely proportional to the probability of responding. So in stratum i , each person would have a non-response weight of n_i/n_{ir} , where n_i is the number of people selected from stratum i and n_{ir} is the number of people responding from that stratum. This and more sophisticated methods are discussed in Lehtonen and Pahkinen (2004).

The weights for sample design and the weights for non-response can both be used at once, by multiplying the two columns of weights together and rescaling so that the new weights add to the sample size.

In a multi-cultural survey, in which different sampling designs may have been used to select participants in different countries, different weight calculations will be needed for respondents from each constituent country, and this requires detailed knowledge of the different survey designs which were used to select the samples. In practice this information may not be readily available.

10.3. Elementary analysis

10.3.1. Descriptive analysis

Any statistical analysis of data should begin with simple data description and presentation using summary statistics, tables and plots. While the main interest may well be in modelling, the initial analysis is still an essential first step. The results of such analysis with

well-designed graphs and/or tables can reveal unsuspected patterns in the data, and will ensure that the obvious characteristics of the data are clearly understood by readers of any resulting report.

Responses to any simple survey question clearly require this approach, e.g. to determine the proportion of respondents in a postal survey who are currently beekeepers, where this cannot be determined in advance of choosing the sample, or the proportion wishing to remain anonymous, or the proportion experiencing any colony losses over a specified period. The data on any categorical variable can also be presented in a bar chart and/or a contingency table, with frequencies and relative frequencies, for an overview of the responses, the range of values and the most common category. This will also help in identifying invalid responses.

Extending this analysis to more than one categorical variable, e.g. to compare the proportions of losses experienced by respondents in different countries, or by geographical area within a single country, or for different sizes of beekeeping operation, two-way tables are useful. Relevant follow-up tests include chi-squared tests of association or homogeneity, which will permit the statistical investigation of the possible significance of differences in sample proportions. Even if observations contributing to each cell in the table are not all independent, the results of this can inform any subsequent modelling, by identifying potential risk factors for colony loss, for example, to be included in the model.

For questions with a quantitative response, of most interest is some measure of a typical or central value. The most appropriate measure depends on the distribution of the numerical responses. Where these are fairly symmetrically distributed, and there are not many extreme atypical values, the best measure is the *mean* or arithmetical average of the observations. However if the distribution of the data is very skewed, and/or there is a fairly large proportion of extreme atypical values, then the mean can be seriously misleading. For example, in the distributions of number of colonies kept per beekeeper, or honey yield, the existence of a few very large numbers of colonies kept or correspondingly high honey yields has the effect that the mean will give a grossly inflated idea of what is a typical value. The number of lost colonies per beekeeper also tends to have a highly positively skewed distribution. For such cases the *median* is preferred. This is the central observation, or the mid-point between the two central observations, after the data have been arranged in increasing order of magnitude.

Almost as important is some measure of dispersion of the observations around the mean or median, whichever has been chosen as being most appropriate. The usual choices are either the *standard deviation* for variables for which the mean is used, or the *inter-quartile range* for situations where the median is the appropriate measure of a typical value. (Any first level statistics textbook, such as Ott and Longnecker (2009) or Samuels *et al.* (2010), will describe the computation

of these quantities). Confidence intervals based on the mean are z -intervals in the case of a large sample, or t -intervals for smaller samples. Population means may be compared using Analysis of Variance (ANOVA; see Ott and Longnecker (2009), and Pirk *et al.* (2013)), assuming independent observations and independent samples from normal distributions. For medians, nonparametric confidence intervals and tests are available, including the Kruskal-Wallis test as a nonparametric equivalent of ANOVA. Nonparametric procedures generally are robust to data which does not conform exactly to the assumptions of the test procedure.

Histograms are essential graphical tools to study the nature of the probability distribution of a quantitative variable such as number of colonies kept or number of colonies lost or honey yield, and hence to determine whether this is symmetric or skewed. Boxplots can also be useful in this regard. Comparing these between countries for example can indicate differences.

Comparing a histogram visually with the theoretical density functions of a range of possible probability distributions is also a simple first step in selecting and justifying a plausible model for use in more advanced statistical modelling of a dataset. The most frequently used probability model for the distribution of continuous numerical data is the symmetric bell-shaped *Normal* or *Gaussian* distribution. However for data which are clearly asymmetrical and skewed, the choice is wider. For continuous positive data, the *Gamma* distribution provides a large family of shapes of probability distribution, or the *Beta* distribution can be used for positive data over a finite range between 0 and some given positive value a . For skewed count data, the *Negative Binomial* distribution may be appropriate. For example, data describing number of colony losses contains many zeroes, but may also have some rather high numbers lost. Various tests for goodness of fit can be used to see if any of these models can be clearly ruled out, but often the final choice is governed by considerations of convenience and mathematical tractability.

10.3.2. Loss calculations and Confidence Intervals

1. Regarding loss rates, rather than the raw numbers of colonies kept and number of colonies lost which are used in their calculation, different quantities are of interest. The overall loss rate is the proportion calculated as the total number of lost colonies in the sample of beekeepers divided by the total number of colonies at risk of loss in the sample. (VanEngelsdorp *et al.* (2013) refer to this as "total loss". As this suggests to us the total number of colonies lost rather than any kind of rate or proportion, we prefer the terms overall loss rate or overall proportion of colonies lost). Adjustments can be made to this calculation to take account of colony management (VanEngelsdorp *et al.*, 2012). The overall loss rate is influenced disproportionately by the larger beekeepers, who are fewer in number. Using this approach,

confidence intervals for proportions may be calculated. There are several ways to do this. Alternatively, the average loss rate is the average of the individual loss rates (number of colonies lost divided by number of colonies at risk) experienced by different beekeepers in the sample. Using this approach, confidence intervals should be those for an average, not a proportion. However, a difficulty of using the average loss rate is that the loss rates experienced by beekeepers with different sizes of operation are not equally variable, yet they are weighted equally in the calculation of this average. While the loss rates can only range between 0 and 1 (0 to 100%), larger scale beekeepers have many more colonies which can be lost, and can experience a much larger set of possible loss rates within this range; therefore, their loss rates are subject to greater variation. Also, there are many ties in the individual loss rates, for example due to the large number of beekeepers with no losses. The median individual loss rate could well be zero. Average individual loss rate is often higher than overall loss rate, owing to the larger number of small scale beekeepers present in many populations of beekeepers, who can suffer extreme individual loss rates. For this reason, the use of medians and Kruskal-Wallis tests to compare loss rates should be avoided. Owing to these various difficulties, we recommend use of the overall loss rate.

