



TURKS & CAICOS ISLANDS

Report on

(Census 2001)

POST ENUMERATION

SURVEY

TURKS & CAICOS ISLANDS

Report on

(Census 2001)

POST ENUMERATION

SURVEY

Consulting Statistician: David N. Thomas

TABLE OF CONTENTS

	Page No.
I Introduction	1
II Survey Methodology	2
III Data Capture	7
IV Analysis of PES Results	8
V Conclusion	14
 Appendix I	
Table A. Turks and Caicos Islands – Data Files Specifications	17
Table B. Distributions of Households Listed and Response/Non-response Rates	20
Table C. Estimates of Coverage and Other Statistics of Characteristics of Interest by Regional Strata	21
 Appendix II	
Table 3.0. Percentage of Observations Unreconciled by Selected Characteristics of Interest and Regions	24
Table 3.1. Percent Identically Reported in the Census and the Post Enumeration Survey by Selected Characteristics of Interest and Regions	24
Table 3.2. Gross difference Rates between the Census and the Post Enumeration Survey by Selected Characteristics of Interest and Regions	25
Table 3.3. Ratio of Gross Difference to Identically Identified between the Census and the PES by Selected Characteristics of Interest and Regions	26
Table 3.4. Simple Random Response Variance between the Census and the Post Enumeration Survey by Selected Characteristics of Interest and Regions	27
Table 3.5. Index of Gross Shift between the Census and the Post Enumeration Survey by Selected Characteristics of Interest and Regions	28
Table 3.6. Index of Inconsistency between the Census and the Post Enumeration Survey by Selected Characteristics of Interest and Regions	29
Table 3.7. Net Difference Rates between the Census and the Post Enumeration Survey by Selected Characteristics of Interest and Regions	30
Table 3.8. Index of Net Shift between the Census and the Post Enumeration Survey by Selected Characteristics of Interest and Regions	31

	Page No.
Appendix III	
Chart 3.0. Percentage of Observations Unreconciled by Characteristics of Interest and Regions	33
Chart 3.3(a). Ratio of Gross Difference to Identically Identified for Selected age Groups by Regions	34
Chart 3.3(b). Ratio of Gross Difference to Identically Identified for Place of Birth – TCI by Regions	35
Chart 3.3©. Ratio of Gross Difference to Identically Identified for Ethnic Group – African by Regions	36
Chart 3.3(d). Ratio of Gross Difference to Identically Identified for Fertility – Zero Live Birth by Regions	37
Chart 3.3(e). Ratio of Gross Difference to Identically Identified for Highest Level of Education Primary/Secondary by Regions	38
Chart 3.5(a). Index of Gross Shift between the Census and the PES for Selected Age Groups by Regions	39
Chart 3.5(b). Index of Gross Shift between the Census and the PES for Place of Birth – TCI by Regions	40
Chart 3.5©. Index of Gross Shift between the Census and the PES for Ethnic group – African by Regions	41
Chart 3.5(d). Index of Gross Shift between the Census and the PES for Number of Live births equals Zero by Regions	42
Chart 3.5(e). Index of Gross Shift between the Census and the PES for Highest Education Level given as Primary/Secondary by Regions	43
Chart 3.6(a). Index of Inconsistency between the Census and the PES for Selected Age Groups by Regions	44
Chart 3.6(b). Index of Inconsistency between the Census and the PES for Place of Birth – TCI by Regions	45
Chart 3.6©. Index of Inconsistency between the Census and the PES for Ethnic group – African by Regions	46
Chart 3.6(d). Index of Inconsistency between the Census and the PES for Number of Live births equals Zero by Regions	47
Chart 3.6(e). Index of Inconsistency between the Census and the PES for Highest Education Level given as Primary/Secondary by Regions	48
Chart 3.8(a). Index of Net Shift between the Census and the PES for Selected Age Groups by Regions	49
Chart 3.8(b). Index of Net Shift between the Census and the PES for Place of Birth – TCI by Regions	50
Chart 3.8©. Index of Net Shift between the Census and the PES for Ethnic group – African by Regions	51
Chart 3.8(d). Index of Net Shift between the Census and the PES for Number of Live births equals Zero by Regions	52
Chart 3.8(e). Index of Net Shift between the Census and the PES for Highest Education Level given as Primary/Secondary by Regions	53
Appendix IV	
Model for the Evaluation of Response Errors	55

List of Appendices:

Appendix I

Table A

Table B

Table C

Appendix II

Table 1

Table 2

Table 3.1 to 3.8

Appendix III

Model and Methodology of Response Error Measurement

Appendix IV

Chart 2

Chart 3.3

Chart 3.5

Chart 3.6

Chart 3.8

I. Introduction

1.1 Background of the Survey

During the period August 20 to September 10, 2001, the Department of Economics Planning and Statistics (DEPS) of the Turks and Caicos Islands conducted its decennial Census. Following in the tradition of sound statistical principles (as mainly practiced in the developed countries), the DEPS, with the assistance of a Survey Statistician, planned and executed a Post Enumeration Survey (PES) of the Census. For the official engagement of the services of the Survey Statistician a contractual agreement was entered into, of which the main terms of reference were:

- i. Document the methodology, which the Survey Statistician developed and was used in the PES
- ii. Process the PES data and produce statistical information therefrom
- iii. Provide statistical analysis, including an interpretation of the survey results; and make recommendations for the conduct of similar survey activities in the future

1.2 Survey Objective

The primary purpose of the PES is to examine the quality (in terms of contents of the survey instrument) and accuracy (in terms of the coverage of population) emanating from the Census data collected over the census period. It is a well-known fact that a census is an enormous undertaking. The level of efficiency and accuracy, which can be attained from a smaller group of enumerators, who have been strictly, skillfully and uniformly trained to conduct a count of a proportion of the population, is far superior than what can be obtained from the much larger group of enumerators, who were similarly trained to conduct a 100 % count of the same population.

The main idea then, behind the conducting of the PES, is to investigate if a more accurate count of the population than what was obtained from the Census can be achieved, under the assumption that the PES was

designed and executed under sound statistical methods and applications.

II. Survey Methodology

2.1 Survey Design

The methodology discussed here is based on the well-known procedure of Dual System of Estimation (DSE). The DSE requires an independent and representative sample of the population (designated the P-Sample) to be taken, following which a case-by-case comparison of the results of the Census sample (designated E-Sample) and the P-Sample for identical persons or units is made and their differences highlighted.

Significant differences between the Census and PES results are indicative of errors made in the conduct of the Census. Errors may be classified as follows:

1. Errors in *coverage* of dwelling units as well as households
2. Errors in *coverage* of persons in dwelling units and/or in households
3. Errors in schedule content data.

In using the DSE method, one is able to quantify those errors. Implicit in the assumption that differences in results of the Census and PES are mainly due to weaknesses in the Census, is the superiority of the PES over the Census. In fact, the composition of the field staff to conduct the PES is predicated on the assumption that the best supervisors and/or the best interviewers who participated in the Census were selected.

As was pointed out above, the main objective of the PES, is to shed light on the causes of under-coverage or over-coverage since it identifies dwellings, households and persons that have either been missed or counted in error in the Census. In addition, the PES facilitates the examination of the quality of the content data on selected characteristics of interest reported in the Census. Generally, the PES helps to determine

whether interviewers were carrying out procedures correctly and how these procedures can be improved for future surveys and Censuses.

2.2 The PES Process

The PES consisted of three main aspects. Firstly, block clusters of households called Enumeration Districts (EDs), were identified and listed. Identification of EDs was done by scientific procedures as outlined in the sampling plan in the next section of this paper. The listing exercise was conducted using techniques of listing and within household coverage similar to those used in the Census.

Secondly, following the listing exercise, households are identified for re-interviews. The purpose of these re-interviews was to conduct within household coverage in more detail than in the listing exercise, and to conduct the schedule content data check on certain characteristics of interest.

During the re-interviews, the re-interviewers were expected to reconcile any discrepancies between responses in the Census and the PES and to state reasons for those discrepancies. The re-interviewers with the help of the respondents were expected to provide the most accurate responses, to be used as the final results.

Thirdly, utilizing the framework of DSE, summary measures were constructed for the estimation of coverage (see Table A of appendix I). In addition, following receipt of additional data from the P-Sample, summary measures were constructed for analyzing the content errors of the Census.

2.3 Sample Design and Plan of the PES

Two basic requirements of the PES are:

- That probability sampling be used. In probability sampling, each *sampling unit* has a known non-zero chance of being selected in the sample. The chance of selection must also be calculable.

- That a nationally representative sample of the *population* be selected.

The PES sample utilizes an equal probability of selection method (epsem), whereby each *sampling unit* (household) has an equal chance of being selected from the *population* of Non-Institutional households throughout the Islands.

Access to the population is made possible through a *Frame*, a listing of households within Enumeration Districts (ED's). ED's are the smallest geographic units into which the Islands were sub-divided for the purpose of Census 2001. These units have been demarcated to fit within non-overlapping boundaries based on easily identifiable features as far as possible. The size of an ED ranges between 1 - 325 households. The EDs are designated the Primary Sampling Units (**PSUs**)

2.4 Sample Size

The determination of the sample size was largely guided by practical considerations in terms of limited human and financial resources and time constraint on the one hand, and on the other hand, the desire for reliable information at the level of regional stratification. It was also decided that the total sample of PSUs should be allocated proportionately among the Islands regions of stratification.

In terms of stratification, each PSUs was located in one of three strata:

Stratum I comprised PSUs belonging to Grand Turk and Salt Cay; Stratum II comprised PSUs from South, Middle and North Caicos, and Stratum III comprised PSUs from Providenciales, Pine Cay and Parrot Cay. (For the allocation of PSUs, see table 1 below).

TABLE 1. DISTRIBUTION OF PRIMARY SAMPLING UNITS (PSUs) AND ULTIMATE SAMPLING UNITS (USUs) AMONG STRATA OF ISLANDS

Stratum of Islands	NO. of (PSUS) EDs	Number of (USUs) Households	Percent Distribution of (USUs) Households	No. of Clusters	Av. Size of Cluster	No. of Selected EDs
Grand Turk	19	1290	16.9%	257	5	4
Salt Cay	1	48	0.6%	10	5	0
Stratum I	20	1338	17.5%	266	5	4
South Caicos	5	372	4.9%	74	5	1
Middle Caicos	1	74	1.0%	15	5	0
North Caicos	2	478	6.3%	95	5	2
Stratum II	8	924	12.1%	184	5	3
Providenciales	42	5256	68.9%	1046	5	16
Pine Cay	1	67	0.9%	13	5	0
Parrot Cay	1	42	0.6%	8	5	0
Stratum III	44	5365	70.3%	1068	5	16
Grand Total	72	7627	100.0%	1518	5	23

In order to achieve those ends mentioned above, it was decided to select approximately 32 % of all PSUs at a first stage. PSUs selected at this stage were listed and intensive coverage checks were undertaken. Approximately twenty-three PSUs were selected, which yielded 3385 households.

At the second stage clusters of households were selected from the PSUs, which were selected at the first stage. Based on the above decisions, clusters of approximately five households were selected from each selected PSU, so that, the overall rate of sampling households was approximately 1.5% of all households or 1 in 66 households, giving an overall sample size of approximately 115 households. The second stage sample of 115 households was used to investigate the quality of the content of the Census data collected.

2.5 Sample Selection

2.5.1 Selection of the PSUs

At the first stage, PSUs were selected with probability approximately proportional to size (PPS), the size measure being the number of clusters of households that constitutes the PSU. The selection of PSUs was achieved as follows. Note that the following procedure was applied in each region of stratification:

Firstly, beginning with the first PSU on the list of PSUs listed in sequence, the number of clusters of households for that PSU was recorded adjacent to that first PSU in a variable designated "Cumulative Clusters". A cumulative sum of the number of clusters of households of the first and the second PSUs is recorded in the said variable, corresponding to the second PSU. A cumulative sum of the number of clusters of households of the first, second and third PSUs is recorded in the said variable adjacent to the third PSU, and so on.

Secondly, a random number (designated a *random start*) between 1 and the inverse of the sampling fraction, 66 (inclusively), was determined (from a book of random numbers). The random start was compared with the first cumulative total in the variable "Cumulative Clusters". If the random start was less than or equal to the said cumulative total, then the PSU associated with the said cumulative total was selected. Otherwise, the random start was compared with consecutive cumulative totals and selection of the first PSU was made where the associated cumulative total was less than or equal to the random start.

Once the first selection was made, the number 66 ($=1 \cdot 66$) was added to the random start and the new number thus formed was compared with the cumulative total associated with the first selected PSU. If the new number was less than or equal to the cumulative total associated with the first selected PSU, then the same PSU was selected again. Otherwise, the new number was compared with consecutive cumulative totals and the second selection was made where the associated cumulative total was less than or equal to the new number.

In general, in order to select an additional PSU following the selection of the k th PSU, the number k times 66 was added to the random start and the new number thus formed was compared with the cumulative total associated with the k th PSU selected. If the new number was less than or equal to the cumulative total associated with the k th selected PSU, then the k th PSU was selected again. Otherwise, the new number was compared with consecutive cumulative totals and the $(k+1)$ th selection was made where the associated cumulative total was less than or equal to the new number. The process of selection of PSUs continued until the sum of all clusters of households for the region of stratification was exceeded.

2.5.2 Selection of Ultimate Sampling Units

The selection of Ultimate Sampling Units (USUs) – households from selected PSUs, was done in two stages. Following the selection of PSUs at the first stage, at the second stage, households were selected with probability inversely proportional to size (PPS^{-1}) using the same size measures (number of clusters of households) of the selected PSUs. To identify such households the process used was similar to that described above. So that, to select the first household a random number between 1 and the size measure was drawn. Based on this random start the $(k+1)$ selection was systematically done by adding k times the size measure ($k = 1, 2, \dots$) to the random start until the total number of households within the PSU was exceeded.

III. Data Capture

3.1 Source of Data

Information collected in the Field was recorded in documents referred to as Listing records. The listing records used in this exercise came in pairs since they consisted of information based on two independent sources, namely the E-Sample and the P-Sample. As already noted, the E-Sample, which originated from the Census was formed as a result of the P-Sample. Once the sampling units of the P-Sample were identified the E-Sample was generated from only PSUs selected in the P-Sample.

A third source of information utilized in this exercise was the reconciliation forms for each sampled PSU. Reconciliation was made after the entire PSU was listed and the enumerator returned back to the office with the listing record used in the P-Sample for the specific PSU. The enumerator's supervisor in collaboration with an officer in the Statistical Office and the enumerator compared the data from the original E-sample listing record with what was recorded in the P-sample listing record.

Where apparent discrepancies existed attempts were made to determine the correct response which invariably resulted in the return to the respective households for guidance in the matter. Reasons for any observed differences were recorded in the reconciliation forms.

3.2 Data Entry and Editing

Window Excel spreadsheet was used to capture the information contained in the listing records of both the E-Sample and P-Sample. Editing of the data was done as much as possible. Wherever possible, the information in the reconciliation forms was used to assist in the editing.

IV. Analysis of PES Results

4.1 Non-Response

Apart from errors due to sampling (*sampling errors*), another type of error is the non-sampling error. A major source of this latter type of error is *non-response*. Non-response could be problematic in any survey, since large non-response rates could introduce *selection* bias, which can seriously undermine the confidence in the validity or accuracy of the results of the survey.

An examination of Table B of Appendix I reveals that the total non-response rate was 27.7%. However, when adjustments for vacant and closed dwellings are made, the effective non-response rate stood at 11.6%, less than 50% of the total non-response. All responsibility for the high effective non-response rate is attributed to the number of no contacts. Quite unusual, if not suspect, is the high vacant rates (as high as 65% in a PSU in Providenciales) observed in some cases.

There appears to be some correlation between the number of “vacant dwellings” and the number of “non contacts”. Dependent on the strength of that correlation, the accuracy of the estimate of the population count by means of the P-Sample may be seriously undermined.

One may determine the strength of the correlation between “vacant” and “non contact” non-response rates, by conducting an analysis of variance to test for interaction between interviewers, as replicates, and the types of non-responses (“vacant”, “non contact”, “closed”, etc.). However, data on non-responses by interviewer were not available.

In order to assess the risk of obtaining large response biases from the data, one needs to determine whether the sample is representative of the population or not. This may be achieved by comparing the distribution of the population with that of the effective sample.

Table 2. Comparison of Distributions of Households: Population vs Effective Sample by Strata

Percent Distribution of Households	Stratum		
	I	II	III
Population	17.5%	12.1%	70.3%
Effective Sample	14.5%	11.5%	74.0%

From Table 2 above, it can be observed that Stratum I is under represented in the sample. On the other hand, Stratum III is over represented in the sample.

4.2 Analysis of Coverage

In Table A of Appendix I, the file specification of an exhaustive list of variables with their description, which were used in this study, is provided. However, the most important variable among all the others given, that is to say, the main one upon which the entire analysis hangs is the **coverage rate**. Coverage relates to questions such as:

How many dwelling units/households/persons were either omitted in error or included in error in the Census?

