

**ESCASSEZ DE ÁGUA E MORTALIDADE INFANTIL: NOVAS EVIDÊNCIAS  
A PARTIR DE UM PAINEL QUANTÍLICO ROBUSTO**

**AUTORES:**

**Fernanda Leite Santana**

Doutoranda em Economia Aplicada  
Programa de Pós Graduação em Economia  
Universidade Federal da Paraíba, Brasil  
E-mail: [nandinha.economia@gmail.com](mailto:nandinha.economia@gmail.com)

**Wallace Patrick Santos de Farias Souza**

Doutorando em Economia Aplicada  
Programa de Pós Graduação em Economia  
Universidade Federal do Rio Grande do Sul, Brasil  
E-mail: [wpsfarias@gmail.com](mailto:wpsfarias@gmail.com)

**Erik Alencar de Figueiredo**

Doutor em Economia pela UFRGS e Pós-Doutor pela University Of Tennessee  
Programa de Pós Graduação em Economia,  
Universidade Federal da Paraíba, Brasil  
E-mail: [eafigueiredo@gmail.com](mailto:eafigueiredo@gmail.com)

## **ESCASSEZ DE ÁGUA E MORTALIDADE INFANTIL: NOVAS EVIDÊNCIAS A PARTIR DE UM PAINEL QUANTÍLICO ROBUSTO**

**RESUMO:** Este artigo propõe averiguar as implicações de restrições pluviométricas sobre os períodos iniciais de vida dos indivíduos. Em particular, se investigam os resultados de saúde expressos na taxa de mortalidade infantil, tal qual abordado em Rocha e Soares (2015), para um painel de município-a-mês (1996-2010) do semiárido brasileiro, levando em consideração importantes efeitos: confiabilidade dos dados de precipitação, a endogeneidade existente, bem como seus efeitos individuais com as variáveis independentes, até o momento negligenciados pela literatura. Para tanto, emprega-se o método de regressão quantílica em dados em painel, desenvolvido por Canay (2011), como também o método de regressão quantílica em dados em painel com variáveis instrumentais, proposto por Harding e Lamarche (2009), em cenários com dados de reanálise e observacionais de precipitação. Os resultados, em conformidade com a literatura vigente, mostram um efeito negativo da precipitação e um efeito positivo da seca sobre a taxa de mortalidade infantil dos municípios pertencentes ao semiárido brasileiro. Contudo, a contribuição do presente trabalho consiste em ressaltar que a escolha de dados de clima e de métodos inadequados pode estar superestimando essas relações.

**Palavras-chaves:** Taxa de Mortalidade Infantil; Precipitação; Saúde; Semiárido Brasileiro

**ABSTRACT:** This article proposes to investigate the implications of rainfall restrictions on the early life periods of individuals. In particular, we investigate health outcomes expressed in the child mortality rate, like addressed in Rocha and Smith (2015) for a panel of municipalities-by-month (1996-2010) the Brazilian semiarid, taking into account important effects: the reliability of rainfall data, existing endogeneity as well as their individual effects with the independent variables, until now neglected by the literature. Therefore, we employ the quantile regression method on panel data addressed by Canay (2011), as well as the quantile regression method on panel data instrumental variables proposed by Harding and Lamarche (2009) in scenarios with data reanalysis and observational precipitation. The results, according to the literature, show a negative effect of precipitation and a positive effect of drought on child mortality rate of municipalities in the Brazilian semiarid region. However, the contribution of this work is to emphasize that the choice of climate data and of inappropriate methods may be overestimating these relationships.

**Keywords:** Infant Mortality Rate; Precipitation; Health; Brazilian semiarid

**JEL:** R10, C23, I10

## 1 INTRODUÇÃO

Há um crescente corpo de pesquisa que investigam os impactos das condições iniciais sobre os determinantes dos resultados socioeconômicos dos indivíduos quando adultos (BARCKER, 1998; GLUCKMAN e HANDSON, 2005; CUNHA e HECKMAN, 2007; HECKMAN, 2007). Dentro desse contexto, choques climáticos nos anos iniciais de vida, sobretudo relacionados às quantidades pluviométricas, tem desempenhado papel relevante nos estudos sobre o desenvolvimento humano (MACCINI e YANG, 2006; UGAZ e ZANOLINI, 2011; BURGESS et. al, 2014).

Dentre as principais consequências têm-se ressaltado o efeito negativo em processos produtivos (AGUILAR e VICARELLI, 2011; BEZABIH, DI FALCO e MEKONNEN, 2014), na saúde (CURRIE e VOGL, 2012; CURRIE et. al., 2013; CURRIE e ROSSIM-SLTER, 2013; ROCHA e SOARES, 2015), bem como na acumulação de capital humano (SHAH e STEINBERG, 2013). O pressuposto básico é de que os choques climáticos propiciam ambientes iniciais (des) favoráveis, uma vez que há efeito negativo/positivo na cadeia produtiva, na saúde e na qualidade de vida da região (CIRILO, 2008).

Nesse ínterim, a questão pluviométrica vem ganhando destaque em estudos recentes. Por exemplo, Shah e Stenberg (2013) ressaltam que exposição a períodos de seca no início da vida tem efeitos prejudiciais nos salários, na saúde e na escolaridade dos indivíduos adultos da área rural da Índia. Burgess, et. al. (2014) ratificam esses resultados para a Índia em termos de mortalidade em populações rurais e urbanas, bem como altas temperaturas reduzindo salários agrícolas e reais. No Brasil, Rocha e Soares (2015) enfatizam a escassez da água na região semiárida do nordeste como fator determinante para a mortalidade infantil mais elevada, menor peso ao nascer, e períodos de gestação mais curtos.

Por outro lado, Maccini e Yang (2008) investigam o efeito das condições climáticas na época do nascimento sobre a saúde, educação e resultados socioeconômicos dos adultos indonésios e concluem que uma maior precipitação no início da vida tem grandes efeitos positivos sobre os resultados de mulheres adultas, mas não para homens. Este conjunto de evidências vem atribuindo determinado grau de importância para as questões pluviométricas no desenvolvimento humano destas regiões, contudo, é relevante ressaltar que estas estimativas podem estar viesadas se não forem observadas questões como confiabilidade de dados de clima e a endogeneidade existente nessas relações.

Em termos do preditor clima, verifica-se que boa parte dos estudos na economia da saúde vem considerando dados históricos de variações das chuvas como covariável das mudanças climáticas, uma vez que a precipitação é uma das principais para caracterizar o clima do continente, como também para o planejamento de inúmeras atividades produtivas (FREITAS, FRANCHITO e RAO, 2010). Neste cenário, alguns trabalhos vêm utilizando dados de Reanálise<sup>1</sup> - chamados de dados modelados – e outros, os observacionais<sup>2</sup>. Ambos reproduzem a climatologia e as variabilidades das regiões de interesse, embora possam sub/superestimar a intensidade na média. Na literatura meteorológica, sobretudo climatológica, há um grande debate acerca da precisão dos dados de reanálise no que concerne a exploração dos métodos utilizados

---

<sup>1</sup> O processo de reanálise captam dados oriundos de satélites, no qual acontece por duas etapas, primeiro, são interpolados através de modelos utilizados para prever tempo e, posteriormente, novos dados são interpolados para áreas em que não possuem dados disponíveis (FREITAS, FRANCHITO e RAO, 2010).

<sup>2</sup> Dados observacionais são oriundos de estações meteorológicas de superfície. De acordo com Vieira e Piculli (2009), há dois tipos de estações: i) convencionais – onde os dados são coletados e organizados por um observador; e ii) automáticas – a coleta é feita através de sensores.

para sua obtenção e a comparação com dados observacionais (ver GUEDES et. al., 2005; PINTO et. al., 2009; FREITAS, FRANCHITO e RAO, 2010). Para fins desta pesquisa optou-se pelos dados observacionais obtidos no BDMEP.

Não obstante, alguns desafios econométricos devem, também, aqui serem ressaltados, principalmente no que tange a consistência dos parâmetros de interesse, ou seja, o parâmetro relacionado à precipitação. Primeiramente, cabe esclarecer que podem ocorrer erros de medição uma vez que tanto os *inputs* quanto os *outputs* só podem ser obtidos de forma aproximada, e.g., taxa de mortalidade infantil utilizada em Rocha e Soares (2015) pode não ser uma boa aproximação para os resultados de saúde dos indivíduos na idade adulta. Além disso, os *inputs*, especificamente dados de precipitação, podem ser endógenos e não observados.

O ponto é a omissão de variáveis, uma vez que muitas pesquisas vêm atribuindo ao fator clima especial ênfase, por ora, deixando de levar em conta outros fatores que podem afetar a mudança dos padrões de saúde que não seja o clima. Por exemplo, a distribuição geográfica e a sazonalidade de doenças transmitidas podem alterar resultados de saúde e não necessariamente a causa disso estar nas mudanças climáticas. Além do que, há correlação entre mudanças climáticas e outros atributos individuais, e.g., a renda dos pais, já que onde tem climas mais secos tendem a ter piores níveis socioeconômicos. Ambos os problemas supracitados podem causar além do viés, a inconsistência dos parâmetros (Wooldrige, 2011), necessitando, portanto, rigor metodológico adequado para suprir tais restrições.