2. Another difficulty is that the usual procedure to calculate standard errors and confidence intervals for the overall loss rate (the proportion of colonies lost) is based on the binomial distribution, as the number of losses is limited by the number of colonies at risk. This assumes that each bee colony is lost or not independently of any other colony, and also that the probability of loss is the same for all colonies. Within apiaries, whether or not a colony is lost is likely dependent on whether or not neighbouring colonies are lost. Furthermore, the probabilities of losing a colony are likely to differ between beekeepers. One way to account for that extra source of variation in the data is to model the data using a generalisation of the binomial distribution. There are different ways to do this. One approach uses generalised linear modelling using a quasi-binomial distribution and a logit link function, and derives a confidence interval for the overall loss rate based on the standard error of the estimated intercept in an intercept-only model (see VanEngelsdorp *et al.* (2012) and below).
3. Another approach to calculating confidence intervals, when it is felt that formulae based on parametric models are not appropriate, is to use the nonparametric bootstrap approach, based on resampling the data (Efron and Tibshirani, 1994). This avoids the need to specify any particular model for the data. This is easy to implement in a software package such as R.

10.3.3. Loss rate per factor including stratification on the operation size

The loss calculations and confidence intervals described above can be used as a means to identify risk factors for colony loss, by looking for confidence intervals that do not overlap each other. Total loss of operations reporting or not reporting a particular management type (e.g. transport of colonies) can be compared using the chi-square test (as in VanEngelsdorp *et al.*, 2010, 2011, for example). The loss rates of operations grouped by factors presumed to be involved in colony mortality (starvation, high varroa infestation etc.) can also be compared. Of course, this analysis does not give any information on, or account for, interdependencies of different factors, for which model fitting is needed (as described below).

To account for known or obvious differences among beekeeping operations, a first stratification, for example on operation size, can be accomplished, by classifying operations as hobby, side-line or commercial. Alternatively the number of colonies per beekeeper can be used as a basis for stratification.

Depending on the size of the survey and cultural differences between the target populations, beekeeping operations can be split into three operation size classes, for example

- small operations (≤ 50 colonies),
- intermediate operations (51-500 colonies),
- large scale operations (> 500 colonies).

If the scale of beekeeping in the survey population is limited mainly to small and intermediate operations, the classes can be split further as:

- small hobbyist beekeepers (≤ 15 colonies),
- large hobbyist beekeepers (16-50 colonies),
- small-commercial beekeepers (51-150 colonies),
- larger-commercial beekeepers (150-500 colonies).

When comparing several operation size classes, a chi-square test can be used first to compare all size classes, and if the result of this is significant, it can be followed up by pairwise multiple comparisons, again using the chi-square test or a z-test of the difference in two proportions. In each such pairwise test, the significance level to reject the null hypothesis should be Bonferroni adjusted (i.e. divided by the number of tests being conducted) to reduce the rate of false rejections of the null hypothesis that operations of different sizes have equal rates of loss. It should be borne in mind that the chi-squared test and z-test assume independent observations and therefore have their limitations.

10.4. Advanced analysis; identification of risk factors by logistic regression

10.4.1. Logistic regression

Elementary analysis of the answers to the essential questions, regarding colony losses, can yield an estimate of the overall loss rate

for the observations (beekeepers or operations) grouped together by a single factor (such as country, or involvement (or not) in commercial pollination). Comparing these loss rate estimates and confidence intervals for the loss rates can indicate differences between the groups and hence potential risk factors relating to the risk of colony loss. The overall loss rate is a problematic estimator when the contribution of multiple factors to the risk of loss has to be determined, since factor responses may be associated, not independent of each other. For example, commercial pollination is more common in certain countries than others. Larger scale beekeepers contribute more to the overall loss rate than smaller scale beekeepers.

A statistical approach that deals with the difficulties of overall loss rate and enables conclusions on how factors (bee race, pollination practices, size of operation, honey yield, location etc.) influence colony losses is regression analysis (see Zuur *et al.* (2009) and Pirk *et al.* (2013)). In regression analysis, the numerical outcome of the essential questions (number of colonies lost, number of colonies alive or the calculated population at risk) is linked to the factors through a linear model. In the analysis of bee colony losses, many of the response variables of interest are positively skewed (having a long tail to the right) and so generalized linear regression models (GZLMs) are appropriate. These models assume that the observations y_i arise independently from a specified family of probability distributions, and independent variables or factors $x_{j,i}$, $j=1, \dots, k$, are used to provide a set of linear predictors

$$\eta_i = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}$$

such that $g(\mu_i) = \eta_i$, where μ_i is the mean of y_i , and the β_i are model coefficients to be estimated. Using GZLMs requires the specification of an appropriate probability distribution for the response variable y and also an appropriate form for the link function g (Krzanowski, 1998; McCullagh and Nelder, 1983).

The dependent variable of interest, the loss rate, is binary in the nature of its components (the number of lost colonies divided by the number of colonies at risk makes up the loss rate). This property leads to models that use a binomial distribution for the dependent variable. Each colony can be regarded as a "Bernoulli trial" resulting in no loss or a loss (0 or 1 respectively), and the number of lost colonies for a beekeeper can be regarded as a "binomial trial" of a certain size n (total number of colonies at risk, or number alive before the winter rest period) with a certain probability (p) of any one colony being lost after winter (an "event") and probability $1-p$ of the colony being alive after winter (a "non-event"). If x is the number of events per beekeeper, then the binomial probability distribution describing the probability of x events has the formula

$$p(x, n, p) = \binom{n}{x} p^x (1-p)^{n-x},$$

with the mean value of x given by np and variance of x by $np(1-p)$.

Groups of beekeepers or operations can be seen as series of binomial trials which vary in size, and also with different probabilities

of an event, p . Hence it is of interest to model the probability of loss for (groups of) beekeepers or operations characterized by different values of the risk factors involved, such as country or operation size or migratory practice.

Probabilities cannot be used directly as a response variable in a classical linear regression model, as probabilities can only have values ranging from 0 to 1, whereas continuous response variables can have any value. The solution for this problem is moving from the probability to the "odds" ($p/(1-p)$) and calculating the logarithm of the odds, the "logit", to be used as the dependent variable. The first step, taking the odds, removes the boundary of 1 as the odds can have any positive value, while taking the logit in the second step removes the boundary of 0 as the logarithm can be negative (for odds less than 1). A probability of 50% has an odds of 1 and a logit of 0, with negative and positive logits corresponding to probabilities of less than and more than 50% respectively.