How significant were those omissions/inclusions and what impact, if any, did they have on regional or total population count?

If we are to be guided by the American standards as to what constitutes a significant error in coverage, then an under count rate of at least 2% will be considered significantly different from 0 % under count. Table C of Appendix I contains the results of coverage.

By means of the DSE, the True Population count for the Turks and Caicos Islands was estimated to be 19497. The distribution of the population among the regional strata was as follows:

Providenciales/Parrot/Pine Cay	13507
North/Middle/South Caicos	1979
Grand Turk/Salt Cay	4011

Furthermore, the Census data seemed to have covered approximately (94.5%) of the true population due to an omission rate of 5.5 %.

From the perspective of the regional strata the omission rates for North/Middle/South Caicos, and Grand Turk/salt Cay were 2.1 % and 1.5% respectively. By the USA standards those omissions rates are not significant from zero. Hence, we may conclude that the sub-populations of those two mentioned regional strata were approximately 100 % covered by the Census count. However, for the regional strata of Providenciales/Pine/Parrot Cay, there appeared to be a significant omission rate of 7.2 % (according to USA standards).

An interesting feature of the analysis is the relationship between the variables “In-Mover” and “Matched In-mover” for the three regional strata.

While there is close comparability between “In-Mover” and “Matched In-Mover” for both regional strata Grand Turk/Salt Cay (237 compared

with 233) and North/Middle/South Caicos (214 compared with 208) respectively, “In-Mover” and “Matched In-Mover” for Providenciales/Pine/Parrot Cay were significantly different (2717 compared with 2158). The great disparity between the variables as shown here is consistent with the high “no contact” and “vacant” rates for the latter regional strata.

4.3 Measurement of Content Errors.

As mentioned earlier, one of the primary purposes of the PES is to examine the quality of the data collected in the Census. In order to accomplish this goal, comparisons are made between data on ***characteristics of interest*** collected in the Census and the PES. Differences in responses to the two are used as a measure of random response errors and other associated statistical biases.

4.3.1 Summary Measures of Characteristics of Interest

In order to conduct the analysis of content errors of the Census, summary measures or ***response error*** indicators were constructed for the following characteristics of interest:

Age:

- Less than 1 year
- 1 to 4 Years
- 15 to 19 Years
- 35 to 39 Years
- 45 to 49 Years

Place of Birth – Turks and Caicos

Ethnicity – African/Negro/Black

Number of Live Births equals Zero

Highest Level of Education:

- Primary School
- Secondary School

The following ***response error*** indicators were constructed:

1. **Gross difference rate**
2. **Net difference rate**
3. **Index of gross shift**
4. **Index of net shift**
5. **Ratio of Identically Identified to PES Total**
6. **Ratio of Gross difference to Identically Identified**
7. **Index of Inconsistency**

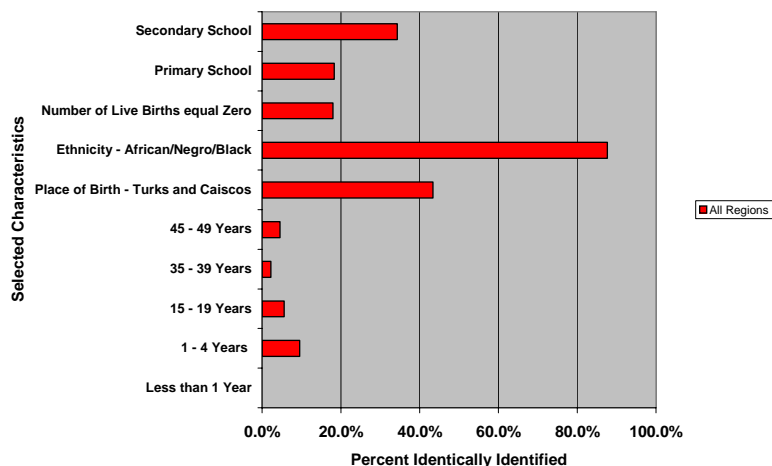
Appendix IV provides the Model from which the **response error** indicators were developed for use in the analysis.

4.3.2 Analysis of response error Indicators

It should be pointed out at the onset, that this aspect of the study has been fraught with many problems associated with the limitations of the data collected. Field activities with respect to re-interviews of households may not have been effectively performed. This assessment of the Fieldwork is easily borne out by an examination of the degree of reconciliation that was achievable from the data collected from the PES and linked to the Census.

From a total of 347 respondents in the PES a maximum 178 or 51.3% could be traced back to corresponding respondents in the Census. That implies that a significant percentage (48.7%) of the total observations could not be reconciled between the PES and Census (see Table 3.0 and Chart 3.0 of Appendices II and III respectively).

Chart 1.0 Percent Identically Identified in Census and PES for all Regions by Selected Characteristics



In terms of the identification of the selected characteristics of interest under review, of those respondents that could be traced back from PES to Census, it was observed that, except for the characteristic “Ethnicity – African”, less than 50% was identically identified between Census and PES. Chart 1.0 above displays the percent distributions of selected characteristics of interest that were identically identified.

It is highly probable that the main reason for the low frequency of identical matches between Census and PES data contents is the elapsed time in the administration of the PES following the Census. A common practice is to conduct the PES at most two weeks after the Census. In this case the PES was conducted a couple months after the Census.

Another rule of thumb is that the ratio of **gross difference** to **identically identified** should not exceed 10.0%. For almost all characteristics of interest (across all regions), except for the characteristic “Ethnicity – African”, the ratio ranged between 50.0% to 250.0% (see table 3.3 and chart 3.3 of Appendices II and III respectively). The ratio of **gross difference** to **identically identified** for “Ethnicity-African” was 1.9% well below the maximum acceptable value. The **Gross Difference Rate** or rate of different observations on both occasions (Census and PES) for selected characteristics of interest, ranged between 1.7% and 43.4% across all regions (Table 3.2 of Appendix II refers).

Among the indicators of response errors (for all indicators constructed see Appendix II), is the **index of inconsistency** (table 3.6 of appendix II and charts 3.6 (a)-(e) of appendix III refer). This index measures the proportionate increase in element variance due to lack of repetition in responses over the two occasions (Census and PES). Large values in the index indicate the instability of the estimated values of the characteristics of interest in the Census.

An examination of the **index of inconsistency** across all regions reveals that, apart from the value of the index for the characteristic “Ethnicity – African” - (7.8%), values of the index for other selected characteristics are significantly large, all exceeding 19.0%.

V. Conclusion

5.1 Summary

The main objectives of the PES were to examine and measure the Census coverage of TCI's population and the degree of stability or repeatability of responses to selected characteristics of interest obtained during the Census. In order to achieve those objectives summary measures were constructed within the framework of the Dual System of Estimation (DSE). Those summary measures formed the basis of the analysis for the exercise.

5.1.1 Main Findings

Analysis of the PES data under the DSE system revealed the following main findings:

- The PES is indeed an invaluable mechanism for assessing the performance of the Census data collection efforts.
- Total coverage of the population, based on the Census count, was approximately 95.5%. That implies that the population was undercounted by 5.5%.
- When viewed at the level of the regions, the undercount was primarily restricted to the stratum of regions, Providenciales/Pine/Parrot Cay, which registered an undercount rate of 7.2%
- Low frequencies in matched responses between Census and PES were most probably due to the long elapsed time between the conduct of the PES and the Census.
- The index of inconsistency for all but one of the selected characteristics of interest was significantly large, indicating the instability or lack of repeatability of responses for those characteristics over surveys.

An interesting feature of response error measurement that was not considered in this analysis was that relating to the evaluation of ***correlated response variance*** or ***interviewer variance***. This latter form of response error is due to differential average effects of individual interviewer on his/her workload. In order to measure the effect of interviewer the sample of first stage units must first be replicated and an interviewer assigned to each replication. This sample was not constructed in that manner.

5.1 Recommendations

The recommendations proffered here emanate from the main findings above. They are based on the assumption that the PES will be viewed as an integral part of future Censuses.

- 1) PES activities should be planned in conjunction with Census activities, so that the PES may be fielded not later than two weeks after the beginning of the Census. This will have implication for the total number of field interviewers to be deployed for both the Census and the PES. It will also have implication for coordinating field activities around both events simultaneously.
- 2) The PES survey instruments must be strengthened to accommodate more detailed information at the level of the household individuals in order to obtain greater dis-aggregation of data on coverage as well as content.
- 3) Among other variance measurements, the ***interviewer variance*** must be measured. This will necessitate the arrangement of enumeration districts into replications and the random assignment of interviewers to replications.
- 4) PES field activities need stricter monitoring from start to ending. In addition, procedure for the reconciliation of the PES data must be reviewed.

APPENDIX

I

TRUKS AND CAICOS ISLANDS PES OF CENSUS 2001

TABLE A. DATA FILE SPECIFICATION

Variable Number	Variable Name	(Specifications based on data from PES and Census VR and Reconciliation between the two instruments).								
.....										
Sample Design Variable										
1	Island	a 2-digit code								
2	Region	= 1 if Island is Grand Turk or Salt Cay = 2 if Island is North, Mid., or South Caicos = 3 if Island is Providenciales, Pine or Parrot Cay. Region is a classification variable								
3	SAMPL-RATE	= number of sample ED's in Island divided by the number of frame ED's in Island; carry 4 implicit decimals								
		<table border="1"> <thead> <tr> <th>REGION</th> <th>RATE</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0.1945</td> </tr> <tr> <td>2</td> <td>0.3902</td> </tr> <tr> <td>3</td> <td>0.3148</td> </tr> </tbody> </table>	REGION	RATE	1	0.1945	2	0.3902	3	0.3148
REGION	RATE									
1	0.1945									
2	0.3902									
3	0.3148									
4	ENUM	Enumeration District, a 4-digit code								
5	WEIGHT	Inverse of fraction of allocation of households to selected ED's, 5 digits with four implicit decimals								
6	HHOLD	Household number, a 3-digit code								

TABLE A. DATA FILE SPECIFICATION

Variable	Variable	
Number	Name	(Specifications based on data from PES and Census VR and Reconciliation between the two instruments).
.....		
Analysis Variables (calculate value at the level of household) Cont'd		
7	NON-MOVER	= number of non-movers from PES, 5 Digits
8	OUT-MOVER	= number of out-movers (incl. Exit by death), either part or all household persons from PES, 5 Digits
9	IN-MOVER	= number of in-movers (incl. Entry by birth), either part or all household persons from PES, 5 Digits
10	NM-MATCHED	= number of PES non-movers that could be matched to Census persons, 5 DIGITS
11	OM-MATCHED	= number of PES out-movers that could be matched to Census persons, 5 DIGITS
12	IM-MATCHED	= number of PES in-movers that could be matched to Census persons, 5 DIGITS
13	ERR-INCLUDED	= number of persons erroneously included in the Census, 5 DIGITS
14	POP-CENSUS	= NM-MATCHED + OM-MATCHED + ERR-INCLUDED, 5 DIGITS
15	PES-POP	= NON-MOVER + IN-MOVER, 5 DIGITS
16	OMISSION	= PES-POP – NM-MATCHED – IM-MATCHED, 5 DIGITS
17	OMISSION RATE	= OMISSION/PES-POP *100, 5 DIGITS
18	COVG-RATE	= 100-OMISSION RATE/100, 5 DIGITS

TABLE A. DATA FILE SPECIFICATION

Variable Number	Variable Name	(Specifications based on data from PES and Census VR and Reconciliation between the two instruments).
.....		
Analysis Variables (calculate value at the level of household)-CONCLUDED		
19	TRUE-POP	= POP-CENSUS – ERR- INCLUDED/COVG-RATE, 5 DIGITS
20	NET-ERR	= TRUE-POP – POP-CENSUS, 5 DIGITS
21	GROSS-ERROR	= OMITTED + ERR-INCLUDED, 5 DIGITS

Note: - The Weights for First stage sampling units within respective units are as follows:

Region	Weight
Grand Turk or Salt Cay	3.5871
South, Middle, or North Caicos	1.3276
Providenciales, Pine or Parrot Cay	2.3205

T

Table B. Distributions of Households Listed and Response/Non-response Rates

Regional Strata	LOCALITY	E.D NUMBER	NO. HHOLDS LISTED	Distr. of listed Households	CLOSED	VACANT	NO CONTACT	REFUSALS	Effective Sample	Distr. Of Effec. Smpl	Effective Non-resp.	Total Non-resp.	Effective Non-resp. Rate	Total Non-resp. Rate
1	2	2	142	4.2%	2	8	0	0	132	4.8%	0	10	0.00	7.04
1	2	3	115	3.4%	1	9	1	0	105	3.8%	1	11	0.95	9.57
1	4	1	55	1.6%	0	14	0	0	41	1.5%	0	14	0.00	25.45
1	9	1	71	2.1%	13	17	0	0	41	1.5%	0	30	0.00	42.25
Grant Turks			383	11.3%	16	48	1	0	319	11.5%	1	65	0.31	16.97
3	6	1	222	6.6%	0	40	0	0	182	6.6%	0	40	0.00	18.02
5	1	1	208	6.1%	6	65	18	0	137	4.9%	18	89	13.14	42.79
5	2	1	165	4.9%	34	49	3	0	82	3.0%	3	86	3.66	52.12
North/South Caicos			595	17.6%	40	154	21	0	401	14.5%	21	215	5.24	36.13
6	1	1	340	10.0%	3	46	107	0	291	10.5%	107	156	36.77	45.88
6	1	2	441	13.0%	0	8	48	0	433	15.6%	48	56	11.09	12.70
6	3	1	81	2.4%	2	41	10	0	38	1.4%	10	53	26.32	65.43
6	5	4	175	5.2%	1	18	0	0	156	5.6%	0	19	0.00	10.86
6	6	2	63	1.9%	0	8	1	0	55	2.0%	1	9	1.82	14.29
6	7	1	97	2.9%	0	32	16	0	65	2.3%	16	48	24.62	49.48
6	7	3	130	3.8%	0	12	17	0	118	4.3%	17	29	14.41	22.31
6	7	4	176	5.2%	3	35	20	0	138	5.0%	20	58	14.49	32.95
6	9	1	107	3.2%	0	20	0	0	87	3.1%	0	20	0.00	18.69
6	9	4	186	5.5%	2	33	43	0	151	5.5%	43	78	28.48	41.94
6	9	5	205	6.1%	1	30	31	0	174	6.3%	31	62	17.82	30.24
6	11	4	105	3.1%	0	25	0	0	80	2.9%	0	25	0.00	23.81
6	15	3	156	4.6%	0	14	3	0	142	5.1%	3	17	2.11	10.90
6	15	4	145	4.3%	2	23	2	0	120	4.3%	2	27	1.67	18.62
Providenciales			2407	71.1%	14	345	298	0	2048	74.0%	298	657	14.55	27.30
All Islands (Grand Total)			3385	100.0%	70	547	320	0	2768	100.0%	320	937	11.56	27.68

Table C. ESTIMATES AND OTHER STATISTICS OF CHARACTERISTICS OF INTEREST BY REGIONAL STRATA

		STRATUM I	STRATUM II	STRATUM III	Total
		Grand Turk/ Salt Cay	North,Middle, South Caicos	Providenciales, Pine/Parrot Cay	ALL STRATA/ REGIONS
Sample Size	No. of Households	384	594	2407	3385
Variable of Analysis	Statistics				
(A) Population Totals:					
Non-Mover	Estimated Number	3530	1682	10456	15668
	Standard Error	835	211	1381	1628
	Coefficient of Variation (%)	23.6	12.6	13.2	10.4
	95% Confidence Interval				
	Lower Limit	1893	1268	7749	12477
	Upper Limit	5166	2097	13164	18858
Out-Mover	Estimated Number	473	284	2458	3215
	Standard Error	92	70	319	339
	Coefficient of Variation (%)	19.4	24.6	13	10.5
	95% Confidence Interval				
	Lower Limit	293	147	1833	2551
	Upper Limit	654	421	3082	3879
In-Mover	Estimated Number	237	214	2717	3168
	Standard Error	109	31	430	444
	Coefficient of Variation (%)	46	14.5	15.8	14.0
	95% Confidence Interval				
	Lower Limit	23	153	1875	2297
	Upper Limit	450	274	3559	4039
Matched Non-Mover	Estimated Number	3476	1648	10071	15195
	Standard Error	822	204	1321	1570
	Coefficient of Variation (%)	23.6	12.4	13.1	10.3
	95% Confidence Interval				
	Lower Limit	1865	1248	7481	12117
	Upper Limit	5087	2047	12661	18272
Matched Out-Mover	Estimated Number	473	288	2460	3221
	Standard Error	92	71	317	337
	Coefficient of Variation (%)	19.4	24.8	12.9	10.5
	95% Confidence Interval				
	Lower Limit	293	148	1839	2560
	Upper Limit	654	428	3080	3883
Matched In-Mover	Estimated Number	233	208	2158	2599
	Standard Error	106	30	304	323
	Coefficient of Variation (%)	45.6	14.2	14.1	12.4
	95% Confidence Interval				
	Lower Limit	25	151	1562	1966
	Upper Limit	441	266	2754	3232
E-Sample	Estimated Number	3949	1936	12531	18416
	Standard Error	895	194	1509	1766
	Coefficient of Variation (%)	22.7	10	12	9.6
	95% Confidence Interval				
	Lower Limit	2194	1556	9573	14955
	Upper Limit	5704	2315	15489	21876
P-Sample	Estimated Number	3767	1896	13173	18836
	Standard Error	934	200	1712	1960
	Coefficient of Variation (%)	24.8	10.6	13	10.4
	95% Confidence Interval				
	Lower Limit	1936	1504	9818	14994
	Upper Limit	5597	2288	16528	22678

