taps

Working Paper Series

Interviewer bias: Lessons from panel and cross-sectional surveys from a native Amazonian society

Victoria Reyes-García, Vincent Vadez
Ricardo Godoy, Elizabeth Byron, Tomás Huanca,
William R. Leonard

Tsimane' Amazonian Panel Study Working Paper # 15, Nov-2005

Interviewer bias:

Lessons from panel and cross-sectional surveys from a native Amazonian society

Victoria Reyes-García, Vincent Vadez

Ricardo Godoy, Elizabeth Byron, Tomás Huanca,

William R. Leonard

Address correspondence to:

Victoria Reyes-García (c/o Ricardo Godoy)

Sustainable International Development Program MS 078

Heller School for Social Policy and Management

Brandeis University, Waltham, MA 02454-9110, USA

Tel. 1-781-736-2784

E-mail: vreyes@brandeis.edu

Keywords: interviewer bias, panel survey, cross-sectional survey, Tsimane' (Bolivia)

Interviewer bias:

Lessons from panel and cross-sectional surveys from a native Amazonian society

Abstract: Data collected through surveys is subject to interviewer bias, which may worsen when more than one researcher collects data. Interviewer bias can account for a large share of the variation in data. In this article we estimate differences between two pairs of interviewers: one pair conducted five consecutive quarterly surveys, and another pair conducted a cross-sectional survey. Both pairs worked among Tsimane' Amerindians in the Bolivian Amazon. A t-test analysis of 14 variables showed statistically significant differences in two of the variables collected during the panel survey, and in seven of the variables collected during the cross-sectional survey. Additionally, the difference between interviewers was higher among those in the cross-sectional survey than among those in the panel survey. A detailed analysis of one of the variables over the five quarters of the panel survey showed that differences between interviewers were high and statistically significant during the first quarter, and decreased and became statistically insignificant after the second quarter. The findings suggest that, although interviewers can introduce errors in measurement, one can attenuate the errors by improving interviewer training, team exchanges, and ethnographic knowledge. The finding have implications for researches using surveys in which more than one person collects data, and for one-time large cross-sectional surveys where researchers have limited ethnographic knowledge.

Author's statement

Victoria Reyes-García /a/, Vincent Vadez /a/ Ricardo Godoy /a/, Elizabeth Byron /b/, Tomás Huanca /a/, William R. Leonard /c/

/a/ Sustainable International Development Program, Heller School for Social Policy and Management, Brandeis University, Waltham, MA 02454-9110, USA

/b/ International Food Policy Research Institute, Washington, DC 20006-1002, USA

/c/ Department of Anthropology, Northwestern University, Evanston, IL 60208, USA

Acknowledgements

This research was funded by two grants from the National Science Foundation (SBR-9731240 and SBR-9904318), a grant from the John D. and Catherine T. MacArthur Foundation, and a grant from the Conservation, Food and Health Foundation. We would like to thank people who participated in data collection: L. Apaza, E. Pérez, S. Tanner, Z. Foster, B. Sandstrom, A. Yakhedts, Y. Gutierrez, and M. Alvarado. We thanks the following people and institution for logistical support during fieldwork: J. Pache, A. Nate, D. Ista, P. Pache, E. Tayo, L. Pache, N. Canchi, M. Roca, and Gran Consejo Tsimane'.

Interviewer bias: Lessons from panel and cross-sectional surveys from a native Amazonian society

Introduction

The collection of data through structured surveys is subject to at least two main sources of error. The first measurement error is generated by a lack of accuracy in informant's responses. Reasons for lack of accuracy range include faulty recall and mistrust of interviewers (Bernard et al. 1984). The second source of error stems from how interviewers interpret the answers of informants. The second type of error can worsen when more than one interviewer collects data.

Researchers have long noticed that interviewers can differ in how they record and interpret the same phenomena (Cohen 1960, 1968; Krippendorff 1980; Lewis and Cook 1968; Rosenthal 1966; Sipes 1976; Yelton, Wildman, and Erickson 1977). Studies to test whether different people judge the same phenomena in similar ways consists of confronting coders with the same phenomena and then comparing their answers (Braakhius et al. 2003; Lacy and Riffe 1996; Lombard, Snyder-Duch, and Bracken 2002; Peter and Lauf 2002; Ryan 1999). In disciplines such as medicine, psychology, and mass communications it is becoming more and more common to report inter-coder level of agreement to assess the reliability of the data presented (Riffe and Freitag 1997; Yeaton and Wortman 1993).

