

Health Insurance Estimates from the U.S. Census Bureau: Background for a New Historical Series

By

Cheryl Hill Lee and Sharon M. Stern

Poverty and Health Statistics Branch
Housing and Household Economic Statistics Division
U.S. Census Bureau

June 2007

This report is released to inform interested parties of ongoing research and to encourage discussion.

I. Introduction

The Current Population Survey (CPS) has been producing health insurance coverage estimates since 1987 as part of the Annual Social and Economic Supplement (ASEC, previously called the March Supplement)¹. Since then, the supplement has changed several times to accommodate changes to technology of data collection and changes to the health insurance framework. The purpose of this paper is to explain an important adjustment to the health insurance edit process that assigns private coverage to non-policy holders.

This working paper will provide detailed information to allow users to fully understand the implications of the change to the edit process. With this information users should be able to choose whether or not to use the original data as published. Section II provides some background on the CPS ASEC and describes the editing process in more detail. Section III gives specifics about the adjustment, including the variables involved, and describes the effect on the overall estimates. Section IV presents the Census Bureau's resolution. Section V provides some illustrations of how the adjustment may impact trends in private coverage over time.

II. Background

The current CPS ASEC is a survey of approximately 78,000 households, which includes detailed health insurance questions. The survey questions are asked about all people in the household, but usually a single person within the household responds.

¹ The CPS ASEC also asked health insurance questions from 1980 to 1986, but these questions were quite different from the questions asked from 1987 to the present.

Respondents are asked to report health insurance coverage status in the previous calendar year for themselves and all other household members. Using this information the U.S. Census Bureau provides a historical time series at the national level, as well as data that can be used to follow state-level trends (with multi-year averages). The health insurance estimates are one component of the formula used to allocate funds to states under the State Children's Health Insurance Program (SCHIP). The CPS ASEC is a popular source of health insurance coverage data in the United States due to the other critical characteristics available for analysis—employment, poverty, etc.

Initially, the CPS ASEC was conducted using a paper questionnaire and children's coverage was only reported through their parent's coverage status. In March 1994, the paper questionnaire was converted into a computerized questionnaire or a computer-assisted interview (CAI) format, but the survey questions remained the same. In March 1995, new questions were included in the health insurance section of the CPS ASEC, and one of the main differences was that questions about the coverage status for children were asked directly instead of through the coverage status of their parents. This change allowed for more detailed information about children's health insurance coverage to be collected. In March 2000, a verification question was asked of people who responded 'no' to all questions about specific types of health insurance coverage in order to verify whether they were actually uninsured. The inclusion of the verification question resulted in identification of health insurance coverage for more people.

The CPS ASEC health insurance variables go through an edit process where the questionnaire items feed into edited variables that determine a set of recoded health insurance variables. The questionnaire items pertinent to the edit process that assign

employment-based and private health insurance coverage are SHI2, SHI3, SHI4, SHI7, SHI8, and SHI9. Item SHI4 is on the record of an employment-based policyholder and indicates which household members are covered on the policy. Similarly, item SHI9 is on the record of the policyholder of direct purchase plans and records which household members are covered on the policy.² The specific flow of these questions (and their output) is below. Item SHI2 is a household-level question asking whether anyone in the household was covered by a group health or employment-based health insurance plan. On the internal version of the data file, that answer is stored on the household-level data set as a variable named SHI2. If the answer to SHI2 is ‘yes’ then the instrument goes to item SHI3, which asks which person (or persons) is a policyholder on employment-based health insurance plans. The instrument allows the field representative (FR) to enter up to 16 policyholders. On the data file, the answer is stored on the household-level data set as 16 variables named SHI31 to SHI316. These variables contain as values the line number of the policyholder and each household member. For example, item SHI31 is the line number of the first policyholder; item SHI32 is the line number of the second policyholder, and so forth. In item SHI4, the respondent reports who else in the household was covered on each policyholder’s plan. The FR can enter the line numbers of up to 16 household members or enter ‘A’ for ‘All’ or ‘X’ for ‘None.’ On the data file, the answer is stored on the policyholder’s person record as 16 variables labeled SHI41 through SHI416. According to the internal data dictionary, the answer ‘A’ is stored as ‘-4’ and an answer of ‘X’ is stored as ‘-5’.

² See Appendix A for an excerpt from the items booklet that covers the relevant questionnaire items. The full items booklet is available online as part of the Technical Documentation at <<http://www.census.gov/apsd/techdoc/cps/cps-main.html>> and a question flow chart is available online at <<http://www.census.gov/hhes/www/hlthins/hlthinsseq.pdf>>.

Responses to item SHI3 are linked to the edited variable HI and responses to SHI4 are linked to the edited variable, DEPHI. These are ‘yes’ or ‘no’ variables indicating whether the person is a policyholder (HI) or a dependent (DEPHI) on an employment-based policy.³ Similarly, responses to SHI8 are linked to PRIV and responses to item SHI9 are linked to edited variable, DEPRIV. These are also ‘yes’ or ‘no’ variables indicating whether the person is a policyholder (PRIV) or a dependent (DEPRIV) on a direct purchase health insurance policy. The edited variables HI, DEPHI, PRIV, and DEPRIV feed into one recoded private health insurance variable, COV-HI (a ‘yes’ or ‘no’ person-based variable that reports coverage from a private health insurance plan). Therefore, it is imperative that, when SHI41=-4 and SHI91=-4, everyone within the household be assigned coverage. Otherwise, individuals who should be reported as having employment-based or private health insurance coverage may be improperly assigned.

After production of the 2005 estimates, a programmer found a potential problem in the way that the raw employer-provided and direct purchase health insurance data were being used to assign coverage to non-policyholders. Essentially, when the respondent was asked whether anyone else besides the policyholder was covered by the policy, the program would sometimes assign the correct value of coverage to the non-policyholders (if the line numbers were individually listed by the FR), but in some instances individuals would be marked as missing data (if the respondent indicated that everyone was covered by the plan). During the procedure that assigns coverage status to people with missing data, some were assigned coverage and some were left uninsured. Therefore, the estimates of the uninsured were inflated.

³ See Appendix B for an excerpt from the data dictionary describing the items in question.

III. Recording of Coverage Through SHI4 and SHI9

During the Blaise conversion, programmers found that the two instrument items that allow respondents to report that everyone in the household was covered by employer-provided group-health plans (SHI41) and privately-purchased plans (SHI91) were not being used in the edit that assigns employment-based and private health insurance when such coverage should have been assigned, based on a report of ‘all covered.’ Instead, when the instrument variables SHI41 and SHI91 had values of –4 representing that all people in the household were covered, in most cases the other household members had values for employment-based and private health insurance coverage allocated. If the edit had accepted the report of ‘all covered,’ each member of the household would have been reported as having coverage with an allocation flag reflecting the data are ‘as reported.’ Instead, household members were considered as blanks or non-responses. In many instances, these household members ended up with health insurance coverage (whether private or public) and a flag of ‘allocated.’ In some cases, household members had no coverage. Hence, the total number of people uninsured appears to have been overstated since the health insurance section was revised and the new edits written for 1996 health insurance data. This translates into an uninsured rate that was inflated by approximately half of a percentage point.

A correlated problem is that the flag variable (I-HELSE) is backward in the sense that policyholders get a flag value =1 when SHI41=-4, but others within the household with allocated data have flags=0⁴. Within the original 2006 ASEC data file, there are

⁴ The variable HELSE reports the line number of covered household members. The corresponding allocation flag is I-HELSE = 1 if value is imputed or I-HELSE = 0 otherwise.