Table C. ESTIMATES AND OTHER STATISTICS OF CHARACTERISTICS OF INTEREST BY REGIONAL STRATA

		STRATUM I	STRATUM II	STRATUM III	Total
		Grand Turk/ Salt Cay	North,Middle, South Caicos	Providenciales, Pine/Parrot Cay	ALL STRATA/ REGIONS
Omissions	Estimated Number	57	40	947	1044
	Standard Error	36	6	215	219
	Coefficient of Variation (%)	63.5	16.3	22.8	20.9
	95% Confidence Interval				
	Lower Limit	-14	27	525	616
Upper Limit	129	53	1369	1472	
True Population	Estimated Number	4011	1979	13507	19497
	Standard Error	910	198	1621	1862
	Coefficient of Variation (%)	22.7	10.0	12.0	9.6
	95% Confidence Interval				
	Lower Limit	2226	1591	10330	15847
Upper Limit	5796	2367	16684	23158	
(B) Ratios: Omission Rate = Omissions/P-Sample OR Net Error/True Pop.	Proportion	0.015	0.021	0.072	0.055
	Standard Error	0.009	0.001	0.011	0.009
	Coefficient of Variation (%)	56.6	7.1	14.9	15.3
	95% Confidence Interval				
	Lower Limit	-0.002	0.018	0.051	0.039
Upper Limit	0.032	0.024	0.093	0.072	

APPENDIX II

Table 3.0 Percentage of Observations Unreconciled by Characteristics of Interest and Regions

Characteristic of Interest	Regions				
	Providenciales	Grand Turk	North Caicos	South Caicos	ALL Regions
Age	60.96%	20.90%	32.26%	28.57%	48.70%
Place of Birth - Turks and Caicos	60.96%	49.25%	32.26%	28.57%	54.18%
Ethnicity - African/Negro/Black	60.96%	20.90%	32.26%	28.57%	48.70%
Number of Live Births equal Zero	70.13%	45.00%	57.14%	66.67%	64.55%
Highest Level of Education	67.98%	50.75%	51.61%	52.38%	62.25%

Table 3.1 Percent Identically Reported in the Census and the Post Enumeration Survey by Selected Characteristics of Interest and Regions

Characteristic of Interest	Regions				
	Providenciales	Grand Turk	North Caicos	South Caicos	All Regions
Age					
Less than 1 Year	0.0	0.0
1 - 4 Years	83.3	90.0	100.0	100.0	89.5
15 - 19 Years	66.7	100.0	100.0	66.7	76.9
35 - 39 Years	37.5	100.0	0.0	...	40.0
45 - 49 Years	77.8	50.0	0.0	...	66.7
Place of Birth - Turks and Caicos	50.6	100.0	100.0	100.0	63.9
Ethnicity - African/Negro/Black	96.3	100.0	100.0	100.0	98.1
Number of Live Births equal Zero	100.0	100.0	...	100.0	100.0
Highest Level of Education					
Primary School	61.1	100.0	57.1	60.0	66.7
Secondary School	69.7	76.5	75.0	100.0	73.8

**Table 3.2 Gross Difference Rates between Census and Post Enumeration Survey
by Selected Characterists of Interest and Regions**

Characteristic of Interest	Regions				
	Providenciales	Grand Turk	North Caicos	South Caicos	All Regions
Age					
Less than 1 Year	1.1	1.9	0.0	6.7	1.7
1 - 4 Years	4.5	5.7	4.8	6.7	5.1
15 - 19 Years	3.4	1.9	0.0	6.7	2.8
35 - 39 Years	7.9	1.9	4.8	6.7	5.6
45 - 49 Years	2.2	1.9	9.5	6.7	3.4
Place of Birth - Turks and Caicos	44.9	85.3	0.0	0.0	43.4
Ethnicity - African/Negro/Black	3.4	0.0	0.0	0.0	1.7
Number of Live Births equal Zero	4.3	0.0	0.0	0.0	2.6
Highest Level of Education					
Primary School	20.5	6.1	20.0	20.0	16.8
Secondary School	27.4	12.1	20.0	30.0	22.9

Formulation for Indicator:

Gross Difference Rate:

$$(b + c)/n * 100$$

Table 3.3 Ratio of Gross Difference to Identically Identified between Census and PES by Selected Characteristics of Interest and Regions

Characteristic of Interest	Regions				
	Providenciales	Grand Turk	North Caicos	South Caicos	All Regions
Less than 1 Year
1 - 4 Years	80.0%	33.3%	50.0%	100.0%	52.9%
15 - 19 Years	75.0%	100.0%	0.0%	50.0%	50.0%
35 - 39 Years	233.3%	100.0%	250.0%
45 - 49 Years	28.6%	100.0%	75.0%
Place of Birth - Turks and Caicos	100.0%	1450.0%	0.0%	0.0%	100.0%
Ethnicity - African/Negro/Black	3.9%	0.0%	0.0%	0.0%	1.9%
Number of Live Births equal Zero	33.3%	0.0%	...	0.0%	14.3%
Highest Level of Education					
Primary School	136.4%	33.3%	75.0%	66.7%	91.7%
Secondary School	87.0%	30.8%	50.0%	100.0%	66.7%

Table 3.4 Simple Random Response Variance between Census and Post Enumeration Survey by Selected Characteristics of Interest and Regions

Characteristic of Interest	Regions				
	Providenciales	Grand Turk	North Caicos	South Caicos	All Regions
Age					
Less than 1 Year	0.6	0.9	0.0	3.3	0.8
1 - 4 Years	2.2	2.8	2.4	3.3	2.5
15 - 19 Years	1.7	0.9	0.0	3.3	1.4
35 - 39 Years	3.9	0.9	2.4	3.3	2.8
45 - 49 Years	1.1	0.9	4.8	3.3	1.7
Place of Birth - Turks and Caicos	22.5	42.6	0.0	0.0	21.7
Ethnicity - African/Negro/Black	1.7	0.0	0.0	0.0	0.8
Number of Live Births equal Zero	2.2	0.0	0.0	0.0	1.3
Highest Level of Education					
Primary School	10.3	3.0	10.0	10.0	8.4
Secondary School	13.7	6.1	10.0	15.0	11.5

Formulation for Indicator:

Random Response Variance:

$$(b + c)/2n * 100$$

**Table 3.5 Index of Gross Shift between Census and Post Enumeration Survey
by Selected Characterists of Interest and Regions**

Characteristic of Interest	Regions				
	Providenciales	Grand Turk	North Caicos	South Caicos	All Regions
Age					
Less than 1 Year	1.00	3.00
1 - 4 Years	0.67	0.30	0.50	1.00	0.47
15 - 19 Years	0.50	1.00	0.00	0.33	0.38
35 - 39 Years	0.88	1.00	1.00	...	1.00
45 - 49 Years	0.22	0.50	2.00	...	0.50
Place of Birth - Turks and Caicos	0.51	14.50	0.00	0.00	0.64
Ethnicity - African/Negro/Black	0.04	0.00	0.00	0.00	0.02
Number of Live Births equal Zero	0.33	0.00	...	0.00	0.14
Highest Level of Education					
Primary School	0.83	0.33	0.43	4.00	0.61
Secondary School	0.61	0.24	0.38	1.00	0.49

Formulation for Indicator:

Index of Gross Shift:

$$(b + c)/(a + b)$$

**Table 3.6 Index of Inconsistency between Census and Post Enumeration Survey
by Selected Characteristics of Interest and Regions**

Characteristic of Interest	Regions				
	Providenciales	Grand Turk	North Caicos	South Caicos	All Regions
Age					
Less than 1 Year	0.5	0.5	0.8
1 - 4 Years	0.3	0.2	0.2	0.3	0.2
15 - 19 Years	0.3	0.3	0.0	0.3	0.2
35 - 39 Years	0.7	0.3	...	0.5	0.7
45 - 49 Years	0.2	0.5	1.1	0.5	0.3
Place of Birth - Turks and Caicos	0.9	5.3	0.0	0.0	0.9
Ethnicity - African/Negro/Black	0.144	0.000	0.000	0.000	0.078
Number of Live Births equal Zero	0.2	0.0	0.1
Highest Level of Education					
Primary School	0.5	0.2	0.5	0.5	0.4
Secondary School	0.6	0.3	0.4	0.6	0.5

Formulation for Indicator:

Index of Inconsistency:

$$(b + c)/(2npq)$$

where,

$$p = (a + c)/n, \text{ and } q = 1 - p$$

**Table 3.7 Net Difference Rates between Census and Post Enumeration Survey
by Selected Characteristics of Interest and Regions**

Characteristic of Interest	Regions				
	Providenciales	Grand Turk	North Caicos	South Caicos	All Regions
Age					
Less than 1 Year	-1.1	-1.9	0.0	6.7	-0.6
1 - 4 Years	-2.2	-1.9	-4.8	-6.7	-2.8
15 - 19 Years	1.1	-1.9	0.0	6.7	0.6
35 - 39 Years	3.4	-1.9	4.8	-6.7	1.1
45 - 49 Years	2.2	1.9	0.0	-6.7	1.1
Place of Birth - Turks and Caicos	42.7	-85.3	0.0	0.0	5.7
Ethnicity - African/Negro/Black	3.4	0.0	0.0	0.0	1.7
Number of Live Births equal Zero	-4.3	0.0	0.0	0.0	-2.6
Highest Level of Education					
Primary School	-1.4	-6.1	20.0	20.0	1.5
Secondary School	0.0	12.1	6.7	-30.0	1.5

Formulation for Indicator:

Net Difference Rate:

(b - c)/n *100

**Table 3.8 Index of Net Shift between Census and Post Enumeration Survey
by Selected Characteristics of Interest and Regions**

Characteristic of Interest	Regions				
	Providenciales	Grand Turk	North Caicos	South Caicos	All Regions
Age					
Less than 1 Year	1.0	-0.5
1 - 4 Years	-0.3	-0.1	-0.5	-1.0	-0.2
15 - 19 Years	0.2	-1.0	0.0	0.3	0.1
35 - 39 Years	0.4	-1.0	1.0	...	0.3
45 - 49 Years	0.2	0.5	0.0	...	0.2
Place of Birth - Turks and Caicos	0.5	-14.5	0.0	0.0	0.1
Ethnicity - African/Negro/Black	0.0	0.0	0.0	0.0	0.0
Number of Live Births equal Zero	-0.3	0.0	...	0.0	-0.1
Highest Level of Education					
Primary School	-0.1	0.3	0.4	0.4	0.1
Secondary School	0.0	0.2	0.1	-1.0	0.0

Formulation for Indicator:

Index of Net Shift:

(b - c)/(a + b)

APPENDIX III

Chart 3.0 Percentage of Observations Unreconciled by Characteristics of Interest and Regions

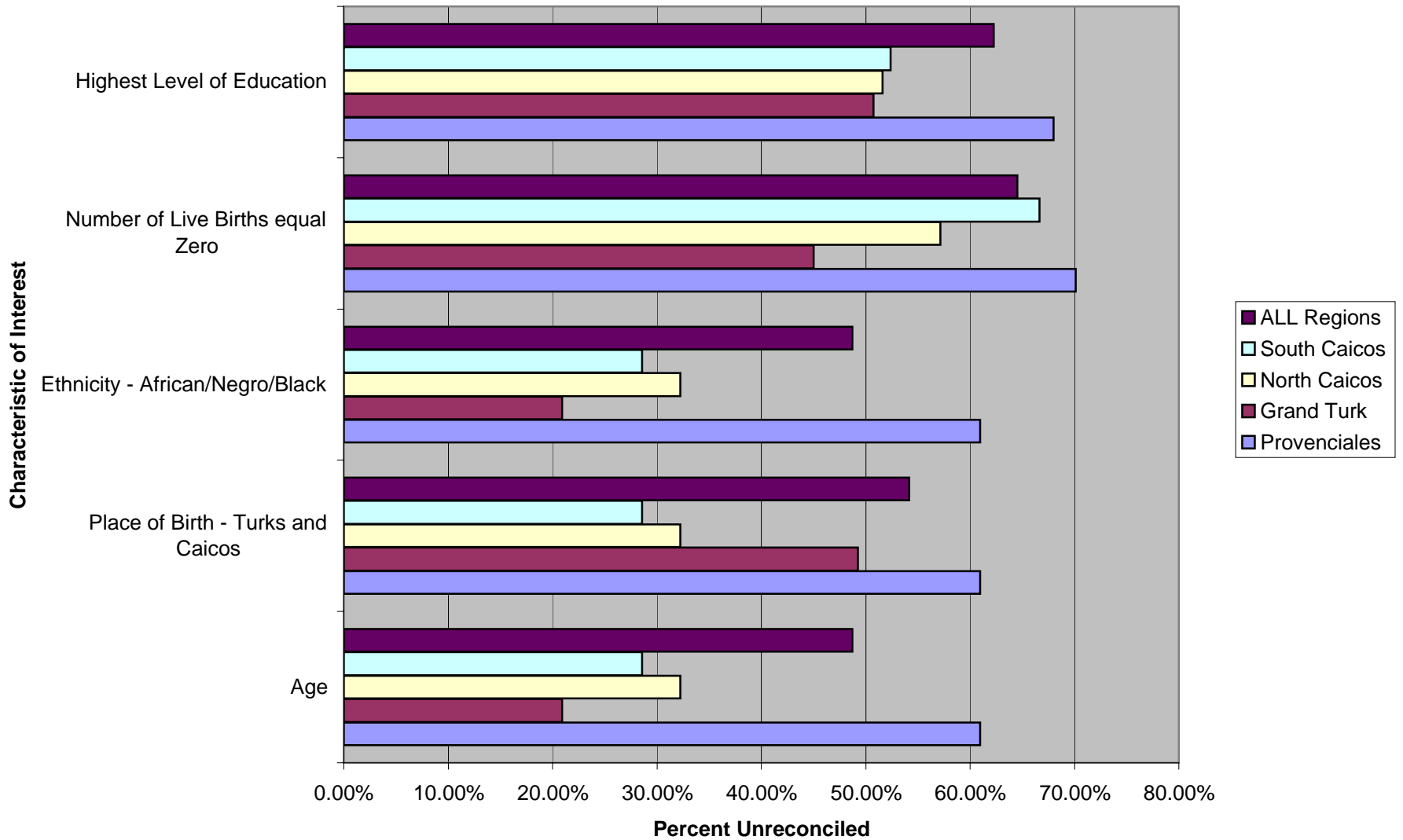
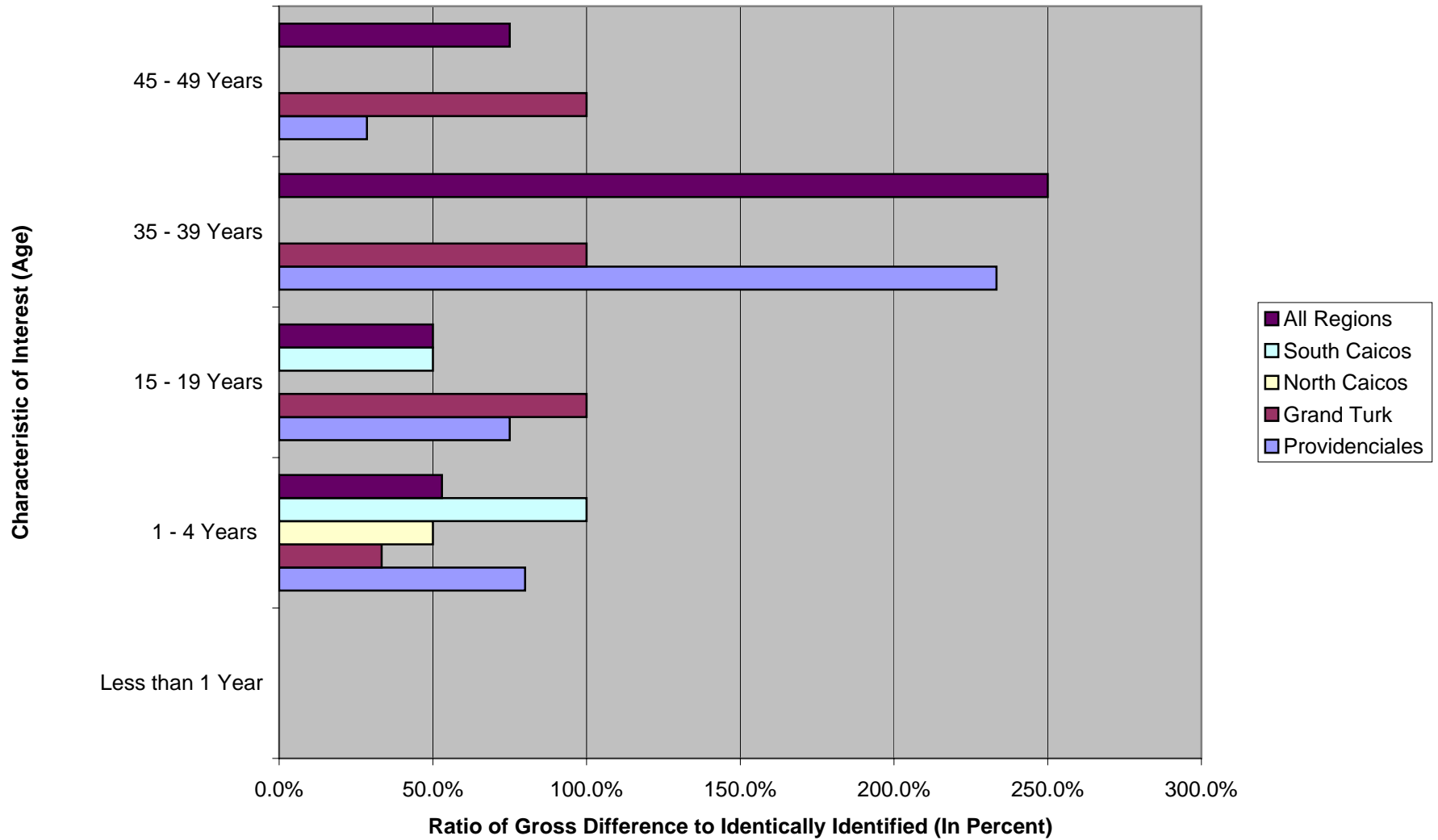