In anthropology, differences in judgments of the same phenomena by different observers were also noticed long ago (Bernard et al. 1986; Heider 1988). The problem of differences in coding have mainly caught the attention of researchers interested in cross-cultural research using the Human Resources Area Files (Beierle and Witkowski 1974;

Bradley 1989; Kay 1957), and researchers using text analysis (Carey, Morgan, and Oxtoby; Jehn and Doucet 1997; Krippendorff 1980; Kurasaki 2000; Ryan 1999). Few anthropologists have assessed the convergence of answers between many interviewers collecting comparable data from different persons in the same sample.

Drawing on repeated observations from the same subjects (or panel survey) and on a one-time survey (or cross-sectional survey) collected from Tsimane' Amerindians in the Bolivian Amazon, in this paper we estimate the reliability of data collected by two pairs of interviewers and make suggestions for how to improve reliability across interviewers. To accomplish the goals, we follow two steps. First, we use bivariate analysis to estimate differences in the data recorded between two pairs of interviewers, one pair collecting panel data, and one pair collecting cross-sectional data. We expect smaller differences between interviewers interacting over time than between interviewers who do not interact over time. We find more differences between interviewers working in the cross-sectional survey than between interviewers working in the panel survey. Second, using data from the panel survey, we analyze the evolution of differences over time. We find that differences in the panel narrow over time, suggesting that researchers can overcome the bias of interviewers by improving training and interviewers exchanges.

Method

Background of the study and the people: The data used for the analysis comes from a bio-cultural study among Tsimane' Amerindians that started in 1999 and continues to the present. The study centers on the effects of markets and modernization on the conservation of natural resources and on well being in a foraging and horticultural population. Recent dissertations and articles contain ethnographic descriptions of the population and the principal research findings from this and other studies on the Tsimane' (Apaza et al. 2003; Byron 2003; Gurven, 2004; Huanca 1999; Kirby et al. 2002; Reyes-García 2001; Reyes-García et al. 2003, 2004; Vadez et al. 2004). Below we summarize some of the main characteristics of the Tsimane'.

The Tsimane' are a foraging and horticultural group of ~8,000 people living in ~100 villages in the Bolivian Amazon (Instituto Nacional de Estadística 2003). The Tsimane' territory spreads from the foothills of the Andes to the northeast, reaching the edges of the Moxos savanna. Until about 1950, Tsimane' were a hunting and gathering society, but at present they show large variation in lifestyles. Some Tsimane' in the upper rivers are nomadic, live in small communities without schools, are monolingual in Tsimane', and rely on hunting, fishing, plant foraging, shifting cultivation, and barter. In areas closer to roads and market towns, Tsimane' are bilingual in Tsimane' and Spanish, live in large settlements reachable by road, and are more likely to live in villages with schools. Besides subsistence agriculture and foraging, those Tsimane' also work for wages and grow rice as a cash crop (Vadez et al. 2004).

Sample: We draw on data collected through a panel and a cross-sectional surveys. In both cases, we used structured interviews to gather the same type of data. In the surveys we focused on social, demographic, and economic data, use of natural resources, subjective health, and nutritional status.

The panel survey consisted of five repeated quarterly surveys done on all persons over 15 years of age in two villages. An average of 139 adults (70 men and 69 women) participated in the study each quarter. Five permanent researchers living in the villages did the surveys with the help of translators. Researchers in a village lived in the same house and shared translators. Although we collected data over six quarters, we used the first quarter to pilot test methods and excluded the first quarter from the analysis. The analysis presented here covers five consecutive quarters lasting 18 months (August 1999-November 2001). During the duration of the study, the five permanent researchers met at least twice each quarter to discuss how to resolve problems that arose during the surveys.

The cross-sectional survey was done among 511 households in 59 villages over the entire Tsimane' territory. Within each village we selected at random 12 households for interviews. In each household selected, we picked at random either the male or the female head of the household for the interview. The cross-sectional survey lasted five months and took place between July and November, 2000. Eight students from American and Bolivian universities did the surveys with the help of translators. Permanent researchers working in the panel survey trained the interviewers doing the cross-sectional survey. Training took place in the villages. Once trained, interviewers were paired to help each other during the survey. After the training ended and once the cross-sectional survey started, there was no further formal discussion of problems encountered by interviewers during the survey.

