



International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report

IR-08-019

An Evaluation of the IIASA/VID Education-Specific Back Projections

Fernando Riosmena (Fernando.Riosmena@colorado.edu)
Isolde Prommer (prommer@iiasa.ac.at)
Anne Goujon (Anne.Goujon@oeaw.ac.at)
Samir KC (kc@iiasa.ac.at)

Approved by

Wolfgang Lutz
Leader, World Population Program

August 13, 2008

Interim Reports on work of the International Institute for Applied Systems Analysis receive only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work.

Contents

Introduction	1
Coverage of the IIASA/VID Education Database.....	2
Particularities of Each Database and Standardization Procedure.....	5
Comparisons	6
Adjustments Between IVEP-Beta and IVEP-1.0	10
Conclusions	11
References	12
Appendix A. List of countries, origin of data, adjustments made and comments relating to the back projections.....	14
Appendix B. Discrepancies in the No Education category between IVEP-Beta and UIS/DHS.....	32
Appendix C. Discrepancies in the Tertiary Education category between IVEP-Beta and UIS/DHS.....	34

Abstract

In 2007, IIASA and the Vienna Institute of Demography of the Austrian Academy of Sciences (VID) released a database reconstructing detailed information on levels of educational attainment by age (in five-year age groups from 15 to 65+ years), sex, and for every five years between 1970 and 2000 for 120 countries (see Lutz et al. 2007). This database was created in two steps. The reconstruction methodology was applied for the first time and generated what is called the Beta version of the database (unpublished). This paper presents the validation procedure that was implemented to check the plausibility of the Beta version against scattered real data from different sources, mostly from the UNESCO collection of levels of educational attainment. The verification was done by using two main indicators where a comparison was possible: Proportion of the population with no education (E1) and proportion with a tertiary education (E4). The validation procedure was a crucial factor in arriving at the present version of the database (called version 1.0). This paper also presents the results of the validation of the published version against real data, and highlights the need for the harmonization of education data to facilitate comparison over time and space.

Acknowledgments

Fernando Riosmena gratefully acknowledges support from IIASA's Luis Donaldo Colosio Fellowship during the completion of this work.

About the Authors

Fernando Riosmena is Assistant Professor of Geography within the Population Program at the Institute of Behavioral Science, University of Colorado at Boulder. He is an alumnus of the World Population Program at IIASA, where he spent one year as a Luis Donaldo Colosio Fellow.

Isolde Prommer is a Guest Research Scholar with the Health and Global Change Program at IIASA.

Anne Goujon is a Research Scholar at the Vienna Institute of Demography of the Austrian Academy of Sciences, and a Research Scholar with the World Population Program at IIASA.

Samir K.C. is a Research Scholar with the World Population Program at IIASA.

An Evaluation of the IIASA/VID Education-Specific Back Projections

Fernando Riosmena, Isolde Prommer, Anne Goujon, and Samir KC

Introduction

IIASA and the Vienna Institute of Demography of the Austrian Academy of Sciences (VID) jointly produced education-specific back projections (IVEP) in 2007. The results of this exercise to reconstruct levels of educational attainment by age and sex for the period 1970-2000 for 120 countries were published in Lutz et al. (2007). The reconstruction was achieved in two main steps that produced two different versions of the database: The first, unpublished version is called IVEP-Beta and the published version is called IVEP version 1.0 (or IVEP-1.0). In this paper, we evaluate the IVEP-Beta version resulting from the uniform application of the reconstruction rules as described in Lutz et al. (2007).¹ By following the procedures described below, we compare these first crude results country-by-country and year-by-year to the (scattered) existing empirical data. Based on this analysis, we exclude and modify those cases deemed most problematic. Besides describing this procedure, we also present the results of our analysis on the published dataset IVEP-1.0. This revised dataset is now being used for all empirical applications of the data to this date and is publicly available on IIASA's website (see Footnote 1). Unless otherwise specified, the acronym IVEP refers to the published database as described in Lutz et al. (2007).

As the very purpose of estimating back projections is to overcome the unavailability of detailed (i.e., sex- and age-specific) data on education that is also comparable across countries and periods, an evaluation of this sort must draw from scattered heterogeneous indicators available from a few sources. Official educational attainment information of this sort has been compiled from censuses and national surveys by UNESCO's Institute for Statistics (UIS, see www.uis.unesco.org). The latest public release of the UIS database (dated December 2002)² includes the distribution of people ages x and over (where x is generally 25 years) in six educational categories (described in the next section). In addition, an additional series of similar distributions for people ages 15 and over is available from a series of Demographic and Health Surveys (DHS), fielded by Macro International (see <http://www.measuredhs.com/>).

We use these data to the best extent possible to identify their main differences with IVEP back projection in the following manner. In Section 2, we explain the procedure used to match the back projection estimates for a given country-year with that available in UIS or DHS, and further assess the coverage of these matches vis-à-vis the space-time covered by

¹ The IVEP full dataset on the population by levels of educational attainment by age (in five-year age groups) and sex for the period 1970-2000 in five-year steps is available at: <http://www.iiasa.ac.at/Research/POP/edu07/index.html> (last accessed on May 19, 2008).

² These data are available at http://www.uis.unesco.org/ev.php?ID=5234_201&ID2=DO_TOPIC (last accessed on June 3, 2006). Other UNESCO data are available in Statistical Yearbooks.

IVEP. Since measures do not have the exact same specificity with respect to age and education categorization, we describe the rationale for determining which measures we deemed more comparable in Section 3. In Section 4, we analyze measures of centrality and spread for the differences between IVEP and UIS/DHS estimates for those country-years and categories where the data were most complete and comparable for the Beta version of the IVEP database. We then look at problematic cases in more detail while attempting to evaluate if these deviations mostly exist due to potential comparability issues between IVEP and UIS/DHS sources, or if they possibly further arose from biases brought by the back projections' assumptions regarding differential mortality and migration by level of education. We further explain some adjustments made to the IVEP database and present descriptive statistics for this version (as indicated above, denoted in version 1.0). In Section 5, we wrap up our discussion of the general validity of IVEP data in light of our results.

Coverage of the IIASA/VID Education Database

The IVEP-1.0 database includes back-projection estimates of the age-, sex- and education-specific population distributions of 120 countries for every five-year period between 1970 and 1995, yielding $120 \cdot 6 = 720$ country-periods (in addition, of course, to the 120 baseline estimates for 2000). Country coverage in IVEP is equivalent 63 percent of the current UN membership (standing at 192) and overall resembles the regional distribution of UN countries (see Table 1). More importantly, at least 40 percent (and as much as 80 percent) of the countries currently forming a given region are represented in the IVEP database. Only North America, which has two countries representing 40 percent of those officially listed in the region by the UN, does not have representation above 50 percent, though the two countries represented – the USA and Canada – account for 91 percent of the region's population. In fact, as shown in the rightmost column of Table 1, the representation of the IVEP database is better in terms of population covered than it is in terms of the number of countries included. Africa is the only region in which population-weighted representation stands at 48 percent while country-wide coverage is 57 percent of the UN countries. All in all, the number of countries and population covered by IVEP thus seem representative of each region and of the world.

Table 1. Distribution of countries covered by IVEP-1.0 versus UN membership by region.

Region	UN membership		IVEP database			Percent of the region's population covered
	No. (<i>N</i>)	Percent	No. (<i>n</i>)	Percent	Percent coverage (<i>n/N</i>)	
Africa	54	28.3	31	25.6	57.4	48.1
East Asia	29	15.2	15	12.4	51.7	77.9
Europe	44	23.0	35	28.9	79.5	90.2
Latin America	33	17.3	22	18.2	66.7	69.7
North America	5	2.6	2	1.7	40.0	91.4
West Asia	26	13.6	16	13.2	61.5	65.4
World	191	100.0	121	100.0	63.4	72.4

As UIS data do not necessarily come from a year ending in a multiple of five, we matched each country-period of available UIS data between 1964 and 2004 only once with the closest back projection period. The matching algorithm allowed a given benchmark estimate to be matched to an IVEP country-period estimate as long as its year of reference was two or fewer years away from the back projection period, except for 1970 and 2000, where we allowed estimates as early as 1964 and as late as 2004 to be correspondingly matched. For instance, if a given UIS estimate for Costa Rica was dated 1973 it would be matched to the 1975 Costa Rican IVEP estimate (and not to that of 1970).

As said before, data were (unevenly) scattered across the space-time under study and, in some cases, not reliable for these comparisons given data quality or data comparability issues (to be explained in more detailed in Section 5). In total, and *after* eliminating the less-comparable records, we were able to match IVEP estimates to at least one benchmark for 224 country-periods, or 31 percent of the cases.³ Section A in Table 2 shows the distribution of matches by decade and region.⁴ From the average of 31 percent for all regions and periods, coverage ranges from 16 percent for African countries in the 1970s to 75 percent of North American countries represented in IVEP in the 1980s. As is clear from Table 2, DHS data serves the main purpose of improving the coverage of benchmark data for the 1990s in Africa, Latin America, and West Asia.

Section B in Table 2 shows the number of countries with *at least* one match. We were able to match an IVEP estimate to either UIS or DHS data in one or more periods in 103 countries, representing 85 percent of the countries in the IVEP database, and thus 54 percent of the total UN membership.

³ In a few instances, we could match more than one source or estimate from a given source with the same period (e.g., we could match a country-period estimate with both UIS and DHS data, or with two different UIS or DHS estimates separated by less than three years). We performed these tests but only count them once regarding the statistics on the number of matches presented above.

⁴ Due to the fact that population censuses are normally carried out closer to the beginning of a decade, there is considerable heaping in the matches towards the years ending in zero. We thus present information for decades.

Table 2. Number of IVEP-1.0 estimates matched to UIS/DHS by region and decade.