Generalized linear models of this nature are called *logistic regression models*, and can be expressed in the form

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i},$$

where the β_j are model coefficients to be estimated and $x_{j,i}$, $j=1, \dots, k$, are the values of the k independent variables or factors used in the model for prediction of the log odds of loss for case i .

Substituting the values and the estimated parameters into the right hand side of the equation enables prediction of the log odds of an event for that beekeeper or operation or group of operations. If this gives a value y , then taking the inverse logit $e^y/(1+e^y)$ gives the prediction of the probability p_i itself.

Kleinbaum and Klein (2002), Hosmer and Lemeshow (2000) and Agresti (2002) give an in-depth explanation of the principles of logistic regression, their interpretation, and the construction of best fitting models.

When honey bee loss data are involved in the analysis, several specific characteristics of these data and their analysis have to be addressed, as are now described.

10.4.2. Dispersion in statistical models

For a binomial distribution, the variance $np(1-p)$ depends on the mean np . When the variance in the observations is bigger or smaller than the expected variance, data are said to show over- or under-dispersion. Both types of dispersion are indicated by the goodness-of-fit tests of fitted models by the ratio of the residual deviance of the fitted model to the number of degrees of freedom, values appreciably larger than 1 indicating over-dispersion and values lower than 1 indicating under-dispersion. Both types can strongly affect and invalidate model hypothesis testing (standard errors, confidence intervals and p -values). See Twisk (2010), Zuur *et al.* (2009), Hardin and Hilbe (2007) and Myers *et al.* (2002) for examples. Causes of under- or over-dispersion can be related to the frequency characteristics of the data, with relatively small and large beekeepers/operations present in

different numbers (heterogeneity of the sample population). An important assumption of a binomial distribution, namely independence of observations (independent Bernoulli trials), might be violated when losses are not independent (are clustered) through an unknown factor (i.e. effects of a certain location, incidence of pathogens) that cannot be used (properly) in the model.

When under- or over-dispersion are not reduced after using the most significant model factors derived from the data and/or stratifying available data according to binomial trial size, the solution is using a different distribution for the dependent variable. A suitable candidate is the quasi-binomial distribution, in which variance is characterised by adding an additional parameter to the binomial distribution, and hypothesis testing can be corrected for the extra-binomial variance. The form of the quasi-binomial probability distribution is:

$$p(x, p, \varphi) = \binom{n}{k} p^x (1-p)^{n-x} (1 + \varphi \frac{x(n-x)}{n})^{-\frac{1}{\varphi}}, \quad x = 0, 1, 2, \dots, n; \quad 0 \leq p \leq 1; \quad -\frac{p}{n} < \varphi < \frac{1-p}{n}$$

See the manual available online by Kindt and Coe (2005) for an excellent example of the use of a quasi-binomial distribution and its differences compared to the standard binomial distribution. An excess of zero values (no loss) can be a cause of over-dispersion. To investigate the relation between predictor variables and the presence of zero values (no loss), zero-inflation techniques can be used (for example, Hall (2000)).

10.4.3. Multilevel analysis

Clustering of losses results in over-dispersed data, but clustering might very well be a biologically relevant phenomenon. A method to investigate correlations between groups of observations is to perform multilevel analysis by means of fitting models that contain random effects (random effects models and mixed models). Classic examples of multilevel analysis include schools or hospitals as random factors in an analysis of dependent variables on the level of students or patients respectively. In the case of colony losses, suitable data levels for random effects are often spatial in nature, as colonies are clustered by beekeepers, beekeepers are clustered in regions or habitat types and the latter are clustered within countries.

See Twisk (2010) and Zuur *et al.* (2009) for practical application of multilevel analysis methods. Rodríguez (2008) is also useful. A good online resource for multilevel analysis can be found at the homepage of the University of Bristol Centre for Multilevel Modelling (at <http://www.bristol.ac.uk/cmm/>).

10.4.4. Software for logistic regression models

Logistic regression can be conducted using generally available statistical software. The software packages R, SAS, STATA and MLwiN are able to perform logistic multilevel regression. The latest version of SPSS (19) also has a mixed model procedure but has no option (at the time of writing) to use quasi-binomial distributions. For an evaluation of different software used for logistic multilevel analysis, see Twisk (2010).

10.4.5. Example of advanced analysis

The analysis below uses the Dutch data collected with the full 2011 COLOSS questionnaire, as an example of how to estimate overall loss rates, calculate confidence intervals and fit GZLMs. It uses the quasi-binomial family of GZLMs, to account for any extra-binomial variation in the data. It is a simple illustration of how model fitting can be done in R, with factors and covariates, rather than a procedure for determining a best fitting model. Guidance on model building may be found, for example, in Dobson (2002) and Zuur *et al.* (2009).

The data was "cleaned" prior to use to remove some inconsistent values. The "glm" procedure in R is sensitive to invalid values in the data, and will generate error messages rather than omit the cases with invalid data values, so it is best to deal with these before attempting model fitting (or any other kind of analysis). The analysis below uses the variables ColOct10 as the number of colonies kept at

1st October 2010, and Loss1011, the stated number of colonies lost over winter 2010/2011, rather than the calculated population at risk or calculated colonies lost. Even so, in one case Loss1011 was missing and in six other cases Loss1011 was greater than ColOct10, causing negative calculated values of a new variable, NotLost, the number of colonies surviving. In some cases, though not all, this was due to winter management (making in/decreases) of colonies. These few cases were also removed before carrying out the analysis shown below.

The analysis does not show all available options for the "glm" procedure. Several diagnostic plots are available, for example.

- Calculation of overall loss rate and confidence interval from a null model (Boxes 12-14).
- Fitting a GZLM with an explanatory term.

Box 12. Reading in and setting up the data for further analysis.