After selecting households for the interviews in the village, interviewers divided the households based on the convenience of the researchers and on the availability of the subjects. The matching of interviewer and subject raises the possibility that differences between interviewers might stem from lack of random assignment between interviewers and subjects. For instance, suppose one interviewer may have preferred to visit only rich households, and the other interviewer may have preferred to visit only poor households. If so, then differences in, say, the mean income of subjects in the sample of each interviewer might be biased by lack of random assignment between subjects and interviewers rather than by disagreement between interviewers. Since convenience drove the matching of interviewers and subjects in *both* the cross-sectional and panel survey, there is no reason to think that errors from interviewer self-selection would affect one type of survey more than the other. However, since we only analyzed data collected by one pair of interviewers in the panel survey and by one pair of interviewers in the cross-sectional survey, we cannot test whether differences stem from lack of random assignment.

Variables: In both surveys we collected data on a total of 14 variables that covered socio-demographic characteristics of subjects (sex, age, schooling, Spanish fluency) and economic characteristics of subject's households (cash income, wealth, credit), use of natural resources by the household (fish and game consumption), adult self-perceived health (health), and anthropometric indices of nutritional status of adult men (stature , weight, biceps and triceps) (Table 1).

INSERT TABLE 1 ABOUT HERE

Socio-demographic variables: **(i)** Sex. **(ii)** Age. At least 45 percent of adults in the cross-sectional survey and 35 percent of adults in the panel survey did not know their exact age. The percent may be higher because few people have birth certificates. In those cases, interviewers estimated the age of the informant in a five-year range (e.g. between 30 and 35 years old) and assigned the lowest value of the range (i.e. 30 for people in the range 30-35). **(iii)** Schooling was defined and measured as the maximum grade completed in school; values ranged from 1 to 13. **(iv)** Spanish fluency was defined as the ability to speak Spanish, and was measured by the interviewer at the moment of the survey. The variable was coded as follow: 0 no ability, 1 some ability, and 2 fluent.

Economic variables**:** **(i)** Cash income included different sources of income (salaries, sale, and barter) from all adults in the household for the 30 days before day of the interview. **(ii)** Wealth referred to the total monetary value of 13 commercial assets (i.e. bicycles) and five traditional physical assets (i.e. canoes) owned by the household. **(iii)** Credit was defined as the total value borrowed by adults in the household, including credit in cash and in kind, and money given in advance for labor or products for the 30 days before the day of the interview. **(iv)** Fish consumption refers to the kilograms of fish brought to the household during the 24 hours before the survey. **(v)** Game consumption refers to the kilograms of game brought to the household during the 24 hours before the survey.

Self-perceived health and nutritional status variables: **(i)** Health was measured as the number of days the informant reported having any ailments during the seven days before the interview. We measured nutritional status by taking anthropometric measurements of subjects. For brevity, when using anthropometric data we restrict the analysis only to adult men. We measured the following anthropometric variables: **(ii)** weight in kilograms without shoes, **(iii)** stature in centimeters without shoes, hair clips, or head gear, **(iv)** triceps skinfolds, and **(v)** biceps skinfolds. We used Lange calipers to measure skinfolds and we used the protocol of Lohman et al. (1988) to measure stature and weight.

Analysis: We tested the reliability of data collected by the two pairs of interviewers: one pair living and working in one village during the panel survey, and one pair working in the cross-sectional study. Since a total of five interviewers collected data in the panel survey and eight interviewers collected data in the cross-sectional survey, we had to decide how to select a sample of pairs to analyze interviewer bias. We decided to select the two pairs who had the highest number of comparable observations. For example, one of the pairs in the panel survey decided to split data collection from the same subjects; one interviewer collected information on self-perceived health, stature, and weight, and the other interviewer asked the same subject about fish and game consumption, and measured biceps and triceps. The division of labor between these interviewers made it impossible to use their data for the analysis. Most of the interviewers of the cross-sectional survey were students from universities in the United States, and only worked in the cross-sectional survey during two of the five months that the survey lasted. For the cross-sectional survey we selected a pair of interviewers composed of Bolivian researchers who collected data during the entire duration of the cross-sectional survey.

To analyze the data collected by each pair of interviewers, we compare and test for differences in means between the two pairs. To compare responses by the two interviewers in the panel survey, we use the average of the data collected over five quarters. To compare responses by the two interviewers in the cross-sectional survey, we exclude data from villages visited by only one interviewer in the team. We estimate differences for the 14 variables described in the previous section. For the analysis, we focus not only on the level of statistical significance when comparing the mean difference between or within pair of interviewers, but also on the absolute value of the difference. We do so because we care not only about the statistical significance of the difference, but also about the economic or social significance of the differences. For each variable, we first estimate the mean for each interviewer, and we then estimate the difference in the means between the two interviewers in each team. We also estimate the error as a percentage of the mean value of the variable. In the discussion that follows we focus only on differences that are statistically significant at the 90 percent confidence level or higher *and* only when the difference between means was over 10 percent of the average value between the two interviewers.