Em que pese o desenvolvimento recente da associação de clima no desenvolvimento humano dos indivíduos, estes questionamentos merecem especial atenção para não incorrer em possíveis sub ou superestimações de resultados, e, conseqüentemente, na definição do correto impacto das mudanças climáticas sob resultados de saúde, crucial para os formuladores de políticas públicas. Deste modo, o objetivo desta pesquisa é adicionar elementos à literatura existente no intuito de reduzir incertezas e fornecer *insights* sobre os resultados de saúde dos indivíduos.

Para fins de estudo, utilizar-se-á como referência o trabalho realizado por Rocha e Soares (2015) sobre a escassez de água do semiárido brasileiro. A ideia é corrigir/ratificar os resultados dos autores apoiando projeções mais sutis de como as taxas de mortalidade poderia ser alteradas com a mudança climática levando em consideração os questionamentos aqui levantados, ou seja, quanto maior/menor seriam os resultados de saúde dados à escassez de água no semiárido com uma base de clima mais precisa e um modelo que considere a endogeneidade existente.

Para tanto, a estratégia empírica adotada basear-se-á em três etapas. Primeiro, com a base dados obtidos de estações meteorológicas observacionais, realizar-se-á estimação de dados de painel com efeitos fixos, similar ao procedimento realizado por Rocha e Soares (2015). Em segundo lugar, baseado em Canay (2011), estimar-se-á por regressão quantílica com dados em painel considerando a heterogeneidade não observada e a dos efeitos das covariáveis em ambientes com efeito fixo. Por último, na terceira etapa, visando endogeneizar a quantidade de precipitações ocorridas no semiárido do Nordeste brasileiro, utilizar-se-á a Temperatura da Superfície do Mar (TSM), como instrumento em um modelo empírico baseado em Harding e Lamarche (2009).

Além desta introdução, o restante do trabalho está organizado em mais quatro seções. A próxima seção apresenta um quadro conceitual teórico. Na terceira seção a estratégia empírica, bem como a descrição dos dados e concepção das variáveis utilizadas nas estimações. Na quarta seção são discutidos os resultados apurados e a quinta parte reservada para as considerações finais.

## 2 CONCEITOS TEÓRICOS

O estudo da literatura da economia da saúde remonta a análise comportamental desenvolvido por Grossman (1972) que, análogo à literatura sobre o capital humano, considera a saúde como resultante de um processo de produção, portanto, uma variável de estoque. Formalmente, a Função de Produção de Saúde (FPS) considera a saúde individual no tempo  $t$ ,  $H_t$ , em função de:

$$H_t = f(H_0, N_t, X, C_t, D_t), \text{ onde } t = 1, 2, \dots, n \quad (1)$$

onde:  $H_0$  é uma dotação inicial de saúde;  $N_t$  é o vetor de *inputs* de saúde;  $X$  é o vetor de variáveis demográficas e invariantes no tempo;  $C_t$  variáveis relacionadas ambientes externos, como a infraestrutura do local ou ambiente de doença.

A partir deste escopo, desenvolvimentos recentes relacionados a literatura da psicologia, neurociência cognitiva, economia e epidemiologia tem destacado que a condição inicial,  $H_0$  depende das características genéticas ( $G$ ), mas também das condições iniciais vivenciadas pelos indivíduos  $R_0$ :

$$H_0 = v(G, R_0) \quad (2)$$

Nesse contexto, o modelo teórico de Heckman (2007) é aqui utilizado e adaptado para fins desta pesquisa, e tem como base a teoria do ciclo da vida, em que apresenta como principal característica o desenvolvimento de um processo dinâmico de múltiplos estágios, ou seja, o ciclo de vida dos indivíduos é analisado considerando vários períodos e cada qual com uma taxa de retorno diferente. A ideia é mostrar que o desenvolvimento da criança depende de um conjunto de fatores. Assim, o modelo pode ser expresso como:

$$\theta_{t+1}^k = f_t^k(\theta_t^H, I_t^k, h^H), \quad k \in \{H\} \quad (3)$$

onde  $\theta_t$  denota o vetor de capacidades de saúde:  $\theta_t = (\theta_t^H)$ . Deixe  $I_t$  denotar o vetor de investimentos em capacidades de saúde:  $I_t = (I_t^H)$ . Use  $h = (h^H)$  para denotar habilidades parentais<sup>3</sup>. A equação (3) traduz que em cada estágio  $t$ , pode-se definir uma tecnologia recursiva para saúde ( $k = H$ ).

Conforme Heckman (2007) esta tecnologia é suficientemente rica para descrever a evidência de que a capacidade da criança em ser mais saudável afeta o resultados de saúde posteriores. Tal constatação é decorrente de duas características importantes da tecnologia: *i) autoproductividade* – habilidades produzidas em um estágio aumentam as habilidades desenvolvidas nos estágios posteriores, e surge quando  $\partial f_t(h, \theta_t, I_t) / \partial \theta_t > 0$ ; e *ii) complementariedade dinâmica* – habilidades produzidas em um estágio aumentam a produtividade do investimento em estágios subsequentes e surge quando  $\partial^2 f_t(h, \theta_t, I_t) / \partial \theta_t \partial I_t' > 0$ .

Além disso, os autores afirmam que esta tecnologia capta os períodos sensíveis e críticos postulados na literatura psicológica e da neurociência. Considere  $t^*$  ser o período sensível para  $\theta_{t+1}$ , assim:

$$\frac{\partial \theta_{t+1}}{\partial I_s} = \frac{\partial m_t(h, \theta_1, I_1, \dots, I_t)}{\partial I_s} \equiv 0 \quad \text{para todo } h, \theta_1, I_1, \dots, I_s, \quad s \neq t^*$$

Mas,

<sup>3</sup> Por exemplo, QI, genes, educação, renda.

$$\frac{\partial \theta_{t+1}}{\partial I_{t^*}} = \frac{\partial m_t(h, \theta_1, I_1, \dots, I_t)}{\partial I_{t^*}} > 0 \text{ para alguns } h, \theta_1, I_1, \dots, I_t$$

Esta condição diz que os investimentos em  $\theta_{t+1}$  são produzidos no período  $t^*$ , mas não em qualquer outro período  $s \neq t^*$ . Período  $t^*$  é o período sensível para  $\theta_{t+1}$  se:

$$\left. \frac{\partial \theta_{t+1}}{\partial I_s} \right|_{h=\bar{h}, \theta_1=\theta, I_1=i_1, \dots, I_t=i_t} < \left. \frac{\partial \theta_{t+1}}{\partial I_{t^*}} \right|_{h=\bar{h}, \theta_1=\theta, I_1=i_1, \dots, I_t=i_t}$$

Em outras palavras, o período de  $t^*$  é um período sensível em relação ao período de  $s$  se, ao mesmo nível de insumos, o investimento é mais produtivo na fase  $t^*$  do que em outro estágio  $s \neq t^*$ . A ideia é de que no **período sensível** haja certa estabilidade na conectividade dos neurônios decorrente da influência do ambiente, assim indivíduos nascidos em ambientes restritos tendem a produzir habilidades menos eficazes. Mesmo que se invistam períodos mais tardes, denominados de **períodos críticos**, este nunca será um substituto perfeito. Nesse contexto, o objetivo da pesquisa é investigar os efeitos das mudanças pluviométricas, corroboradas pelos longos períodos de seca no nordeste, nos resultados de saúde de jovens nascidos no semiárido brasileiro na década de 90.

### 3 METODOLOGIA E DADOS

Esta seção destina-se a apresentar a estratégia adotada neste estudo. Inicialmente, apresenta-se o modelo empírico para avaliar o impacto das restrições pluviométricas dos primeiros anos de nascimento nos resultados de saúde dos indivíduos nascidos no semiárido na década de 90. A seguir, discorre-se sobre dois métodos adotados. O primeiro, baseado em Canay (2011) visa apresentar modelo de regressão quantílica com dados em painel com efeitos fixos. Por sua vez, o segundo baseia-se em Harding e Lamarche (2009) numa abordagem de regressão quantílica para modelos de dados em painel com variáveis endógenas e efeitos individuais associados com as variáveis independentes. Por fim, expõem-se os dados utilizados, bem como a variável instrumental Temperatura da Superfície do Mar (TSM).

