

**cepr**

Center for Economic and Policy Research

**Health Insurance  
Data Brief  
*Appendix***

**Technical Documentation**

By Heather Boushey and Joseph Wright<sup>1</sup>

April 13, 2004

CENTER FOR ECONOMIC AND POLICY RESEARCH 1611 CONNECTICUT AVE., NW, SUITE 400  
WASHINGTON, D.C. 20009 (202) 293-5380 <WWW.CEPR.NET> EMAIL: CEPR@CEPR.NET

---

<sup>1</sup> Heather Boushey is an economist and Joseph Wright a research assistant at the Center for Economic and Policy Research.

*This is the Technical Documentation for a five-part series by the Center for Economic and Policy Research (CEPR) on health insurance coverage in the United States.*

*The data used in this series come from CEPR's analysis of the Survey of Income and Program Participation. CEPR created user-friendly Data Sets from this survey. The data and programs are available to the public via our website ([www.cepr.net](http://www.cepr.net)).*

*David Maduram provided valuable research assistance on this project.*

*This project was funded by a generous grant from the Rockefeller Foundation.*

## **Technical Documentation: Health Insurance Data Briefs**

The data in the Health insurance Data Briefs come from CEPR's analysis of the Survey of Income and Program Participation (SIPP). This documentation outlines the key concepts used in the Data Briefs and explains our models. For more information on how we constructed our variables and consistency with other published data on health insurance coverage, please see *Set H Memo* available on our website, [www.cepr.net](http://www.cepr.net)

### **Key concepts:**

*Any insurance* includes any type of government-provided insurance (Medicare, Medicaid, Veterans, Military) or any type of private insurance (employer-provided or privately purchased).

*Private insurance* includes all non-government provided health insurance.

*Medicare* refers to the Federal Health Insurance Program for the Aged and Disabled as provided for by Title XVIII of the Social Security Act. The phrase "Medicare covered" refers to persons enrolled in the Medicare program, regardless of whether they actually utilized any Medicare covered health care services during the survey reference period.

*Medicaid* refers to the Federal-State program of medical assistance for low-income individuals and their families as provided for by Title XIX of the Social Security Act. The phrase “Medicaid covered” refers to persons enrolled in the Medicaid program, regardless of whether they actually utilized any Medicaid covered health care services during the survey reference period. Medicaid coverage also includes children enrolled in the State Children’s Health Insurance Program (SCHIP), begun in 1997. In the 2001 panel, respondents were asked a specific question about including SCHIP, although this was not the case in earlier panels.

*Employer-Provided health coverage in own name* is when an individual is covered by insurance provided by his or her employer, a former employer, or a union. This coverage is under that individual’s own name.

*Employer-Provided health coverage from any source* is when an individual’s health insurance is provided by an employer, be that his or her own or a family member’s employer. This coverage can be in the individual’s own name or a family member’s name.

*Provider* is an individual who uses their employer-provided health insurance to cover someone else in his or her family who lives with that person. This person receives employer-provided health insurance in his or her own name.

*Dependent* is an individual who is covered by the employer-provided health insurance plan of another family member, with whom they live. This coverage is by definition in someone else’s name.

## **Simulations for health insurance coverage in 2002**

In Health insurance Data Briefs #1, #2 and #4, we report the results of a logit analysis using data for 2002, from the 01 panel of the SIPP. This analysis is conducted to determine whether the trends we find in the descriptive statistics hold across groups, once we control for differences among individuals.

Health insurance Data Brief #1 Figures 1 and 2: Universe is all adults

$hiayr = f(wage, gender, race, employment, marital status, adults, foreign, age, age^2)$

Health insurance Data Brief #2 Figures 1 and 2: Universe is all employed adults

$emphiayr = f(wage, gender, race, employment, marital status, adults, foreign, age, age^2)$

Health insurance Data Brief #4, Figure 1: Universe is all adults

$anyempayr = f(wage, gender, race, employment, marital status, adults, foreign, age, age^2)$

where :

*hiayr* is a dichotomous variable for whether or not the individual has health insurance coverage, of any type, all year

*emphiayr* is a dichotomous variable for whether or not the individual has employer provided insurance in his or her own name

*anyempayr* is a dichotomous variable for whether or not the individual has employer-provided health insurance all year; this includes coverage in one's own name or through a family member

*wage* is the hourly wage

*gender* is a dummy for female

*race* dummies are for African-American, Hispanic, Other, and White (omitted)

*employment* status is whether or not the individual has a job during that month

*marital status* dummies are for married (omitted), divorced, separated, widowed, cohabit, and never married

*adults* is the number of individuals, 18 and older, in the family

*foreign* is a dummy for whether or not the individual is foreign-born

*age* and *age*<sup>2</sup> are included to allow for non-linear effects of age on the probability of having health insurance all year

We conduct simulations where we calculate a distribution of 1000 expected values of the probability of having health insurance for particular values of the independent variables. For example, we calculate 1000 expected values for each age between 18 and 64 (for both women and men) in Health insurance Data Brief #2, Figure 1. We set all other explanatory variables at their mean value. We then plot the median value of the expected value over a range of independent variable values (age and wage). This simulation provides us with a substantively meaningful assessment, using controls, of the effect of certain individual characteristics on the probability of having health insurance coverage. We can also plot the 95% confidence intervals expected value simulations. Because we are using large samples (generally more than 30,000), the 95% confidence intervals and the median of the expected values are indistinguishable in the graphs over such a large range of independent variable values.

### **Pooling Latinos across two years**

In each of the tables and figures that look at racial/ethnic breakdowns of health insurance coverage, we pool the observations for Latinos across two years (instead of

for just one year). For the 1992 sample, we pool across calendar years 1992 and 1993 (both from 92 SIPP panel); for the 1999 sample we pool across calendar years 1998 and 1999 (both from the 96 SIPP panel); and for the 2002 sample we pool across the calendar year 2001 and 2002 (both from the 01 SIPP panel). We do this for two reasons. First it increases our sample size for Latinos, producing more robust estimates. Second, by averaging over two years we ensure more accurate *trend* estimates because any single year error will be averaged out.

### **Using data from 1992, 1999, and 2002**

We conducted analysis for every year between 1992 and 2002 (excluding 2000—when SIPP data is unavailable). We chose the year 1999 because it was the closest year (of available data) to the peak of the 1990s economic boom. Similarly, we chose 2002 because it is the most recent year (of available data) of economic downturn. We chose 1992 because it was also a year of economic contraction, at a similar point in the business cycle as 2002, providing a useful comparison for longer-term trends in health coverage.