

Avaliando o impacto do financiamento governamental federal em saúde na eficácia da atenção primária: evidências para o Brasil mediante internações hospitalares *

Julyan Gleyvison Machado Gouveia Lins **
Tatiane Almeida de Menezes ***

Resumo

O objetivo deste trabalho é verificar, mediante uma estratégia de identificação causal, se choques positivos no financiamento governamental federal em saúde primária afetam negativamente a probabilidade de uso (internação) do sistema hospitalar devido a melhorias da saúde coletiva. Usaram-se dados do Brasil que, no que diz respeito a políticas de saúde, possui um sistema de atenção básica que é universal, público e gratuito. Os resultados encontrados sugerem, de forma geral, que maiores financiamentos federais da atenção primária geraram maiores médias nesse indicador nos pequenos municípios do país. Tal resultado evidencia um importante efeito de transbordamento desse financiamento nas internações do sistema hospitalar do país, por meio de um maior acesso a estes serviços por parte da população. Porém, esses achados também evidenciam uma ineficácia desses serviços em diminuir os repasses de indivíduos ao sistema de média e alta complexidade mediante ações de saúde preventiva de doenças.

Palavras-chave: Saúde pública, Financiamento governamental federal em saúde, Cuidados primários em saúde, Brasil.

Abstract

Evaluating the impact of federal government health financing on the effectiveness of primary care: evidence for Brazil through hospital admissions

The objective of this work is to verify, through causal inference, if positive shocks in federal government funding in primary health care negatively affect the probability of use (hospitalization) of the hospital system due to improvements in collective health. Data from Brazil, which, with regard to health policies, has a basic health care system that is universal, public and free, was used. The results largely suggest that greater federal funding of primary care generated higher averages for this indicator in small municipalities in the country. This result indicates that this financing resulted in a significant effect of overflows in hospitalizations in the country's hospital system, through greater access to these services by the population. However, the findings also show an ineffectiveness of these services to reduce the transfer of individuals to medium and high complexity care through preventive health care.

Keywords: Public health, Federal government health financing, Primary health care, Brazil.

JEL: I18, H51, O54.

Introdução

Muitos estudos sugerem que uma boa política de atenção primária de saúde poderia beneficiar a população porque garantiria uma maior proximidade aos serviços de saúde, reduzindo ou eliminando uma possível dificuldade de acesso a esses serviços; melhoraria a qualidade dos cuidados, aumentando o foco na prevenção e administrando problemas simples; e, por fim, diminuiria a

* Artigo recebido em 14 de janeiro de 2019 e aprovado em 15 de janeiro de 2021.

** Professor do Departamento de Economia da Universidade Federal da Bahia (PPGE-UFBA), Salvador, BA, Brasil. E-mail: julyanlink@hotmail.com. ORCID: <https://orcid.org/0000-0002-3207-6363>.

*** Professora do Departamento de Economia da Universidade Federal de Pernambuco (PIMES-UFPE), Recife, PE, Brasil. E-mail: tatianedemenezes@gmail.com. ORCID: <https://orcid.org/0000-0002-9737-6084>.

necessidade de cuidados de especialistas (Starfield; Shi; Macinko, 2005). Como destacam Baicker e Chandra (2004), uma boa cobertura de cuidados primários poderia resultar também em uma diminuição com gastos em saúde. Neste sentido, quanto melhores os cuidados básicos, menor seria o surgimento de doenças e a necessidade de internações, o que diminuiria as pressões de uso e os custos do sistema hospitalar.

O problema é que boa parte dos trabalhos empíricos que tentam validar essa proposição tendem a não usar estratégias empíricas de identificação causal, evidenciando apenas correlações que não permitem fazer inferência sobre os resultados, abrindo, assim, margem para que essa questão ainda não tenha um resultado claro. Além disso, é importante diferenciar claramente a existência de serviços de atenção primária da eficiência destes serviços, ou seja, se atendem de forma adequada todas as necessidades da população. Como é mais complexo mensurar a qualidade, a adequada cobertura e o acesso, responder esse tipo de pergunta se torna uma tarefa difícil.

Uma possível estratégia para respondê-la é utilizar-se do financiamento dessas ações, uma vez que é plausível supor que maiores financiamentos tenderão a melhorar os serviços de atenção primária, uma vez que há uma maior oferta de recursos econômicos. Dado o exposto, será que choques positivos no financiamento governamental em saúde primária afetam a qualidade dos serviços de saúde e implicam em alguma mudança na probabilidade de uso do sistema hospitalar devido a melhorias da saúde coletiva? Será que as regiões que recebem maiores financiamentos governamentais para execução de ações em saúde primária são mais saudáveis, ou seja, apresentam menores níveis de internações?

Segundo Starfield, Shi e Macinko (2005), os cuidados primários de saúde são caracterizados por um conjunto de ações que focam na adoção de medidas de assistência em nível familiar e comunitário, nos quais o processo de acompanhamento do paciente é contínuo no tempo e baseado em medidas de prevenção. Esses autores ainda destacam que a Organização Mundial de Saúde definiu em 1978 as quatro características essenciais desse modelo: *i*) é o primeiro contato ao acesso as necessidades de saúde por parte da população; *ii*) foca-se no cuidado personalizado; *iii*) dá ênfase à prevenção; *iv*) coordena os cuidados necessários quando os pacientes devem ser direcionados para unidades de tratamento de saúde mais complexas.

Neste contexto, as Condições Sensíveis a Atenção Primária (CSAP) são um conjunto de doenças passíveis de serem evitadas (ou administradas) pelos adequados cuidados primários de saúde, ou seja, são agravos cuja morbidade e mortalidade podem ser totalmente (ou parcialmente) minimizadas, pela presença de serviços efetivos de saúde da atenção primária. A ideia é que quando a atenção primária não garante acesso suficiente e adequado, ela pode gerar uma demanda excessiva para os níveis de média e alta complexidade de tratamento, implicando em fortes pressões de custo ao sistema hospitalar, além de deslocamentos desnecessários na alocação dos recursos para indivíduos em situação de saúde facilmente evitável (Starfield; Shi, 2002).

Segundo Macinko, Dourado e Guanais (2011), as taxas de internações por condições sensíveis a atenção primária são comumente usadas na avaliação empírica da efetividade dos cuidados básicos de saúde, de modo que, de forma geral, a literatura aponta uma relação negativa entre a efetividade da atenção primária e essas taxas de internações. Esses autores ainda destacam que o indicador tem sido tradicionalmente uma referência de avaliação do desempenho dos sistemas de

saúde em muitos países do mundo, justamente pelo fato de que essas internações podem capturar falhas no sistema de saúde, seja pela ineficiência das ações, seja pela pouca cobertura, ou acesso da população aos serviços prestados. Assim, uma vez que uma boa atenção primária deveria detectar a condição da doença antes de sua progressão, reduzindo sua severidade, prevenindo o aparecimento de complicações e evitando a necessidade de internação hospitalar, o indicador pode ser um importante instrumento para o planejamento das ações de saúde.

Dado que a atenção primária objetiva melhorar a saúde da população, via monitoramento da morbidade já instalada ou com práticas preventivas de saúde atreladas ao estilo de vida das pessoas, é muito provável que essa boa administração da saúde deva afetar a probabilidade de internação da população de uma determinada localidade (Dusheiko et al., 2011). Se práticas preventivas podem reduzir a probabilidade de aparecimento de doenças ou a progressão da enfermidade, então é de se supor que o procedimento de internação hospitalar deve ser menor em locais que apresentam um melhor desempenho da atenção primária. Uma outra suposição é que se há um aumento do financiamento dessas ações, é provável também que as internações aumentem, uma vez que uma das funções da atenção primária é descobrir e repassar pacientes em estado mais grave para os devidos cuidados no sistema hospitalar. Qual dos dois efeitos prevalece?

Deve-se salientar, no entanto, que tais ações de saúde apresentam um custo de financiamento. Em muitos países, o desembolso para o acesso aos cuidados primários é feito diretamente pelos pacientes, de forma privada. Em outros, o Estado pode fornecer a totalidade ou uma parte desses serviços gratuitamente (via financiamento por impostos ou contribuições sociais). Nesta segunda linha de oferta dos serviços, o objetivo deste trabalho é verificar até que ponto o financiamento governamental da atenção primária é um importante determinante da probabilidade de uso de internações por condições sensíveis a atenção primária no Brasil. No que diz respeito ao monitoramento e controle epidemiológico, tais grupos de enfermidade são de responsabilidade da atenção básica do país, que é pública e gratuita. Desse modo, busca-se verificar se tais choques financeiros afetam significativamente esse indicador hospitalar, por melhoria da saúde coletiva.

A justificativa para este trabalho tem como pano de fundo destacar políticas de saúde pública. Em parte considerável dos países do mundo, principalmente os pobres e em desenvolvimento, o Estado pode ter sérias dificuldades em oferecer serviços gratuitos de saúde a toda sua população. Este trabalho tem o objetivo de mostrar evidências empíricas da relevância em priorizar ações neste sentido, uma vez que isso pode representar muito em termos de melhoria das condições de vida, principalmente das populações mais necessitadas. Neste sentido, a literatura empírica recente tem evidenciado como modelos de intervenções em saúde primária apresentam uma importante efetividade na melhoria da saúde de muitas populações ao redor do planeta como destaca o trabalho de Bailey e Goodman-Bacon (2015), por exemplo. Assim, o trabalho poderá gerar um valioso conjunto de informação que incentive a elaboração de ações públicas de saúde ao redor do mundo.

O restante do ensaio é estruturado como segue: Na seção 1, têm-se as principais ações de atenção primária executadas no país e sua relação com as internações por condições sensíveis. Na seção 2, a apresentação do banco de dados e, na seção 3, a estratégia empírica adotada. Na seção 4 são apresentados os resultados do trabalho. Por fim, têm-se as considerações finais.

1 Atenção primária de saúde no Brasil: ações implementadas e sua relação com internações hospitalares

No Brasil, os cuidados primários públicos de saúde originam-se da Política Nacional de Atenção Básica (PNAB). Esta política é desenvolvida com alto grau de descentralização, primeiro com ações direcionadas aos municípios, segundo às comunidades e, por fim, às famílias. A base estrutural dessa política são as Unidades Básicas de Saúde (UBS), que são estruturas físicas e organizacionais dentro dos municípios. Essas unidades constituem o primeiro nível de acesso à saúde pública no país e são responsáveis pelo atendimento de uma população específica, em uma certa área geográfica. Elas ofertam serviços de forma integral e permanente às populações locais, além de realizarem atividades educativas para prevenção de doenças. Além disso, no intuito de minimizar a carência de estrutura pública de saneamento, coleta de lixo e a falta de acesso à água potável, por exemplo, também prestam assistência para minimizar diversos fatores de risco causadores de doenças.

No que diz respeito ao financiamento federal de ações primárias de saúde, até o ano de 2017, os recursos eram repassados diretamente da esfera federal para a municipal, por meio de conta bancária específica, denominadas de Bloco de Financiamento da Atenção Básica. Essa conta poderia receber duas modalidades de transferências: o Piso da Atenção Básica Fixo (PAB fixo) e o Piso da Atenção Básica Variável (PAB variável), que são transferências monetárias que representam a maior fonte de financiamento das ações que integram a atenção primária do Brasil¹. O nosso interesse aqui é no financiamento do PAB fixo que, como o nome diz, é uma verba fixa estabelecida por um piso (limite inferior) *per capita* anual de financiamento da saúde básica dos municípios.

Estes recursos são transferidos mensalmente, de forma regular e automática, do Fundo Nacional de Saúde aos fundos municipais e, atualmente, são quatro pisos a depender do município: R\$ 23,00, R\$ 24,00, R\$ 26,00 e R\$ 28,00 anuais. Esses valores foram definidos em Portaria no ano de 2013, de modo que este recurso constituísse a garantia de um financiamento mínimo *per capita* anual destinado a ações de saúde de nível primário em cada município do país². De forma geral, há maiores transferências *per capita* para municípios menos populosos, e isto é proposital no estabelecimento dos pisos. O objetivo é tornar mais igualitária a capacidade de financiamento da atenção primária local entre municípios pouco populosos (que tendem a ser mais pobres) e muito populosos (que tendem a ser mais ricos). As estatísticas descritivas do financiamento do PAB fixo estão expostas na Tabela A1 (Anexo) para os anos 2014-2016.