#### 3.1 Estratégia empírica

##### A. Modelo Empírico

O painel é composto de municípios por mês segundo a região semiárida nordestina, para o período de 1996-2010, construído com as variáveis de interesse. Assim, a análise dos impactos das variações pluviométricas na mortalidade infantil e saúde durante o período de gestação segue a especificação do painel de municípios por mês de nascimento abordado em Rocha e Soares (2015):

$$Y_{itm} = \alpha + \beta R_{itm} + \phi_{im} + \lambda_t + \varphi t_{gtm} + \pi T_{itm} + \epsilon_{itm} \quad (4)$$

onde  $Y_{itm}$  é um resultado de saúde médio (por município) para as crianças nascidas no município  $i$ , no ano  $t$  e mês  $m$ . A variável dependente utilizada é a mortalidade infantil.  $R_{itm}$  é a variável de precipitação;  $\phi_{it}$  é um efeito-fixo para município  $i$  e mês  $m$  (com  $m = 1, 2, \dots, 12$ );  $\lambda_t$  é um efeito-fixo ano;  $T_{itm}$  é a temperatura média no município  $i$  no mesmo período de 12 meses antes do nascimento;  $t_{gtm}$  é a tendência temporal linear e  $\epsilon_{itm}$  é o termo de erro aleatório.

Cabe aqui destacar que a variável relacionada ao clima,  $R_{itm}$ , é análoga a realizada em Rocha e Soares (2015), ou seja, são concebidas de mesma forma, embora difiram na natureza do problema, uma vez que a fonte de dados e sua tabulação são diferentes entre si. Esta variável é construída de duas maneiras, ambas para captar os efeitos das secas no semiárido através da flutuação das chuvas durante o período de gestação dos indivíduos:

- I. **Log desvio das chuvas nos últimos 12 meses** – conforme os autores, esta variável por ser interpretada como uma porcentagem de desvio da precipitação média. Definida por:

$$R_{it\tau}^1 = \ln \left( \sum_{t=\tau-11}^{\tau} r_{im} \right) - \ln(\bar{r}_i) \quad (5)$$

onde  $r_{im}$  indica a precipitação mensal no município  $i$  e mês  $m$ ,  $\bar{r}_i$  é a precipitação anual média histórica no município  $i$ , e  $\tau$  indica o mês de nascimento de um indivíduo. Diferente ao realizado em Rocha e Soares (2015), o cálculo do  $\sum_{t=\tau-11}^{\tau} r_{im}$  será realizado apenas para os municípios pertencentes ao semiárido, uma vez que contabilizar estados onde tem muita chuva, por exemplo, os estado do Norte, pode superestimar esta covariável.

- II. **Seca dos últimos 12 meses** – uma variável *dummy* que assume o valor 1 se houve seca no período, em síntese esta variável busca captar os eventos extremos, calculado por:

$$R_{it\tau}^2 = 1 \text{ se } \sum_{t=-11}^{\tau} r_{im} < (\bar{r}_i - r_i^{SD}), \text{ e } 0 \text{ para outros} \quad (6)$$

onde  $r_i^{SD}$  é o histórico anual do desvio padrão de chuvas para o município  $i$  (calculado sobre o período de 1961-1990). Em outras palavras, o  $R_{it\tau}^2 = 1$  indica que a precipitação ao longo dos 12 meses anteriores ao nascimento de um indivíduo era mais do que um desvio padrão abaixo da média histórica para o município  $i$ .

Em suma, a equação (4) estima os efeitos das variações pluviométricas, aqui representadas nas equações (5) e (6), sobre a taxa de mortalidade infantil em dados em painel de município-a-mês de 1996 à 2010.

### B. Métodos Empíricos

Para estimar a equação (4) é necessário considerar que os efeitos fixos por município-mês controlam variações climáticas em determinados períodos do ano assim como diferenças entre os municípios, enquanto que o efeito fixo por ano tenta captar choques agregados e tendências que impactam toda a região. Por sua vez, a tendência linear controla a dinâmica da evolução das mudanças climáticas e socioeconômicas da região e a temperatura é responsável por outras variações climáticas que podem estar relacionadas com a quantidade de chuvas.

No entanto, como se trata de uma estimativa do efeito médio na variável dependente, este pode não retratar a realidade da maioria dos municípios. A explicação

para tal é a existência de uma gama de heterogeneidade entre os municípios. Potencialmente, modelos de regressão quantílica poderiam explicar a heterogeneidade não observada e efeitos de localização, como também modelos de dados em painel permitiriam a inclusão dos efeitos fixos. A junção desses dois métodos de regressão permite estimadores mais robustos.

Uma grande literatura sobre estimação de regressões quantílicas com dados em painel tem surgido nos últimos anos, tal como Koenker (2004) que introduziu uma abordagem geral para estimação com dados longitudinais. Controlando para os efeitos fixos municipais e a heterogeneidade das covariáveis, a estimação por *quantis* é uma abordagem mais flexível para a análise de dados em painel do que os estimadores de efeitos fixos e aleatórios gaussianos clássicos. Em adição, considerar os efeitos de tratamento em modelos quantílicos de dados em painel constitui mais recente robustez das análises desta natureza. Não obstante, é necessário ir além e também captar outros efeitos na média. Portanto, neste arcabouço, apresentam-se os dois estimadores aqui utilizados.

### *I - Estimador assintoticamente consistente de Canay (2011) – (EACC)*

Canay (2011) apresenta um estimador consistente em que utiliza uma simples transformação de dados para eliminar os efeitos fixos dos *quantis* condicionais, conforme o modelo:

$$y_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it} \quad E(\varepsilon_{it}/X_{it}, \alpha_i) = 0 \quad (7)$$

Onde  $(y_{it}, x_{it}) \in \mathbb{R} \times \mathbb{R}^k$  são variáveis observáveis,  $\theta(U_{it}) = \beta$  e  $(U_{it}, \alpha_{it}) \in \mathbb{R} \times \mathbb{R}$  são não observados.  $U_{it}/x_{it}, \alpha_{it} \approx U[0,1]$ . A função  $\tau \rightarrow x'\theta(\tau)$  é assumida como sendo estritamente crescente em  $\tau \in (0,1)$  e o parâmetro de interesse é assumida para ser  $\theta(\tau)$ . A principal restrição é que as heterogeneidades individuais associadas a  $x_{it}$  deve ter um efeito local puro.

Em suma, o método versa na questão em que as condições  $(U_{it}, \alpha_{it})$  do parâmetro  $\theta(\tau)$  podem ser identificadas e consistentemente calculadas a partir dos dados. Para estimar a equação (7), Canay (2011) propõe um estimador simples de dois passos, sendo  $E_T(\cdot) \equiv T^{-1} \sum_{t=1}^T(\cdot)$  e  $E_{nT}(\cdot) \equiv (nT)^{-1} \sum_{t=1}^T \sum_{i=1}^n(\cdot)$ , como segue:

**Passo 1.** Deixe  $\hat{\theta}_\mu$  para  $\sqrt{nT}$  – estimador consistente de  $\theta_\mu$ . Estimam-se os efeitos fixos como:  $\hat{\alpha}_i \equiv E_T[y_{it} - x'_{it}\hat{\theta}_\mu]$ .

**Passo 2.** Sendo  $\hat{y}_{it} \equiv y_{it} - \hat{\alpha}_i$  e definir o estimar do dois-estágios  $\hat{\theta}(\tau)$  como:

$$\hat{\theta}(\tau) \equiv \underset{\theta \in (\cdot)}{\operatorname{argmin}} E_{nT}[\rho_r(\hat{y}_{it} - x'_{it}\theta)]$$

O método consente na obtenção de *insights* sobre os efeitos dos regressores sobre a variável dependente. Isto porque permite uma leitura dos efeitos heterogêneos através da transformação dos dados, uma vez que modelos de regressão padrão que seguem a média, a heterogeneidade muitas vezes não é captada nestes modelos.

### *II - Estimador de regressões quantílicas e dados em painel com variáveis instrumentais (EQDIV)*

Alternativamente, considerar-se-á para efeitos de resultados, o estimador proposto por Harding e Lamarche (2009), no qual apresentam uma abordagem de regressão quantílica para modelos de dados em painel com variáveis endógenas. Este



método é mais robusto do que o método de Canay (2011), uma vez que potencializa as estimativas ao considerar na regressão efeitos além da heterogeneidade individual. O método é uma extensão do procedimento de Chernozhukov e Hansen (2008) permitindo a inclusão de efeitos fixos introduzidos em Koenker (2004). Denote o vetor de variável endógena,  $d$ , na equação (7):

$$y_{it} = d'_{it}\delta + x'_{it}\beta + \alpha_i + u_{it} \quad (8)$$

$$d_{it} = h(x_{it}, w_{it}, v_{it}) \quad (9)$$

$$\alpha_i = g(x_{i1}, \dots, x_{iT}, d_{i1}, \dots, d_{iT}, \varepsilon_{it}) \quad (10)$$

onde a equação (9) define a variável endógena  $d$  relacionada a um vetor de instrumentos  $w$  que estão estocasticamente independente de  $u$ . Já a variável  $v$  é estocasticamente dependente de  $u$ . A equação (10) expressa a correlação entre as variáveis e os efeitos individuais. Neste caso, assume-se que  $\varepsilon$  é independente de  $u$  e  $v$ . Considere a função objetivo para o relacionamento quantílico instrumental condicional:

$$R(\tau, \delta, \beta, \gamma, \alpha, \mu) = \sum_{t=1}^T \sum_{i=1}^N \rho_{\tau}(y_{it} - d'_{it}\delta - x'_{it}\beta - \alpha_i - z'_{it}\alpha - \widehat{w}'_{it}\gamma) \quad (11)$$

