

Geolocalização de internações cadastradas no Sistema de Informações Hospitalares do Sistema Único de Saúde: uma solução baseada no programa estatístico R

doi: 10.5123/S1679-49742018000400016

Geolocation of hospitalizations registered on the Brazilian National Health System's Hospital Information System: a solution based on the R Statistical Software

Geolocalización de admisiones registradas en el Sistema de Informaciones Hospitalarias del Sistema Único de Salud de Brasil: una solución basada en el programa estadístico R

Thiago Augusto Hernandes Rocha¹ –  orcid.org/0000-0002-6262-3276

Núbia Cristina da Silva¹

Pedro Vasconcelos Maia Amaral²

Allan Claudius Queiroz Barbosa¹

João Ricardo Nickenig Vissoci³

Erika Bárbara Abreu Fonseca Thomaz⁴

Rejane Christine de Sousa Queiroz⁴

Matthew Harris⁵

Luiz Augusto Facchini⁶

¹Universidade Federal de Minas Gerais, Faculdade de Ciências Econômicas, Belo Horizonte, MG, Brasil

²Universidade Federal de Minas Gerais, Centro para o Desenvolvimento e Planejamento Regional, Belo Horizonte, MG, Brasil

³Duke University, Divisão de Medicina de Emergência, Carolina do Norte, Estados Unidos

⁴Universidade Federal do Maranhão, Departamento de Saúde Pública, São Luís, MA, Brasil

⁵Imperial College London, Faculdade de Medicina, Londres, Reino Unido

⁶Universidade Federal de Pelotas, Faculdade de Medicina, Pelotas, RS, Brasil

Resumo

Objetivo: apresentar uma solução para permitir a geolocalização de internações processadas junto ao Sistema de Informações Hospitalares do Sistema Único de Saúde do Brasil. **Métodos:** para espacializar as AIHs, estruturou-se um *script* em linguagem R, com base nos pacotes *microdatasus* e *CepR*; o *script* foi aplicado para levantamento das AIHs em Goiás, referentes ao ano de 2015; após o *download* e pré-processamento dos dados, foi detalhado o procedimento para espacialização das AIHs. **Resultados:** das 361.213 AIHs processadas, foi possível extrair 24.220 códigos de endereçamento postal (CEPs) diferentes; deste conjunto de CEPs, foram geolocalizados 23.910 (98,7%); estes CEPs geolocalizados permitiram a espacialização de 97,7% das AIHs processadas para o estado de Goiás. **Conclusão:** é possível espacializar as AIHs com alto índice de sucesso; o método detalhado neste trabalho abre um novo leque de possibilidades para o desenho de estudos avaliativos, formulação de políticas e planejamento de ações de cuidado em saúde.

Palavras-chave: Hospitalização; Sistemas de Informação; Processamento Automatizado de Dados; Análise Espacial; Sistemas de Informação Geográfica.

Endereço para correspondência:

Thiago Augusto Hernandes Rocha – Organização Pan-Americana da Saúde, Setor de Embaixadas Norte, lote 19, Brasília, DF,

Brasil. CEP: 70800-400

E-mail: rochahernandes3@gmail.com



Introdução

O Brasil se configura como um espaço privilegiado para discussões acerca da avaliação de serviços de saúde, em função da publicização de diversas informações sobre o aparato prestador de serviços. Nesse sentido, destaca-se o papel do Departamento de Informática do Sistema Único de Saúde (Datasus),¹ responsável por disponibilizar o acesso aos diversos sistemas de informações em saúde. Apesar da relevância dessa iniciativa, persistem dificuldades quanto à documentação das informações divulgadas, possibilidade de desagregação dos dados e manejo do grande número de arquivos.^{2,3}

A despeito das potencialidades associadas ao SIH/SUS, a maior parte dos dados divulgados possui nível de desagregação máximo até a esfera municipal. Assim, realizar análises voltadas para o exame de características intramunicipais é uma tarefa desafiadora.

Dentro do universo de sistemas de informações que compõem o rol de dados divulgados pelo Datasus, merece destaque o Sistema de Informações Hospitalares do SUS (SIH/SUS). O SIH/SUS é responsável pelo processamento das informações levantadas das autorizações de internação hospitalar (AIHs).⁴ O mesmo sistema monitora os recursos pagos a cada hospital do SUS, os códigos definidos na Décima Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde (CID-10) associados às internações, tempo de permanência média, código de endereçamento postal (CEP) de cada AIH, entre outras informações importantes. Esse sistema tem sido amplamente utilizado para subsidiar estudos avaliativos em saúde no Brasil.⁵⁻⁹

A despeito das potencialidades associadas ao SIH/SUS, a maior parte dos dados divulgados possui nível de desagregação máximo até a esfera municipal.⁴ Assim, realizar análises voltadas para o exame de características intramunicipais é uma tarefa desafiadora, implica a necessidade de coleta de dados primários, mais granulares que aqueles usualmente divulgados pelo Datasus. Além disso, a necessidade de manipular um grande volume de dados, pouco documentado, acaba por exigir do pesquisador o

manejo de ferramentas com capacidade de flexibilização restrita, permeadas por dificuldades de acesso e limitadas do ponto de vista de geração de informações, como é o caso do TabWin e do TabNet.¹⁰ Alguns esforços têm sido realizados para melhoria da documentação relacionada ao SIH/SUS.¹¹ Entretanto, desafios persistem.

