

Commentary paper

An international perspective on the undercount of young children in the U.S. Census

J. Gregory Robinson

U.S. Census Bureau (retired), Alexandria, VA, USA

E-mail: robin0007@verizon.net

This paper is a very useful contribution to understanding the complex issue of the undercount of children. It should prompt further investigation of the coverage of young children (systematically document the relative high undercounts, begin to explain why the rates are high, and do something about it in census-taking). The net undercount rates of children are high in the U.S. and unlike older ages groups, the measured rates have not been reduced in recent censuses. The paper gives an international perspective – comparing the coverage of children in the U.S. to other countries. Is poor coverage of children an issue elsewhere as well as in the U.S., adding urgency to understanding the whys of the relative high undercount of children? *Yes it is.*

The author asks three questions: (1) Is a net undercount of children under age 5 common? (2) Is the net undercount of young children higher than that of older children? and (3) Is the net undercount of children greater than that of “any other age group”? The paper focuses on the first two and provides solid documentation to answer those. The third question about coverage of children relative to adults for later research is answered for the U.S. but not from an international perspective. Estimates for more age groups and countries are needed for a definitive assessment.

Further research should also mention the impact of measurement error on the estimates of net undercount (apart from the “real” undercounts we often assume the estimates are showing). The different structure of er-

rors across countries could affect the comparisons with the U.S. results (such as the impact of relatively high levels of duplication in the U.S. census that could distort the comparisons with other countries). And some acknowledgement should be given to the methodology used to produce the estimates of each country – is it based on coverage measurement surveys or analytic methods like Demographic Analysis?

The paper clearly builds a case that the undercount of young children (age 0–4) is high relative to older children; research will help separate the measurement error of this from the “real” undercount that needs to be explained. A systematic uncovering of patterns in the coverage rates across countries (e.g., the difference of the 0–4/5–9 gap in certain Asian countries with other countries shown in the paper) will help the investigation.

Most of my comments are discussion in nature – pointing to further research on the important topic of understanding differential undercounts and the accurate measurements of coverage.

Section 2, measuring net undercount – The two methods used in the U.S. (Demographic Analysis and Dual Systems estimates) reveal very different results regarding the net coverage of young children. The methods used for assessing coverage of children in other countries are not documented – later research should indicate the method used to help evaluate the results. Do other countries find different results in cover-

age of the demographic and survey methods? I suspect the survey method is generally used.

Section 2.1, Fig. 2 – Here, historical DA estimates by single years of age could be used to study how the sharp age gradient has shifted along with the infusion of coverage improvements techniques and difficulty of estimating the growing international migration component. The slope in the DA estimates was much less pronounced before 1970, when the Census Bureau began introducing a range of coverage improvement programs (some of which tended to duplicate certain groups, including young adults who may live more than one place during the year).

Also, international migration has become a much larger component in the DA estimates since 1970, and the number of young immigrants may be understated. As a case in point, the 2010 DA estimates for Hispanics exhibit “overcounts” for ages 14–17 (versus large undercounts for young Hispanics) – a somewhat implausible result. The point is that measurement error may be influencing the patterns seen in Figs 1, 2, 3, and 4 (especially the post-1990 results) and this needs to be considered in investigating the coverage issue of children. I know that this is not a methodological paper, but these sources of error should be noted and included in later research.

Section 4.1, Fig. 3 – The sharp drop in net undercount from –1.7 percent in 1990 to a negligible undercount of –0.1 in 2000 and overcount of 0.1 in 2010 brings up a factor not discussed in the paper, that is, the impact of duplicates in the United States census and how it can affect the interpretation of net undercounts – especially for time series or cross-country comparisons. Duplicates were very large in 2000 (at least 5.8 million according to Census Bureau estimates) and even larger in 2010 (at least 10 million). These offset omissions of people of about the same amount, leading to the nets of near zero. Research ideas for another day – to better understand how much of the calculated net undercounts represent actual omissions of children and adults and how much reflects measurement/census process issues.

Section 4.2, Fig. 4 – The trend for ages 18+, leading to overcounts by 2010, could be influenced by duplicates, and perhaps measurement error of international migration. In studying the net undercount results of other countries, comparison to the U.S. estimates will be affected by the differing levels of duplication/measurement errors in those countries. A case study could be comparisons of the “high” net undercounts at age 20–24 in Canada (–9.6 percent in 2011:

Table 3) to the net *overcounts* at 20–24 in the U.S. (+1.3 percent in 2010: Fig. 1). Are measurement errors contributing to this (as opposed to “real” undercount differences)?

