LTA 2/01 • p. 260–273

JAN WALLIN

Auditors' Selection of Tolerable Error and Risk Levels in the Context of Sample Size Decisions: A Cross-cultural Experiment

ABSTRACT

As cross-border mergers and acquisitions become more frequent, cross-cultural research into the practices and attitudes of auditors becomes more warranted. Since many large, national auditing firms in northern Europe nowadays are associated with one of the American Big 5 firms, harmonisation of practices and adoption of the same sophisticated methods can be expected. The present study focuses on monetary unit sampling (MUS) or dollar-unit sampling (DUS), a method that has been recommended by the American Institute of Certified Public Accountants (AICPA).

The study tests the effects of population size, internal control, analytical review, audit object and culture on auditors' selection of acceptable upper-precision (materiality) and risk levels, judgements required when MUS is used to determine the sample size. The results indicate that internal control and analytical review have statistically significant effects on the acceptable upper-precision limits and also the acceptable risk level, suggesting that if internal control is known to be strong, or if an analytical review gives no warning of possible errors, auditors across the two cultures tend to accept lower precision and risk limits, the result of which is smaller sample sizes as determined by MUS. In addition, culture had an almost significant effect on the acceptable risk levels, with Swedish auditors more risk-averse than their Finnish colleagues, resulting in the choice of smaller sample sizes by the latter group.

Keywords: *Auditing, sample size decisions, monetary unit sampling, precision and risk judgements, cross-cultural, experiment*

JAN WALLIN, Professor
Swedish School of Economics and Business Administration, Vaasa • e-mail: jwallin@wasa.shh.fi

INTRODUCTION

Several American articles dealing with monetary unit sampling (MUS) – for example, Neter et al., 1978; Harwood et al., 1982; Andrews and Mayper, 1983; Dworin and Grimlund, 1984; Willits and McCaslin, 1985; Harwood et al., 1987 – have been published in the last two decades, but very few empirical studies focus on the subjective judgements needed for sample size determination in connection with MUS. To my knowledge only one American study dealing with auditors' sample size decisions in the context of MUS has been published in a top-tier journal during the 1990s: Kachelmeier and Messier, Jr. (1990). Dividing subjects into three experimental groups: (1) an intuitive judgement group, (2) a decision aid group who calculated sample sizes using the so-called AICPA Guide formula (1983: 59–64), and (3) a group who provided only formula parameters, which enabled the researchers to compute the sample sizes, Kachelmeier and Messier tested the hypotheses that the sample sizes of group 2 would be larger than those of group 1, and that those of group 3 would be larger than those of group 2. Neither hypothesis could be rejected. Repeating the experiment for two levels of internal control did not change the results.

The present study aims to explore the sample size issue further by extending the investigation to European cultures – in this case, Sweden and Finland – where auditors can be expected to have less experience with MUS than their American colleagues. The study will also focus on choices required of the auditor if the formula for the mean of the Poisson approximation to the binomial distribution – a formula which is equivalent to the AICPA formula – is used. The study will also focus on factors expected to influence these choices. Because of the obvious paucity of knowledge about auditors' judgement behaviour in the context of MUS, the study is of an explorative nature. The aim is to make an attempt to pinpoint interesting data patterns in the responses of a small sample of professional auditors, data patterns that could form the basis for formulating hypotheses for further investigation in the future.

RESEARCH QUESTIONS

The formula for the mean of the Poisson approximation to the binomial distribution ($m = n \times p$; n = number of trials, and p = error probability or tolerable error as a percentage of the population book value) requires three different choices of the auditor: (1) maximum tolerable difference between the book value and the audited value (upper-precision), (2) desired confidence level (risk = 1 – confidence level), and (3) the expected number of errors occurring in the sample. There are many factors that can be expected to influence these choices (cf. a list presented in IFAC International Auditing Guideline 19). The following factors were selected

for study here: population size, internal control effectiveness, results of analytical review, audit object and culture. Because of the explorative nature of the study, the formulation of research questions is preferred to the formulation of hypotheses. By sample size judgements I mean selection of acceptable precision and risk levels. The number of expected errors in the sample is expected to be zero.

> RESEARCH QUESTION 1: Does population size have an effect on auditors' sample size judgements?