A. Countries with matches by region and period					
Period	Region	Number of matches	Percent of IVEP cases in region-period	Number of matches with UIS	Number of matches with DHS
1970s	Africa	10	16.1	10	0
	East Asia	12	40.0	12	0
	Europe	12	17.1	12	0
	Latin America	19	43.2	19	0
	North America	1	25.0	1	0
	West Asia	8	25.0	8	0
	All regions	62	25.6	62	0
1980s	Africa	10	16.1	10	0
	East Asia	8	26.7	8	0
	Europe	15	21.4	15	0
	Latin America	15	34.1	15	0
	North America	3	75.0	3	0
	West Asia	6	18.8	6	0
	All regions	57	23.6	57	0
1990s	Africa	33	53.2	10	26
	East Asia	14	46.7	13	2
	Europe	23	32.9	22	1
	Latin America	20	45.5	11	12
	North America	3	75.0	3	0
	West Asia	12	37.5	4	9
	All regions	105	43.4	63	50
1970-1995	Africa	53	28.5	30	26
	East Asia	34	37.8	33	2
	Europe	50	23.8	49	1
	Latin America	54	40.9	45	12
	North America	7	58.3	7	0
	West Asia	26	27.1	18	9
	All regions	224	30.9	182	50

B. Countries with at least one match by region					
Period	Region	Number of countries with at least one match	Percent of IVEP countries	Number of countries with at least one UIS match	Number of countries with at least one DHS match
1970-1995	Africa	31	100.0	15	17
	East Asia	14	93.3	13	1
	Europe	21	60.0	20	1
	Latin America	21	95.5	20	1
	North America	2	100.0	2	0
	West Asia	14	87.5	10	4
	All regions	103	85.1	80	24

Particularities of Each Database and Standardization Procedure

The IVEP database includes the education-specific population estimates by sex and five-year age groups, 15-60 years old, as well as an open-ended 65+ interval.⁵ The four education categories used in IVEP are: No education (E1); incomplete and completed primary plus incomplete lower secondary (E2); completed lower secondary, incomplete and completed higher secondary, and incomplete tertiary (E3); and completed tertiary (E4). The categories mainly correspond to the categorization existent in the most recent data provided by UNESCO (which differs from the 2002 database used for comparisons) and used for baseline estimates. A smaller but significant portion of IVEP baseline estimates are based on data from population censuses, or nationally-representative surveys (see Lutz et al. 2007). In addition to any attempt on our part to standardize educational categories (described below), IVEP population figures were simply aggregated into the appropriate reference age group and then converted to proportions by each of the aforementioned education groups.

Most UIS data were available for *six* education groups in one single open-ended age group. The education categories are: No schooling; uncompleted primary; completed primary; entered lower secondary; entered higher secondary; and post-secondary. There are two main differences between UIS categories and those from IVEP. The first refers to the treatment of those with *incomplete* lower secondary, who are combined with those having some primary in IVEP but allegedly indistinguishable from those with completed lower secondary in UIS. The second relates to those with *incomplete* tertiary education, who are included in the same category as those with higher secondary and completed lower secondary education in IVEP but not separable from those with complete tertiary in the UIS database.

Table 3 shows a comparison between each of the definitions in the different databases, and their closer (if not perfect) equivalence to the IVEP database. As we found no sensible way to further separate UIS/DHS figures to make them comparable to IVEP or the other way around for IVEP categories E2 and E3, we only compared the lowest and highest education groups in order to minimize the difference in category definitions. We did so as there is no fundamental difference in the definition of the “no education” group in the two databases. Thus, a direct comparison with UIS and DHS estimates seemed plausible. In addition, as the percent of people with completed tertiary should be naturally lower than those with any level of tertiary education, we should expect the figure for IVEP-E4 to be *lower* than those with “post-secondary” education in UIS and DHS in the absence of any systematic biases in the IVEP or UIS/DHS estimates.⁶

As UIS estimates refer to one broad age group (most commonly 25+ and more specifically in just over 80 percent of the *valid* cases as just described), we aggregated age-specific IVEP figures into one open-ended interval where the starting age matched that of the corresponding UIS estimate. This was not a problem as most common groups next to the 25+ were 15+ (4.7 percent) and 20+ (3.8 percent) and 30+. As mentioned before, all DHS estimates refer to the 15+ age group. Thus, we aggregated IVEP figures to refer to the same group whenever there was a match with a DHS estimate.

⁵ Whenever possible, age groups above 65+ were further broken down, but for most countries the open-ended interval starts at 65.

⁶ The only fundamental difference arises from the definition of post-secondary used by UIS. We assume this category does *not* include any technical or vocational education that only requires lower secondary (as opposed to higher secondary) studies. In other words, we assume the post-secondary education category includes people who *entered* into ISCED levels 4 and over.

Table 3. Definition of education categories and equivalence between IVEP and UIS/DHS.

Category / Data source	IVEP	UIS	DHS	Closer equivalence		
No education	E1	A	1	E1	A	1
Some primary	E2	B	2	E2	B+C	2
Completed primary		C				
Some lower secondary	E3	D	3	E3	D+E	3
Completed lower secondary		E				
Some higher secondary		F				
Completed higher secondary	E4	F	4	E4	F	4
Incomplete post-secondary						
Complete post-secondary						

In addition to the category standardization, we performed some adjustments and checks to the UIS database. On occasion, UIS reported two or more categories lumped together as it was not possible to separate them in the original source. Whenever these sums were calculated across the no education, low-high, or tertiary categories (e.g., primary and lower secondary; no education with incomplete primary; or higher secondary and post-secondary, respectively) we eliminated the *whole* record from the originally-matched database. In addition, we eliminated country-periods of the analysis as the data from UIS was flagged in ways that would have significantly affected comparisons. For instance, problematic cases were typically those where a) unknowns were added to the no education category; b) counts excluded people with no education; c) percentages in the original data summed to significantly less than 100 percent without any apparent reason recorded.

Comparisons

Table 4 shows descriptive statistics for the difference between the proportion of people in each comparable education category and that of UIS and DHS. Figures are shown for all matched *back projections* (i.e., excluding any matches with IVEP baseline estimates for 2000). For the most part, the centrality of these differences is close to zero (as one would want for these estimates) and have a relatively reasonable spread around them. Centrality measures became closer to zero between the Beta and 1.0 versions, while changes in spread across versions varied according to the types of outliers being dealt with.

Overall, the mean and median differences between UIS/DHS and back projections are as close to zero as one would want for these estimates. For instance, the median for the contrasts between the IVEP E1 category and the UIS back projections are -0.60 for the Beta version and lie exactly at zero for version 1.0. These values are in the same range but negative for contrasts with DHS: The median difference between the Beta version and the available DHS estimates is -0.54, while the corresponding figure for version 1.0 is slightly smaller at -0.34 (see Table 4, sections A.1. and A.2.). The means of these distributions are slightly larger. For UIS, they are -0.52 and 0.95 for the Beta and 1.0 versions, respectively, while these figures are -3.42 and -1.57 for DHS data.

Table 4. Differences between the IVEP E1 and E4 categories and UIS/DHS for all matched periods, 1970-1995.

	10th	25th	Median	75th	90th	Mean	Std. Dev.	n
A. No education category								
A.1. Comparisons with Beta version								
E1-UIS	-11.37	-4.27	-0.60	2.87	11.95	-0.52	14.11	188
E1-DHS	-14.86	-3.94	-0.54	1.00	4.18	-3.42	10.38	49
A.2. Comparisons with version 1.0								
E1-UIS	-7.22	-2.90	0.00	3.73	9.46	0.95	15.16	196
E1-DHS	-7.64	-3.74	-0.34	1.61	4.52	-1.57	5.81	51
B. Tertiary education category								
B.1. Comparisons with Beta version								
E4-UIS	-2.53	-0.11	1.11	3.18	6.05	1.57	4.71	188
E4-DHS	-7.28	-3.20	-0.14	0.49	3.68	-1.34	5.29	49
B.2. Comparisons with version 1.0								
E4-UIS	-5.20	-1.59	0.18	1.44	2.95	-0.35	6.57	196
E4-DHS	-7.83	-4.48	-0.71	-0.08	0.62	-2.67	4.88	51

Contrasts with the E4 group yield slightly larger differences of magnitude than the E1 group, which tends to be significantly larger (hence, the relative differences in the E4 category are larger). The median difference between IVEP-Beta and UIS is 1.11, while it is only -0.14 for DHS. After adjustments were made to the IVEP estimation procedure which yielded IVEP-1.0, the median difference for UIS was reduced substantially to 0.81, while differences for DHS increased slightly to -0.71 (see Table 4, section B.2.). The means of contrast with the E4 category yielded a slightly larger range than the median. The mean difference between IVEP-BETA and UIS is 1.57 and -0.35 when comparing with IVEP-1.0. The corresponding contrasts with DHS are -1.34 and -2.67.

In addition to the satisfactory centrality of these distributions around zero, there is of course the issue of their spread. The inter-quartile range (i.e., 75th percentile – 25th percentile) of the distribution of differences is below seven points for all estimates, while the middle 80 percent of the distribution (that is, the difference between the 90th and 10th percentiles) is mostly below 20 points. All measures of spread examined (i.e., the inter-quartile range, the middle 80 percent of the distribution, and the standard deviation) are larger for the group with the largest proportions (i.e., E1). The standard deviation of UIS contrasts increased slightly between the Beta and 1.0 versions (14.11 versus 15.16 for E1 and 4.71 versus 6.57 for E4), while they decreased for contrasts with DHS (10.38 versus 5.81 for E1 and 5.29 versus 4.88 for E4). In the case of UIS, this was a result of being able to deal more effectively with extreme negative than with extreme positive outliers (see next section). At any rate, the extremes of the distribution (e.g., the 10th and 90th percentiles) decreased for both UIS and DHS, decreasing between the Beta and 1.0 versions for the most part.