Read in the data, in this case in csv format:

```
>dutch<-read.csv("cleaner_dutchdata.csv", header=T,sep=",")
```

Check the first few rows and columns:

```
>dutch[1:5,1:6]
```

| | Validity | COLOSSID2011 | IDBeekeeper | Country | Region | City |
|---|----------|--------------|-------------|---------|--------|-----------|
| 1 | 0 | 6528 | 1426 | 167 | NA | |
| 2 | 0 | 6529 | 1607 | 167 | NA | |
| 3 | 0 | 6531 | 5048 | 167 | NA | Den Hoorn |
| 4 | 0 | 6532 | 5296 | 167 | NA | Amsterdam |
| 5 | 0 | 6533 | 5396 | 167 | NA | Amsterdam |

Load the data into the memory, so variables can be identified by name:

```
>attach(dutch)
```

Check the descriptive data for the variables:

```
>summary(ColOct10)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|---------|
| 1.000 | 3.000 | 5.000 | 8.932 | 9.000 | 401.000 |

```
> summary(Loss1011)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|--------|
| 0.000 | 0.000 | 1.000 | 1.905 | 2.000 | 67.000 |

Calculate a new variable, the number of colonies not lost (alive), combine this with the data set and check the descriptive statistics of this variable:

```
> NotLost<-ColOct10 - Loss1011
```

```
> dutch<-cbind(dutch,NotLost)
```

```
> summary(NotLost)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|---------|
| 0.000 | 2.000 | 4.000 | 7.027 | 7.000 | 354.000 |

Box 13. Fitting a quasibinomial intercept-only model for estimation of the overall loss rate.

Estimate the overall loss rate by fitting the null (intercept only) model, omitting any missing values. The overall loss rate is the predicted probability of loss.

```
> dutch.glm1<-glm(cbind(Loss1011,NotLost)~1,
+family=quasibinomial(link="logit"),data=dutch,na.action=na.omit)
```

```
> summary(dutch.glm1)
```

Call:

```
glm(formula = cbind(Loss1011, NotLost) ~ 1,
+ family = quasibinomial(link = "logit"),data = dutch, na.action = na.omit)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|---------|
| -9.0730 | -1.3851 | -0.6480 | 0.8932 | 10.7613 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|------------|
| (Intercept) | -1.30553 | 0.03732 | -34.98 | <2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 3.196163)

Null deviance: 4979 on 1530 degrees of freedom
Residual deviance: 4979 on 1530 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4

```
> dutch.glm1$fitted.values[1]
```

```
1
0.2132358
```

The overall loss rate 0.213 can also be calculated directly:

```
> overall_loss<-sum(Loss1011)/sum(ColOct10)
> overall_loss
[1] 0.2132358
```

Or it can be calculated as the inverse logit of the estimated coefficient (the intercept) of the model. For this the inverse logit function of the bootstrap library is used:

```
> library(boot)
```

```
> inv.logit(coef(dutch.glm1))
(Intercept)
0.2132358
```

Box 14. Calculating a confidence interval for the overall loss rate, using results from Box 12.

To calculate the 95% confidence interval (CI) for the overall loss rate, the standard error of the null model intercept is stored and used in the formula for the normal approximation interval or t-interval, and then the inverse logit of the result is calculated. If using the t-interval, the value of df below is the residual degrees of freedom from the model fitting above. The t-interval is recommended for smaller samples. While it makes little difference for this data, it is used in the further analysis for greater generality.

```
> se.glm1<-0.03732
```

```
> inv.logit(coef(dutch.glm1)+c(-1,1)*1.96*se.glm1)
```

```
[1] 0.2012216 0.2257647
```

```
> inv.logit(coef(dutch.glm1)+c(-1,1)*qt(0.975, df=1530)*se.glm1)
```

```
[1] 0.2012125 0.2257746
```

The second step in model building is the use of explanatory variables. Explanation of the methods for evaluating model fit and determining optimal models is outside the scope of this document. For this example analysis, the variable Region is used. The region variable is one that is largely outside of the beekeeper's control, rather like pesticide use by farmers, yet for various reasons may be associated with the loss rate. In some countries, region may be a substitute for meteorological variables. Boxes 15 to 18 and Fig. 3 show the analysis.

Box 15. Fitting a quasibinomial GZLM with one explanatory factor.

The categorical predictor variable "Region" is added to the model by means of the as.factor command. A continuous predictor variable would be added in the same way, but omitting the use of as.factor(). Significant effects of several regions are found. The intercept corresponds to the first level of the factor, i.e. region 2073.

Note that there are large differences in the number of observations between regions (shown by the tabulation of Region below, giving frequencies for each of regions 2073 to 2086) so differences in loss have to be interpreted cautiously.

```
> region.glm1<-glm(cbind(Loss1011,NotLost)~ as.factor(Region),
+ family=quasibinomial(link="logit"),data=dutch,na.action=na.omit)
```

```
> region.summary<-summary(region.glm1)
>region.summary
```

Call:

```
glm(formula = cbind(Loss1011, NotLost) ~ as.factor(Region),
+ family = quasibinomial(link = "logit"), data = dutch, na.action = na.omit)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|--------|
| -8.0276 | -1.2511 | -0.6123 | 0.8868 | 9.3824 |

Coefficients:

| | Estimate | Std. Error | T value | Pr(> t) |
|-----------------------|----------|------------|---------|-------------|
| (Intercept) | -1.36639 | 0.12719 | -10.743 | < 2e-16 *** |
| as.factor(Region)2074 | 0.41251 | 0.32822 | 1.257 | 0.20902 |
| as.factor(Region)2075 | 0.49789 | 0.18588 | 2.679 | 0.00747 ** |
| as.factor(Region)2076 | -0.16554 | 0.15226 | -1.087 | 0.27713 |
| as.factor(Region)2077 | 0.52147 | 0.20211 | 2.580 | 0.00997 ** |
| as.factor(Region)2079 | -0.11394 | 0.18460 | -0.617 | 0.53720 |
| as.factor(Region)2080 | 0.36701 | 0.15172 | 2.419 | 0.01568 * |
| as.factor(Region)2081 | 0.01721 | 0.18930 | 0.091 | 0.92759 |
| as.factor(Region)2082 | 0.08012 | 0.18101 | 0.443 | 0.65809 |
| as.factor(Region)2083 | 0.05483 | 0.19392 | 0.283 | 0.77742 |
| as.factor(Region)2084 | -0.78537 | 0.39354 | -1.996 | 0.04615 * |
| as.factor(Region)2086 | -0.21667 | 0.19578 | -1.107 | 0.26861 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 3.106299)

Null deviance: 4895.9 on 1509 degrees of freedom
Residual deviance: 4734.8 on 1498 degrees of freedom
(21 observations deleted due to missingness)
AIC: NA

Number of Fisher Scoring iterations: 4

```
>table(Region)
```

Region

| | | | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|------|------|
| 2073 | 2074 | 2075 | 2076 | 2077 | 2079 | 2080 | 2081 | 2082 | 2083 | 2084 | 2086 |
| 116 | 25 | 108 | 327 | 76 | 95 | 208 | 118 | 151 | 115 | 37 | 134 |

Box 16. Testing factor significance, and obtaining confidence intervals for log odds of loss per region.

To determine if the fitted model gives better prediction than the null model, the models are compared by means of an ANOVA. In this case the model with the factor region is a significantly better predictor of loss than the null model:

