UNIVERSIDADE FEDERAL DE CIÊNCIAS DA SAÚDE DE PORTO ALEGRE – UFCSPA

PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E GESTÃO EM SAÚDE

Amauri Duarte da Silva

Vigilância Epidemiológica: Uso de Aprendizado de Máquina na Predição de Surtos de Síndrome Respiratória Aguda Grave

Porto Alegre

2022

Amauri Duarte da Silva

Vigilância Epidemiológica: Uso de Aprendizado de Máquina na Predição de Surtos de Síndrome Respiratória Aguda Grave

Dissertação submetida ao Programa de Pós-Graduação em Tecnologias da Informação e Gestão em Saúde da Universidade Federal de Ciências da Saúde de Porto Alegre como requisito para a obtenção do grau de Mestre.

Orientadora: Profª Drª Ana Beatriz Gorini da Veiga

Co-orientador: Prof. Dr. Marcelo Ferreira da Costa Gomes

Porto Alegre

2022

Catalogação na Publicação

Silva, Amauri Duarte da
Vigilância epidemiológica: uso de aprendizado de máquina na predição de surtos de síndrome respiratória aguda grave / Amauri Duarte da Silva. -- 2022.
130 p. : il., graf., tab. ; 30 cm.

Dissertação (mestrado) -- Universidade Federal de Ciências da Saúde de Porto Alegre, Programa de Pós-Graduação em Tecnologias da Informação e Gestão em Saúde, 2022.

Orientador(a): Ana Gorini da Veiga ; coorientador(a): Marcelo Ferreira da Costa Gomes.

1. SRAG. 2. Inteligência Artificial. 3. Aprendizado de Máquina. 4. Modelos Preditivos. 5. Redes Neurais. I. Título.

Sistema de Geração de Ficha Catalográfica da UFCSPA com os dados fornecidos pelo(a) autor(a).

Amauri Duarte da Silva

Vigilância Epidemiológica: Uso de Aprendizado de Máquina na Predição de Surtos de Síndrome Respiratória Aguda Grave

Dissertação submetida ao Programa de Pós-Graduação em Tecnologias da Informação e Gestão em Saúde da Universidade Federal de Ciências da Saúde de Porto Alegre como requisito para a obtenção do grau de Mestre.

Porto Alegre, xx de março de 2022. [completar após a aprovação do trabalho]

BANCA EXAMINADORA

_____

Prof. Dr. Flávio Codeço Coelho

Fundação Getulio Vargas

_____

Prof. Dr. Airton Tetelbom Stein

Universidade Federal de Ciências da Saúde de Porto Alegre

_____

Profª Drª Juliana Silva Herbert

Universidade Federal de Ciências da Saúde de Porto Alegre

DEDICATÓRIA

Dedico esses dois anos de pesquisa e os resultados dela aos meus pais Diva (in memorian) e Artêmio.

AGRADECIMENTOS

Não há como caminhar sozinho na realização de um projeto. Ao longo dessa jornada, contei com a ajuda, incentivo e colaboração de inúmeras pessoas. Recebam meu agradecimento e reconhecimento:

Profa. Dra. Ana Gorini da Veiga, que me deu a grande oportunidade e a honra de ser seu orientando, de ter acreditado em alguém que ela não conhecia. Agradeço a ela também pelo aprendizado adquirido e à tanta atenção, apoio e incentivo à qualquer hora ou dia da semana.

Prof. Dr. Marcelo Ferreira da Costa Gomes, pela confiança em ter aceitado o convite para orientar alguém que ele não conhecia e que, na época, teria que co-orientar à distância. Agradeço por ter me passado tanto conhecimento numa área que era pouco conhecida por mim.

Dra. Tatiana Schäffer Gregianini (LACEN/RS) e Leticia Garay Martins (CEVS/RS) pela colaboração constante e disponibilidade para as reuniões iniciais do projeto.

Professores do PPGTIG que, além do conhecimento transmitido, trataram-me sempre com muito carinho e atenção e que, além disso, me deram exemplos da valiosa arte de ensinar.

SERPRO - Serviço Federal de Processamento de Dados que, através do seu Programa de Educação Pós-graduada, disponibilizou precioso tempo para que eu pudesse realizar o mestrado com mais dedicação. Nesse contexto, agradeço aos meus gerentes Joseane Pedroso, Fábio Cunha e Cleber Sana, pelo apoio dado antes e durante o mestrado.