In summary, the “corrections” made to outliers generally (though not always) resulted in smaller differences between the IVEP and UIS/DHS. In particular, the maximum differences of the 5-percentage tails increased for the back projection for both educational categories for the UIS modifications. However, the modifications done for the outliers classified in the Beta version compared to the DHS data led to large decreases in the

differences between the Beta and 1.0 versions, especially for the negative differences of the E1 and E4 baseline and back projections. But overall, central tendency and spread statistics decreased after the data were modified based on additional empirical data. Nevertheless, expert judgment was necessary to decide which data are more reliable and trustworthy, although they might not decrease the calculated versus the empirical data of comparisons (see next section for more discussion).

While the magnitude of difference between IVEP and UIS/DHS is relatively reasonable, it is also worth investigating if there are any systematic deviations between them through time that could suggest that the retro-projections are inducing some sort of bias. Figures 1 and 2 show box-plots for the differences in IVEP E1 and E4, respectively (Figures 1a and 1b show the contrasts for the Beta and 1.0 versions separately for the E1 group, while Figures 2a and 2b present similar estimates for the E4 group). Figures 1a and 1b confirm that the differences in the E1 category are close to zero, yield a relatively narrow inter-quartile range for both IVEP versions, and generate fewer outliers in version 1.0 (see discussion of outliers in the next section). Moreover, these Figures suggest that there is no observable time trend in the differences. Figures 2a and 2b show a slightly different picture. While the centrality of differences is close to zero and their spread is relatively narrow (although less than that of the E1 category), there seems to be a time trend in the differences with the Beta version (see Figure 2a). However, this time trend is much less pronounced when looking at the differences with IVEP-1.0 (see Figure 2b).

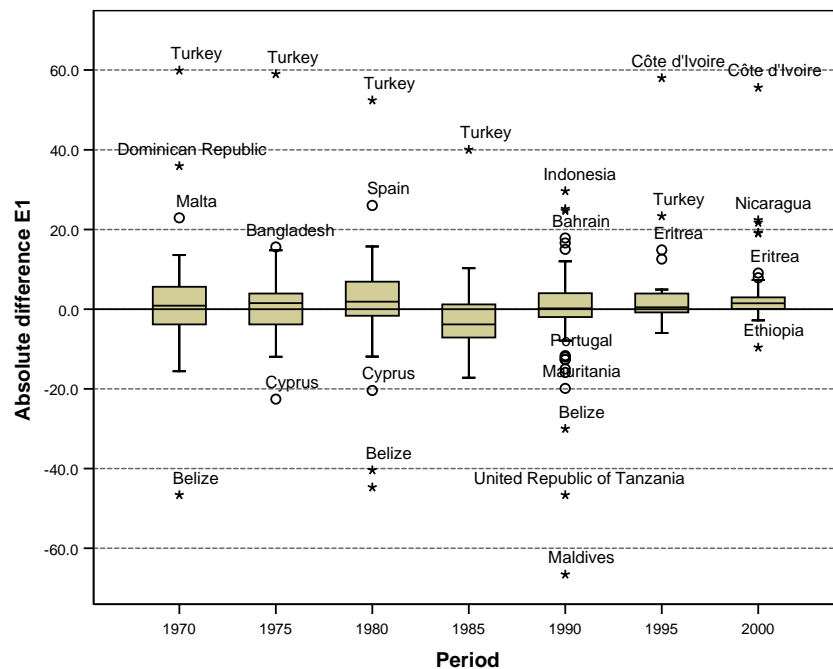


Figure 1a. No education E1 – IVEP-Beta: Absolute differences between IVEP back projections and empirical data. Sources: Lutz et al. (2007); UIS and DHS.

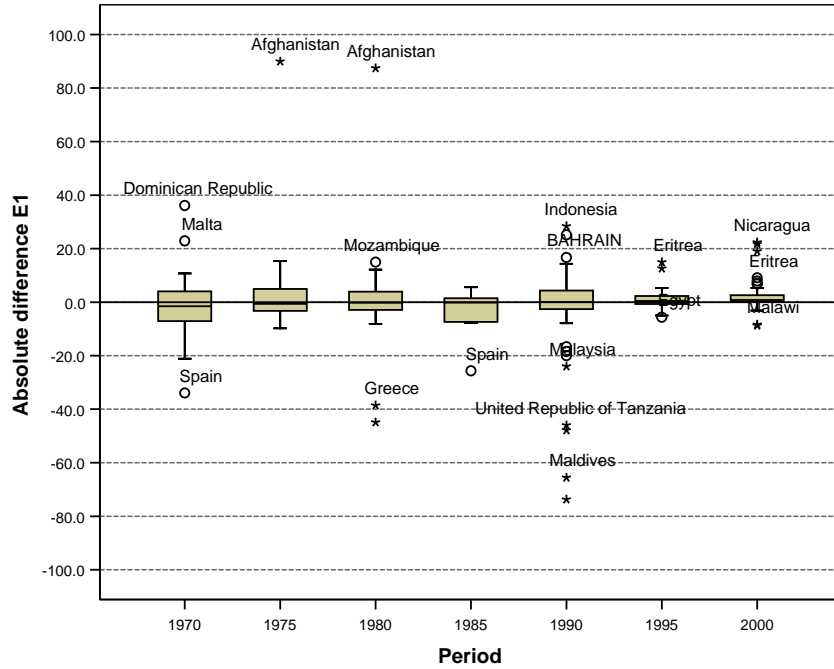


Figure 1b. No education E1 – IVEP-1.0: Absolute differences between IVEP back projections and empirical data. Sources: Lutz et al. (2007); UIS and DHS.

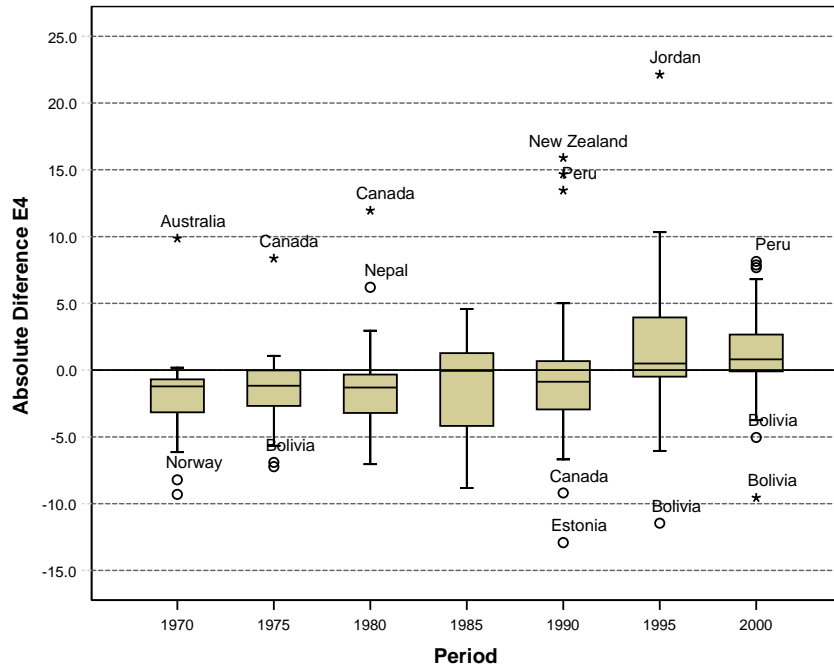


Figure 2a. Tertiary education E4 – IVEP-Beta: Absolute differences between IVEP back projections and empirical data. Sources: Lutz et al. (2007); UIS and DHS.

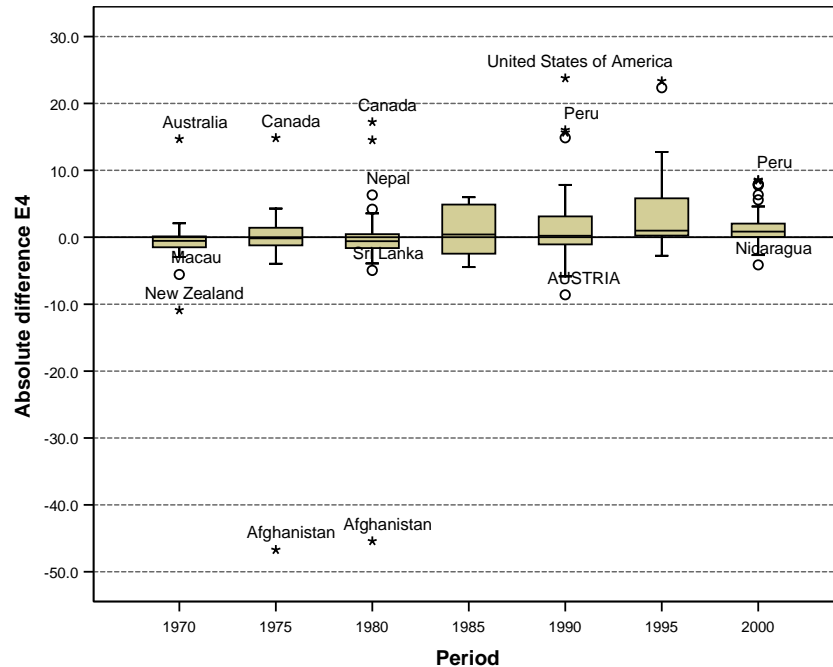


Figure 2b. Tertiary education E4 – IVEP-1.0: Absolute differences between IVEP back projections and empirical data. Sources: Lutz et al. (2007); UIS and DHS.

