WP. 17
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Vienna, Austria, 21–23 April 2008)

Topic (iii): Improvement of quality through data editing

ESTABLISHING E&I CAPABILITY AND BEST PRACTICES AT STATISTICS NEW ZEALAND

Invited Paper

Prepared by Vera Costa and Tracey Savage, Statistics New Zealand, New Zealand

I. INTRODUCTION

1. Editing and imputation (E&I) of sample survey, census and administrative data is one of the key statistical issues that Statistics New Zealand is currently facing. There are four main reasons for this:

- The context in which editing takes place has changed. Where we used to have data edited unit by unit as it was manually captured in-house, this has been replaced by computer assisted interviewing, administrative data, questionnaire scanning, etc.
- Our desire to improve the efficiency of data processing, including E&I, so that more resources can be put into data analysis and dissemination.
- The need to take a wider view of the objectives for E&I, so that they include continuous improvement of our end-to-end business processes and overall data quality, through learning from the E&I process.
- Growing demand from users for access to metadata to support our outputs, particularly microdata outputs.

2. Statistics NZ therefore decided, in 2005, to develop and document E&I objectives and principles for Statistics NZ as a whole, along with a strategy for implementing these across the organisation. The aim of this work was to guide the future development of our E&I processes, methods and systems, and to inform final decisions on the solutions and tools to be used, eventually, across all collections (eg surveys, administrative data) and all types of data (eg cross-sectional, longitudinal, integrated).

3. Statistics NZ's E&I Methodology Network, which was set up in 2004, led the development of the E&I objectives, principles and strategy. The activities and role of this network are presented in section 2.

4. Section 3 of this paper contains a statement of Statistics NZ's strategic goals for E&I. The E&I Principles are covered in section 4. This is followed, in section 5, by the E&I Strategy itself. The experience gained in E&I training through the Pacific Island Attachment Programme, which aims at developing statistical capability in the Pacific Region, is outlined in section 6. Section 7 contains some of our current concerns related with the implementation of these ideas and some conclusions are outlined in section 8.

II. THE EDITING & IMPUTATION METHODOLOGY NETWORK

5. Statistics NZ Methodology Networks are groups within the Statistical Methods Division set up to further the knowledge that Statistics NZ has relating to particular statistical methodologies (eg Data Integration, E&I). The groups are usually made up of about 10 people from Statistical Methods and Subject Matter Areas who fulfil their network role in conjunction with their regular output related work. These methodology networks:

- document existing methodological knowledge to manuals
- do research to advance the methodologies
- make their knowledge available to all of Statistics NZ and the greater NZ Official Statistics System (OSS).

6. In the past, these networks have also been responsible for managing research projects but currently they are being refocused to have a stronger emphasis on strategic direction and capability building within Statistics NZ. Responsibility for methodological research has been moved to the newly created Methodological Development Unit (MDU), which also sits within the Statistical Methods Division.

7. The Editing & Imputation Methodology Network was established in 2004. The specific aims of the E&I Network are to:
   - establish, document and promote E&I best practice across the full spectrum of Statistics NZ data sources
   - provide information to users of output data on E&I procedures and the quality of the data they are using
   - coordinate E&I skill development across Statistics NZ
   - understand and research E&I in terms of:
     - the impact it has on our outputs
     - how editing can help us to improve our survey processes
     - forming better E&I methods and systems
     - international best practice trends.

8. Since 2004, the E&I Network has led the following projects that encompass E&I objectives, principles and strategy:
   - development of an E&I plan template to be used for each family of outputs
   - documentation of the E&I methods used across Statistics NZ, by collection;
   - research into generalised E&I tools
   - development of E&I standards and guidelines
   - offering E&I training.

9. The Network has therefore played a critical role in establishing E&I capability and best practices at Statistics NZ.

III. OBJECTIVES FOR E&I

10. The development of the E&I objectives was led by Statistics NZ's E&I Methodology Network. Their work incorporated international best practice and guidelines for E&I at the time, and drew on the work of other statistical agencies in the area of E&I objectives, principles and strategies, in particular Statistics Canada and the Australian Bureau of Statistics (ABS).

11. Statistics NZ's objectives for E&I, as documented in the Statistics NZ Editing & Imputation Principles & Strategy (Statistics NZ, 2007) are to:

1) Provide users with fit for purpose, plausible data and outputs by the most effective and efficient means.