Resumidamente são duas as principais funções da atenção primária no Brasil: resolução dos agravos que são de sua competência e repasse ao sistema hospitalar, dos casos que não são. No primeiro, o caráter resolutivo da maior parte dos problemas de saúde da população (que são mais simples e fáceis de administrar) tem o intuito de desafogar os leitos hospitalares dos agravos de saúde

(1) No ano de 2014, por exemplo, dados do Fundo Nacional de Saúde mostram que o investimento federal total em atenção primária foi da ordem de 14,072 bilhões de reais, o que representou um investimento *per capita* de R\$ 69,25 ou de US\$ 29,47 dólares pela cotação média da moeda americana nesse ano. O PAB fixo representou cerca de um terço desse investimento.

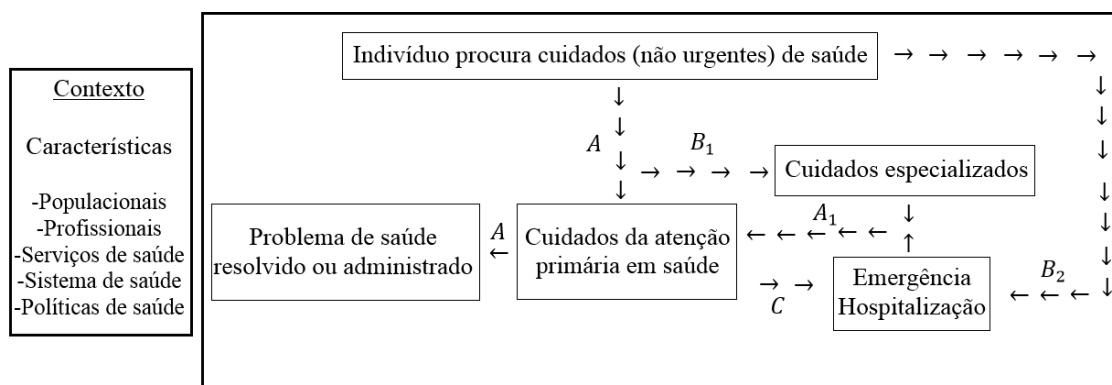
(2) Disponível em: http://bvsmis.saude.gov.br/bvsmis/saudelegis/gm/2013/prt1409_10_07_2013.html.

evitáveis. No segundo, tem a função de encaminhar aos sistemas de média e alta complexidade (hospitalar) os casos mais graves que exigem tratamento mais intenso, especialização médica e uso de uma maior complexidade biotecnológica.

Mas será mesmo que os cuidados primários da atenção básica estariam relacionados com hospitalizações? Como já apresentado, o principal pressuposto por trás do indicador de hospitalização evitável é que o acesso a cuidados de saúde primários de qualidade deve ter resultado na prevenção ou melhor gerenciamento de saúde, de modo a prevenir hospitalizações ou reduzir sua frequência. Mas como isso deve funcionar na prática? Quais seriam os caminhos que fazem com que a atenção primária (de boa qualidade ou não) afete as internações do sistema hospitalar?

Macinko, Dourado e Guanais (2011) nos mostram um relevante esquema explicativo, adaptado de Caminal e Casanova (2003), de todo esse processo. A figura 1, mostra que há diferentes fatores contextuais associados ao risco de hospitalização, que vão desde características populacionais até políticas de saúde.

Figura 1
Quadro conceitual da relação entre os cuidados primários de saúde e as internações hospitalares



Fonte: Adaptado de Caminal e Casanova (2003) e Macinko, Dourado e Guanais (2011).

Neste esquema, esses autores mostram que se um indivíduo procura o sistema de atenção primária por uma condição sensível não emergencial, o ideal é que seu problema seja resolvido (ou administrado) nos cuidados primários. Neste caso, a sequência na Figura 1, seria $A \rightarrow A$, especialmente se a localidade apresenta uma boa oferta de serviços. Uma outra situação é aquela em que ele procura atenção primária, mas devido às características de sua demanda de saúde, é preciso que seja encaminhado para o cuidado de especialistas ou o sistema de hospitalização. Assim, na sequência $A \rightarrow A_1$, como destacam os autores anteriormente citados, os cuidados primários têm um papel de coordenação e orientação dos pacientes das informações geradas pelo encontro com o especialista. Além disso, como espera-se que há uma melhor administração da doença, o sistema de emergência e hospitalização receberia menos pacientes.

Por outro lado, os autores mostram que as sequências B_1 e B_2 representam padrões de utilização de serviços de saúde nos quais a atenção primária não é facilmente acessível ou não desempenha de forma adequada seu papel como primeiro contato do sistema de saúde. Na sequência B_1 , os indivíduos procuram diretamente, e de maneira errônea, os cuidados de especialistas sem passar pela intermediação do nível primário (um clínico geral, por exemplo). Por outro lado, na sequência B_2 , os pacientes procuram de imediato o sistema hospitalar, sobrecarregando de maneira desnecessária esse nível de atenção. Por fim, a sequência C mostra que mesmo que os pacientes passem pela atenção primária, isso ainda pode resultar em uma via subótima. Nesse cenário, os serviços dos cuidados primários são ruins, há falta de capacidade desse nível de atenção de resolver os agravos que seriam de sua competência, o que faz com que o indivíduo seja repassado imediatamente para o sistema hospitalar, novamente sobrecarregando esse nível. Assim, todos os caminhos mostram que não é possível entender a dinâmica das internações hospitalares sem antes verificar qual o estado de eficácia dos cuidados primários da localidade.

2 Dados das internações por condições sensíveis a atenção primária no Brasil

A variável de internação foi construída por meio de dados do Sistema de Informações Hospitalares (SIH/DataSus) do Ministério da Saúde. Tais dados estão agregados em nível de município. O SIH do Brasil recebe informações, a nível de pessoa, sobre internações de hospitais públicos e privados conveniados ao SUS mediante formulários padrões denominados de AIH (Autorização de Informação Hospitalar). Nestes formulários estão incluídas informações por causa de internação (em CID-10), e esses dados cobrem a maior parte dos dados de hospitalizações do sistema de saúde do país.

Um ponto deve ser esclarecido no que diz respeito à delimitação da variável de interesse. Para se ter uma visão ampla do efeito causal de análise, é necessário utilizar um indicador de internação hospitalar que contenha diferentes dimensões de internações afetadas pela atenção primária. Porém, como destaca Alfradique (2009), o problema é que não há consenso na literatura internacional, com uma lista fechada, de quais as condições potencialmente causadoras de hospitalização que podem ser consideradas sensíveis a este nível.

Para contornar esse problema, o presente trabalho utilizará a Lista Oficial Brasileira de Internações por Condições Sensíveis a Atenção Primária. Esta lista foi estabelecida pelo Ministério da Saúde do Brasil na Portaria n. 221 (Ministério da Saúde), de 17 de abril de 2008³. O órgão elaborou uma lista própria, que expressa a realidade brasileira, e que pode servir de instrumento de avaliação de atenção básica em saúde pelo poder público. A lista é composta por 19 diferentes grupos de causas, com mais de 70 diagnósticos classificados pelo CID-10, e está na Tabela 1.

(3) Disponível em: http://bvsmis.saude.gov.br/bvsmis/saudelegis/sas/2008/prt0221_17_04_2008.html.

Tabela 1
Lista brasileira de internações por condições sensíveis a atenção primária

Grupo de diagnóstico	CID 10
1. Doenças preveníveis por imunização.	A37; A36; A33-A35; B26; B06; B05; A95; B16; G00.0; A17.0; A19; A15.0; A15.3; A16.0-A16.2; A15.4-A15.9; A16.3-A16.9, A17.1-A17.9; A18; I00-I02; A51-A53; B50-B54.
2. Gastroenterites infecciosas e complicações.	E86; A00-A09.
3. Anemias.	D50.
4. Deficiências nutricionais.	E40-E46; E50-E64.
5. Infecções de ouvido, nariz e garganta.	H66; J00-J03; J06; J31.
6. Pneumonias bacterianas.	J13-J14; J15.3, J15.4; J15.8, J15.9; J18.1.
7. Asma.	J45-J46.
8. Doenças pulmonares.	J20-J21; J40-J44; J47.
9. Hipertensão.	I10-I11.
10. Angina.	I20.
11. Insuficiência cardíaca.	I50; J81.
12. Doenças cerebrovasculares.	I63-I67; I69, G45-G46.
13. Diabetes <i>melitus</i> .	E10.0-E10.1; E11.0-E11.1; E12.0-E12.1; E13.0-E13.1; E14.0-E14.1; E10.2-E10.8; E11.2-E11.8; E12.2-E12.8; E13.2-E13.8; E14.2-E14.8; E10.9; E11.9; E12.9-E13.9; E14.9.
14. Epilepsias.	G40-G41.
15. Infecção no rim e trato urinário.	N10-N12; N30; N34; N39.0.
16. Infecção da pele e tecido subcutâneo.	A46; L01-L04; L08.
17. Doença inflamatória dos órgãos pélvicos femininos.	N70-N73; N75-N76.
18. Úlcera gastrointestinal.	K25-K28; K92.0-K92.2.
19. Doenças relacionadas ao pré-natal e parto.	O23; A50; P35.

Fonte: Portaria SAS/MS n. 221, de 17 de abril de 2008 (Ministério da Saúde do Brasil).

As variáveis de internação hospitalar são definidas pela taxa de internação por essas doenças para cada grupo de 10 mil habitantes, e as estatísticas descritivas dessas variáveis, por sua vez, estão exposta na Tabela A2 (Anexo), que reporta os valores para os anos 2014-2016 (que é o horizonte temporal de análise deste trabalho). Ela exhibe a distribuição das internações (por percentis) tanto para o Brasil como um todo, quanto para os pequenos municípios do país, ou seja, as localidades com menos de 50 mil habitantes e que representam a maior parte dessas unidades administrativas (cerca de 90% do total) e detêm próximo de um terço da população do país.

No período 2014 a 2016, segundo dados do SIH/DataSus, o sistema hospitalar do Brasil apresentou um total de 34,94 milhões de internações dos mais variados motivos, resultando em uma média de cerca de 11,65 milhões de internações por ano. As internações por condições sensíveis à atenção primária, por sua vez, representaram um total de 4,42 milhões de internações (12,67% do total de internações do período), nos dando uma média anual de cerca de 1,47 milhões de internações. Além disso, a incidência (medida em casos para cada grupo de 10 mil habitantes) variou

substancialmente entre o porte populacional dos municípios. No período 2014-2016, por exemplo, a incidência apresentou uma média de cerca 60 internações naqueles municípios com até 50 mil habitantes, 52 internações para os municípios entre 50 mil e 100 mil habitantes, 43 internações nos municípios que tem entre 100 mil e 500 mil habitantes; e por fim, foi de cerca de 45 internações para os municípios que tem mais de 500 mil habitantes. O que se percebe, é que à medida que o porte populacional do município aumenta, tendem a diminuir as taxas de internação por condições sensíveis, embora acima de 100 mil habitantes não seja possível diferenciar tamanho de cidade e internações, como feito na comparação com os municípios menores de 100 mil habitantes, uma vez que se torna bem maior a heterogeneidade entre os tamanhos populacionais das unidades.

Esse é um resultado, de certa forma, esperado. No Brasil, devido à estrutura organizacional e hierárquica do Sistema Único de Saúde, os serviços hospitalares de média e alta complexidade (cujos tratamentos necessitam de mais intensidade biotecnológica) tendem a ser regionalizados e focados nas grandes cidades do país, o que pode explicar boa parte dessa concentração de internações sensíveis nos pequenos municípios. Além disso, um outro fator que pode gerar essa diferença é a capacidade do aporte financeiro dos municípios nas ações de saúde. Como os municípios pequenos tendem a não possuir grande capacidade de arrecadação tributária própria, isso se traduz em uma maior dificuldade financeira de sustentar uma boa cobertura de atenção primária, e isso inevitavelmente pode se refletir em maiores internações por esta causa nestas localidades. Aliado a isso, existem outras características que estão associadas ao tamanho do município, que são as melhores oportunidades (em âmbito educacional, emprego, renda, estrutura da oferta de serviços de saúde, etc.) e que inevitavelmente influenciam no indicador de internação.