Meus colegas de mestrado que, mesmo com o distanciamento da pandemia, construíram comigo relações de amizade e apoio mútuo.

Minha esposa Márcia e minha filha Nicole, pelo incentivo, apoio e compreensão durante esses dois anos.

Com certeza não estou agradecendo aqui a todos que eu gostaria; seriam necessárias muitas páginas, por isso expresso minha profunda gratidão àqueles que, de alguma forma e em algum momento, estiveram junto comigo nessa caminhada.

RESUMO

Introdução: Surtos de Síndrome Respiratória Aguda Grave (SRAG) ocorrem anualmente em diversas regiões geográficas, com o período dos picos de notificação de casos de internação variando de região para região. Os gestores da saúde necessitam de ferramentas para planejamento de recursos adequados a cada temporada de SRAG. **Objetivo**: Com o uso de Aprendizado de Máquina (AM), gerar modelos para a predição de surtos de SRAG para um próximo período (ano). **Método**: Foram utilizados dados de notificação de casos de internação por SRAG no Brasil, do período de 2013 a 2021, excluindo casos de SRAG ocasionados por Covid-19. Esses dados foram preparados de forma a alimentar uma rede neural configurada para gerar modelos preditivos para séries temporais. A rede neural foi implementada com uma ferramenta de pipeline que permitiu, de forma rápida, a simulação de inúmeros modelos. Foram gerados modelos para as cinco regiões brasileiras, porém são destacados os modelos preditivos para as regiões Sul e Sudeste. A validação dos modelos gerados foi feita para diferentes anos de ocorrência de surtos de SRAG. Na validação de temporadas com menor número de períodos (anos), os modelos preditivos tiveram menor desempenho devido à menor quantidade de dados disponíveis para treinamento. **Resultados**: Com o uso de redes neurais foi possível gerar modelos preditivos que possibilitam predizer curvas com picos de SRAG, volume de casos por temporada e início do período pré epidêmico com boa qualidade (R^2 = 0,91, considerando a temporada observada de 2019). **Conclusão**: Mesmo com uma quantidade deficiente de dados é possível, com o uso de AM, gerar modelos preditivos para SRAG, contribuindo para medidas de controle de surtos e epidemias.

Palavras-chave: Aprendizado de Máquina, Inteligência Artificial, Modelos Preditivos, Redes Neurais, SRAG, Vigilância Epidemiológica.

ABSTRACT

Introduction: Outbreaks of Severe Acute Respiratory Infection (SARI) occur annually in several geographic regions, with the period of peak notification of hospitalizations varying between regions. Health managers need adequate resource planning tools for each SARI season. **Objective**: Using Machine Learning (ML), generate models for the prediction of SARI outbreaks for a next period (year). **Method**: Data from notification of SARI hospitalization cases in Brazil, from 2013 to 2021, were used, excluding SARI cases caused by Covid-19. These data were prepared in order to feed a neural network configured to generate predictive models for time series. The neural network was implemented with a pipeline tool that quickly allowed the simulation of numerous models. Models were generated for the five Brazilian regions, and the predictive models for the South and Southeast regions are highlighted. The validation of the generated models was performed for different years of SARI outbreaks. In the validation of seasons with fewer periods (years), the predictive models had lower performance due to the smaller amount of data available for training. **Results**: Using neural networks, it was possible to generate predictive models that enabled the prediction of curves with SARI peaks, volume of cases per season and the beginning of pre-epidemic periods with good quality (R^2 = 0.91, considering the observed season of 2019). **Conclusion**: Even with a deficient quantity of data, it is possible, with the use of ML, to generate predictive models for SARI, contributing for the management of outbreaks and epidemics.

Keywords: Artificial Intelligence, Epidemiological Surveillance, Neural Networks, Machine Learning, Predictive Models, SARI.

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS

AM	Aprendizado de máquina
ARIMA	*Autoregressive integrated moving average*
CI	*Credible interval*
CNES	Cadastro nacional de estabelecimentos de saúde
CNNs	*Convolutional neural networks*
IA	Inteligência artificial
IAE	Inteligência artificial explicável
IAV	Vírus Influenza A
IBV	Vírus Influenza B
IFD	Imunofluorescência direta
KNN	*K-nearest neighbor*
LACEN	Laboratório central
LSTM	*Long short term memories*
MAE	*Mean absolute error*
MLP	*Multi-layer perceptron*
MLR	*Multiple linear regression*
MSE	*Mean square error*
OMS	Organização mundial da saúde
PCR	*Polymerase chain reaction*
qPCR	PCR em tempo real
R^2	Coeficiente de determinação
RF	*Random forest*
RMSE	*Root mean square error*
RNDS	Rede nacional de dados em saúde
RT-qPCR	Transcrição reversa seguida por qPCR
SARI	*Severe acute respiratory infection*
SARS-CoV-2	*Severe acute respiratory syndrome coronavirus 2*
SE	Semana epidemiológica
SG	Síndrome gripal
SINAN	Sistema de informação de agravos de notificação
SIVEP-Gripe	Sistema de informação da vigilância epidemiológica da gripe
SRAG	Síndrome respiratória aguda grave