Adjustments Between IVEP-Beta and IVEP-1.0

Appendix A summarizes the results of our analyses of differences between IVEP and the existing historical data that were considered outliers. In general, we followed the rule that an IVEP-UIS or IVEP-DHS difference larger than 5 percentage points and larger than 20 percent of the actual IVEP value was considered an outlier (all outliers for the E1 and E4 groups are shown in Appendices B and C). Appendix A shows the adjustments that were implemented between the Beta and 1.0 versions in response to the validation exercise. The table also provides more details on the source and nature of the data. After identifying the outliers in the comparison between the existing data and the IVEP reconstruction data in the lowest (no education E1) and highest (tertiary education E4) educational categories, we performed an in-depth analysis for all the outliers to determine the source of the discrepancy. We found the differences to have several origins:

- Definition of the education categories: As mentioned in Section 1, the comparison was done on the no education and tertiary education categories, which are less subject to differences in the definition. But even there, we found that the definition could vary quite a bit, especially for the tertiary level in terms of completion and level at completion. This probably explains most of the differences among cohorts across time.
- Mortality/migration education differentials that deviate from our assumptions (see Lutz et al. 2007): In Israel, the share of highly educated people in the migration flows makes it difficult to reconstruct along cohort lines. Therefore, we decided to remove that country from the sample. For other countries, in the absence of clear information on the education composition of the migrants, we decided not to take this parameter into account, as the work necessary to correct and adjust the data would have been too daunting.

- Inconsistency in the starting IVEP data: It might happen that the dataset chosen to be the base for our reconstruction would not correctly picture the educational share of the population. In some cases and when available, we changed the dataset when we found that the reconstructed data was too far from the existing historical data.
- Inconsistency in the historical data: This is probably the main source of differences between the reconstruction and the historical data. This is especially visible when comparing similar cohorts across time. For instance, in many cases, we found that the educational share of the population was based on a smaller sample than the total population with large missing groups.
- Error or inadequacy in the extrapolation procedure in closing the open-ended interval: Reconstruction requires the repartitioning by educational category of elderly cohorts, which in most cases is not available. As shown in Lutz et al. (2007), this is done by extrapolating the five empirical age groups before the age group with the constraint is given empirically by the proportions for the entire open-ended age group. In a few cases, we changed the extrapolation procedure to provide a better fit with the existing historical data. In some cases, such as in Costa Rica, Malawi, Malaysia, Paraguay, Russian Federation, and Sri Lanka, it was enough to find a better fit between the historical and the reconstructed data and to remove the country from the list of outliers in version 1.0.⁷

As shown in Appendix A, most of the implemented adjustments included choosing a different dataset or adjusting the dataset with another source. As a simple rule, we preferred census data to sample survey data (LFS, DHS, etc.) and more recent data to older data. As we checked the data with different datasets across time, we paid attention especially when all data series were in conflict with our reconstruction. When only a period or two were classified as outliers, we usually kept the reconstructed data as such. We also used more historical data from older surveys to visually confirm or disprove the reconstruction based mainly on the archives available at Statistics Austria.

When the outlier was found in the years around 1970, we ignored it when the differences with the closest matched estimate for the country (usually in the 1980s) were not sizable. This seemed to be especially true of developing countries and places with under-invested statistical offices, where census quality might not have been optimal.

In many cases, we did not implement any correction of the reconstructed data, giving more credit to the recently measured data than to the historical data, especially when we were not able to explain (by mortality or migration) the differences between cohorts across several years in the historical dataset, for instance, when comparing the proportion with no education at age 20-24 in 1970 and at age 40-44 in 1990.

Conclusions

This paper highlights the need for more detailed validation exercises. This should be done in direct collaboration with UIS, the main provider of data on educational attainment, in order to eliminate differences in category definitions across databases to the best extent possible. As mentioned in the introduction, we tested version 1.0 of the IVEP dataset in addition to an initial Beta version. The next version should try “to resolve all discrepancies so that in the end, a corrected and completed (based on comparison to our reconstruction) UIS historical

⁷ Other countries for which the extrapolation procedure was modified between the Beta and 1.0 versions are: Bahrain, Bangladesh, Chile, Macao, Dominican Republic, Egypt, El Salvador, and Indonesia.

dataset and our further validated reconstruction dataset become identical” (Lutz et al. 2007: 229).

In practice, some discrepancies among historical data and the back projections will probably remain unexplained due to factors such as data quality; biased survey samples; changes in the definition of educational categories over time and across datasets; the error range of the base year information for the back projections by education groups; and the model errors and errors in the assumption making and the calculated information to deal with the educational levels of in- and out-migrants; and the transition rates and fertility and mortality differentials. Despite all these limitations the current dataset by age group, sex, and educational attainment, and the average years of schooling is the most robust, comprehensive and detailed dataset today and seems to be reasonably close to those estimates we deemed more comparable and reliable.

References

- Australian Bureau of Statistics. 2001. *2001 Census of Population and Housing*. Highest level of education by age, by sex, count of persons aged 15 years and over (excluding overseas visitors), based on location on census night. Canberra: Australian Bureau of Statistics. Available online at <http://www.censusdata.abs.gov.au/ABSNavigation/prenav/ProductSelect> [accessed 26/11/2007]
- Bahrain, Ministry of Health. 2002. *National Nutrition Survey for Adult Bahrainis aged 19 Years and Above*. Manama: Ministry of Health, Kingdom of Bahrain.
- Belgium, Institut National de Statistique. 1981. *Recensement de la population et des logements au 1^{er} mars 81, Résultats généraux, population scolaire et niveau d’instruction*. Bruxelles: Institut National de Statistique, Ministère des Affaires économiques, Royaume de Belgique.
- Burkina Faso, Ministère du Plan et de la Coopération et Ministère de l’Intérieur et de la Sécurité. 1995. *Recensement général de la population et de l’habitat d’avril 1993*. Ouagadougou: Ministère du Plan et de la Coopération et Ministère de l’Intérieur et de la Sécurité.
- Cameron, R.J. 1983. *Census of Population and Housing, 30 June 1981*. Cross-classified characteristics of persons and dwellings, Australia. Canberra: Australian Bureau of Statistics.
- Cote d’Ivoire, Direction de la Statistique. 1984. *Population de la Cote d’Ivoire: Analyse des données démographiques disponibles*. Abidjan: Direction de la Statistique, Ministère de l’économie et des Finances, République de Cote d’Ivoire.
- Denmark, Danmarks Statistik. 1981. *Folke-og bolitaellingen 1. januar 1981*. Copenhagen: Danmarks Statistik.
- Ethiopia, Office of the Population and Housing Census. 1991. *The 1984 Population and Housing Census of Ethiopia. Analytical report at national level*. Addis Ababa: Office of the Population and Housing Census, Transitional Government of Ethiopia.
- Indonesia, BPS. 2001. *Statistical Yearbook of Indonesia 2001*. Jakarta: Badan Pusat Statistik.

- Lutz, W., A. Goujon, S. KC, and W. Sanderson. 2007. Reconstruction of populations by age, sex and level of educational attainment for 120 countries for 1970-2000. *Vienna Yearbook of Population Research 2007*, pp. 193-235.
- Mauritania, Direction de la Statistique. 1968. *Annuaire Statistique 1968*. Nouakchott: Direction de la Statistique, Ministère de la Planification et du Développement Rural, République Islamique de Mauritanie.
- Statistical Centre of Iran. 1996. *National Census of Population and Housing 1996. Selected tables (national results)*. Teheran: Statistical Centre of Iran, Plan and Budget Organization, Islamic Republic of Iran.
- Statistical Office of Estonia, Central Statistical Bureau of Latvia, Statistics Lithuania. 2003. *2000 Round of Population and Housing Censuses in Estonia, Latvia, and Lithuania*. Vilnius: Statistics Lithuania.

Appendix A. List of countries, origin of data, adjustments made and comments relating to the back projections

Country	Data source Year of origin <u>Beta version</u>	Data source Year of origin <u>IVEP-1.0</u>	Checked against ⁸	Outlier <u>Beta</u> version	Outlier <u>IVEP-</u> <u>1.0</u>	Comment/Data Adjustment/Outlier in <u>Beta version</u>	Outlier in <u>IVEP-1.0</u>
Argentina	UIS 2001	Ibid.	1991				
			1970				
Armenia	UIS 2001	Ibid.				- No data for comparison	
Australia	LFS ⁹ 2000	Ibid.	2001 ¹⁰			- LFS data only includes data for E3 and E4. E1 and E2 shares for age groups 35-64 were inferred from UIS data for 1971. - LFS data available on share by education categories only for age groups from 15 to 64 years of age. The share for 65+ was taken from New Zealand.	<u>Keep reconstruction</u> Categories do not match
			1981 ¹¹				
			1971	E4	E1 E4		
Austria	Census 2001 ¹²	Eurostat ¹³ 2001	1971			- E3 is the first education category and includes all those who have not completed compulsory education (Hauptschule) - Available Census data from Statistics Austria for 1971, 1981,	<u>Keep reconstruction</u> ISCDED 5/6 corresponded until 2005 to the achievement of a Master degree (Magister) explaining
			1981				
			1991	E4	E4		

⁸ When not indicated otherwise, the data on educational attainment originates from the UNESCO statistical yearbooks (different years) or from the UNESCO historical database file.