Analisar os eventos ocorridos na esfera municipal é crucial para a realização de estudos avaliativos em saúde. Incorre em equívoco quem considera o território geográfico como expressão de características uniformes, especialmente no que concerne a eventos de saúde. Cidades como Belo Horizonte, por exemplo, possuem bairros com índices de desenvolvimento humano (IDHs) idênticos aos da Suíça, bem como regiões com IDH equivalente àqueles observados na África Subsaariana.¹² Ao se analisar indicadores municipais, ambos os extremos são considerados de forma agregada, produzindo um efeito de homogeneização de diferenças, catalisando erros associados à falácia ecológica.

Ponderar sobre determinantes sociais e geográficos é fundamental para melhorar a compreensão da díade saúde-doença.¹³ Levando-se em conta a importância de estudos baseados na localização geográfica de eventos em saúde¹⁴⁻¹⁶ e a relativa incipiência de trabalhos amparados nesse tipo de metodologia, no contexto brasileiro,¹⁵ pesa a necessidade do desenvolvimento de estratégias mais efetivas para a espacialização de eventos de saúde.

A importância de se desenvolver soluções capazes de permitir o exame de características intramunicipais, associada às limitações e dificuldades inerentes ao manejo do TabWin e do TabNet, contribuiu para o desenho do presente trabalho, cujo objetivo foi apresentar uma solução capaz de permitir a geolocalização de internações processadas junto ao SIH/SUS.

Desenvolvimento da solução

O presente estudo se configura como de natureza metodológica.¹⁷ Para permitir a espacialização de internações do SIH/SUS, foram geolocalizados, com a ajuda da linguagem R, os CEPs divulgados em cada AIH.

Para permitir a espacialização das internações processadas junto ao SIH/SUS, alguns desafios precisaram ser superados, a destacar: necessidade do manejo de múltiplos arquivos 'RD*.dbc', pelos quais são divulgados os dados; e extração dos dados do CEP referente a cada AIH, assim como sua geolocalização. Uma vez cumpridas essas etapas, é possível obter as coordenadas de latitude e longitude de cada AIH, com precisão restrita à rua de

origem do doente, possibilitando a realização de análises mais granulares que aquelas realizadas apenas com indicadores municipais agregados.

Para estruturação da solução, optou-se pela utilização do *software* R, o qual permite a inclusão de pacotes voltados à realização de funções específicas, a exemplo do manejo de dados do Datasus.

A primeira etapa do processo de geolocalização de AIHs perpassa o *download* dos dados divulgados junto ao SIH/SUS. As informações sobre as internações são divulgadas no endereço eletrônico `ftp://ftp.datasus.gov.br/dissemin/publicos/SIHSUS/200801_`; os arquivos são nomeados segundo o padrão RDUFAAMM, no qual RD é a abreviação para 'reduzida', UF para Unidade da Federação, AA para ano e MM para mês. Assim, para cada Unidade da Federação, há um arquivo reduzido mensal contendo os dados referentes às internações realizadas pelos estabelecimentos da respectiva UF. Santos¹¹ descreveu as informações contidas em cada variável divulgada nos arquivos 'RD*.dbc' do SIH/SUS. Para analisar um único ano, é necessário o manejo de 12 arquivos – um por mês – para cada uma das 27 UFs. É necessário, ainda, descompactar cada um desses arquivos, pois eles são divulgados em formato *.dbc e precisam ser convertidos ao formato *.dbf. Para abordar os desafios vinculados a esse processo oneroso, foi desenvolvido por Saldanha¹⁸ o pacote *microdatasus*.

O pacote do R *microdatasus* apresenta funções para *download* dos arquivos de microdados do Datasus (formato *.dbc), sua leitura pelo pacote *Read.dbc* e pré-processamento para utilização.¹⁸ A descompactação do formato *.dbc para *.dbf é realizada de forma automatizada, pelo pacote *Read.dbc*.³ Esses pacotes podem ser instalados a partir do *GitHub*, por meio do pacote *devtools*, diretamente na *front-end/GUI*, a partir do R ou do RStudio,¹⁹ mediante os seguintes comandos:

```
install.packages("devtools") # instala o pacote devtools
devtools::install_github("rfsaldanha/microdatasus")
install.packages("read.dbc")
```

O pacote *microdatasus* possui apenas duas funções: *fetch_datasus* e *process_sib*. A primeira se destina a automatizar os *downloads* de arquivos a partir do Datasus, já realizando o processo de descompactação e agregação dos arquivos individuais em um arquivo único. Ao realizar tal tarefa, essa função reduz de modo substancial

o número de operações necessárias à criação de um banco processável para os dados disponíveis no Datasus. A segunda função destina-se ao pré-processamento dos dados baixados do Datasus, atribuindo rótulos às variáveis brutas obtidas com a utilização da função *fetch_datasus*.