Section 4.2, 3rd paragraph (first question not answered by the profession) – Indeed, we need to understand what contributed to the more flat-line trend of relative high undercoverage of young children, which is so different than the reduced NET undercount (now overcounts) of older ages. Here again, a full investigation will consider the effect of the changing pattern of measurement errors. For young children, one still unknown factor is the effect of imputation. In the U.S., are some young children really counted but are misassigned through imputation to an older age? Research at the Census Bureau has been addressing this.

Section 4.2, 2nd question not answered by profession – Yes, we need to understand the making of the diverse trends in coverage of children and adults. What is the confluence of factors, including those that affect real coverage and those that contribute to measurement error?

Section 4.2, last paragraph – Yes, the 3 patterns are demonstrated for the U.S. with the available data on coverage measurement, including the higher net undercount of young children than any other age group.

Section 5, list of three questions – Unlike the U.S., I will note how this question (young children versus adults) is *not* authoritatively answered on an international perspective by the data presented in the tables.

Section 5, discussion of international coverage estimates – To help evaluate the results, some reference to methods used to derive the estimates is needed. How many countries use DA versus survey-based estimation? Any evaluations of the accuracy of the estimates, like in the U.S.? For later research – but I had to mention it here!

Section 5.1, Table 3 – A small number of countries and ages are reported here for adults (only 4 have estimates for ages 20–24 in Table 3, one for 15–29, one for 20–29, and none for 25 or older). Nor are any estimates over age 10 given in Tables 4 or 5. This undermines the 3rd stated question of “is the net undercount rate for young children commonly higher than the net undercount rate for *any other age group*”. The paper does not answer this with the data provided. Grounds for later research.

Section 5.1, Table 5 – The child undercounts in China are indeed very high, especially the 26.2 percent in 2000. This estimate is so extraordinary that it deserves some “explanation”. What may have con-

tributed to such a high net undercount (compared to other years in China)? Were the rates high for other ages as well? Was it a problem with data or methods that elevated the undercount for children only (here, the total undercount rate could help explain)? Instinctively I do not trust this number without some context. And knowing nothing about the methodology for the China estimates (as well as the other high Asian estimates), I wonder if there is some bias for these countries. Just for clarification.

Section 5.1, data on Asian countries in Table 4 – Yes, the estimates for the Asian countries are revealing and there are two patterns in the numbers. For the first 8 countries the 0–4 undercount estimates are high and average 5.9 points higher than for 5–9 – this is a much greater difference than the “smaller” 1.0 difference for the European/North American countries in Table 3. The last three Asian countries (South Korea, Japan, Taiwan) show a pattern distinctly different than the first eight, with the 0–4 estimates being marginally higher than the 5–9 estimates (more similar to the Table 3 patterns). So there are two stories. Why are the differences so much larger for the first eight Asian countries (and the similar pattern for China in Table 5)? Future research can investigate the methodological, census-taking, cultural, and economic influences.

Section 5.3 – As noted earlier, the data are not provided to definitively answer the question about net undercount of young children relative to the rate for “any other age group”. The statement seems to come from the examination of the rates of 6 countries in Table 3 – and usually only age 20–24 is listed for “adults”. No

age data for adults are shown for the Asian countries (only estimates under age 10 are given). So the results of children in this paper are really compared to that of young adults and the question for Section 5.3 could be restated to reflect this. More study is needed to assess the relative coverage of children compared to any other age group (including all adults).

Section 5.3, last paragraph – The fact that the U.S. is an outlier regarding the high net undercount of children compared to young adults may be attributable in part to census taking and measurement differences. For example, the high level of duplication in the last two U.S. censuses may “drive down” the measured *net* undercount of young adults relative to other countries (where duplication is less prevalent?). Grounds for more research!

Section 6, discussion – Yes, the sample is haphazard and limited, but does seem to support the first 2 questions (there is a net undercount of children, and it is commonly highest for young children).

Section 6/7, discussion and summary – The reasons for the net undercount of children clearly deserve joint collaboration. More countries and age groups need to be added and patterns in the estimates identified to foster the investigation. And the impact of coverage measurement method used (survey-based versus demographic analysis-based), census taking effects (such as duplicates) and other factors can be considered in the evaluation of the results.

Good luck in the continued research on this important topic!