```
>anova(region.glm1,test="F")
```

Analysis of Deviance Table

Model: quasibinomial, link: logit

Response: cbind(Loss1011, NotLost)

Terms added sequentially (first to last)

| | Df | Deviance | Resid. Df | Resid. Dev | F | Pr(>F) |
|-------------------|----|----------|-----------|------------|--------|--------------|
| NULL | | | 1509 | 4895.9 | | |
| as.factor(Region) | 11 | 161.09 | 1498 | 4734.8 | 4.7143 | 3.912e-07*** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Odds, probabilities and corresponding CIs can be calculated for the factor levels. For example, to get just the predicted loss rate for the region coded 2074:

```
> predict(region.glm1, data.frame(Region=2074),type="response")
```

```
1
0.2781065
```

Or for all the regions, and requesting standard errors for calculation of confidence intervals:

```
> values<-predict(region.glm1,data.frame(Region=levels(as.factor(Region))),
```

```
+ type="link",se.fit=T)
```

```
> logodds<-values$fit
```

```
> lowerlim<-values$fit-qt(0.975, df= 1498)*values$se.fit
```

```
> upperlim<-values$fit+qt(0.975, df= 1498)*values$se.fit
```

Approximate 95% CIs for the log odds of loss per region, given as the lower limit, log odds and upper limit respectively:

```
> cbind(lowerlim, logodds, upperlim)
```

| | lowerlim | logodds | upperlim |
|----|-----------|------------|------------|
| 1 | -1.615869 | -1.3663880 | -1.1169065 |
| 2 | -1.547393 | -0.9538734 | -0.3603536 |
| 3 | -1.134385 | -0.8685001 | -0.6026148 |
| 4 | -1.696136 | -1.5319275 | -1.3677194 |
| 5 | -1.153013 | -0.8449141 | -0.5368147 |
| 6 | -1.742775 | -1.4803233 | -1.2178717 |
| 7 | -1.161645 | -0.9993760 | -0.8371073 |
| 8 | -1.624211 | -1.3491820 | -1.0741526 |
| 9 | -1.538904 | -1.2862660 | -1.0336283 |
| 10 | -1.598711 | -1.3115591 | -1.0244077 |
| 11 | -2.882287 | -2.1517622 | -1.4212369 |
| 12 | -1.875024 | -1.5830575 | -1.2910912 |

Box 17. Obtaining confidence intervals for the odds of loss and the model coefficients.

Approximate 95% CIs for the odds of loss per region, given as the lower limit, odds and upper limit respectively:

```
> odds<-exp(logodds)
> cbind(exp(lowerlim),odds, exp(upperlim))
```

| | odds | | |
|----|-----------|-----------|-----------|
| 1 | 0.1987178 | 0.2550265 | 0.3272907 |
| 2 | 0.2128020 | 0.3852459 | 0.6974296 |
| 3 | 0.3216197 | 0.4195804 | 0.5473785 |
| 4 | 0.1833908 | 0.2161187 | 0.2546871 |
| 5 | 0.3156840 | 0.4295943 | 0.5846074 |
| 6 | 0.1750340 | 0.2275641 | 0.2958592 |
| 7 | 0.3129710 | 0.3681091 | 0.4329611 |
| 8 | 0.1970670 | 0.2594524 | 0.3415871 |
| 9 | 0.2146163 | 0.2763006 | 0.3557140 |
| 10 | 0.2021570 | 0.2693997 | 0.3590091 |
| 11 | 0.0560065 | 0.1162791 | 0.2414152 |
| 12 | 0.1533513 | 0.2053463 | 0.2749706 |

Approximate 95% CIs for the odds ratios per region, relative to the reference region, can be obtained from 95% CIs for the coefficients in the model, which we find first:

```
> coeffs<-region.summary$coef[,1]
> se.coeffs<-region.summary$coef[,2]
> coeffs.lowerlim<-coeffs-qt(0.975, df=1498)*se.coeffs
> coeffs.upperlim<-coeffs+qt(0.975, df=1498)*se.coeffs
> coeffs.CIs<-cbind(coeffs.lowerlim, coeffs, coeffs.upperlim)
```

These are the CIs for the model coefficients:

```
> coeffs.CIs
```

| | coeffs.lowerlim | coeffs | coeffs.upperlim |
|-----------------------|-----------------|-------------|-----------------|
| (Intercept) | -1.61586947 | -1.36638799 | -1.1169065 |
| as.factor(Region)2074 | -0.23130745 | 0.41251455 | 1.0563366 |
| as.factor(Region)2075 | 0.13328402 | 0.49788793 | 0.8624918 |
| as.factor(Region)2076 | -0.46421217 | -0.16553956 | 0.1331331 |
| as.factor(Region)2077 | 0.12503197 | 0.52147392 | 0.9179159 |
| as.factor(Region)2079 | -0.47604280 | -0.11393532 | 0.2481722 |
| as.factor(Region)2080 | 0.06940126 | 0.36701199 | 0.6646227 |
| as.factor(Region)2081 | -0.35411884 | 0.01720602 | 0.3885309 |
| as.factor(Region)2082 | -0.27493680 | 0.08012204 | 0.4351809 |
| as.factor(Region)2083 | -0.32556165 | 0.05482889 | 0.4352194 |
| as.factor(Region)2084 | -1.55732512 | -0.78537421 | -0.0134233 |
| as.factor(Region)2086 | -0.60070769 | -0.21666949 | 0.1673687 |

Box 18. Obtaining confidence intervals for odds ratios and probability of loss, per region.

