

Municipal-level covariates of health status in Brazil: a proposed method for data interpolation

Frederico C. Guanais¹

Suggested citation

Guanais FC. Municipal-level covariates of health status in Brazil: a proposed method for data interpolation. *Rev Panam Salud Publica*. 2013;34(3):190-7.

ABSTRACT

Objective. To propose a method for the interpolation of yearly local-level covariates of health status that is suitable for panel data analysis of the effect of health services.

Methods. The proposed method distributes the yearly rate of growth of covariates at the regional level (e.g., state) from household survey data, and applies it to interpolate yearly data at the local level (e.g., municipality) between two consecutive census surveys. The method was applied to municipal-level socioeconomic covariates of health status in Brazil for every year between 2001 and 2009. The data was tested on a previously validated analysis of the effects of the Family Health Program on post-neonatal infant mortality in Brazil.

Results. A total of 895 628 values were generated for 20 socioeconomic predictors of health status. Valid data were obtained for 5 057 municipalities in the Northeast, Southeast, South, and Center-West regions of Brazil, from 2001 to 2009, covering 98.89% of the municipalities in these regions and 90.87% of municipalities in the country. A supplemental annex includes the interpolated data from 2001 to 2009, plus the 2000 and 2010 census data, for all 5 057 municipalities. An application on a fixed-effect regression model suggested that, compared to linear interpolation, the proposed method reduced multi-collinearity and improved the precision of the estimates of the effects of health services.

Conclusions. The advantages of the proposed interpolation method suggest that it is a feasible solution for panel data analysis of health services at the local level in Brazil and other countries.

Key words

Socioeconomic factors; family health; infant mortality; models, statistical; Brazil.

The increasing availability of administrative and epidemiological data on health services provision in Latin America and the Caribbean creates a very attractive opportunity for analyses by public health researchers. Such studies could provide invaluable contributions to policy makers, health professionals, and users of health systems, contributing to the design and implementation

of higher-quality, more accessible, and lower-cost services. Nevertheless, an important limitation of the analyses of administrative health service data at the local level (e.g., a municipality or city) is that they often lack proper control for socioeconomic characteristics of the population served, thus generating biased results (1, 2). In the last decade, ever-increasing attention has been given to the importance of the social determinants of health, and the subsequent need for controlling for socioeconomic covariates in health systems evaluation research (3, 4). Ignoring such aspects could lead to

potentially misleading results and incorrect conclusions.

With some level of variation, yearly household sample survey data are available for Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, El Salvador, Honduras, Jamaica, Mexico, Paraguay, Uruguay, and Venezuela, and census data are generally available in 10-year intervals. Within this setting, detailed cross-section studies that include socioeconomic covariates at the local level are only possible for either 1) cross-sections of a single census year or 2) longitudinal analyses of data from the

¹ Inter-American Development Bank, Washington, District of Columbia, United States of America. Send correspondence to: Frederico C. Guanais, frederico@iadb.org

years occurring during the 10-year gaps between national census surveys. Typically, yearly household-sample survey data can be considered representative of other aggregate levels such as country, state, or province, and, in some cases, metropolitan regions. In contrast, yearly panel data that are representative for the local level and have broad coverage are rarely available for countries in Latin America and the Caribbean.

Lack of yearly, local-level data on socioeconomic variables generates a few known problems for ecological panel data analysis. First, the level of intervention may not be the same as the level of analysis. For example, in the case of Brazil, municipal administrations are responsible for managing primary care services, which suggests that the analysis should be done at this level or lower if possible (5). Second, the lower number of states or provinces compared to municipalities or cities reduces the number of available observations in a panel data set. This can lead to analyses with low statistical power, and effects of useful health interventions may go undetected.

For researchers who want to analyze the relationship between data on health service provision and health outcomes at the municipal level, a few different approaches may be used. First, the analyst may decide to ignore socioeconomic covariates in the statistical models. In this case, the omitted variable bias may lead to an overestimation (or underestimation) of the variable of interest. For example, during 2000–2010, expansion of primary health care in Brazil occurred simultaneously with an increase in educational attainment, particularly in the lower-income regions of the country (6). Omitting education variables from a multivariate model that estimates the impact of primary health care provision on infant mortality would thus overestimate the impact of the health programs on infant mortality, because parental education is an accepted predictor of child health (7, 8).

Second, in the absence of yearly socioeconomic data, the researcher may decide to conduct a linear interpolation, assuming that the rate of change of the variable of interest is constant across the 10-year gaps between census surveys. Although linear interpolation has been used in several studies published in peer-reviewed journals (9–11), and

despite the appeal of its simplicity, use of this approach creates a few methodological issues. One is that having all variables change at a constant rate will lead to problems with multi-collinearity, which is likely to reduce the precision of the estimation and cause large changes in the estimates when new variables or observations are introduced (12). Another problem is that the assumption that the rate of change is uniform across all years may not be realistic. Yet another issue is that the approach fails to take advantage of all available data for inferring the value of socioeconomic covariates, given that many countries have yearly state-level household survey data.

