44
DISCUSSION PAPER
Originally published by Ipea in July 1993 as
number 307 of the series Texto para Discussão.

UNEQUAL OPPORTUNITY TO SURVIVE, EDUCATION AND REGIONAL DISPARITIES IN BRAZIL

Ricardo Barros
Diana Sawyer

44

DISCUSSION PAPER

Originally published by Ipea in July 1993 as
number 307 of the series Texto para Discussão.

Brasília, January 2015

UNEQUAL OPPORTUNITY TO SURVIVE, EDUCATION AND REGIONAL DISPARITIES IN BRAZIL

Ricardo Barros
Diana Sawyer

DISCUSSION PAPER

A publication to disseminate the findings of research directly or indirectly conducted by the Institute for Applied Economic Research (Ipea). Due to their relevance, they provide information to specialists and encourage contributions.

Discussion paper / Institute for Applied Economic

Research.- Brasília : Rio de Janeiro : Ipea, 1990-

ISSN 1415-4765

1. Brazil. 2. Economic Aspects. 3. Social Aspects. I. Institute for Applied Economic Research.

CDD 330.908

SUMMARY

UNEQUAL OPPORTUNITY TO SURVIVE, EDUCATION AND REGIONAL DISPARITIES IN BRAZIL*

Ricardo Barros**
Diana Sawyer***

* This study received financial support from the DHS Small Grant Program. The study was developed while Diana Oya Sawyer was a Visiting Researcher in the Department of Epidemiology and Public Health at Yale University. We would like to thank Dedobrah Levison and Rosane Mendonça for very useful comments on several versions of this study .

** Department of Economics,Yale University and Instituto de Pesquisa Econômica Aplicada - IPEA, Brasil.

*** Center for Regional Development and Planning-Cedeplar at Federal University of Minas Gerais-UFMG.

This paper was financed by project GESEP (LOAN BIRD 2347 -BR) and elaborated with the support of project UNDP/IPEA BRA 93/011.

1. INTRODUCTION

The two most striking features of the Brazilian social-economic development are: the slow pace of the social development when compared to the Brazilian rapid process of economic growth and industrialization, and the existence of extremely large and temporally stable regional differences in virtually all social indicators: these differences in social indicators being much larger than the corresponding differences in most economic indicators [Albuquerque and Villela (1991)].

Among all indicators of social development, the infant mortality rate (IMR) is perhaps the single most important.[1] Societies, however, are not made of a homogeneous group of families. In general, the opportunity to survive of children is strongly influenced by their family background. The stronger is the association between the risk of infant mortality and family background, the higher will be the degree of inequality of opportunity prevailing in the society and consequently the deeper will be the social concern. As a result, attention must be given to both: **a)** the national level of the IMR; and **b)** the extent to which the IMR varies with family background among other dimensions.

As Table 1 reveals, Brazil is among the Latin-American countries with the largest IMR. Brazil ranks sixth in decreasing order of IMR being between the Dominican Republic and El Salvador.

This study investigates the extent to which this large IMR varies with and so can be explained by the level of education of parents and their region of residence.[2] More specifically, the study has three main objectives. First, to determine and contrast the relative contribution to the degree of inequality in survival opportunities of two dimensions of family background: **a)** parents' education; and **b)** parents' region of residence. We want to consider questions like: Are the risk of infant mortality more sensitive to parent's education or to region of residence?

[1] Infant mortality rate is defined as the probability that a newborn will die before completing 12 months.

[2] Related studies are: Beltrão and Sawyer (1991), Henriques, Strauss and Thomas (1989), McCraken (1990), McCraken and Rodrigues (1991) and IBGE (1989).

Table 1

Infant Mortality Rate for Latin-American Countries, 1990 (o/oo)

Country	Infant Mortality Rate (IMR)
Argentina	31
Bolivia	102
Brazil	60
Chile	20
Colombia	39
Costa Rica	18
Cuba	11
Dominican Republic	61
El Salvador	59
Guatemala	54
Haiti	92
Honduras	63
Mexico	40
Nicaragua	56
Panama	22
Paraguay	41
Peru	82
Uruguay	22
Venezuela	35

Source: United Nations (1992).

Secondly, we aim to estimate the impact on the aggregated level of the IMR of different types of educational expansion and regional reallocation of the population. In particular, we investigate the curvature of the relationship between the risk of infant mortality and parents' education. The more convex this relationship is, the large will be the reduction in infant mortality which would follow from an education expansion reducing the dispersion in the distribution of education. As a matter of fact, if this relationship were linear, inequality in education would have no impact whatsoever on the level of infant mortality.

Thirdly, based on the assumption that the region of residence of parents is partially responsible for their education level, we decompose the variability in the IMR between regions into two components: **a)** the direct impact of region of residence; and **b)** the impact of region which operates through its effect on parents' education.

Precise answers to these three questions rely on the estimation of the causal relationship between the aggregated IMR, on the one hand, and parents' education and region of residence, on the other hand. Based on non-experimental data and in the absence of a comprehensive theoretical model for the determination of the survival probability, the estimation of this relationship can not be done without strong untested assumptions.

The assumptions underlying the analysis conducted in this study are not new, but they have often been left implicit in most of the studies in the literature. In this study we explicit these assumptions in great detail. However, since the analysis of these assumptions is lengthy and requires a great deal of notation, we decided to present the empirical findings first and discuss their meaning at a intuitive level. Later on, we formally describe the assumption implicit in the analysis.

The study is organized in 11 sections. The next section describes the data set and how the variables were constructed. Sections 3 to 7 presents the empirical findings of the study. In Section 8 we introduce the framework required to give a causal interpretation to our empirical findings. In Section 9 we introduce the orthogonality assumptions we use to identify the causal effect of region of residence and mother's education on the risk of infant mortality from the information available in our data set. Section 10, based on the framework and assumptions introduced in the previous section, carefully defines what is meant by the impact of region of residence and mother's education on the risk of infant mortality and how they relate to the empirical analysis conducted in Sections 3 to 8. Finally, in Section 11 the main results of the study are summarized.