A função *fetch_datasus* possui sete argumentos:

- *year_start* : ano inicial
- *month_start* : mês inicial
- *year_end* : ano final
- *month_end* : mês final
- UF : Unidades da Federação
- *information_system* : sistema de informações
- *vars* : variáveis de interesse

Os argumentos iniciados com '*year*' ou '*month*' são destinados à definição do ano e mês de início e fim do conjunto de dados a serem baixados, respectivamente. O argumento UF define quais estados devem ser baixados. O argumento *information_system* define de quais sistemas os dados devem ser baixados. Neste trabalho, apesar de terem sido utilizados os dados referentes ao SIH/SUS, o pacote *microdatasus* é capaz de realizar *downloads* automatizados de outros sistemas, tais como o Sistema de Informações sobre Mortalidade (SIM), o Sistema de Informações sobre Nascidos Vivos (Sinasc) e o Cadastro Nacional de Estabelecimentos de Saúde (CNES). Futuramente, será incluído o Sistema de Informações Ambulatoriais do SUS (SIA/SUS).¹⁸ Por último, o argumento *vars* define quais variáveis serão baixadas.

Responsável pelo pré-processamento dos dados baixados do Datasus pela função *fetch_datasus*,¹⁸ a função *process_sib* possui apenas três argumentos:

- *data* : arquivo criado pela função *fetch_datasus*
- *information_system* : subsistema do SIH/SUS
- *municipality_data* : dados do município de residência

O argumento *data* define o arquivo que recebeu o resultado da função *fetch_datasus* sem nenhuma modificação. A função *process_sib* deve ser usada logo após a função *fetch_datasus*. O argumento *information_system* detalha de qual fonte são os dados armazenados no arquivo a ser processado. Por fim, o argumento *municipality_data* acrescenta ao arquivo informações sobre o município de residência, como nome completo, latitude e longitude.¹⁸ Para que as funções operem de modo satisfatório, é preciso

se conectar à internet e que o FTP do Datasus esteja funcionando. Informações mais detalhadas sobre o pacote podem ser obtidas na Wikipedia: <https://github.com/rfsaldanha/microdatasus/wiki>

A seguir, encontra-se detalhado um exemplo prático dos comandos utilizados para carregamento do pacote *microdatasus* no ambiente de trabalho do R e *download* dos dados do SIH/SUS para o estado de Goiás em 2015:

```
library(microdatasus) # carrega o pacote
microdatasus no ambiente de trabalho do R.

internacoes <- fetch_datasus(year_start = 2015,
month_st = 1, year_end = 2015, month_end =
12, uf = "GO", information_system = "SIH-RD")

str(internacoes)
```

Estes comandos geraram um arquivo com 361.213 internações para o ano de 2015 no estado de Goiás. Ao todo, foram reunidas 113 variáveis associadas a cada internação. O descritivo completo das variáveis pode ser obtido do trabalho de Santos.¹¹ A estrutura do banco de dados referente às seis primeiras variáveis é apresentada a seguir.

```
> str(internacoes)

'data.frame':   361213 obs. of  113 variables:

 $ UF_ZI:Factor w/ 172 levels
"520000", "520013": 1 1 1

 $ ANO_CMPT: Factor w/ 1 level "2015": 1 1 1

 $ MES_CMPT: Factor w/ 12 levels "01","02","03": 1 1 1

 $ ESPEC : Factor w/ 12 levels "01","02","03",...: 1 1 1

 $ CGC_HOSP : Factor w/ 106 levels
"00029827000128": NA NA NA

 $ CEP: Factor w/ 24220 levels
"03145010", "03728210": 6464 6447 6458
```

Cabe destacar que as internações não permitem a identificação unívoca dos doentes, pois um mesmo

doente pode ter passado por múltiplas internações. Há soluções para o *linkage* probabilístico de internações como aquela realizada pelo *OpenRecLink*. Entretanto, a aplicação desse tipo de técnica foge ao escopo do presente trabalho.²⁰ A função *process_sib* foi aplicada ao objeto 'internações' gerado, sendo responsável por rotular as variáveis baixadas do Datasus. Entre as variáveis registradas na AIH, está o CEP de referência da residência do doente admitido no estabelecimento de saúde, informação de cunho obrigatório para registro da AIH. Com a utilização do pacote *CepR*, é possível geolocalizar cada AIH com base no CEP fornecido pelo doente.

A base de dados do CEP brasileiro é monopolizada pela Empresa Brasileira de Correios e Telégrafos e sua utilização está condicionada ao pagamento de licença. Mesmo assim, a versão paga não contém os parâmetros de latitude e longitude associados a todos os CEPs do país. Para contornar essas dificuldades, foi desenvolvido, de modo colaborativo, o projeto CEP Aberto.²¹ Este projeto visa conceder acesso gratuito, bem como construir, de maneira colaborativa, uma base de dados com os CEPs geolocalizados de todo o Brasil. Atualmente, encontram-se registrados um total de 980.955 CEPs no país. O projeto CEP Aberto também elaborou uma *application programming interface* (API), de uso gratuito, para permitir pesquisar, a partir de um CEP, dados como: estado de localização, município, código telefônico, bairro, logradouro, latitude, longitude e altitude. Para utilizar a API, é necessário fazer um cadastro no endereço <http://cepaberto.com/>. Uma vez realizado o cadastro, é fornecido um *token* de acesso. Como o projeto é mantido de forma colaborativa, a API apresenta limitações de volume de consulta, sendo possível realizar uma consulta a cada três segundos e no máximo 10 mil consultas por *token*/dia. Considerando-se as funcionalidades do projeto CEP Aberto, Robert Myles desenvolveu um pacote R, denominado *CepR*, que consulta os dados de CEP junto ao projeto CEP Aberto diretamente do R.²² Integrando a solução do *CepR* e do *microdatasus*, foi desenvolvida uma rotina de *download*, pré-processamento e geolocalização de AIHs.