Results

Comparison between interviewers within and across teams: In the panel survey, interviewers differed in their measure of two of the 14 variables (Table 2); in the cross-sectional survey they differed in seven of the 14 variables (Table 3). Put in percentage terms, interviewers in the panel and in the cross-sectional survey differed in 14 percent and in 50 percent of the variables they measured.

INSERT TABLE 2 ABOUT HERE

In both the panel and the cross-sectional surveys, we found no significant difference between the two pairs of interviewers in their measure of demographic variables (sex and age). In the panel, we did not find differences between interviewers in their measure of variables for human capital (schooling and Spanish fluency). In the cross-sectional survey, the average level of schooling was similar for both interviewers, but the mean for the variable that measured the ability to speak Spanish was 56 percent higher for one interviewer than for the other ($p<0.0001$) (Table 3). That is, one interviewer found subjects being, on average, fluent in Spanish, but the other found them, on average, to be only moderately fluent.

INSERT TABLE 3 ABOUT HERE

In the panel survey, we found no difference between interviewers on variables measuring economic status or the use of natural resources. In the cross-sectional survey, data collected on cash income by one interviewer was 69 percent higher than data collected by the other interviewer ($p<0.0003$). The difference in values represents more than one month's salary, so the difference is not only statistically significant but also carries economic significance. In the cross-sectional survey we also found that the mean value of the following variables differed between interviewers: household wealth (30 percent; $p<0.003$), credit (166 percent; $p=0.05$), and fish consumption (75 percent; $p<0.02$).

In the two teams, we found that the only other variable where interviewers differed was in the measure of skinfolds. In the panel survey, we found a difference between interviewers that accounted for 72 percent of the value of biceps and for 45 percent of the value of triceps, and the difference was statistically significant at the 99 percent confidence level. In the cross-sectional survey, we also found significant differences between interviewers measuring biceps (79 percent ; $p<0.0001$) and triceps (77 percent; $p<0.0001$).

We found additional differences when comparing data across teams. Out of the 14 variables, we found that in 12 cases the value of the difference of means between interviewers was higher among the interviewers in the cross-sectional survey than among the interviewers in the panel survey. Only in one variable (age) did we find that the value of the mean was higher for interviewers in the panel, and in another variable (health) values were equal. Further, in the cross-sectional survey, for all the variables in which we found differences between interviewers, one of the interviewers always had a higher mean than the other, suggesting a general overestimation by one interviewer compare with the other.

Interviewer's reliability over time: Why would interviewers working in the panel survey show less differences than interviewers working in the cross-sectional survey? One possible explanation is that as the panel unfolds, a pair of interviewers working in the same village gain more ethnographic understanding of the people whom they interview, and get to discuss with each other doubts and reach consensus on how to resolve the doubts, thus increasing convergence in the data they collect. To explore the idea, we analyze the evolution of differences between interviewers doing the panel survey over the five quarters of the survey.

When examining data over time, we found that differences between interviewers decreased over time. Panel data for the variable cash income suggests that the difference between the two interviewers was highest during the first quarter than the mean cash income value for that quarter (dif= 58.8; p=0.01) (Figure 1). Additionally, during the first quarter we found no overlap in the standard deviation for the variable cash income between the two interviewers, suggesting important differences in the values collected by the two interviewers. During the second quarter the absolute value of the difference between interviewers decreased from 58.8 to 42.2 and lost its statistical significance. In later quarters the difference of means continued to decrease both in absolute and in relative value. During the last two quarters, the absolute difference between data reported by the two interviewers was less than 10 percent of the mean value, and the confidence interval of data collected by the two interviewers overlapped.

INSERT FIGURE 1 ABOUT HERE

Another possible reason for convergence in the information collected by the two interviewers in the panel would attribute convergence to time-in-sample bias or panel conditioning. The biases arise when subjects in a panel do one or more of the following: 1) learn from answering repeated questions or doing repeated tasks and thus supply more accurate information, 2) feel more comfortable with interviewers and thus provide more accurate information, or 3) get tired of answering question and, in response to the burden of the survey, provide terse answers or answers that do not lead to follow-up questions. Points (1)-(3) would increase convergence, but points (1)-(2) would enhance accuracy and point (3) would decrease it. We doubt panel conditioning explains the convergence of answers we observed in the panel because both interviewers worked with the same people during the same time. Subjects who learned how to supply more/less accurate answers would have supplied them to both interviewers; hence, difference between interviewers cannot stem from conditioning bias.