Por esse motivo, este trabalho foca-se no efeito do maior financiamento da atenção básica (por meio do PAB fixo) nas menores localidades do país, precisamente os pequenos municípios (que têm menos de 50 mil habitantes). Estes totalizam cerca de 90% dessas unidades administrativas no país e detêm próximo de um terço da população do Brasil. Como destacam Brollo et al. (2013), tais localidades tendem a ter baixa capacidade de arrecadação tributária (ou seja, recursos próprios) para financiar seus serviços públicos. Este fato, faz com que seja possível verificar de maneira mais clara os efeitos das transferências federais do PAB fixo, dado que é uma amostra mais homogênea em várias dimensões sociais e econômicas. Além disso, incluir médios e grandes municípios na amostra (que tendem a ter uma maior capacidade de investir em serviços de saúde com receitas próprias) poderia mascarar o efeito do maior financiamento do PAB fixo sobre as internações locais.

Por fim, vale salientar que as internações por condições sensíveis foram computadas apenas para os residentes, ou seja, os indivíduos que moram na localidade, dado que são estes que recebem a atenção primária local e são suscetíveis aos efeitos do maior financiamento do PAB fixo.

3 Estratégia empírica

Como será mostrado, a estratégia empírica deste trabalho atacará duas possíveis fontes de endogeneidade que podem enviesar a resposta da pergunta de interesse: tais elementos são a variável omitida e variável omitida espacial. O trabalho se baseia no de Gibbons, Overman e Patachinni (2015) e parte da seguinte especificação (na forma reduzida) para entender o processo em que há estas duas fontes de viés:

$$Y_i = x'_i\gamma + \mu_i + m(Y, j)_i\beta + m(x, j)'_i\theta + m(v, j)'_i\lambda + \varepsilon_i \quad (1)$$

Na equação acima, Y , que é a variável dependente, será função de um conjunto de características observáveis (x') e de características não observáveis (μ)⁴ para cada indivíduo i . Partindo da argumentação de que Y é derivado de um processo em que há dependência espacial, esta variável também será função de características (observáveis e não observáveis) da vizinhança j , de cada i . Estas características podem ser decompostas em três partes: em $m(Y, j)$ têm-se a média de Y na vizinhança j ; em $m(x, j)'$ têm-se o valor médio das características do vetor x' dessa vizinhança; e, em $m(v, j)'_i$, têm-se as características não observadas da vizinhança, ou seja, variáveis não mensuráveis, mas que são correlacionadas no espaço geográfico. Por fim, o termo de erro aleatório $\varepsilon \sim N(0,1)$.

Em modelos que não incorporam processos de dependência espacial, $m(Y, j)$, $m(x, j)'$, $m(v, j)'$ recairão no termo de erro, introduzindo viés de endogeneidade por variável omitida espacial. Além disso, sem estratégias de identificação, μ_i também recairá no termo de erro, induzindo, por sua vez, a viés de endogeneidade por variável omitida. Dado o exposto, as próximas subseções mostrarão, em detalhes, como será realizado o tratamento para cada um desses problemas.

3.1 Atacando o viés de endogeneidade oriundo de variável omitida: o instrumento

O trabalho tem o objetivo de testar a hipótese de que choques positivos no financiamento governamental federal da atenção primária de saúde se refletem em uma variação negativa nas taxas de internações por condições sensíveis a atenção básica. Como acima discutido, é plausível supor que variáveis não observáveis como a qualidade das instituições municipais e a competência dos gestores possam vir a influenciar tanto os recursos que os municípios destinam à saúde como sua capacidade de investimento. Como tais variáveis não são passíveis de serem observadas pelo pesquisador surge o problema de variáveis omitidas que tornam o tradicional Estimador de Mínimos Quadrados Ordinários (OLS) viesado. Para contornar esse problema emprega-se o estimador de Variável Instrumental (IV). A estratégia de identificação consiste em utilizar o PAB fixo teórico (definida pela Portaria do PAB fixo n. 1.409⁵) como instrumento para o PAB que de fato é repassado ao município.

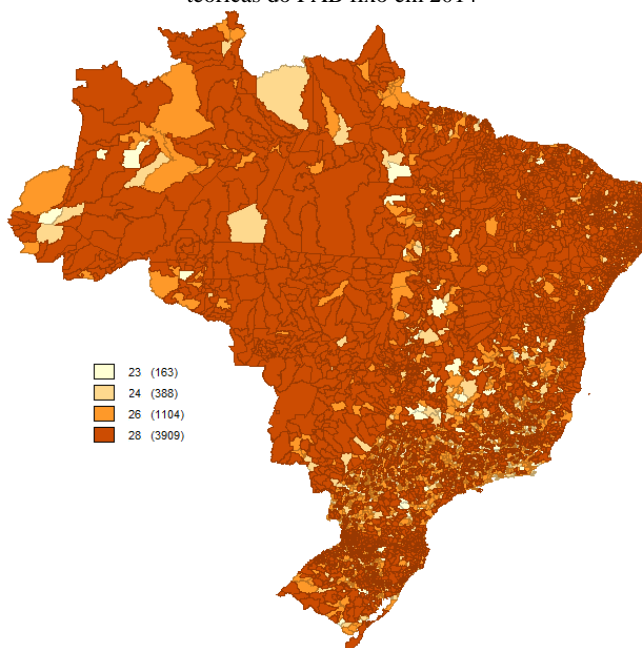
Já foi explicitado anteriormente que a Portaria que determinou os atuais repasses do PAB fixo definiu uma transferência mínima *per capita* para os municípios do país. Para isso, ela estabeleceu um montante de verba, que baseado na população municipal estimada do ano anterior à Portaria (2012), determinou cada um dos quatro pisos. Do ponto de vista do critério populacional, quanto menor a população, mais o município deve receber em termos *per capita*. Esse padrão é proposital no estabelecimento dos pisos. O objetivo é tornar mais igualitária a capacidade de financiamento da atenção primária local entre municípios pouco populosos (que tendem a ser mais pobres) e muito populosos (que tendem a ser mais ricos).

(4) Ao imputarmos uma dimensão temporal ao modelo, $\mu = \mu_F + \mu_V$, em que μ_F é um componente não observável fixo no tempo, enquanto μ_V é um componente não observável variante no tempo.

(5) Disponível em: http://bvsm.s.saude.gov.br/bvs/saudelegis/gm/2013/prt1409_10_07_2013.html.

No entanto, a transferência *per capita* que de fato ocorre não condiz com a que foi estabelecida na Portaria. O motivo é que, ao longo dos anos posteriores ao documento, enquanto o montante destinado a cada município não mudou em termos nominais, a população mudou e, conseqüentemente, o valor *per capita* mudou. Enquanto pela Portaria um determinado município deveria receber em 2015, por exemplo, R\$ 26,00 *per capita*, ele de fato recebe um pouco mais ou um pouco menos, a depender se sua população diminuiu ou aumentou entre a data da Portaria (ano de 2013) e o recebimento da verba (anos de 2014 em diante). Assim, o que temos são transferências efetivas do PAB fixo, que mudam em termos *per capita* dependendo da dinâmica populacional de cada localidade ao longo dos anos⁶. As figuras 2 e 3 exemplificam bem o que expomos. Na Figura 2, têm-se a distribuição geográfica do repasse das transferências teóricas garantidas pelo Ministério da Saúde. Por outro lado, na Figura 3, têm-se as transferências efetivas direcionadas aos municípios no ano de 2014, o que mostra uma clara distinção entre os dois tipos de PAB fixo.

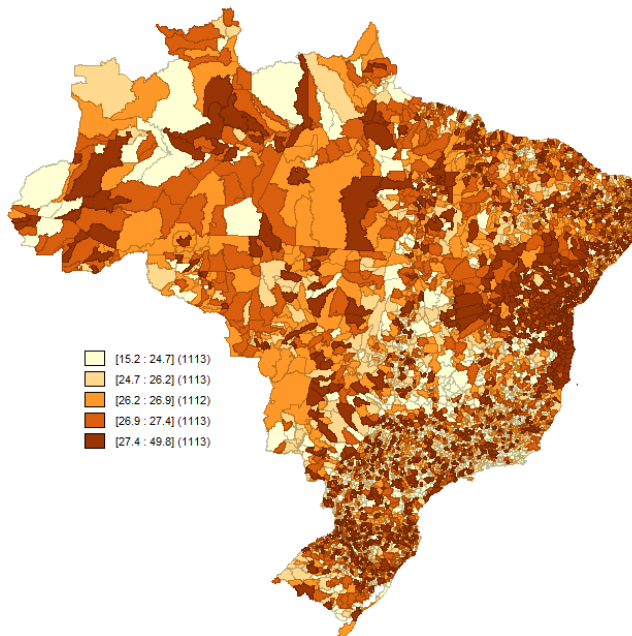
Figura 2
Distribuição espacial das transferências
teóricas do PAB fixo em 2014



Fonte: Portaria do PAB fixo.

(6) Vejamos um exemplo prático para o entendimento dessa dinâmica. São Paulo é a maior cidade do país e em 2012 possuía 11.376.685 habitantes. Em 2013, o governo federal estabeleceu o repasse do PAB fixo em R\$ 261.663.755,04, que enquadrava a cidade no piso de R\$ 23,00 *per capita*, baseando-se na população estimada do ano anterior. No ano de 2014, que foi o primeiro ano de repasse das transferências completamente por essa nova regra, o valor total da transferência não foi alterado em termos nominais, mas a população passou a ser de 11.967.825 (um aumento de 5,19% em relação a 2012). Isso fez com que a transferência efetiva *per capita* do PAB fixo fosse de R\$ 21,86, ou seja, um valor menor do que foi estabelecido pela Portaria. Isso sem levar em conta o efeito da inflação.

Figura 3
Distribuição espacial das transferências
efetivas do PAB fixo em 2014



Fonte: Dados do Fundo Nacional de Saúde.

Dado o exposto, para a estratégia de identificação por variável instrumental, o trabalho utilizará as transferências teóricas *per capita* do PAB fixo (pisos de R\$ 28,00, R\$ 26,00, R\$ 24,00 e R\$ 23,00) como instrumento para extrair a endogeneidade das transferências efetivas *per capita* do PAB fixo. Como já exposto, os municípios não recebem o PAB fixo teórico *per capita* (o piso estabelecido na Portaria), e sim o PAB fixo efetivo *per capita*. Isto é decorrente da diferença populacional entre a população usada na definição do piso (2012) e a população corrente dos municípios a partir de 2014 (ano em que se iniciaram os repasses totais pela Portaria), dado que o montante financeiro transferido do PAB fixo manteve-se fixo em termos nominais. Como as duas variáveis estão fortemente correlacionadas, do ponto de vista da estratégia de identificação, o PAB fixo teórico só tem a capacidade de afetar os indicadores de saúde locais, via definição do PAB fixo efetivo. Assim, há um claro instrumento: o PAB fixo teórico *per capita* (piso) apenas afeta a atenção básica do município via a influência que exerce na definição do tamanho do PAB fixo efetivo *per capita*.

3.2 Identificação causal: a estimação por variável instrumental e a eliminação da endogeneidade por variável omitida

O objetivo do trabalho é verificar o efeito das maiores transferências do PAB fixo efetivo (*PABE*) sobre as internações municipais (*I*). Dada a simples especificação econométrica:

$$I_{it} = \beta_0 + \beta_1 PABE_{it} + \gamma_t + \tau_r + \phi X_{it} + \varepsilon_{it} \quad (2)$$

Têm-se que γ_t representa efeito fixo de ano, τ_r efeito fixo de macrorregião e X uma matriz de variáveis de controle. Por fim, ε é o termo de erro aleatório, em que $\varepsilon_{it} \sim N(0,1)$. Na estimação da equação 2 feita por OLS, o coeficiente β_1 capturaria o impacto que o maior financiamento da saúde resulta no indicador de internação e representa o efeito médio de interesse. Mas, o principal problema acima é que $E(\varepsilon|PABE) \neq 0$, e isso exige uma estratégia de identificação de modo a garantir variações exógenas nessas transferências. Uma vez que já foi explicitada a estratégia de identificação por Variável Instrumental, passemos a ver como ela se desenvolve.