A seguir, são definidos os passos para integração das soluções de ambos os pacotes R. Primeiramente, é preciso sintetizar os CEPs obtidos junto ao SIH/SUS obtido junto à função *fetch_datasus*, para minimizar o número de consultas na API do CEP Aberto. Isto diminui o tempo necessário para o processo de geolocalização, visto que um CEP, geralmente, se repete várias vezes nos registros da AIH. Nesse sentido, o primeiro

passo é a criação de uma lista unívoca de CEPs. Os comandos a seguir exemplificam como fazê-lo, a partir de um arquivo resultante da função *fetch_datasus*.

Utiliza-se o mesmo arquivo de interações definido anteriormente. Para a realização das consultas, é necessário iterar sobre uma listagem de CEPs unívocos (*cep_unicos*). Para a realização dessa operação, faz-se mister a configuração de uma rotina de iteração (*loop*). Esse comando deverá consultar, para cada entrada de CEP listada no vetor *cep_unicos*, quais são as respectivas informações de localização geográfica. A criação de um arquivo para receber os dados de cada consulta de CEP realizada, assim como a parametrização do *loop*, são definidas nos passos descritos a seguir:

```
install.packages("cepR") # instala o pacote CepR
#cria um data frame temporário que será
utilizado para ser preenchido durante as
iterações relativas a cada CEP.
geo_coded<- data.frame(estado=character(),cid
ade=character
bairro=character(),cep=character(),logradouro
=character(),latitude=character(),longitude=ch
aracter(),altitude=character(),ddd=character()
,cod_IBGE=character(),quality = logical(),cep_
buscado = character(),stringsAsFactors=FALSE)

cep_unico <- as.character(unique(internacoes$C
EP)) # cria um vetor com os CEPS sem repetição
para minimizar a realização das consultas na API
do CEP Aberto
#Loop de consulta de cada CEP listado no vetor:
cep_unico
for (i in 1:10000) {
  sys1 <- Sys.time()
  consulta <-busca_cep(cep=(cep_unico[[i]]),
token='seu token')
  consulta$quality <- anyNA(c(consulta$latitude,
consulta$longitude))
  consulta$cep_buscado <- cep_unicos[[i]]
  geo_coded <- rbind(geo_coded, consulta,
make.row.names=FALSE)
  if(Sys.time()-sys1 <=4.0) Sys.sleep(4.0-(Sys.
time()-sys1))}
```

O *loop* em questão já considera o intervalo mínimo de tempo a ser respeitado para a realização de

consultas na API do CEP aberto. É preciso apenas modificar os parâmetros iniciais em negrito da seção (i in 1:10000), para que sejam fornecidos os índices de no máximo 10 mil CEPs por dia. Ademais, deve-se substituir esse trecho – seu *token* – pelo *token* fornecido ao se realizar o cadastro no sítio eletrônico do projeto CEP Aberto. O código de consulta guarda um registro de sucesso ou não, referente à busca realizada na coluna *quality*. Um valor *TRUE* indica sucesso na obtenção das coordenadas de latitude e longitude do CEP checado. Uma vez terminado o processo de consulta, é gerado o arquivo *geo_coded* com a seguinte estrutura:

```
str(geo_coded)
Classes 'tbl_df', 'tbl' and 'data.frame':
24220 obs. of 12 variables:
 $ estado   : chr "GO" "GO" "GO"
 $ cidade   : chr "São Simão" "Mineiros"
"Quirinópolis"
 $ bairro   : chr NA NA NA
 $ cep      : chr "75890000" "75830000"
"75860000"
 $ logradouro : chr "São Simão" "Mineiros"
"Quirinópolis" "AC Jataí, Avenida Dorival de
Carvalho, 1007"
 $ latitude  : chr "-18.9964906" "-17.5624415"
"-18.4476442"
 $ longitude : chr "-50.547432" "-52.5489206"
"-50.4551598"
 $ altitude  : chr "478.400000" "787.900000"
"512.900000"
 $ ddd       : chr NA NA NA "64" ...
 $ cod_IBGE  : chr "5220405" "5213103"
"5218508"
 $ quality   : logi TRUE TRUE TRUE
 $ cep_buscado: chr "75890000" "75830000"
"75860000"
```

Os passos detalhados foram aplicados junto aos dados das AIHs para o estado de Goiás relativas a 2015. Os resultados do processo de espacialização de AIHs foram plotados com o auxílio do *software* ARCMAP.²³ Apesar da opção pelo uso do ARCMAP, qualquer *software* de GIS pode ser utilizado com o mesmo fim, tal como as soluções livres QGIS, GEODA e TerraView.