Third, the researcher may rely on multiple imputation, typically when data is missing at random (13, 14). Use of this method requires that predictors are available for each missing variable. The imputation generates multiple versions of completed data sets, the statistical model of interest is fitted to each one, and the results are combined to obtain an estimate of coefficients and standard errors. Because of these requirements, this method is not appropriate for interpolating a large number of observations that are systematically missing over time, and for which there are few reliable predictors.

Fourth, the researcher may decide to use a more sophisticated method, such as the one that has been used to construct small-area estimation of poverty rates, commonly known as “poverty maps” (15, 16). This approach is complex and difficult to implement, very computationally intensive, and requires access to individual-level census data, which are often protected due to issues of confidentiality of personal information. As a result, while poverty maps could be considered the gold standard for imputation of small-area data, for all the reasons mentioned above there is typically a lag of several years between the publication of census data and the preparation of imputed data sets. In addition, these data sets are often only available for poverty rates or income data, and for a limited number of countries.

The solution proposed in this report is to combine municipal- (or city-) level data, which is available for census years, with state- (or province-) level data, which is available for every year in several countries in Latin America and the

Caribbean. The approach is an adaptation of a suggestion originally made for analysis of economic growth in the United States, and distributes the rate of growth in the covariate of interest according to the patterns observed in a higher hierarchical level, such as a state, province, or metropolitan region (17). Use of this approach is advantageous versus alternative approaches because 1) it eliminates the issues that would be created by omitting socioeconomic variables; 2) it reduces multi-collinearity across variables, and uses more available subnational data, compared to linear interpolation; and 3) its implementation is simpler and faster than the “poverty maps” method.

This report describes the results of the use of the proposed interpolation method for 1) socioeconomic covariates in Brazil between 2000 and 2010 and 2) analysis of the effect of the Brazil’s Family Health Program (FHP) on infant mortality in Brazil compared to results obtained using linear interpolation. Yearly interpolated data at the municipal level are also provided in a [supplemental electronic annex](#). The case of primary health care in Brazil was chosen for the study because it has been well documented by several studies published in peer-reviewed public health literature (18–20).

MATERIALS AND METHODS

The interpolation of municipal-level data from Brazil between two census years (2000 and 2010) followed a three-step method beginning with Formula A:

$$X_{M,2000+t} = X_{M,2000+t-1} \times \left(\frac{\ln\left(\frac{X_{S,2000+t}}{X_{S,2000+t-1}}\right)}{\ln\left(\frac{X_{S,2010}}{X_{S,2000}}\right)} \times \ln\left(\frac{X_{M,2010}}{X_{M,2000}}\right) \right) \quad (A)$$

In Formula A, X is the variable to be interpolated (i.e., the percentage of the population in the municipality with access to clean water supply), and the subscripts are defined as follows: M, municipal level; S, state level; 2000, year of the first of two consecutive census surveys; and 2010, year of the second of two consecutive census surveys. The index *t* ranges from 1 to 9 for inter-census years. The values of $X_{M,2000}$, $X_{M,2010}$, and the set of 11 elements from $X_{S,2000}$ to $X_{M,2010}$ are known a priori.

The proposed formula generates a curve for each municipality that is an-

chored on the data points of the two census years, and that deviates from the linear interpolation according to the shape of the state-level variation. However, due to the mathematical properties of the formula, there are a few cases in which the imputed data may grow exponentially and result in implausibly high (or low) results. To eliminate this possibility, a second step is recommended to smoothen the deviation from the linear interpolation. This second step is shown in Formula B:

$$Y_{M,2000+t} = \frac{1}{10}(X_{M,2010} - X_{M,2000}) + Y_{M,2000+t-1} \quad (B)$$

In Formula B, Y corresponds to a linear interpolation of the same concept expressed by X (e.g., access to water). The values of $X_{M,2000}$ and $X_{M,2010}$ are known a priori, and $Y_{M,2000} = X_{M,2000}$. The third and final step is to adjust the imputed value to limit the possible deviation from the linear interpolation, which can be accomplished using Formula C below:

$$Z_{M,2000+t} = Y_{M,2000+t} + \frac{X_{M,2000+t} - Y_{M,2000+t}}{e^{(0.03 \cdot |X_{M,2000+t} - Y_{M,2000+t}|)}} \quad (C)$$

The factor 0.03 in Formula C limits the deviation between the proposed interpolation and the linear interpolation (the value of the fraction term in the formula) to a value of approximately 12.26. This is an arbitrary limit, and other values could be used. For example, a factor of 0.02 would limit the maximum deviation to approximately 18.39, and a factor of 0.04 would limit the result to approximately 9.20. The specification of Formula C, with an inverse of an exponential term, seeks to counterbalance the tendency for exponential growth in some circumstances in Formula A while interfering less for smaller values of the interpolated value.