2. EMPIRICAL PRELIMINARIES

Data Source: This study is based on the PNSMIPF-Pesquisa Nacional sobre Saúde Materno-Infantil e Planejamento Familiar. PNSMIPF is a household survey conducted by Bemfam and IRD as part of the DHS research program. This survey covers 95% of the Brazilian population[3] and was conducted from May to August of 1986. The main goal of the survey was to collect

[3]The survey do not cover the rural areas of the following states: Amazonas, Pará, Goiás, Tocantins, Mato Grosso do Sul, and Mato Grosso. Also the survey do not cover either the urban or the rural areas of the states of Rondônia, Acre, Roraima, and Amapá. All excluded areas, as a whole, represent less than 5% of the Brazilian population according to the 1980 Demographic Census [Arruda et alii (1987)].

information on reproductive behavior. Nonetheless, the survey also collected information on mortality. Of fundamental importance for this study is the availability for each women in the survey of retrospective information about the day of birth of every offspring born before the day of the interview and also the day of death of all offspring who have die before the day of the interview.

The survey is based on a random sample of households. A self-weighted, regionally stratified, and multistage sample design has been used.[4] Using a sampling rate varying across regions from 1/1,700 to 1/6,400 near 9,000 households have been sampled. In each sampled household all women aged 15 to 44 have been interviewed. The survey ended up interviewing close to 6,000 women [see Arruda et alii (1987, Appendix A)].

Unit and Universe of Analysis: In this study children will be the unit of analysis. To avoid the confounding effects of urban-rural differentials, age of the mother, and birth cohort, the universe of analysis was constrained along three lines: a) the child's mother had to be living in a urban household at the time of the interview; b) the child's mother had be between 20 and 34 years old when the child was born; and c) the child had to be born between January of 1979 and December of 1982. This procedure generates a sample of 1,547 children born from 1,122 mothers.[5] When expanded this generates an universe of approximately 15 millions children. In segments of the study we use as the universe of analysis a subset of the overall universe obtained by eliminating all children whose mother's finished at least a year of college. This subuniverse comprises 1,032 mothers.

Defining the Variables: To study infant mortality we construct, for each child in the sample, an indicator, **D**, which equals to one when the child died before completing 12 months and equals to zero otherwise. Let **R** denote region of residence. Following the sample stratification, we divide Brazil into six regions: The

[4]The sample was construted in four stages.

5 Distribution of Mothers by the Number of Offspring

Number of Offspring/Mother	Number of Mothers
01	769
02	290
03	54
04	9

Under certain assumptions, these regional disparities measure the total causal effect of region of residence on children's survival probabilities. The precise meaning of this total causal effect as a counter-factual experiment and the statement of the required assumptions are presented in Sections 9 and 10. Therefore, to the extent that these assumptions are valid, Table 2 reveals that regional disparities in Brazil lead to a considerable degree of inequality in survival opportunities. To summarize this degree of inequality we estimate the degree regional dissimilarity in IMR, D. The index of dissimilarity used is defined as the smallest fraction of deaths which needs to be redistributed among regions to ensure an equal IMR in all regions. The index can be obtained via:

$$D \equiv \frac{1}{2 \cdot P[D=1]} \cdot \sum_r P[R=r] \cdot \left| P[D=1|R=r] - P[D=1] \right| .$$

Based on the information contained in Table 2 a degree of dissimilarity of 30%, was estimated (see Table 3). Therefore, at least 30% of all deaths need to be regionally reallocated for the IMR to reach the same level in all regions. Table 2 identifies the Northeast Region as the main responsible for this large degree of regional dissimilarity. To confirm this conjecture we recompute the index of dissimilarity six times, excluding at each time, one and only one region.[6] The results in Table 3 clearly indicate the Northeast as responsible for more than 50% of all regional dissimilarity in mortality. As a matter of fact, when the Northeast is excluded the index of dissimilarity drops from 30 to 13%. In all other cases, the exclusion of each region never led to a variation in the index of dissimilarity superior to 4%.

Table 3
Index of Dissimilarity

Region	Index of Dissimilarity (%)	Variation
RJ Excluded	29	-1
SP Excluded	34	+4
SO Excluded	28	-2
CE Excluded	31	+1
NE Excluded	13	-17
FR Excluded	30	0
All Brazil	30	-

[6] $D(i) \equiv \frac{1}{2 \cdot P[D=1|R \neq i]} \cdot \sum_{r \neq i} P[R=r|R \neq i] \cdot \left| P[D=1|R=r] - P[D=1|R \neq i] \right|$

States of Rio de Janeiro (RJ) and São Paulo (SP), the South (SO), Central-East (CE) and Northeast (N) Regions, and the Frontier (FR) area. For a precise definition of these areas see Arruda et alii (1987, p.8). To measure family education, E, we use mother's number of years of schooling.

An important warning about the use of these variables constructed to represent region of residence and mother's education is that both refer to the time of the interview not to the time of the birth of the child. Moreover, it should be emphasized that all information available on children is condition on their mother being alive at the date of the interview. The impact of these two facts on our estimates remain to be further investigated.

3. UNEQUAL OPPORTUNITY TO SURVIVE AND REGIONAL DISPARITIES IN BRAZIL

Table 2 presents estimates of the distribution of children by region, $\{P[R=r]:r \in R\}$, and the IMR for each region, $\{P[D=1|R=r]:r \in R\}$. $R=\{RJ,SP,SO,CE,NE,FR\}$.