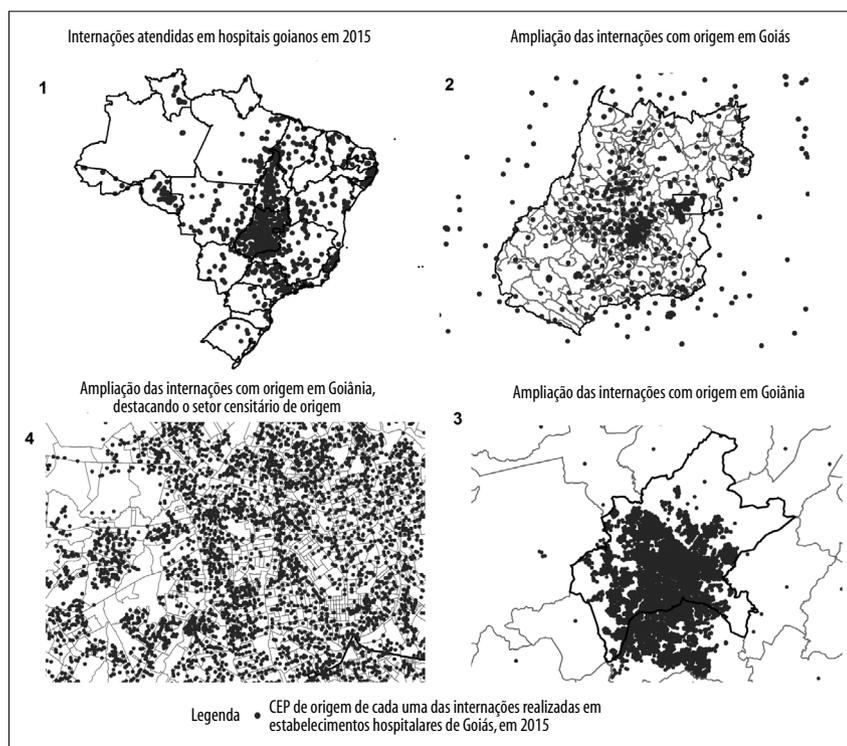


Figura 1 – Distribuição de autorizações de internação hospitalar em estabelecimentos de saúde, por residência do doente, realizadas em Goiânia, Goiás, 2015

Resultados

Dos 24.220 CEPs únicos, 23.910 (98,7%) foram geolocalizados, correspondendo a 353.004 AIHs de 361.213 internações possíveis (97,7%). Não houve ocorrência de CEP extraído do SIH/SUS com número de dígitos inferior a oito. Os CEPs não geolocalizados corresponderam a CEPs genéricos de cidades inteiras ou grandes regiões, e aos CEPs que ainda não haviam sido inseridos nas bases do projeto CEP Aberto. Quanto à espacialização dos pontos referentes às AIHs, observou-se que doentes oriundos de todos os estados brasileiros foram admitidos em algum hospital goiano ao longo de 2015 (Figura 1); cada ponto indica o local de residência do doente admitido em um estabelecimento de saúde goiano no ano de 2015. Cada item da Figura 1 exhibe as AIHs geolocalizadas, com diferentes níveis de granularidade de visualização.

A Figura 2 apresenta, somente para a região de Goiânia e adjacências, pontos na cor cinza correspondentes a cada AIH geolocalizada, frente às unidades básicas de saúde (UBS), malha de arruamento e rede fluvial. Ao sobrepor diferentes camadas de informação

geográfica, é possível analisar, com o uso de técnicas de geoprocessamento e vigilância, prováveis relações entre padrão de internações e relacionamento com elementos geográficos. Como o SIH/SUS disponibiliza os códigos da CID-10, é possível analisar a dependência espacial de doenças específicas e elementos geográficos. Outrossim, como o SIH/SUS cobre dados desde 2008, é possível desenvolver análises em formato de série temporal.

A Figura 3 demonstra como determinantes sociais em saúde, do tipo de cuidado prestado por equipes de atenção básica e volume de internações em uma região específica, podem ser analisados conjuntamente, com um grau de granularidade até então não reportado na literatura. Os círculos ao redor de cada cruz desenhada na Figura 3 correspondem a uma distância de 3km de cada UBS. Considerando-se técnicas de captura, é possível atribuir uma carga de internações às equipes da Atenção Primária à Saúde, ponderada por informações sociodemográficas vinculadas aos setores censitários, por exemplo. Essa carga de internações pode ser utilizada no delineamento de estudos avaliativos quase experimentais.

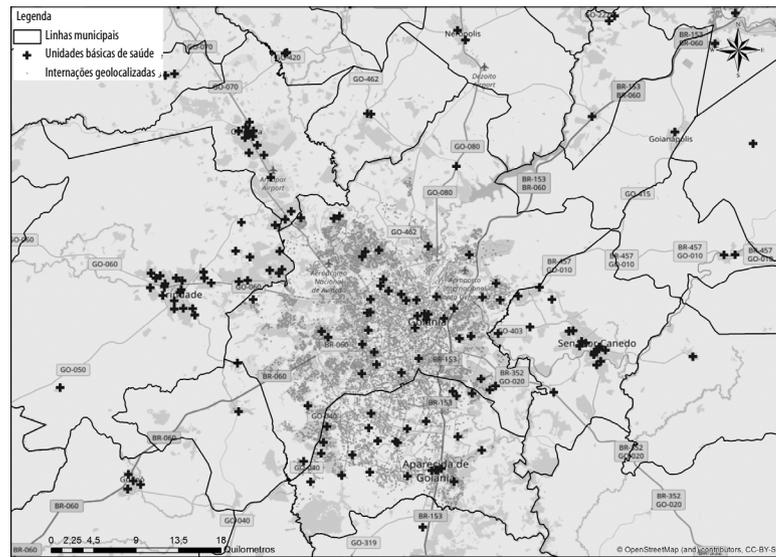


Figura 2 – Autorizações de internação hospitalar geolocalizadas e sua relação com outros elementos geográficos em Goiânia e adjacências, Goiás, 2015

