

Covering California's Uninsured: Three Practical Options

Appendix A: Data Sources

Prepared for
California HealthCare Foundation

Prepared by
M. Susan Marquis, Ph.D.

October 2006

Acknowledgments

This appendix is a supplement to a report authored by Rick Curtis and Ed Neuschler of the Institute for Health Policy Solutions. Rick Curtis of IHPS served as project director, identified alternative approaches for analysis in consultation with experts based in California, and provided substantive guidance for the project as a whole and for the main report. Ed Neuschler of IHPS worked extensively with RAND to identify the public budget and other implications of the alternatives coverage models that are presented in the main report.

Cathi Callahan of the Actuarial Research Corporation produced actuarial values for alternative benefit plans for incorporation into the estimation models, and Jim Mays of ARC provided a range of insightful inputs that were invaluable to the analysis.

About the Author

M. Susan Marquis, Ph.D., of the RAND Corporation developed the database and estimation models and produced the cost estimates for the alternative coverage models.

About the Foundation

The California HealthCare Foundation, based in Oakland, is an independent philanthropy committed to improving California's health care delivery and financing systems. Formed in 1996, our goal is to ensure that all Californians have access to affordable, quality health care.

For more information about CHCF, visit us online at www.chcf.org.

Data Sources

In order to estimate the cost and coverage effects of an array of coverage-expansion scenarios involving varying degrees of employer participation, the project required a database including comprehensive information on establishments and firms in California and their employees and dependents. Unfortunately, no such data exist. Addressing this data limitation required the creation of a database that describes employers and their employees in California based on a number of sources, including:

- Data from the Employment Development Department (EDD) in California;
- The 2002 and 2003 surveys of private employers in California conducted by the Kaiser Family Foundation and the Health Research and Educational Trust (KFF/HRET);
- The 1997 Robert Wood Johnson Family Foundation (RWJF) Employer Health Insurance Survey;
- The 2001 panel of the Survey of Income and Program Participation (SIPP);¹ and
- The 1997 Medical Expenditure Panel Survey.

This final database represents the best information available about businesses and workers in 2003; however all dollar figures have been inflated to 2006 using the CPI for wages and income and the medical component of the CPI for health spending. Premium levels for 2003 were first adjusted to match the average 2004 premium found in the California Employer Health Benefits Survey, 2004 (CHCF/HRET), then increased by 9.2 percent per year to reach 2006 (the latter being the 2004-2005 national growth in the KFF/HRET annual employer survey).

Each of these data sources is necessary for the construction of the analytic employer-employee linked database. The EDD data provide current estimates of the number of firms and workers in California. However, these are aggregate data without any micro-level information on firm or workers characteristics. The KFF/HRET data and the RWJF data provide information on employers in California, including health plan offerings and worker composition. The KFF/HRET data have been recently collected, but they lack several crucial variables and need to be supplemented with RWJF data. The SIPP data provide comprehensive micro-level data on workers, but lack information on medical expenditures. The MEPS data are used to supplement the SIPP with medical expenditure data. What follows is a description of how each of these sources were used in constructing the synthetic database of employers and employees in California and the non-working population.

PRIVATE EMPLOYER DATABASE

The authors of *Covering California's Uninsured: Three Practical Options* obtained special analyses from the EDD of the counts of private employers in California for eight employer size classes, with employer defined to include all establishments in the state that are part of a single firm.² Separate counts were provided for multi-site businesses and single site businesses. EDD also provided total employment counts in each employer size class. The EDD data were used as benchmarks and new weights were developed for the sample of employers included in the 1997

RWJF survey so that the weighted count of employers and employees in the adjusted RWJF sample matched the EDD benchmark totals. These new weights were then used to generate one record for each of the 1,013,506 private employers in California. The 1997 RWJF sample was selected as the base for generating employers because each record provides a count of the number of employees in the state as well as the number of employees in the country. It is this latter measure that is available from the population surveys, so this nationwide employment count is necessary to match workers and businesses.

Information on characteristics of the business—including whether the business offered insurance, the share of employees enrolled in the group plans, its industry, whether it has union employees, and the age composition of the business—were then linked by randomly selecting these characteristics from businesses sampled in the 2002 and 2003 KFF/HRET surveys of the same employer size and multi-site status as the business in our employer population. That is, the KFF/HRET survey is effectively re-weighted to match the EDD benchmarks. Where possible, the pooled 2002 and 2003 surveys were used. However, because data on the age composition was only available from the 2002 survey, the decision was made to select an age mix from the 2002 KFF/HRET sample controlling for business size, multi-site status, and industry.