This table reveals dramatic regional disparities in IMR. Most striking is the level of the IMR in the Northeast Region. In this region the IMR is from 2.5 to 6 times as large as in any other region! The State of São Paulo and the Central-East Region are areas with moderate levels of infant mortality. Finally, the Frontier followed by the State of Rio de Janeiro and the South Region are, in this order, the areas with the smallest levels of infant mortality.

Table 2

Distribution of Children According to Their Region of Residence and Infant Mortality Rate by Region of Residence

Region (R)	Distribution of Children (%)	Infant Mortality (0/00)	
		All Mothers	Non-College Educated
Rio de Janeiro (RJ)	12(0.8)	28(11)	31
São Paulo (SP)	23(1.1)	50(15)	50
South (SO)	13(0.8)	21(12)	16
Central-East (CE)	15(0.9)	40(12)	37
Northeast (NE)	29(1.2)	129(15)	128
Frontier (FR)	8(0.7)	31(11)	33
Urban Brazil	100	64(6)	64

Note: Numbers in parentheses are the standard error of the estimates.

4. THE IMPACT OF POPULATION MOVEMENTS ON THE LEVEL OF THE IMR AND ON DEGREE OF REGIONAL INEQUALITY IN IMR

The overall IMR, P[D=1] and the degree of regional dissimilarity in IMR, D, can both be obtained from two elements: a) The regionally specific mortality rates, $\{P[D=1|R=r]:r\in\mathcal{R}\}$; and b) the regional distribution of the population, $\{P[R=r]:r\in\mathcal{R}\}$. In fact they can be written as

$$P[D=1] = \sum_r P[R=r]\cdot P[D=1|R=r]$$

and

$$D \equiv \frac{1}{2\cdot P[D=1]}\cdot\sum_r P[R=r]\cdot\left| P[D=1|R=r]-P[D=1]\right|.$$

In this section we explore these two identities. Based on them, we estimate the impact on the overall IMR, P[D=1], and on the degree of dissimilarity, D, of marginal changes in the regional distribution of the population, $\{P[R=r]:r\in\mathcal{R}\}$, holding constant all regionally specific IMR, $\{P[D=1|R=r]:r\in\mathcal{R}\}$.

The marginal effect of having less children in region r^1 and more in region r^2 on the level of the IMR, $\Delta M(r^1 \to r^2)$, and on the degree of regional inequality in IMR, $\Delta D(r^1 \to r^2)$, are given respectively by

$$\Delta M(r^1 \to r^2) = P[D=1|R=r^2] - P[D=1|R=r^1)]$$

and

$$\Delta D(r^1 \to r^2) = \begin{cases} \frac{D+P[R=NE]}{P[D=1]}\cdot\left(P[D=1|R=r^1]-P[D=1|R=r^2]\right) & \text{if } r^2\neq NE \\ \frac{D+P[R=NE]}{P[D=1]}\cdot\left(P[D=1|R=r^1]-P[D=1|R=NE]\right)+\frac{P[D=1|R=NE]}{P[D=1]}-1 & \text{if } r^2=NE \end{cases}$$

Tables 4 and 5 present estimates of these effects. Since the risk of infant mortality is much larger in the Northeast than in any other region, reductions in the proportion of children in the Northeast have the greater impact in reducing the level of the IMR.

The State of São Paulo and the Central-East Region are areas with intermediate levels of IMR. As a consequence, increments in the proportion of the population in these two areas, particularly in the State of São Paulo, are the type of change which generates the greater reductions in inequality in infant mortality.

Finally, since the South Regional followed by the State of Rio de Janeiro are the areas with the smallest levels of IMR, increments in the proportion of the population in these regions reduce the IMR. Moreover, such increments also tend to increase the degree of regional inequality in infant mortality, except when these increments are accompanied by a corresponding decrease in the proportion of the population in the Northeast region.

Table 4
The Impact of Population Movements on the Level of the IMR
$\Delta M(r^1 \rightarrow r^2)$

	Region (r^2)				
Region (r^1)	São Paulo	South	Central-East	Northeast	Frontier
Rio de Janeiro	22	-7	12	101	3
	(19)	(17)	(17)	(19)	(16)
São Paulo	-	-29	-10	79	-19
		(20)	(20)	(22)	(19)
South		-	18	108	10
			(17)	(20)	(16)
Central-East			-	90	-8
				(20)	(16)
Northeast				-	-98
					(19)

Note: Numbers in parentheses are the standard error of the estimates.

Table 5
The Impact of Population Movements on the Index of Dissimilarity
$\Delta D(r^1 \rightarrow r^2)$

	Region (r^2)				
Region (r^1)	São Paulo	South	Central-East	Northeast	Frontier
Rio de Janeiro	-20	6	-11	9	-3
	(18)	(15)	(16)	(11)	(15)
São Paulo	-	26	10	29	17
		(18)	(18)	(14)	(18)
South		-	-17	3	-9
			(16)	(12)	(15)
Central-East			-	20	8
				(11)	(15)
Northeast				-	-12
					(10)

Note: Numbers in parentheses are the standard error of the estimates.

5. THE IMPACT OF REGION OF RESIDENCE ON MOTHER'S EDUCATION

To the extent that mother's education affects child survival and that region of residence affects mother's education, at least part of the impact of region on infant mortality operates indirectly via the impact of region on mother's education.

In this section we examine how mother's distribution of education varies according to region of residence. In the next section, we investigate the relationship between mother's education and infant mortality. In Section 7, we then combine these two relationships to identified the faction of the total impact of region of residence on IMR which is channeled trough its impact on education.

Table 6 presents, by region of residence, estimates of: **a)** the distribution of children according to their mother's education, $\{P(e_1 \leq E \leq e_2 | R=r) : r \in \mathbb{R}\}$, and **b)** mother's average number of years of schooling, $\{E[E | R=r] : r \in \mathbb{R}\}$.