Discussion and conclusion

Interviewer bias can be large and significant. For example, in the cross-sectional survey we found that interviewer bias represented 69 percent of the mean value of cash earning during the last month, 166 percent of the value of the mean amount of credit received by households in the last month, 72-79 percent of the mean value of biceps skinfolds, and 46-77 percent of the mean value of triceps skinfolds. As the analysis of panel data over time suggests, differences declined with training and discussion among interviewers. In the panel survey, where interviewers had good ethnographic knowledge, many interactions with each other, long training, and long presence in the research site, we found fewer differences between interviewers than in the cross-sectional survey. In the panel survey, interviewers learned over time and exchanged information about problems they encountered, generated solutions of mutual accord, and thus achieved convergence in the data they collected.

The analysis of variables across pairs of interviewers suggests that some variables are less sensitive to interviewer's measurement errors than other variables. We did not find any statistically significant differences in the collection of demographic variables between the panel and the cross-sectional survey. We found differences in human-capital variable were interviewers had to interpret data from subjects. For instance, we found differences when interviewers had to assess the language fluency of subjects. Variables related to economic status and the use of natural resources were sensitive to interviewer errors in the cross-sectional survey, but not in the panel survey, suggesting that long-term presence in the field improves accuracy in the data, probably because interviewers help reduce informant's recall errors through prompts. Interviewers in the panel survey lived continuously in the village, and they participated and were aware of many of the events that took place in the village (i.e. arrival of a trader, hunting expeditions). They used the information to prompt subjects in the surveys. Interviewers in the cross-sectional survey remained no more than five days in the village, and did not have enough time to be aware of ordinary activities in the village. Variables used to measure self-perceived health and nutritional status were, in general, comparable across interviewers, except for the measure of biceps and triceps skinfolds. A possible explanation for lower accuracy in skinfolds is that they require more training to use calipers than the standard measures of stature and weight.

In sum, our findings suggest that interviewer bias can account for an important part of measurement error in survey data. Anthropologists have developed techniques to increase ethnographic knowledge of societies, but this requires time and resources, and the trade-off is a smaller sample size. Cross-sectional surveys can reach more households, but the trade-off is less reliable data. An alternative model to reach large sample with accurate data is to combine panel and cross sectional data, with panel data colleted before cross-sectional data. In this way, the experience gained during the panel survey can be passed on to the interviewers of the later cross-sectional survey. To reduce measurement error in one-time cross-sectional surveys with many interviewers, researchers should rely on initial training and subsequent meetings as the study unfolds so interviewers can discuss the solution to mutual problems. In many cross-sectional surveys it is standard to have training up-front before the study begins, but not as the study unfolds. Based on the results of the panel study, these simple steps could enhance the accuracy of data collected in cross-sectional surveys.

References Cited

Apaza, Lilian, R. Godoy, D. Wilkie, E. Byron, T. Huanca, W. Leonard, E. Perez, V. Reyes-García, and V. Vadez
2003 Markets and the use of wild animals for traditional medicine among the Tsimane' Amerindians, Bolivia. Journal of Ethnobiology
23(1):47-64.

Beierle JM, and S. Witkowski
1974 HRAF coding reliability. Behavior Science Research
9(1):57-65.

Bernard HR, P.D. Killworth, D. Kronenfeld, and L. Sailer
1984 The problem of informant accuracy. The validity of retrospective data. Annual Review of Anthropology
13:495-517.

Bernard, HR, P.J. Pelto, O. Werner, J.S. Boster, A.K. Romney, A. Johnson, C.R. Ember, and A. Kasakoff.
1986 The construction of primary data in cultural anthropology. Current Anthropology
27(4):382-96.

Braakhius, A.J., K. Meredith, G.R. Cox, W.G. Hopkins, and L.M. Burke
2003 Variability in estimation of self-reported dietary intake data from elite athletes resulting from coding by different sports dietitians. International Journal of Sport Nutrition and Exercise Metabolism
13(2):152-65.

Bradley, C.
1989 Reliability and inference in the cross-cultural coding process. Journal of Quantitative Anthropology
1(4):353-71.