A estimação do IV é desenvolvida em dois estágios. No primeiro estágio, o regressor endógeno de interesse é regredido em relação ao instrumento. Obtêm-se o valor previsto dessa regressão, e em seguida essa nova variável entra como regressor no segundo estágio. A ideia do método é que dada a exogeneidade do instrumento, ao se obter o valor previsto da regressão do primeiro estágio, obtêm-se também uma variação exógena do regressor de interesse, uma vez que a partir disso, ele passa a ser função apenas do instrumento. Duas coisas são importantes nesse método: a primeira, como já foi dito, é a exogeneidade da Variável Instrumental. A segunda, por outro lado, diz respeito ao poder de previsão do instrumento no regresso endógeno. Nesse caso, é preciso que grande parte da variação do regressor de interesse seja atribuída ao instrumento, ou seja, precisamos de um alto coeficiente de determinação (conhecido como R^2) no primeiro estágio.

A especificação do nosso modelo IV pode ser vista nas equações 3 e 4. Com apenas um instrumento, o IV é equivalente a um OLS em dois estágios: no primeiro estágio (equação 2), a variável $PABE_{it}$, que é o PAB fixo efetivo do município i no ano t , é instrumentalizada pela variável $PABT_{it}$, que corresponde ao PAB fixo teórico destinado ao município. No segundo estágio (equação 3), por sua vez, a variável I_{it} , que corresponde ao nosso indicador hospitalar, é regredida em relação ao valor previsto do PAB fixo efetivo do primeiro estágio:

$$I_{it} = \beta_0 + \beta_1 \widehat{PABE}_{it} + \gamma_t + \tau_r + \phi X_{it} + \varepsilon_{it} \quad (3)$$

$$PABR_{it} = \delta_0 + \delta_1 PABT_{it} + u_{it} \quad (4)$$

Em ambas as equações, temos que γ_t representa efeito fixo de ano, τ_r efeito fixo de macrorregião e X uma matriz de controles. Por fim, μ e ε são os termos de erro aleatório, em que $u_{it}, \varepsilon_{it} \sim N(0,1)$. Na equação 3, o coeficiente β_1 representa o efeito médio que uma unidade monetária adicional do PAB fixo tem na taxa de internação hospitalar.

3.3 Atacando o viés de endogeneidade oriundo de variável omitida espacial: o modelo de efeitos fixos espacial

Algo que o modelo apresentado na seção anterior pode não comportar, são questões relativas às interações dos comportamentos sociais expressos no espaço geográfico. Em sistemas públicos universais, como o SUS, é natural haver uma concentração da estrutura hospitalar de média e alta complexidade em municípios de maior porte como uma medida de obtenção de ganhos de escala e diminuição dos custos médios por paciente. Assim, estes serviços tendem a ser regionalizados, visando a concentração do volume de leitos, o que explica a relevância da regionalização na análise do sistema. Por exemplo, um problema que pode surgir na presente estratégia de identificação é a

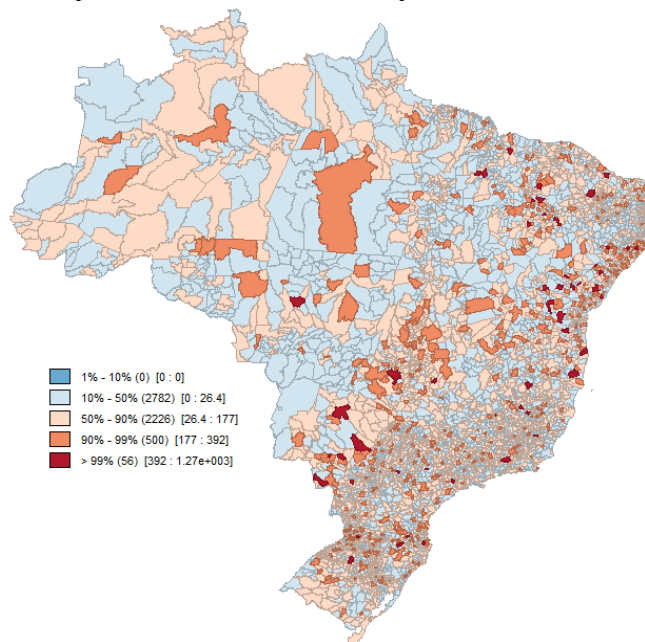
questão dos fluxos de migração para internações hospitalares. No Brasil, devido à carência dos serviços hospitalares (nos pequenos municípios, principalmente), a população dessas localidades tende a se internar em municípios próximos de sua residência (em geral, centros urbanos maiores). Assim, parte do total das internações municipais pode ser de indivíduos não residentes da sede do hospital e, neste caso, estas pessoas não recebem a atenção básica da região em que se internam, gerando um viés de variável omitida espacial, uma vez que o nível das internações locais é função também de características da vizinhança do município.

Com essa constatação, é possível admitir que há um notório efeito sistemático das características da vizinhança nas taxas de internações locais de um determinado município i , gerando um problema de autocorrelação espacial no modelo. A justificativa é a seguinte: se essa pequena localidade está perto de outros municípios (vizinhança j) com melhor estrutura hospitalar ou qualquer outro fator atrativo, é plausível supor que i tenderá a ter menor nível de internação de residentes, uma vez que a população terá menos custos associados à procura do leito desejado no município vizinho mais estruturado, independente de qual seja o tamanho do financiamento do PAB fixo efetivo. Se isso ocorre de forma sistemática, o nível das internações locais será função das características determinantes das internações na vizinhança, ou seja, os pacientes podem escolher o local de internação por meio de características observadas e não observadas da vizinhança. E este efeito de proximidade geográfica de outros municípios não pode ser deixado de lado na análise econométrica.

Outra possível fonte de endogeneidade decorrente de dependência espacial é a distribuição geográfica do PAB fixo efetivo. Embora a estratégia de identificação do IV permita que no segundo estágio o PAB fixo efetivo seja função apenas do PAB fixo teórico, devido a questões de interações espaciais entre municípios vizinhos, ainda pode haver variáveis omitidas espaciais correlacionadas com esse regressor e que também determinam os níveis de internações locais. Se, por exemplo, ao longo dos anos a vizinhança passa a receber um PAB efetivo *per capita* maior devido a uma queda da população (por algum motivo não observável), esse fator provavelmente estará correlacionado com as transferências efetivas *per capita* também do município vizinho (devido também à sua mudança populacional), uma vez que muitos fenômenos sociais e econômicos variam conjuntamente no espaço (as ligações estabelecidas pelas localidades próximas são endógenas), gerando novamente viés de endogeneidade no regressor de interesse.

A fim de exemplificar como o efeito da autocorrelação espacial das internações hospitalares de interesse não pode ser descartado no Brasil, basta ver a Figura 4. Pode-se constatar a distribuição espacial das taxas de internações por condições sensíveis a atenção primária no ano de 2014. Nota-se que as internações não estão dispersas pelo espaço geográfico, mas tendem a ter um padrão de regiões com altos índices (regiões Sudeste, Sul e Nordeste) e regiões com baixos índices nesse indicador (partes do Centro-Oeste, Nordeste e Norte). No modelo especificado nas equações 3 e 4, essa observação traz luz ao problema de variável omitida espacial, pois a especificação não leva em consideração o efeito de variáveis omitidas espaciais da região de entorno de cada município, que acaba recaindo nos resíduos do modelo e gerando viés de endogeneidade.

Figura 4
Distribuição espacial (por percentis) da taxa de internações hospitalares por condições sensíveis
(por 10 mil habitantes) nos municípios do Brasil (2014)



Fonte: Dados do Ministério da Saúde.

O problema é que se fosse possível incluir todos os regressores das características da vizinhança (para cada município), isso inviabilizaria a estimação devido à falta de graus de liberdade, dada a quantidade imensa de variáveis explicativas que seria preciso incluir na equação. Pior ainda é que mesmo que o problema de graus de liberdade não existisse, ainda poderiam existir variáveis relevantes espaciais que não são observáveis, ou seja, que não podem ser mensuradas, o que novamente faria o modelo cair em viés de endogeneidade.

Por esses argumentos, o modelo IV anteriormente descrito não é capaz de resolver o problema de viés de variável omitida espacial, ou seja, ele parte do princípio de que questões espaciais não são relevantes ou não afetam os resultados estimados, o que no caso deste trabalho é uma suposição forte. Embora a técnica de Variável Instrumental resolva o problema de endogeneidade de origem não espacial, a mesma técnica pode não garantir que existam variáveis omitidas espaciais no termo de erro correlacionadas com este regressor.

Para testar a existência de dependência espacial das internações por condições sensíveis nos municípios do Brasil, pode ser estimado um Modelo SAR. Este modelo inclui, na equação de regressão da equação 2, a defasagem espacial da variável dependente (WI), ou seja, a média da vizinhança dessa variável. No caso do presente trabalho teria-se a seguinte especificação:

$$I_{it} = \beta_0 + \rho WI_{it} + \beta_1 PABE_{it} + \gamma_t + \tau_r + \phi X_{it} + \varepsilon_{it} \quad (5)$$

Esta nova variável WI foi construída usando de técnicas estatísticas da Econometria Espacial, notadamente a matriz de vizinhança, também conhecida como matriz de pesos espaciais, representada por W . Gibbons (2004) destaca que a matriz de ponderação captura a interação de interesse entre as

regiões distintas e, como é apontado também por Tyszler (2006), essa matriz de peso espacial pode ser uma boa medida de representação da estrutura espacial e da interação social ou econômica que se busca avaliar em um trabalho empírico. Por outro lado, o coeficiente ρ mede o grau de transbordamento espacial das interações da vizinhança de i na taxa de crescimento desta região. LeSage e Fischer (2008) argumentam que a especificação do modelo SAR pode ser útil para capturar a dependência espacial, diminuindo o problema de endogeneidade, resultado de variáveis omitidas espaciais.

Segundo Elhorst (2001), a estimação da equação 5 por OLS é inadequada, uma vez que havendo autocorrelação espacial na variável dependente, o estimador de mínimos quadrados mantém-se enviesado e perde consistência. Tradicionalmente, os trabalhos empíricos tendem a estimar esse tipo de modelo por Máxima Verossimilhança (ML)⁷. No entanto, embora a estimação por ML permita a correção da endogeneidade mecânica ocasionada pela utilização da defasagem espacial da variável dependente (Anselin, 1988), essa técnica apresenta o inconveniente de que o parâmetro ρ não pode ser interpretado de forma tradicional (Lesage e Fischer, 2008). Neste caso, ρ não representa a influência da variável explicativa sobre a variável dependente, e isso ocorre devido ao efeito de realimentação de choques entre regiões vizinhas⁸. Um outro ponto importante é que, em geral, impõe-se uma restrição de que $|\rho| < 1$, ou seja, que os efeitos de transbordamento diminuam à medida que há afastamento do epicentro do choque. Por isso, é comum utilizar-se de uma matriz de ponderação espacial normalizada na linha.

Uma outra forma de testar a existência de dependência espacial do modelo da equação 2 é a utilização da estatística I de Moran (1948). Se há existência de dependência espacial e essa dependência não é modelada, ela cairá nos resíduos e estes tenderão a ter algum padrão de autocorrelação espacial. Em geral, o que os trabalhos empíricos fazem é estimar seus modelos sem a incorporação de variáveis espaciais e verificar se os resíduos dessa especificação apresentam alguma autocorrelação geográfica, mediante o I de Moran. Esta estatística testa a hipótese de que alguma variável apresenta um padrão de associação linear espacial. A estatística é apresentada abaixo:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2} \quad (6)$$

I é calculado mediante uma medida de autocovariância, variância e uma matriz de ponderação espacial, que pode ser normalizada na linha. Nesta estatística, n é o número de observações espaciais, z_i é a variável de interesse na região i e \bar{z} a sua média. Por fim, têm-se que w_{ij} são os elementos da matriz de ponderação espacial. Como já foi dito, a estatística testa a hipótese nula de aleatoriedade espacial. Valores positivos e estatisticamente significativos implicam que a variável apresenta autocorrelação espacial positiva. Valores negativos e estatisticamente significativos significam que a variável apresenta autocorrelação espacial negativa. Nos dois casos, a força dessa associação é dada pela magnitude do coeficiente.

(7) Devido à propriedade do estimador ser consistente e ter eficiência assintótica.

(8) O município i afeta o município vizinho j , mas como j também afeta i , qualquer variação em i afetará ele mesmo via efeito que tem em j .