Under a set of assumptions which will be described in Section 9, these regional disparities measure the causal effect of region of residence on mother's education. The precise meaning of this causal effect as a counter-factual experiment are presented in Section 10.

Assuming the assumptions are valid, this table reveals two facts: on the one hand, the table reveals some regional disparities with the Northeast being the region with the worse opportunities for education. On the other hand, the table reveals that regional disparities in education are considerably smaller than regional disparities in IMR. As a matter of fact, all indices of dissimilarity for regional disparities in the distribution of education which are presented in table 6, indicates that regional disparities in infant mortality (D=30) are from two to three times larger than regional disparities in education. In summary, regional disparities in Brazil seem to generate more inequality in survival opportunities than in educational opportunities.

Table 6
Distribution of Children According to their Mothers Education and Mothers Average Years of Schooling by Region of Residence

Region ($\underline{R}$)	Percent with				Average
	No Schooling	Lower Primary	Upper Primary	High School	
Rio de Janeiro	3	41	42	14	5.6
São Paulo	9	57	23	11	4.6
South	12	37	34	18	5.2
Central-East	10	48	22	19	5.2
Northeast	15	53	19	13	4.0
Frontier	18	45	25	13	4.3
Dissimilarity	15	6	11	9	-

6. THE IMPACT OF MOTHER'S EDUCATION ON INFANT MORTALITY

In this section we investigate the impact of mother's education on the risk of infant mortality. Since as shown in Table 6, mothers tend to be less educated in the Northeast where the IMR is greater, the direct relationship between IMR and the level of mother's education would probably overestimate the correct impact of mother's education on infant mortality. In an attempt to estimate the effect of mother's education without this confounding effect of region of residence, we compute the IMR by level of mother's education which would be observed if the mother's distribution of education were the same in all regions, i.e., $P[\mathbf{E}=e|\mathbf{R}=r]=P[\mathbf{E}=e]$. This counter-factual IMR by educational level is denoted by $P^*[D=1|E=e]$ and can be obtained via

$$P^*[D=1|E=e] = \sum_r P[D=1|R=r,E=e]\cdot P[R=r].$$

Notice the contrast between this counter-factual IMR and the IMR by mother's educational level actually observed which is given by

$$P[\mathbf{D}=1|E=e] = \sum_r P[\mathbf{D}=1|\mathbf{R}=r,\mathbf{E}=e]\cdot P[\mathbf{R}=r|\mathbf{E}=e].$$

Table 7 presents estimates of $P^*[D=1|E=e]$ when education is measured by the number of years of schooling. Under the set of assumption which will be described in Section 9, these differentials in mortality rate by educational level measure the causal effect of mother's education on children's survival probabilities, holding regional of residence constant. The precise meaning of this conditional causal effect as a counter-factual experiment will be presented in Section 10.

Assuming the assumptions are valid, this table reveals two important facts. First, it reveals that despite mother's education having a strong impact on the probability of survival of their children, this impact is, however, much weaker than that associated to region of residence. In fact, as demonstrated in Section 6, reductions in the proportion of the population in the Northeast will generate reductions in infant mortality of approximately 100 per thousand (see Table 3). Table 6, on the other hand, reveals that even an increment in the proportion of mothers with 10-11 years of schooling with a concomitant decrease in the proportion with 0-1 year of schooling would reduce the mortality in approximately 60 per thousand.[7]

Table 7
Infant Mortality according to Mothers' Schooling Regionally Standardized

Mothers' Schooling Category	Infant Mortality
0-1 year	88
2-3 years	82
4-5 years	63
6-7 years	60
8-9 years	35
10-11 years	25
Dissimilarity	13

Secondly, this table reveals that the relationship between infant mortality and mothers'education is almost linear and certainly not convex, as one could expect. As a result, a change in the distribution of education which keeps the mean constant but reduces the variance would not reduce the overall IMR. To illustrate the lack of convexity we conduct the following change in the distribution of mothers' education:

[7] 88-25=63.

$$P^{*}[0 \le E \le 1] = P[0 \le E \le 1] - 0.01$$

$$P^{*}[2 \le E \le 4] = P[2 \le E \le 4] + 0.01$$

$$P^{*}[8 \le E \le 9] = P[8 \le E \le 9] - 0.01$$

$$P^{*}[10 \le E \le 11] = P[10 \le E \le 11] + 0.01$$

After this transformation, the total number of schooling years would remain constant while the inequality in schooling would diminish. The consequence of this change on the overall IMR is surprising. Accordingly to the estimates in Table 7 the IMR will increase by 0.04![8]

7. THE DIRECT AND INDIRECT CONTRIBUTIONS OF REGION OF RESIDENCE ON THE DEGREE OF UNEQUAL OPPORTUNITY TO SURVIVE

Part of the regional variability in IMR is due to regional differences in the mothers' distribution of education. To isolate the direct contribution of region, we estimate for each region what the IMR would be if the distribution of education of mothers in all regions were the same. In other words, we compute

$$P^{**}[D=1|R=r] = \sum_{e} P[D=1|R=r,E=e] \cdot P[E=e]$$

and contrast with the actually observed regionally specific mortality rates

$$P[D=1|R=r] = \sum_{e} P[D=1|R=r,E=e] \cdot P[E=e|R=r].$$

Table 8 presents the unstandardized and the standardized IMR and their respective dissimilarity indices. The fact that all standardized and unstandardized IMR are very similar indicates that regional differences in education explain a negligible fraction of the regional disparities in IMR. In fact, the standardization for the distribution of education reduces the index of dissimilarity only from 31 to 28%.

[8] ((-25+35)+(-88+82))x0.01=0.04.