Characteristics of the plans offered from the 2002 KFF/HRET survey were also selected. A comparison of the 2002 and 2003 KFF/HRET surveys found that the average value of most plan characteristics such as deductibles and copayments remained unchanged from 2002 to 2003. However, premiums had increased about 15 percent for all types of plans. The 2002 premiums were adjusted for this inflation. The 2002 survey data was chosen for two reasons. First, the 2002 survey contains some information about the age of the workers that could be used in matching workers to business; characteristics of plans offered and premiums are likely to differ in businesses that have different age profiles. Second, actuarial values were imputed to the plans offered, and the 2002 KFF/HRET survey provided more measures of the plan characteristics that could be used for this purpose.

In order to obtain more detail about the distribution of characteristics of workers for the employer database, the authors again used the 1997 RWJF survey, which includes a more detailed age breakdown, as well as a distribution of workers by four wage categories,³ by gender, and by union affiliation. Matching these characteristics from the RWJF sample employers to employers in the constructed database included controls for size, multi-site status, does/does not offers insurance, does/does not have any union employees, and the share of young workers.

The resultant database provides basic information about the business, some information about the composition of its workers on several socio-economic dimensions, and information about health insurance plans offered and worker enrollment decisions. Specifically, it yields information about the number of workers by age, gender, and whether in a union and by four age groups,⁴ four wage groups, and five health insurance enrollment decisions (in a non-offering business, offered but not enrolling, enrolling in single coverage, enrolling in two-party coverage, and enrolling in family coverage). These measures provide information about the number of workers in the business on each of several dimensions separately, but they don't provide information about the number of workers in each cell of the joint distribution of these characteristics. For instance, it is possible to know how many workers are in each of four wage groups and how many are in each of four age groups, but not how many people are in each of the 16 groups defined by both wages and ages. The analysis for *Covering California's Uninsured*:

Three Practical Options applies the technique of “iterative proportional fitting” to estimate the number of workers in each cell of the cross-classification of all of our dimensions (Causey, 2003). This technique uses information about the joint distribution of these characteristics in a population database and iteratively adjusts this distribution to match the marginal counts for the employer on each of the dimensions. The resultant solution gives an estimate of the total number of workers in each of the cells described by the full cross-classification of the dimensions listed above. The initial joint distribution of these characteristics is from the SIPP, which is the source of the population data and described in more detail later. The results from the iterative proportional fitting process are used to determine the mix of workers to select to populate each of the businesses.

GOVERNMENT EMPLOYER DATABASE

The sources of data for government employers were EDD and the 1997 RWJF survey. The EDD provided information about the number of local government businesses by eight size groups and the number of workers in each class, and the total number of federal and state workers in California. The KFF/HRET data only sampled private employers and were not used here.

As in constructing the private employer database, this analysis re-weighted local government establishments included in the 1997 RWJF survey so that the number of businesses and the count of employees in each size group matched the EDD benchmark totals. The 3,702 local government establishments were then created. Each establishment has information about the distribution of characteristics of workers and information about health insurance plan offers from the 1997 RWJF establishment that represented it. The technique of iterative proportional fitting described above was used to determine the mix of local government workers to assign to each local government establishment.

A single business entity for all federal workers in California and a single unit for state workers was created, because health insurance decisions are by and large made for the group as a whole. Information about the characteristics of federal and state workers and the health insurance offers were taken from the 1997 RWJF survey. EDD counts of federal and state workers were used as the measure of the unit employment. Because federal and state governments are represented here as a single entities, all workers reporting that they are employed by the federal or state government in the population database were used to populate these two “employers,” with adjustments in the weights in the population database to correspond to the EDD totals.

ACTUARIAL VALUE OF PLANS OFFERED

The benefits of the plans that are offered by employers in our constructed database are represented here in terms of the actuarial value—the share of the total expenses incurred by the group that would be paid for by the plan. Insurance typically covers large medical bills more generously than small medical bills. Therefore, the actuarial value of the plan benefits for workers in four groups classified are calculated according to spending levels based on the quartiles of the distribution of spending in the entire population, as well as an overall measure of the actuarial value. This latter measure is heavily weighted by coverage for large medical bills and does not represent what an average or median member of the group would expect to have covered.