2) Ensure all users are better informed about the quality of our data and statistical outputs.
3) Continuously improve our end-to-end business processes and overall data quality.

12. In the past, Statistics NZ has not placed much, if any, priority on objective 2 as one of the roles of E&I, and even less on objective 3. However, we need to increase our focus on both of these objectives going forward, as they are critical to achieving objective 1, which is our highest priority as an official statistical agency. This will require an E&I culture change across the organisation.

IV. E&I PRINCIPLES

13. This section lists the E&I principles to be applied across all of our outputs, as documented in the Statistics NZ Editing & Imputation Principles & Strategy (Statistics NZ, 2007). These principles align with the strategic goals for Statistics NZ E&I outlined in section 3.

14. Although some of these principles may seem obvious for E&I experts, having them clearly stated, and approved by a governance board, can be very helpful when discussing possible E&I approaches with people from a different background.

15. The over-arching E&I principle is the key principle that should be foremost in our thoughts when deciding on our approach to E&I. This over-arching principle is supported by ten editing principles and nine imputation principles.

Over-arching E&I principle

Statistics NZ should maintain, wherever possible, the original data provided by the respondent or data supplier.

16. In other words, if the intentions of the respondent or data supplier are clear, we should respect that, and only edit the data if it is absolutely essential from a data quality perspective (eg imputing missing data for key output variables). This is to ensure we do not impose our own expectations on the data, forcing the data to comply with these. Doing this carries the risk of biasing the results towards our view of the world and masking the real world effects or variations in the data, especially as values that fall outside the expected range may, in fact, be correct.

17. This requires a shift in Statistics NZ's approach to E&I of our outputs. We must firstly question whether we need to edit and impute, and, if the answer is yes, justify why. Only then should we put thought and effort into developing an E&I plan, and deciding on the most appropriate E&I methods.

18. An example of this principle in action is the *2006 Census Quality Management Strategy* (Statistics NZ, 2004). The approach to editing outlined in that document is to respect respondent's intentions, except in situations where inconsistent combinations of responses involve impossible situations.

Editing principles

- *All Statistics NZ outputs must have a documented E&I plan.*
- *Ensure that anticipated or likely errors are considered during development or re-development of Statistics NZ outputs and, where the potential impact of these is significant, put processes in place to eliminate or reduce the impact (ie design quality in by focusing on error prevention).*
- *Choose editing methods carefully, considering the type of data to be edited (eg numeric vs categoric) and its end use(s), the availability of auxiliary and historical data, and alignment with Statistics NZ's imputation standards and guidelines, and international best practice.*
- *Ensure that edit specification to identify errors that are not anticipated is done in a systematic and orderly manner, and that the resulting edits are consistent and reproducible.*

- *Attempt as much editing as is practicable at the point of contact with respondents (eg via CAI tools*), to realise the maximum benefit from the initial contact.*
  **CAI is Computer Assisted Interviewing, and includes interviewing in person (CAPI), via telephone (CATI) or via the web (CAWI).*
- *Implement efficient editing techniques, such as macro-editing and selective editing, to ensure that editing resources are used as efficiently and effectively as possible.*
- *Automate the editing process where possible, ensuring a balance with clerical (manual) intervention.*
- *For each Statistics NZ output, keep an editing audit trail, including input (unedited) and output (edited) data, along with editing flags at both the unit record and item level, so that the sources, types and distribution of errors can be monitored.*
- *For each Statistics NZ output, produce, monitor and analyse editing diagnostics and measures to evaluate and understand the editing process, including its cost effectiveness and efficiency.*
- *Ensure all Statistics NZ staff involved in the editing process have access to documentation of and training in editing principles, methods and procedures.*