Table 8
Standardized and Unstandardized IMR by Region of Residence

Region (R)	Unstandardized	Standardized
Rio de Janeiro	31	44
São Paulo	50	47
South	16	18
Central-East	37	36
Northeast	128	120
Frontier	33	32
Dissimilarity	31	28

8. A FRAMEWORK FOR CAUSAL INTERPRETATION

In the previous sections we investigate the empirical association among infant mortality, region of residence and mothers' education. To give a causal interpretation to these empirical associations requires that certain orthogonality conditions must be satisfied. To describe these orthogonality conditions and the precise meaning of the causal interpretation they imply, the following framework is required:

Agents: There exists two agents: children and families. Let $\mathbb{C}$ denote the set of all children and $\mathbb{F}$ the set of all families. Let c denote a child in $\mathbb{C}$ and f a family in $\mathbb{F}$.

Children's Attributes: At the time of birth each child is endowed with a biological health status and is assigned to a family. Hence, let $B:\mathbb{C}\to(-\infty,\infty)$ be a function defining the biological health status of each child, and $F:\mathbb{C}\to\mathbb{F}$ be a function which assigns children to families.

Family Environment: The family environment is made of three components: a health environment, a region of residence, and a education level. Let $\mathbb{R}$ denote the set of possible regions of residence, and $\mathbb{E}$ the set of possible levels for family education. Moreover, let $H^*:\mathbb{F}\to(-\infty,\infty)$, $R^*:\mathbb{F}\to\mathbb{R}$, and $E^*:\mathbb{F}\to\mathbb{E}$, be functions which assing to each family a health environment, a region of residence, and a level of education, respectively.

Transfering Characteristics from families to children: Since each child is associated to one family, each child is also uniquely associated to a family environment. For example, $E^*(F(c))$ is the education attainment of the family to which children c has been assigned. To simplify the notation we define $H(c)\equiv H^{*}\circ F(c)$, $R(c)\equiv R^{*}\circ F(c)$, and $E(c)\equiv E^{*}\circ F(c)$. As a general rule we eliminate the *-superscript when denoting families characteristics assigned to children.

Child Mortality: Let $D:\mathbb{C}\rightarrow\{0,1\}$ be an indicator of whether or not the child life is shorter than 12 months. We assume that a child, $c\in\mathbb{C}$, will live less than 12 months, $D(c)=1$, if and only if the sum of his/her biological health status, $B(c)$, with his/her family health environment, $H(c)$, is negative, i.e., for all $c\in\mathbb{C}$, $D(c)=1$, if and only if $B(c) + H(c) < 0$.

Response Functions: The framework underlying this study is constructed to investigate a sequential causal chain: Region of residence is permitted to influence family health environment and family education. By its turn, family education is permitted to influence family health environment but not region of residence.

Accordingly, each family $f\in\mathbb{F}$ is endowed with two response functions: $\mathcal{H}(f):\mathbb{R}\times\mathbb{E}\rightarrow(-\infty,\infty)$, and $\mathcal{E}(f):\mathbb{R}\rightarrow\mathbb{E}$. For each $f\in\mathbb{F}$ and $(r,e)\in\mathbb{R}\times\mathbb{E}$, $\mathcal{H}(f;r,e)$ determines the health environment of family f if this family were to live in region r and to attain education level e. Moreover, for each $f\in\mathbb{F}$ and $r\in\mathbb{R}$, $\mathcal{E}(f;r)$ determines the education level of family f if this family were to live in region r.

Shifters and the Separability Assumptions: To simplify the analysis we introduce two shifters: **a)** $S_h^{*}:\mathbb{F}\rightarrow(-\infty,\infty)$, the family health shifter; and **b)** $S_e^{*}:\mathbb{F}\rightarrow(-\infty,\infty)$ the education shifter. These shifters are used to summarize the family effect on its health environment and education level, in the sense that

$$\mathcal{H}(f,r,e) = h(r,e) + S_h^{*}(f), \tag{S1}$$

$$\mathcal{E}(f,r) = e(r,S_e^{*}(f)). \tag{S2}$$

Except for the separability assumption in (S1), there is no essential loss of generality since we could always choose $S_h^{*}(f)=S_e^{*}(f)=f$.

Same Important Implications: It follows immediately from the procedure used to construct the shifters and from the separability assumptions S1-S2 that

$$H^{*} = h(R^{*},E^{*}) + S_h^{*},$$

$$E^{*} = e(R^{*},S_e^{*}).$$

Moreover and more importantly, it follows that

$$H = h(R,E) + S_h,$$

$$E = e(R,S_e)$$

and

$$D=1 \quad \text{if and only if } B+H<0.$$

These three equations form a recursive system. The construction of this system was the main objective of this section.

9. IDENTIFICATION CONDITIONS

In this section we introduce the orthogonality assumptions used to identify the causal effect of region of residence and mother's education on the risk of infant mortality from the information available in our data set.

Observable Entities: The fundamental observable entity in this study consists of the joint-distribution of (D,R,E) and consequently all marginal and conditional distributions which can be derived from it. Hence, in particular and of importance to us, the conditional probabilities $P[D=1|R=r,E=e]$, $P[D=1|R=r,$ and $P[E=e|R=r]$ are part of the set of observable entities.

Identification Concept: Our goal is to encounter conditions under which the parameters which we want to estimate - the result of the five counter-factuals to be described in the next section - could be expressed as functions of observable entities.

Same important Relationships: Note that

$$P[D=1|R=r,E=e] = P[B+S_h<-h(r,e)|R=r,E=e] = G(-h(r,e,w)|R=r,E=e)$$

where, as before, G is the cumulative distribution function of $B+S_h$. Moreover,

$$P[D=1|R=r] = E[G(-h(r,e(r,S_e))|S_e,R=r)|R=r]$$

and

$$P[E=e|R=r] = P[e(r,S_e)=e|R=r]$$

Identification Hypothesis: Identification is obtained by a suitable set of orthogonality conditions involving the biological endowement, B, and the shifters (S_h,S_e), on the one hand, and the family characteristics **(R,E)** on the other hand. The identifying assumptions are:

(A1) $(B, S_h, S_e) \perp\!\!\!\perp R$

(A2) $(B, S_h) \perp\!\!\!\perp (R, E)$

Remark 1: Notice that **(A1)** and **(A2)** are satisfied if and only if

(A3) $(B, S_h) \perp\!\!\!\perp S_e \perp\!\!\!\perp R$

10. PARAMETERS OF INTEREST

In this section, five counter-factual experiments will be introduce. Moreover, the connection between these counter-factuals and the empirical analysis in Section 3 to 7 is established assuming that assumptions A1 and A2 hold.