The proposed method was used to interpolate socioeconomic covariates per year from 2001 to 2009 in 5 117 municipalities in Brazil's Federal District and in the following 19 Brazilian states: Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte, Sergipe, Espírito Santo, Minas Gerais, Rio de Janeiro, São Paulo, Paraná, Rio Grande do Sul, Santa Catarina, Goiás, Mato Grosso, and Mato Grosso do Sul. For the country's remain-

ing seven states (Acre, Amapá, Amazonas, Pará, Rondônia, Roraima, and Tocantins), all in the Northern region, available household survey data before 2003 were not representative of the entire state population (only the urban population was included in the survey) and therefore the interpolation method was not applied. State-level data were substituted for metropolitan-region data in eight cases (Belo Horizonte, Curitiba, Fortaleza, Porto Alegre, Recife, Rio de Janeiro, Salvador, and São Paulo) because the metropolitan regions are clusters of municipalities and are hierarchically lower than states.

Municipal-level data for 2000 and 2010 were obtained from Brazil's demographic censuses (21, 22), and state and metropolitan region-level data from 2001 to 2009 were obtained from the national household sample surveys (23). The censuses and surveys were conducted by the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística*, IBGE).

A total of 20 socioeconomic covariates were interpolated and divided into three categories of socioeconomic variables: household characteristics, education, and income. For household characteristics, the following covariates were interpolated: "has clean water supply"; "has garbage collection"; "has clothes washer"; "has refrigerator"; "has television"; and "has radio." For education, the percentage of people aged 5 years or older who can read and write, and the percentage of people aged 10 years or older, were interpolated in each of the following categories: "no schooling" or "less than primary school education"; "completed primary school"; "completed secondary school"; and "completed college." For income, the percentage of people aged 10 years or older in each municipality was interpolated in each of the following categories: "no earnings"; "≤ one-half minimum wage"; "> one-half minimum wage" and "≤ one minimum wage"; "> one and ≤ two minimum wages"; "> two and ≤ three minimum wages"; "> three and ≤ five minimum wages"; "> five and ≤ 10 minimum wages"; "> 10 and ≤ 20 minimum wages"; and "> 20 minimum wages." The "no earnings" category included families that only received cash benefits so can be understood as a proxy of poverty incidence.

To demonstrate the usefulness of the proposed formulas, a multivariate analysis of the association between the expansion of the FHP in Brazil between 2000 and 2010 and post-neonatal infant mortality rate was conducted. The FHP is a community-oriented model of delivery of primary health care services that includes multidisciplinary teams of health professionals. The program has expanded rapidly since 1994 to its current coverage of 109.3 million registered users or 57.3% of the Brazilian population (24). The current application is an extension of analyses conducted by several authors in the peer-reviewed literature, and is presented in the current study only as an illustration of the proposed interpolation method (i.e., not as a demonstration of the impacts or effects of the FHP).

Fixed-effect regression models were estimated for the municipal level from 2000 to 2010, with the post-neonatal mortality rate as the dependent variable, and coverage of FHP as the main independent variable of interest, controlling for socioeconomic covariates. Both coverage of the FHP and the socioeconomic covariates that were included are known predictors of post-neonatal infant mortality. Municipal fixed-effects control for all time-invariant characteristics specific to each municipality throughout the period of analysis. Dummy variables for all but one year were included in all models to control for secular trends that affected all municipalities simultaneously, such as national economic growth. The full specification of the population model is presented in Equation D:

$$PNIMR_{i,t} = \beta_0 + \beta_1 fhp_{i,t} + \beta_2 water_{i,t} + \beta_3 garb_{i,t} + \beta_4 educ_{i,t} + \beta_5 inc_{i,t} + \alpha_i + \lambda_t + \varepsilon_{i,t} \quad (D)$$

In Equation D, $PNIMR$ is the post-neonatal infant mortality rate at the municipality i in the year t , fhp is the proportion of population covered by FHP, $water$ is the percentage of the population with access to safe water supply, $garb$ is the percentage of the population with garbage collection at home, $educ$ is the percentage of the population with less than primary school, inc is the percentage of the population with no income, α_i is a dummy variable for each municipality, λ_t is a dummy variable for every year but one, and $\varepsilon_{i,t}$ is a population error term.

All analyses and interpolations were conducted using STATA software (25).

RESULTS

Using the proposed interpolation algorithm, a total of 895 628 values were generated for 20 variables representing socioeconomic predictors of health status or their proxies. Valid data were generated for a maximum of 5 057 municipalities in the Northeast, Southeast, South, and Center-West regions of Brazil, representing 98.89% of the municipalities in these regions, and 90.87% of the municipalities in Brazil. The data set generated by the algorithm is a few hundredths of a percentage point short of entire coverage due to missing data in the original data set used for the interpolation algorithm, caused by the division or creation of municipalities during the decade, and other reasons. Table 1 shows population-weighted, average results for the imputed socioeconomic covariates. The [supplemental electronic annex](#) includes a detailed table with the values of all variables for all 5 057 municipalities.