O problema deste teste é que ele é uma estatística global, ou seja, considera todo o conjunto de regiões de análise em seu cômputo. Isso implica que muitos padrões locais de similaridades podem ser compensados por outros padrões locais de dissimilaridades, e isso pode fazer com que não se possa rejeitar a hipótese nula do I de Moran. Sabendo desse fato, o que os trabalhos empíricos têm feito é também apresentar, como complemento à análise do I de Moran, a estatística LISA. Esta estatística nada mais é do que uma decomposição do I de Moran, ou seja, é um I de Moran Local, que algebricamente pode ser expresso por:

$$I_i = n \frac{y_i \sum_j^n w_{ij} y_j}{\sum_j^n y_j^2} \quad (7)$$

Em que $y_i = z_i - \bar{z}$ e $y_j = z_j - \bar{z}$, onde z_i é a variável de interesse na região i e \bar{z} representa a média desta variável.

Uma vez constatada a autocorrelação espacial global e local nos resíduos da equação 2, é necessário modelar essa dependência de modo a capturá-la em um (alguns) parâmetro(s) que corrijam esse problema. Um dos modelos espaciais mais simples é o SAR, que já foi exposto anteriormente. No entanto, um dos problemas da especificação do modelo SAR é que ele impõe endogeneidade mecânica (simultaneidade) ao introduzir-se o *lag* espacial da variável dependente na equação. Uma das formas de tentar contornar esse problema é por meio do modelo SLX, que no caso do presente trabalho assume a forma:

$$I_{it} = \beta_0 + \beta_1 PABE_{it} + \varphi WPABE_{it} + \gamma_t + \tau_r + \phi X_{it} + \varepsilon_{it} \quad (8)$$

Ou seja, introduz-se ao modelo inicial (equação 2) a defasagem espacial do regressor de interesse e estima-se por OLS. A especificação acima pode ser útil em capturar a dependência espacial por meio dessa característica observável da vizinhança, principalmente quando se têm uma justificativa teórica (Almeida, 2012). Nesse caso, φ é o efeito dos vizinhos que surge a partir de certa característica observável. Mas o problema desse modelo é que continua não existindo garantia de que $E(\varepsilon|PABE) = 0$, o que implica que o parâmetro estimado de interesse (β_1) não pode ser usado para inferência de efeito causal.

Um outro problema do modelo SLX é que, como destaca Almeida (2012), é muito provável que $E(\varepsilon|WPABE) \neq 0$, devido ao fato de que $E(WI|WPABE) \neq 0$. É simples de entender: a inclusão de *WPABE* e exclusão de *WI* na equação 2, faz com que possa ocorrer viés de variável relevante omitida, tornando o estimador inconsistente (uma vez que *WI* recairá no termo de erro), e o modelo não capturará todo o processo de dependência espacial. Além disso, pode surgir viés de endogeneidade a partir de um processo de *sorting*, descrito por Gibbons, Overman e Patacchini (2015). No caso do presente trabalho, ele pode ocorrer quando determinantes não observados da vizinhança (e, portanto, não mensuráveis e que recaem no termo de erro) podem afetar conjuntamente a variável explicativa de interesse e a variável dependente, fazendo novamente com que o estimador seja enviesado e inconsistente. Neste caso, teria-se que $\varepsilon = Wv + \mu$, em que o termo de erro pode ser decomposto em um componente oriundo de características não observadas da vizinhança (Wv) e um componente que é oriundo de características também não observadas, mas não espaciais (μ).

Para contornar os problemas expostos no parágrafo anterior, o presente trabalho segue Gibbons, Overman e Patacchini (2015) e emprega o Estimador de Efeitos Fixos Espacial (EFE). Neste método, a variável de interesse não representará as internações de residentes do município i no tempo t (representado por I_{it}), mas sim o diferencial entre essas internações e a média das internações de sua vizinhança no tempo t , de modo que:

$$\widetilde{I}_{it} = I_{it} - WI_{it} \quad (9)$$

Na equação 8, têm-se que WI_{it} é a média de internações de residentes na vizinhança do município i em uma determinada unidade temporal t . Para a construção dessa média utilizou-se a matriz de pesos do inverso da distância. Nela, são atribuídos pesos aos municípios: quanto mais próximo geograficamente, maior o peso, quanto mais distante geograficamente, menor o peso. Por fim, normalizou-se estes pesos para que eles somem um, ou seja, a matriz é normalizada na linha. Do mesmo modo, aplica-se essa diferença ao regressor endógeno de interesse, uma vez que é preciso que a imposição dessa diferença espacial seja realizada nos dois lados da equação:

$$PABE_{it} = PABE_{it} - WPABE_{it} \quad (10)$$

Assim, o que foi feito foi simplesmente multiplicar o vetor de internações de residentes de cada localidade por uma matriz de ponderação espacial, obtendo a média de internações da vizinhança de cada município. Esta técnica permite que sejam eliminados problemas de autocorrelação espacial no modelo, que tornariam as estimativas enviesadas e inconsistentes.

Por fim, para eliminar o problema de $E(\varepsilon = \mu|PABE) \neq 0$, foi associado ao EFE a estratégia de IV. Tudo isso faz com que se tenha uma nova estrutura de identificação, dada pelas equações 11, 12 e 13, que correspondem à versão espacial de efeitos fixos associada à técnica de Variável Instrumental (que será denominada de IV-EFE):

$$\widetilde{I}_{it} = \beta_0 + \beta_1 \widetilde{PABE}_{it} + \gamma_t + \tau_r + \phi X_{it} + \varepsilon_{it} \quad (11)$$

$$\widetilde{PABE}_{it} = \delta_0 + \delta_1 PABT_{it} + u_{it} \quad (12)$$

$$\widetilde{I}_{it} = \beta_0 + \beta_1 PABT_{it} + \gamma_t + \tau_r + \phi X_{it} + \varepsilon_{it} \quad (13)$$

A equação 11 representa o segundo estágio, a equação 12 o primeiro estágio e a equação 13 a forma reduzida do modelo. Essa nova especificação é uma estratégia de identificação causal do IV que incorpora Efeitos Fixos Espaciais. Essa metodologia garante que nenhuma característica espacial (observável ou não) esteja correlacionada com o regressor de interesse (o PAB fixo efetivo recebido pelos municípios), eliminando problemas de autocorrelação espacial. Adicionalmente, o IV permite variações exógenas nessas transferências.

Por outro lado, o uso da normalização na linha se dá pelo fato de que, como destaca Tyszler (2006), a literatura que usa matrizes de pesos espaciais encontra-se, de forma preponderante, em estudos com ela normalizada. Não por menos, Anselin (1998) faz algumas considerações importantes sobre a matriz normalizada na linha. Primeiro, ela permite uma comparação mais fácil entre diferentes pesos de uma dada matriz. Segundo, possibilita uma comparação mais fácil entre diferentes pesos entre diferentes matrizes. Terceiro, facilita a leitura da influência geral em i atribuída à vizinhança j . Como a normalização na linha consiste em fazer com que a soma dos elementos da linha da matriz

de pesos seja 1, isso faz com que cada vizinho de i tenha um peso inferior a 1 e que a soma dos pesos dos k vizinhos de i seja 1. Do ponto de vista matemático, a consequência é que isso restringe a magnitude do parâmetro, fazendo com que ele esteja no intervalo $(-1; +1)$, o que também é uma vantagem teórica na estimação do modelo, como visto anteriormente.

Vale salientar que, para estimação dos modelos, os dados foram organizados para os anos de 2014, 2015 e 2016 em um painel balanceado cuja unidade de corte é o município. Município é a menor área administrativa do sistema político do Brasil e atualmente são 5.570 unidades. Destes, como já foi dito, cerca de 90% correspondem a pequenos municípios com menos de 50 mil habitantes (em 2012, ano da Portaria do PAB fixo). Como o período de estudo compreende vários anos, a fim de considerar o surgimento de novos municípios ao longo desses três anos, utilizou-se da criação de Áreas Mínimas Comparáveis (AMCs), totalizando 5.564 AMCs em cada corte anual. Como o trabalho objetiva verificar o efeito causal nas pequenas localidades do país (com menos de 50 mil habitantes), os modelos serão estimados para 4.952 AMCs em cada um dos três cortes anuais, totalizando 14.856 observações.

4 Análise dos resultados

Como já foi apresentado, a pergunta a ser respondida aqui é: será que mais dinheiro público destinado às ações de saúde da atenção primária geram uma variação na probabilidade de uso hospitalar por condições sensíveis a atenção primária?

Antes de prosseguir para a análise das estimativas, cabem algumas considerações importantes sobre os resultados teóricos esperados das estimativas dos indicadores. Em uma primeira situação, supõe-se que, se a atenção primária do município receber um maior financiamento do PAB fixo, sendo, tudo o mais constante, isso poderá representar uma melhora da cobertura ou intensificação das ações executadas pelas unidades básicas de saúde. Ter-se-ia, assim, uma melhora da saúde da população, uma vez que os problemas de saúde seriam melhor administrados ou controlados neste nível de atenção. Como consequência, diminuiria o repasse de doentes ao sistema hospitalar. Isso porque o choque no financiamento, se por um lado pode diminuir o surgimento da doença em indivíduos saudáveis, por outro lado, também pode diminuir o ritmo de progressão das doenças dos indivíduos já afetados, de modo que caso a pessoa precise ser internada, seu grau de deterioração de saúde poderá não ser tão grave, resultando em menor pressão de uso do sistema hospitalar.

Por outro lado, supondo que a atenção primária não apresenta nenhum caráter de resolução dos problemas da comunidade, funcionando mais como intermediador entre a população e o sistema hospitalar (o que é uma suposição plausível), é de se esperar que com maior financiamento, maior o repasse de enfermos ao sistema de média e alta complexidade. Isso poderia resultar em uma elevação da média dos indicadores de interesse, devido a uma sobreutilização de leitos e congestionamento da capacidade resolutiva do sistema hospitalar. É de se esperar também que, uma vez que a cobertura e as ações da atenção primária melhorem, parte do contingente populacional que não tinha cobertura ou que não recebia os cuidados adequados, passe a ser levado em consideração pelas equipes de saúde. No entanto, parte desse contingente, outrora não coberto, pode estar em um grau avançado do agravo de saúde, portanto, a solução da atenção básica seria impossível. Isso representaria, novamente, um maior repasse de doentes ao sistema hospitalar. Neste caso, independentemente se a

atenção primária apresenta eficácia, ou funciona apenas como um atravessador de agravos ao sistema hospitalar, esse resultado final positivo de elevação dos indicadores se manteria.

Uma terceira situação ocorre quando o choque no financiamento não gera nenhum efeito nos indicadores em decorrência da má administração dos recursos pelos gestores locais, ou devido à corrupção. Também pode ser devido a uma situação diferente, em que a atenção primária é muito boa e tem ampla cobertura, de modo que mais financiamento não gera diferença nas ações implementadas na localidade. Dessa forma, percebe-se que há a possibilidade de várias direções nos indicadores, a depender do efeito que se sobressai quando ocorrer um choque positivo do financiamento. Em geral, podem coexistir todos esses efeitos ao mesmo tempo entre os municípios, embora apenas o mais forte prevaleça. Desse modo, o que constata-se aqui é apenas o efeito líquido global, mediante o sinal e magnitude dos parâmetros encontrado nas estimativas.

Antes de seguir à estimação do modelo IV-EFE, será testada a existência de dependência espacial das interações por condições sensíveis no Brasil. Primeiramente utilizar-se-á um modelo SAR para verificar a existência de efeito de transbordamento das interações entre regiões vizinhas. Depois, serão apresentados os resultados dos testes de hipótese de I de Moran Global e Local. No caso do modelo SAR, o resultado está na Tabela 2.

Tabela 2
Estimativa do Modelo SAR. Estimação realizada
para todo o Brasil (ano de 2014)

Variável	
<i>WI</i>	0,1083*** (0,0209)
Observações	5.564
R-squared	0,005
Log likelihood	-12.226,8

*significativo a 10%, ** a 5%, *** a 1%. Nota: a estimação foi realizada por Máxima Verossimilhança (ML).