Counter-Factual 1: The Dependence of Infant Mortality on Region of Residence, Ψ. Let a child be chosen from $\mathbb{C}$ and assigned to his/her family as is currently done, i.e., let child c be assigned to family $\mathbf{F}(c)$. Let us, however, interfere on the region of residence of his/her family by moving (if necessary) his/her family to region r, i.e., we change his/her family region of residence from $R(c)$ to r. In this counter factual experiment the child's family education level is permitted to adjust to the (possible) change in the region of residence. For child c the new family education will be given by $\mathcal{E}(\mathbf{F}(c),r)$.

Notice that $\mathcal{E}(\mathbf{F}(c),r)$ can be quite distinct from the original family education of child c, $\mathbf{E}(c)$, but it has not been pre-specified by the experiment. The final value will depend on which child has been chosen. Under this experiment, if child c is chosen he/she will survive at least 12 months if and only if

$$B(c) + \mathcal{H}(\mathbf{F}(c), r, \mathcal{E}(\mathbf{F}(c), r)) > 0.$$

Let $\Psi(r)$ denote the probability that a randomly chosen child submitted to this experiment will die before exact age 1. This probability is given by

$$\begin{aligned}\Psi(r) &= P[\mathbf{B} + \mathcal{H}(\mathbf{F}, r, \mathcal{E}(\mathbf{F}, r)) < 0] \\ &= P[\mathbf{B} + S_h + h(r, e(r, S_e)) < 0] \\ &= E[G(-h(r, e(r, S_e)) \mid S_e)].\end{aligned}$$

Notice that $\Psi(r)$ is determined by the functions h and e and by the joint-distribution of $(\mathbf{B}, \mathbf{S}_h, \mathbf{S}_e)$. Ψ=

$\{(r,\Psi(r)):r\in\mathbb{R}\}$ is our measure for the dependence of the risk of infant mortality on region of residence. Finally, notice that if A1 and A2 are satisfied then $\Psi(r)=P[\mathbf{D}=1|\mathbf{R}=r]_r]$. This implies that under the hypothesis that A1 and A2 are valid, estimates of $\Psi=\{(r,\Psi(r)):r\in\mathbb{R}\}$ can be found in Table 2.

Counter-Factual 2: The Dependence of Family Education on Region of Residence, λ. Let a child be chosen from $\mathbb{C}$ and assigned to his/her family as is currently done, i.e., let child c be assigned to family $\mathbf{F}(c)$. Let us, however, interfere on the region of residence of his/her family by moving (if necessary) his/her family to region r. As in the previous counter-factual, the education level of the family will adjust to the (possible) change in the region of residence. For child c the new family education will be given by $\mathcal{E}(\mathbf{F}(c),r)$.

Let the probability that the family education of a randomly chosen child submitted to this counter-factual experiment equals to e be denoted by $\lambda(r,e)$. This probability is given by

$$\lambda(r,e) = P[e,(r,\mathbf{S}_e)=e]$$

Notice that $\lambda(r,e)$ is determined by e and the marginal distribution of $\mathbf{S}_e$. $\lambda=\{(r,e,\lambda(r,e)):r\in\mathbb{R}$ and $e\in\mathbb{E}\}$ is our measure for the dependence of the distribution of education on the region of residence.

Finally, notice that if A1 and A2 are satisfied then $\lambda(r,e)=P[\mathbf{E}=e|\mathbf{R}=r]$. Hence, under the hypothesis that A1 and A2 are valid, estimates of $\lambda=\{(r,e,\lambda(r,e)):r\in\mathbb{R}$ and $e\in\mathbb{E}\}$ can be found in Table 6.

Counter-Factual 3: The Dependence of Infant Mortality on Family Education and Region of Residence, Λ. Let a child be chosen from $\mathbb{C}$ and assigned to his/her family as is currently done, i.e., let child c be assigned to family $\mathbf{F}(c)$. Let us, however, interfere on the region of residence of his/her family and on their education level. More specifically, we move (if necessary) his/her family to region r and get it (dis)educated up to level e. i.e., for child c we change $\mathbf{R}(c)$ for r and $\mathbf{E}(c)$ for e. According to our framework, this child, c, will now, if submitted to this experiment, survive at least 12 months if and only if $\mathbf{B}(c)+\mathcal{H}(\mathbf{F}(c),r,e) > 0$. Let $\Lambda(r,e)$ denoted the probability that a randomly chosen child submitted to this experiment will die before exact age 1. This probability is given by

$$\Lambda(r,e) = P[\mathbf{B}+\mathcal{H}(\mathbf{F},r,e)<0] = P[\mathbf{B}+\mathbf{S}_h < -h(r,e)] = G(-h(r,e))$$

where G is the cumulative distribution function of **$B+S_h$**, the sum of the child biological endowment and his/her family health shifter. Based on this counter-factual we construct the following two experiments.