Chart 3.3(a) Ratio of Gross Difference to Identically Identified for Selected Age Groups by Regions



**Chart 3.3(b) Ratio of Gross Difference to Identically Identified for Place of Birth - TCI
by Regions**

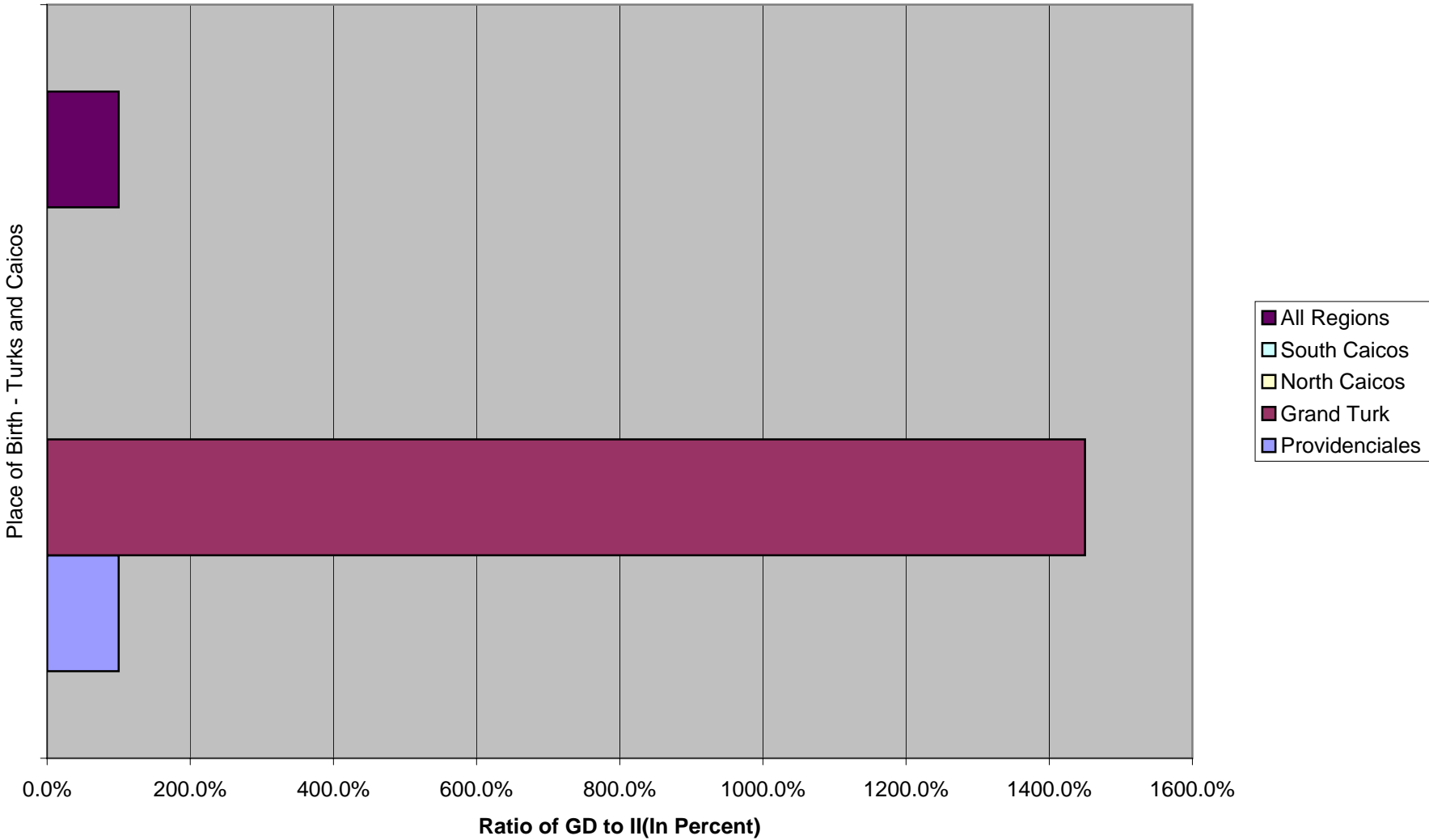


Chart 3.3(c) Gross Difference to Identically Distributed for Ethnic Group-African by Regions

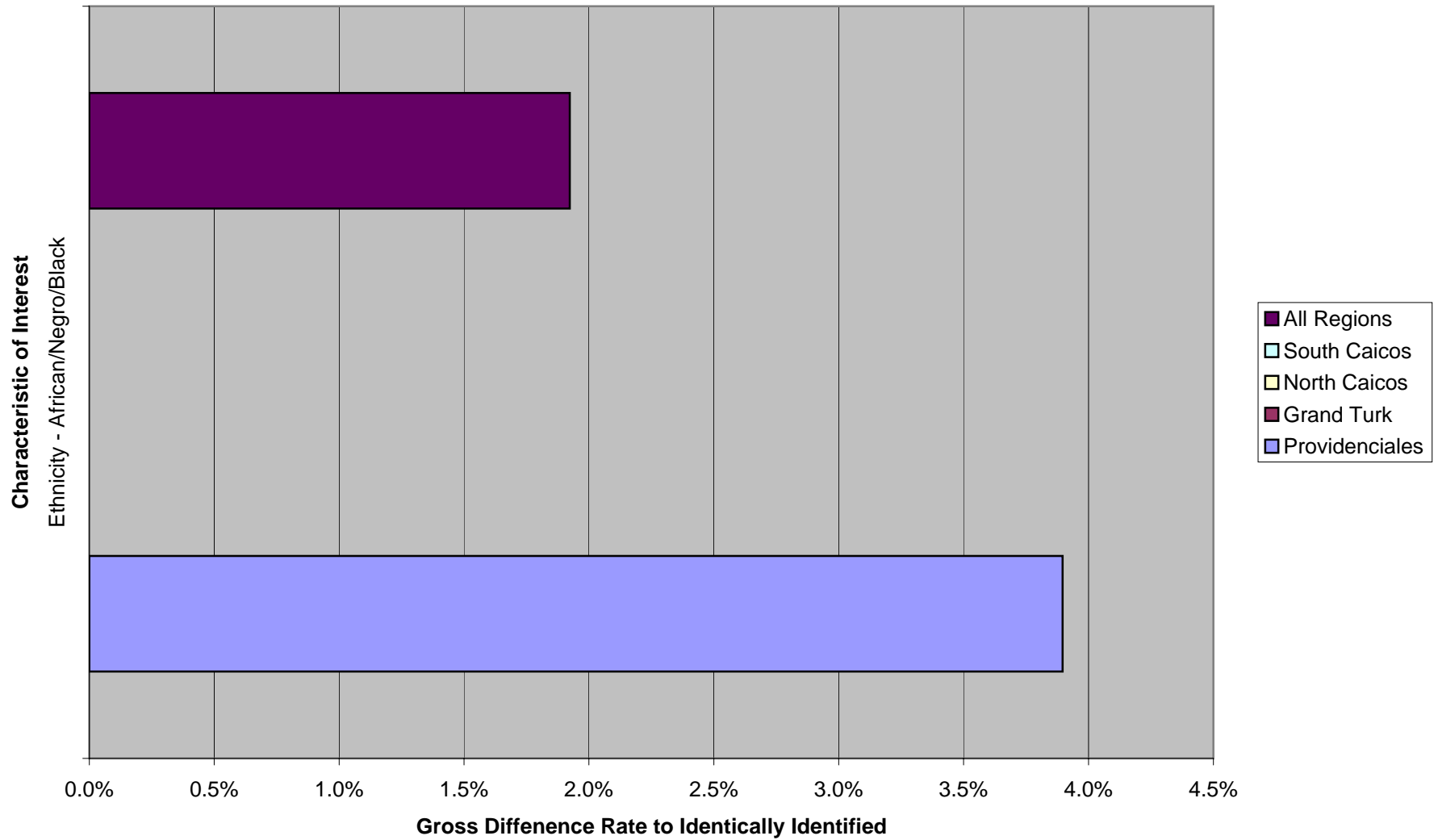


Chart 3.3(d) Ratio of Gross Difference to Identically Identified for Fertility-Zero Life Birth by Regions

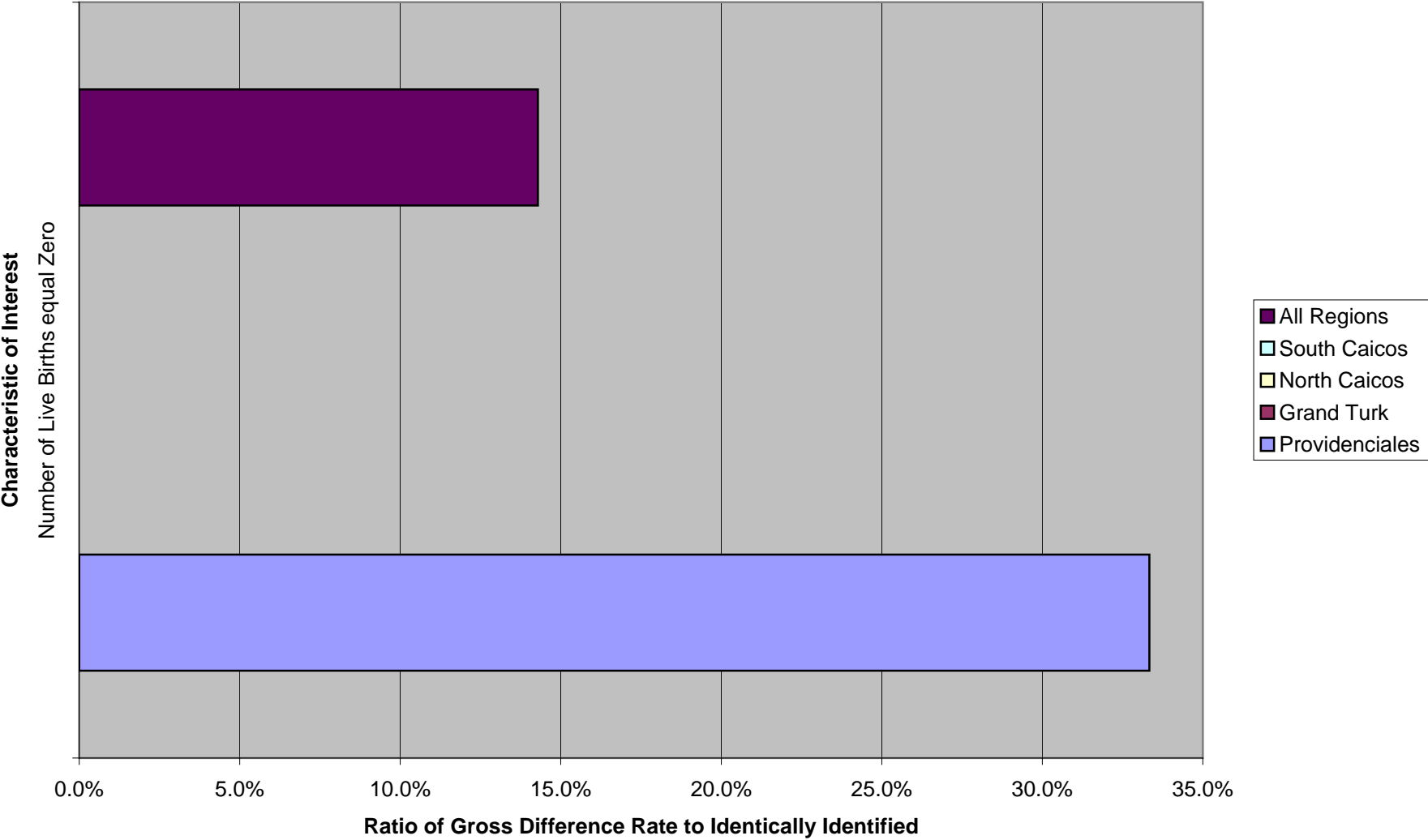


Chart 3.3(e) Ratio of Gross Difference Rate to Identically Identified for Highest Level of Education - Primary/Secondary by Regions

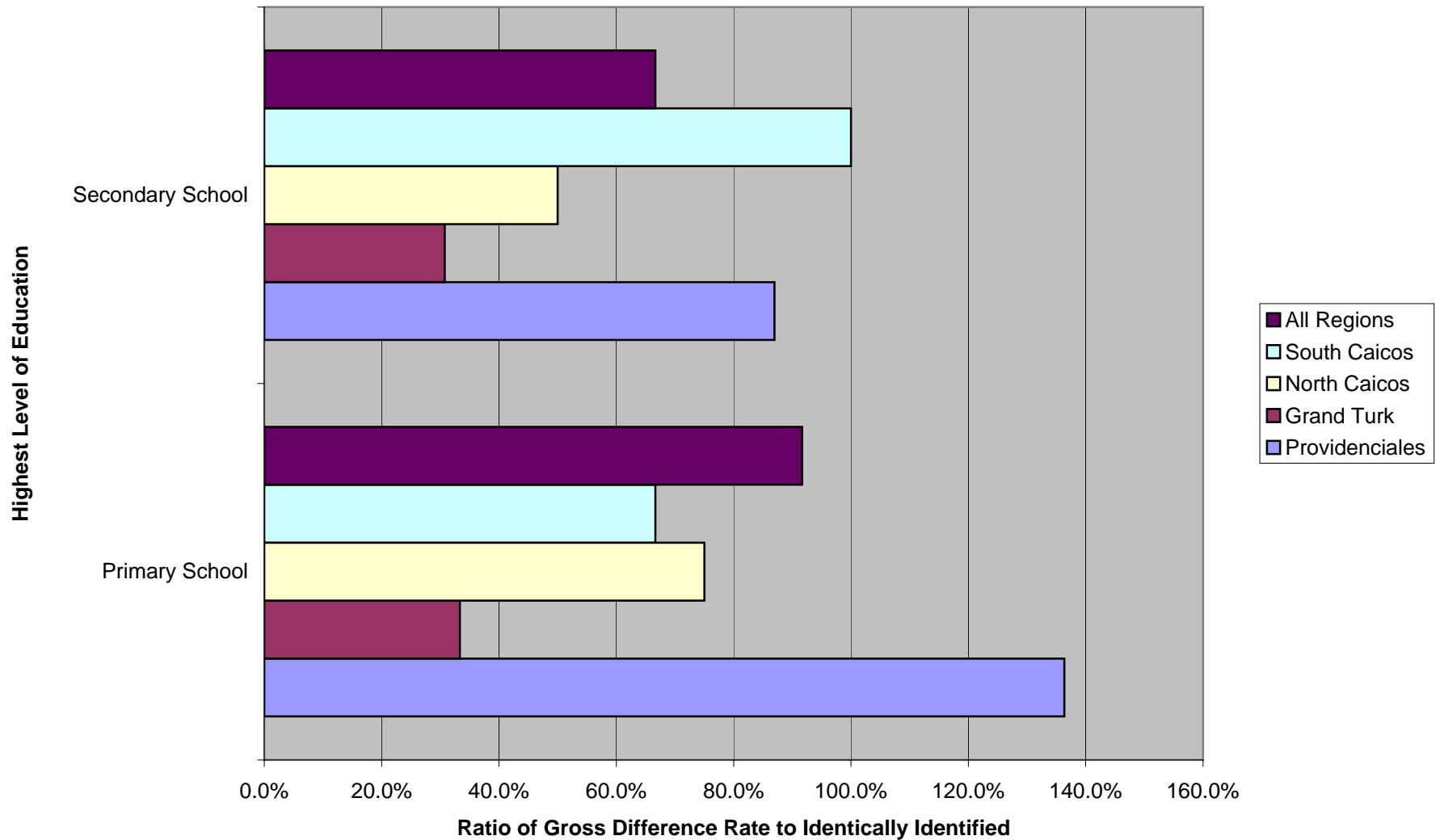


Chart 3.5(a) Index of Gross Shift Between the Census and the PES for Selected Age Group and Regions (In Percent)