```
>odds.ratios<-exp(coeffs)
>odds.ratios.CIs<-cbind(exp(coeffs.lowerlim), odds.ratios, exp(coeffs.upperlim))
```

The CIs for the odds ratios (excluding the baseline category) are as follows, given as the lower limit, odds ratio and upper limit respectively :

```
>odds.ratios.CIs[-1,]

as.factor(Region)2074 0.7934955 1.5106115 2.8758163
as.factor(Region)2075 1.1425745 1.6452427 2.3690566
as.factor(Region)2076 0.6286302 0.8474363 1.1424020
as.factor(Region)2077 1.1331847 1.6845087 2.5040662
as.factor(Region)2079 0.6212369 0.8923157 1.2816806
as.factor(Region)2080 1.0718662 1.4434152 1.9437570
as.factor(Region)2081 0.7017916 1.0173549 1.4748125
as.factor(Region)2082 0.7596201 1.0834193 1.5452425
as.factor(Region)2083 0.7221217 1.0563598 1.5453021
as.factor(Region)2084 0.2106989 0.4559490 0.9866664
as.factor(Region)2086 0.5484234 0.8051960 1.1821901
```

Note that the CIs excluding 1 correspond to the significant regions in the summary.region model output. 95% confidence intervals and point estimates of the probability of loss for each region, given as the lower limit, estimated probability and upper limit respectively:

```
> library(boot)

> prob<-inv.logit(logodds)

> cbind(inv.logit(lowerlim), prob, inv.logit(upperlim))