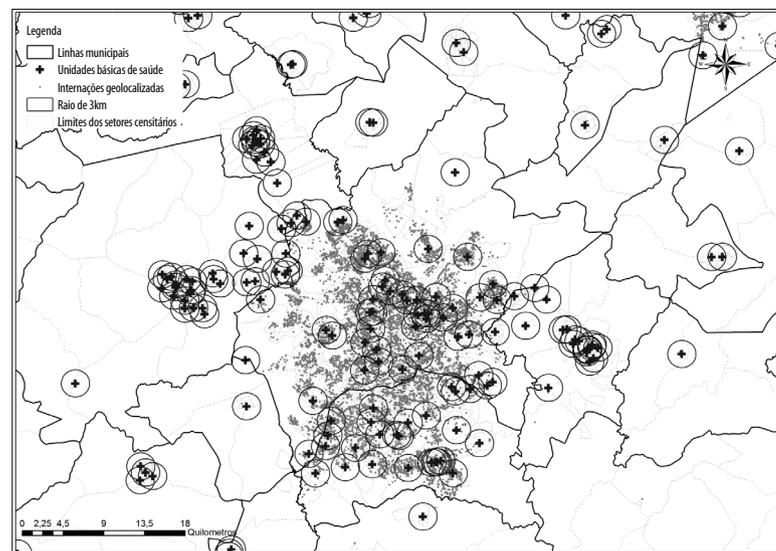


Figura 3 – Relação entre unidades básicas de saúde, internações e setores censitários de Goiânia e adjacências, Goiás, 2015

Discussão

A aplicação de métodos da geografia da saúde para fins de solução de problemas metodológicos associados a estudos avaliativos configura-se como uma agenda inconclusa, todavia. O presente trabalho pretendeu dar uma pequena contribuição à ampliação das possibilidades analíticas vinculadas ao SIH/SUS. Os achados fornecidos são promissores, considerando-se

a potencialidade para o desenvolvimento de estudos de maior qualidade, a proposição de políticas públicas amparadas por evidências locais e a qualificação do processo de planejamento de oferta do cuidado.

Graças ao delineamento proposto, abre-se a possibilidade para melhor investigar uma miríade de questões. Análises sobre determinantes sociais de saúde poderão ser mais bem embasadas. A geolocalização das AIHs permite a análise concomitante de perfis de incidência

de agravos e outros elementos sociais, como perfil de população residente, nível socioeconômico, grau de urbanização e outros dados divulgados, por exemplo, a partir dos setores censitários. Quanto a variáveis ambientais, é possível confrontar bases de dados sobre poluentes, arruamento e cursos fluviais, entre outras, com a ocorrência de doenças em regiões específicas.

A análise das redes de cuidado poderá se beneficiar do método aqui proposto. Será possível analisar o local de residência de um doente e comparar essa informação com a localização do estabelecimento de saúde que lhe prestou atendimento. Esse dado pode fornecer um *insight* poderoso sobre os fluxos reais de doentes em busca de cuidado. Fluxos regulares e constantes de doentes sob determinadas condições de saúde, de acessibilidade inclusive, podem fornecer a profissionais de saúde informações sobre regiões que lidam com carência de serviços.

Análises de equidade e acesso a serviços podem ser mais bem detalhadas e áreas de abrangência a serviços podem ser definidas, com base em técnicas de geoprocessamento. A qualidade do cuidado prestado por hospitais, unidades básicas de saúde e unidades de pronto atendimento (UPAs) podem ser avaliadas de acordo com parâmetros de proximidade geográfica e áreas de influência potencial. Técnicas de *linkage* probabilístico podem ser otimizadas com a inclusão de parâmetros de localização geográfica, aumentando seu grau de precisão. Cargas de internações podem ser vinculadas a serviços de saúde específicos, como no caso da Atenção Primária à Saúde. Ações de planejamento da oferta de cuidado podem ser feitas de forma mais detalhada, haja vista a possibilidade de analisar o perfil epidemiológico de uma determinada população adscrita a uma região geográfica. A disponibilização de dados mais granulares abre um leque de opções para gestores, profissionais de saúde, formuladores de políticas públicas e acadêmicos.