Nesta tabela, pode-se observar que o parâmetro associado à defasagem espacial da variável dependente é positivo e estatisticamente significativo, e sugere que quando são maiores as interações da região vizinha j , isto está associado à maior taxa de interação de determinado município i . O mesmo raciocínio se mantém no sentido contrário: quanto menores as interações na vizinhança, menor a interação do município em questão. Como já foi explicitado anteriormente, o parâmetro ρ , associado à variável WI , não é interpretado como um efeito de impacto, ou seja, não há uma relação de causa e efeito entre as interações na vizinhança e as interações locais. Neste modelo, a interpretação do ρ é apenas uma constatação de que há um efeito de vizinhança ou um transbordamento das interações entre municípios próximos.

Vejam agora os resultados dos testes de I de Moran e LISA. Como já foi explicitado anteriormente, é usual testar a autocorrelação dos resíduos de modelos econométricos em que se suspeita haver algum processo de dependência espacial. Assim, o que foi feito foi testar a autocorrelação espacial dos resíduos da especificação exposta na equação 1 (estimado por OLS),

cujos resultados estão nas figuras 5 e 6. Na Figura 5, a estatística I de Moran apresentou valor de 0,0447, em que se pôde rejeitar a hipótese nula de que não há autocorrelação espacial.

Figura 5
Diagrama de dispersão do teste I de Moran dos
resíduos do modelo 1: efeito do PAB efetivo sobre as interações
por condições sensíveis nos municípios do Brasil (2014)

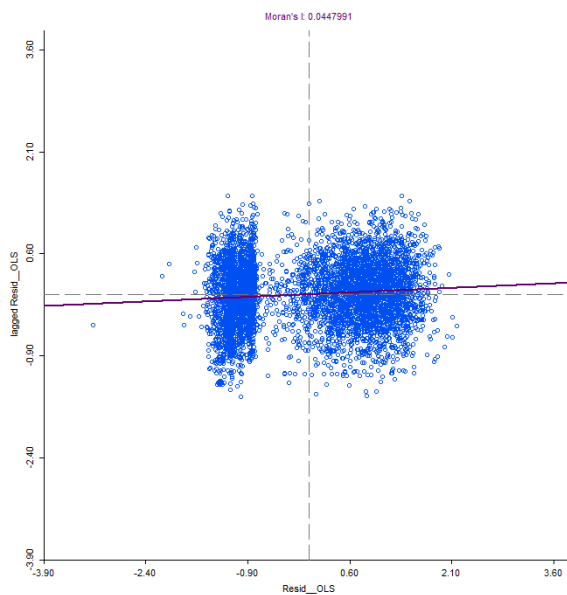
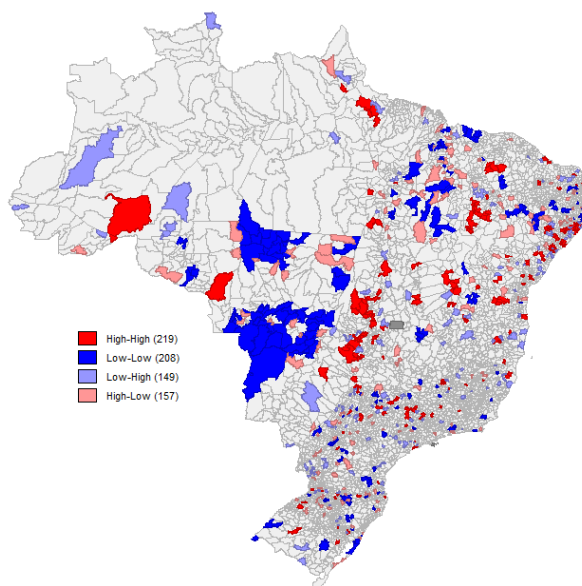


Figura 6
Moran Map do teste I de Moran Local (LISA) dos
resíduos do modelo 1: efeito do PAB efetivo sobre as interações
por condições sensíveis nos municípios do Brasil (2014)



Pelo sinal da estatística (e padrão do diagrama) os resíduos apresentam um comportamento predominante de similaridade espacial. O problema é que esta estatística pode esconder padrões locais de dissimilaridade além de *outliers* espaciais. Dessa forma, é conveniente ver o resultado da estatística LISA, expresso na Figura 6. Nesse mapa é possível perceber padrões de similaridade (*High-High* e *Low-Low*), dissimilaridades (*High-Low* e *Low-High*) dos resíduos da regressão. Esses três resultados revelam a necessidade de acomodar a dependência espacial por meio de alguma modelagem.

Vejamos agora, a estimação do modelo de interesse deste trabalho. Na Tabela 3, vê-se o efeito das transferências efetivas sobre a variável de interesse. Têm-se na coluna 5 o modelo IV-EFE. Para efeito de comparação, também foram estimados os modelos 1, 2, 3 e 4. Os modelos 1 e 2 são modelos que não incorporam dependência espacial, ou seja, que não incorporam Efeitos Fixos Espaciais. Na coluna 1, a estimação foi feita mediante simples OLS e na coluna 2 têm-se a estratégia de identificação com o IV. Na coluna 1 percebe-se uma correlação positiva entre as transferências efetivas do PAB e as internações locais. Na coluna 2, o efeito causal obtido também é positivo, embora a elasticidade tenha caído bastante quando comparado ao modelo OLS.

Tabela 3
Estimativas dos modelos do efeito do financiamento do PAB
fixo efetivo sobre as internações por condições Sensíveis. Regressões
estimadas para os pequenos municípios do Brasil (2014-2016)

Variável	(1) OLS	(2) IV	(3) SLX	(4) OLS-EFE	(5) IV-EFE
PAB Efetivo	3,7384*** (0,4642)	1,0852*** (0,6170)	3,5268*** (0,4642)	0,0802*** (0,0011)	0,1263*** (0,0133)
Observações	14.856	14.856	14.856	14.856	14.856
R-squared	0,3072	0,3025	0,3112	0,7672	0,6881

Erros-padrão robustos entre parênteses. *significativo a 10%, ** a 5%, *** a 1%.

Nota: as variáveis estão em logaritmo, de modo que o parâmetro expressa elasticidade.

Nas colunas 3, 4 e 5 têm-se os modelos que tentam controlar a dependência espacial. Na coluna 3, o modelo SLX exibe um coeficiente com magnitude parecida com o modelo OLS, embora ele seja incapaz de acomodar a dependência espacial de características da vizinhança que não são observáveis. Finalmente nas colunas 4 e 5 é incorporada a técnica de Efeitos Fixos Espaciais. No modelo mais robusto (IV-EFE) o parâmetro estimado mostrou-se positivo e estatisticamente significativo, evidenciando um efeito de impacto do PAB fixo efetivo sobre as internações por condições sensíveis nos pequenos municípios do Brasil. No caso do modelo IV-EFE, que é o modelo mais completo, os resultados mostram que um aumento de 1% no PAB efetivo gera, em média, um aumento de 0,1263% nas internações por condições sensíveis.

Quando se compara o IV com o IV-EFE, percebe-se que quando não se incorpora Efeitos Fixos Espaciais o Estimador de Variável Instrumental superestima o efeito do PAB sobre as internações locais. Em um modelo não espacial, o efeito do PAB do município *i* se mistura ao efeito do PAB de sua vizinhança *j* (e inevitavelmente ao da atenção primária), além de características não observáveis dessa vizinhança que inevitavelmente afetarão o nível de internações da localidade *i*. Isso

faz com que se superestime o efeito do financiamento em cada localidade. Quando se desconta o efeito das características espaciais observáveis e não observáveis da vizinhança, o efeito cai bastante, embora ainda se mantenha positivo.

Mas por que o efeito do PAB cai ao se passar do modelo IV para o IV-EFE? Conforme exposto, o efeito do maior financiamento do PAB sobre as internações de residentes pode ser positivo ou negativo. Em ambos os modelos IV e IV-EFE verifica-se que prevalece o efeito dele gerar maior acesso ao sistema hospitalar. Dado o exposto, se por exemplo, um município i está rodeado de uma vizinhança j , maiores repasses do PAB fixo em i , provavelmente gerariam um efeito de maior acesso ao sistema hospitalar dessa localidade i , mas tenderia a diminuir o acesso dos residentes da localidade j . Isso aconteceria pois, primeiro, parte da população de i tenderia a procurar esses serviços fora de sua residência, a um custo de deslocamento baixo, dada a proximidade geográfica definida pela vizinhança. Segundo, parte dos leitos locais de j seriam direcionados para esses não residentes (oriundos de i), diminuindo a disponibilidade de leitos para os nativos em j . Do mesmo modo, um maior financiamento do PAB fixo na região j , geraria um aumento das internações de residentes na região j , e uma diminuição do acesso dos residentes da região i , pelo mesmo efeito de substituição de leitos entre residentes e não residentes. Assim, ter-se-ia:

$$\Delta PABE_i > 0 \rightarrow \Delta I_{res_i} > 0 \text{ e } \Delta I_{res_j} < 0$$

$$\Delta PABE_j > 0 \rightarrow \Delta I_{res_j} > 0 \text{ e } \Delta I_{res_i} < 0$$

Como análise complementar, verificou-se a existência dessa elasticidade-cruzada mediante os seguintes modelos:

$$WI_{it} = \beta_0 + \beta_1 PABE_{it} + \gamma_t + \tau_r + \phi X_{it} + \varepsilon_{it} \quad (14)$$

$$I_{it} = \beta_0 + \beta_1 WPABE_{it} + \gamma_t + \tau_r + \phi X_{it} + \varepsilon_{it} \quad (15)$$

Estas estimações foram feitas por OLS. Os resultados encontrados (mas não reportados em tabela) mostram que, ao nível de significância de 1%, no primeiro modelo, um aumento de 1% no PAB fixo de determinado município i está associado a uma queda média de 3,87% das internações dos residentes da vizinhança j . No segundo modelo, um aumento de 1% no PAB fixo médio da vizinhança j , está associado ao fato de que as internações dos residentes declinam 0,005%, em média, em i . Os dois modelos demonstram elasticidades cruzadas com resultados esperados teoricamente. Vale salientar que uma vez que não há estratégia de identificação em 14 e 15, as elasticidades têm que ser interpretadas em termos de correlações e não impacto causal.

Dada a constatação dessa elasticidade cruzada negativa, supondo, por exemplo, uma melhor infraestrutura de saúde dessa vizinhança j , quando comparado a i , ocorreria uma maior migração hospitalar de i para j , do que de j para i . Isso implicaria um menor nível de internação local de residentes em i relativamente a j , quando em ambas as regiões o PAB cresce. Por isso, no modelo IV-EFE, quando é descontado o efeito das características observáveis e não observáveis da vizinhança, é natural que o impacto do PAB seja menor nas internações dos residentes que no modelo IV.

O modelo IV-EFE é um modelo de Variável Instrumental associado a um modelo de Efeitos Fixos Espacial. Na Tabela 3, tanto no modelo da coluna 2, quanto no da coluna 5, as transferências

teóricas foram utilizadas no primeiro estágio para expurgar a endogeneidade do PAB efetivo, de modo que, por meio do valor previsto do modelo do primeiro estágio, foi possível gerar variação exógena nessa verba. Na Tabela 4 (coluna 1), vê-se os resultados das estimativas do efeito do PAB teórico sobre o PAB efetivo, ou seja, o primeiro estágio da identificação do IV-EFE. O modelo nos dá as estimativas em termos de elasticidades e a coluna evidencia que a transferência teórica é um excelente preditor da transferência efetiva (de modo que o instrumento é altamente correlacionado com o regressor endógeno de interesse), uma vez que cerca de 63% da variação do PAB fixo efetivo é decorrente de variações no PAB fixo teórico.

Tabela 4
Estimativas do primeiro estágio e da forma reduzida do
Modelo IV-EFE: Regressões estimadas para os pequenos municípios
do Brasil (2014-2016)

Variáveis	(1) Primeiro Estágio	(2) Forma Reduzida
PAB Efetivo	73,2677*** (10,2919)	- -
PAB Teórico	- -	9,2541*** (1,2744)
Observações	14.856	14.856
R-squared	0,6330	0,5317

Erros-padrão robustos entre parênteses. *significativo a 10%, ** a 5%, *** a 1%.

Nota: as variáveis estão em logaritmo, de modo que o parâmetro expressa elasticidade.

Um alto poder de previsão do instrumento sobre o regressor endógeno é um requisito necessário para que o instrumento seja bom, porém não é suficiente. A exogeneidade estrita do instrumento, por sua vez, não pode ser testada. No entanto, como foi exposto anteriormente, o argumento levantado aqui (que o PAB teórico só pode afetar os indicadores de internação unicamente via determinação do PAB efetivo) parece ser coerente, uma vez que os municípios nunca recebem o PAB teórico. Assim, há um único caminho da transferência teórica chegar ao município, que é pela via da transferência efetiva.