Examples from selected municipalities by population size are included in Figures 1 and 2 to illustrate the results of using the proposed interpolation method versus linear interpolation. Figure 1 shows 1) the percentage of people aged 5 years or older who can read and write, in selected Brazilian municipalities, states, and metropolitan regions, from 1998 to 2010, according to linear interpolation, and 2) the same data obtained using the proposed interpolation method. It also includes the values of the variable of interest at the state level, and at the metropolitan region, for the municipality of São Paulo. As shown, the initial and final values of each municipal series are the same for the years 2000 and 2010. The yearly values obtained using the proposed method vary according to the shape of the state-level data for the three smaller municipalities, and according to the shape of the metropolitan region-level data for one municipality in one metropolitan region (the metropolitan region and municipality of São Paulo).

Figure 2 presents the percentage of households with access to clean water supply, in selected Brazilian municipi-

palities (a different set than those used in Figure 1), states, and metropolitan regions. As in Figure 1, the selected municipalities vary in population size, for illustration purposes.

To apply the data obtained using the proposed interpolation method and compare it with data obtained using linear interpolation, the relationship between the expansion of Brazil's FHP and infant mortality was estimated using both data sets. The results are shown in Table 2.

Model 1 is a fixed-effects model at the municipal level that shows strong association between the FHP and lower infant mortality rate. As dummy variables representing all but one year in the series are introduced in Model 2, the magnitude of the association is considerably reduced. This was expected because the expansion of the FHP is correlated with the passage of time, and the coefficient in Model 1 overestimated its effects.

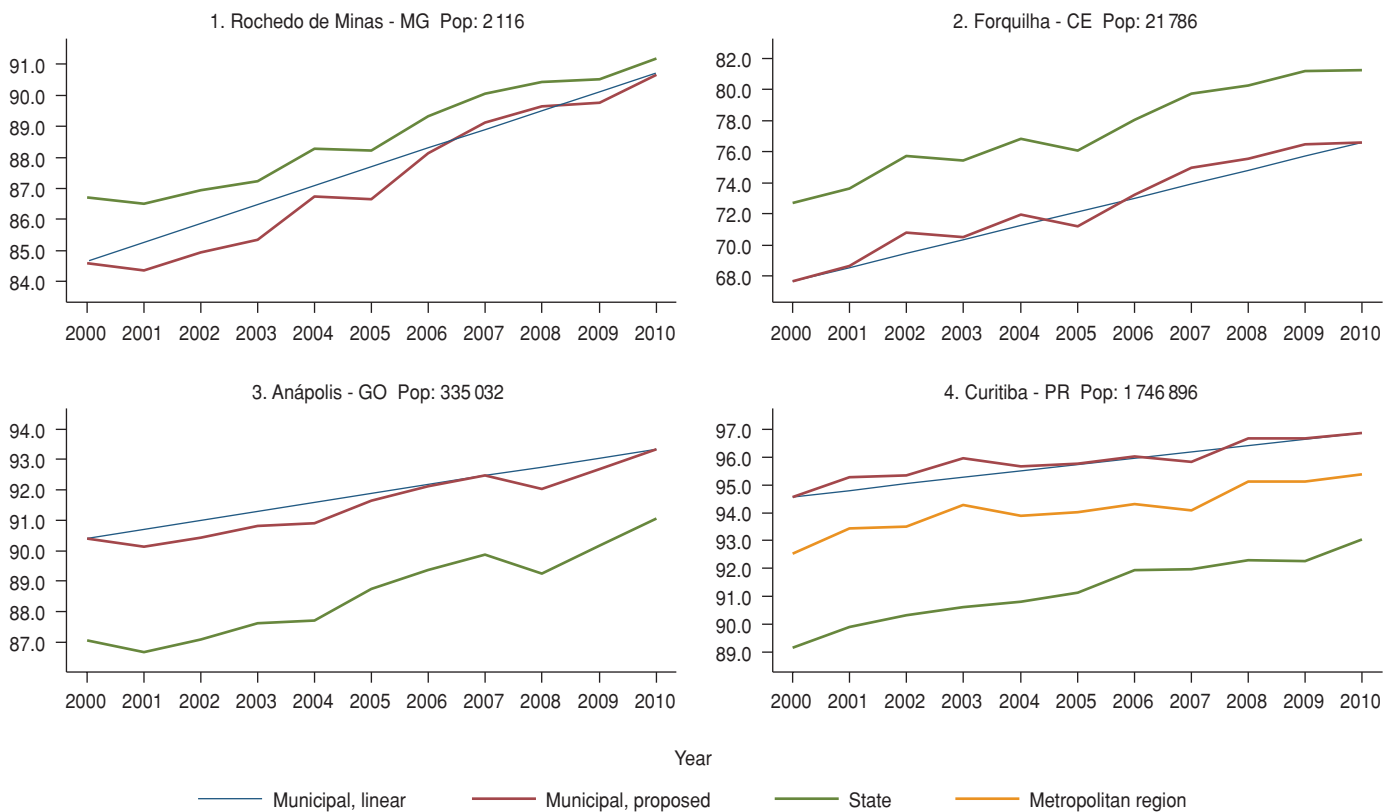
Models 3–6 progressively add socioeconomic covariates obtained through simple, linear interpolation. As more variables are introduced, the magnitude of the association between the FHP and