SUS	Sistema único de saúde
SVM	*Support vector machine*
SVR	*Support vector regression*
TSA	Teste de sensibilidade aos antivirais
UFCSPA	Universidade Federal de Ciências da Saúde de Porto Alegre
UV	Luz ultravioleta
VRS	Vírus respiratório sincicial

SUMÁRIO

1 APRESENTAÇÃO

Este trabalho é resultado de dois anos de pesquisa, tendo como foco a busca de modelos preditivos que descrevessem curvas de surtos de SRAG para períodos futuros. Para isto foram identificados e analisados dados de interesse que pudessem agregar qualidade aos modelos. Buscou-se, também, selecionar algoritmos de Aprendizado de Máquina (AM) que pudessem gerar modelos preditivos com bom desempenho.

A estrutura desta dissertação passa primeiramente por uma revisão da literatura, onde são apresentados os assuntos que contribuíram para o atingimento dos objetivos propostos e suas referências bibliográficas. Na seção seguinte são apresentados os objetivos do trabalho. Em seguida é apresentada a situação do manuscrito, enviado a um periódico, com os principais resultados da pesquisa. Na sequência está incluída a íntegra do artigo. Na seção 6 são apresentadas as principais conclusões da pesquisa e, na seção 7, são apresentados os resultados adicionais, que foram entregas de capítulos de livros, outro artigo submetido a periódico e a descrição de uma ferramenta para a visualização dos resultados dos modelos preditivos gerados.

2 REVISÃO DA LITERATURA

2.1 VIGILÂNCIA DE INFECÇÕES RESPIRATÓRIAS

2.1.1 REDE DE VIGILÂNCIA

A vigilância de infecções respiratórias no Brasil é dividida em síndrome gripal (SG) e Síndrome Respiratória Aguda Grave (SRAG). A identificação de SG considera qualquer indivíduo que apresente febre, acompanhada de tosse ou dor de garganta sem diagnóstico adicional; enquanto a SRAG é caracterizada por um quadro de febre (acima de 37,5ºC), mesmo que referida, acompanhada de tosse ou dor de garganta e dificuldade respiratória (dispneia) ou saturação de oxigênio menor que 95% em ar ambiente ou, ainda, desconforto respiratório (MARTINS et al., 2011; FREITAS, 2013).

O Brasil adota uma rede de vigilância sentinela para influenza e outros vírus respiratórios que começou com uma única unidade sentinela em 2000 e, gradualmente, se expandiu para 60 unidades em 2010, com pelo menos uma unidade em 26 das 27 unidades federativas (26 estados e Distrito Federal). A rede de unidades sentinela é composta de ambulatórios, unidades de emergência e hospitais gerais, com registro no Cadastro Nacional de Estabelecimentos de Saúde (CNES) (BRASIL, 2020a), que relatam semanalmente o número total agregado de visitas e o total de visitas devido a sintomas de SG por meio de um sistema online chamado Sistema de Informação de Vigilância Epidemiológica da Gripe (SIVEP-Gripe) (BRASIL, 2020b). Essas unidades atuam na identificação, registro, investigação e diagnóstico de casos suspeitos e confirmados. Como meta estabelecida para as unidades sentinelas de influenza temos: para SG, realizar, no mínimo, 80% de registros (SIVEP-Gripe) e coleta de material por semana epidemiológica (SE), cinco amostras de secreção da nasofaringe; alimentar o Sistema SIVEP-Gripe semanalmente e informar proporção de atendimentos por SG, em relação ao total de atendimentos no serviço, semanalmente. Já para SRAG, devem registrar, no mínimo 80% dos casos de SRAG internados em UTI, com a devida coleta de amostra, envio ao Laboratório Central de Saúde Pública (LACEN) e digitação no SIVEP-Gripe, semanalmente; coletar, do total dos casos registrados, 80% de amostra; realizar em 90% das semanas epidemiológicas registro do número

de internações que ocorreram na instituição por CID 10: J09 a J18[1]. Quando essas metas não forem atingidas, deverá ser elaborada uma justificativa e a avaliação dessa justificativa poderá acarretar suspensão de repasses financeiros. (BRASIL, 2019)