Onde  $\widehat{w}_{it}$  é um vetor de instrumentos e  $\rho_{\tau} = u(\tau - 1(u \leq 0))$  é uma função perda de regressão quantílica. Seguindo o procedimento de dois passos de Chernozhukov e Hansen (2006, 2008), temos:

**Passo 1.** Minimizar a função objetivo (11) para  $\beta, \alpha, \mu$  e  $\gamma$  em função de  $\tau$  e  $\delta$ :

$$\{\widehat{\beta}(\tau, \delta), \widehat{\gamma}(\tau, \delta), \widehat{\alpha}(\tau, \delta)\} = \arg \min_{\beta, \gamma, \alpha} R(\tau, \delta, \beta, \gamma, \alpha) \quad (12)$$

**Passo 2.** Em seguida, estima-se o coeficiente da variável endógena para encontrar o valor de  $\delta$ , no qual minimiza a função distância ponderada definida em  $y$ :

$$\widehat{\delta}(\tau) = \arg \min_{\delta} \widehat{\gamma}(\tau, \delta)' A \widehat{\gamma}(\tau, \delta) \quad (13)$$

Onde  $A$  é uma matriz definida positiva. De acordo com Harding e Lamarche (2009) este estimador é assintoticamente normal podendo ser executada para mais de um quantil simultaneamente.

### 3.2 Descrição dos dados

#### A. Dados de Clima

O uso de variáveis relacionadas ao clima é utilizado para suportar uma ampla gama de estudos de investigação sobre o impacto destas variáveis sobre os resultados de saúde dos indivíduos. Neste sentido, o uso de dados de precipitação e temperatura vem ganhando espaço da literatura, pois a variabilidade destas afetam diversas atividades econômicas (FREITAS, FRANCHITO e RAO, 2010) e costumam representar a variabilidade de sinais geofísicos e regimes de precipitação (GUEDES et. al., 2004),

sobretudo na literatura climatológica. O Quadro 1 apresenta algumas fontes e tipo de análise.

Na literatura meteorológica e climatológica, destacam-se três tipos de dados: i) dados observacionais - estações; ii) dados de reanálise e iii) dados de sensoriamento. A primeira é oriunda de estações compostas por sensores isolados que registram continuamente os parâmetros meteorológicos que são lidos e anotados por um observador a cada intervalo. Contudo, apesar de reconhecer a fidedignidade deste tipo de observação, questões geográficas e econômicas tem dificultado a produção de séries temporárias de longo prazo (PINTO et. al., 2009).

Por sua vez, dados de reanálise são muito convenientes para completar as lacunas onde não possuem informações, como por exemplo, onde não há estação meteorológica, ou mesmo, devido a problema de obtenção de dados. No entanto, as séries históricas são geradas a partir de modelos e são construídas com base em algum tipo de observação em superfície, inclusive de estações. Por fim, os dados de sensoriamento remoto que combinam dados de superfície com estimativas feitas por satélite para calibrar o banco de dados.

**Quadro 1** – Diversas Fontes e Metodologias de dados de Precipitação

Tipos de Dados	Fonte	Grade	Autores que utilizaram os dados
Observacionais	ANEEL	-	Freitas, Franchito e Rao, 2010;
	SUDENE	-	Menezes, et. al (2008)
	BDMEP	-	Rao et. al., 1996;
	DNOCS	-	Rao, Lima e Franchito (1993)
Reanálise	NCEP/NCAR	~1,9° x 1,9°	Guedes et. al., 2005; Pinto et. al., 2009; Freitas, Franchito e Rao, 2010;
	CPTEC	0,4° x 0,4°	Pinto et. al., 2009;
	CPC/NOAA Terrestrial Air Temperature and Terrestrial Precipitation	0,5° x 0,5°	Rocha e Soares (2015)
Sensoriamento	GPCP	2,5° x 2,5°	Pinto et. al., 2009; Freitas, Franchito e Rao, 2010;
	CMAP	2,5° x 2,5°	Pinto et. al., 2009;
	TRMM	1,0° x 1,0°	Pinto et. al., 2009;

Fonte: Elaboração dos autores.

É possível perceber que existe um *trade-off* no uso de variáveis clima. Se por um lado, os dados observacionais são fontes mais confiáveis, a ausência de informações e as quebras das séries de longo prazo tem dificultado as investigações climatológicas. Todavia, os dados de reanálise dependem de fontes primárias confiáveis, modelos bem ajustados e técnicas de obtenção robustas. Neste sentido, Silva, et. al. (2007), considerando as novas calibrações efetuadas no modelo de obtenção dos dados de reanálise da CPC/NCAR, realizado em 2005, constataram através dos diagramas de dispersão tendências para valores mais altos da CPC em relação aos dados observacionais. Não obstante, os autores ainda ressaltam, consistentemente aos estudos anteriores, que as comparações indicam menor número de dias secos, maior número de dias de baixa precipitação, e menos eventos extremos de precipitação por grade analisada, apesar da boa correlação existente entre as mesmas.

Outros estudos também visam comparar alguns produtos de precipitação. Por exemplo, Freitas, Franchito e Rao (2010) que compararam os dados ANEEL (observacionais); GPCP (sensoriamento) e NCEP/NCAR (reanálise) e concluíram que

os dados de reanálise NCEP/NCAR superestimam a precipitação no Nordeste Brasileiro e que os dados observacionais ANNEL, apesar da ausência de dados e falhas estas reproduziram melhor a climatologia. Guedes et. al. (2005) em um estudo para Fortaleza-CE compararam dados NCEP/NCAR com os dados das estações e concluem que os dados observados e de reanálise são comparáveis, mas o ciclo anual da reanálise parece estar um mês adiantado em relação aos dados observados e os desvios padrões mensais apresentam grandes diferenças. Tais resultados são corroborados em Pinto et. al. (2009) ao concluir que os dados de reanálise CPTEC não são capazes de representar os principais regimes de precipitação existentes no continente, com forte tendência a superestimar a intensidade média no interior do Nordeste Brasileiro.

Nesse arcabouço, recentemente Rocha e Soares (2015) utilizaram os dados *Terrestrial Precipitation: 1900-2010 Gridded Monthly Times Series* (versão 3.02). Este conjunto de séries temporais é procedente das estações do *Global Historical Climatology Network* (GHCN) e *Global Surface Summary of day* (GSOD) através de interpolação espacial. Portanto, trata-se de dados de reanálise e que de acordo com os estudos supracitados podem-se incorrer em superestimações de valores de precipitação e temperatura para o nordeste brasileiro. Para fins de comparação com os resultados recentes dos autores e em virtude da ausência de séries mais longas, neste artigo empregam-se os dados observacionais do Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP).

A ideia é verificar se os impactos das restrições pluviométricas na época de nascimento do indivíduo são realmente relevantes nos resultados de saúde quando adultos em cenário com dados mais confiáveis de precipitação, embora isso acarrete na redução da amostra. Isto porque este banco não possui estações para todos os municípios pertencentes do semiárido brasileiro como também os inícios das séries históricas divergem entre elas.

Neste escopo foram criadas as duas variáveis expressas nas equações (5) e (6) utilizando os dados obtidos na BDMEP, primeiro localizando o município através da latitude e longitude; e depois, gerando os dados mensais de precipitação para cada município que compõe o semiárido brasileiro para o período de 1977 à 2010. No total, foram geradas 8.370 observações distribuídas nos 49 municípios analisados.

### *B. Variável Instrumental*

Para empregar o método de Harding e Lamarche (2009), que seja um estimador de regressão quantílica de dados em painel com variáveis instrumentais necessitar-se-á de um vetor que denote uma variável que afeta o nível de precipitação, mas não tem efeito direto sobre os resultados de saúde dos indivíduos. Pesquisas tem mostrado que as condições oceânicas, sobretudo a Temperatura da Superfície do Mar (TSM), apresentam-se como a principal variável física a influenciar as condições climáticas.

Um fenômeno climático de interação oceano-atmosfera bastante investigado é o *El Niño* – Oscilação Sul (ENOS). Esta variabilidade climática é cíclica, oriunda do oceano Pacífico, com periodicidade que varia de 3 a 7 anos, duração igual ou superior a cinco meses consecutivos (TRENBERTH, 1997), e podem ser detectadas através das anomalias persistentes da TSM. Anos de *El Niño* (ou *El Niña*) estão normalmente associados à escassez (ou chuvas abundantes). A influência dos ENOS, sobretudo no Nordeste Brasileiro, tem sido objeto de estudo, no qual se destaca os estudos de Kayano, Rao e Moura (1988), Rao, Lima e Franchito (1993), Xavier et. al. (2003), Alves e Souza (2003), Santos e Brito (2007), entre outros.