Byron, Elizabeth
2003 Market Integration and Health: The Impact of Markets on the Nutritional Status, Morbidity, and Diet of the Tsimane' Amerindians of Lowland Bolivia. PhD dissertation
University of Florida, Gainesville.

Carey, J.W., M. Morgan, M.J. Oxtoby
1996 Intercoder agreement in analysis of responses to open-ended interview questions: examples from tuberculosis research. Cultural Anthropology Methods
8(3):1-5.

Cohen, J.
1960 A coefficient of agreement for nominal scales. Education and Psychological Measurement 20:37-48.
1968 Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin 70(4):213-20.

Gurven, Michael
2004 Economic games among the Amazonian Tsimane': exploring the roles of market access, costs of giving, and cooperation on pro-social game behavior. Experimental Economics 7:5-24.

Heider, K.G.
1988 The Rashomon Effect. When ethnographers disagree. American Anthropologist 90(1):73-81.

Huanca, Tomas
1999 Tsimane' Indigenous Knowledge. Swidden Fallow Management and Conservation. Ph. D. dissertation University of Florida, Gainesville.

Instituto Nacional de Estadística
2003 Bolivia: Características sociodemográficas de la población indígena. La Paz: INE.

Jehn, K.A., and L. Doucet
1997 Developing categories for interview data: consequences of different coding and analysis strategies in understanding text. Cultural Anthropology Methods 9(1):1-7.

Kay, B.R.
1957 The reliability of HRAF coding procedures. American Anthropologist 59:524-7

Kirby, K.N., R. Godoy, V. Reyes-García. E. Byron, L. Apaza, W. Leonard, E. Perez, V. Vadez, and D. Wilkie
2002 Correlates of delay-discount rates: Evidence from Tsimane' Amerindians of the Bolivian rain forest. Journal of Economic Psychology 23:291-316.

Krippendorff, K.

1980 Content analysis: An introduction to its methodology. Beverly Hills: Sage Publications.

Kurasaki, K.S.
2000 Intercoder reliability for validating conclusions drawn from open-ended interview data. Field Methods
12(3):179-94.

Lacy, S, and D. Riffe
1996 Sampling error and selecting intercoder reliability samples for nominal content categories. Journalism & Mass Communication Quarterly
73(4):963-73.

Lewis, B.N., and J.A. Cook
1968 Some comments and recommendations concerning inter-coder reliability. Psychological Reports
22(3):1213.

Lohman, T.G., A.F. Roche, and R. Martorell
1988 Anthropometric Standardization Reference Manual (Abridged Edition). Windsor, Ontario: Human Kinetics Publishers.

Lombard, M., J. Snyder-Duch, and C.C. Bracken
2002 Content analysis in mass communication: Assessment and reporting of intercoder reliability. Human Communication Research
28(4):587-604.

Peter, J., E. Lauf
2002 Reliability in cross-national content analysis. Journalism & Mass Communication Quarterly
79(4):815-32.

Reyes-García, Victoria
2001 Indigenous People, Ethnobotanical Knowledge, and Market Economy. A Case Study of the Tsimane' Amerindians, Bolivia. PhD dissertation
University of Florida, Gainesville.

Reyes-García, V., R. Godoy, V. Vadez, L. Apaza, E. Byron, E. Perez, W. Leonard, and D. Wilkie
2003 Is ethnobotanical knowledge held communally? Evidence from Bolivian Amerindians. Science
299:1707.

Reyes-García, V., V. Vadez, E. Byron, L. Apaza, W. Leonard, E. Perez, and D. Wilkie
2005 Market economy and the loss of ethnobotanical knowledge: Estimates from Tsimane' Amerindians, Bolivia. Current Anthropology

Accepted, pending revisions.

Riffe, D., and A. Freitag
1997 A content analysis of content analyses: Twenty-five years of journalism quarterly. Journalism & Mass Communication Quarterly
74(3):515-24.

Rosenthal, R.
1966 Experimenter Effect in Behavioral Research. New York: Irvington.

Ryan, G.W.
1999 Measuring the typicality of text: Using multiple coders for more than just reliability and validity checks. Human Organization
58(3):313-22.

Sipes, R.G.
1976 Test for coder bias. Behavior Science Research
11(3):149-68.

Vadez, Vincent, V. Reyes-García, R. Godoy, W. Leonard, T. Huanca, and E. Byron
2004 Income diversification of rural households: What role for agriculture? Household evidence from the Tsimane' Amerindians of the Bolivian Amazon. World Development
Under review.