No Brasil, a SRAG tornou-se agravo de notificação universal em resposta à pandemia de influenza A H1N1 em 2009 (COSTA, 2016). Até 2018 a notificação era realizada no Sistema de Informação de Agravos de Notificação (SINAN) (BRASIL, 2021b), quando então migrou para o SIVEP-Gripe. O sistema contém diversos dados relacionados a cada caso de SRAG, como por exemplo: idade, sexo, sintomas, comorbidades e outros fatores de risco; data dos primeiros sintomas; informações sobre vacina; exames de PCR para diversos tipos de vírus; data de hospitalização, data de alta ou óbito; unidade de internação e localização geográfica do paciente. Outros tipos de informações podem ser obtidos com o cruzamento da base de dados do CNES com os dados do SIVEP-Gripe. Desta forma, além de ser possível realizar uma análise epidemiológica da SRAG, é possível traçar um perfil das internações com caracterização do tipo de leito (particular ou SUS), grupo jurídico gestor da unidade, tipo de gestão por região, etc.

2.1.2 VIGILÂNCIA DE INFLUENZA NA REDE LABORATORIAL

No Brasil, as informações da vigilância da influenza são geradas pela rede de vigilância epidemiológica e também pela rede laboratorial (vigilância virológica). Nessa rede laboratorial estão os LACENs, que fazem a análise das amostras coletadas em pacientes que têm SG ou SRAG. Essas coletas são feitas de forma sistemática, permitindo a identificação e caracterização dos vírus que circulam no Brasil. Os Laboratórios de Referência fazem análises complementares para a identificação viral, como a caracterização antígena e genética. Também fazem teste de sensibilidade a antivirais e isolamento viral.

A base da informação utilizada para vigilância, a partir da identificação do agente etiológico, tipagem e subtipagem de vírus influenza circulantes, são de responsabilidade dos LACENs. Os LACENs realizam o processamento inicial das amostras coletadas, incluindo aliquotagem, estocagem e diagnóstico laboratorial viral. Os Laboratórios de Referência nacionais (Instituto Evandro Chagas, Fiocruz e Instituto Adolfo Lutz) recebem as amostras já processadas pelos LACENs para

[1] Influenza [gripe] e pneumonia

análises complementares, como por exemplo sequenciamento genômico do vírus (BRASIL, 2019).

2.1.3 VIGILÂNCIA EPIDEMIOLÓGICA

No início do século XX ocorreram as primeiras intervenções do Estado na prevenção e controle de doenças, utilizando bases científicas modernas. Eram orientadas pelos avanços da era bacteriológica e também pela descoberta dos ciclos epidemiológicos de algumas doenças infecciosas e parasitárias. Na época, as intervenções eram feitas em termos de organização de grandes campanhas sanitárias que objetivavam o controle de doenças que comprometem as atividades econômicas. Na década de 50 surgiu o termo vigilância epidemiológica que servia para designar o controle das doenças transmissíveis da época e também designar as etapas posteriores à etapa de ataque da campanha de erradicação da malária, designando uma de suas fases constitutivas. Tratava-se da vigilância das pessoas de forma individual e não de forma coletiva, com base em medidas de isolamento ou quarentena (BRASIL, 2005).

O propósito da vigilância epidemiológica é fornecer orientações técnicas aos profissionais de saúde que decidem sobre a execução de ações de controle de doenças e agravos. A vigilância epidemiológica também é um importante instrumento no planejamento, na organização e na operacionalização dos serviços de saúde e, também, na normatização de atividades técnicas relacionadas. Operacionalmente, a vigilância epidemiológica possui um ciclo de funções específicas e complementares que são desenvolvidas continuamente, permitindo que se saiba, a cada momento, o comportamento de uma doença ou agravo. As funções da vigilância epidemiológica são: coleta de dados, processamento dos dados coletados, análise e interpretação dos dados processados, recomendação das medidas de controle apropriadas, promoção das ações de controle indicadas, avaliação da eficácia e efetividade das medidas adotadas e divulgação de informações pertinentes. Os níveis municipais, estaduais e federal da vigilância epidemiológica desempenham todas as funções citadas da vigilância epidemiológica, porém com graus de especificidade variáveis (BRASIL, 2005).