At issue here is whether the auditor will stick to a constant upper-precision level, even if the population size increases. When the population size increases, a constant precision level – that is, tolerable error as a percentage of population size – will require the acceptance of larger absolute errors. This is evident from the AICPA formula:

$$\text{Sample Size} = \frac{\text{Population's Book Value}}{\text{Tolerable Error}} \times \text{Assurance Factor}$$

The assurance factor is identical to the mean of the Poisson distribution. This is tabled for several combinations of confidence levels and expected number of errors in the sample. By selecting the latter two, the auditor automatically determines the value of the assurance factor. To complete the MUS-procedure s/he also has to select an upper limit for precision.

To some auditors it may be natural to uphold an upper precision limit of, say, 5%, regardless of population size. However, 5% of, say, 2 million is much less in absolute terms than 5% of 20 million. Many auditors may therefore feel compelled to lower their relative precision limit when the population increases in size. If this is the case, the sample size will automatically be affected: a lower acceptable precision limit will lead to a higher sample size. The required risk level decision, on the other hand, need not be affected.

> RESEARCH QUESTION 2: Does internal control efficiency have an effect on auditors' sample size judgements?

If internal control is considered to be strong, the external auditor may feel it justified to select a smaller sample. As observed above, a smaller sample can be achieved in three ways: (1) by increasing the upper precision limit, leaving the risk level unchanged, (2) by increasing the acceptable risk level, leaving the precision limit unchanged, or (3) by changing both upwards simultaneously. Thus, internal control could have an effect on precision as well as risk.

> RESEARCH QUESTION 3: Do the results of an analytical review have an effect on auditors' sample size judgements?

In the USA and the UK the use of simple rather than sophisticated analytical review techniques prior to substantive tests is reportedly high (Ameen and Strawser, 1994; Fraser et al., 1997). Nevertheless, many practising auditors consider analytical review to be an important part of the auditing process. Analytical review may therefore be expected to have an effect on sample size judgements. If significant trend deviations are detected, the auditor may be expected to select a larger sample.

RESEARCH QUESTION 4: Does the object of the audit have an effect on auditors' sample size judgements?

Monetary unit sampling is primarily designed for use on accounts receivable and inventories. If accounts receivable are to be audited, it would not be surprising if auditors selected higher sample sizes than when the raw materials inventory is to be audited. One reason for this could be that the value of the latter is perceived to be difficult to determine correctly with or without a thorough audit because of valuation problems.

RESEARCH QUESTION 5: Is there a significant difference between the judgements of Swedish auditors and Finnish auditors?

Although the Swedish and Finnish national cultures overlap to a certain degree, it is a well-known fact that there are differences between the management styles of Swedish and Finnish companies: Swedish management culture is perceived as being democratic and thorough, requiring a long time to arrive at decisions, whereas Finnish management culture is perceived as being autocratic, resulting in quick and possibly risky decisions (Nurmi, 1998; von Brunow, 1998; Kinturi, 1998). When it comes to the attitudes of auditors, too little is known for the formulation of a hypothesis, but in view of the cultural differences concerning management style, it would not be very surprising if there turned out to be cultural differences concerning auditors' attitudes towards tolerable error and risk acceptance as well.

EXPERIMENTAL DESIGN

Independent variables

Table 1 shows of a list of the independent variables selected for the experiment. For each variable two levels were chosen. The design is a full-factorial repeated measures analysis of variance design (ANOVA), with three within-subjects variables and two between-subjects variables.

TABLE 1. Independent variables.

	Level	
	1	2
Within-subjects variables		
A. Population size	Large: 5 000 units with a book value of 20 million (SEK/FIM)	Small: 500 units with a book value of 2 million (SEK/FIM)
B. Internal control	Strong	Weak
C. Analytical review	No warning: No significant trend deviations detected	Warning: Significant trend deviations detected
Between-subjects variables		
E. Audit object	Raw materials inventory	Accounts receivable
D. Culture	Sweden	Finland

Subjects

Research instruments were sent to 200 Swedish and 40 Swedish-Finnish (Finnish citizens with Swedish as their native language) randomly selected members of the respective associations of chartered accountants. A native Finnish-speaking group was also contemplated but was excluded because of a very low response rate revealed in a pilot study. The Swedish-Finnish group of respondents should be reasonably representative of Finnish auditors seeing that (a) most of these respondents are thoroughly immersed in Finnish culture, born and bred as they are in Finland, and (b) they work for large auditing firms together with native Finnish-speakers.