8,262 cases of people reporting that everyone within the household was covered by the policyholder's employment-based plan (which correspond to SHI41=-4 for those 8,262 persons).

Within the original 2006 ASEC data file, there were 639 cases of people reporting that everyone within the household was covered by the policyholder's direct purchase plan (which correspond to SHI91=-4 for those 639 persons). There were 1,088 sample people (representing 1.8 million weighted people) that were not covered by group health that lived within households where at least one person had an employment-based policy that covered everyone within the household (This inconsistency directly comes from the misallocation of the values from variable SHI41). For direct purchase plans, 131 people lived in households where someone had a private health insurance plan that covered everyone, but they were uninsured after the edit process (this inconsistency directly comes from the misallocation of the values from variable SHI91).

IV. Resolution

The first step in addressing this issue was to revisit the edit procedure and make sure that when the variables SHI41 and SHI91 have values of -4 all household members are assigned coverage. The second step was to make sure that no other estimates would be adversely impacted by the change to the program logic intended to correct private coverage and allocation flags. This was completed within a few months of the discovery of the problem.

The Census Bureau then determined the feasibility of implementing the correction for previous years. The information needed for a correction to items SHI41 and SHI91 is

on all files since the 1997 ASEC, but the time and resources for a full re-run would be immense. The Census Bureau developed and implemented the following plan for revised products:

- (1) Rerun data files for 2005 and 2006 ASEC.
 - a. Release new public use files (completed 4/2/07)
 - b. Release new detailed tables (completed 5/30/07)
 - c. Release a User's Note with tables summarizing the effect of the revision (completed 3/23/07)
- (2) Release extracts allowing advanced users the ability to approximate the correction for 1997 through 2004 ASEC (completed 6/28/07)
- (3) Create a new historical data series that captures the best information from 1999 (ASEC 2000) through the present. Including the following characteristics:
 - a. Estimates from 1999 include the verification question and use 2000-based weights.⁵
 - b. Estimates from 1999 to 2003 approximating the correction, hence private coverage estimates will differ from original, but government coverage will be unchanged.
 - c. Estimates, starting with 2004, will be from the new edit process.

The historical health insurance tables will have an updated series beginning in 1999 that will continue forward to show the effect of the revision. The data for the years 1999-2003 will be an approximated series, but the data from 2004 onward will have revised health insurance data. The approximated series comes from correcting the process that misallocated the cases of private coverage for non-policyholders, but the entire health insurance processing system was not rerun the way the data for 2004 and

⁵ The CPS Annual Social and Economic Supplement (ASEC) Health Insurance Verification Question and its Effect on Estimates of the Uninsured (incomplete footnote reference)

2005 were completely revised. The original series will stop at the 2005 data year and it will continue to be available online, along with the updated series of data.

V. Impact on Historical Series

Appendix C presents a series of charts illustrating the impact of this revision on the private health insurance coverage rates over time. The figures show that the trends in private coverage from 1996 to 2003 look similar for the original data and the approximated correction. These figures show a graphical comparison between the original series (prior to the correction) and the approximate series (after the correction). The approximate series has the correction written into the estimation process to demonstrate how the coverage rates would change if the health edit process were altered to reflect the correction to the problem with non-policy holders not receiving coverage. Each series includes a break in 1999, due to the verification questions that were added in March 2000. Estimates from 1996 to 1999 present a consistent series without the verification questions. The estimates from 1999-2003 present a consistent series that includes the verification questions.

The figures illustrate the trends in coverage over time before and after the revision. They further demonstrate the impact on various groups, based on age, race, Hispanic origin and type of coverage. There is a consistent trend throughout the charts that shows the impact is minimal and the approximate series is always slightly higher in coverage rates compared with the original series. There is not a large difference in coverage for any of the analyzed demographic groups.

Figures 11 and 12 show that approximated values for types of private health insurance coverage for 2004 and 2005 are extremely close to the revised values. This confirms that if data users choose to use the extract file to approximate the values of health insurance coverage they will get estimates that are virtually the same as the revised figures. For instance, the direct purchase coverage rate was 9.2 percent for the approximated and revised values in 2005. The coverage rate for employment-based was 60.2 percent in 2005 for both the approximated and revised values.

Appendix A

Excerpt from the items booklet.

>SHI2< **At any time in 2005, (were you/was anyone in this household) covered by a health insurance plan provided through (their/your) current or former employer or union? (MILITARY HEALTH INSURANCE WILL BE COVERED LATER IN ANOTHER QUESTION.)**

- <1> Yes
- <2> No

===>

NOTE: THIS ITEM DOES NOT APPEAR FOR SINGLE PERSON HOUSEHOLDS

>SHI3@a<

Who in this household were policyholders?	LN NAME	RELATION
	(person 1)	
	(person 2)	
	(person 3)	
	(person 4)	
	(person 5)	
	(person 6)	
	(person 7)	
	(person 8)	
	(person 9)	
PROBE: Anyone else?	(person 10)	
	(person 11)	
ENTER LINE NUMBER <N> No more	(person 12)	
_ _ _ _ _	(person 13)	
_ _ _ _ _	(person 14)	
_ _ _ _ _	(person 15)	
_ _ _ _ _	(person 16)	

NOTE: THIS ITEM DOES NOT APPEAR FOR SINGLE PERSON HOUSEHOLDS

>SHI4@a<

In addition to (you/name), who else in this household was covered by (name's/your) plan?	LN NAME	RELATION
	(person 1)	
	(person 2)	
	(person 3)	
	(person 4)	
PROBE: Anyone else?	(person 5)	
	(person 6)	
ENTER LINE NUMBER <N> No more	(person 7)	
ENTER <A> FOR ALL	(person 8)	
ENTER <X> FOR NONE	(person 9)	
_ _ _ _ _	(person 10)	
_ _ _ _ _	(person 11)	
_ _ _ _ _	(person 12)	
_ _ _ _ _	(person 13)	
_ _ _ _ _	(person 14)	
_ _ _ _ _	(person 15)	
_ _ _ _ _	(person 16)	

>SHI7< **At anytime during 2005, (were you/was anyone in this household) covered by a health insurance plan that (you/they) PURCHASED DIRECTLY FROM AN INSURANCE COMPANY, that is, not related to current or past employment?**

- <1> Yes
- <2> No

NOTE: THIS ITEM DOES NOT APPEAR FOR SINGLE PERSON HOUSEHOLDS
>SHI8@a<

Who in this household were policyholders?	LN NAME	RELATION
	(person 1)	
	(person 2)	
	(person 3)	
	(person 4)	
	(person 5)	
	(person 6)	
	(person 7)	
	(person 8)	
	(person 9)	
	(person 10)	
PROBE: Anyone else?	(person 11)	
	(person 12)	
ENTER LINE NUMBER <N> No more	(person 13)	
— — — — — — — — —	(person 14)	
	(person 15)	
— — — — — — — — —	(person 16)	

NOTE: THIS ITEM DOES NOT APPEAR FOR SINGLE PERSON HOUSEHOLDS
>SHI9@a<

In addition to (you/name), who else in this household was covered by (name's/your) plan?	LN NAME	RELATION
	(person 1)	
	(person 2)	
	(person 3)	
	(person 4)	
PROBE: Anyone else?	(person 5)	
	(person 6)	
ENTER LINE NUMBER <N> No more	(person 7)	
ENTER <A> FOR ALL	(person 8)	
ENTER <X> FOR NONE	(person 9)	
	(person 10)	
	(person 11)	
	(person 12)	
	(person 13)	
	(person 14)	
— — — — — — — — —	(person 15)	
— — — — — — — — —	(person 16)	

