QUESTIONNAIRE TRANSLATION AS A SOURCE OF DATA ERROR
- AN EXAMPLE FROM THE PHILIPPINES

Lee J.-A.[1,2,3], More S.[1], Cotiw-An B.[2]

Les erreurs au niveau des données constituent un menace permanente pour la validité des études épidémiologiques. Un niveau d'erreurs inacceptable lors de la collecte des informations a récemment été découvert en examinant quelque grandes bases de données. Dans le cadre d'échanges inter-culturels, l'étape de traduction puis d'adaptation aux outils de l'épidémiologie est reconnue dans le domaine médical, comme une source d'erreur supplémentaire pour les données. Dans un exemple d'étude transversale chez des petites exploitations porcines aux Philippines, nous avons estimé l'importance d'erreurs de données dues à la traduction d'un questionnaire. Ce questionnaire comportait 120 questions permettant la collecte de l'information puis leur saisie en 361 rubriques. L'évaluation de la traduction du questionnaire a été évaluée par une re-traduction, et par un test pilote. Chaque procédure a été utilisée pour identifier et quantifier les types d'erreurs soit : de traduction littérale, d'omission, ou de mauvaise traduction. Des erreurs ont été relevées dans 44 (36.7%) questions et 147 (40.7%) enregistrements. 84 (57.1%) de ces 147 erreurs ont été identifiées à l'aide de la re-traduction uniquement, 36(24.5%) en utilisant le test pilote, et 27(18.4%) grâce aux deux procédures. La majorité des erreurs découvertes par ces procédures ne l'auraient pas été avec les techniques habituelles de vérification de la base de données. Bien que certaines erreurs puissent encore subsister dans le questionnaire final, il n'y a guère de doute que ces méthodes d'évaluation améliorent en substance la validité de l'étude. Les vétérinaires épidémiologistes devraient utiliser les méthodes les plus usuelles dans le domaine médical en vue de minimiser les erreurs liées à la traduction du questionnaire

INTRODUCTION

Data errors are a continual threat to the validity of veterinary epidemiological studies. Although this has been well understood for many years, it is only recently that studies have quantified the number of data errors in large animal datasets (Naipospos-Hutabarat, 1995; Pollari et al., 1996). These studies have clearly identified levels of error arising from problems in data collection sufficient to adversely affect research results.

A further entry point for data error in epidemiological studies is well recognised in the medical field. In cross-cultural settings, error can occur during the preliminary step of translation and cultural adaptation of an epidemiological tool. Indeed, international medical studies have clearly demonstrated, as a consequence of language and cultural differences, that simple translation of a survey questionnaire does not ensure data quality. For example, Berkanovic (1980) examined the responses of Hispanics to English and Spanish versions of a survey and found that poor Spanish translation prompted misleading answers from the respondents. Guidelines for the translation and cross-cultural adaptation of human epidemiological tools have been established to minimise this type of error. Procedures suggested in these guidelines include forward- and backward-translation, a committee review, and pre-testing.

This paper describes the evaluation of a translated questionnaire that was used to survey smallholder pig raisers in The Philippines. This work was undertaken to quantify the importance of questionnaire translation as a source of data error in the veterinary field.

METHODOLOGY

The research site

The research site is a temperate mountainous region in northern Benguet, The Philippines where the main agricultural activity is small scale vegetable production. Pigs are the major livestock species in this region, and small herds of pigs are kept by many households. These animals are raised to supplement household income, and for slaughter during religious and other special occasions.

The majority of people in this region belong to one indigenous highland tribe with a common dialect and customs. In 1990, 85% of people in the study area reported their mother tongue as the regional Malayo-Polynesian language of Kankana-ey.

General aspects of the study design

A cross-sectional study of smallholder pig raisers was conducted between April and October 1996. The objective was to record the health and productivity of smallholder pigs and to identify constraints and opportunities for pig production under the existing management system. In this study, a questionnaire was used to collect data on measures of sow and grower pig productivity and potentially related factors at pig raiser, herd and pig levels. Randomly selected smallholder pig raisers were interviewed once using a Kankana-ey translation of this questionnaire.

---

[1] School of Veterinary Science, University of Queensland, Brisbane, Australia.
[2] College of Veterinary Medicine, Benguet State University, La Trinidad, Benguet, The Philippines.
[3] 1994 Inaugural Dunlop Asia Fellow.