Outro fenômeno oceano-atmosfera é o dipolo do Atlântico. Destacam-se os estudos seminais de Hastenrath e Heller (1977), Moura e Slukla (1981), mais

recentemente, Aragão (1998), Moraes Neto e Araújo (2007), Andreoli e Kayano (2007), entre outros. Identifica-se o Dipolo quando as águas do Atlântico tropical estão mais quentes e do Atlântico Equatorial e Tropical Sul estão mais frias culminando para movimentos descendentes transportando o frio e seco inibindo a formação de nuvens e diminuindo a precipitação (ARAGÃO, 1998).

Com base na literatura climatológica, optou-se pelos dados de temperatura dos Oceanos Pacíficos e Atlântico, conforme as pesquisas retromencionadas, uma vez que tanto os ENOS quanto o Dipolo Atlântico afetam a ocorrência de secas no semiárido brasileiro, mas não afetam diretamente a taxa de mortalidade infantil nem as condições de saúde ao nascer. Desta feita, os dados de TSM observados no Oceano Pacífico e Atlântico foram extraídos do *International Comprehensive Ocean–Atmosphere Dataset* (ICOADS) e *National Ocean and Atmospheric Administration* (NOAA), respectivamente.

O primeiro<sup>4</sup> arquivo contém dados mensais em pontos de grade de 2° x 2° de latitude-longitude para um período de 1854 – até o presente composto por quatro arquivos a seguir: i) Extremo Oriente - Niño 1 + 2 (0-10S; 90W-80W); ii) Leste Tropical - Niño 3 (5N-5S; 150W-90W); iii) Leste Central Tropical - Niño 3.4 (5N-5S; 170W-120W); e iv) Tropical Central – Niño 4 (5N-5S; 160E – 150W). Por sua vez, o segundo<sup>5</sup> arquivo são dados com resolução 0,25° x 0,25° obtidos através de uma interpolação ótima composto de dois tipos: i) Atlântico Norte (5N-20N; 60W-30W); e ii) Atlântico Norte (0-20S; 30W-10E). A Tabela 1 apresenta os coeficientes de correlação de *Spearman* para os dados de precipitação e os dados de TSM.

Para fins desta pesquisa, utilizar-se-á os dados de TSM com maior correlação com os dados de precipitação. Logo, observa-se na Tabela 1 que o Niño 3.4 correspondente ao leste central tropical do Oceano Pacífico apresenta maior correlação com as duas bases de dados de precipitação.

**Tabela 1** – Correlação de *Spearman* entre dados de Precipitação de TSM

PRECIPITAÇÃO/TSM	OCEANO PACÍFICO				OCEANO ATLÂNTICO	
	Niño 1 + 2	Niño 3	Niño 3.4	Niño 4	Atlântico Norte	Atlântico Sul
<b>BMEP</b>	0,229 (0,000)*	0,0047 (0,673)	0,239 (0,000)*	0,0098 (0,377)	-0,461 (0,000)*	0,469 (0,000)*
<b>Rocha e Soares (2015)</b>	0,211 (0,000)*	-0,0211 (0,056)	0,219 (0,000)*	-0,019 (0,089)	-0,480 (0,000)*	0,472 (0,000)*

Fonte Elaboração dos Autores

\* Estatisticamente significativo a um nível de significância de 5%.

A escolha temporal do estudo foi de 1996 à 2010, para efeito de comparação com os resultados de Rocha e Soares (2015).

#### 4 RESULTADOS EMPÍRICOS

Ao analisar a relação entre a precipitação no início da vida e os resultados de saúde dos adultos – aqui expressos pela taxa de mortalidade infantil, consideram-se três importantes inovações desta pesquisa. Primeiro, busca-se testar a base de dados apresentados em Rocha e Soares (2015) com métodos mais robusto, baseado em Canay

<sup>4</sup> Ver detalhes são encontrados em Huang, et. al (2014) e Liu, et. al. (2014).

<sup>5</sup> Ver detalhes em Reynolds, et. al (2007).

(2011) e Harding e Lamarche (2009), considerando uma regressão quantílica de dados em painel para captar além dos efeitos fixos por município-mês e por ano, a heterogeneidade dos indivíduos por meio de regressão quantílica, como também tirar efeito da endogeneidade através de variáveis instrumentais. Trata-se, portanto, de regressão quantílica com dados em painel apresentado na Tabela 2.

Outro desafio central deste artigo é o fato da variável precipitação ser uma medida suscetível a erro. O nível de precipitação é obtido através de estações meteorológicas que muitas vezes possuem bases históricas sem continuidade. Como alternativa surge na literatura às chamadas séries de reanálise baseados em modelos matemáticos, como por exemplo, interpolação linear, entre outros. Contudo, esse tipo de dados incorre em erro de medição clássica da variável de precipitação superestimando tais valores, e, com vista a evitar tal problema, optou-se por dados observacionais obtidos através da BMEP, embora isso represente perda de quantidade de observações a serem analisadas. Tais evidências encontram-se na Tabela 3.

Terceiro, para fins de comparação, reduziu-se a amostra de Rocha e Soares (2015) para a mesma quantidade de municípios da base de dados BMEP com a finalidade de testar os efeitos das variáveis de precipitação considerando os mesmos municípios, mas variáveis de precipitação construídas a partir de séries diferentes. Conforme se verifica na Tabela 4.

Cabe aqui enfatizar que as Tabelas ora aqui apresentadas, possuem *layout* semelhante ao realizado em Rocha e Soares (2015). Além disso, adiciona-se que fora utilizado também as seguintes variáveis de controle: *i*) temperatura média dos últimos doze meses; *ii*) tendência temporal baseada na grade formada pelos municípios pertencentes ao quadrado 5° x 5° de latitude/longitude; *iii*) precipitação dos 13 a 24 meses antes do nascimento; *iv*) precipitação nos doze meses posteriores ao nascimento dos indivíduos. Não obstante, é relevante destacar que para cada painel construiu-se as variáveis de precipitação e seca, expressos nas equações (5) e (6) deste estudo, respectivamente.

Para este estudo, os dados foram coletados a partir de áreas geográficas dos quais as chuvas são bastante irregulares e o solo é raso. O semiárido brasileiro possui essas características, e embora exista ocorrência de chuvas, historicamente são permeados por longos períodos de secas, com características edafoclimáticas e socioeconômicas restritivas à população em razão de sua escassez. O saldo global dos efeitos sobre a saúde é provável que seja negativo, principalmente quando acometem indivíduos no período intrauterino ou nos primeiros anos de nascimento (BANCKER, 1998; HECKMAN, 2007), onde tendem a ser particularmente vulneráveis aos efeitos adversos.

As correlações entre as secas e os resultados de saúde dos indivíduos do semiárido brasileiro são objeto de estudo de Rocha e Soares (2015) ao utilizarem como *proxy* do resultado de saúde a taxa de mortalidade infantil dos municípios e a construção de uma variável de precipitação interpretada como uma porcentagem da precipitação média, ambas apresentados no Painel A da Tabela aqui analisada. Os autores concluíram que existe correlação estatisticamente significativa e negativa como se verifica na Coluna 1 da Tabela 2. Em termos de magnitude, isto representa que o aumento no nível de precipitação no período gestacional reduz em cerca de 3% a taxa de mortalidade infantil. Tais valores aumentam quando controlado pela temperatura média durante a gravidez (-5,48%) e nível de precipitação antes da concepção e depois do nascimento (-6,14%), conforme a Coluna (2) e (3), respectivamente.

**Tabela 2: Regressões Quantílicas para dados em Painel com Efeitos fixos. Mediana (quantil 0,50). Período de 1996-2010. Região do Semiárido do Nordeste**

	Taxa de Mortalidade Infantil								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	FE(RS)			FEQR			FEQRIV		
<b>Painel A – Precipitação</b>									
Precipitação antes do nascimento	-3,252*	-5,475***	-6,141***	1,088***	-2,082***	-3,020***	0,648***	-2,487***	-2,627***
	(1,775)	(1,507)	(1,989)	(0,129)	(0,134)	(0,141)	(0,134)	(0,132)	(0,133)
Precipitação 13-24 meses antes do nascimento			-1,488			-0,914***			-1,941***
			(2,218)			(0,127)			(0,126)
Precipitação 1-12 meses após o nascimento			-2,737			-2,879***			-1,981***
			(2,339)			(0,142)			(0,131)
<b>Painel B – Seca</b>									
Seca antes do nascimento	2,535	3,343***	3,728***	18,830***	2,668***	2,952***	18,810***	2,826***	2,858***
	(1,801)	(1,175)	(1,227)	(0,231)	(0,147)	(0,161)	(0,2358)	(0,152)	(0,156)
Precipitação 13-24 meses antes do nascimento			-0,980			-1,150***			-2,028***
			(2,154)			(0,125)			(0,127)
Precipitação 1-12 meses após o nascimento			-1,812			-2,658***			-1,794***
			(2,176)			(0,132)			(0,120)
Observações	182.458	180.659	168.267	182.458	182.458	169.946	182.458	182.458	169.946
Número de Municípios	1.048	1.037	1.037	1.048	1.048	1.048	1.048	1.048	1.048
Efeito fixo de município x mês	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Efeito fixo de ano de nascimento	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Temperatura antes do nascimento	Não	Sim	Sim	Não	Sim	Sim	Não	Sim	Sim

Fonte: Elaboração dos autores com base nas estimações.