A despeito dos avanços e da potencialidade do método desenvolvido, pesam algumas limitações a serem abordadas. A primeira delas remete à presença de algum grau de imprecisão na localização da origem do doente, em função do uso do CEP. Como a geolocalização é feita com base no código do arruamento, ela é marcada por algum grau de incerteza com poder de afetar análises mais sensíveis, gerando imprecisões que remetem a problemas em pequenas áreas. Outra limitação da geolocalização a considerar: em função das limitações do volume de consultas da API, o processo desenvolvido é de realização lenta, de maneira que grandes volumes de

dados podem consumir um tempo substancial até serem geolocalizados. Para os CEPs genéricos, que abrangem cidades inteiras e grandes regiões e possuem final '000', a técnica é incapaz de proceder a localização precisa. Por último, cabe destacar a dependência de uma solução colaborativa sem financiamento regular. O projeto CEP Aberto é fundamental para a operacionalização da solução aqui proposta. Ademais, nem todos os CEPs brasileiros encontram-se nas bases do projeto CEP Aberto, contribuindo para algum nível de imprecisão que deve ser analisado, caso a caso. Sem o projeto CEP Aberto, o método aqui definido perde sua capacidade propositiva.

Não obstante as limitações apresentadas, a solução objeto deste trabalho é marcada por mais potencialidades que limitações. A aplicação da técnica, descrita em detalhes, mostra-se capaz de fomentar novas pesquisas com dados do SIH/SUS. A perspectiva histórica do Sistema de Informações Hospitalares do Sistema Único de Saúde e o novo grau de granularidade obtido com o método de geolocalização podem servir de subsídio para o delineamento de pesquisas em Saúde Pública quase experimentais. Esse tipo de desenho é capaz de fornecer evidências mais robustas sobre políticas e programas de saúde, e deve ser incentivado.

Agradecimentos

A Fúlvio Nedel, Danieal Petruzalek, Daniel Saldaña e Robert Myles, pela doação de tempo, esforço e dedicação no desenvolvimento de soluções voltadas para a promoção de melhorias no SUS. À CAPES, pela concessão de bolsa de doutorado modelo “sanduíche” ao primeiro autor do trabalho. À Organização Pan-Americana da Saúde.

Contribuição dos autores

Rocha TAH participou da concepção, programação e redação, revisão bibliográfica, análise de dados, discussão dos elementos textuais e revisão crítica do manuscrito. Silva NC, Amaral PVM, Barbosa ACQ e Visoci JRN participaram da redação, revisão bibliográfica e discussão dos elementos textuais do manuscrito. Thomaz EBAF, Queiroz RCS, Harris M e Facchini LA participaram da discussão dos elementos textuais e revisão crítica do manuscrito. Todos os autores aprovaram a versão final do manuscrito e são responsáveis por todos os aspectos do trabalho.

Referências

1. Ministério da Saúde (BR). Departamento de Informática do SUS – Datasus. Informações de saúde (TabNet) [Internet]. 2015 [citado 208 nov 8]. Disponível em: <http://www2.datasus.gov.br/DATASUS/index.php?area=02>
2. Rocha TAH, Rocha J, Silva N, Amaral P, Facchini L, Thumé E, et al. Cadastro nacional de estabelecimentos de saúde: evidências sobre a confiabilidade dos dados. *Ciêns Saúde Coletiva*. 2017 jan;23(1):229-40. doi: 10.1590/1413-81232018231.16672015.
3. Petruzalek D. READ.dbc – um pacote para importação de dados do Datasus na linguagem R [Internet]. In: Anais do XV Congresso Brasileiro de Informática em Saúde; 2016 27 nov – 30 nov [citado 2018 nov 8]; Goiânia, Brasil. Disponível em: http://docs.bvsalud.org/biblioref/2018/07/906543/anais_cbis_2016_artigos_completos-601-606.pdf
4. Ministério da Saúde (BR). Departamento Nacional de Auditoria do SUS. Coordenação-Geral de Desenvolvimento Normatização e Cooperação Técnica. Auditoria no SUS: noções básicas sobre sistemas de informação [Internet]. Brasília: Ministério da Saúde; 2004 [citado 2018 nov 8]. 94 p. Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/auditoria_sus.pdf
5. Gerhardt TE, Pinto JM, Riquinho DL, Roese A, Santos DL, Lima MCR. Utilização de serviços de saúde de atenção básica em municípios da metade sul do Rio Grande do Sul: análise baseada em sistemas de informação. *Ciêns Saúde Coletiva*. 2011; 16(suppl 1):1221-32. doi: 10.1590/S1413-81232011000700054.
6. Bittencourt SA, Camacho LAB, Leal MC. O Sistema de informação hospitalar e sua aplicação na saúde coletiva hospital. *Cad Saúde Pública*. 2006 jan;22(1):19-30. doi: 10.1590/S0102-311X2006000100003.
7. Loyola Filho AI, Leite Matos D, Giatti L, Afradique ME, Viana Peixoto S, Lima-Costa ME. Causas de internações hospitalares entre idosos brasileiros no âmbito do Sistema Único de Saúde. *Epidemiol Serv Saúde*. 2004 dez;13(4):229-38. doi: 10.5123/S1679-49742004000400005.
8. Escosteguy CC, Portela MC, Medronho RA, Vasconcellos MT. The Brazilian hospital information system and the acute myocardial infarction hospital care. *Rev Saúde Pública*. 2002 Aug;36(4):491-9.
9. Schramm JM, Szwarcwald CL. Sistema hospitalar como fonte de informações para estimar a mortalidade neonatal e a natimortalidade. *Rev Saúde Pública*. 2000 jun;34(3):272-9. doi: 10.1590/S0034-89102000000300010.
10. Silva NP. A utilização dos programas TabWin e TabNet como ferramentas de apoio a disseminação das informações em saúde [dissertação]. Rio de Janeiro (RJ): Fundação Oswaldo Cruz; 2009.
11. Santos AC. Sistema de informações hospitalares do Sistema Único de Saúde: documentação do sistema para auxiliar o uso das suas informações [dissertação]. Rio de Janeiro (RJ): Fundação Oswaldo Cruz; 2009.
12. Fonseca B, Silva K. Atribuição de IDH aos bairros de Belo Horizonte [Internet]. *Rev Transite*. 2017 [cited 2017 set 5]. Disponível em: <http://transite.fafich.ufmg.br/idh-bairros-de-belo-horizonte/>
13. Guimarães RB. Geografia e saúde coletiva no Brasil. *Saúde e Soc*. 2016 out-dez;25(4):869-79. doi: 10.1590/s0104-12902016167769.
14. Kearns R, Moon G. From medical to health geography: novelty, place and theory after a decade of change. *Prog Hum Geogr*. 2002 Oct;26(5):605-25. doi: 10.1191/0309132502ph389oa.
15. Macintyre S, Ellaway A, Cummins S. Place effects on health: How can we conceptualise, operationalise and measure them? *Soc Sci Med*. 2002 Jul;55(1):125-39.
16. Dummer TJB. Health geography: supporting public health policy and planning. *CMAJ*. 2008 Apr;178(9):1177-80. doi: 10.1503/cmaj.071783.
17. Lima DVM. Research design: a contribution to the author. *Online Brazilian J Nurs*. 2011;10(2):1-18. doi: 10.17665/1676-4285.2011v10n2.
18. Github. Download de dados do DataSUS e pré-processamento no R [Internet]. 2017 [citado 2018 nov 7]. Disponível em: <https://github.com/rfsaldanha/downloadDataSUS>
19. Nedel FB. csapAIH: uma função para a classificação das condições sensíveis à atenção primária no programa estatístico R*. *Epidemiol Serv Saúde*. 2017 jan-mar;26(1):199-209. doi: 10.5123/S1679-49742017000100021.
20. Camargo Júnior KR, Coeli CM. Going open source: some lessons learned from the development of OpenReLink. *Cad Saúde Pública*. 2015 Feb;31(2):257-63.