The complete questionnaire contained three sections and 120 questions. Some questions gathered a number of separate pieces of data, and a computerised database, which was constructed in Paradox for Windows version 5.0 (Borland International Inc., Scotts Valley, California, USA) to enter and manage these data, had 361 data entry fields. The first section (PIG RAISER) contained 10 questions on the household and pig herd and provided data for entry in 41 data entry fields in the computerised database. The second and third sections (GROWER and SOW) contained questions about pig herd management and individual grower pigs and sows. There were 53 and 57 questions, and 143 and 177 data entry fields in the grower and sow sections, respectively.

Construction and evaluation of the questionnaire

Five Australian and Filipino veterinarians constructed the questionnaire in English after a thorough review of published international and Filipino literature on sow and grower pig productivity. A forward-translation of this document into Kankana-ey was performed by a bilingual animal scientist who was a native Kankana-ey speaker. The accuracy of the translation was evaluated using two procedures: 1) back-translation into English and 2) a pilot test.

1. Back-translation into English

Back-translation into English was performed by a bilingual Kankana-ey livestock technician who had not seen the English version. Corresponding questions in the original English and back-translated English versions were examined by the senior author and the two bilingual translators to detect differences in the wording and meaning of questions. Discrepancies were categorised as either errors of literal translation, omission and/or mistranslation (Table I) and quantified according to the percentage of questions and data entry fields affected. Using this information, the Kankana-ey version was then revised to ensure equivalent meaning with the original English version of the questionnaire.

2. Pilot test

The revised Kankana-ey questionnaire was piloted during interviews with five smallholder pig raisers similar to raisers chosen for the official survey. Answers given by participants during the pilot test were compared with the type of answers expected for each question to identify errors. The identified errors were categorised and quantified as described previously.

Table I
Definition of errors identified in the study

Literal translation	Direct transfer of a word, phrase or sentence construction that changed or confused the meaning of a question and the intent of the corresponding data entry field or fields
Omission	Loss of a word or phrase that changed the meaning of a question or the intent of the corresponding data entry field or fields
Mistranslation	Incorrect translation of a word or phrase that changed the meaning of part or all of a question and the intent of the corresponding data entry field or fields

RESULTS

Forty-four (36.7%) of the 120 questions in the questionnaire were shown to be affected by error in this study (Figure 1). Twenty-six (59.1%) of the 44 error-affected questions were identified using back-translation alone, 14 (31.8%) using pilot testing alone, and four (9.1%) questions contained errors identified by both procedures. These questions obtained data for entry in 147 (40.7%) of the 361 data entry fields. Eighty-four (57.1%) of the 147 error-affected data entry fields were identified using back-translation alone, 36 (24.5%) using pilot testing alone, and 27 (18.4%) by both procedures.

Of the 147 error-affected data entry fields, 62 (42.2%) were affected by literal translation error alone, 10 (6.8%) by omission alone, 73 (49.7%) by mistranslation alone, and two (1.3%) by both literal translation and mistranslation errors (Figure 2). Table II presents examples of questions from the questionnaire affected by each error type.

DISCUSSION

These results clearly indicate that a considerable proportion of the data collected using the translated questionnaire would have certainly been invalid, had evaluation of the questionnaire not been performed. The majority of the identified data errors would not have been detected using routine epidemiological methods of data checking such as random error checks and computer data entry checks. Although a number of errors may have remained in the final questionnaire following evaluation, there is no doubt that the validity of the survey was substantially improved as a result of this process.

In this study, forward-translation was performed by a bilingual animal scientist, and not a qualified translator. This may in part explain the high incidence of errors that were identified during the evaluation of the questionnaire. In many cross-cultural studies in the veterinary field, questionnaires have frequently been translated 'on-the-spot' at the time of interview by relatively poorly qualified research assistants. In such situations, even higher rates of data errors might be expected.