The actuarial values to the plans offered by the private businesses in our database are imputed based on specific coverage provisions of the plans. The Actuarial Research Corporation previously constructed measures of the actuarial values of all plans that were offered by employers in the 1997 RWJF survey. This was done by simulating what each plan would pay for the spending reported by persons with group insurance in the 1987 Medical Expenditure Panel (adjusted to 1997 spending totals). Thus, the actuarial values of two plans differ only because of the benefit design and not because of characteristics of persons enrolled in the plan.

In order to account for changes in benefit design between 1997 and 2003, it is necessary to estimate the relationship between actuarial values and benefit characteristics and then impute actuarial values based on current plan benefit designs. The analysis fits regression models relating the actuarial values to specific features of the plans in the 1997 RWJF survey data, including the amount of the deductible, the amount of the copayment or coinsurance rate, whether coverage was provided for prescription drugs and mental health care, the copayment or coinsurance rate for these services, the out-of-network coinsurance rate for PPO and POS plans, and the maximum out-of-pocket expenditure for PPO and conventional plans.⁵ Separate regressions are fitted for the four types of plan: HMOs, PPOs, POS, and conventional. Because actuarial values fall between 0 and 1, a logistic model is fitted to the data to constrain the predictions to the appropriate range. This provided 20 regression models (the four quartiles of the actuarial values and the overall actuarial value for each of the four types of plan). Also available are four sets of five empirical residual distributions—that is, the residuals from fitting the five equations for each of the four plan types. Actuarial values for each of the plans in the KFF/HRET survey were then imputed by using the regression to predict the expected value of each of the five actuarial values associated with each plan and drawing randomly from the residual distribution to account for unobserved characteristics. The five residuals are jointly selected to preserve the correlation in the distribution of errors across the actuarial values. That is, cases that have above-average actuarial value for the lowest quartile of expenditures, given the observed characteristics of the plan, also have an above average actuarial value for the other expenditure quartiles. The average correlation among errors for the HMO and conventional plans was 0.71 and it was 0.61 for the PPO and POS plans.

POPULATION DATABASE

The primary data source for information about the population and its demographic, economic, and insurance characteristics is the 2001 SIPP panel. The SIPP provides information about work, insurance coverage, and access to group insurance for all family members—information that is necessary to accurately characterize who will be affected by reforms. [In contrast, the Census Bureau’s Current Population Survey’s Annual Social and Economic Supplement (CPS) does not provide information about access to group insurance, and the California Health Interview Survey (CHIS) does not provide data about all family members.] In addition, the SIPP provides demographic data and some information about health status, while the CPS lacks equivalent information about health status. All 8,163 California respondents were selected from the May 2002 cross-section of the SIPP panel. The May cross-section was chosen because it is the wave of data collection from the panel that includes questions about whether the employer offered insurance.

Respondents in the SIPP who are workers are used to populate each of the private and public employers with employees. Private sector workers are sorted into one of 960 strata—defined by gender, whether union member, four wage groups, four age groups, five health insurance groups and three business size groups—fewer than 50 workers, 50-199 workers, and 200 or more workers. Workers from these strata were randomly selected to populate each private employer, given the characteristics of the workers estimated as described earlier. Sampling of workers was done with replacement, so that a single worker may be assigned to more than one business, and also may be assigned more than once to a single business. In effect, this reweights the SIPP sample to accord with the characteristics of the worker population based on the employer database.

A similar matching was carried out for local government workers, but only two size classes were used (fewer than 50 and 50 or more). As noted earlier, this analysis has only one employer representing federal employment and one employer representing state employment, and re-weight these observations in the SIPP to match the total employment counts provided by EDD. The distribution of characteristics were compared using the SIPP weights and in the re-weighted database, and they were quite similar. That is, the development of the synthetic database did not substantially alter the estimates of the distribution of characteristics of workers.

The employer measure of size used here is the number of workers employed in the state of California. The SIPP includes two size measures—the size of the establishment and the size of the firm nationwide. In order to obtain a measure of size of the employer that matches the authors’ concept and to obtain size classifications for the SIPP that match those used here, a random selection was made of a business in 1997 RWJF database (which includes measures of establishment size, statewide employment in the firm, and national employment) from the same nationwide firm and establishment size group, industry, and multi-site status as the SIPP worker and attributed the statewide employment size from the selected business to the worker.

Dependents are linked to workers, and so all dependents are re-weighted according to the selection of the worker. Insured dependents are linked to the worker who provides their group coverage. If the dependent does not have group coverage, the dependent is linked to the worker in the largest business size group. In the case of ties, the dependent is linked with the male worker, since the majority of dual worker families provide coverage to dependents through the male’s employer (Marquis and Kapur, 2004).