Forty-six auditors responded to three survey questions, but only thirty-five (20 Swedes and 15 Swedish-Finns) of these (15.3%) participated in the experiment proper. Subjects were randomly assigned to two groups based on the between-subjects variable 'audit object' before the research instrument was sent. One group was told that the object of auditing was raw materials inventory and the other group was told that accounts receivable were to be audited. The distribution of the 35 usable responses between these two groups was:

the raw materials inventory group:	10 Swedes, 8 Swedish-Finns
the accounts receivable group:	10 Swedes, 7 Swedish-Finns

Experimental task and dependent variables

As indicated earlier, acceptable upper-precision limits and risk levels were chosen as dependent variables, for which the subjects were asked to select a percentage from a matrix that also included sample sizes in monetary units for different combinations of precision and risk. Each

subject was presented with a questionnaire in which eight framed information boxes – one for each combination of within-subjects factor levels – containing client characteristics and requests for judgements were included. The precision and risk percentages were presented to the subjects on scales with unequal distances between points. To use the points on the scales directly would therefore bias the analysis. Instead the precision and risk percentages presented to the subjects on the research instruments were coded on an ascending scale from 1 to 9 and 1 to 7, respectively, with '1' representing a precision and risk of 0%. The respective scale transformations are shown in Appendix 1.

RESULTS

Survey questions

The research instrument included three survey questions, the aim of which was to provide a background picture of the extent to which the AICPA formula, monetary unit sampling and analytical review techniques are used by auditors in Finland and Sweden.

The responses of the 35 experimental subjects to the first survey question are summarised in Table 2.

TABLE 2. Responses to survey question 1.

Survey question 1: If you choose not to check the whole population, how do you determine how many units to check?			
	Sweden	**Finland**	**Weighted average**
a) Intuition or experience	**70.0%**	**64.7%**	**67.6%**
b) Poisson mean formula	**0.0%**	**0.0%**	**0.0%**
c) Database search	**55.0%**	**41.2%**	**48.6%**
d Other methods	**30.0%**	**11.8%**	**21.6%**

It can be observed that none of the respondents use the formula for the mean of the Poisson approximation to the binomial distribution. The experiment must therefore have seemed quite hypothetical to those who participated in the experiment – and especially to those who chose not to participate. On the whole, it seems as if statistical methods do not play a major role in determining sample sizes in Sweden and Finland. Intuition and experience are much more important. The use of database search techniques, such ACL (Audit Command Language), also seems to be quite popular. To some extent, other methods, such as computer-based techniques for stratified sampling, are used.

TABLE 3. Responses to survey question 2.

Survey question 2: How well do you know a method called Monetary unit sampling (also known as Dollar unit sampling)?	Sweden	Finland	Weighted average
a) I'm not familiar with the method at all.	65.0%	52.9%	59.5%
b) I'm familiar with the method, but I don't use it.	10.0%	17.6%	13.5%
c) I use the method occasionally.	25.0%	29.4%	27.0%
d) I use the method on a regular basis.	0.0%	0.0%	0.0%

As can be seen from Table 3, surprisingly many respondents are not familiar with MUS at all. Obviously, auditing courses taught at business schools and universities in Sweden and Finland do not emphasise this method, a method that – as mentioned above – has received a lot of attention in American research and professional journals as well as in auditing guidelines.

TABLE 4. Responses to survey question 3.

Survey question 3: Which of the following methods for analytical review do you use?	Sweden	Finland	Weighted average
a) I don't carry out analytical review.	0.0%	0.0%	0.0%
b) Scanning	35.0%	64.7%	48.6%
c) Trend analysis (not regression analysis)	40.0%	58.8%	48.6%
d) Ratio analysis	90.0%	82.4%	86.5%
e) Reasonableness tests	100.0%	94.1%	97.3%
f) Regression analysis	5.0%	0.0%	2.7%
g) Other methods	5.0%	23.5%	13.5%

Table 4 shows that analytical review is viewed as an important part of the auditing process. Judgmental and simple quantitative techniques are predominant, whereas an advanced quantitative technique such as regression analysis is almost not used at all. In addition to the techniques mentioned in the question, some respondents indicated that they make comparisons with budgets and last year's figures, and some analyse relationships between items in the financial statements. On the whole, the emphasis seems to be on methods that are not statistically sophisticated. This is in line with the findings of the aforementioned American and British surveys of the use of analytical review techniques (Ameen & Strawser, 1994; Fraser et al., 1997).