Ainda na Tabela 4 (coluna 2), verifica-se o efeito das transferências do PAB teórico sobre o indicador geral de internações por condições sensíveis (ICSA), ou seja, a coluna nos reporta as estimativas da forma reduzida do modelo IV-EFE (as estimativas dão as respostas em termos de elasticidade) Os resultados nos mostram que uma variação positiva em 1% no PAB teórico gera, em média, uma variação positiva de 9,25% nas internações.

Será que esse resultado se mantém em uma análise regional? Na Tabela 5, pode-se ver a heterogeneidade da resposta do modelo anteriormente analisado para cada uma das grandes regiões do país. Nessa tabela, vê-se o efeito das transferências efetivas do PAB fixo em cada uma das grandes regiões, em que se percebe uma nítida diferença na magnitude, sinal e significância dos parâmetros

estimados ao incorporar ao IV, o efeito fixo espacial. O que se percebe também é que o padrão de respostas encontrados para o Brasil se mantêm nas regiões, ao nível de significância de 5%, apenas no modelo IV-EFE. A esse mesmo nível de significância, choques positivos no financiamento da atenção básica parecem resultar em um aumento no indicador hospitalar de interesse nas regiões Norte, Sudeste e Centro-Oeste. Além disso, a variação percentual é maior na região Sudeste, a mais rica do país.

Tabela 5
Estimativas dos modelos do efeito do financiamento do PAB fixo efetivo sobre as internações por condições sensíveis em cada macrorregião.
Regressões estimadas para os pequenos municípios do Brasil (2014-2016)

Variáveis	(1) IV	(2) IV-EFE
PAB Efetivo_Norte	-0,2847 (1,4444)	0,0485** (0,0222)
PAB Efetivo_Nordeste	6,9919* (3,7446)	0,1039 (0,2536)
PAB Efetivo_Sudeste	0,9461 (0,7711)	0,1422*** (0,0235)
PAB Efetivo_Sul	1,1355 (0,9643)	0,0795 (0,3355)
PAB Efetivo_Centro-Oeste	2,6692 (2,1248)	0,0933** (0,0409)
Observações	14.856	14.856
R-squared	0,2296	0,7106

Erros-padrão robustos entre parênteses. *significativo a 10%, ** a 5%, *** a 1%.

Nota: as variáveis estão em logaritmo, de modo que o parâmetro expressa elasticidade.

Um ponto importante e que já foi destacado neste trabalho, é que supõe-se que qualquer variação no financiamento do PAB efetivo, deva afetar a taxa de internações por condições sensíveis a atenção primária, por dois motivos. O primeiro efeito seria positivo, e decorrente de uma das funções da atenção básica, que é a de descobrir e encaminhar os casos que não são de sua atribuição resolutiva para o sistema hospitalar. Espera-se que quando o município receba mais verba isso se traduza em uma maior cobertura das ações, ou melhoramento da triagem, resultando em maiores taxas de internação, como já foi exposto anteriormente. É importante destacar que isso pode ser fruto de um subfinanciamento crônico, que faz com que exista forte contingente de demanda reprimida por cuidados de média e alta complexidade, que só aparecem quando se amplia o acesso a pacientes que estavam fora do sistema de saúde, sem qualquer tipo de cuidado. O segundo efeito, por outro lado, seria negativo e resultado do caráter resolutivo da atenção básica que, em tese, resolveria cerca de 80% dos problemas de saúde da população, diminuindo naturalmente o repasse de enfermos ao sistema hospitalar.

De forma geral, os resultados encontrados nesse trabalho sugerem que um aumento do financiamento do PAB fixo resultou em maiores taxas de internações, o que evidencia uma

prevalência do efeito resultado de um provável contingenciamento de demanda reprimida por serviços de saúde. Este dado evidencia que uma melhora das ações da atenção primária, resultado de maior aporte de recursos federais, implicou em um maior acesso a internação por parte da população local nessas pequenas localidades, de pessoas que provavelmente tinham alguma restrição de acesso. Isso é um dado preocupante, pois parece que a oferta de atenção básica, apesar de maior financiamento, não tem sido capaz de ampliar a resolutividade de problemas de saúde dos usuários neste nível de atenção, ainda levando a maior parte dos problemas de saúde da população para o sistema hospitalar.

O que os dados indicam é que pode estar ocorrendo uma subutilização da capacidade resolutiva ou ineficiência da atenção primária, comprometendo a efetividade dos recursos públicos destinados a essas ações. Assim, a oferta de atenção básica, apesar de sofrer um incremento financeiro, não tem tido a capacidade de organizar melhor os fluxos de atendimentos nos centros hospitalares. Tal quadro evidencia a ineficácia do gasto público em atenção básica no que diz respeito à sua função primordial, que é a de resolver a maior parte dos problemas de saúde da população (que são mais simples), desafogando leitos e encaminhado ao sistema hospitalar apenas os casos mais graves e que não são de sua atribuição. O que se verifica, é que parece haver apenas uma variação positiva do acesso dos indivíduos aos serviços hospitalares à medida que a cobertura ou ações implementadas melhoram via maior financiamento. Deste modo, a atenção básica tem funcionado mais como um encaminhador ou mediador de acesso entre o usuário e o serviço hospitalar, de agravos que ela deveria resolver, pelo menos nos pequenos municípios do Brasil.

O problema se agrava porque os municípios de análise embora possuam uma população menor, são localidades com condições econômicas e sociais mais precárias e tendem a apresentar um conjunto de equipamentos de saúde menos complexo, ou uma pior gestão administrativa do sistema de atenção primária de caráter resolutivo, devido a inúmeras questões já citadas anteriormente. Isso pode ser uma justificativa para essa relação positiva entre as ações primárias e as internações por condições sensíveis: quando ocorre uma melhora dessas ações (via maior financiamento do PAB efetivo), a cobertura melhora, fazendo que as equipes de atenção primária apenas encaminhem os indivíduos para o sistema hospitalar e não diminuam os casos de internação via ações preventivas, acompanhamento e orientação médica.

Neste sentido, por exemplo, Bhalotra, Rocha e Soares (2016) destacam que ações de atenção primária são importantes porque podem reduzir o surgimento de muitas doenças crônicas⁹ (como diabetes, por exemplo) por meio de programas de prevenção e de mudanças no estilo de vida dos indivíduos. Como destacam os autores, essas ações poderiam resultar em uma melhora da saúde preventiva (melhor alimentação, práticas de atividades físicas, consumo moderado de tabaco e álcool, etc.) resultando em uma diminuição da incidência dessas doenças por meio de um estilo de vida mais saudável.

Além disso, mesmo que o indivíduo já tenha desenvolvido alguma doença crônica, por exemplo, as Unidades Básicas de Saúde poderiam ajudar consideravelmente na vida e no bem-estar dos pacientes pela melhor administração dos agravos mediante ações ambulatoriais (como, por

(9) Doenças crônicas são caracterizadas na literatura de saúde com uma altíssima incidência na população. Segundo Macinko, Dourado e Guanais (2011), esse tipo de doença já mata 3 em cada 5 pessoas no mundo.

exemplo, a injeção regular e controlada de insulina em diabéticos e a vacinação), o que diminuiria a necessidade frequente de internações. Neste sentido, como destacam Malta et al. (2011), as equipes de saúde da atenção primária podem fazer acompanhamento longitudinal dos usuários, e isso pode ser fundamental na melhoria da resposta ao tratamento de enfermidades. Os autores ainda destacam a rede farmacêutica e a distribuição gratuita de medicamentos para hipertensão e diabetes, por exemplo, nas farmácias das unidades básicas de saúde, o que permitiria que a população pobre tenha acesso a medicamentos e ao adequado tratamento químico.

Na contramão, o que se verifica nos achados empíricos deste trabalho é que as ações e organização da oferta de serviços de atenção básica tem contribuído mais para um maior acesso aos hospitais (provavelmente devido a uma demanda contingenciada), por problemas de saúde cuja resolução deveria ter sido feita pelas equipes e unidades de atenção primária. Assim, este resultado sugere que a atual organização da oferta dos serviços de atenção básica do Brasil está sobrecarregando os serviços hospitalares dos pequenos municípios, resultando em uma alocação ineficiente de leitos. Parece haver uma duplicidade de ações entre os centros de atenção básica (ou equipes de saúde) e os hospitais. Ao que parece, a atenção primária está funcionando primordialmente como um intermediador entre a população e o sistema hospitalar, não apresentando resolução ou administrando os problemas de saúde que são de sua competência. Este resultado sugere uma atual desarticulação entre as atribuições dos níveis de atenção do sistema de saúde no Brasil.

Há margem para muitas discussões teóricas sobre os resultados empíricos aqui encontrados (financiamento insuficiente, ineficiência dos gastos, corrupção, etc.), mas focaremos aqui em uma que parece ser premente e de longa data, que é o subfinanciamento crônico, que tem fortes implicações em uma demanda contingenciada por serviços de saúde. Para se ter uma ideia, o gasto governamental em saúde no Brasil, no ano de 2013, foi da ordem de cerca de R\$ 946 *per capita* (US\$ 591), valor que é de quatro a sete vezes menor do que o de outros países que também apresentam sistemas universais de saúde, como França, Reino Unido e Canadá (Brasil, 2015). Isto deverá gerar uma maior preocupação fiscal por parte do governo brasileiro se este desejar garantir adequadamente os serviços do sistema.

Desde a constituição do SUS, o debate sobre o adequado financiamento (e vinculação de recursos ao sistema) tem fomentado discussões e ações tanto do Poder Executivo quanto do Congresso Nacional ao longo de três décadas. Com a criação do SUS em 1988 (através da Constituição Federal) e sua regulamentação em 1990 (através das Leis n. 8.080 e n. 8.142), estabeleceu-se que deveria ser destinado ao sistema 30% do Orçamento da Seguridade Social. O problema é que mesmo com determinação legal, esta prática não se sustentou historicamente. Só no ano 2000, com a Emenda Constitucional n. 29, é que houve uma determinação mais incisiva de modo a realmente garantir o financiamento da rede através da vinculação de receitas das três esferas administrativas de governo. Nesta emenda, caberia aos Estados destinar à saúde um mínimo de 12% de suas receitas próprias, enquanto aos Municípios, caberia destinar um mínimo de 15% destas. À União, estabeleceu-se uma regra que levaria em conta um financiamento baseado nas ações empenhadas em 1999, acrescido de no mínimo 5%, para o ano 2000. No período 2001-2015, a aplicação foi baseada no ano anterior corrigida pela variação nominal do produto interno bruto (crescimento real mais inflação).

Embora tenhamos tido um processo lento de avanço nas garantias de financiamento do SUS durante estas décadas, recentemente esta dinâmica parece ter cessado. Em 2016 a Emenda Constitucional n. 95 (mais conhecida como PEC do Teto dos Gastos) redefiniu a Emenda Constitucional n. 85, modificando os escalonamentos anuais e os pisos de financiamento da União da Lei anteriormente citada. Além disso, definiu também que a partir de 2018, o investimento federal mínimo seria o valor empenhado em 2017 atualizado pela inflação até 2036, sujeitando-se à regra do teto dos gastos com as despesas primárias, a despeito das necessidades da população, seu envelhecimento, novas demandas de saúde, entre outros fatores ao longo das próximas duas décadas.

Vieira e Benevides (2016) argumentam que este novo modelo fiscal poderá ser perverso ao SUS, dado que provavelmente aumentará as iniquidades de adesão ao sistema, afetando os grupos sociais mais vulneráveis e as populações das regiões/estados mais pobres e com menor capacidade de financiar tais ações. Some-se a isto um problema que tem sido reportado constantemente na literatura, que é a pressão ou esforço fiscal crescente dos municípios em financiar a saúde: entre 2003-2015, por exemplo, a proporção dos gastos municipais em saúde subiu de 25,4% para 31%, enquanto a proporção dos gastos do governo federal em saúde caiu de 50,1% para 43%. Com menores capacidades de financiamento, os municípios tenderão a sofrer mais com as pressões de demanda por ações e serviços em saúde por parte de suas populações.