TABLE 1. Population-weighted, average values of socioeconomic covariates for 5 057 municipalities (excluding the Northern region), based on proposed (nonlinear) interpolation method, Brazil, 2001–2009^a

Covariate	2001	2002	2003	2004	2005	2006	2007	2008	2009
Household characteristics (% of households)									
Has clean water supply	79.00	80.16	80.79	81.86	82.10	83.04	83.77	84.42	84.98
Has garbage collection	74.12	74.78	76.54	76.76	78.83	79.42	79.51	78.24	81.47
Has clothes washer	34.21	37.95	42.56	47.38	53.65	60.72	67.00	73.36	82.09
Has refrigerator	84.01	81.74	78.12	74.32	69.89	66.62	63.30	58.62	55.00
Has television	87.96	87.84	86.74	86.41	86.14	86.49	86.34	85.39	84.82
Has radio	90.04	90.86	92.05	93.47	94.92	95.90	97.26	98.26	98.46
Education									
Can read and write (% of people ≥ 5 yrs old)	85.32	85.76	86.30	86.55	86.96	87.78	88.32	88.68	88.97
Less than primary school education (% of people ≥ 10 yrs old)	62.43	60.59	58.62	56.86	55.81	54.05	52.45	50.90	50.06
Completed primary school (% of people ≥ 10 yrs old)	17.59	15.31	15.87	16.44	16.24	16.39	17.20	17.00	16.36
Completed secondary school (% of people ≥ 10 yrs old)	16.72	18.23	19.42	20.60	21.67	22.83	23.63	24.67	25.64
Completed college (% of people ≥ 10 yrs old)	4.76	5.06	5.30	5.52	5.72	6.25	6.62	7.00	7.51
Income (number of BMWs ^b earned) (% of people ≥ 10 yrs old)									
None	31.72	30.29	29.54	29.34	28.39	27.32	27.09	25.96	26.32
> 0 and ≤ ½	5.07	6.64	7.34	7.53	7.82	8.19	7.32	8.27	8.26
> ½ and ≤ 1	14.29	15.28	15.58	15.72	17.31	17.46	17.10	17.20	17.51
> 1 and ≤ 2	16.97	17.49	17.29	19.28	19.63	20.69	20.81	21.39	21.89
> 2 and ≤ 3	9.10	8.92	9.34	7.68	7.94	8.43	9.06	9.15	8.43
> 3 and ≤ 5	7.43	7.00	7.08	7.71	6.89	5.83	6.27	6.57	6.67
> 5 and ≤ 10	5.43	5.13	4.51	4.81	4.45	4.61	4.49	4.07	3.90
> 10 and ≤ 20	2.53	2.19	2.09	2.07	1.74	1.69	1.75	1.63	1.58
> 20	1.22	1.13	1.00	0.79	0.75	0.66	0.66	0.64	0.56

^a Source: calculation by author using data from census and household-sample surveys conducted by the Brazilian census bureau (*Instituto Brasileiro de Geografia e Estatística*, IBGE).

^b BMW: Brazilian minimum wage.

FIGURE 1. Percentage of population ≥ 5 years old who can read and write in four municipalities, according to linear and proposed (nonlinear) interpolation methods, Brazil, 2000–2010

Note: Lines represent percentages of people who can read and write obtained by two methods of interpolation in each municipality, and the corresponding percentages in the states and metropolitan regions (when applicable) where the municipalities are located. Names of municipalities are displayed with two-letter state codes and population size.

infant mortality is reduced, which was also expected, given the omitted variable bias. However, the precision of the estimate also reduces sharply, and the coefficients on FHP in Models 5–7 are no longer significant. This was likely to occur because the linearly interpolated variables are collinear among themselves, and because the expansion of the FHP is correlated with the passage of time.

Models 7–10 repeat the progressive introduction of socioeconomic covariates but use the ones obtained by the proposed interpolation method. As more variables are introduced, the magnitude of the association between the FHP and infant mortality is reduced once more, but at a slower rate. Moreover, the precision of the estimate does not fall by the same amount. The growth in the variance inflation factor (VIF), which is a measure of multi-collinearity, is also slower for Models 7–10 compared to Models 3–6, suggesting that multi-collinearity was mitigated by this approach.

Unlike Model 6, Model 10 shows that the FHP is still significantly associated with lower post-neonatal infant mortality, after controlling for water supply, garbage collection, education, and income.

DISCUSSION

The application of the effects of the FHP on the infant mortality rate demonstrates the usefulness of the proposed interpolation method. The coefficients of the association between the FHP and the infant mortality rate estimated in this study are larger than previous results in the literature (19, 20). It should be noted, however, that the present analysis covers a longer period than the previous studies, and uses more recent data.