The experiment

Since the number of participating auditors was rather low, an ANOVA should be regarded with caution. The basic assumption of an ANOVA is that the within-group data are samples from normal populations with the same variance. In this case tests indicate that the latter part of the assumption is satisfied, whereas the first part is not. This is not unusual when repeated measures are involved. This may not be very serious since F-tests are known to be quite robust to nonnormality (Lipe and Salterio, 2000: 293). In addition, visual inspection of stem-and-leaf plots and normal Q-Q probability plots indicate that the data distributions approximate to a reasonable extent normal distributions. Although outliers exist, these should not have a significant effect on the size of the means because there are very small differences between the observed cell means and the 5% trimmed cell means. In view of this, an ANOVA may be reliable enough to be able to pinpoint possible significant main effects and interactions, thus providing a basis for the generation of hypotheses worthy of investigation in future research.

Table 5 shows the repeated-measures ANOVA results for the subjects' upper-precision level judgments.

TABLE 5. Experimental results for auditors' upper-precision selections.

Source of variation	SS	df	MS	F	p-value
Between subjects					
Culture	1.103	1	1.103	0.117	0.735
Audit object	5.947	1	5.947	0.628	0.434
Culture* Audit object	133.077	1	133.077	14.054	0.001
Error	293.542	31	9.469		
Within subjects					
Population size	0.158	1	0.158	0.092	0.764
Internal control	10.094	1	10.094	8.864	0.006
Analytical review	6.406	1	6.406	10.303	0.003
Error (Population size)	53.242	31	1.717		
Error (Internal control)	35.299	31	1.139		
Error (Analytical review)	19.274	31	0.622		

Note: Non-significant (p > 0.05) first- and higher-order interactions are excluded.

Since the p-values are lower than 0.05 for the variables 'internal control' and 'analytical review', the null-hypothesis of equal means can be rejected for these two variables, suggesting that these have an effect on subjects' upper-precision selections. In other words, weak internal control induces auditors to lower their tolerable error levels, and so does a warning signal resulting from an analytical review. The other experimental variables did not have a signifi-

cant effect on the upper-precision judgments. Thus, when the upper-precision limit is the dependent variable, the results indicate negative answers to research questions 1, 4, and 5, whereas the results indicate affirmative answers to research questions 2 and 3.

In addition to these results, it is interesting to note that a strong interaction can be observed between the two between-subjects variables 'culture' and 'audit object' (p-value = 0.001). The Swedish auditors who participated in the experiment tolerated a higher error level for raw materials inventory than their Finnish colleagues, whereas the reverse was true for accounts receivable (Appendix 2, Figure 1).

The ANOVA results for the dependent variable 'risk' are displayed in Table 6. The low p-values for the variables 'internal control' and 'analytical review' indicate significant main effects, whereas the effect of 'culture' is almost significant (p-value = 0.07). It is evident that (1) the Swedish participants were more risk-averse than their Finnish colleagues; (2) weak internal control resulted in a lower acceptable risk level; and (3) a lower acceptable risk level was also the result if the analytical review was said to detect a warning signal about possible errors.

TABLE 6. Experimental results for auditors' acceptable risk judgments.

Source of variation	SS	df	MS	F	p-value
Between subjects					
Culture	23.339	1	23.339	3.518	0.070
Audit object	0.042	1	0.042	0.006	0.937
Error	205.636	31	6.633		
Within subjects					
Population size	0.077	1	0.077	0.181	0.673
Internal control	77.701	1	77.701	45.460	0.000
Analytical review	17.638	1	17.638	27.139	0.000
Population size* Culture	2.186	1	2.186	5.096	0.031
Internal control* Culture	9.447	1	9.447	5.527	0.025
Error(Population size)	13.300	31	0.429		
Error(Internal control)	52.986	31	1.709		
Error(Analytical review)	20.147	31	0.650		

Note: Non-significant (p>0.05) first- and higher-order interactions are excluded.