Appendix B

Excerpt from internal DLL

New health insurance questions first asked in the outgoing rotations
c: in march 1995. The questions deal with coverage in the previous calendar
year.

c: They apply to all persons of all ages. These fields will be recoded
c: into umcare, umcaid, uchamp, uhiyn, uhiown, uhiemp, uhipaid, uhielse.

c:

SHI4 (16) int 5 (-5:16) .

c: HHES/POP char 2

c: Who in this household was covered by ...'s plan? .

c: -5 = none

c: -4 = all

c: -3 = refused

c: -2 = don't know

c: 0-16 = line number(s) of person(s) covered by this person's

c: employer/union health plan if this person's line

c: number is in hhld field shi3

SHI9 (16) int 5 (-5:16) .

c: HHES/POP char 2

c: Who in this household was covered by ...'s plan? .

c: -5 = none

c: -4 = all

c: -3 = refused

c: -2 = don't know

c: 0-16 = line number(s) of person(s) covered by this person's

c: private health plan if this person's line

c: number is in hhld field shi8

Appendix C

Figure 1: Health Insurance Coverage Rates: 1996-2003

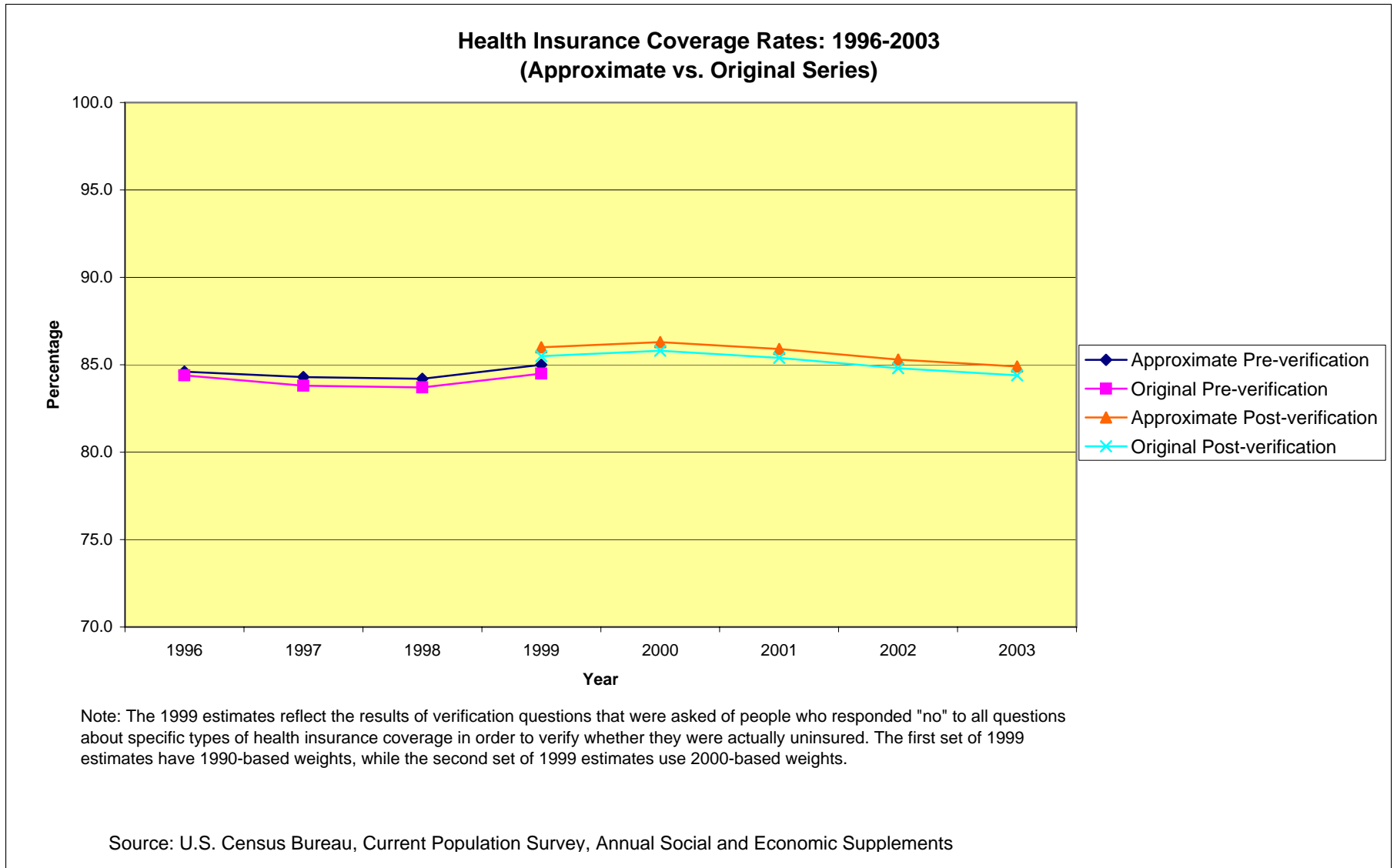


Figure 2: Private Coverage Rates: 1996-2003