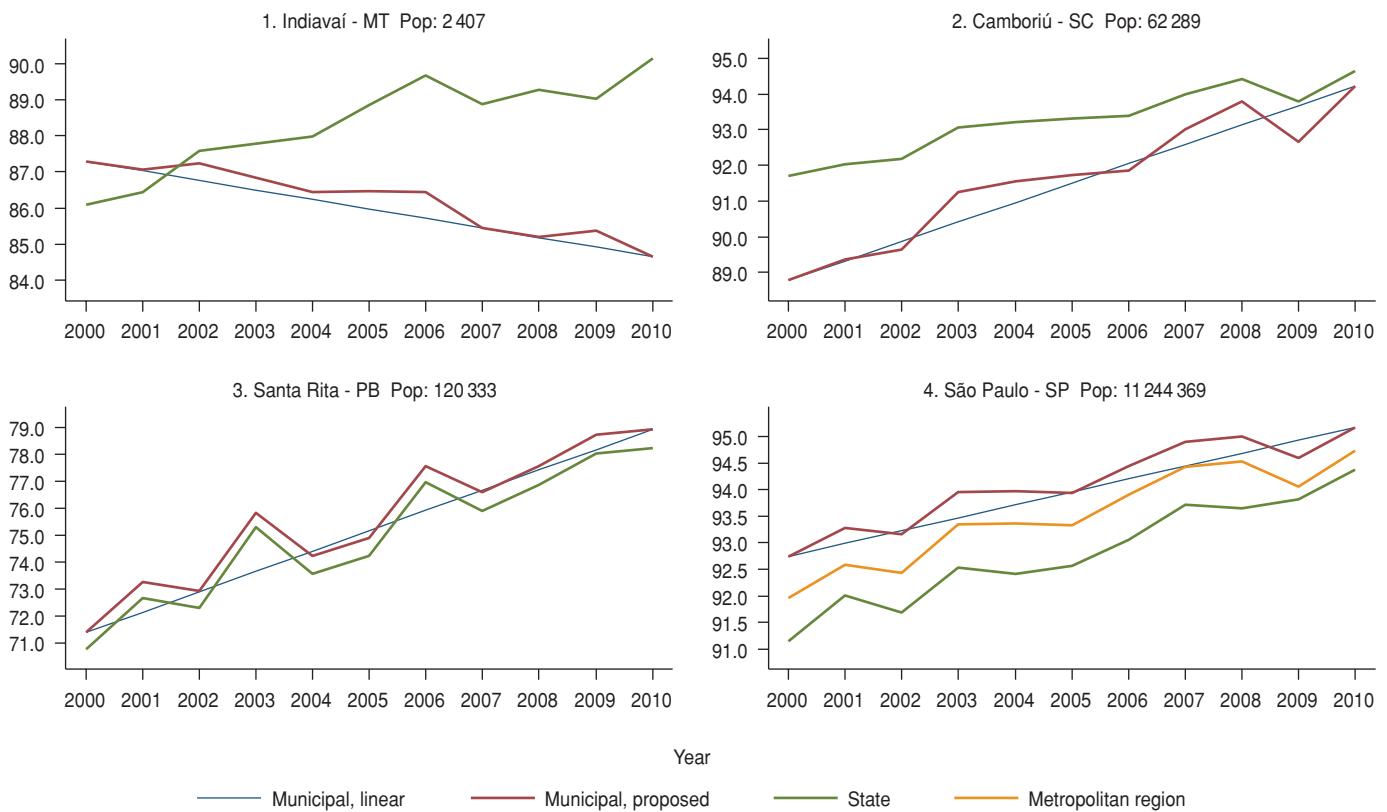
Making socioeconomic covariate data at the municipal level available for inclusion in multivariate analyses strengthens the case for the effectiveness of health programs. This is particularly impor-

tant for the case of Brazil and several other countries in Latin America and the Caribbean, where the expansion of health services in the last decade has occurred simultaneously with other forms of social progress, such as improvements in sanitation infrastructure, educational attainment, and economic development. Disentangling the effects of these different variables is essential for greater financial support for the expansion of health services.

One advantage of the proposed method is its simplicity and potential to be replicated in several countries other than Brazil, as long as equivalent variables are available in both census survey data and household survey data. Furthermore, making the data on Brazilian municipalities directly available should encourage more rigorous studies by Brazilian researchers, policy analysts, and public health students.

Given that progress in socioeconomic variables may be simultaneous, finding

FIGURE 2. Percentage of households with access to clean water supply in four municipalities, according to linear and proposed (nonlinear) interpolation methods, Brazil, 2000–2010



Note: Lines represent percentages of households with access to clean water supply obtained by two methods of interpolation in each municipality, as well as the corresponding percentages in the states and metropolitan regions (when applicable) where the municipalities are located. Names of municipalities are displayed with two-letter state codes and population size.