Thus, when acceptable risk levels constitute the dependent variable, the results indicate affirmative answers to research questions 2, 3, and 5, but negative answers to research questions 1 and 4.

Significant interaction effects between the variables 'culture' and 'population size', as well as between 'culture' and 'internal control', are also noteworthy (p-value = 0.031 and 0.025, respectively). The difference between the cultures in terms of mean acceptable risk levels were

larger for a large population size than for a small population size, with the Swedish respondents consistently being more risk-averse, that is, indicating lower acceptable risk levels, than the Finnish respondents (Appendix 2, Figure 2). As for the interaction between culture and internal control, it is evident that Swedish and Finnish auditors' mean acceptable risk levels are almost identical when internal control is strong but significantly different when internal control is weak, with Swedish auditors again more risk-averse than their Finnish colleagues (Appendix 2, Figure 3).

LIMITATIONS

The results of the study should be viewed with caution because of the low response rate and also because none of the respondents use MUS on a regular basis to determine sample size. The experimental setting must therefore have seemed very hypothetical to those who participated in the experiment – in fact, comments to this effect were written on some of the returned research instruments. The fact that some of the participants were familiar with monetary unit sampling, whereas some were not, may also have had an effect on the results. In a large-sample study this should be controlled for in the form of a covariate independent variable.

The experimental setting could be made more realistic in a number of ways. For example, several subjects suggested that the population book value should be related to total assets; otherwise it is difficult to assess the materiality – or acceptable tolerable error – of a deviation from the book value of the account to be audited. In addition, the inclusion of information concerning revenues, earnings, financial position and trend development could be considered.

CONCLUSIONS, IMPLICATIONS AND FUTURE RESEARCH ISSUES

The main purpose of the study was to carry out a preliminary investigation of the issue whether culture, audit object, population book value, internal control and analytical review have an effect on Swedish and Finnish professional auditors' selection of tolerable error and risk levels in the context of sample size decisions. Because the response rate was low (15.3%), no firm conclusions can be drawn. The results should be viewed as preliminary evidence providing a starting point for future research.

The results suggest that internal control and analytical review have an effect on subjects' precision limit selections. Culture, audit object and population size alone do not seem to affect these selections, but there were indications of effects of culture interacting with audit object. The Swedish respondents seem to tolerate higher upper-precision limits for the audit ob-

ject raw materials inventory than the Finnish respondents, whereas the reverse seems to be true for the audit object accounts receivable.

The results are to some extent different for selected risk levels. As in the case of precision limit selections, internal control and analytical review show significant main effects, but the effect of culture is also almost significant. In addition, there are significant interaction effects for culture and population size, as well as for culture and internal control. There is a larger difference between the cultures when the population size is large than when it is small, with Swedish auditors displaying more cautious judgements. Swedish auditors also seem to be more cautious in that they pay more attention to weak internal control than Finnish auditors.

In conclusion, the experimental results – especially the observed, significant interactions – pinpoint several interesting issues for future research:

a) Are there cultural differences in auditors' attitudes towards the perceived importance of different audit objects?
b) Are there cultural differences in auditors' attitudes towards the perceived need for readjusting materiality limits when population sizes increase?
c) Are there cultural differences in auditors' attitudes towards the perceived importance of internal control quality?
d) Do risk attitudes among auditors reflect observed cultural differences in management style?

The results of the study suggest affirmative answers to all of these questions, but studies larger in scope are needed before the questions can be considered fully answered. Obviously, statistical experiments based on small samples can only provide a scan of promising research issues. Large-scale surveys and in-depth interviews are needed before a full and reliable picture can emerge. ■

REFERENCES

AMEEN, E.C. and STRAWSER, J.R. (1994) 'Investigating, the Use of Analytical Procedures: An Update and Extension', *Auditing: A Journal of Practice and Theory*, 13(2): 69–76.

American Institute of Certified Public Accountants *(AICPA), Statistical Sampling Subcommittee Audit and Accounting Guide* (1983) 'Audit Sampling'.

ANDREWS, W.T. and MAYPER, A.G. (1983) 'Dollar-Unit Sampling', *The Internal Auditor*, April: 31–34.