It is clear that the stage of transfer from one language to another is an additional entry point for data error in the data collection process. Back-translation and pilot tests are used widely in international human epidemiological studies to ensure the linguistic and conceptual equivalence of cross-cultural research tools. Veterinary epidemiologists should also consider translation as a potential source of data error and implement proven quality control procedures to overcome this problem when conducting studies in a second language setting.
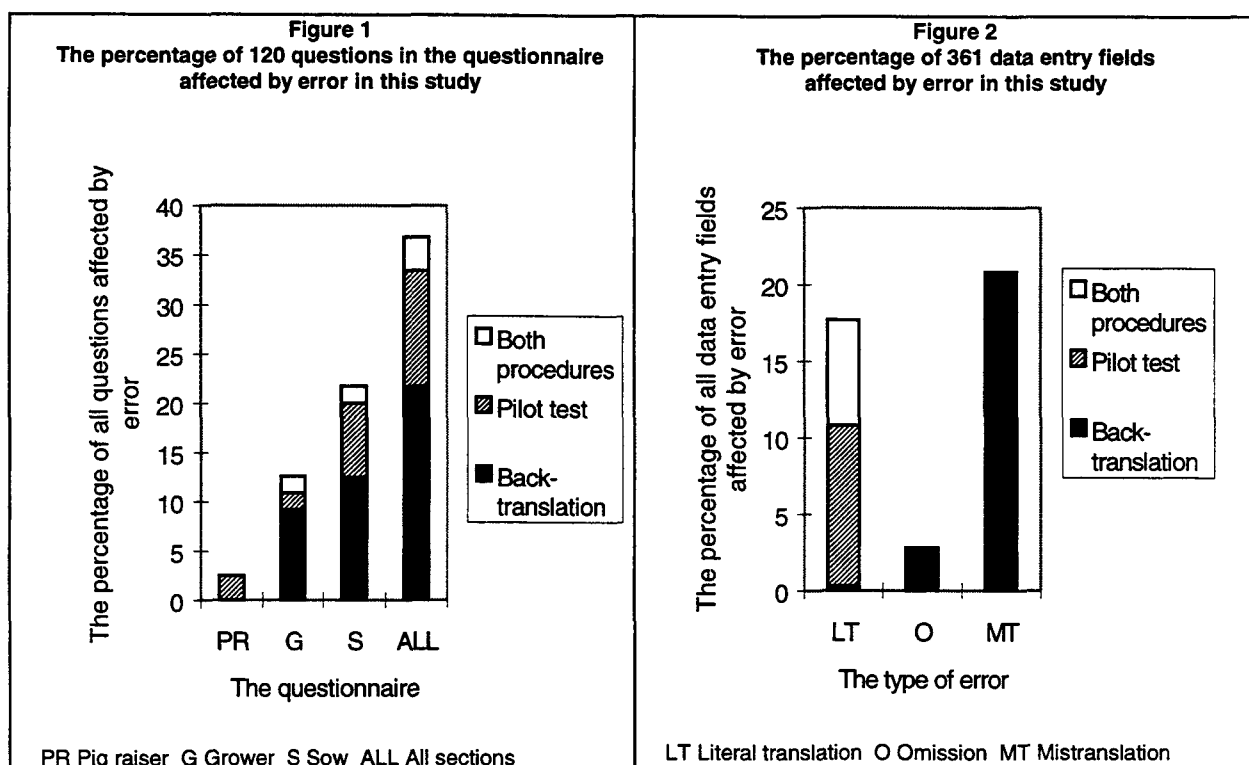
Figure 1
The percentage of 120 questions in the questionnaire affected by error in this study

PR Pig raiser G Grower S Sow ALL All sections

Figure 2
The percentage of 361 data entry fields affected by error in this study

LT Literal translation O Omission MT Mistranslation

Table II
Examples of questions affected by the three types of error

	English version	Kankana-ey version	Effect of error on data
Literal translation	Did the piglets have access to **creep feed**?	Were the piglets given **creep feed**?	Missing or invalid data as no equivalent Kankana-ey term for 'creep feed'
Omission	**In the last 12 months** have you sought advice **from outside the household** on issues of grower management?	Did you ask for help with growers?	Extra information due to loss of reference time period and advice source outside household
Mistranslation	Did piglets have access to **dirt**?	Were the piglets in **dirty matter**? eg. soil, faeces	Extra information due to expanded meaning of dirt

REFERENCES

Berkanovic, E., 1980. The effect of inadequate language translation on Hispanics' responses to health surveys. American Journal of Public Health 70(12), 1273-1276.

Guillemin, F., Bombardier, C., Beaton, D., 1993. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. Journal of Clinical Epidemiology 46(12), 1417-1432.

Naipospos-Hutabarat, T.S.P., 1995. The development of an animal health information system for Indonesia. UnpublishedPhD thesis, Massey University, New Zealand.

Pollari, F.L., Bonnett, B.N., Allen, D.G., Bamsey, S.C., Martin, S.W., 1996. Quality of computerised medical record abstract data at a veterinary teaching hospital. Preventive Veterinary Medicine 27, 141-154.