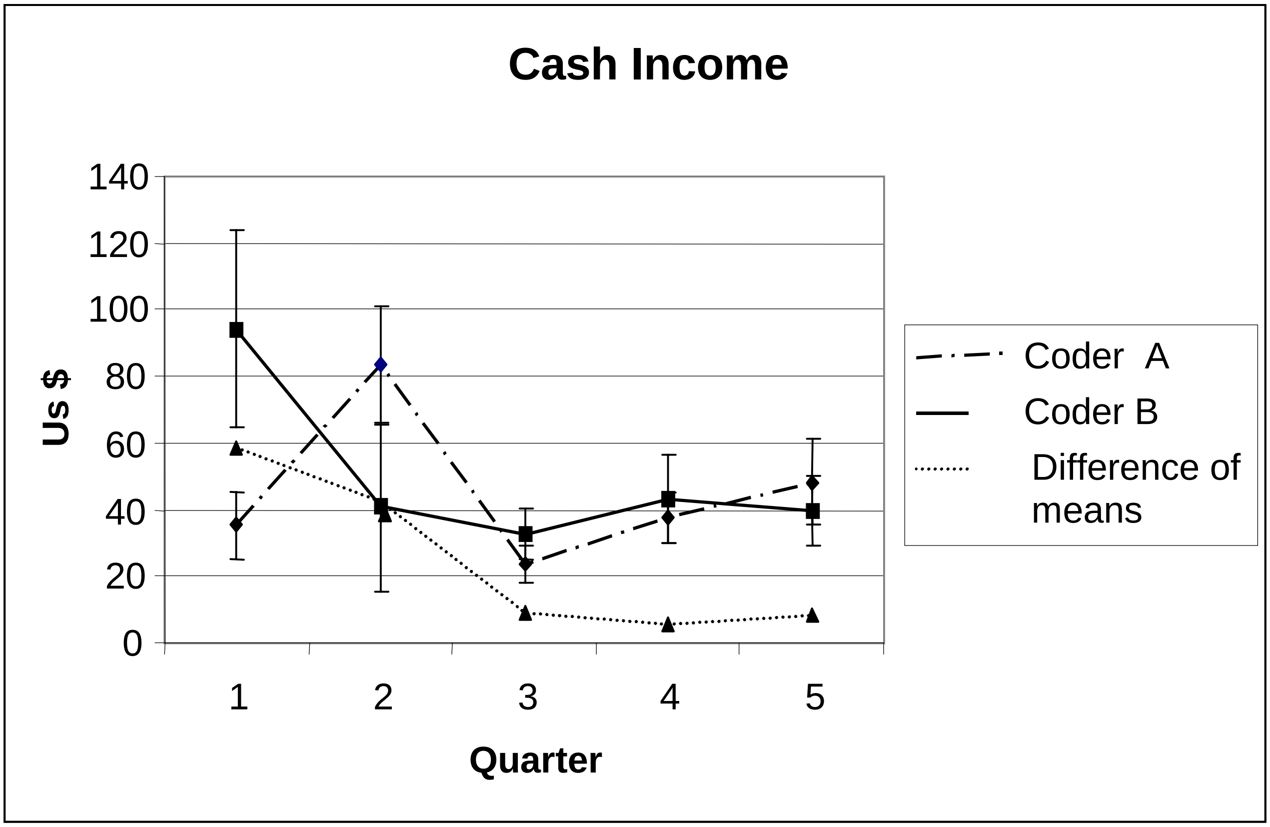
Yeaton, W.H., and P.M. Wortman
1993 On the reliability of meta-analytic reviews. The role of intercoder agreement. Evaluation Review
17(3):292-309.

Yelton, A.R., B.G. Wildman, and M. T. Erickson
1977 A probability-based formula for calculating interobserver agreement. Journal of Applied Behavioral Analysis
10:127-31.

Figure captions

Figure 1: Differences in cash income on data collected by two interviewers over five quarters

Figure 1



Cash Income
140
120
100
80
60
40
20
0
Us $
1
2
3
4
5
Quarter
Coder A
Coder B
Difference of means