BRUNOW VON, S. (1998) 'Management by perkele rubbar svenska cirklar', *Hufvudstadsbladet*, February 22: 18.

DWORIN, L. and GRIMLUND, R. (1984) 'Dollar Unit Sampling for Accounts Receivable and Inventory', *The Accounting Review*, April, LIX(2):218–241.

FRASER, I.A.M., HATHERLY, D.J. and LIN, K.Z. (1997) 'An Empirical Investigation of the Use of Analytical Review by External Auditors', *British Accounting Review*, 29: 35–47.

HARWOOD, G.B., SHIFFLER, R.E. and WOOD, J.M. (1982) 'Putting Dollar-Unit Sampling to Work', *The Internal Auditor*, February: 50–55.

HARWOOD, G.B., DAVIS, G.J. and BOROWSKI, P. (1987) 'Save Time with Computerized Dollar-unit Sampling', *The Internal Auditor*, April: 41–44.

IFAC International Auditing Guideline 19 (1985) 'Audit Sampling', February. Kachelmeier, S. J. and Messier, Jr. W. (1990) 'An Investigation of the Influence of a *The Accounting Review*, January, 65(1): 209–226.

KINTURI, M.-L. (1998) '1+1=3 eli Kulttuurisalamoista energiaa yritystoimintaan', *Yritystalous*, 56(1): 5–9.

LIPE, M.G. and SALTERIO, S.E. (2000) 'The Balanced Scorecard: Judgmental Effects of Common and Unique Performance Measures', *The Accounting Review*, July, 75(3): 281–298.

NETER, J., LEITCH, R.A. and FIENBERG, S.E. (1978) 'Dollar Unit Sampling: Multinomial Bounds for Total Overstatement and Understatement Errors', *The Accounting Review*, January, LIII(1): 77–93.

NURMI, E. (1998) 'Johtaja lahden takaa', *Helsingin Sanomat*, February 22: D1–D2.

WILLITS, S.D. and MCCASLIN, T.E. (1985) 'Micro Simplifies Dollar-unit Sampling', *The Internal Auditor*, June: 62–69.

APPENDIX 1. Scale transformations.

Dependent variable: 'precision'

1	2	3	4	5	6	7	8	9	
0%	0.1%	0.25%	0.5%	1.0%	2.5%	5.0%	10.0%	15.0%	> 15%
1	–	–	–	2	3	4	5	6	7

Dependent variable: 'risk'

APPENDIX 2. Plots of a selection of significant or almost significant first-order interactions.

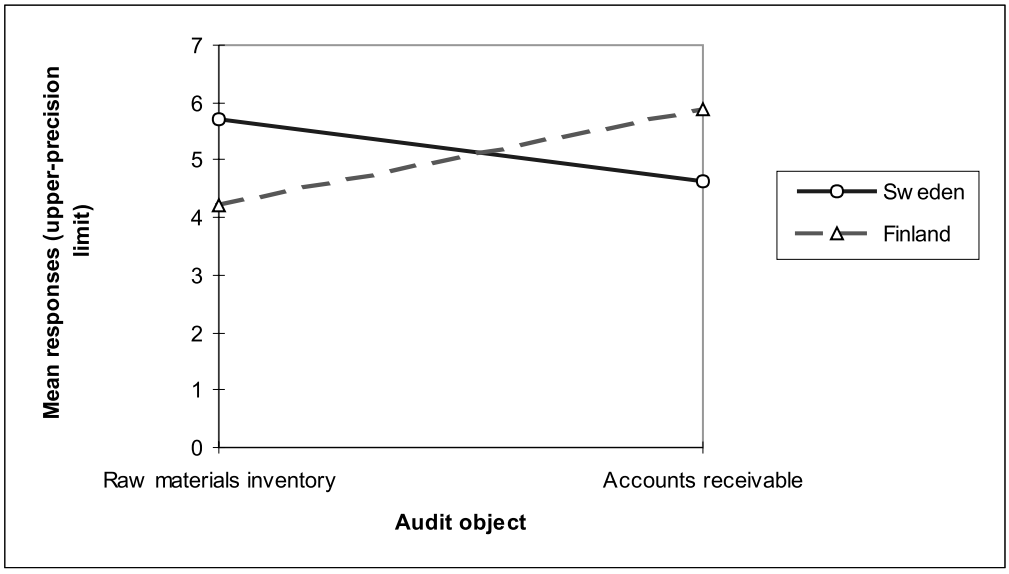

FIGURE 1. Interaction between the variables culture and audit object.