21. CEP Aberto. O programa CEP aberto [Internet]. 2017 [citado 2017 set 5]. Disponível em: <http://cepaberto.com/>
22. Github. Um pacote R para buscar informações sobre CEPs, endereços, bairros e cidades. (An R package for accessing Brazilian postal code data). [Internet]. 2017 [citado 2018 nov 7]. Disponível em: <https://github.com/RobertMyles/cepR>
23. Environmental Systems Research Institute. ArcGIS desktop: release 10.3 [Internet]. 2014 [cited 2018 Nov 8]. Available in: <https://www.esri.com/esri-news/releases/15-1qtr/arcgis-10-3-and-arcgis-pro-modernize-gis-for-organizations-and-enterprises>

Abstract

Objective: to describe a solution enabling geolocation of hospital admissions (AIH), processed on the Brazilian National Health System's Hospital Information System. **Methods:** in order to spatialize AIHs an R language script was written, based on the microdatasus and CepR packages; the script was applied to identify all AIHs in Goiás state in the year 2015; after downloading and pre-processing the data, the procedure for AIH spatialization was detailed. **Results:** of the 361,213 AIHs processed, we were able to retrieve 24,220 different ZIP codes (CEPs); from this set of ZIP codes, 23,910 (98.7%) were geolocated; these geolocated ZIP codes enabled spatialization of 97.7% of AIHs processed for the state of Goiás. **Conclusion:** it is possible to spatialize AIHs with a high success rate; the method detailed in this paper opens a new range of possibilities for the design of evaluative studies, formulation of policies and planning of health care actions.

Keywords: Hospitalization; Information Systems; Automatic Data Processing; Spatial Analysis; Geographic Information Systems.

Resumen

Objetivo: presentar una solución para permitir la geolocalización de internaciones procesadas junto al Sistema de Informaciones Hospitalarias del Sistema Único de Salud de Brasil. **Métodos:** con la finalidad de espacializar las AIHs, se elaboró un script en lenguaje R, con base en los paquetes microdatosus y CepR; el script se aplicó a todos los AIHs realizados en el estado de Goiás referentes a 2015; después de descargar y pre-procesar los datos, se detalló el procedimiento para la espacialización de AIHs. **Resultados:** de los 361.213 AIHs procesados, fue posible extraer 24.220 diferentes códigos postales (CEP); de este conjunto de códigos postales, 23.910 (98,7%) fueron geolocalizados; estos CEPs geolocalizados permitieron la espacialización del 97,7% de los AIHs registrados en Goiás. **Conclusión:** es posible espacializar las AIHs con una alta tasa de éxito; el método detallado en este documento abre una nueva gama de posibilidades para el diseño de estudios evaluativos, la formulación de políticas y la planificación de acciones de salud.

Palabras clave: Hospitalización; Sistemas de Información; Procesamiento Automatizado de Datos; Análisis Espacial; Sistemas de Información Geográfica.

Recebido em 22/02/2018
Aprovado em 01/11/2018