TABLE 2. Results from 10 fixed-effects regression models using post-neonatal mortality rates as the dependent variable and Family Health Program (FHP) coverage and other covariates as independent variables, according to linear and proposed (nonlinear) interpolation methods, for 5 057 municipalities nationwide (excluding Northern region), Brazil, 2000–2010^a

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
FHP coverage (% of people) (SE ^b)	-4.014 ^c (0.193)	-0.902 ^c (0.233)	-0.505 ^d (0.214)	-0.340 (0.197)	-0.346 (0.199)	-0.299 (0.191)	-0.578 ^e (0.216)	-0.496 ^d (0.209)	-0.496 ^d (0.213)	-0.472 ^d (0.206)
Access to safe water supply (% of households) (SE)			-0.123 ^c (0.015)	-0.104 ^c (0.015)	-0.109 ^c (0.015)	-0.104 ^c (0.015)	-0.089 ^c (0.010)	-0.081 ^c (0.010)	-0.086 ^c (0.010)	-0.086 ^c (0.010)
Garbage collection (% of households)				-0.045 ^c (0.010)	-0.049 ^c (0.009)	-0.045 ^c (0.009)		-0.026 ^c (0.006)	-0.028 ^c (0.006)	-0.027 ^c (0.006)
Less than primary school education (% of people ≥ 10 yrs old)					-0.089 (0.046)	-0.097 ^d (0.042)			-0.068 ^c (0.031)	-0.067 ^d (0.031)
No earnings (% of people ≥ 10 yrs old)						0.092 ^c (0.027)				0.035 ^d (0.014)
Covariate interpolation method			Linear	Linear	Linear	Linear	Proposed	Proposed	Proposed	Proposed
Dummy variables used for year	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dummy variables used for municipality	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	55 334	55 334	54 694	54 025	54 025	54 025	54 692	54 008	54 008	53 999
Number of municipalities included	5 111	5 111	5 107	5 105	5 105	5 105	5 107	5 105	5 105	5 105
Coefficient of determination (R ²) ^f	0.2767	0.2966	0.3042	0.3072	0.3075	0.3080	0.3035	0.3061	0.3064	0.3066
Mean VIF ^g	1.00	1.29	2.47	7.04	7.83	11.42	2.46	7.17	7.76	10.31

^a Models are adjusted for clustering of observations at the municipal level and weighted by population size.

^b SE: standard error.

^c P < 0.001.

^d P < 0.05.

^e P < 0.01.

^f Including municipality effects; variation of the dependent variable explained by the dummy variables representing all but one municipality.

^g VIF: variance inflation factor.

an interpolation strategy that minimizes multi-collinearity is relevant. The presence of collinearity does not reduce the predictive power of the model as a whole, but the individual coefficients estimated for each collinear variable could be unstable when new variables or observations are introduced to the model, especially for VIF values ≥ 10 (26). For a program evaluation application, such as the example given in this study, collinearity among control variables (socioeconomic variables) is not likely to affect the interpretation of the coefficient of the main variable of interest (coverage of FHP). However, because of multi-collinearity among the interpolated variables, it is recommended that they not be used as a study's main variable of interest. For

example, if the interpolated data are used to answer research questions such as the effect of educational attainment on mortality or morbidity, including other interpolated covariates, the predictive power of the model is not affected, but the estimated value of each individual coefficient may be misleading.

Caution should also be used in interpreting the standard errors associated with the coefficients of the interpolated variables. The household sample surveys are complex survey designs, and the computations proposed here further complicate the issue. It should be noted that this problem should not interfere with precision of the estimate of the health services variables of interest, and that the presence of the interpolated

covariate should provide control for socioeconomic aspects included in the multivariate analysis.

Finally, the interpolation is a second-best solution for a situation in which directly observed data are not available. The net gain of the proposed method should be positive, however, as it makes use of additional information to control for important determinants of health status.

Disclaimer. The opinions expressed in the manuscript are the author's and do not necessarily reflect the views of the Inter-American Development Bank, its board of directors, or its technical advisors.

Conflicts of interest. None.