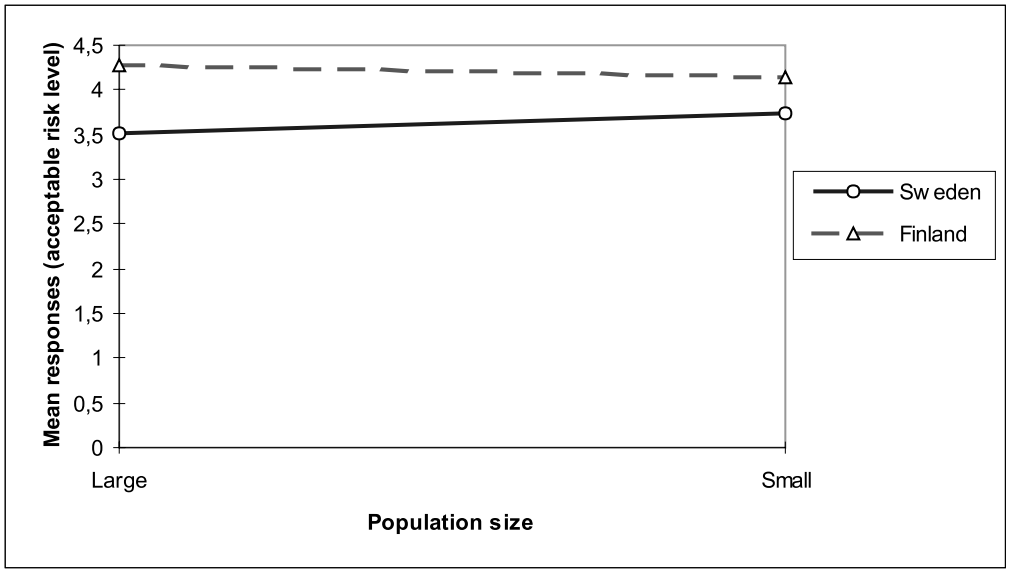

FIGURE 2. Interaction between the variables culture and population size.

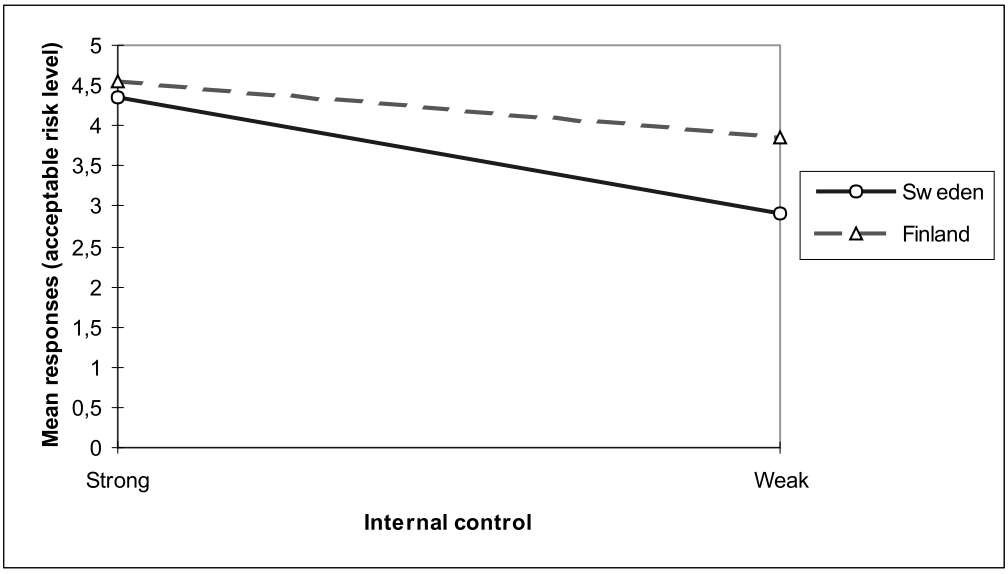

FIGURE 3. Interaction between the variables culture and internal control.