People in families that do not include any workers are retained in the data set with their SIPP sample weight.

Workers declining group coverage

The resultant database identifies workers who are in firms that offer coverage and workers who enroll in offered coverage; however, it does not make it possible to distinguish between those workers who are ineligible for a plan that the firm offers and workers who are eligible but decline to participate. This distinction is important, because it assumes that all eligible workers will participate in employer-sponsored plans under the mandates. This analysis used data from the California Health Interview Survey on turn-down rates by family income and insurance status to impute a value of ineligible for coverage versus declined to accept to workers in our database in offering firms who do not have employer coverage.

Legal status of immigrants

The legal status of non-citizen residents affects their eligibility for federal matching funds under public health insurance programs and is generally of policy concern. Therefore, it is desirable to be able to distinguish immigrants who are in the country legally from those who are not. Length of time in the U.S. (more or less than five years) is also of policy interest.

The SIPP interview asks respondents whether they are citizens and if not, how long they have been in the country. All of those reporting that they are not citizens and have been here less than five years were identified as legal immigrants in the country for fewer than five years. The weighted SIPP count of cases was 0.5 million in contrast to a count of about 1.1 million based on data from the Yearbook of Immigration Statistics. To match the 1.1 million count, therefore, some other non-citizen respondents were randomly assigned to the category of legal immigrants who have resided in the U.S. for fewer than five years.

To identify undocumented workers, estimates were developed for the distribution of undocumented immigrants by poverty status and insurance status (insured vs uninsured) based on work previously done by the Institute for Health Policy Solutions (Taylor and Wong, 2003). Data from the Migration Policy Institute indicate that there were about 2.3 million undocumented immigrants in California. A status of “undocumented immigrant” was imputed to respondents reporting that they were non citizens who were not otherwise identified as being legal citizens to achieve a weighted count of 2.3 million persons with the appropriate insurance and income distribution.

EXPECTED HEALTH SPENDING BY THE POPULATION

The SIPP collects information about the health status of respondents, but it does not include information about expected health spending. This analysis used expected health spending by group members to estimate premium costs and to estimate out-of-pocket spending. Health care spending was imputed to people in our SIPP sample based on health care spending reported by the sample in the 1997 MEPS. In an earlier study, ARC aged the health care spending reported by all privately insured respondents to the 1997 MEPS to the year 2002 using the CMS National Health Accounts as the benchmark. Those numbers were inflated here to 2006 using the medical component of the CPI. For each person in the MEPS, data is available for their spending, their age, and their self-reported health status. The analysis stratified the MEPS sample into 30 groups based on six age categories and the five levels of reported health status. A level of health care spending for respondents in the SIPP sample is randomly selected from the appropriate stratum. This is a measure of actual spending in one period. In order to obtain a measure of expected spending for each person and the distribution of spending for each person, this process is repeated five times. Expected health care spending for the individual is then measured as the average over the five measures.

REFERENCES

Marquis, M.S. and K. Kapur, “Family Decision Making When Two Workers Are Offered Coverage”, 2004, <http://www.dol.gov/ebsa/>.

Causey, B.D. “Dual System Estimation Based on Iterative Proportional Fitting”, Census/SRD/RR-84/03, 2003, <http://www.census.gov/srd/papers/pdf/rr84-03.pdf>

Taylor, L.Q. and L. Wong, “Participation Considerations for Non-Citizen Workers in the “Working for Health” Program,” Institute for Health Policy Solutions, December 2003.

ENDNOTES

¹ The SIPP was chosen as the source for population data, in preference to the Census Bureau's Current Population Survey (CPS) or the California Health Interview Survey (CHIS) for reasons discussed in the "Population Database" section below.

² The size classes provided by the EDD were 0-19, 20-49, 50-99, 100-199, 200-499, 500-999, 1000-4999, and 5000+.

³ The categories in 2003 dollars were less than \$8 per hour, \$8-11.5 per hour, \$11.5-17 per hour, and above \$17 per hours. The 1997 survey dollars were inflated to 2003 dollars using the Consumer Price Index.

⁴ Under age 30, 30-40, 40-50 and older than 50.

⁵ Dollar values in 1997 were inflated to current dollars using the medical services component of the Consumer Price Index. Since the actuarial value is a ratio of dollars, it is not adjusted.