Imputation principles

- *All Statistics NZ outputs must have a documented E&I plan.*
- *Fields that fail an edit checks and are not resolved through respondent follow-up or other means of verification, may need to be imputed to ensure the data and outputs are "fit for purpose" and plausible (ie replace an erroneous or suspect response only where its retention would render the outputs unfit for use).*
- *Choose imputation methods carefully, considering the type of data to be imputed (eg numeric vs categoric) and its end use(s), the availability of auxiliary and historical data, and alignment with Statistics NZ's imputation standards and guidelines, and international best practice.*
- *The records from which imputed values are calculated or obtained should be similar to or closely resemble the record to be imputed.*
- *The imputed record should closely resemble the failed edit record, thereby preserving as much respondent data as possible.*
- *Imputed records should satisfy all edits.*
- *For each Statistics NZ output, keep an imputation audit trail, including input (unimputed) and output (imputed) data, along with imputation flags at both the unit record and item level, so that the degree, methods and sources of imputation can be monitored.*
- *For each Statistics NZ output, produce, monitor and analyse imputation diagnostics and measures to evaluate and understand the imputation process, including its cost effectiveness and efficiency.*
- *Ensure all Statistics NZ staff involved in the imputation process have access to documentation of and training in imputation principles, methods and procedures.*

19. The detailed implications of each one of these principles must be acknowledged. For example, some of the issues addressed by the second Editing principle are:

- Prevention shifts the focus from the errors themselves to the **source** of the errors. Draw on subject matter expertise, and past experience, to determine the likely sources of error and aim to eliminate or reduce these.
- Consider lessons from previous E&I of similar data / outputs.
- Evaluate, update, or entirely rework the design, layout, and wording of the questionnaire to ensure maximum clarity.
- Build in procedures to prevent the types of errors known to occur at the imaging step, such as optimistic tickmarks, improper date formats, or column-strikethroughs being read improperly by the scanners.
- All data collection processes, methods and systems must be tested prior to going ‘live’. Dual production runs are highly recommended when processes, methods and systems for existing outputs have changed.

20. Once documented, the Statistics NZ E&I principles and strategy were peer reviewed by E&I experts within Statistics NZ, and also from Statistics Canada, the ABS and the Office for National Statistics (ONS), and the resulting feedback incorporated. The final E&I principles and strategy document was then endorsed by the Statistics NZ Statisticians' Advisory Forum.

V. E&I STRATEGY

21. The strategy for achieving the objectives outlined in section 2 includes development, implementation and promotion stages as well as additional support and maintenance in the future. Such tasks are an attribution of the E&I Network.

22. The main deliverables related to the E&I strategy are an E&I Plan template, E&I Standards and Guidelines and some E&I Training. Each one of them is detailed below.

V.1 E&I Plan

23. The process of developing an end-to-end E&I plan for each family of outputs will ensure that:

- issues related to prioritisation and balancing of different user needs and different fitness for purpose standards that users would like Statistics NZ to meet are raised, discussed and debated
- we clearly document the final agreed approach to E&I, along with both the reasoning behind it, and the factors that influenced it
- we manage user expectations about the quality of the final outputs / data they will eventually receive, by sharing the documented E&I plan with users in advance.

24. However, there is no 'one size fits all' E&I plan, as varying approaches need to be used to produce different outputs, depending on the needs of the users, the prioritisation of these needs, and what constitutes fit for purpose outputs for each. So in planning a specific project to develop an E&I plan for a family of outputs, or a project which includes this as a task, the time, thought and effort required to do this should not be underestimated.

25. The development of such plan must be done via discussions, workshops etc. involving input from the Subject Matter Area(s) responsible for the outputs, the users of the outputs – both internal and external, across the Official Statistics System, the Statistical Methods Division, the different Data Collection teams as well as the Product Development and Publishing and the Application Services Divisions. Representatives of each of these groups will need to review, and sign-off on, the final E&I plan.

26. In developing an end-to-end E&I plan for each family of outputs, there are a wide range of factors to consider and information to include:

- user needs – including objectives for the outputs; key data items; quality requirements; etc
- operational constraints – such as the resources, time and systems available; workload; etc
- expectations of the data, at both the unit record and aggregate level
- lessons learnt from past experience.

27. Besides the high-level template document for an end-to-end E&I plan for a family of outputs, supporting material needs to be provided in relation to the practical aspects of such plan development. The supporting material may include documentation of the suggested process for developing the E&I plan, guidelines for running E&I plan workshops, training sessions, and examples of completed E&I plans.

28. The development of an E&I Plan template so far has been based on similar work done at the ABS. Further progress in this area is expected by the middle of the current calendar year, probably in combination with the development of the E&I Standards and Guidelines, described below.

V.2 E&I Standards and Guidelines

29. Having E&I standards and guidelines will ensure that a consistent approach is taken across similar outputs. Much of the subjective judgement around E&I will then be made at a corporate level in the standard, rather than by individuals within individual output areas.