      prob
1 0.16577531 0.2032040 0.2465855
2 0.17546308 0.2781065 0.4108740
3 0.24335270 0.2955665 0.3537457
4 0.15497064 0.1777118 0.2029886
5 0.23993910 0.3005008 0.3689289
6 0.14896082 0.1853786 0.2283112
7 0.23836856 0.2690641 0.3021444
8 0.16462490 0.2060041 0.2546142
9 0.17669472 0.2164855 0.2623813
10 0.16816191 0.2122261 0.2641697
11 0.05303613 0.1041667 0.1944677
12 0.13296150 0.1703629 0.2156682
```

The row numbers 1-12 of the above output correspond with the order of the region names in the Figure below, which shows the estimated probability and its 95% CI for each region. Some of the confidence intervals overlap each other, indicating that there is no significant difference between these pairs of regions in terms of probability of loss. However, significant differences between some groups of regions can be seen. Gelderland, Limburg, Zeeland and Zuid-Holland have a lower probability than Friesland and Groningen and Noord-Brabant.

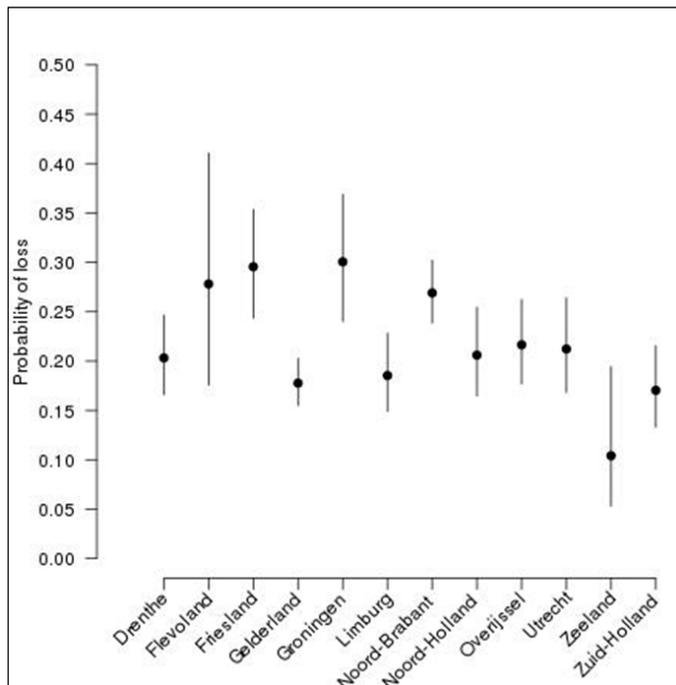


Fig. 3. Estimated probability of loss and 95% confidence interval per region.

11. Conclusions

Estimating colony loss rates reliably depends both on selecting representative samples of beekeepers and also on using suitable methods of estimation. Examining potential risk factors for losses also requires use of suitable statistical methodology. Standardisation of methodology will enable valid comparisons of loss rates to be made across time and between/within countries. Standardisation of terminology avoids confusion and facilitates the required comparisons.

In this manuscript, we have defined terminology associated with colony losses, and have presented the concepts involved in conducting a survey. The latter ranges from choosing the method of data collection to designing the questionnaire and how to select a representative sample and guidelines for choosing the sample size. Practical suggestions and examples are given.

We have examined many of the difficulties of conducting surveys and the important or most likely sources of error in surveys. Being aware of the potential for error makes it more likely that the survey organiser will be careful to avoid practices which are likely to introduce error into a survey, and therefore should achieve a more reliable result. We have also reviewed relevant methods for assessing the quality of the data and for statistical analysis, and have illustrated the more advanced techniques proposed for analysis.

12. Acknowledgements

The COLOSS (Prevention of honey bee COLony LOSSes) network aims to explain and prevent massive honey bee colony losses. It was

funded through the COST Action FA0803. COST (European Cooperation in Science and Technology) is a unique means for European researchers to jointly develop their own ideas and new initiatives across all scientific disciplines through trans-European networking of nationally funded research activities. Based on a pan-European intergovernmental framework for cooperation in science and technology, COST has contributed since its creation more than 40 years ago to closing the gap between science, policy makers and society throughout Europe and beyond. COST is supported by the EU Seventh Framework Programme for research, technological development and demonstration activities (*Official Journal L 412, 30 December 2006*). The European Science Foundation as implementing agent of COST provides the COST Office through an EC Grant Agreement. The Council of the European Union provides the COST Secretariat. The COLOSS network is now supported by the Ricola Foundation - Nature & Culture.

13. References

- AAPOR (2011) *Standard definitions: final dispositions of case codes and outcome rates for surveys*. The American Association for Public Opinion Research. Available at: http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR_Standard-Definitions-Final-Dispositions-of-Case-Codes-and-Outcome-Rates-for-Surveys.pdf
- AGRESTI, A (2002) *Categorical data analysis (2nd Ed.)*. John Wiley and Sons, Inc., Hoboken, NJ, USA.
- BRODSCHNEIDER, R; MOOSBECKERHOFER, R; CRAILSHEIM, K (2010) Surveys as a tool to record winter losses of honey bee colonies: a two year case study in Austria and South Tyrol. *Journal of Apicultural Research* 49(1): 23-30. <http://dx.doi.org/10.3896/IBRA.1.49.1.04>
- CAMPBELL, M J; WATERS, W E (1990) Does anonymity increase the response rate in postal questionnaire surveys about sensitive subjects? A randomised trial. *Journal of Epidemiology and Community Health* 44: 75-76. <http://dx.doi.org/10.1136/jech.44.1.75>
- CHARRIERE, J D; NEUMANN, P (2010) Surveys to estimate winter losses in Switzerland. *Journal of Apicultural Research* 49(1): 132-133. <http://dx.doi.org/10.3896/IBRA.1.49.1.29>
- COX-FOSTER, D L; CONLAN, S; HOLMES, E C; PALACIOS, G; EVANS, J D; MORAN, N A; QUAN, P; BRIESE, T; HORNIG, M; GEISER, D M; MARTINSON, V; VANENGELSDORP, D; KALKSTEIN, A L; DRYSDALE, A; HUI, J; ZHAI, J; CUI, L; HUTCHISON, S K; SIMONS, J F; EGHOLM, M; PETTIS, J S; LIPKIN, W I (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 318: 283-287. <http://dx.doi.org/10.1126/science.1146498>
- DAHLE, B (2010) The role of *Varroa destructor* for honey bee colony losses in Norway. *Journal of Apicultural Research* 49(1): 124-125. <http://dx.doi.org/10.3896/IBRA.1.49.1.26>

- DE LEEUW, E D (2008) Choosing the method of data collection. In *E D De Leeuw, J J Hox, and D A Dillman (Eds). International handbook of survey methodology*. Lawrence Erlbaum Associates, Taylor & Francis Group; New York, USA. pp 113-135.
- DE LEEUW, E D; HOX, J J; DILLMAN, D A (Eds) (2008) *International handbook of survey methodology*. Lawrence Erlbaum Associates, Taylor & Francis Group; New York, USA.
- DOBSON, A J (2002) *An introduction to Generalized Linear Models (2nd Ed.)*. Chapman & Hall / CRC; Boca Raton, Florida, USA.
- EFRON, B; TIBSHIRANI, R J (1994) *An introduction to the bootstrap*. Chapman and Hall / CRC; New York, USA.
- ELLIS, J D; GRAHAM, J R; MORTENSEN, A (2013) Standard methods for wax moth research. In *V Dietemann; J D Ellis; P Neumann (Eds) The COLOSS BEEBOOK, Volume II: standard methods for Apis mellifera pest and pathogen research. Journal of Apicultural Research* 52(1): <http://dx.doi.org/10.3896/IBRA.1.52.1.10>
- GRAY, A; PETERSON, M; TEALE, A (2010) An update on colony losses in Scotland from a sample survey covering 2006-2008. *Journal of Apicultural Research* 49(1): 129-131. <http://dx.doi.org/10.3896/IBRA.1.49.1.28>
- GRAY A; PETERSON, M (2012) SBA member surveys 2010/2011 (main results). *The Scottish Beekeeper* 89(12): 312-316.
- GROVES, R M (2006) Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly* 70(5): 646-675. <http://dx.doi.org/10.1093/poq/nfl033>
- HALL, D B (2000) Zero-inflated Poisson and Binomial regression with random effects: a case study. *Biometrics* 56: 1030-1039.
- HARDIN, J W; HILBE, J M (2007) *Generalized Linear Models and extensions (2nd Ed.)*. Stata Press; Texas, USA.
- HATJINA, F; BOUGA, M; KARATASOU, A; KONTOTHANASI, A; CHARISTOS, L; EMMANOUIL, C; EMMANOUIL, N; MAISTROS, A-D (2010) Data on honey bee losses in Greece: a preliminary note. *Journal of Apicultural Research* 49(1): 116-118. <http://dx.doi.org/10.3896/IBRA.1.49.1.23>