2.1.4 PROTOCOLOS PARA DETECÇÃO DE VÍRUS RESPIRATÓRIOS

A Imunofluorescência Direta (IFD) é um dos ensaios utilizados para a pesquisa

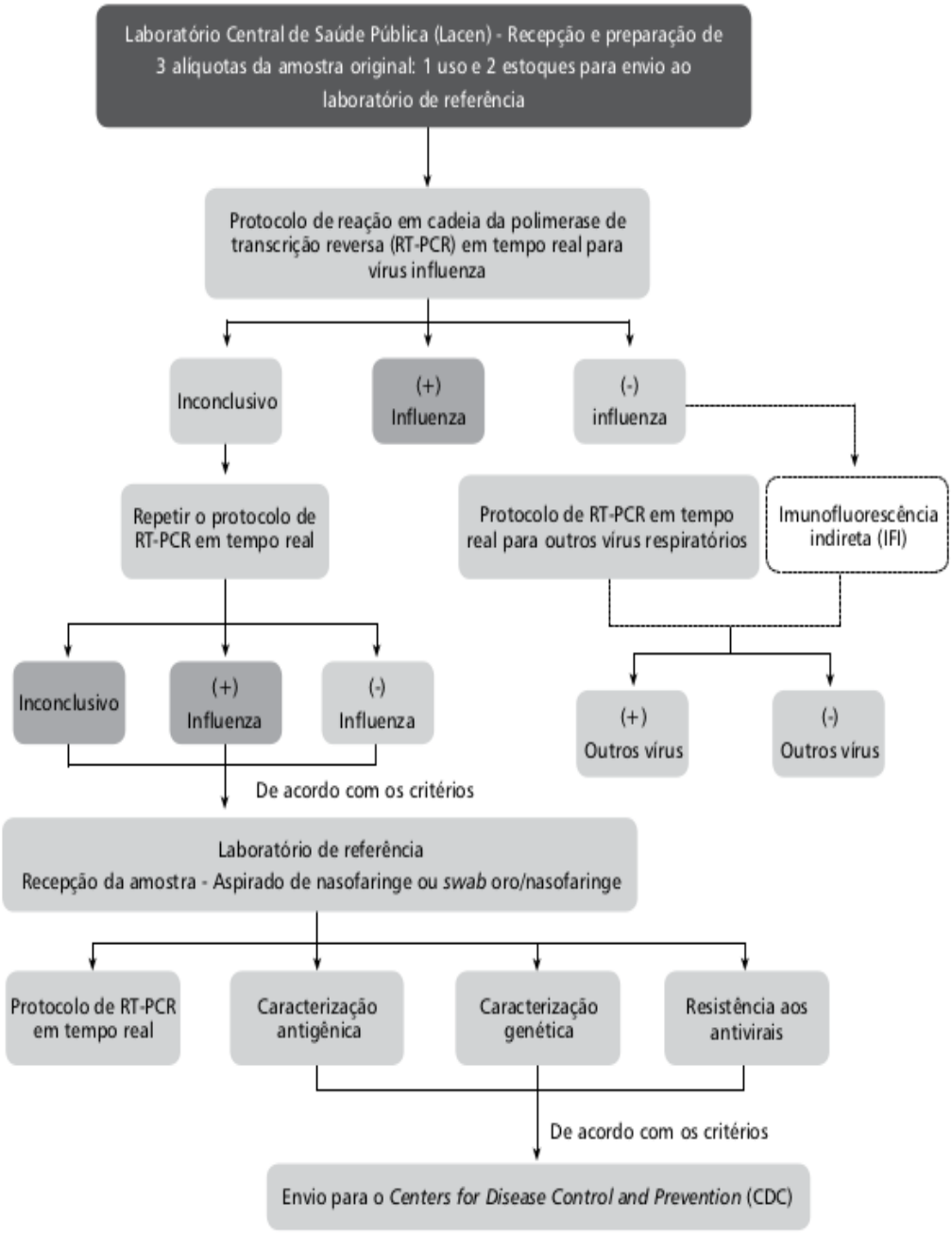
e a detecção de vírus respiratórios. Os vírus cobertos por esse protocolo são: vírus influenza A (IAV) e influenza B (IBV), parainfluenzavírus tipos 1, 2 e 3, vírus respiratório sincicial (VRS) e adenovírus humano. Este protocolo é utilizado na identificação rápida e específica de antígenos virais em células infectadas, através de reação imunológica com antígenos e anticorpos monoclonais específicos para cada vírus, produzidos em camundongos e marcados com fluoresceína. O imunocomplexo do resultado pode ser visualizado em um microscópio de imunofluorescência com luz ultravioleta (UV) ou halogênio, onde o cromógeno é ativado, permitindo a sua visualização. A grande vantagem desse teste é a rapidez na identificação de vírus respiratórios. A sensibilidade deste ensaio depende da carga viral e da qualidade das amostras coletadas em pacientes com SG ou SRAG (BRASIL, 2016).