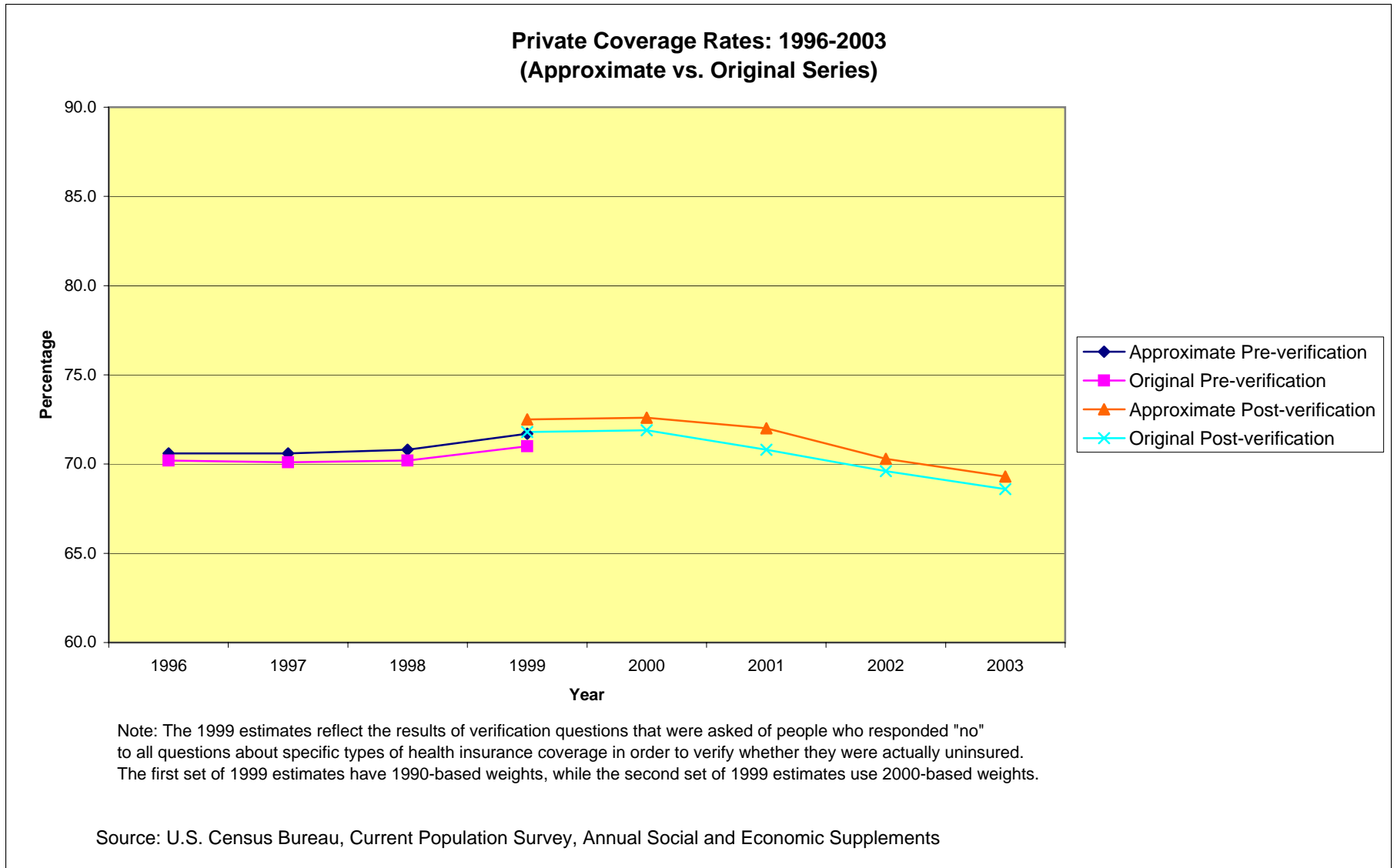


Figure 3: Employment-based Coverage Rates: 1996-2003

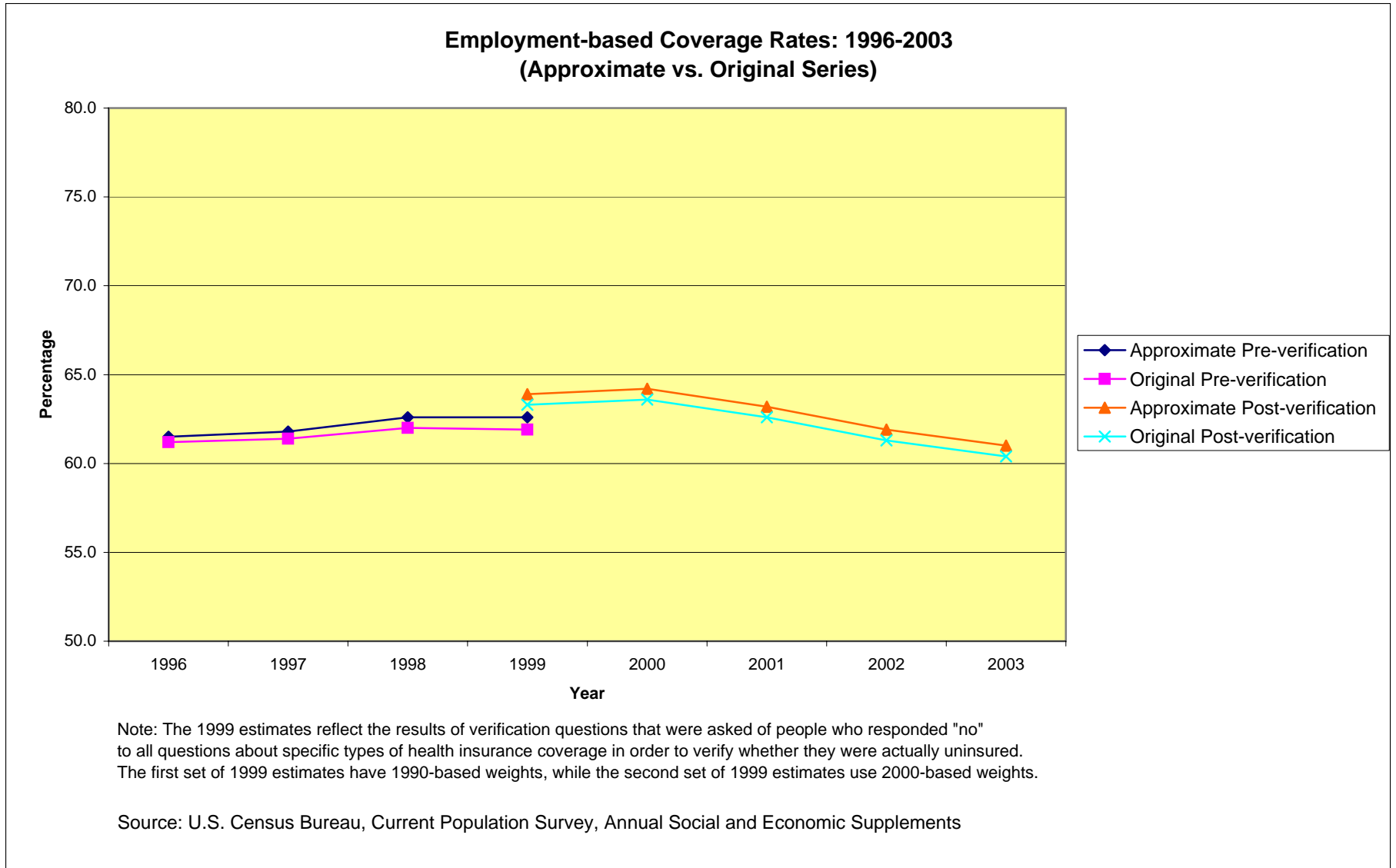


Figure 4: Direct Purchase Coverage Rates: 1996-2003

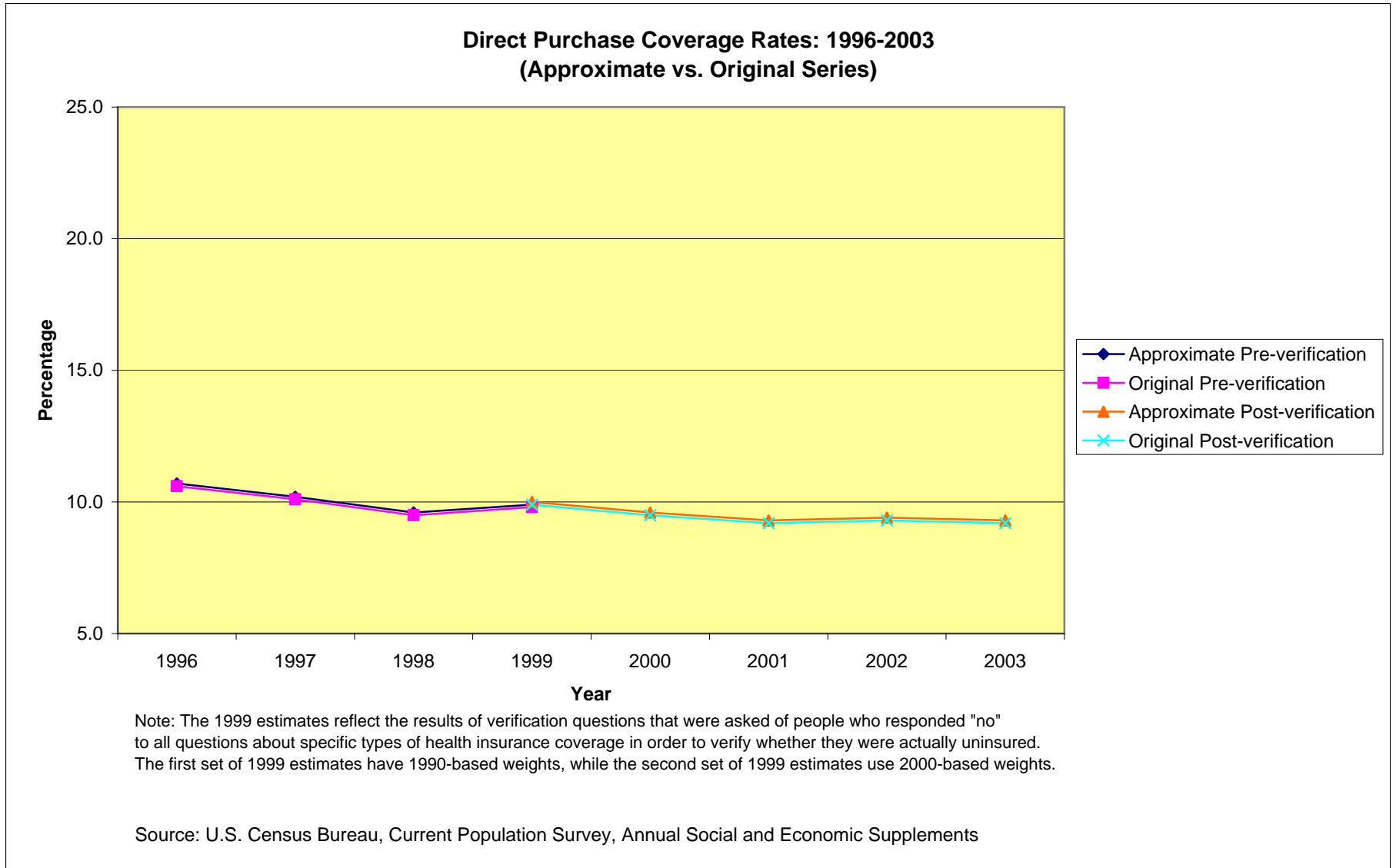


Figure 5: Overall Coverage Rates for Children Under 18 Years: 1996-2003

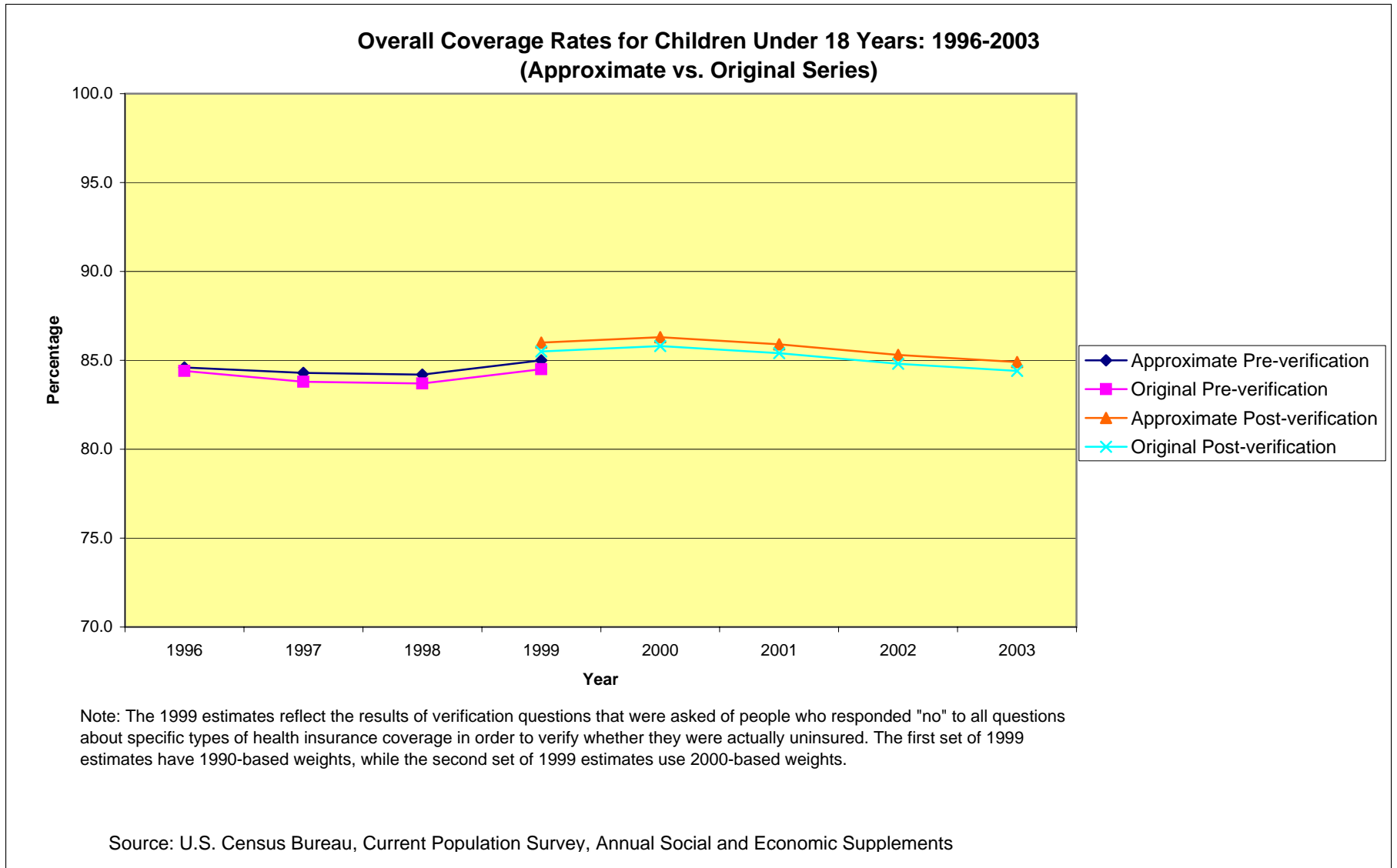