Além disso, os autores anteriormente citados argumentam que o SUS já parte, na nova regra fiscal, de uma situação de forte subfinanciamento, dado que entre 2002-2015, por exemplo, verificou-se uma estabilização dos gastos federais em ações de saúde (como proporção do PIB) e caiu a relação de gastos com ações em saúde como proporção das despesas primárias do governo federal. Assim, verificou-se que o gasto público federal no SUS, em relação ao PIB, não aumentou com a expansão econômica verificada no período. Isto, em um cenário em que a população cresceu e envelheceu. Como destacam Piola et al. (2013), isto foi resultado de uma política do governo federal de não investir (ou investir muito pouco) além do “mínimo obrigatório” estabelecido pela Emenda Constitucional n. 29.

Obviamente, no que diz respeito ao PAB fixo, foco deste trabalho, a questão do subfinanciamento não foge à regra. Como destacam Mendes e Marques (2014), o valor *per capita* instituído na criação do PAB (em 1997 com a Portaria n. 1.882 do Ministério da Saúde) já estava 20% defasado em termos das reais necessidades da população. Além disso, embora venha sofrendo reajustes periódicos ao longo dos anos, estes são sempre menores que a inflação, e é muito pouco provável que este problema será corrigido com a nova regra fiscal. Outro problema é que os reajustes, historicamente, também não têm levado em conta a dinâmica populacional dos municípios, ferindo um dos princípios básicos do PAB fixo, que é o de tentar equalizar o financiamento *per capita* da atenção primária entre municípios com diferentes tamanhos populacionais e capacidades de arrecadação tributária.

Considerações finais

O objetivo deste trabalho foi verificar a eficácia da atenção primária nos indicadores de internações hospitalares nos pequenos municípios do Brasil. Para isso, foi utilizada uma estratégia de identificação que permitiu garantir choques exógenos de financiamento dessas ações, de modo a

verificar a existência de efeito causal no indicador de interesse livre de processos espaciais resultado de *sorting*. Esse fenômeno causa viés de endogeneidade na variável de interesse por meio da correlação desse regressor com variáveis espaciais não observáveis incluídas no termo de erro. Assim, foi utilizado um método de estimação por Variável Instrumental associado a um estimador de Efeitos Fixos Espaciais.

De forma geral, os resultados evidenciam que quanto maior o financiamento das ações primárias executadas nos pequenos municípios, maiores os repasses de indivíduos ao sistema hospitalar, contrariando, assim, o resultado teórico esperado de que a atenção primária reduziria em grande medida as internações por melhoria da saúde coletiva. Ao que parece, a atenção primária, da forma como está atualmente formada, funciona apenas como um intermediário entre a população e o sistema hospitalar, ou seja, parece que o Brasil encontrou dificuldades em estruturá-la nessas unidades administrativas menos populosas.

Além disso, deve-se salientar que os resultados aqui apresentados não se contrapõem aos apresentados na literatura sobre intervenções locais de saúde comunitária apresentados na introdução deste trabalho. Muito pelo contrário, os complementa sobre uma nova ótica. Em geral, esses trabalhos mostram apenas que a introdução de ações de atenção primária diminuiu consideravelmente as causas de mortes em muitos lugares do planeta, principalmente nas pequenas e mais pobres localidades. O que este trabalho inova trazendo ao debate para literatura de saúde pública, é a discussão sobre a possibilidade de ineficácia dos gastos governamentais nesse nível de atenção, e suas consequências sobre a alocação eficiente de leitos do sistema hospitalar. No caso dos pequenos municípios do Brasil, mesmo que a atenção primária possa ser ineficiente no sentido de solucionar os agravos de saúde que seriam de sua competência, o simples repasse desses doentes para o sistema hospitalar já é uma forma de permitir que essas pessoas tenham acesso, bem ou mal, a alguns cuidados médicos, o que inevitavelmente se refletirá nos indicadores de mortalidade.

A discussão posta aqui é que nestas pequenas localidades parece haver uma ineficácia resolutive da atenção primária (a capacidade de evitar, tratar ou administrar doenças), que ainda parece funcionar estruturada em um modelo essencialmente voltado para a solução do problema de saúde em âmbito hospitalar. Enquanto esse fato não for melhor analisado e corrigido, o simples argumento de que mais verbas públicas solucionarão a superlotação de leitos dos hospitais do país, pode ser falacioso.

Referências bibliográficas

ALFRADIQUE, M. E. et al. Ambulatory care sensitive hospitalizations: elaboration of Brazilian list as a tool for measuring health system performance (Project ICSAP-Brazil). *Cadernos de Saúde Pública*, v. 25, n. 6, p. 1337-1349, 2009.

ALMEIDA, E. *Econometria espacial*. Campinas, SP: Alínea, 2012.

ANSELIN, L. *Spatial econometrics: methods and models*. Dordrecht: Kluwer Academic, 1988.

BAICKER, K.; CHANDRA, A. Medicare spending, the physician workforce, and beneficiaries' quality of care. *Health Affairs*, v. 23, n. 3, p. 291-291, 2004.

- BAILEY, M. J.; GOODMAN-BACON, A. The War on Poverty's experiment in public medicine: Community health centers and the mortality of older Americans. *The American Economic Review*, v. 105, n. 3, p. 1067-1104, 2015.
- BHALOTRA, S.; ROCHA, R.; SOARES, R. R. *Does universalization of health work? Evidence from health systems restructuring and maternal and child health in Brazil*. Institute for Social and Economic Research, 2016.
- BRASIL. Instituto Brasileiro de Geografia e Estatística. *Conta-satélite de saúde Brasil: 2010 a 2013*. Rio de Janeiro: IBGE, 2015.
- BROLLO, F. et al. The political resource curse. *The American Economic Review*, v. 103, n. 5, p. 1759-1796, 2013.
- CAMINAL, H.; CASANOVA, M. C. Primary care evaluation and hospitalization due to ambulatory care sensitive conditions. *Conceptual Framework. Atencion Primaria*, v. 31, n. 1, p. 61-65, 2003.
- DUSHEIKO, M. et al. Does better disease management in primary care reduce hospital costs? Evidence from English primary care. *Journal of Health Economics*, v. 30, n. 5, p. 919-932, 2011.
- ELHORST, J. P. Dynamic models in space and time. *Geographical Analysis*, v. 33, n. 2, p. 119-140, 2001.
- GIBBONS, S. The costs of urban property crime. *The Economic Journal*, v. 114, n. 499, p. F441-F463, 2004.
- GIBBONS, S.; OVERMAN, H. G.; PATACCHINI, E. Spatial methods. In: *HANDBOOK of Regional and Urban Economics*. Elsevier, 2015. p. 115-168.
- LESAGE, J. P.; FISCHER, M. M. Spatial growth regressions: model specification, estimation and interpretation. *Spatial Economic Analysis*, v. 3, n. 3, p. 275-304, 2008.
- MACINKO, J.; DOURADO, I.; GUANAIS, F. C. *Chronic diseases, primary care and health systems performance*. Inter-American Development Bank (IDB), 2011.
- MALTA, D. C.; MORAIS NETO, O. L.; SILVA JUNIOR, J. B. Apresentação do plano de ações estratégicas para o enfrentamento das doenças crônicas não transmissíveis no Brasil, 2011 a 2022. *Epidemiologia e Serviços de Saúde*, v. 20, n. 4, p. 425-438, 2011.
- MENDES, A.; MARQUES, R. M. O financiamento da atenção básica e da estratégia saúde da família no Sistema Único de Saúde. *Saúde em Debate*, v. 38, n. 103, p. 900-916, 2014.
- MORAN, P. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 10, n. 2, p. 243-251, 1948.
- PIOLA, S. F.; PAIVA, A. B.; SÁ, E. B.; SERVO, L. M. S. *Financiamento público da saúde: uma história à procura de rumo*. Brasília: IPEA, 2013. (Texto para Discussão, n. 1846).
- STARFIELD, B.; SHI, L. Policy relevant determinants of health: an international perspective. *Health Policy*, v. 60, n. 3, p. 201-218, 2002.

STARFIELD, B.; SHI, L.; MACINKO, J. Contribution of primary care to health systems and health. *Milbank Quarterly*, v. 83, n. 3, p. 457-502, 2005.

TYSZLER, M. *Econometria espacial: discutindo medidas para a matriz de ponderação espacial*. 2006. Tese (Doutorado).

VIEIRA, F. S.; BENEVIDES, R. P. *Os impactos do novo regime fiscal para o financiamento do Sistema Único de Saúde e para a efetivação do direito à saúde no Brasil*. Brasília, DF: IPEA 2016. (Nota Técnica, n. 28).

Anexo

Tabela A1
Estatísticas descritivas das transferências efetivas do PAB fixo *per capita* anual (R\$) no período 2014-2016 (Brasil)

Brasil			Pequenos Municípios		
Ano de 2014	Percentil	Média	Ano de 2014	Percentil	Média
1%	21,70	26,22	1%	22,03	26,54
5%	22,67		5%	24,11	
10%	23,66	Mediana	10%	24,77	Mediana
25%	25,18	26,76	25%	25,96	26,89
50%	26,76		50%	26,89	
75%	27,28	Desvio padrão	75%	27,35	Desvio padrão
90%	27,71	1,75	90%	27,74	1,49
95%	27,93		95%	27,99	
99%	29,41		99%	29,62	
Ano de 2015	Percentil	Média	Ano de 2015	Percentil	Média
1%	21,34	26,11	1%	21,75	26,45
5%	22,40		5%	23,66	
10%	23,25	Mediana	10%	24,52	Mediana
25%	25,05	26,59	25%	25,68	26,75
50%	26,59		50%	26,75	
75%	27,26	Desvio padrão	75%	27,35	Desvio padrão
90%	27,75	1,90	90%	27,80	1,65
95%	28,13		95%	28,19	
99%	29,77		99%	29,97	
Ano de 2016	Percentil	Média	Ano de 2016	Percentil	Média
1%	20,97	26,00	1%	21,46	26,36
5%	22,13		5%	23,30	
10%	23,03	Mediana	10%	25,25	Mediana
25%	24,88	26,43	25%	25,44	26,61
50%	26,43		50%	26,61	
75%	27,25	Desvio padrão	75%	27,35	Desvio padrão
90%	27,83	2,06	90%	27,89	1,83
95%	28,36		95%	28,47	
99%	30,50		99%	30,86	
Observações em cada ano	5.564		Observações em cada ano	4.952	

Fonte: Dados do Fundo Nacional de Saúde.

Tabela A2
Estatísticas descritivas das internações por condições sensíveis no período 2014-2016 (Brasil)

Brasil			Pequenos Municípios		
Ano de 2014	Percentil	Média	Ano de 2014	Percentil	Média
1%	0,00	62,33	1%	0,00	63,08
5%	0,00		5%	0,00	
10%	0,00	Mediana	10%	0,00	Mediana
25%	0,00	26,40	25%	0,00	15,82
50%	26,40		50%	15,82	
75%	90,49	Desvio padrão	75%	96,77	Desvio padrão
90%	177,38	92,98	90%	187,45	96,93
95%	246,25		95%	255,77	
99%	392,17		99%	397,47	
Ano de 2015	Percentil	Média	Ano de 2015	Percentil	Média
1%	0,00	60,05	1%	0,00	60,91
5%	0,00		5%	0,00	
10%	0,00	Mediana	10%	0,00	Mediana
25%	0,00	23,97	25%	0,00	10,79
50%	23,97		50%	10,79	
75%	86,58	Desvio padrão	75%	92,60	Desvio padrão
90%	171,35	91,01	90%	181,27	95,37
95%	241,30		95%	251,31	
99%	393,76		99%	405,08	
Ano de 2016	Percentil	Média	Ano de 2016	Percentil	Média
1%	0,00	55,31	1%	0,00	55,81
5%	0,00		5%	0,00	
10%	0,00	Mediana	10%	0,00	Mediana
25%	0,00	19,44	25%	0,00	6,07
50%	19,44		50%	6,07	
75%	80,53	Desvio padrão	75%	85,53	Desvio padrão
90%	156,18	87,86	90%	161,85	92,10
95%	216,26		95%	225,54	
99%	399,33		99%	412,32	
Observações em cada ano	5.564		Observações em cada ano	4.952	

Nota: as internações representam as taxas de internações para cada grupo de 10 mil habitantes.

Fonte: DataSus (Ministério da Saúde).