⁹ All Labor Force Survey (LFS) data are accessible at http://epp.eurostat.ec.europa.eu/portal/page?_pageid=0,1136184,0_45572595&_dad=portal&_schema=PORTAL

¹⁰ Source: Australian Bureau of Statistics (2001)

¹¹ Source: Cameron (1983)

¹² Available through ISIS database at Statistics Austria: <http://www.statistik.at/index.shtml>

¹³ All Eurostat data are accessible at http://epp.eurostat.ec.europa.eu/portal/page?_pageid=0,1136184,0_45572595&_dad=portal&_schema=PORTAL

Country	Data source Year of origin <u>Beta version</u>	Data source Year of origin <u>IVEP-1.0</u>	Checked against ⁸	Outlier <u>Beta</u> version	Outlier <u>IVEP-</u> <u>1.0</u>	Comment/Data Adjustment/Outlier in <u>Beta version</u>	Outlier in <u>IVEP-1.0</u>
			2001			1991, 2001 shows a share in ISCED 5/6 levels judged too low; we trusted the Eurostat dataset more.	the very low percentage found in the census data.
Bahamas	Census 2000	Ibid.	1990				
Bahrain	UIS 2001	Ibid.	2002 ¹⁴				<u>Keep reconstruction</u> Only UIS data for 1991 is an outlier, whereas 1981 and 1971 are perfectly in line with the reconstruction.
			1991	E1	E1		
			1981				
			1971				
Bangladesh	DHS 1999/2000	Ibid.	1981			- DHS adjustment factor (see methodology) No census data available from Bangladesh Statistical Services.	<u>Keep reconstruction</u> Only UIS data for 1974 is an outlier, whereas 1981 is in line with the reconstruction.
			1974	E1	E1		
Belgium	LFS 2003	Ibid.	1981 ¹⁵			- E2 is the first education category and includes E1.	
			1977				
			1970				
Belize	Census 2000 ¹⁶	Ibid.	1991	E1		- 2000 dataset was corrected between Beta and 1.0 versions due to mistakes in categorization.	<u>Keep reconstruction</u> The outlier is found in years 1970 and 1980 compared to UIS. However the UIS data
			1980	E1	E1		

¹⁴ Source: Bahrain, Ministry of Health (2002)

¹⁵ Source: Belgium, Institut National de Statistique (1981)

¹⁶ Available from the Central Statistical Office

Country	Data source Year of origin <u>Beta version</u>	Data source Year of origin <u>IVEP-1.0</u>	Checked against ⁸	Outlier <u>Beta</u> version	Outlier <u>IVEP-</u> <u>1.0</u>	Comment/Data Adjustment/Outlier in <u>Beta version</u>	Outlier in <u>IVEP-1.0</u>
			1970	E1	E1		for these years seem too low compared to the 1991 data: 11% in 1970, 12% in 1980, and 13% in 1991.
Benin	DHS 2001	Ibid.	1992			- DHS adjustment factor (see methodology)	
			1979				
Bolivia	Census 2000 ¹⁷	Ibid.	1992			- 2000 dataset was corrected between Beta and 1.0 versions due to mistakes in categorization.	
			1976	E4			
Brazil	UIS 2003	Ibid.	1989				
			1980				
			1976				
			1970				
Bulgaria	Census 2001 ¹⁸	Ibid.	1992				<u>Keep reconstruction</u> Only UIS data for 1975 is an outlier, whereas 1985 and 1992 are in line with the reconstruction.
			1985				
			1975	E1 E4	E1		
Burkina Faso	DHS 2003	Ibid.	1993 ¹⁹			- DHS adjustment factor (see methodology)	<u>Keep reconstruction</u> Missing population was judged too high in the 1993 data.
Cambodia	DHS 2000	Ibid.	1993			- DHS adjustment factor (see methodology)	

¹⁷ Available from Instituto Nacional de Estadística

¹⁸ Available from National Statistical Institute of Bulgaria

¹⁹ Source: Burkina Faso, Ministère du Plan et de la Coopération et Ministère de l'Intérieur et de la Sécurité (1995)

Country	Data source Year of origin Beta version	Data source Year of origin IVEP-1.0	Checked against ⁸	Outlier Beta version	Outlier IVEP- 1.0	Comment/Data Adjustment/Outlier in Beta version	Outlier in IVEP-1.0
Cameroon	DHS 2004	Ibid.	1976			- DHS adjustment factor (see methodology)	
Canada	UIS 2001	LFS 2000	1991	E4	E4	- Categories E1 and E2 are adjusted with UIS data for 1981 from Statistical Yearbook 1989 - Source 2000 dataset was changed from UIS to LFS between Beta and 1.0 versions.	<u>Keep reconstruction</u> Years 1976, 1981, and 1991 are outliers compared to UIS. However UIS data for these years seem volatile: 31%, 37%, and 21%, respectively.
			1986	E4			
			1981	E4	E4		
			1976	E4	E4		
			1972				
			1970				
Central African Republic	DHS 1994/95	Ibid.	1988			- DHS adjustment factor (see methodology) - Forward projection to 2000	
			1975				
Chad	DHS 1996/97	Ibid.				- DHS adjustment factor (see methodology)	
Chile	Census 2002	Ibid.	1982				<u>Keep reconstruction</u> Only UIS data for 1970 is an outlier, whereas 1982 data is in line with the reconstruction.
			1970	E1	E1		
China	Microcensus (1%) 2000	Ibid.	1990 ²⁰		E1		<u>Keep reconstruction</u> Although proportion E1 was not consistent between 1982 and 1990 on one hand, and 2000 on the other, we trusted the newest data more.
			1982 ²¹				
China, Hong	UIS 2001	Census 2001	1996			- Data available from Census and	

²⁰ Source: China, micro datasets of 1% sample of the 1982 Census (courtesy of Leiwen Jang)

²¹ Source: China, micro datasets of 1% sample of the 1990 Census (courtesy of Leiwen Jang)

Country	Data source Year of origin Beta version	Data source Year of origin IVEP-1.0	Checked against ⁸	Outlier Beta version	Outlier IVEP- 1.0	Comment/Data Adjustment/Outlier in Beta version	Outlier in IVEP-1.0
Kong SAR			1991			Statistics Department for 2001 for both sexes. Data for 1996 (UNESCO historical database) was used to distribute between male and female. - Source 2000 dataset was changed from UIS to Census 2001 between Beta and 1.0 versions.	
			1986	E1			
			1981	E1			
			1976	E1			
			1971	E1			
China, Macao SAR	UIS 2001	Ibid.	1991				We copied the educational attainment proportion of the 65+ population from China to Macao.
			1970	E1	E1 E4		
Colombia	DHS 2000	Ibid.	1993			- DHS adjustment factor (see methodology)	
			1973				
Comoros	DHS 1996	Ibid.				- DHS adjustment factor (see methodology)	
Costa Rica	Census 2000	Ibid.	1984 ²²			- 65+ disaggregation into more age groups was changed between Beta and 1.0 versions.	
			1973 ¹⁶	E1			
			1968	E1			
Côte d'Ivoire	DHS 1998/99	Ibid.	1988		E1 E4	- DHS adjustment factor (see methodology)	<u>Keep reconstruction</u> Our reconstruction is in line with the 1984 data, whereas the UIS 1988 data seems questionable (0% with no education).
			1984 ²³				

²² Source: Costa Rica, Instituto Nacional de Estadística y Censos, Retadam software for tabulations from censuses 1973 and 1984 available on line at <http://www.inec.go.cr/> [accessed 23/11/2006]

²³ Source: Cote d'Ivoire, Direction de la Statistique (1984)

Country	Data source Year of origin Beta version	Data source Year of origin IVEP-1.0	Checked against ⁸	Outlier Beta version	Outlier IVEP- 1.0	Comment/Data Adjustment/Outlier in Beta version	Outlier in IVEP-1.0
Croatia	UIS 2001	Census 2001 ²⁴	1991	E4	E1	- Source 2000 dataset was changed from UIS to Census 2001 between Beta and 1.0 versions.	<u>Keep reconstruction</u> E4 is no more an outlier and E1 is more in line with 2001 census data.
			1981	E1 E4	E1		
			1971	E1 E4	E1		
Cuba	UIS 2002	Ibid.	1981				
Cyprus	UIS 2001	Census 2001 ²⁵	1992	E1	E4	Source 2000 dataset was changed from UIS to Census 2001 between Beta and 1.0 versions.	<u>Keep reconstruction</u> 1987, 1991, and 1992 are outliers, whereas other 4 data points are in line with our reconstructed data.
			1991	E1 E4	E4		
			1989	E1 E4			
			1987	E1	E4		
			1984	E1			
			1980	E1 E4			
Czech Republic	Census 2001 ²⁶	Ibid.	1970 ²⁷				
			1980				
			1991				
Denmark	Eurostat 2001	Ibid.	1995			- E2 is the first education category and includes E1.	
			1994				
			1991				
			1981 ²⁸				
Dominican Republic	Census 2002	Ibid.	1970	E1	E1		<u>Keep reconstruction</u> The 1970 UIS data for the 25+ group is too far

²⁴ Available from Croastat

²⁵ Available from CYSTAT

²⁶ Available from Czech Statistical Office

²⁷ Data for years 1970, 1980 and 1991 were made available from Czech Statistical Office through Tomas Sobotka (on staff at the VID)

²⁸ Source: Denmark, Danmarks Statistik (1981)