List of table headings

Table 1

Variable	**Definition**	**Cross-sectional**	**Panel**
Socio-demographic			
Female	Sex of subject. *Female=1; male=0*	Observed	Observed
Age	Subject age. *Years*	If unknown, estimated by the interviewer	If unknown, estimated by the interviewer
Schooling	Maximum school attainment. *School grades*	Reported by subject	Reported by subject
Spanish fluency	Ability to communicate in Spanish. *From 0 (not able) to 2 (fluent)*	Measured by the interviewer	Measured by the interviewer
Economics			
Cash	Earnings (wages, sale & barter) obtained by adults in a household. *Bolivianos/ household /30 days*	Household level. Two-week recall	Individual level. 30 days recall
Wealth	Value of a stock of items owned by adults in the household. *Bolivianos/ household*	Household level	Individual level
Credit	Amount of debt incurred by adults in a household. *Bolivianos/ household/30 days*	Household level. Two-week recall	Individual level. 30 days recall
Fish consumption	Amount of fish brought to the household by adults. *Kg/household/24h*	Main fisher. 24h recall	Individual level. 24h recall. Weekly
Game consumption	Amount of game meat brought to the household by adults. *Kg/household/24h*	Main hunter. 24h recall	Individual level. 24h recall. Weekly surveys
Subjective health and nutritional status			
Health	Total number of person days ill in last 7 days. *Person/week*	Individual level. One week recall.	Individual level. One week recall.
Men stature	Stature of men (>15 years). *Centimeters*	Without shoes, flat against a vertical board.	Without shoes, flat against a vertical board.
Men weight	Weight of men (>15 years). *Kilograms*	Without shoes.	Without shoes.
Men biceps	Skinfold of men (>15 years), on the interior part of the arm. *Millimeters*	Measured at the mid point of the arm.	Measured at the mid point of the arm.
Men triceps	Skinfold of men (>15 years), on the back of the arm. *Millimeters*	Measured at the mid point of the arm.	Measured at the mid point of the arm.

Table 2

Variable	Interviewer A n	Interviewer A Mean	Interviewer B n	Interviewer B Mean	Difference of means X(A)-x(B)	%	T-Test P
Socio-demographic							
Female[a]	53	0.53	13	0.46	0.1	14.1	
Age *Years*	53	34.9	13	37.7	-2.8	-7.7	
Schooling *Years*	53	2.2	13	2.3	-0.1	-13.3	
Spanish fluency *0 to 2*	53	0.72	13	0.77	-0.1	-6.7	
Economics							
Cash *US $/household/month*	84	38.1	23	35.3	2.8	7.6	
Wealth *US $/household*	84	382	23	383	1	-0.1	
Credit *US $/household/month*	84	5.9	23	4.8	1.1	20.6	
Fish Consumption *Kg/household/24h*	209	1.9	465	1.4	0.5	30.3	
Game Consumption *Kg/household/24h*	209	0.31	465	0.49	-0.18	-45.0	
Subjective health and nutritional status							
Health *Days ill/person/week*	131	1.9	51	2.2	-0.3	-14.6	
Men stature *Cms*	97	163.1	14	162.5	0.6	0.4	
Men weight *Kgs*	97	62.8	14	62.5	0.3	0.5	
Men biceps *mms*	97	3.7	14	7.9	-4.2	-72.4	<0.0001
Men triceps *mms*	97	8.3	14	13.2	-4.9	-45.6	<0.0001

[a] Dummy variable. Name of the variable =1, omitted category=0

Table 3

Variable	Interviewer C n	Interviewer C Mean	Interviewer D n	Interviewer D Mean	Difference of means X(C)-x(D)	Difference of means %	T-Test p
Socio-demographic							
Female[a]	103	0.55	96	0.47	0.1	15.7	
Age *Years*	103	35.7	96	35	0.7	2.0	
Schooling *Years*	103	0.71	96	0.61	0.1	15.2	
Spanish fluency *0 to 2*	103	0.66	96	1.17	-0.5	-55.7	<0.0001
Economics							
Cash *US $/household/month*	103	57.5	96	118.1	-60.6	-69.0	0.0003
Wealth *US $/household*	103	192	96	260	-68	-30.1	0.003
Credit *US $/household/month*	103	0.63	96	6.9	-6.27	-166.5	0.05
Fish Consumption *Kg/household/24h*	103	0.5	96	1.1	-0.6	-75.0	0.02
Game Consumption *Kg/household/24h*	103	1	96	0.5	0.5	66.7	
Subjective health and nutritional status							
Health *Days ill/person/week*	103	1.7	96	2	-0.3	-16.2	
Men stature *Cms*	46	161.8	51	160.4	1.4	0.9	
Men weight *Kgs*	46	62.3	51	61	1.3	2.1	
Men biceps *mms*	46	1.9	51	4.4	-2.5	-79.4	<0.0001
Men triceps *mms*	46	4.3	51	9.7	-5.4	-77.1	<0.0001

[a] Dummy variable. Name of the variable =1, omitted category=0