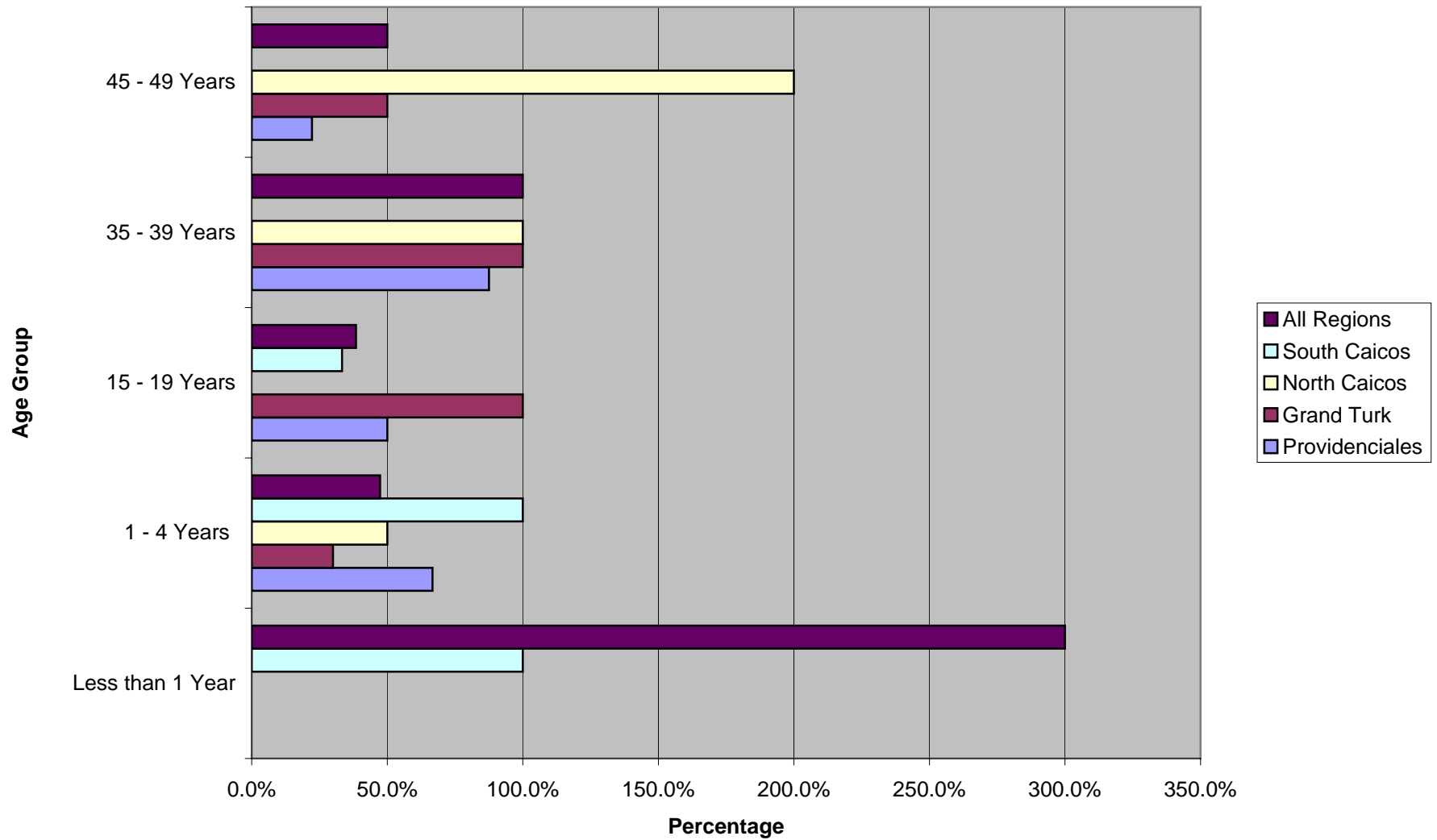


Chart 3.5(b) Index of Gross Shift between Census and PES for Place of Birth-TCI and Regions

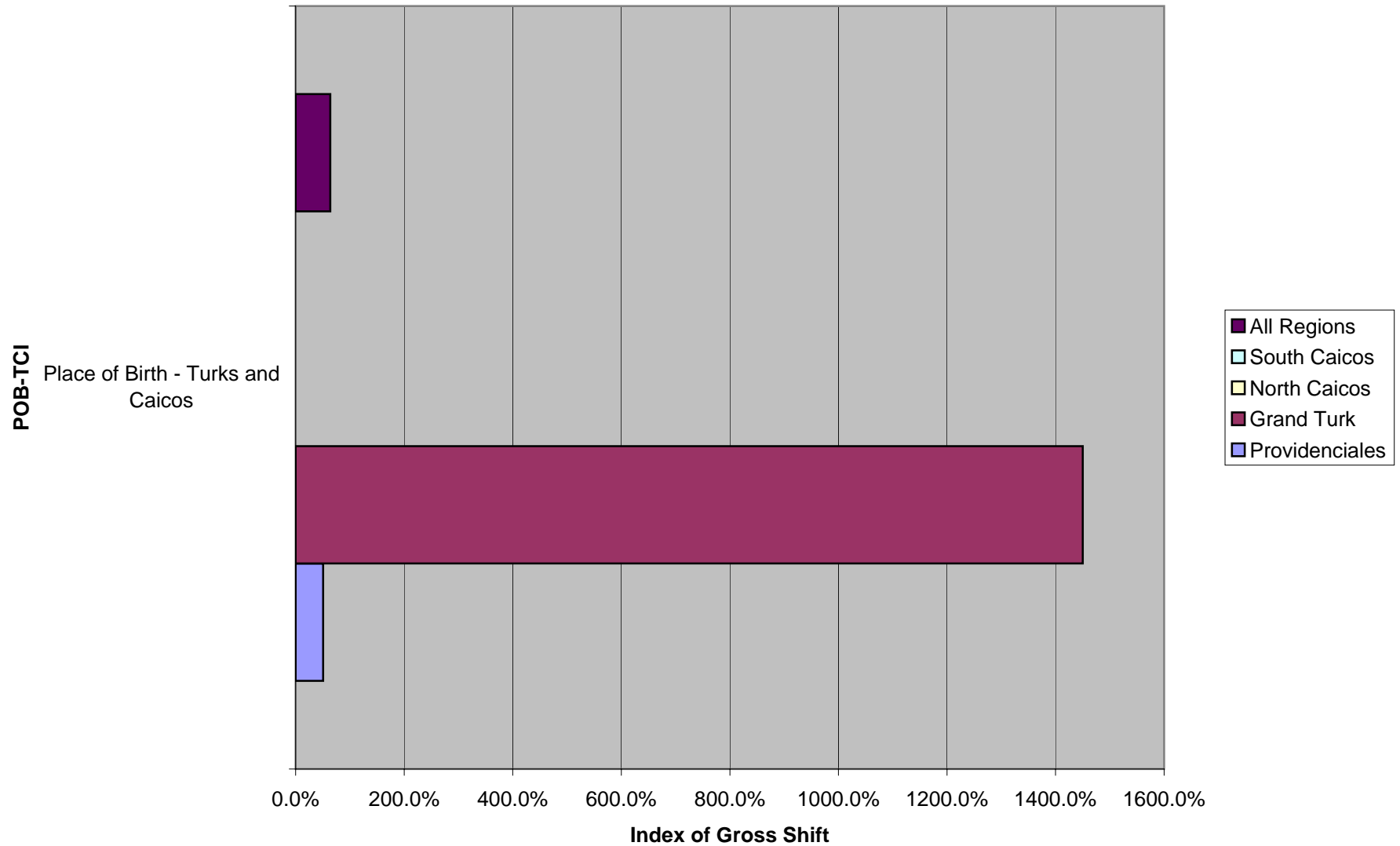


Chart 3.5(c) Index of Gross Shift between Census and PES for Ethnic Group-African

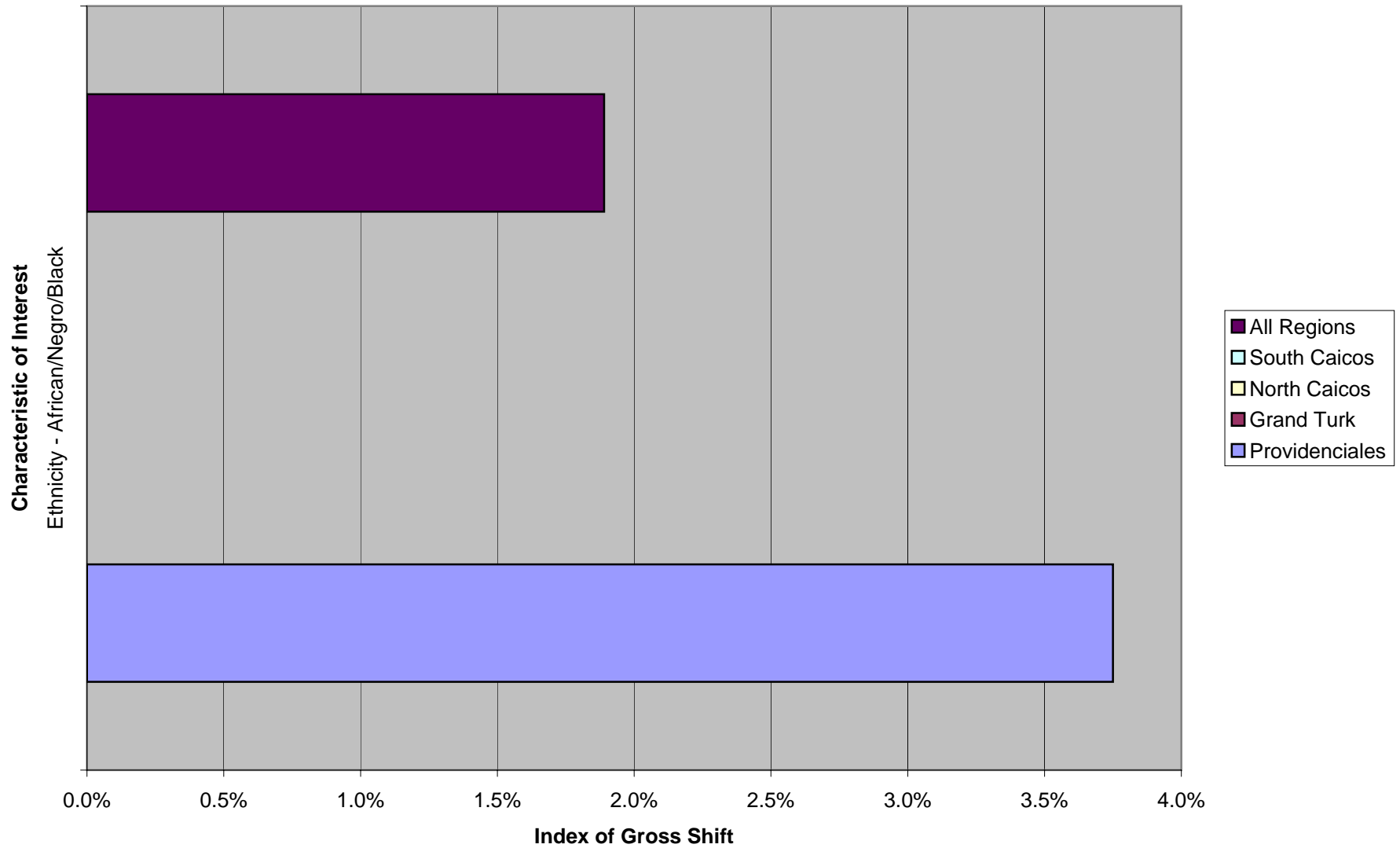


Chart 3.5(d) Index of Gross Shift between Census and PES for Number of Live Births equals Zero and Regions

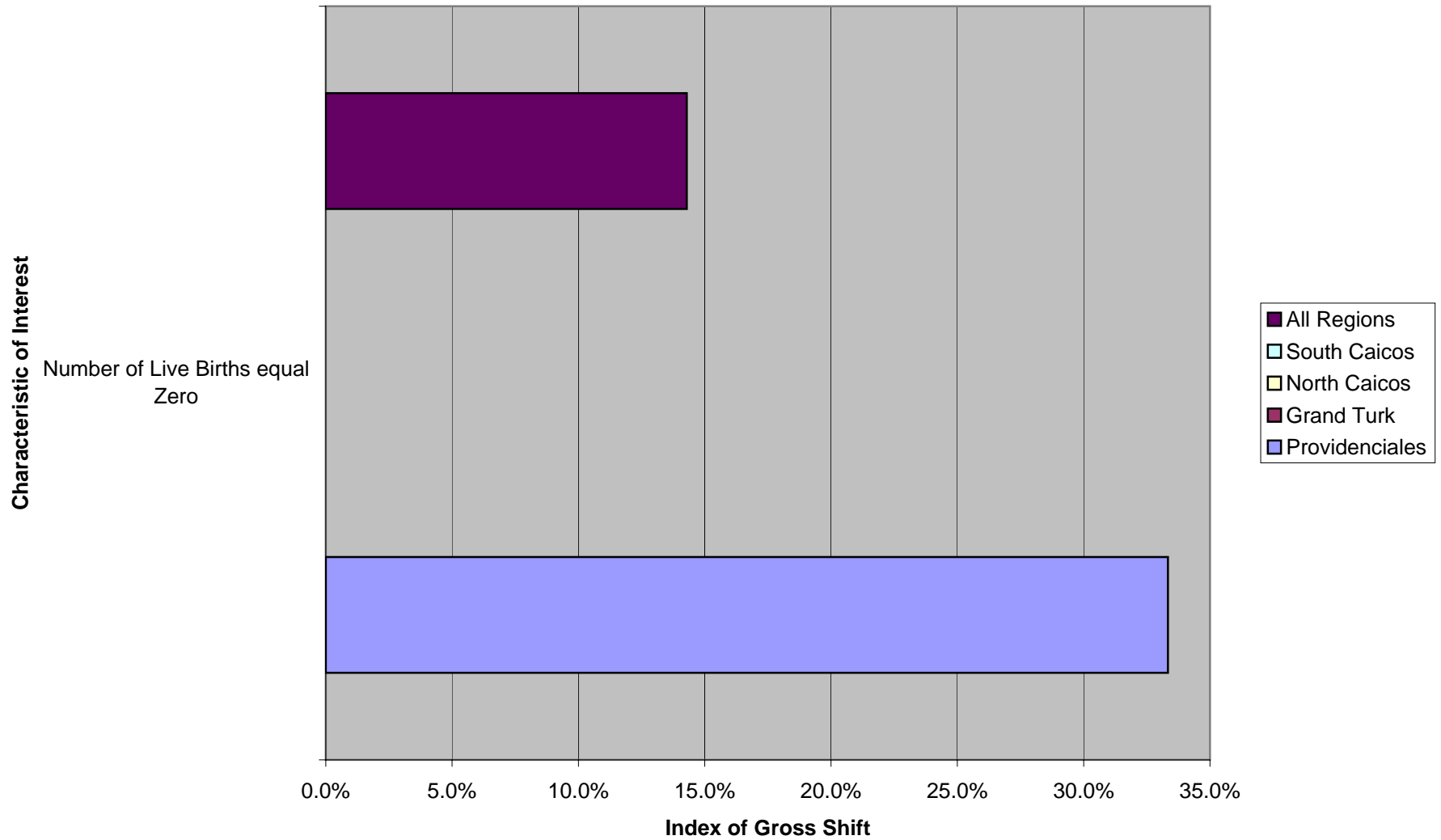


Chart 3.5(e) Index of Gross Shift between Census and PES for Highest Education Level given as Primary/Secondary School