Counter-Factual 4: The Dependence of Infant Mortality on Family Education Holding Region of Residence Constant, Γ. This counter-factual is constructed by adding an extra step to the previous experiment. Instead of pre-specifying the family region of residence, we independently randomly choose for each randomly selected child a region of residence for his/her family with probabilities $\{q(r):r\in\mathbb{R}\}$. The other features of the experiment remain the same. The child is still allocated to the same family and his/her family (dis)educated up to level e. Let the probability of death of a randomly chosen child submitted to this experiment be denoted by $\Gamma(e)$. This probability is given by

$$\Gamma(e) = \sum_r q(r)\cdot\Lambda(r,e)$$

Notice that $\Gamma(e)$ is determined by the function h, the joint distribution of B and S_h, and the probabilities of selecting regions $\{q(r):r\in\mathbb{R}\}$. $\Gamma=\{(e,\Gamma(e)):e\in\mathbb{E}\}$ is our measure of dependence of the risk of infant mortality on family education holding constant region of residence. Alternatively, we refer to Γ as the intra-regional relationship between infant mortality and education.

Finally, notice that if A1 and A2 are satisfied then $\Gamma(e)=P^*[D=1|E=e]$. Hence, under the hypothesis that A1 and A2 are valid, estimates of $\Gamma=\{(e,\Gamma(e)):e\in\mathbb{E}\}$ can be found in Table 7.

Counter-Factual 5: The Dependence of Infant Mortality on Region of Residence Holding Family Education Constant, Ξ. Similarly to the previous counter-factual, this one is constructed by adding an extra step to counter-factual 3.

Table 8 presents the unstandardized and the standardized IMR and their respective dissimilarity indices. The fact that all standardized and unstandardized IMR are very similar indicates that regional differences in education explain a negligible fraction of the regional disparities in IMR. In fact, the standardization for the distribution of education reduces the index of dissimilarity only from 31 to 28%. Instead of pre-specifying the family education, we independently randomly choose for each randomly selected child a level of education for his/her family with probabilities $\{p(e):e\in\mathbb{E}\}$. The other features of the

$\{(r,\Psi(r)):r\in\mathbb{R}\}$ is our measure for the dependence of the risk of infant mortality on region of residence. Finally, notice that if A1 and A2 are satisfied then $\Psi(r)=P[\mathbf{D}=1|\mathbf{R}=r]$. This implies that under the hypothesis that A1 and A2 are valid, estimates of $\Psi=\{(r,\Psi(r)):r\in\mathbb{R}\}$ can be found in Table 2.

Counter-Factual 2: The Dependence of Family Education on Region of Residence, λ. Let a child be chosen from $\mathbb{C}$ and assigned to his/her family as is currently done, i.e., let child c be assigned to family $\mathbf{F}(c)$. Let us, however, interfere on the region of residence of his/her family by moving (if necessary) his/her family to region r. As in the previous counter-factual, the education level of the family will adjust to the (possible) change in the region of residence. For child c the new family education will be given by $\mathcal{E}(\mathbf{F}(c),r)$.

Let the probability that the family education of a randomly chosen child submitted to this counter-factual experiment equals to e be denoted by $\lambda(r,e)$. This probability is given by

$$\lambda(r,e) = P[e,(r,\mathbf{S}_e)=e]$$

Notice that $\lambda(r,e)$ is determined by e and the marginal distribution of $\mathbf{S}_e$. $\lambda=\{(r,e,\lambda(r,e)):r\in\mathbb{R} \text{ and } e\in\mathbb{E}\}$ is our measure for the dependence of the distribution of education on the region of residence.

Finally, notice that if A1 and A2 are satisfied then $\lambda(r,e)=P[\mathbf{E}=e|\mathbf{R}=r]$. Hence, under the hypothesis that A1 and A2 are valid, estimates of $\lambda=\{(r,e,\lambda(r,e)):r\in\mathbb{R} \text{ and } e\in\mathbb{E}\}$ can be found in Table 6.

Counter-Factual 3: The Dependence of Infant Mortality on Family Education and Region of Residence, Λ. Let a child be chosen from $\mathbb{C}$ and assigned to his/her family as is currently done, i.e., let child c be assigned to family $\mathbf{F}(c)$. Let us, however, interfere on the region of residence of his/her family and on their education level. More specifically, we move (if necessary) his/her family to region r and get it (dis)educated up to level e. i.e., for child c we change $\mathbf{R}(c)$ for r and $\mathbf{E}(c)$ for e. According to our framework, this child, c, will now, if submitted to this experiment, survive at least 12 months if and only if $\mathbf{B}(c)+\mathcal{H}(\mathbf{F}(c),r,e) > 0$. Let $\Lambda(r,e)$ denoted the probability that a randomly chosen child submitted to this experiment will die before exact age 1. This probability is given by

$$\Lambda(r,e) = P[\mathbf{B}+\mathcal{H}(\mathbf{F},r,e)<0] = P[\mathbf{B}+\mathbf{S}_h < -h(r,e)] = G(-h(r,e))$$

where G is the cumulative distribution function of $B+S_h$, the sum of the child biological endowment and his/her family health shifter. Based on this counter-factual we construct the following two experiments.

Counter-Factual 4: The Dependence of Infant Mortality on Family Education Holding Region of Residence Constant, Γ. This counter-factual is constructed by adding an extra step to the previous experiment. Instead of pre-specifying the family region of residence, we independently randomly choose for each randomly selected child a region of residence for his/her family with probabilities $\{q(r):r\in\mathbb{R}\}$. The other features of the experiment remain the same. The child is still allocated to the same family and his/her family (dis)educated up to level e. Let the probability of death of a randomly chosen child submitted to this experiment be denoted by $\Gamma(e)$. This probability is given by

$$\Gamma(e) = \sum_r q(r)\cdot\Lambda(r,e)$$

Notice that $\Gamma(e)$ is determined by the function h, the joint distribution of B and S_h, and the probabilities of selecting regions $\{q(r):r\in\mathbb{R}\}$. $\Gamma=\{(e,\Gamma(e)):e\in\mathbb{E}\}$ is our measure of dependence of the risk of infant mortality on family education holding constant region of residence. Alternatively, we refer to Γ as the intra-regional relationship between infant mortality and education.