Country	Data source Year of origin Beta version	Data source Year of origin IVEP-1.0	Checked against ⁸	Outlier Beta version	Outlier IVEP- 1.0	Comment/Data Adjustment/Outlier in Beta version	Outlier in IVEP-1.0
							from the observed 55+ Census data..
Ecuador	UIS 2001	Ibid.	1990				
			1982				
			1974				
Egypt	DHS 2000	Ibid.	1986			- DHS adjustment factor (see methodology)	<u>Keep reconstruction</u> The 1976 UIS data for the 25+ group is too far from the observed 50+ DHS data.
			1976	E1	E1		
El Salvador	UIS 2003	Ibid.	1992	E1	E1		<u>Keep reconstruction</u> 1992 is an outlier, whereas 1980 and 1971 data are in line with our reconstructed 1970 data.
			1980				
			1971				
Eritrea	DHS 2002	Ibid.				- DHS adjustment factor (see methodology)	
Estonia	LFS 2000	Census 2000 ²⁹	2000 ³⁰			Source 2000 dataset was changed from LFS to Census 2000 between Beta and 1.0 versions.	
			1989	E4			
Ethiopia	UIS 2002	Ibid.	1984 ³¹				
Finland	UIS	Statistics 1970- 2000	1992			- 2000 dataset was changed from UIS to Census 2000 between Beta and 1.0 versions. Other data points from statistical office. - In IVEP-1.0, E3 is the first	<u>Keep reconstruction</u> 1990 is an outlier, whereas 1992, 1985 and 1980 data are in line with our reconstructed
			1990	E1	E4		
			1985	E1			

²⁹ Available from Statistics Estonia

³⁰ Source: Statistical Office of Estonia et al. (2003)

³¹ Source: Ethiopia, Office of the Population and Housing Census (1991)

Country	Data source Year of origin <u>Beta version</u>	Data source Year of origin <u>IVEP-1.0</u>	Checked against ⁸	Outlier <u>Beta</u> version	Outlier <u>IVEP-</u> <u>1.0</u>	Comment/Data Adjustment/Outlier in <u>Beta version</u>	Outlier in <u>IVEP-1.0</u>
			1980	E1		education category and includes E1 and E2.	data.
France	LFS 2003 ³²	Ibid.	1990			- E2 is the first education category and includes E1. - Too many unknowns in census data.	
Gabon	DHS 2000	Ibid.	1993			- DHS adjustment factor (see methodology)	
Germany	Census 2000	Ibid.				-E2 is the first education category and includes all those who have not completed compulsory education.	
Ghana	DHS 2003	Ibid.	1970			- DHS adjustment factor (see methodology)	
Greece	Eurostat 2001	Ibid.	1991		E1		- Data from 2001 census were recovered after publication of the reconstruction and will be published in Version 1.2.
			1981		E1		
Guatemala	UIS 2002	Ibid.	1981				
			1973				
Guinea	DHS 1999	Ibid.				- DHS adjustment factor (see methodology)	
Guyana	Census 2002	Ibid.	1980				
			1970				
Haiti	DHS 2000	Ibid.	1986			- DHS adjustment factor (see methodology)	
			1982				
			1971				

³² Available from INSEE <http://www.insee.fr/en/>

Country	Data source Year of origin Beta version	Data source Year of origin IVEP-1.0	Checked against ⁸	Outlier Beta version	Outlier IVEP- 1.0	Comment/Data Adjustment/Outlier in Beta version	Outlier in IVEP-1.0
Honduras	Census 2001	Ibid.	1983				
			1974				
Hungary	Eurostat 2001	Ibid.	1990				
			1980				
			1970				
India	Census 2001	Ibid.	1981	E1	E1		<u>Keep reconstruction</u> 1981 is an outlier, whereas 1971 data are in line with our reconstructed data.
			1971				
Indonesia	DHS 2002/03	Ibid.	2001 ³³			- DHS adjustment factor (see methodology)	<u>Keep reconstruction</u> 1990 is an outlier, whereas 1980, 1976 and 1971 data are in line with our reconstructed data.
			1990	E1	E1		
			1980				
			1976				
			1971				
Iran	Census 1996 ³⁴	Ibid.	1966				
Ireland	Eurostat 2001	Ibid.	1991				
			1981				
			1971				
Israel	UIS	Ibid.	1983	E4		Country removed between Beta and 1.0 versions.	
			1982	E4			
			1972	E4			
Italy	Eurostat 2001	Ibid.	1991				
			1981				
			1971				
Japan	Census 2000	Ibid.	1990			- E3 includes the E2 category	

³³ Source: Based on 2001 National Socio-Economic Survey (Indonesia, BPS 2001)

³⁴ Source: Statistical Centre of Iran (1996)

Country	Data source Year of origin Beta version	Data source Year of origin IVEP-1.0	Checked against ⁸	Outlier Beta version	Outlier IVEP- 1.0	Comment/Data Adjustment/Outlier in Beta version	Outlier in IVEP-1.0
			1980			(primary and junior secondary education are grouped in the census)	
			1970				
Jordan	UIS	Ibid.	1997	E4	E4		<u>Keep reconstruction</u>
			1990	E4	E4		
Kazakhstan	DHS 1999	Ibid.	1989			- DHS adjustment factor (see methodology)	
Kenya	DHS 2003	Ibid.	1979			- DHS adjustment factor (see methodology)	
			1969				
Kyrgyzstan	DHS 1997	Ibid.				- DHS adjustment factor (see methodology)	
Latvia	Eurostat 2000	Ibid.	2000 ³⁵			- E2 is the first education category and includes E1.	
		Ibid.	1989				
Lithuania	UIS 2003	Eurostat 2001	2000 ³⁵			- 2000 dataset was changed from UIS to Eurostat between Beta and 1.0 versions.	<u>Keep reconstruction</u> The 1989 UIS data for the 25+ group is too far from the observed 35+ Eurostat data.
			1989	E1	E1		
Luxembourg	UIS 2003	Ibid.	1991			- E2 is the first education category and includes E1. - No data for comparison.	
Madagascar	DHS 2003/04	Ibid.				- DHS adjustment factor (see methodology)	
Malawi	DHS 2000	Ibid.	1987	E1		- DHS adjustment factor (see methodology) - 65+ disaggregation into more age groups was changed between Beta and 1.0 versions	
			1977				

³⁵ Source: Statistical Office of Estonia et al. (2003)

Country	Data source Year of origin <u>Beta version</u>	Data source Year of origin <u>IVEP-1.0</u>	Checked against ⁸	Outlier <u>Beta</u> version	Outlier <u>IVEP-</u> <u>1.0</u>	Comment/Data Adjustment/Outlier in <u>Beta version</u>	Outlier in <u>IVEP-1.0</u>
Malaysia	UIS 2000	Ibid.	1991	E1	E1 E4		<u>Keep reconstruction</u> Our reconstruction for E1& E4 is in line with the 1970 data, whereas the UIS 1991 data seems questionable for E1 (0.1% with no education).
			1980	E1			
			1970				
Maldives	UIS 2000	Ibid.	1990	E1	E1	- 65+ disaggregation into more age groups was changed between Beta and 1.0 versions	- Data from 2000 census was recovered after publication of the reconstruction and will be published in version 1.2.
Mali	DHS 2001	Ibid.	1976			- DHS adjustment factor (see methodology)	
Malta	LFS 2003	Ibid.	1967			- E2 is the first education category and includes E1.	
Mauritania	DHS 2000/01	Ibid.	1988	E1	E1	- DHS adjustment factor (see methodology)	<u>Keep reconstruction</u> Lack of reliable data
			1968 ³⁶				
Mauritius	UIS 2000	Ibid.	1990			Country added in IVEP-1.0	<u>Keep reconstruction</u> The 1983 and 1972 UIS data for the 25+ group is too far from the observed 40+ and 55+ UIS data.
			1983		E1		
			1972		E1		
Mexico	UIS 2000	Ibid.	1990				
			1980				
			1970				

³⁶ Source: Mauritania, Direction de la Statistique (1968)

Country	Data source Year of origin <u>Beta version</u>	Data source Year of origin <u>IVEP-1.0</u>	Checked against ⁸	Outlier <u>Beta</u> version	Outlier <u>IVEP-</u> <u>1.0</u>	Comment/Data Adjustment/Outlier in <u>Beta version</u>	Outlier in <u>IVEP-1.0</u>
Mongolia	UIS 2000	Ibid.	1969				
Morocco	DHS 2003/04	Ibid.	1971			- DHS adjustment factor (see methodology)	
Mozambique	DHS 2003	Ibid.	1980	E1	E1	- 65+ disaggregation into more age groups was changed between Beta and 1.0 versions	<u>Keep reconstruction</u> The 1980 UIS data for the 25+ group is too far from the observed 40+ DHS data.
Namibia	UIS 2001	Ibid.	1991			- E2 is the first education category and includes E1.	
Nepal	DHS 2001	Ibid.	1991			- DHS adjustment factor (see methodology)	<u>Keep reconstruction</u> 1981 is an outlier, whereas 1971 and 1981 data are in line with our reconstructed data.
			1981	E1 E4	E1 E4	- 65+ disaggregation into more age groups was changed between Beta and 1.0 versions	
			1971				
Netherlands	LFS 2000	Ibid.	1971			- E2 is the first education category and includes E1.	
New Zealand	LFS 2001	Ibid.	1981			- E3 is the first education category and includes E1 and E2.	<u>Keep reconstruction</u> The UIS data is high for E4 and most likely includes post secondary studies that are normally in E3.
			1991	E4	E4		
			1966	E4	E4		
Nicaragua	UIS 2001	Ibid.	1995 ³⁷			- E2 is the first education category and includes E1.	
			1971				
Niger	DHS 1998	Ibid.	1977			- DHS adjustment factor (see methodology)	
Nigeria	DHS 2003	Ibid.				- DHS adjustment factor (see methodology)	

³⁷ Source: Nicaragua, Instituto Nacional de Estadística y Censos. Censos Nacionales de Población y Vivienda Nicaragua 1995, available online at <http://censos.ccp.ucr.ac.cr/cgi-bin/consulta>