Notas: Erros padrão robustos entre parênteses. \*\*\*p-valor < 0,01. \*\* p-valor < 0,05. \* p-valor < 0,10.

Variável dependente é a mortalidade infantil por cada 1000 nascimentos, calculado em nível municipal por mês de nascimento. Variável independente no Painel A é o desvio-log da precipitação nos últimos 12 meses e no Painel B uma *dummy* que indica seca nos últimos 12 meses. Todas as regressões incluem efeitos fixos de município por mês e efeitos fixos de ano de nascimento.

No entanto, os autores não consideram a heterogeneidade não-observada. Ainda de difícil identificação por não serem diretamente observáveis ou mensuráveis, não se pode negligenciá-la. Isso se deve ao fato de que como as estimativas se tratam de efeito médio, portanto, é plausível admitir que estas não retratem em sua totalidade a realidade dos 1.048 municípios analisados. Além do mais, a literatura epigenética, na condução dos estudos de interações gene-ambiente, ratifica a clara distinção entre os atributos adquiridos ao longo da vida e aqueles relacionados aos genes, muitas vezes negligenciados na literatura da economia da saúde.

Portanto, características não observáveis correlacionadas com a probabilidade de exposição a ambientes desfavoráveis poderiam ser responsáveis por uma parcela dos resultados de saúde. Outrora, condições inerentes aos indivíduos são particularmente associadas aos seus resultados de saúde independente do ambiente em que convive. Não obstante, é relevante considerar, também, que os indivíduos podem responder a choques externos como a seca de forma distinta entre eles. Além do mais, questões como distribuição geográfica e sazonalidade também dever-se-iam ser consideradas.

Uma alternativa para contornar estes problemas é o método de regressão quantílica com dados em painel proposto por Canay (2011) – denominado aqui de estimador FEQR, apresentados nas Colunas (4), (5), (6). Com vistas a esses resultados, frisa-se que os efeitos de tratamento desse modelo ratifica o efeito negativo entre precipitação e mortalidade infantil, contudo, em magnitude inferior ao apresentado em Rocha e Soares (2015).

Ao comparar-se os estimadores FE(RS) e FEQR, cabem aqui alguns apontamentos. Primeiro, verifica-se na coluna (3) que os controles utilizados não foram estatisticamente significativos corroborando para a conclusão de Rocha e Soares (2015) de que níveis de chuva em outros períodos que não seja no gestacional não afetam a mortalidade infantil. O segundo, diz respeito a implicação do primeiro, uma vez que afirmar que as variáveis de controle não exercem influência sob a mortalidade infantil seria concluir que nutrição materna no ano antes da concepção, bem como a nutrição e doenças no primeiro ano de vida não são afetados pelas chuvas. Tal conclusão vai justamente em oposição a literatura vigente (GLEWWE e KING, 2001; HECKMAN, 2007; CUNHA e HECKMAN, 2007, KUDAMATSU et. al., 2010).

Terceiro, tais evidências são refutadas com o estimador FEQR, mostrando que as variáveis de controle são estatisticamente significativas e negativas e a principal implicação é a redução da magnitude da variável resposta. Além do mais, verifica-se através da coluna (6) que o efeito da precipitação de 1-12 meses após o nascimento (-2,88) é maior do que o efeito da precipitação de 13-24 meses antes do nascimento (-0,914), corroborando para as evidências da literatura de que os primeiros anos de vida dos indivíduos são períodos sensíveis<sup>6</sup> e respondem mais precisamente a choques adversos (CUNHA e HECKMAN, 2007; HECKMAN, 2007). Por exemplo, privação nutricional no início da vida decorrentes das secas pode retrair indicadores antropométricos persistindo até a vida adulta (AGUILAR E VICARELLI, 2011).

Quarto, e não menos importante, é notório que a falta de rigor metodológico imprime um papel demasiadamente relevante para o clima nos resultados de Rocha e Soares (2015). Notem que as colunas (1), (2) e (3) são, em números, quase o dobro das colunas (4), (5) e (6). Em síntese, tem-se um efeito superestimado da precipitação, o que

---

<sup>6</sup> De acordo com Cunha e Heckman (2007) e Heckman (2007) mostram que etapas importantes no desenvolvimento físico dos indivíduos pode ocorrer na infância e ter consequências para os resultados de saúde quando adultos. São os chamados períodos sensíveis que segundo os autores decorrem da primeira infância que vai até os primeiro quatro anos de idade.

nos leva a concluir *a priori* que existe condição endógena entre as variáveis relacionadas à precipitação que afetam resultados de saúde dos indivíduos, muito embora seja consensual de que esse efeito é negativo.

Nesse cenário, outra alternativa metodológica, via de regra mais robusta, baseia-se no método proposto por Harding e Lamarche (2009) visando endogeneizar a quantidade de precipitações através de um modelo empírico de regressão quantílica de dados em painel com variáveis instrumentais, ora aqui chamado de estimado FEQRIV. Os resultados encontram-se nas Colunas (7), (8) e (9) da Tabela 2. A inserção de variáveis instrumentais dado que estas não pertencem a equação explicativa, mas está correlacionada com a variável de precipitação é aqui justificada pelos fenômenos oceano-atmosfera ENOS (RAO, LIMA e FRANCHITO, 1993) e do Dipolo Atlântico (ARAGÃO, 1998), expressos na variável Temperatura da Superfície do Mar (TSM).

Observa-se na Tabela 2 com os resultados controlados, coluna (9), intuitivamente tem-se as mesmas conclusões do estimador de Canay (2011) - FEQR, as magnitudes dos parâmetros diminuem em absoluto e as variáveis de controle são estatisticamente significativas. Enquanto que Rocha e Soares (2015) concluíram que aumentos de precipitação nos primeiros anos de vida pode reduzir em torno de 6% a taxa de mortalidade infantil, o estimador FEQRIV demonstra que seu poder redutor resume em apenas 2,6%, ou seja, menos da metade.

Sob outro prisma, o Painel B da Tabela 2 apresenta uma variante do nível de precipitação. Denota-se a variável seca uma *dummi* que quando assumido valor 1 significa que a precipitação ao longo dos 12 meses anteriores ao nascimento era mais do que um desvio padrão abaixo da média histórica. Como esta medida a interpretação se inverte, espera-se sinal positivo desta relação, conforme se verifica nos resultados dos três estimadores aqui utilizados. Contudo, análogo a análise anterior, constata-se que os resultados de Rocha e Soares (2015) também foram superestimados provavelmente pelos problemas de endogeneidade não contabilizados.

Em suma, apesar dos resultados entre os estimadores serem semelhantes qualitativamente, é relevante destacar que as evidências correspondentes apontam um trato demasiadamente elevado ao *status* clima nos resultados de saúde dos indivíduos. Ademais, suspeita-se que esse *quórum* superestimado dos parâmetros possa também ser decorrente do uso inapropriado de dados de precipitação. É consensual que dados históricos de chuvas e temperatura são boas medidas para verificar seu impacto sobre os resultados de saúde dos indivíduos, principalmente em regiões como a semiárida brasileira que é acometida por longos períodos de secas, e, que, conseqüentemente deve gerar ônus para sua população residente. Uma das justificativas para seu uso é que a variabilidade destas afetam diversas atividades econômicas (FREITAS, FRANCHITO e RAO, 2010), o que potencializa seu impacto, principalmente em regiões subdesenvolvidas e dependentes da agricultura (CIRILO, 2008).

Portanto, ao analisar a relação entre a precipitação no início da vida e os próprios resultados de saúde dos adultos, um desafio central é que os dados de precipitação utilizados na literatura são medidos com erro. Mas como isso ocorre? Infelizmente as séries históricas disponíveis decorrentes de estações meteorológicas possuem muitas lacunas em virtude de problemas de coleta e inexistência de estações em regiões de difícil acesso, com isso, para suprir tais dificuldades surgiram os dados de reanálise, também chamados de dados modelados, nos quais se utilizam informações de satélites e observacionais em modelos matemáticos para obter séries longas e de maiores alcances. Contudo, a contra gota é que essa medida contém erros que podem superestimar as variabilidades climatológicas estudadas.