REFERENCES

- Giuffrida A, Gravelle H, Roland M. Measuring quality of care with routine data: avoiding confusion between performance indicators and health outcomes. *BMJ*. 1999;319(7202):94-8.
- Blustein J, Hanson K, Shea S. Preventable hospitalizations and socioeconomic status. *Health Aff (Millwood)*. 1998;17(2):177-89.
- World Health Organization. The World Health Organization on Health inequality, inequity, and social determinants of health. *Pop Dev Rev*. 2007;33(4):839-43.
- Marmot M, Friel S, Bell R, Houweling TA, Taylor S; Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet*. 2008;372(9650):1661-9.
- Paim J, Travassos C, Almeida C, Bahia L, Macinko J. The Brazilian health system: history, advances, and challenges. *Lancet*. 2011;377(9779):1778-97.
- Victora CG, Barreto ML, do Carmo Leal M, Monteiro CA, Schmidt MI, Paim J, et al. Health conditions and health-policy innovations in Brazil: the way forward. *Lancet*. 2011;377(9782):2042-53.
- Houweling TA, Kunst AE. Socio-economic inequalities in childhood mortality in low- and middle-income countries: a review of the international evidence. *Br Med Bull*. 2010;93:7-26.
- Bicego GT, Boerma JT. Maternal education and child survival: a comparative study of survey data from 17 countries. *Soc Sci Med*. 1993;36(9):1207-27.
- Aquino R, de Oliveira NF, Barreto ML. Impact of the family health program on infant mortality in Brazilian municipalities. *Am J Public Health*. 2009;99(1):87-93.
- Rasella D, Aquino R, Santos CA, Paes-Sousa R, Barreto ML. Effect of a conditional cash transfer programme on childhood mortality: a nationwide analysis of Brazilian municipalities. *Lancet*. 2013;382(9886):57-64.
- Bixby LR. Evaluación del impacto de la reforma del sector de la salud en Costa Rica mediante un estudio cuasiexperimental. *Rev Panam Salud Publica*. 2004;15(2):94-103.
- Greene WH. *Econometric analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall; 2003.
- Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.
- Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol*. 2010;171(5):624-32.
- Emwanu T, Hoogeveen JG, Okiira Okwi P. Updating poverty maps with panel data. *World Dev*. 2006;34(12):2076-88.
- Elbers C, Lanjouw JO, Lanjouw P. Micro-level estimation of poverty and inequality. *Econometrica*. 2003;71(1):355-64.
- Brown SPA, Hayes KJ, Taylor LL. State and local policy, factor markets, and regional growth. *Rev Reg Stud*. 2003;33(1):40-60.
- Guanais F, Macinko J. Primary care and avoidable hospitalizations: evidence from Brazil. *J Ambul Care Manage*. 2009;32(2):115-22.
- Macinko J, Marinho de Souza Mde F, Guanais FC, da Silva Simões CC. Going to scale with community-based primary care: an analysis of the family health program and infant mortality in Brazil, 1999-2004. *Soc Sci Med*. 2007;65(10):2070-80.
- Macinko J, Guanais FC, de Fátima M, de Souza M. Evaluation of the impact of the Family Health Program on infant mortality in Brazil, 1990-2002. *J Epidemiol Community Health*. 2006;60(1):13-9.
- Instituto Brasileiro de Geografia e Estatística, Sistema IBGE de Recuperação Automática (BR). Banco de dados agregados. Censo demográfico 2000 [Internet]. Available from: <http://www.sidra.ibge.gov.br/cd/default/cd2000.asp?o=15&i=P> Accessed 7 July 2012.
- Instituto Brasileiro de Geografia e Estatística, Sistema IBGE de Recuperação Automática (BR). Banco de dados agregados. Censo demográfico 2010 [Internet]. Available from: <http://www.sidra.ibge.gov.br/cd/default/cd2010.asp?o=4&i=P> Accessed 5 August 2012.
- Instituto Brasileiro de Geografia e Estatística, Sistema IBGE de Recuperação Automática (BR). Banco de dados agregados. Pesquisa Nacional por Amostra de Domicílios 2001 a 2011 [Internet]. Available from: <http://www.sidra.ibge.gov.br/pnad/default.asp> Accessed 9 June 2012.
- Harris M, Haines A. Brazil's Family Health Programme. *BMJ*. 2010;341:c4945.
- StataCorp. *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP; 2011.
- Kennedy P. *A guide to econometrics*. Cambridge, MA: MIT Press; 2003.

Manuscript received on 30 August 2012. Revised version accepted for publication on 26 August 2013.

RESUMEN**Covariables a escala municipal del estado de salud en el Brasil: propuesta de un método para la interpolación de datos**

Objetivo. Proponer un método para la interpolación de las covariables locales anuales del estado de salud que sea apropiado para el análisis de datos longitudinales del efecto de los servicios de salud.

Métodos. El método propuesto, a partir de los datos de las encuestas llevadas a cabo en los hogares, distribuye la tasa anual de crecimiento de las covariables a escala regional (por ejemplo, un estado) y la utiliza para interpolar los datos anuales a escala local (por ejemplo, un municipio) entre dos encuestas censales consecutivas. El método se aplicó a las covariables socioeconómicas a escala municipal del estado de salud en el Brasil para cada año de los comprendidos entre el 2001 y el 2009. Los datos se sometieron a prueba mediante un análisis previamente validado de los efectos del Programa de Salud Familiar sobre la mortalidad posneonatal en lactantes del Brasil.

Resultados. Se generaron un total de 895 628 valores correspondientes a 20 factores predictivos socioeconómicos del estado de salud. Se obtuvieron datos válidos de 5 057 municipios de las regiones del nordeste, sudeste, sur y centro-oeste del Brasil, del 2001 al 2009, que comprendían el 98,89% de los municipios de estas regiones y el 90,87% de los municipios del país. Un anexo suplementario incluye los datos interpolados del 2001 al 2009, y los datos de los censos del 2000 y del 2010 correspondientes a los 5 057 municipios. La aplicación de un modelo de regresión de efectos fijos indicó que, en comparación con la interpolación lineal, el método propuesto redujo la multicolinealidad y mejoró la precisión de los cálculos de los efectos de los servicios de salud.

Conclusiones. Las ventajas del método de interpolación propuesto indican que constituye una solución factible para el análisis de datos longitudinales de los servicios de salud a escala local en el Brasil y en otros países.

Palabras clave

Factores socioeconómicos; salud de la familia; mortalidad infantil; modelos estadísticos; Brasil.