Figure 6: Private Coverage Rates for Children Under 18 Years: 1996-2003

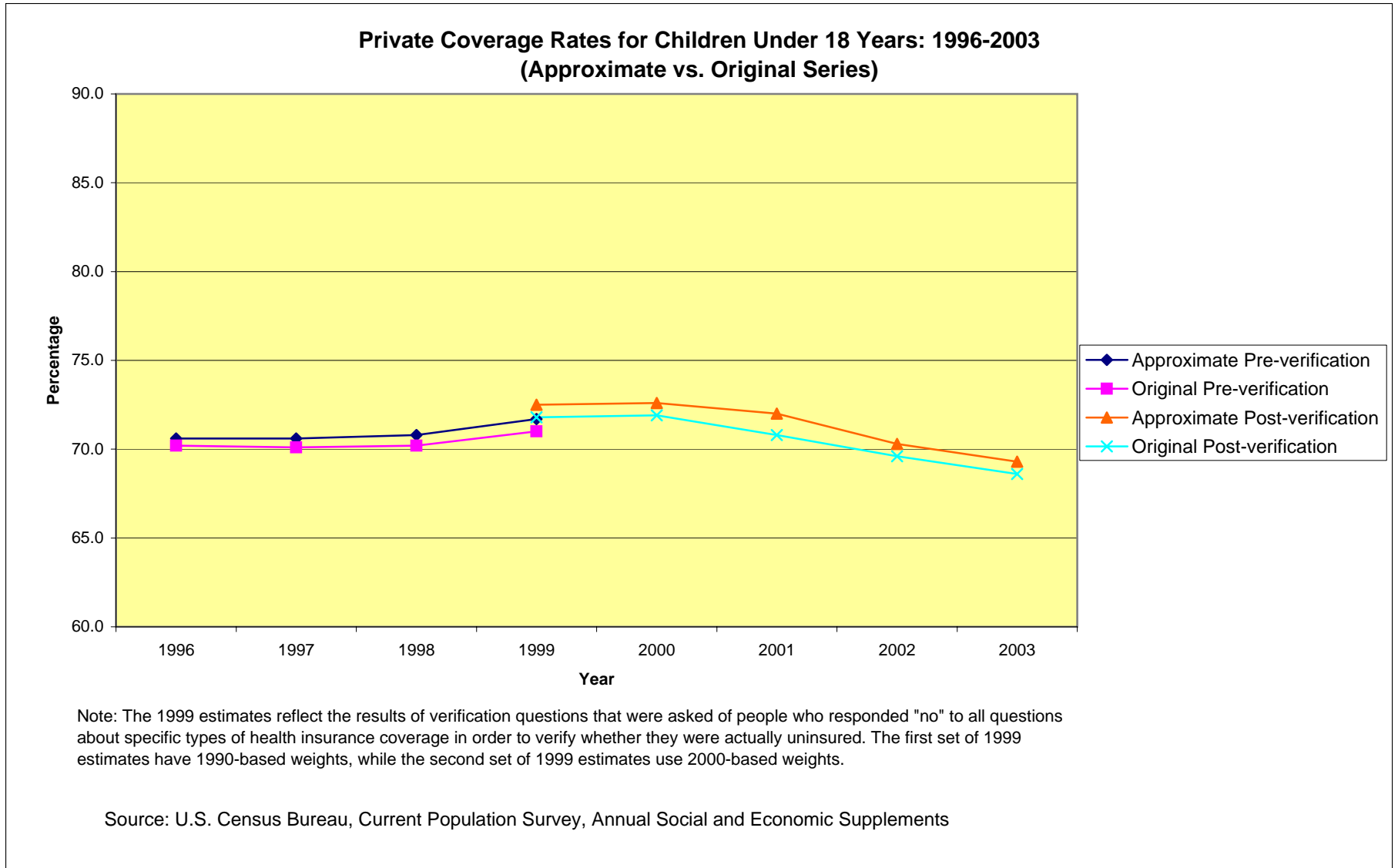


Figure 7: Private Coverage Rates for Non-Hispanic Whites: 1996-2003

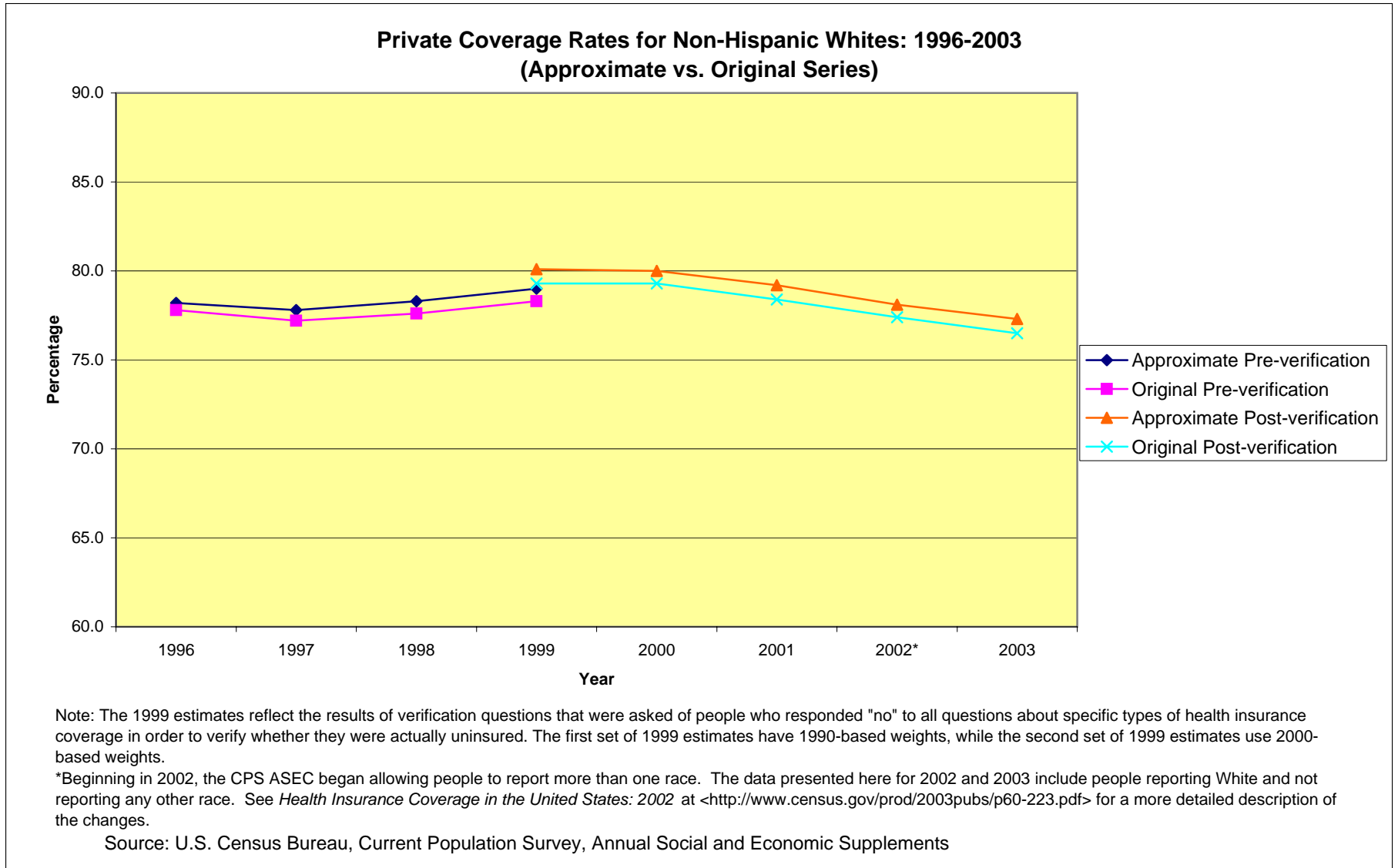


Figure 8: Private Coverage Rates for Blacks: 1996-2003

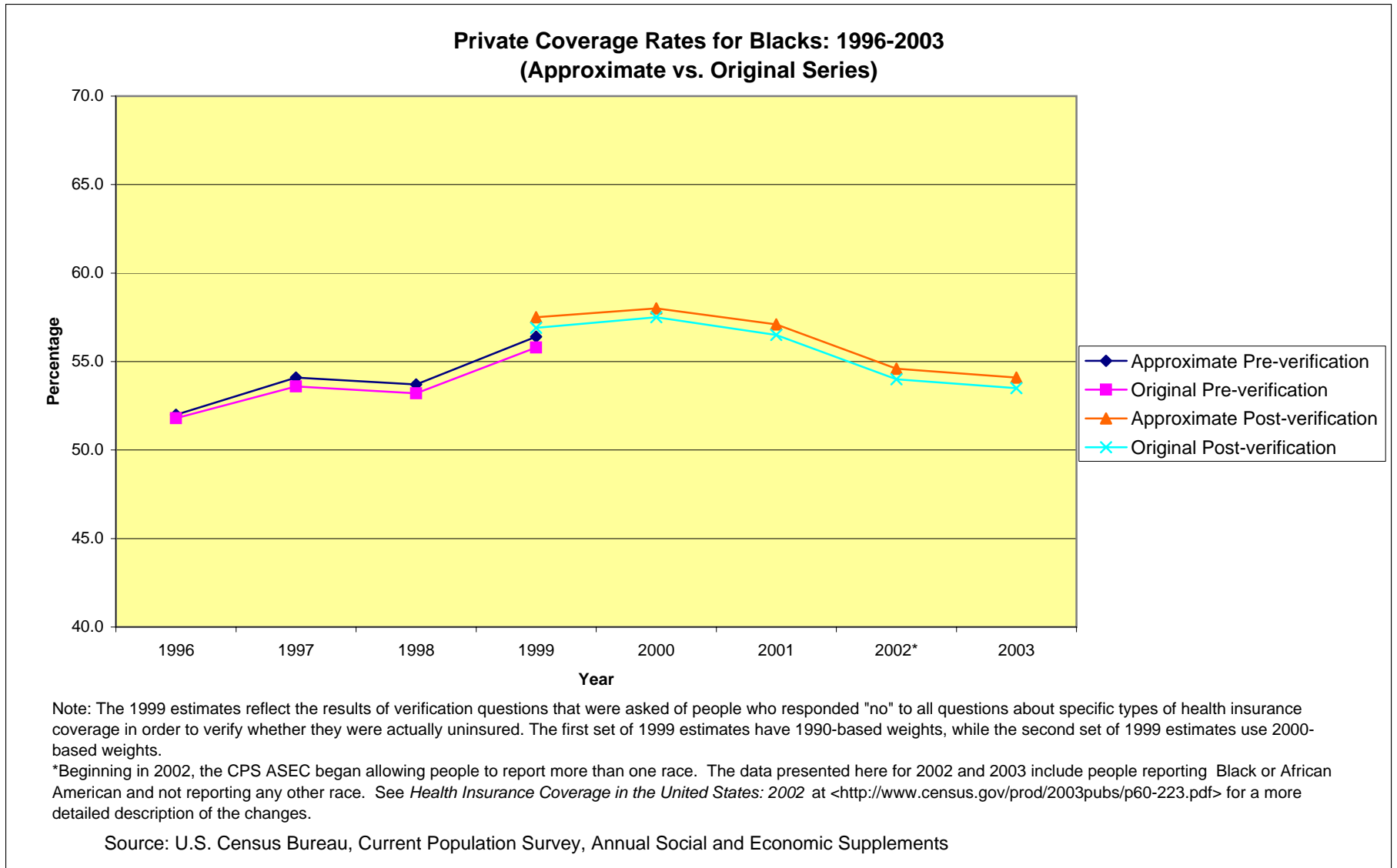


Figure 9: Private Coverage Rates for Asians: 1996-2003