**Tabela3: Regressões Quantílicas para dados em Painel com Efeitos fixos. Mediana (quantil 0,50). Período de 1996-2010. Região do Semiárido do Nordeste**

	Taxa de Mortalidade Infantil								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	FE(RS)			FEQR			FEQRIV		
<b>Painel A – Precipitação</b>									
Precipitação antes do nascimento	4,251 (4,872)	4,812 (5,089)	3,781 (6,407)	-3,985*** (0,940)	-4,122*** (0,981)	-3,931*** (1,020)	-4,421*** (0,924)	4,556*** (0,972)	-3,799*** (1,040)
Precipitação 13-24 meses antes do nascimento			-9,193 (8,857)			-2,971*** (1,116)			-2,830** (1,124)
Precipitação 1-12 meses após o nascimento			-8,997 (7,152)			-7,380*** (1,104)			-7,719*** (1,239)
<b>Painel B – Seca</b>									
Seca antes do nascimento	-5,619 (3,590)	-5,792 (3,701)	-5,722 (3,576)	10,923*** (1,469)	4,250*** (1,596)	4,088*** (1,353)	10,722*** (1,477)	4,351*** (1,454)	4,089*** (1,385)
Precipitação 13-24 meses antes do nascimento			-9,041 (8,727)			-2,868** (1,123)			-2,731** (1,126)
Precipitação 1-12 meses após o nascimento			-9,343 (6,663)			-7,523*** (1,149)			-7,523*** (1,178)
Observações	8.070	8.070	7.541	8.070	8.070	7.541	8.070	8.070	7.541
Número de Municípios	47	47	47	47	47	47	47	47	47
Efeito fixo de município x mês	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Efeito fixo de ano de nascimento	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Temperatura antes do nascimento	Não	Sim	Sim	Não	Sim	Sim	Não	Sim	Sim

Fonte: Elaboração dos autores com base nas estimações.

Notas: Erros padrão robustos entre parênteses. \*\*\*p-valor < 0,01. \*\* p-valor < 0,05. \* p-valor < 0,10.

Variável dependente é a mortalidade infantil por cada 1000 nascimentos, calculado em nível municipal por mês de nascimento. Variável independente no Painel A é o desvio-log da precipitação nos últimos 12 meses e no Painel B uma *dummy* que indica seca nos últimos 12 meses. Todas as regressões incluem efeitos fixos de município por mês e efeitos fixos de ano de nascimento.

Neste cenário, optou-se pelos dados observacionais obtidos no BDMEP, para testar se erro de medição clássica na variável precipitação no início da vida leva a estimativa dos coeficientes atenuados. Muito embora isto implique em redução significativa da amostra estudada, resumidas em 47 municípios analisados em uma amostra de 8.070 observações. Os resultados são apresentados na Tabela 3 – Painel A. As colunas (1), (2) e (3) referem-se ao coeficiente FE(RS), método utilizado em Rocha e Soares (2015), no qual todas as estimativas foram estatisticamente iguais a zero, ou seja, para esta amostra não há correlação entre as variáveis.

Entretanto, as colunas (4), (5) e (6) com o estimador FEQR apresentam-se estatisticamente significativo e negativo ratificando os resultados apresentados na Tabela 2. Análogo ocorre nas colunas (7), (8) e (9) da Tabela 3, considerando o estimador com variáveis instrumentais FEQRIV. As conclusões são: *i*) quanto maior a precipitação antes da concepção, durante o período gestacional e nos primeiros anos de vida dos indivíduos nascidos no semiárido menor a taxa de mortalidade infantil nesta região; *ii*) quando controlado a heterogeneidade não observada e a endogeneidade existente nesta relação, a precipitação tem peso menor indicando que existem outras variáveis que afetam os resultados de saúde; *iii*) o método utilizado em Rocha e Soares (2015) não é adequado para estimar essa relação. Sem embargo, o Painel B, que considera a variável seca, apresentam as mesmas conclusões sob ângulo oposto.

Para consolidar as conclusões aqui expostas, utilizou-se os dados de reanálise a *Terrestrial Air Temperature and Terrestrial Precipitation: 1900–2010 Gridded Monthly Time Series*, versões 3.01 e 3.02, respectivamente (Matsuura and Willmott, 2009) utilizados em Rocha e Soares (2015), considerando apenas os municípios do semiárido que tem informações na base de dados BDMEP, conforme apresentado na Tabela 4. Para efeitos desta análise, ignoram-se as colunas (1), (2) e (3), uma vez que constatada a conclusão de que o método simples de dados em painel não é adequado para estimar o impacto de precipitação nos resultados de saúde, conquanto aqui o eixo são os métodos de Canay (2011) e Harging e Lamarche (2009), ou seja, os estimadores FEQR e FEQRVI, respectivamente. Portanto, observam-se as colunas 4-9 das Tabelas 3 e 4.

A primeira conclusão, *a priori*, é de que de fato variáveis oriundas de reanálise tendem a superestimar os parâmetros desta relação. Notem que para cada coluna o parâmetro estimado com dados de reanálise são maiores do que aqueles estimados com dados de precipitação observacionais. No entanto, vale aqui destacar duas importantes reflexões. Primeiro, chama-se atenção para a Tabela 4 (dados de reanálise reduzida) que o controle precipitação de 13-24 meses antes do nascimento não é estatisticamente significativo nos dois estimadores, dando apenas importância para a precipitação após a concepção dos indivíduos. Parece que reduzir a amostra significativamente potencializa erro de medição da variável comprometendo os resultados.

A segunda reflexão se concentra na Tabela 3, no que tange a coluna 8, em que apresenta coeficiente positivo. Quando não controlado pelas variáveis de precipitação antes e depois do nascimento, o resultado é contraditório. Seria plausível pensar que como estamos em um número reduzido de municípios, talvez para os pertencentes o resultado fosse amplamente aceito, no entanto, a coluna (9) mostra ser conivente com a literatura e com os demais resultados aqui explorados de que aumento de chuvas no período de nascimento tendem a reduzir a taxa de mortalidade infantil.

**Tabela 4: Regressões Quantílicas para dados em Painel com Efeitos fixos. Mediana (quantil 0,50). Período de 1996-2010. Região do Semiárido do Nordeste**

	Taxa de Mortalidade Infantil								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	FE(RS)			FEQR			FEQRIV		
<b>Painel A - Precipitação</b>									
Precipitação antes do nascimento	-9,597** (4,554)	-10,277** (4,547)	-11,754* (6,292)	-5,333*** (0,906)	-4,086*** (0,942)	-4,008*** (1,033)	-5,786*** (0,897)	-4,597*** (0,958)	-4,259*** (1,020)
Precipitação 13-24 meses antes do nascimento			-4,336 (12,063)			0,028 (0,989)			0,271 (0,956)
Precipitação 1-12 meses após o nascimento			0,008 (10,295)			-1,908** (0,915)			-2,519*** (0,944)
<b>Painel B – Seca</b>									
Seca antes do nascimento	5,337 (4,302)	4,487 (4,463)	4,972 (4,760)	20,335*** (1,277)	3,810*** (1,253)	5,642*** (1,792)	20,436*** (1,361)	4,358*** (1,194)	5,517*** (1,683)
Precipitação 13-24 meses antes do nascimento			-0,552 (12,085)			0,257 (1,021)			0,338 (0,981)
Precipitação 1-12 meses após o nascimento			-8,863 (11,903)			-1,639* (0,906)			-1,770* (0,950)
Observações	8.518	8.518	7.942	8.518	8.518	7.942	8.518	8.518	7.942
Número de Municípios	48	48	48	48	48	48	48	48	48
Efeito fixo de município x mês	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Efeito fixo de ano de nascimento	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Temperatura antes do nascimento	Não	Sim	Sim	Não	Sim	Sim	Não	Sim	Sim

Fonte: Elaboração dos autores com base nas estimações.

Notas: Erros padrão robustos entre parênteses. \*\*\*p-valor < 0,01. \*\* p-valor < 0,05. \* p-valor < 0,10.

Variável dependente é a mortalidade infantil por cada 1000 nascimentos, calculado em nível municipal por mês de nascimento. Variável independente no Painel A é o desvio-log da precipitação nos últimos 12 meses e no Painel B uma *dummy* que indica seca nos últimos 12 meses. Todas as regressões incluem efeitos fixos de município por mês e efeitos fixos de ano de nascimento.

## 5 CONSIDERAÇÕES FINAIS

O objetivo do trabalho foi avaliar o impacto dos resultados pluviométricos e de períodos de seca na taxa de mortalidade infantil no semiárido nordestino, a fim de ratificar ou não os resultados encontrados por Rocha e Soares (2015). Para tanto, foi aplicada uma metodologia de dados em painel com efeitos fixos, dados em painel com efeitos fixos estimados por regressões quantílicas e dados em painel com efeitos fixos estimados por regressões quantílicas com uso de variáveis instrumentais, sendo esta última à temperatura média do oceano atlântico e pacífico.

Os resultados indicam um efeito negativo da precipitação e um efeito positivo da seca na taxa de mortalidade infantil, para os três métodos utilizados. No entanto, a utilização de regressões quantílicas e variáveis instrumentais revelam possíveis superestimações quando não se consideram tais estratégias. Dessa forma, os efeitos de tais fenômenos naturais na redução da mortalidade infantil existem, mas o seu impacto pode ser menor que o comumente aceito, sendo a maior parte do impacto devido a outros fatores, tais como as condições de saúde dos municípios, entre outros.