- HOSMER, D W; LEMESHOW, S (2000) *Applied Logistic Regression (2nd Ed.)*. John Wiley and Sons, Inc.; New York, USA.
- KINDT, R; COE, R (2005) Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. *Nairobi: World Agroforestry Centre (ICRAF)*. Available at <http://www.worldagroforestry.org/downloads/publications/PDFs/B13695.pdf>.
- KLEINBAUM, D G; KLEIN, M (2002) *Logistic Regression: a self-learning text (2nd Ed.)*. Springer-Verlag; New York, USA.
- KRZANOWSKI, W J (1998) *An introduction to statistical modelling*. Arnold Publishers; London, UK.
- LEHTONEN, R; PAHKINEN, E (2004) *Practical methods for design and analysis of complex surveys (2nd Ed.)*. John Wiley and Sons Ltd; Chichester, UK.
- LOHR, S L (2010) *Sampling: design and analysis (2nd Ed.)*. Brooks / Cole, Cengage Learning; Boston, USA.
- MCCULLAGH, P; NELDER, J A (1983) *Generalized Linear Models*. Chapman and Hall Ltd; London, UK.
- MUTINELLI, F; COSTA, C; LODESANI, M; BAGGIO, A; FORMATO, G; MEDRZYCKI, P; PORRINI, C (2010). Honey bee colony losses recorded in Italy. *Journal of Apicultural Research* 49(1): 119-120. <http://dx.doi.org/10.3896/IBRA.1.49.1.24>
- MYERS, R H; MONTGOMERY, D C; VINING, G G (2002) *Generalized Linear Models: with applications in engineering and the sciences*. John Wiley and Sons, Inc.; New York, USA.
- NEUMANN, P; PIRK, C W W; SCHÄFER, M O; EVANS, J D; PETTIS, J S; TANNER, G; WILLIAMS, G R; ELLIS, J D (2013) Standard methods for small hive beetle research. In *V Dietemann; J D Ellis, P Neumann (Eds) The COLOSS BEEBOOK: Volume II: Standard methods for Apis mellifera pest and pathogen research. Journal of Apicultural Research* 52(4): <http://dx.doi.org/10.3896/IBRA.1.52.4.19>
- NGUYEN, B K; MIGNON, J; LAGET, D; DE GRAAF, D C; JACOBS, F J; VANENGELSDORP, D; BROSTAUX, Y; SAEGERMAN, C; HAUBRUGE, E (2010) Honey bee colony losses in Belgium during the 2008-2009 winter. *Journal of Apicultural Research* 49(4): 337-339. <http://dx.doi.org/10.3896/IBRA.1.49.4.07>
- OTT, R L; LONGNECKER, M (2009) *An introduction to statistical methods and data analysis. (6th Ed.)*. Brooks / Cole, Cengage Learning; Belmont, CA, USA.
- PETERSON, M; GRAY, A; TEALE, A (2009) Colony losses in Scotland in 2004-2006 from a sample survey. *Journal of Apicultural Research* 48(2): 145-146. <http://dx.doi.org/10.3896/IBRA.1.48.2.11>
- PIRK, C W W; DE MIRANDA, J R; FRIES, I; KRAMER, M; PAXTON, R; MURRAY, T; NAZZI, F; SHUTLER, D; VAN DER STEEN, J J M; VAN DOOREMALEN, C (2013) Statistical guidelines for *Apis mellifera* research. In *V Dietemann; J D Ellis; P Neumann (Eds) The COLOSS BEEBOOK, Volume I: standard methods for Apis mellifera research. Journal of Apicultural Research* 52(4): <http://dx.doi.org/10.3896/IBRA.1.52.4.13>
- RODRIGUEZ, G (2008) Multilevel Generalized Linear Models. In *J De Leeuw; E Meijer (Eds). Handbook of multilevel analysis*. Springer; New York, USA. Chapter 9.
- SAMUELS, M L; WITMER, J A; SCHAFFNER, A (2010) *Statistics for the life sciences (4th Ed.)*. Pearson; USA.
- SÄRNDAL, C-E; SWENSSON, B; WRETMAN, J (1992) *Model-assisted survey sampling*. Springer-Verlag; New York, USA.
- SCHAEFFER, R L; MENDENHALL, W; OTT, L (1990) *Elementary survey sampling (4th Ed.)*. PWS-Kent Publishing Company; Boston, USA.
- SOROKER, V; HETZRONI, A; YAKOBSON, B; DAVID, D; DAVID, A; VOET, H; SLABETZKI, Y; EFRAT, H; LEVSKI, S; KAMER, Y; KLINBERG, E; ZIONI, N; INBAR, S; CHEJANOVSKY, N (2010) Evaluation of colony losses in Israel in relation to the incidence of pathogens and pests *Apidologie* 42: 192-199. <http://dx.doi.org/10.1051/Apido/2010047>

- TOPOLSKA, G; GAJDA, A; POHORECKA, K; BOBER, A; KASPRZAK, S; SKUBIDA, M; SEMKIW, P (2010) Winter colony losses in Poland. *Journal of Apicultural Research* 49(1): 126-128. <http://dx.doi.org/10.3896/IBRA.1.49.1.27>
- TWISK, J W R (2010) *Applied multilevel analysis. A practical guide (4th Ed.)*. Cambridge University Press; Cambridge, UK.
- VAN DER ZEE, R (2010) Colony losses in the Netherlands. *Journal of Apicultural Research* 49(1): 121-123. <http://dx.doi.org/10.3896/IBRA.1.49.1.25>
- VAN DER ZEE, R; PISA, L; ANDONOV, S; BRODSCHNEIDER, R; CHARRIERE, J-D; CHLEBO, R; COFFEY, M F; CRAILSHEIM, K; DAHLE, B; GAJDA, A; GRAY, A; DRAZIC, M; HIGES, M; KAUKO, L; KENCE, A; KENCE, M; KEZIC, N; KIPRIJANOVSKA, H; KRALJ, J; KRISTIANSEN, P; MARTIN-HERNANDEZ, R; MUTINELLI, F; NGUYEN, B K; OTTEN, C; ÖZKIRIM, A; PERNAL, S F; PETERSON, M; RAMSAY, G; SANTRAC, V; SOROKER, V; TOPLOSKA, G; UZUNOV, A; VEJSNÆS, F; WEI, S; WILKINS, S (2012) Managed honey bee colony losses in Canada, China, Europe, Israel and Turkey, for the winters of 2008–2009 and 2009–2010. *Journal of Apicultural Research* 51(1): 100-114. <http://dx.doi.org/10.3896/IBRA.1.51.1.12>
- VANENGELSDORP, D; EVANS, J D; SAEGERMAN, C; MULLIN, C; HAUBRUGE, E; NGUYEN, B K; FRAZIER, M; FRAZIER, J; COX-FOSTER, D; CHEN, Y; UNDERWOOD, R; TARPY, D; PETTIS, J S (2009) Colony Collapse Disorder: A Descriptive Study. *PLoS ONE* 4 (8): e6481. <http://dx.doi.org/10.1371/journal.pone.0006481>
- VANENGELSDORP, D; HAYES, J Jr; UNDERWOOD, R M; PETTIS, J (2010) A survey of honey bee colony losses in the United States, fall 2008 to spring 2009. *Journal of Apicultural Research* 49(1): 7-14. <http://dx.doi.org/10.3896/IBRA.1.49.1.03>
- VANENGELSDORP, D; UNDERWOOD, R M; CARON, D; HAYES, J Jr; PETTIS, J S (2011) A survey of managed honey bee colony losses in the USA, fall 2009 to spring 2010. *Journal of Apicultural Research* 50(1): 1-10. <http://dx.doi.org/10.3896/IBRA.1.50.1.01>
- VANENGELSDORP, D; BRODSCHNEIDER, R; BROSTAUX, Y; VAN DER ZEE, R; PISA, L; UNDERWOOD, R; LENGERICHE, E J; SPLEEN, A; NEUMANN, P; WILKINS, S; BUDGE, G E; PIETRAVALLE, S; ALLIER, F; VALLON J; HUMAN, H; MUZ, M; LE CONTE, Y; CARON, D; BAYLIS, K; HAUBRUGE, E; PERNAL, S; MELATHOPOULOS, A; SAEGERMAN, C; PETTIS, J S; NGUYEN, B K (2012) Calculating and reporting managed honey bee colony losses. In *D Sammatro; J A Yoder (Eds). Honey bee colony health: challenges and sustainable solutions*. CRC Press / Taylor and Francis Group; Boca Raton, Florida, USA.
- VANENGELSDORP, D; LENGERICHE, E; SPLEEN, A; DAINAT, B; CRESSWELL, J; BAYLISS, K; NGUYEN, K B; SOROKER, V; UNDERWOOD, R; HUMAN, H; LE CONTE, Y; SAEGERMAN, C (2013) Standard epidemiological methods to understand and improve *Apis mellifera* health. In *V Dietemann; J D Ellis, P Neumann (Eds) The COLOSS BEEBOOK: Volume II: Standard methods for Apis mellifera pest and pathogen research*. *Journal of Apicultural Research* 52(4): <http://dx.doi.org/10.3896/IBRA.1.52.4.15>
- VEJSNÆS, F; NIELSEN, S L; KRYGER, P (2010) Factors involved in the recent increase in colony losses in Denmark. *Journal of Apicultural Research* 49(1): 109-110. <http://dx.doi.org/10.3896/IBRA.1.49.1.20>
- ZUUR, A F; IENO, E N; WALKER, N J; SAVELIEV, A A; SMITH, G M (2009) *Mixed Effects Models and Extensions in Ecology with R*. Springer; New York, USA.
- Online Supplementary Material: Useful resources for survey design. (http://www.ibra.org.uk/downloads/20130805_1/download)