Finally, notice that if A1 and A2 are satisfied then $\Gamma(e)=P^*[D=1|E=e]$. Hence, under the hypothesis that A1 and A2 are valid, estimates of $\Gamma=\{(e,\Gamma(e)):e\in\mathbb{E}\}$ can be found in Table 7.

Counter-Factual 5: The Dependence of Infant Mortality on Region of Residence Holding Family Education Constant, Ξ. Similarly to the previous counter-factual, this one is constructed by adding an extra step to counter-factual 3.

Table 8 presents the unstandardized and the standardized IMR and their respective dissimilarity indices. The fact that all standardized and unstandardized IMR are very similar indicates that regional differences in education explain a negligible fraction of the regional disparities in IMR. In fact, the standardization for the distribution of education reduces the index of dissimilarity only from 31 to 28%. Instead of pre-specifying the family education, we independently randomly choose for each randomly selected child a level of education for his/her family with probabilities $\{p(e):e\in\mathbb{E}\}$. The other features of the experiment remain the same. Let the probability of death of a randomly chosen child submitted to this experiment be denoted by $\Xi(r)$. This probability is given by

$$\Xi(r) = \sum_{e} p(e)\cdot\Lambda(r,e)$$

Notice that $\Xi(r)$ is determined by the function h, the joint distribution of B and S_h, and the probabilities of selecting regions $\{p(e):e\in E\}$. $\Xi=\{(r,\Xi(r)):r\in\Xi\}$ is our measure of dependence of the risk of infant mortality on region of residence holding constant region of residence.

Finally, notice that if A1 and A2 are satisfied then $\Xi(r)=P^{**}[D=1|R=r]$. Hence, under the hypothesis that A1 and A2 are valid, estimates of $\Xi=\{(r,\Xi(r)):r\in\Xi\}$ can be found in Table 8.

11. SUMMARY

This study investigated the relationship between infant mortality rate, IMR, and two dimensions of family background: region of residence and mother's number of years of schooling.

Four main conclusions have been reached. First, it has been shown that the IMR is much more sensitive to region of residence than to the level of mother's education. In fact, while the degree of regional dissimilarity in IMR is 30%, the degree of dissimilarity in IMR between educational categories is only 13%.

Secondly, we have shown that the Northeast is by far the region with the largest IMR. As a result, population movements out of the Northeast are those with the greater impact in reducing the level of the overall IMR.

Thirdly, we have found a non-convex, actually almost linear, relationship between IMR and mother's educational level. As a consequence, changes in the distribution of education which preserve the mean but decrease the variance will not reduce the overall level of the IMR. This result seems to be at odds with what was considered the intuitive result.

Finally, we showed that the large disparities in IMR between the Northeast and other regions do not operate through the effect of region of residence on the level of education of parents. In fact, while the total degree of regional dissimilarity in IMR is 30%, eliminating regional differences in parents education only reduces this degree of regional dissimilarity in 3%.

BIBLIOGRAPHY

ALBUQUERQUE, R.C., VILLELA, R. A situação social no Brasil: um balanço de duas décadas. **A Questão Social no Brasil**, Nobel, 1991.

ARRUDA, J.M. et alii. **Pesquisa nacional sobre saúde materno-infantil e planejamento familiar - PNSMIPF - Brasil 1986**. Rio de Janeiro: Benfam, 1987. 234p.

BELTRÃO, K.I., SAWYER, D.O. **Child mortality in Brazil: the effect of mother's education and Paraty**. S.l., s.ed., 1991, mimeo.

HENRIQUES, M.H., STRAUSS, J., THOMAS, D. Mortalidade infantil, estado nutricional e características de domicilio: a evidência brasileira. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v.19, n.3, p. 427-482, dez. 1989.

HOBCRAFT, J.N., McDONALD, J.W., RUTSTEIN, S.O. Socio-economic factors in infant and child mortality: a cross-national comparison. **Population Studies**, n.38, p.193-223, 1984.

McCRACKEN, S.D. **A multi-level socio-demographic analysis of early childhood mortality in Brazil.** Austin: University of Texas, May 1990 (Dissertação de Doutorado).

McCRACKEN, S.D., RODRIGUES, R.N. **Fertility decline, child health and survival in Brazil, Dominicana Republic and Peru, 1970-74 and 1980-84**. Washington, D.C., Mar. 1991 (Presented in the PAA Conference).

IBGE. **As informações sobre fecundidade, mortalidade e anticoncepção nas PNAD**. Rio de Janeiro, jun. 1989 (Discussion Papers, 15).

Ipea – Institute for Applied Economic Research

PUBLISHING DEPARTMENT

Coordination
Cláudio Passos de Oliveira

Supervision
Everson da Silva Moura
Reginaldo da Silva Domingos

Typesetting
Bernar José Vieira
Cristiano Ferreira de Araújo
Daniella Silva Nogueira
Danilo Leite de Macedo Tavares
Diego André Souza Santos
Jeovah Herculano Szervinsk Junior
Leonardo Hideki Higa

Cover design
Luís Cláudio Cardoso da Silva

Graphic design
Renato Rodrigues Buenos

The manuscripts in languages other than Portuguese published herein have not been proofread.

Ipea Bookstore

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo
70076-900 – Brasília – DF
Brazil
Tel.: + 55 (61) 3315 5336
E-mail: livraria@ipea.gov.br

Composed in Adobe Garamond 11/13.2 (text)
Frutiger 47 (headings, graphs and tables)
Brasília – DF – Brazil

Ipea's mission

Enhance public policies that are essential to Brazilian development by producing and disseminating knowledge and by advising the state in its strategic decisions.

ISSN 1415-4765

ipea Institute for Applied Economic Research

Secretariat of Strategic Affairs

BRAZILIAN GOVERNMENT BRASIL