A conclusão é de que bebês nascidos nas zonas áridas que sofrem secas quando no útero de suas mães residentes no semiárido enfrentam risco maior de morte, especialmente se nasceram quando as secas comprometem os seus primeiros anos de vida e quando a escassez de água afeta a nutrição materna antes da concepção desses indivíduos. Trabalhos como o de Maccini e Yang (2009), Kudamtsu et. al (2010); Aguilar e Vicarelli(2011), Shah e Steberg (2013), Burgess et. Al (2014) e Rocha e Soares (2015) tem ressaltado os efeitos das questões climáticas nos resultados dos adultos, contudo, este estudo vem a contribuir ao ressaltar a importância de escolher métodos adequados que tratem os problemas existentes, bem como a escolha de dados confiáveis e que sejam fidedignos da relação causal a ser estudada.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, José M. Brabo; SOUZA, R. O.; CAMPOS, José Nilson B. Previsão da anomalia de temperatura da superfície do mar (tsm) no atlântico tropical, com a equação da difusão de temperatura. **Revista Climanalyse**, ano, v. 3, p. 163-172, 2003.

ANDREOLI, Rita Valéria; KAYANO, Mary Toshie. A importância relativa do atlântico tropical sul e pacífico leste na variabilidade de precipitação do Nordeste do Brasil. **Revista Brasileira de Meteorologia**, v. 22, n. 1, p. 63-74, 2007.

AGUILAR, Arturo; VICARELLI, Marta. El Nino and Mexican children: medium-term effects of early-life weather shocks on cognitive and health outcomes. **Cambridge, United States: Harvard University, Department of Economics. Manuscript**, 2011.

ARAGÃO, José Oribe Rocha. O Impacto do ENSO e do Dipolo do Atlântico no Nordeste do Brasil. **Bull. Inst. fr. études andines**, v. 27, n. 3, p. 839-844, 1998.

ARAÚJO COSTA, José. O FENÔMENO EL NIÑO E AS SECAS NO NORDESTE DO BRASIL. **EDUCTE: Revista Científica do Instituto Federal de Alagoas**, v. 3, n. 1, 2014.

BARKER, D. J. P. **Mothers, Babies and Health in Later Life**. Editora Churchill Livingstone, edição nº 2, ISBN-10: 0443061653, 1998.

BRAHMANANDA RAO, V.; DE LIMA, Marley C.; FRANCHITO, S. H. Seasonal and interannual variations of rainfall over eastern northeast Brazil. **Journal of Climate**, v. 6, n. 9, p. 1754-1763, 1993.

BEZABIH, Mintewab; DI FALCO, Salvatore; MEKONNEN, Alemu. **On the Impact of Weather Variability and Climate Change on Agriculture: Evidence from Ethiopia**. 2014.

BURGESS, Robin et al. The unequal effects of weather and climate change: Evidence from mortality in india. **Unpublished working paper**, 2013.

CANAY, Ivan A. A simple approach to quantile regression for panel data. **The Econometrics Journal**, v. 14, n. 3, p. 368-386, 2011.

CIRILO, José Almir. Public water resources policy for the semi-arid region. **estudos avançados**, v. 22, n. 63, p. 61-82, 2008.

CUNHA, Flavio; HECKMAN, James. **The technology of skill formation**. National Bureau of Economic Research, 2007.

CURRIE, Janet. **Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development**. National Bureau of Economic Research, 2008.

CURRIE, Janet et al. Something in the water: Contaminated drinking water and infant health. **Canadian Journal of Economics/Revue canadienne d'économique**, v. 46, n. 3, p. 791-810, 2013.

CURRIE, Janet; VOGL, Tom. **Early-life health and adult circumstance in developing countries**. National Bureau of Economic Research, 2012.

CURRIE, Janet; ROSSIN-SLATER, Maya. Weathering the storm: Hurricanes and birth outcomes. **Journal of health economics**, v. 32, n. 3, p. 487-503, 2013.

FREITAS, Ana Carolina Vasques; FRANCHITO, Sérgio Henrique; RAO, Vadlamudi Brahmananda. Análise dos dados de precipitação provenientes de diferentes fontes sobre a América do Sul, com ênfase no Brasil. **CLIMEP-Climatologia e Estudos da Paisagem, Rios Claros, PR, Brasil**, v. 5, n. 1, 2010.

GALVAO, Antonio F.; LAMARCHE, Carlos; LIMA, Luiz Renato. Estimation of censored quantile regression for panel data with fixed effects. **Journal of the American Statistical Association**, v. 108, n. 503, p. 1075-1089, 2013.

GLEWWE, Paul; KING, Elizabeth M. The impact of early childhood nutritional status on cognitive development: Does the timing of malnutrition matter?. **The world bank economic review**, v. 15, n. 1, p. 81-113, 2001.

GLUCKMAN, Peter. e HANSON, Mark. **The fetal Matrix: Evolution, Development, and Disease**. Cambridge: Cambridge University Press, 2005.

- GROSSMAN, Michael. The demand for health: a theoretical and empirical investigation. **NBER Books**, 1972.
- GUEDES, ROBERTO LAGE et al. Série temporal de precipitação mensal de Fortaleza, Brasil: Comparação entre observações e dados de reanálise do NCEP/NCAR. **Revista Brasileira de Meteorologia**, v. 20, p. 83-92, 2005.
- HARDING, Matthew; LAMARCHE, Carlos. A quantile regression approach for estimating panel data models using instrumental variables. **Economics Letters**, v. 104, n. 3, p. 133-135, 2009.
- HASTENRATH, S.; HELLER, L. Dynamics of climatic hazards in north-east Brazil. *Quartely Journal of the Royal Meteorological Society*, v. 110, p. 411-425, 1977.
- HECKMAN, James. The technology and neuroscience of skill formation. **Invest in Kids Working Group, Center for Economic Development, Partnership for America's Economic Success**, July, v. 17, 2007.
- KAYANO, Mary Toshie; RAO, Vadlamudi Brahmananda; MOURA, Antonio Divino. Tropical circulations and the associated rainfall anomalies during two contrasting years. **Journal of Climatology**, v. 8, n. 5, p. 477-488, 1988.
- KOENKER, Roger. Quantile regression for longitudinal data. **Journal of Multivariate Analysis**, v. 91, n. 1, p. 74-89, 2004.
- KUDAMATSU, Masayuki; PERSSON, Torsten; STRÖMBERG, David. Weather and Infant Mortality in Africa. 2012. Manuscrito não publicado.
- MACCINI, Sharon L.; YANG, Dean. **Under the weather: Health, schooling, and economic consequences of early-life rainfall**. National Bureau of Economic Research, 2008.
- MORAES NETO, João M.; BARBOSA, Marx P.; DE ARAÚJO, Alexandre E. Efeito dos eventos ENOS e das TSM na variação pluviométrica do semi-árido paraibano. **R. Bras. Eng. Agríc. Ambiental**, v. 11, n. 1, p. 61-66, 2007.
- MOURA, A. D.; SHUKLA, J. On the dynamics of the droughts in Northeast Brazil: observations, theory and numerical experiments with a general circulation model. *Journal Atmospheric Science*, v. 38, n. 12, p. 2653-2673, 1981.
- PINTO, LUCÍA IRACEMA CHIPPONELLI et al. Comparação de produtos de reanálise para a América do sul. **Revista Brasileira de Meteorologia**, v. 24, n. 4, p. 461-472, 2009.
- ROCHA, Rudi; SOARES, Rodrigo R. Water scarcity and birth outcomes in the Brazilian semiarid. **Journal of Development Economics**, v. 112, p. 72-91, 2015.
- SANTOS, CARLOS ANTONIO COSTA DOS; BRITO, JOSÉ IVALDO BARBOSA. Análise dos índices de extremos para o semi-árido do Brasil e suas relações com TSM e IVDN. **Revista Brasileira de Meteorologia**, v. 22, n. 3, p. 303-312, 2007.

SHAH, Manisha; STEINBERG, Bryce Millett. Could droughts improve human capital? evidence from india. **Unpublished manuscript, last modified**, 2012.

SILVA, Viviane BS et al. An improved gridded historical daily precipitation analysis for Brazil. **Journal of Hydrometeorology**, v. 8, n. 4, p. 847-861, 2007.

TRENBERTH, K. E. The definition of El Niño. **Bulletin of the American Meteorological Society**, v.78, n.12, p.2771-2777. 1997.

UGAZ, Jorge; ZANOLINI, Arianna. Effects of extreme weather shocks during pregnancy and early life on later health outcomes: the case of philippines? typhoons. 2011.

Xavier, T. M. B. S.; Xavier, A. F. S.; Silva-Dias, M. A. F.; Silva-Dias, P. L. Interrelações entre Eventos ENOS (ENSO), a ZCIT (ITCZ) no Atlântico e a Chuva nas Bacias Hidrográficas do Ceará, **Revista Brasileira Recursos Hídricos**, v. 8, n. 2, p. 111-126. 2003.