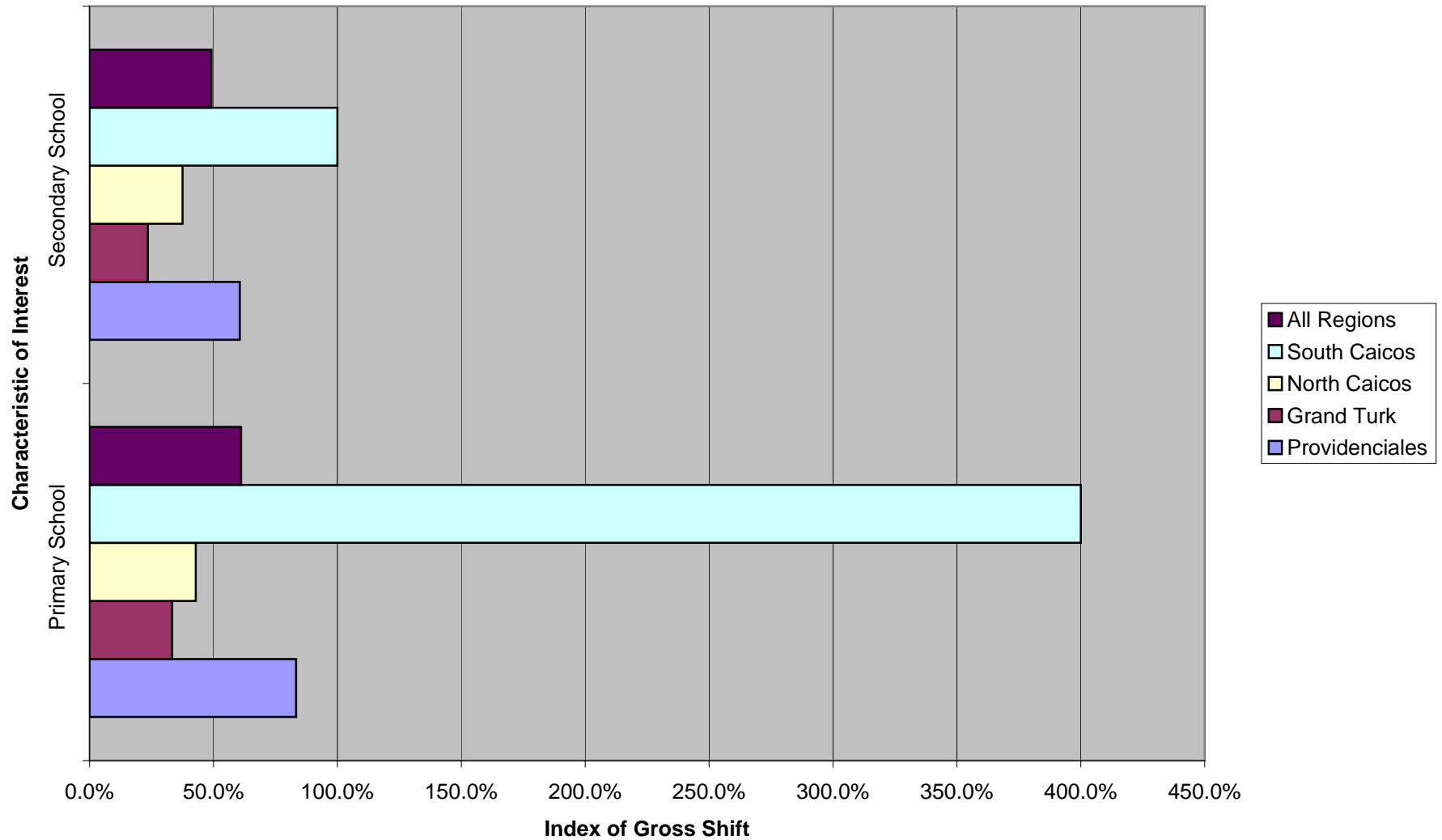


Chart 3.6(a) Index of Inconsistency between Census and PES for Selected Age Groups by Regions

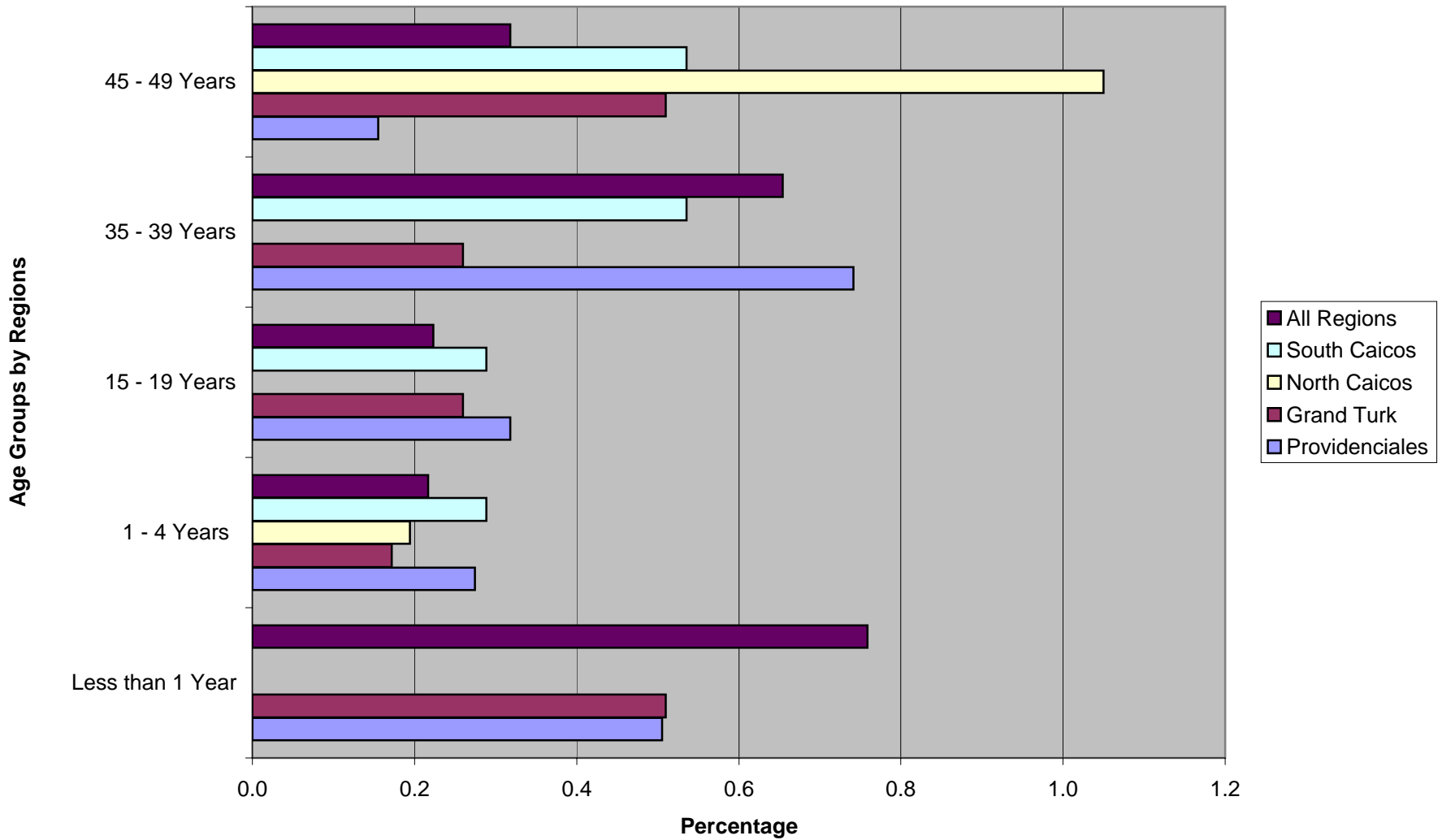


Chart 3.6(b) Index of Inconsistency between Census and PES for Place of Birth - TCI by Regions

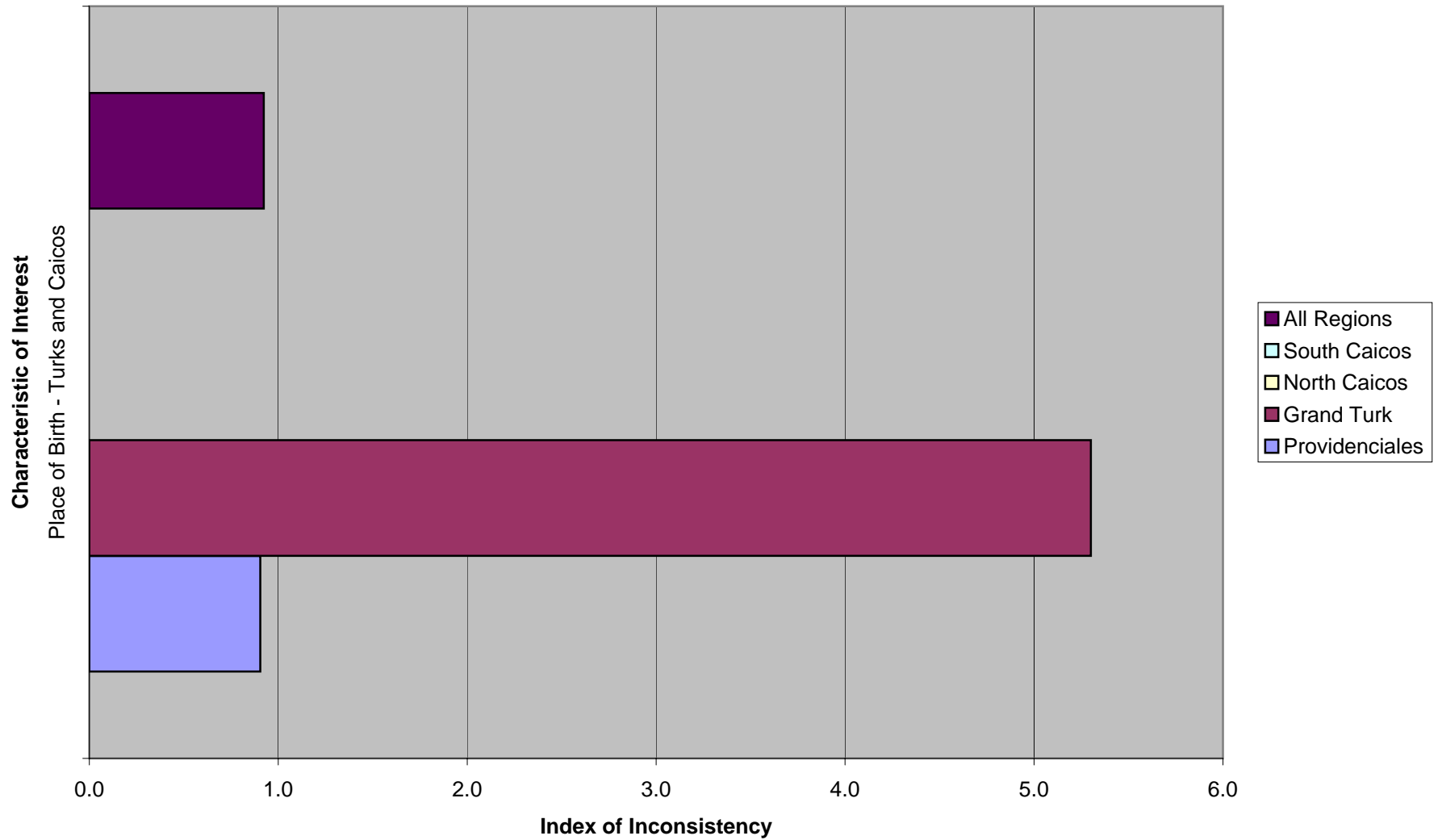


Chart 3.6(c) Index of Inconsistency between Census and PES for Ethnic Group-African by Regions

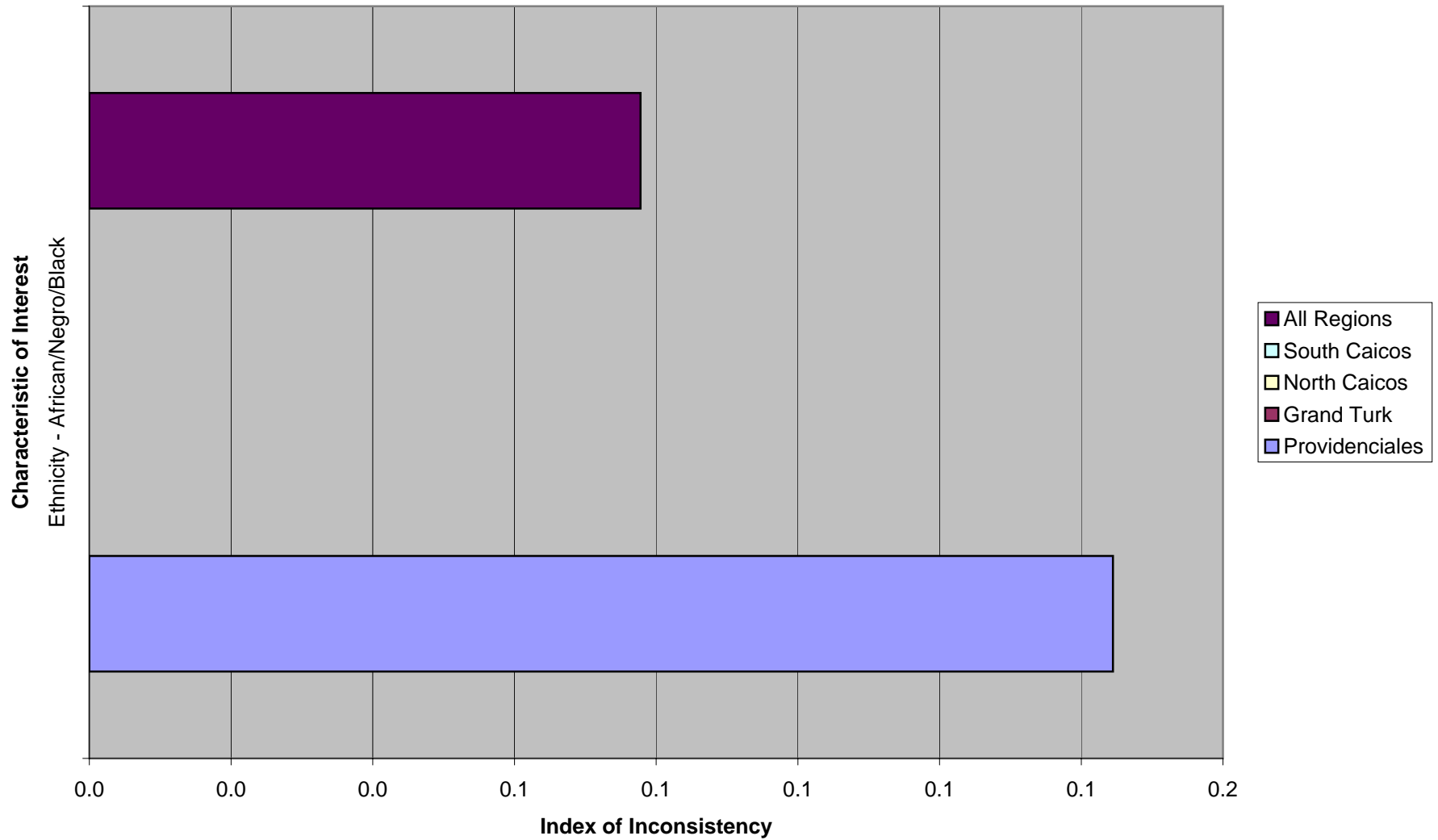


Chart 3.6(d) Index of Inconsistency for the Number of Live Births Equals Zero by Regions