30. It is expected that greater consistency between outputs will lead to a fair savings of time and money. The most obvious saver of time and money would be in training of new or rotated staff; an overall governing set of standards and guidelines for all outputs means that one who is unfamiliar with the specific survey can easily follow the overall reasoning behind the E&I technique, regardless of differing E&I techniques between outputs.

31. Therefore an E&I Standards and Guidelines Project was created within the E&I Network in order to develop and document best practices throughout the Official Statistical System environment in the following areas:

- E&I processes, methods, and tools
- E&I metadata – audit trails, process metadata, diagnostics, performance measures, measures for assessing the impact of E&I etc.

32. The aim is to deliver:

- E&I Standards and Guidelines Document covering E&I processes, methods & tools, and metadata.
- E&I Manual consisting of E&I advice given various survey circumstances. This is meant to be shorter, less technical, and more client-focused than the E&I Standards Document.
- Presentations to promote these standards across Statistics NZ.
- Modular E&I seminar / training series, possibly supported by workshop exercises.

33. The publication of the *Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys* (EDIMBUS Manual, 2007) provides a lot of information that can be used for avoiding ‘reinventing the wheel’, and this is being taken into account in the project development

34. The project is temporarily on hold at the moment. Before its interruption an existing list of E&I methods used in Statistics NZ both for business and social surveys has been updated and this can be of great value when the project resumes. This is planned to happen by the middle of the current calendar year.

V.3 E&I Training

35. Prior to development of the Statistics NZ E&I Principles & Strategy, training in E&I was only provided within the Statistical Methods division – there was no such training available for wider Statistics NZ (eg Subject Matter Areas), unless it was requested and delivered on an ad-hoc basis.

36. However, the E&I Principles state that **all** Statistics NZ staff involved in the E&I process should have access to documentation of and training in E&I principles, methods and procedures. The E&I Strategy also states that training sessions should be included in the material developed to support the practical implementation of E&I Plans. So an introductory E&I training course was developed for Subject Matter Areas implementing an E&I Plan.

37. The course aims to assist with the E&I culture change mentioned above, by:

- providing an overview of & introduction to E&I at Statistics NZ – including standard terminology and definitions, and Statistics NZ's objectives and principles for E&I
- setting the scene for developing an E&I (E&I) plan – including the contents of an E&I plan, and a suggested process for completing one.

38. In the longer term, the E&I Methodology Network is aiming to develop a more comprehensive modular training course in E&I, with a number of the modules available to staff across wider Statistics NZ, rather than just Statistical Methods. This will encompass an introduction to E&I, for a wide audience, along with more in depth modules in E&I approaches, methods and tools, for audiences that need more detailed knowledge.

39. The 2006/2007 Pacific Island Attachment Programme, which is described in more detail in the following section, was the first step in developing this modular E&I training. Further progress in this area is expected by the middle of the current calendar year, in combination of the development of the E&I Standards and Guidelines.

VI. THE PACIFIC ISLAND ATTACHMENT PROGRAMME

40. Each year since 1999, as part of its commitment to developing statistical capability in the Pacific region, Statistics New Zealand has offered a programme of attachments. The programme addresses a specific statistical subject each year with subject matter selection based on key statistical priorities in the region at the time. Applications are then called for from National Statistical Offices across the Pacific Region, with selected participants spending four weeks at Statistics NZ to receive training and undertake a project on the selected topic.

41. The 2006/2007 Pacific Island Attachment Programme was conducted over the period from 16 April 2007 to 11 May 2007. The focus topic was 'data quality and editing strategies encompassing both business and household data', and covered statistical data collected from surveys and from administrative sources. Participants from three countries (Palau, Samoa and Tonga) were accepted.

42. Besides the cultural differences, one reason for concern regarding the programme's organisation was the participants' totally different work environment in comparison to the Statistics NZ one, not only in relation to the size of their organisations but also regarding the sophistication of the systems and methodologies used. Also, the candidates' CVs indicated an heterogeneous range of work experience, which presented an additional challenge.