Country	Data source Year of origin Beta version	Data source Year of origin IVEP-1.0	Checked against ⁸	Outlier Beta version	Outlier IVEP- 1.0	Comment/Data Adjustment/Outlier in Beta version	Outlier in IVEP-1.0
						methodology)	
Norway	LFS 2000	Eurostat 2001	1994	E4		- 2000 dataset was changed from LFS to Eurostat 2001 between Beta and 1.0 versions.	
			1990				
			1980	E4			
			1975				
			1970	E4			
Pakistan	LFS 2003	Ibid.	1990				
			1981				
Panama	Census 2000 ³⁸	Ibid.	1990 ³⁹			Country added in IVEP-1.0	<u>Keep reconstruction</u> The 1970 and 1980 UIS data for the 25+ group are too far from the observed 55+ and 45+ census data.
			1980		E1		
			1970		E1		
Paraguay	UIS 2000/01	Ibid.	1992			- 65+ disaggregation into more age groups was changed between Beta and 1.0 versions.	
			1982	E1			
			1972	E1			
Peru	UIS 2003	Ibid.	1993	E4	E4		<u>Keep reconstruction</u> 1993 is an outlier, whereas 1972 and 1981 data are in line with our reconstructed data.
			1981				
			1972				
Philippines	UIS 2000	Ibid.	1995		E4		<u>Keep reconstruction</u> 1990 and 1995 are outliers, whereas 1970, 1975 and 1980 data are
			1990		E4		
			1980	E1			
			1975	E1			

³⁸ Source: Dirección de Estadística y Censo. Censos Nacionales de Población y Vivienda Panamá 2000, available online at <http://censos.ccp.ucr.ac.cr/cgi-bin/consulta>

³⁹ Source: Dirección de Estadística y Censo. Censos Nacionales de Población y Vivienda Panamá 1990, available online at <http://censos.ccp.ucr.ac.cr/cgi-bin/consulta>

Country	Data source Year of origin <u>Beta version</u>	Data source Year of origin <u>IVEP-1.0</u>	Checked against ⁸	Outlier <u>Beta</u> version	Outlier <u>IVEP-</u> <u>1.0</u>	Comment/Data Adjustment/Outlier in <u>Beta version</u>	Outlier in <u>IVEP-1.0</u>
			1970	E1			in line with our reconstructed data.
Poland	UIS 2000	Eurostat 2002	1988			- 2000 dataset was changed from UIS to Eurostat 2002 between Beta and 1.0 versions.	<u>Keep reconstruction</u> The 1970 and 1978 UIS data for the 25+ group are too far from the observed 55+ and 45+ census data.
			1978		E1		
			1970	E1	E1		
Portugal	LFS 2000	Census 2001 ⁴⁰	1991	E1	E1	- 2000 dataset was changed from LFS to Census 2001 between Beta and 1.0 versions.	<u>Keep reconstruction</u> 1991 is an outlier, whereas 1970 and 1981 data are in line with our reconstructed data.
			1981	E1			
			1970				
Republic of Korea	UIS 2000	Ibid.	1995			- 65+ disaggregation into more age groups was changed between Beta and 1.0 versions. The last age-group in the original dataset is 50+.	<u>Keep reconstruction</u> 1970 and 1975 are outliers, whereas 1980-1995 data are in line with our reconstructed data.
			1990				
			1985				
			1980		E1		
			1975	E1	E1		
			1970		E1		
Romania	Eurostat 2002	Ibid.	1992				
			1977				
Russian Federation	UIS 2002	Ibid.	1994			- E2 is the first education category and includes E1.	
			1989	E4		- 65+ disaggregation into more age groups was changed between Beta and 1.0 versions.	
Rwanda	DHS 2000	Ibid.	1978			- DHS adjustment factor (see methodology)	
Saudi Arabia	UIS 2000	Ibid.				- No data for comparison	

⁴⁰ Available from National Statistical Institute

Country	Data source Year of origin Beta version	Data source Year of origin IVEP-1.0	Checked against ⁸	Outlier Beta version	Outlier IVEP- 1.0	Comment/Data Adjustment/Outlier in Beta version	Outlier in IVEP-1.0
Singapore	Census 2000	Ibid.	1995			- 65+ disaggregation into more age groups was changed between Beta and 1.0 versions.	<u>Keep reconstruction</u> The 1970 UIS data for the 25+ group is too far from the observed 55+ census data in 2000.
			1990				
			1970		E1		
Slovakia	UIS 2001	Ibid.	1991				
Slovenia	LFS 2003	Ibid.	1991			Adjusted with UIS data for 1991 (Statistical Yearbook 1995) for categories E2 and E3.	
South Africa	Census 2001	Ibid.	1994				<u>Keep reconstruction</u> 1985 is an outlier, whereas 1970 and 1980 data are in line with our reconstructed data.
			1985	E1	E1		
			1980				
			1970				
Spain	UIS 2003	Census 2001	1991	E1 E4		- 2000 dataset was changed from UIS to Census 2001 between Beta and 1.0 versions.	<u>Keep reconstruction</u> 1970 and 1986 are outliers, whereas 1981 and 1991 data are in line with our reconstructed data.
			1986		E1		
			1981	E1			
			1970		E1		
Sri Lanka	UIS 2001	Ibid.	1981	E4		- 65+ disaggregation into more age groups was changed between Beta and 1.0 versions.	
			1971				
Sweden	UIS 2003	Ibid.	1986- 1995 ⁴¹			- E2 is the first education category and includes E1.	
			1979				
			1974				

⁴¹ Data is available for each year in that period

Country	Data source Year of origin <u>Beta version</u>	Data source Year of origin <u>IVEP-1.0</u>	Checked against ⁸	Outlier <u>Beta</u> version	Outlier <u>IVEP-</u> <u>1.0</u>	Comment/Data Adjustment/Outlier in <u>Beta version</u>	Outlier in <u>IVEP-1.0</u>
			1970				
Switzerland	Census 2000 ⁴²	Ibid.	1980			- E2 is the first education category and includes all those who have not completed compulsory education.	
			1970				
Syrian Arab Republic	UIS 2002	Ibid.	1970				
TFYR Macedonia	UIS 2002	Ibid.	1994				
Thailand	Census 2000	Ibid.	1980		E1	- 65+ disaggregation into more age groups was changed between Beta and 1.0 versions.	<u>Keep reconstruction</u> 1980 is an outlier, whereas 1970 data are in line with our reconstructed data.
			1970				
Togo	DHS 1998	Ibid.	1970			- DHS adjustment factor (see methodology)	
			1981				
Turkey	UIS 2003	Census 2000	1993			- 2000 dataset was changed from UIS to Census 2000 between Beta and 1.0 versions.	<u>Keep reconstruction</u> 1989 is an outlier for E1, whereas all 4 other data points are in line with our reconstructed data. Same for E4 in 1985.
			1989	E1	E1		
			1985	E1	E4		
			1980	E1			
			1975	E1			
			1965	E1			
Turkmenistan	UIS 1995	Ibid.					
Uganda	DHS 2000/01	Ibid.	1991			- DHS adjustment factor (see	

⁴² Available from Swiss Federal Statistical Office

Country	Data source Year of origin <u>Beta version</u>	Data source Year of origin <u>IVEP-1.0</u>	Checked against ⁸	Outlier <u>Beta</u> version	Outlier <u>IVEP-</u> <u>1.0</u>	Comment/Data Adjustment/Outlier in <u>Beta version</u>	Outlier in <u>IVEP-1.0</u>
		Ibid.	1969			methodology)	
Ukraine	UIS 2001	Ibid.	1970			E2 is the first education category and includes E1.	
United Kingdom	Eurostat 2001	Ibid.	1976			E2 is the first education category and includes E1.	
			1971				
United Republic of Tanzania	UIS 2002	Ibid.	1988	E1	E1		<u>Keep reconstruction</u> 1988 UIS data for E1 is 0% and not in line with the 2002 UIS data.
			1978				
United States of America	UIS 2000	Census 2000 ⁴³	1994		E4	- 2000 dataset was changed from UIS to Census 2000 between Beta and 1.0 versions.	<u>Keep reconstruction</u> 1980, 1990 and 1994 are outliers, whereas 1970, 1979 and 1981 data are in line with our reconstructed data.
			1990		E4		
			1981				
			1980		E4		
			1979				
			1970				
Uruguay	Survey 2004 ⁴⁴	Ibid.	1996		E4	Country added in IVEP-1.0	<u>Keep reconstruction</u> The 1975 UIS data for the 25+ group is too far from the observed 50+ sample survey data.
			1985				
			1975		E1		
Uzbekistan	DHS 1996	Ibid.				- DHS adjustment factor (see methodology)	
Vietnam	Census 1999	Ibid.	1989				
			1979				

⁴³ Available from US Census Bureau

⁴⁴ 2004 data based on a representative survey of the urban population representing around 80 percent of Uruguay, accessible at <http://www.ine.gub.uy/biblioteca/ech/poburbana04.htm>

Country	Data source Year of origin <u>Beta version</u>	Data source Year of origin <u>IVEP-1.0</u>	Checked against ⁸	Outlier <u>Beta</u> <u>version</u>	Outlier <u>IVEP-</u> <u>1.0</u>	Comment/Data Adjustment/Outlier in <u>Beta version</u>	Outlier in <u>IVEP-1.0</u>
Zambia	DHS 2001/02	Ibid.	1993			- DHS adjustment factor (see methodology)	<u>Keep reconstruction</u> 1990 is an outlier, whereas 1969 and 1980 data are in line with our reconstructed data.
			1990	E1	E1		
			1980				
			1969				
Zimbabwe	DHS 1999	Ibid.	1992			- DHS adjustment factor (see methodology)	

Appendix B. Discrepancies in the No Education category between IVEP-Beta and UIS/DHS