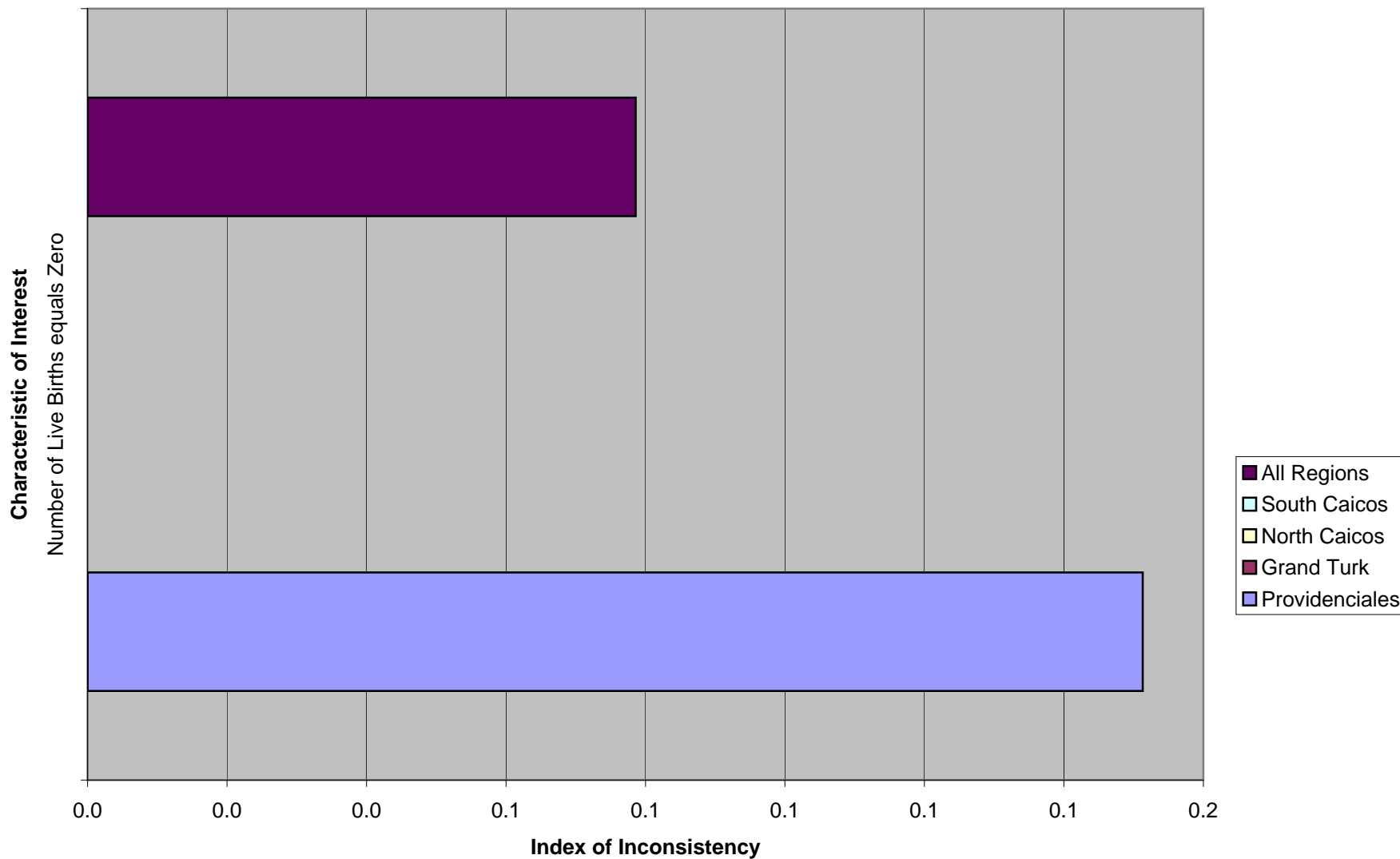


Chart 3.6(e) Index of Inconsistency for Highest Education Level given as Primary/Secondary School by Regions

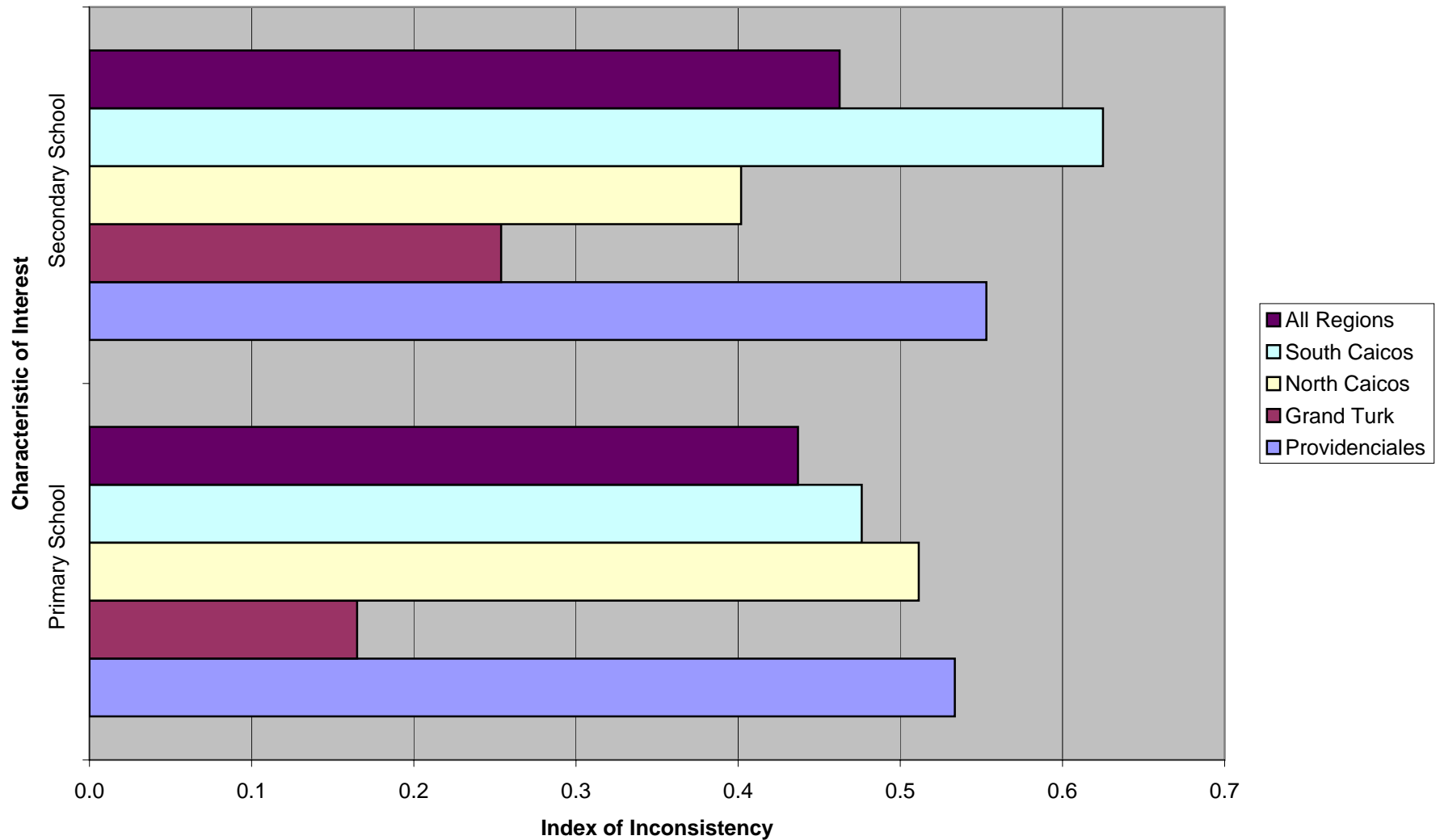


Chart 3.8(a) Index of Net Shift between Census and PES for Selected Age Group and Regions