43. Therefore it was decided to concentrate the training on providing an overview of E&I, focusing on its concepts, rather than on mathematical formulae, and on how the knowledge acquired could be applied when our visitors went back home. According to this approach, the following topics were identified as relevant for the training:

- E&I objectives and principles
- E&I methodology (not too much details or formulas)
- quality and E&I quality indicators
- documentation.

44. Also, a lot of interaction with several subject matter areas was considered very important in order to put the E&I processes into a context and to allow the discussion of several practical issues as seen from different perspectives. This led to workshops involving areas as diverse as the Census, Overseas Trade, Household Economic Survey, Consumer Price Index and the Retail Trade Survey.

45. Overall, the programme schedule can be seen as formed by two types of components: the subject matter (editing and imputation) training and the work on the participant's individual project. The first aspect was covered with the use of interactive presentations and workshops.

46. By the end of the programme each participant made a presentation about their project and what they have achieved through the training. It must be noted that the success of the projects depends a lot on the material brought by the participant, on the setting of realistic expectations about what can be done in the time available as well as on providing the participants most of the methodological knowledge at the beginning of the programme.

47. As seen from the participants' enthusiastic evaluations, they both enjoyed very much and benefited a lot from the programme. Although most of the topics discussed were really new for them, the participants could figure out the advantages of a different approach when implementing E&I processes. Nevertheless, differences in their work culture as well as in the offices' facilities mean that changes are not always going to be straightforward.

48. In addition to providing E&I training for the Pacific Islands participants, this Attachment Programme can be seen as a pilot for future training to be provided to Statistical Methods and Subject Matter Areas in Statistics NZ, as well as to some other agencies within the Official Statistics System. Although the overall structure can be maintained, it is important to make changes in the contents of the presentations / workshops to target the specific needs and level of skills of the future participants.

VII. CURRENT CONCERNS WITH IMPLEMENTATION OF E&I CAPABILITY AND BEST PRACTICE

49. We recognise that deep changes in the culture of our organisation are needed before the best practices described in this paper can be fully implemented. The terminology currently used in many Subject Matter Areas is an additional source of confusion and the exact meaning of different options must be made clear, mainly regarding what constitutes an edit.

50. The high turnover of staff presents a challenge for the development and retention of knowledge in the area. It is recognised that the availability of some training that could be provided at regular basis would be helpful for addressing this issue. The modular approach proposed for the E&I training aims at maximising the resources made available. The E&I Network plays an important role for ensuring that all staff involved with any E&I process has the training needed.

51. Since we have limited resources, it is important to set priorities regarding the work to be done. Currently we are aiming at developing our E&I Standards and Guidelines, an E&I Plan template and creating some E&I training material and course modules. Also, a project is being proposed regarding the implementation of E&I quality indicators so that users can make proper use of the data in subsequent analysis and improvements can be made in our data processing.

VIII. CONCLUSIONS

52. Within a National Statistical Office, setting up dedicated groups to further knowledge relating to E&I, such as Statistics NZ's E&I Network, can play a critical role in establishing E&I capability and best practices. The development of the Statistics NZ objectives, principles and strategy wouldn't have been possible without such group.

53. It is vital to have clear and shared E&I objectives and principles, plus a strategy for implementing these, as this provides direction regarding the areas in which you need to establish E&I capability and best practices.

54. Also, having documented E&I objectives and principles, E&I plans, E&I standards and guidelines, plus training on all of these, is important for developing and retaining E&I knowledge and capability.

55. E&I training should focus firstly on E&I objectives and principles, and the application of these in practice, before delving into statistical / mathematical detail. If we don't get the concepts right in the first place, E&I methods applied are unlikely to achieve the desired result.

56. Finally, a change in E&I culture, through introduction of new E&I objectives and principles, takes time and committed resources to support the change. There is no quick or easy solution.

57. We gratefully acknowledge the discussions, suggestions and peer review work done by our colleagues from several overseas agencies, mainly Statistics Canada, the ABS and ONS.

IX. REFERENCES

EDIMBUS Manual. (2007) Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys. [Available http://edimbus.istat.it]

Statistics New Zealand. (2004). *2006 Census Quality Management Strategy (V4).* Unpublished internal document prepared by Catherine McGill. Wellington: Statistics New Zealand

Statistics New Zealand (2007). *Statistics NZ Editing & Imputation Principles & Strategy (v1.0).*Unpublished internal document prepared by Tracey Savage. Wellington: Statistics New Zealand.