Country	Source IVEP baseline	Source contrast data	Year IVEP estimate	Year contrast	Age group	IVEP estimate	UIS/DHS estimate	Absolute difference	Relative difference
Bahrain	UIS	UIS	1990	1991	25+	20.5	38.4	17.86	87.0
Bangladesh	DHS	UIS	1975	1974	25+	66.7	82.3	15.60	23.4
Belize	Census	UIS	1970	1970	25+	58.8	12.2	-46.62	-79.3
Belize	Census	UIS	1980	1980	25+	51.1	10.7	-40.41	-79.1
Belize	Census	UIS	1990	1991	25+	43.0	13	-30.01	-69.8
Bulgaria	UIS	UIS	1975	1975	25+	3.7	12.9	9.23	251.5
Chile	Census	UIS	1970	1970	25+	19.5	12.4	-7.13	-36.5
Costa Rica	Census	UIS	1970	1968	25+	26.1	20.6	-5.45	-20.9
Costa Rica	Census	UIS	1975	1973	25+	21.2	16.1	-5.12	-24.1
Croatia	UIS	UIS	1970	1971	15+	12.1	17.6	5.47	45.1
Croatia	UIS	UIS	1980	1981	15+	8.0	14.2	6.18	77.1
Cyprus	UIS	UIS	1975	1976	20+	35.6	13	-22.56	-63.4
Cyprus	UIS	UIS	1980	1980	20+	28.4	8	-20.40	-71.8
Cyprus	UIS	UIS	1985	1984	20+	23.2	8	-15.18	-65.5
Cyprus	UIS	UIS	1985	1987	20+	23.2	6	-17.18	-74.1
Cyprus	UIS	UIS	1990	1989	20+	18.6	6	-12.62	-67.8
Cyprus	UIS	UIS	1990	1991	20+	18.6	6	-12.62	-67.8
Cyprus	UIS	UIS	1990	1992	25+	21.0	5.1	-15.89	-75.7
Dominican Republic	Census	UIS	1970	1970	25+	4.2	40.1	35.93	860.8
El Salvador	UIS	UIS	1990	1992	15+	24.1	30.1	6.02	25.0
Finland	UIS	UIS	1980	1980	20+	9.6	1.2	-8.39	-87.5
Finland	UIS	UIS	1985	1985	25+	8.3	0	-8.35	-100.0
Finland	UIS	UIS	1990	1990	25+	6.4	0	-6.42	-100.0
China Hong Kong S.A.R.	UIS	UIS	1970	1971	25+	47.9	33.3	-14.61	-30.5
China Hong Kong S.A.R.	UIS	UIS	1975	1976	25+	40.5	28.5	-11.95	-29.6
Hong Kong	UIS	UIS	1980	1981	25+	31.8	22.5	-9.27	-29.2
China Hong Kong S.A.R.	UIS	UIS	1985	1986	25+	24.7	18.4	-6.32	-25.6
India	Census	UIS	1980	1981	25+	60.2	72.5	12.34	20.5
Indonesia	DHS	UIS	1990	1990	25+	24.9	54.5	29.64	119.2

Country	Source IVEP baseline	Source contrast data	Year IVEP estimate	Year contrast	Age group	IVEP estimate	UIS/DHS estimate	Absolute difference	Relative difference
Republic of Korea	UIS	UIS	1975	1975	25+	32.4	25.2	-7.19	-22.2
Lithuania	UIS	UIS	1990	1989	25+	1.6	9.1	7.53	480.0
Macau	UIS	UIS	1970	1970	25+	42.5	26.9	-15.60	-36.7
Malawi	DHS	UIS	1985	1987	25+	44.7	55	10.29	23.0
Malaysia	UIS	UIS	1980	1980	25+	22.8	36.6	13.78	60.4
Malaysia	UIS	UIS	1990	1991	15+	11.7	0.1	-11.64	-99.1
Malaysia	UIS	UIS	1990	1991	25+	15.1	0.1	-15.00	-99.3
Maldives	UIS	UIS	1990	1990	25+	67.5	0.9	-66.57	-98.7
Malta	LFS	UIS	1970	1967	25+	0.0	22.9	22.90	100.0
Mauritania	DHS	UIS	1990	1988	25+	80.7	60.8	-19.89	-24.6
Mozambique	DHS	UIS	1980	1980	25+	65.3	81	15.73	24.1
Nepal	DHS	UIS	1980	1981	25+	85.9	41.2	-44.71	-52.0
Paraguay	UIS	UIS	1970	1972	25+	9.0	19.6	10.55	116.7
Paraguay	UIS	UIS	1980	1982	25+	8.3	14.1	5.84	70.7
Philippines	UIS	UIS	1970	1970	25+	6.2	19.8	13.58	218.3
Philippines	UIS	UIS	1975	1975	25+	5.6	14.1	8.55	154.0
Philippines	UIS	UIS	1980	1980	25+	4.9	11.7	6.82	139.8
Poland	UIS	UIS	1970	1970	25+	0.0	5.2	5.20	100.0
Portugal	LFS	UIS	1980	1981	25+	39.4	27.5	-11.87	-30.2
Portugal	LFS	UIS	1990	1991	25+	28.1	16.1	-12.01	-42.7
South Africa	Census	UIS	1985	1985	25+	31.9	24.8	-7.11	-22.3
Spain	UIS	UIS	1980	1981	25+	9.1	35.1	26.02	286.8
Spain	UIS	UIS	1990	1991	25+	5.7	30.4	24.67	430.8
Turkey	UIS	UIS	1970	1965	20+	0.0	59.9	59.90	100.0
Turkey	UIS	UIS	1975	1975	25+	0.0	59	59.00	100.0
Turkey	UIS	UIS	1980	1980	25+	0.0	52.4	52.40	100.0
Turkey	UIS	UIS	1985	1985	25+	0.0	40	40.00	100.0
Turkey	UIS	UIS	1990	1989	25+	0.0	16.6	16.60	100.0
Egypt	DHS	UIS	1975	1976	25+	71.5	86.3	14.77	20.7
United Republic of Tanzania	UIS	UIS	1990	1988	25+	46.6	0	-46.64	-100.0
Zambia	DHS	UIS	1990	1990	25+	28.2	40.2	11.99	42.5

Appendix C. Discrepancies in the Tertiary Education category between IVEP-Beta and UIS/DHS

Country	Source IVEP baseline	Source contrast data	Year IVEP estimate	Year contrast	Age group	IVEP estimate	UIS/DHS estimate	Absolute difference	Relative difference
Australia	LFS	UIS	1970	1971	25+	11.6	21.5	9.9	84.7
Austria	LFS	UIS	1990	1991	25+	12.8	6.1	-6.7	-52.3
Bolivia	Census	UIS	1975	1976	20+	12.8	5.9	-6.9	-53.9
Bolivia	Census	UIS	1975	1976	25+	10.7	5.0	-5.7	-53.1
Bulgaria	UIS	UIS	1975	1975	25+	12.7	5.5	-7.2	-56.8
Canada	UIS	UIS	1975	1976	25+	22.5	30.9	8.4	37.1
Canada	UIS	UIS	1980	1981	25+	25.5	37.4	11.9	47.0
Canada	UIS	UIS	1985	1986	25+	28.1	19.3	-8.8	-31.4
Canada	UIS	UIS	1990	1991	25+	30.6	21.4	-9.2	-30.1
Sri Lanka	UIS	UIS	1980	1981	25+	6.7	1.1	-5.6	-83.6
Croatia	UIS	UIS	1970	1971	15+	8.3	2.2	-6.1	-73.6
Croatia	UIS	UIS	1980	1981	15+	10.4	3.6	-6.8	-65.3
Croatia	UIS	UIS	1990	1991	25+	12.4	6.4	-6.0	-48.5
Cyprus	UIS	UIS	1980	1980	20+	15.0	8.0	-7.0	-46.8
Cyprus	UIS	UIS	1990	1989	20+	20.6	14.0	-6.6	-32.0
Cyprus	UIS	UIS	1990	1991	20+	20.6	15.0	-5.6	-27.1
Estonia	LFS	UIS	1990	1989	25+	26.6	13.7	-12.9	-48.5
Finland	UIS	UIS	1990	1990	25+	10.4	15.4	5.0	48.3
Israel	UIS	UIS	1970	1972	25+	31.4	14.8	-16.6	-52.9
Israel	UIS	UIS	1980	1982	25+	36.6	23.1	-13.5	-36.9
Israel	UIS	UIS	1985	1983	15+	33.9	8.9	-25.0	-73.7
Israel	UIS	UIS	1985	1983	25+	37.9	11.2	-26.7	-70.4
Nepal	DHS	UIS	1980	1981	25+	0.6	6.8	6.2	1035.6
New Zealand	LFS	UIS	1970	1966	25+	14.2	4.9	-9.3	-65.5
New Zealand	LFS	UIS	1990	1991	25+	23.2	39.1	15.9	68.5
Norway	LFS	UIS	1970	1970	16+	14.8	6.6	-8.2	-55.5
Norway	LFS	UIS	1980	1980	25+	18.2	11.9	-6.3	-34.7
Norway	LFS	UIS	1995	1994	16+	24.8	18.7	-6.1	-24.5
Peru	UIS	UIS	1995	1993	15+	10.3	20.1	9.8	95.7
Peru	UIS	UIS	1995	1993	20+	11.8	22.1	10.3	87.8

Country	Source IVEP baseline	Source contrast data	Year IVEP estimate	Year contrast	Age group	IVEP estimate	UIS/DHS estimate	Absolute difference	Relative difference
Peru	UIS	UIS	1995	1993	25+	11.0	20.5	9.5	85.9
Russia	UIS	UIS	1990	1989	25+	19.4	14.1	-5.3	-27.3
Spain	UIS	UIS	1990	1991	25+	13.8	8.4	-5.4	-39.0
Jordan	UIS	DHS	1990	1990	15+	0.4	15.0	14.7	3880.3
Jordan	UIS	DHS	1995	1997	15+	0.5	22.7	22.1	4156.3