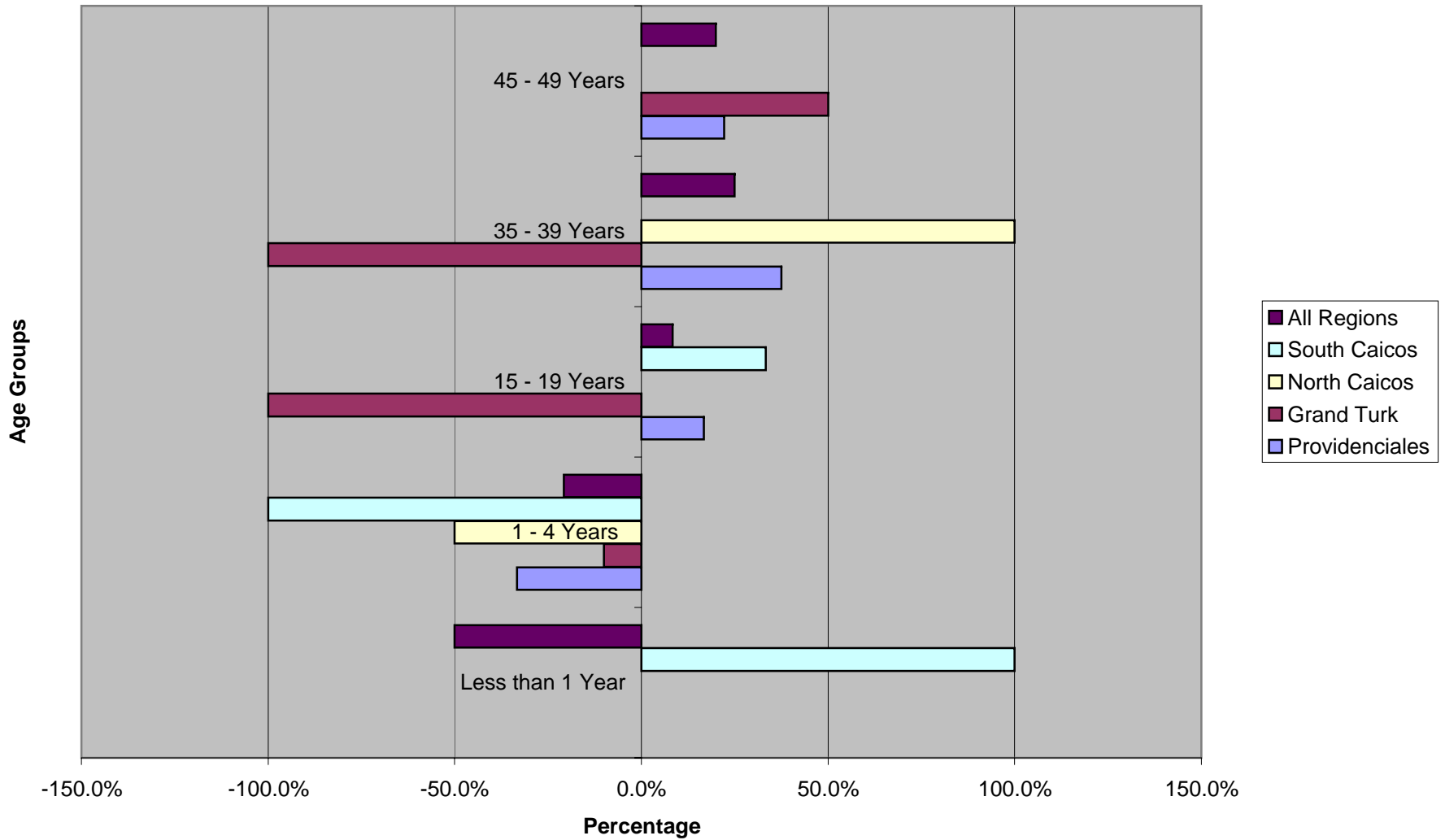


Chart 3.8(b) Index of Net Shift between Census and PES for Place of Birth-TCI

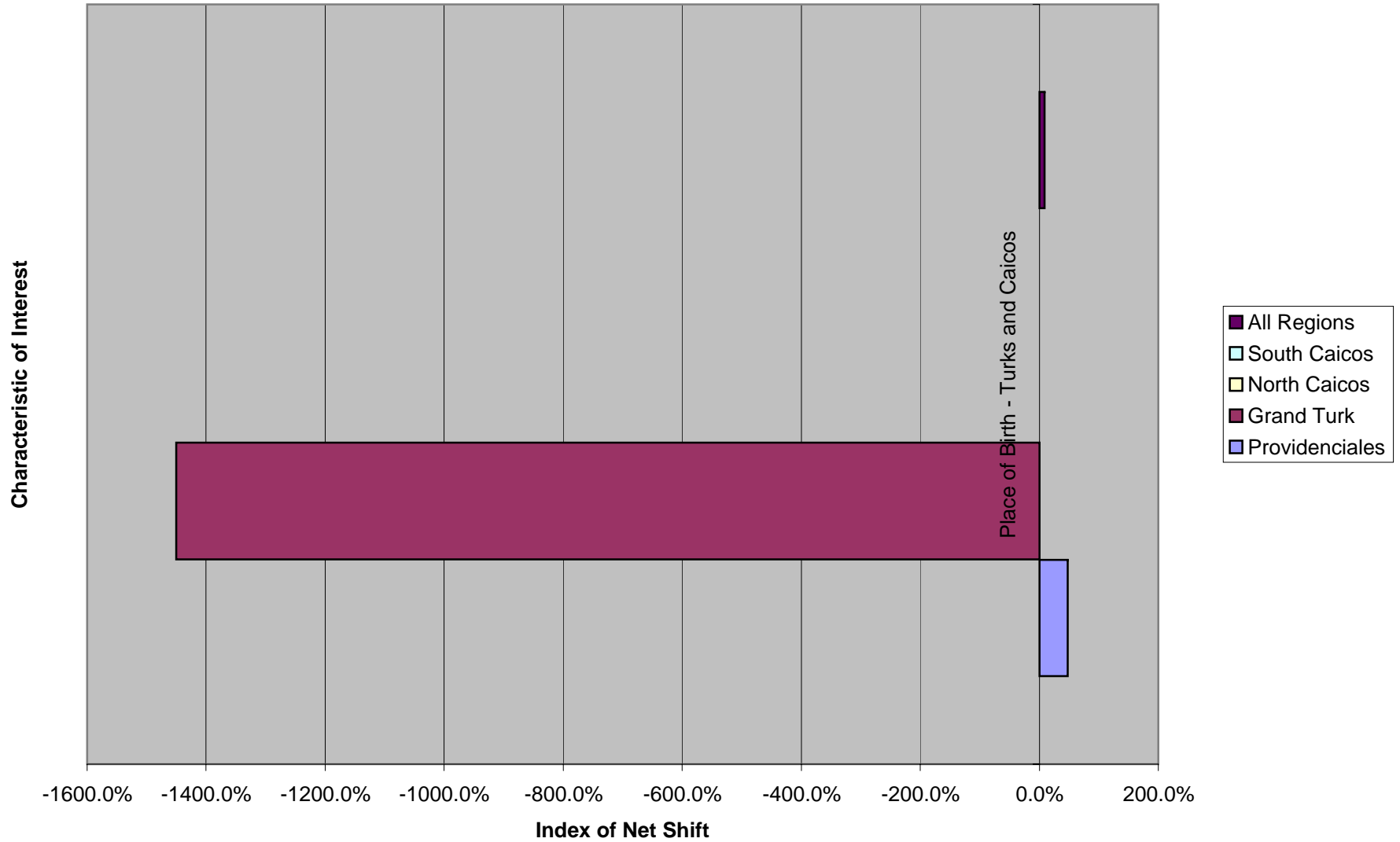


Chart 3.8(c) Index of Net Shift between Census and PES for Ethnic Group-African

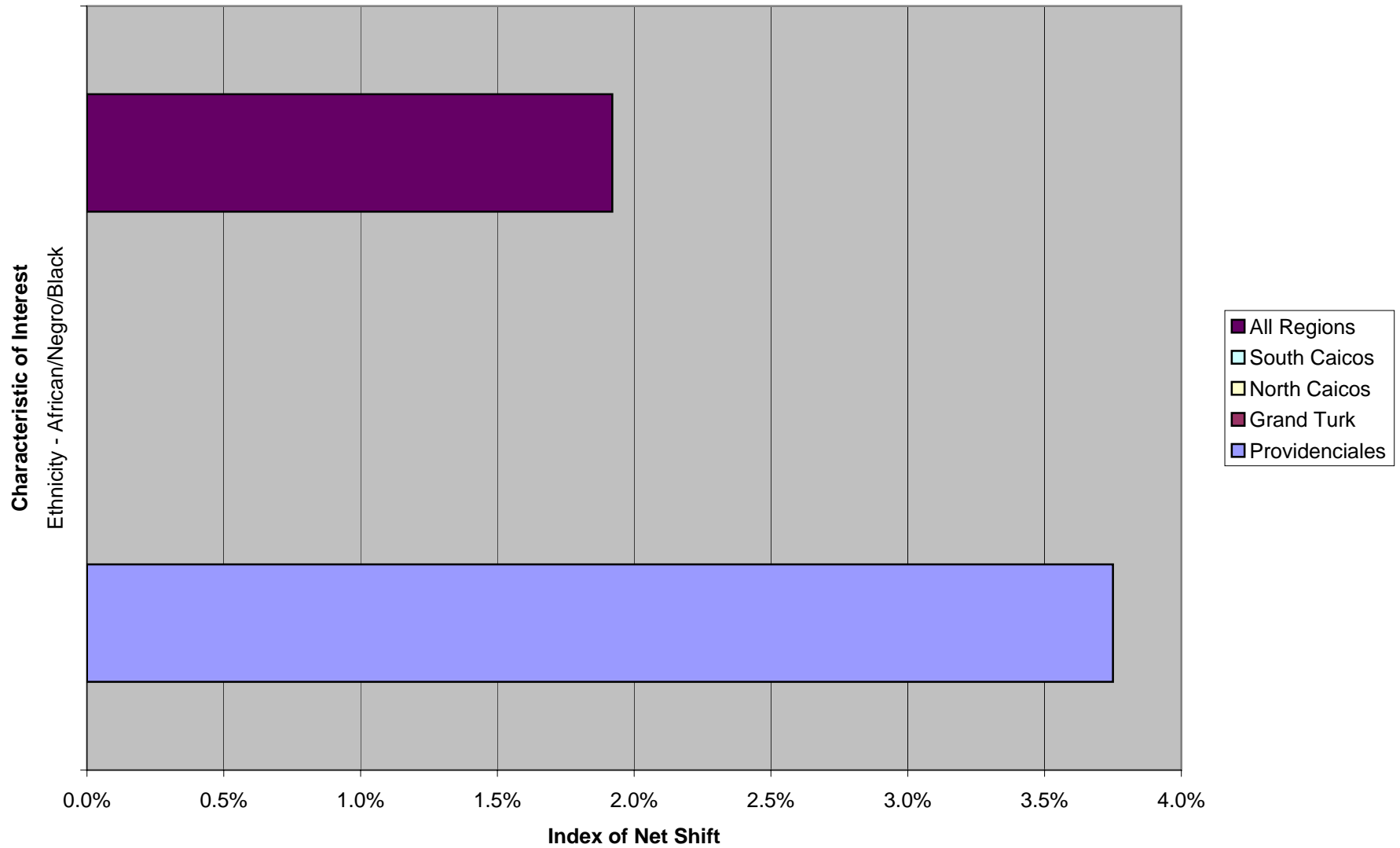
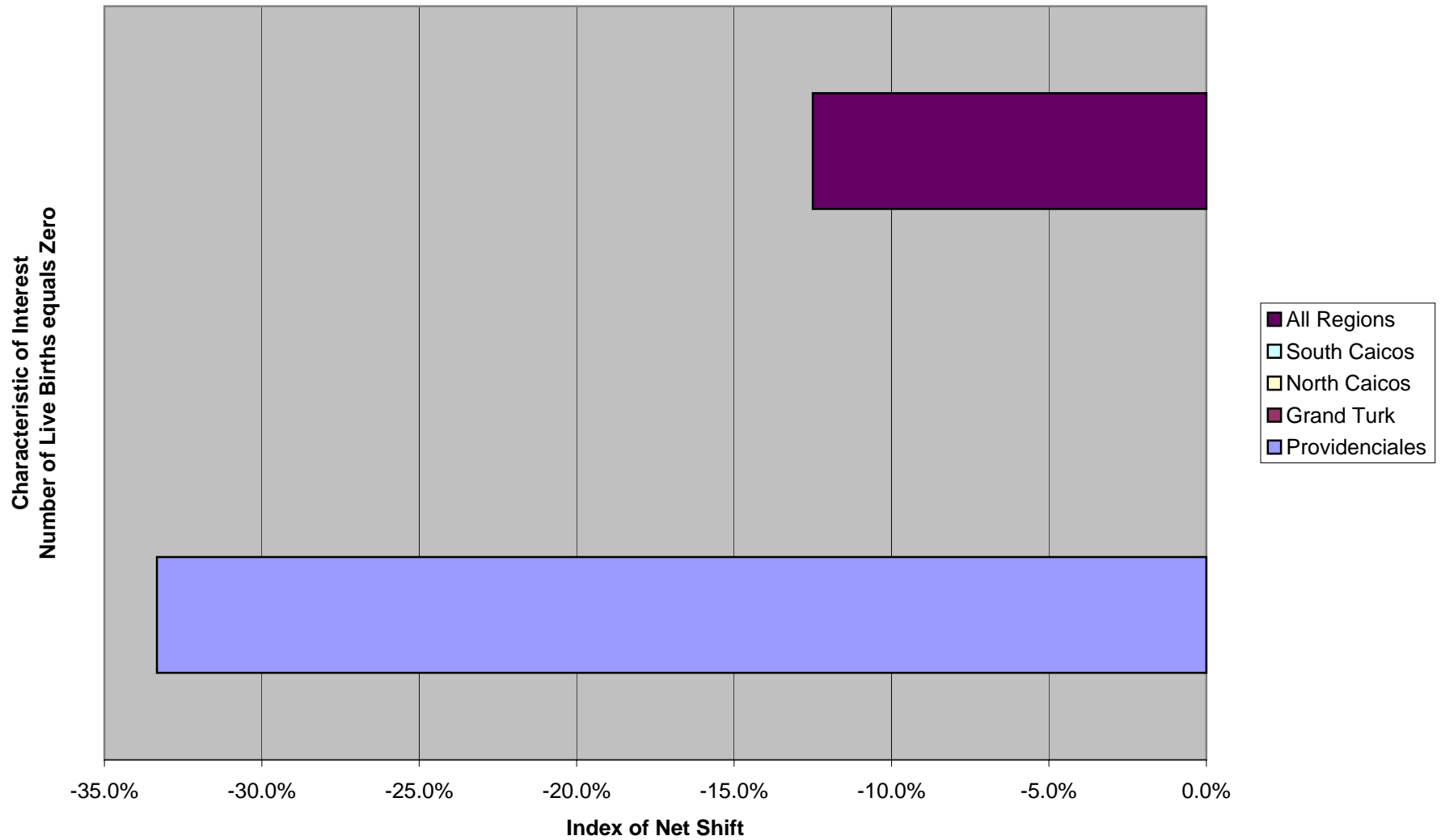
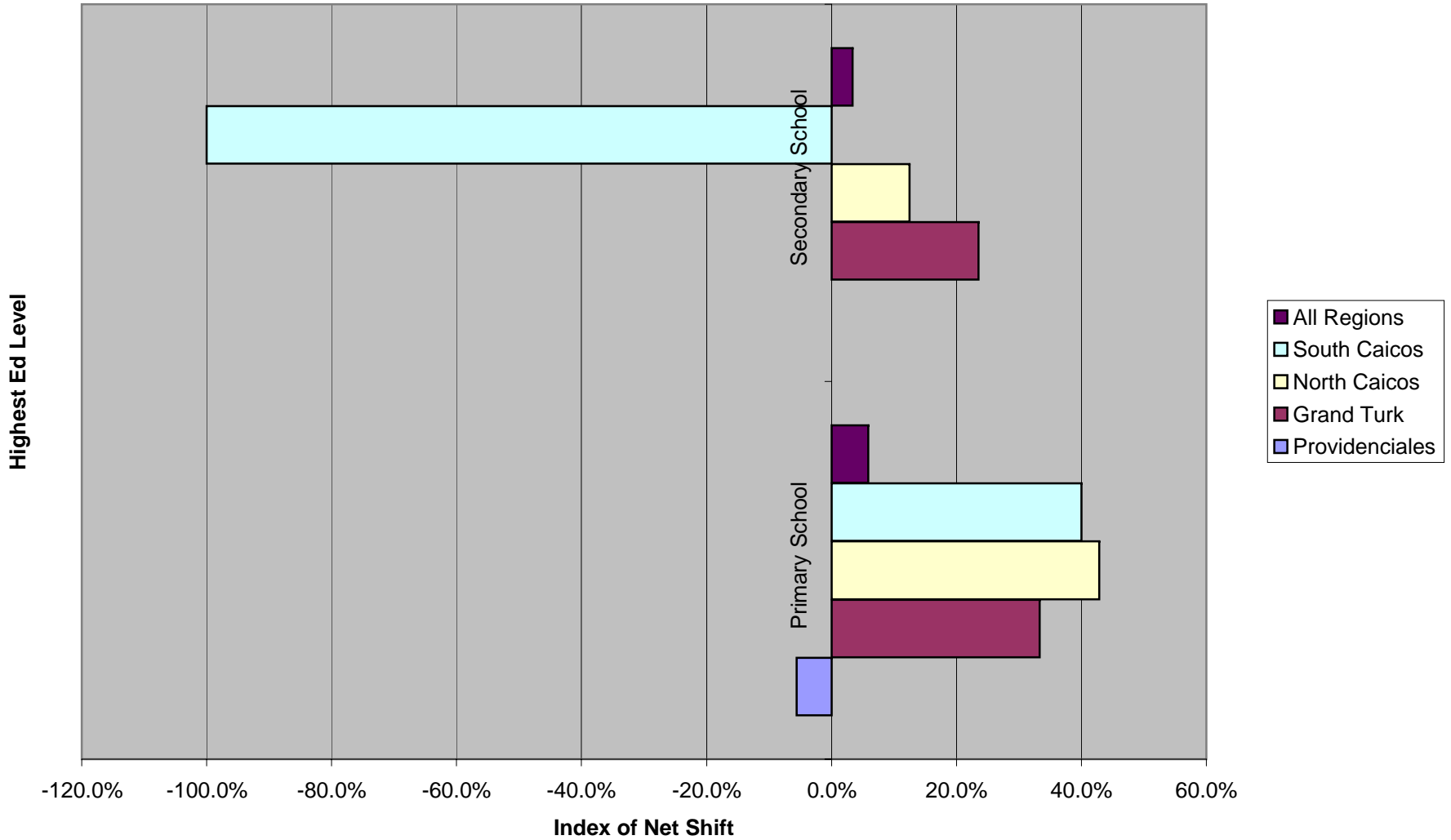


Chart 3.8(d) Index of Net Shift between Census and PES for Number of Live Births Equals Zero



**Chart 3.8(e) Index of Net Shift between Census and PES for Highest Education Level-
Primary/Secondary School**



APPENDIX IV

Model for the Evaluation of Response Errors

The Model

The model used in the evaluation of **response errors** in this study and which is now presented here assumes independence of measurements on identical persons in the Census and the PES. For example, an identical respondent in the PES might have decided to withhold the correct response and repeat the same but erroneous response given in the Census. That course of action might have been predicated on maintaining a consistent response on the part of the respondent. In the above scenario, since the response on the second trial (the PES) was influenced by the response on the first trial (the Census) for the identical person, we can conclude that measurements on both occasions lacked independence. In addition, the model presupposes that the time interval between the conduct of the Census and the PES is relatively short (say, one week).

The essential features of the model are the summary measures or **indicators**, which are constructed to analyze differences affecting the tabulated figures of the Census. The differences affecting census tabulated figures are encountered during a case-by-case comparison of results of characteristics of interest common to both the Census and the PES.

Formulation of the Model

Assume that following the PES one wishes to classify a sample of n^1 individuals as having or not having a characteristic of interest. The classification process may be accomplished as follows:

Classify “a” sample-persons as the number of persons having the characteristic of interest in both the PES and the CENSUS.

Classify “b” sample-persons as the number of persons having the characteristic of interest in the PES and not in the CENSUS.

Classify “c” sample-persons as the number of persons having the characteristic of interest in the CENSUS and not in the PES.

¹ Note that for each individual in the classification there must be a response to the characteristic of interest in both the PES and CENSUS. Otherwise, the individual (observation) is excluded from the analysis.

Classify “d” sample-persons as the number of persons not having the characteristic of interest in both the CENSUS and the PES.
 Below, is a diagrammatic representation of the classification system outlined above.

Diagram A. Representation of Results of the PES and the Census for Identical Persons

Results of the PES	Results of the Census		
	Number having the Characteristic	Number <u>not</u> having the Characteristic	Totals
Number having the characteristic	a	b	a + b
Number <u>not</u> having the characteristic	c	d	
Totals	a + c	b + d	n = a + b + c + d

Definition of Measures

For the model given above, the algebraic expressions can be formulated as follows:

Let $x_{i,p}$ be a random variable whose value is given as follows:

$$x_{i,p} = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ identical person in the PES and Census has the} \\ & \text{characteristic of interest in the PES (p)} \\ 0, & \text{otherwise} \end{cases}$$

where, $i = 1, \dots, n$ (1.1)

Similarly, let $x_{i,c}$ be a random variable whose value is given as follows:

$$X_{i,c} = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ identical person in the PES and Census has the} \\ & \text{characteristic of interest in the Census (c)} \\ 0, & \text{otherwise} \end{cases}$$

where, $i = 1, \dots, n$ (1.2)

Then, by the classification system in Diagram A and definitions (1.1) and (1.2), above, we have,

$$a = \sum_i^n (X_{i,p}X_{i,c}); \quad a + c = \sum_i^n (X_{i,c}) \quad \text{and} \quad a + b = \sum_i^n (X_{i,p})$$
 (1.3)

From the expression in (1.3), note that the measurement variable, **a**, represents the number having the characteristic in both the PES and the Census. Measurement variables, **(a + c)** and **(a + b)** are the Census totals and the PES totals respectively. We can define the index, **r** (“Percent identically reported as having the characteristic”) as:

$$r = a/(a + b)$$
 (1.4)

Note that in expression (1.4), the PES is taken as the base² for measuring the stability of the response.

Furthermore, define the index, **g** (Gross Difference Rate) as:

$$g = 1/n \sum (X_{i,c} - X_{i,p})^2$$
 (1.5)

By expansion of equation (1.5) and using the fact that $X_{i,c}$ and $X_{i,p}$ are indicator random variables, we have:

$$g = 1/n \sum (X_{i,c}) + 1/n \sum (X_{i,p}) - 2/n \sum (X_{i,p}X_{i,c})$$
 (1.6)

By substituting expressions on the right of (1.6) with their corresponding values from (1.3), we have:

² Usually, because of the manageability of the PES due to its size and the engagement of the most skillful interviewing staff to conduct the survey, the PES is considered as providing a better measurement than the Census.

$$\begin{aligned}
g &= 1/n[(a + c) + (a + b) - 2a] \\
&= (b + c)/n
\end{aligned} \tag{1.7}$$

Using the PES as base, we note that the sum $(b + c)$ – “the gross difference”, comprises the number of cases, **b**, erroneously excluded from the classification in the Census and, **c**, the number of cases erroneously included in classification in the Census. Since these differences are in both directions, they are partly or substantially *offsetting*.

On the other hand, the difference $(b - c)$ – “the net difference,” represents the *non-offsetting* part of the gross difference. Using the net difference we can define “the net difference rate” (or “bias”), **s**, as:

$$s = (b - c)/n \tag{1.8}$$

Simple Random Response Variance (V_d)

The simple random response variance (V_d) may be described as an indicator that measures the inherent instability of responses obtained between the Census and the PES. We may obtain the indicator as follows:

Taking expectations of the random variable in (1.6) over i , we have

$$\begin{aligned}
E(g) &= E [(1/n)\sum(X_{i,c}) + (1/n)\sum(X_{i,p}) - (2/n)\sum(X_{i,p}X_{i,c})] \\
&= (1/n)\sum(p_i) + (1/n)\sum(p_i) - (2/n)\sum(p_i^2) \\
&= (2/n)\sum p_i(1-p_i)
\end{aligned} \tag{1.9}$$

Now, we define the simple random response variance (V_d) as:

$$V_d = E(d_{it} - E(d_{it}))^2 \tag{1.10}$$

where, d_{it} , the response deviation of the i -person in the t trial ($t = c, p$), is given as:

$$d_{it} = x_{it} - p_i, \tag{1.11}$$

the difference between the measurement for the i-person on the t trial and the expected value³ for that person over the trials.

From the definition of d_{it} in equation (1.11), it is obvious that $E(d_{it})$ in equation (1.10) is zero, and V_d becomes:

$$V_d = E(x_{it} - p_i)^2 \quad (1.12)$$

$$= E[x_{it}^2 + p_i^2 - 2 x_{it}p_i] \text{ (since } x_{it} \text{ is an indicator variable)} \quad (1.13)$$

Taking expectations first over trials for a fixed person and then averaging over n, we have:

$$V_d = (1/n)\sum p_i(1-p_i) \quad (1.14)$$

Substituting (1.14) in (1.9), we have:

$$E(g) = 2 V_d \quad (1.15)$$

Therefore, from (1.15), it has been established that $g/2$ estimates V_d .

Based on the construction of the basic indicators, one can now derive the following measures:

Index of gross shift

$$I_{GS} = (b + c)/(a + b) \quad (1.16)$$

The ratio of the gross difference rate ($g = (b + c)/n$) to the proportion ($P_p = (a + b)/n$) classified according to the PES.

Index of net shift

$$I_{NS} = (b - c)/(a + b) \quad (1.17)$$

The ratio of the net difference rate ($s = (b - c)/n$) to the proportion ($P_p = (a + b)/n$) classified according to the PES.

Index of Inconsistency

³ The conditional expected value for repeated measurement for the i^{th} person is

$$E(x_{it}) = p_i$$

Where p_i is the proportion of times the i^{th} person had the characteristic of interest.

$$I_1 = (b + c)/(2npq) \quad (1.18)$$

In this index $g/2 = (b + c)/2n$ is viewed as a measure of the simple random response variance and pq as the total random variance (including both random response and sampling error). I_1 represents the proportion of total variability contributed by simple random response error. In this study p is estimated by using $(a+c)/n$ and q is equal to $1-p$.

Ratio of Gross Difference to Identically Identified

$$I_{G/I} = (b + c)/a \quad (1.19)$$

In practice, the ratio to gross different rate to Identically Identified should not be more than 10%.