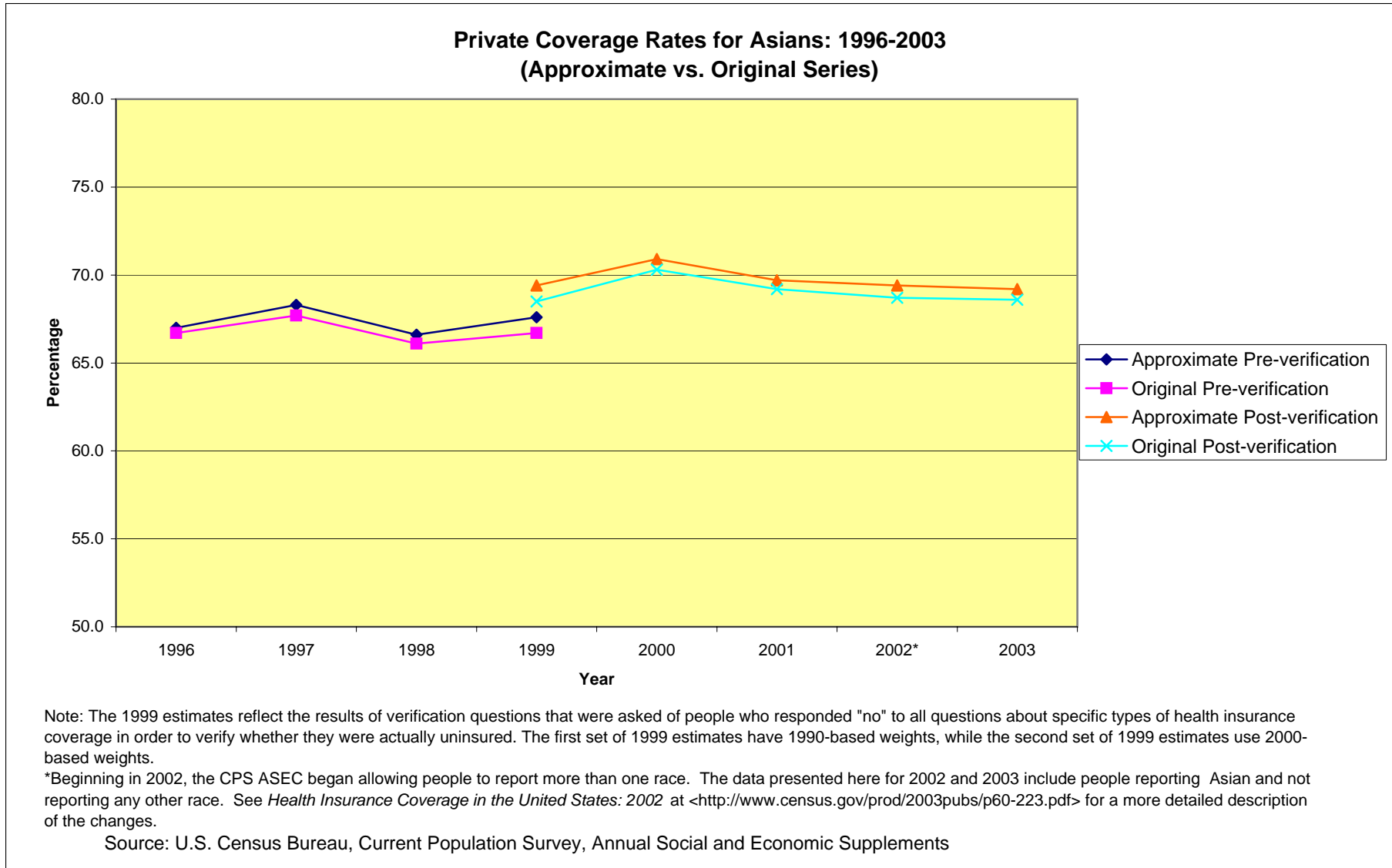


Figure 10: Private Coverage Rates for Hispanics: 1996-2003

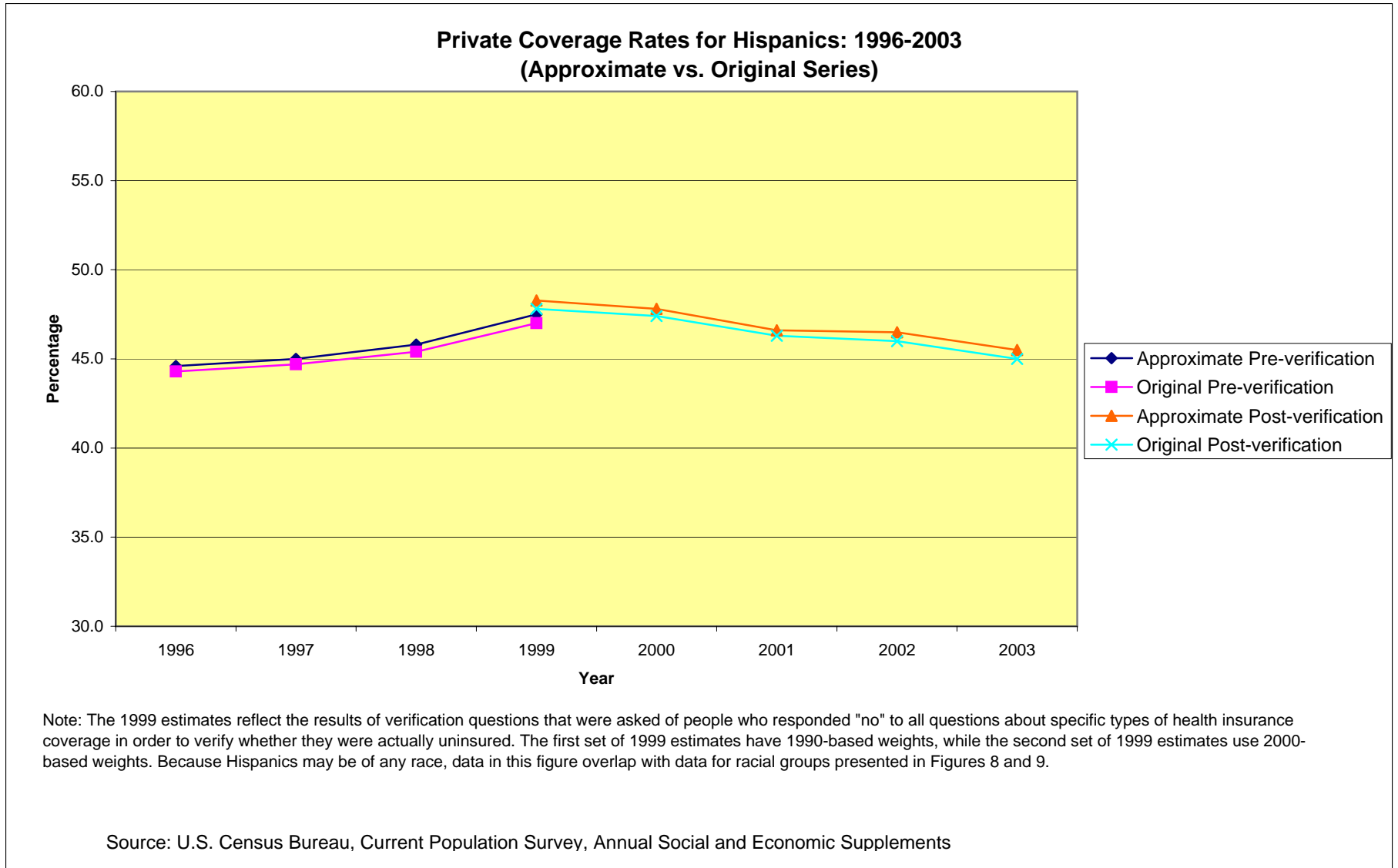


Figure 11: Approximate vs. Revised Values for Types of Health Insurance Coverage: 2004

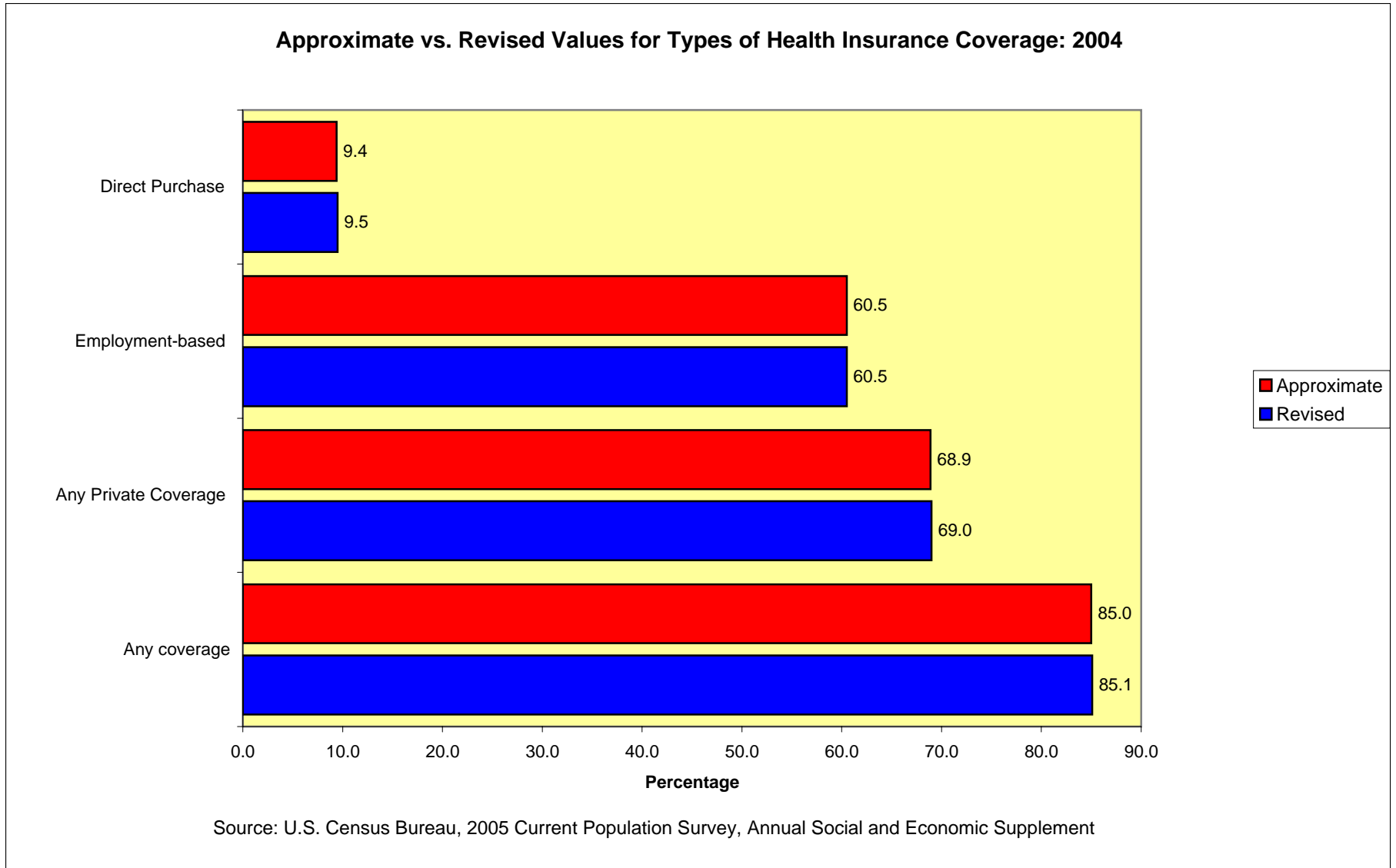


Figure 12: Approximate vs. Revised Values for Types of Health Insurance Coverage: 2005