Outro protocolo utilizado é a detecção molecular baseada em PCR, como a PCR em tempo real (qPCR) e a transcrição reversa seguida por qPCR (RT-qPCR) empregando oligonucleotídeos e sondas específicas para cada tipo viral. Considerada o padrão-ouro para o diagnóstico de infecção respiratória viral, a RT-qPCR é utilizada principalmente para detecção de IAV e IBV e, desde 2020, do novo coronavírus SARS-CoV-2 (BRASIL, 2016).

Mesmo que os sintomas sistêmicos sejam mais intensos na infecção por influenza do que nas demais infecções virais, os quadros clínicos apresentados são muito semelhantes e isto torna difícil um diagnóstico somente por exame clínico. Nesse caso, a orientação é seguir o algoritmo de diagnóstico laboratorial de vírus respiratórios (Figura 1).

Para a utilização eficiente dos dados da vigilância virológica, existe a necessidade de apresentação dos resultados em tempo hábil e, para isto, são definidos prazos máximos para cada metodologia entre o recebimento da amostra no laboratório e a liberação dos resultados. Para a identificação de agente etiológico, tipagem e subtipagem o prazo máximo é de 7 dias. Para o Teste de Sensibilidade aos Antivirais (TSA) este prazo é de 10 dias. Para a caracterização genética o prazo máximo é de 21 dias e, para a caracterização antigênica, 6 dias (BRASIL, 2016).

Figura 1 – Algoritmo de diagnóstico laboratorial.



Fonte: BRASIL, 2005.

2.2 INFLUENZA SAZONAL

A influenza sazonal é uma das principais causas de SRAG. Ela causa a infecção viral aguda do sistema respiratório, com transmissibilidade elevada e distribuição global. Durante a sua vida, um indivíduo pode contrair o vírus diversas vezes e, normalmente, a evolução do quadro é autolimitada, mas pode apresentar-se de forma grave (BRASIL, 2019).

Três importantes epidemias de influenza ocorreram no século XX: a Gripe

Espanhola (1918-20) – considerada a maior pandemia de influenza – a Gripe Asiática (1957-60), e a de Hong Kong (1968-72). Juntas, tiveram uma alta taxa de mortalidade que resultou em quase um milhão de óbitos no mundo todo. Em todas elas ocorreu a substituição de uma cepa de IAV circulante por uma nova cepa. Dentro desse quadro, a influenza é uma das grandes preocupações das autoridades sanitárias no mundo, tendo em vista a morbimortalidade resultante das variações antigênicas cíclicas sazonais. Devido à mutação antigênica do vírus influenza A e à troca genética com vírus não humanos, existe uma grande probabilidade da ocorrência de pandemias. Os casos mais graves são observados em idosos, crianças, pessoas com comprometimento imunológico, cardiopatias e pneumopatias (BRASIL, 2019).

2.3 PANDEMIA DE COVID-19

Com a emergência do novo coronavírus SARS-CoV-2 no final de 2019, desde 2020 os casos de SRAG, na sua maioria, têm sido associados à Covid-19 (doença causada pelo SARS-CoV-2). Em 11 de março de 2020, a Organização Mundial da Saúde (OMS) declarou a Covid-19 como sendo uma pandemia (WORLD HEALTH ORGANIZATION, 2020). No Brasil, o primeiro caso de Covid-19 foi reportado no dia 26 de fevereiro de 2020, no estado de São Paulo, e o segundo caso no dia 29 de fevereiro, também em São Paulo (ALONSO et al., 2020). É importante destacar as dimensões desta pandemia que ainda está em curso: mais de 420 milhões de casos e quase 6 milhões de óbitos em pouco mais de dois anos em todo o mundo (WORLD HEALTH ORGANIZATION, 2022).

A maioria das pessoas infectadas pelo SARS-CoV-2 não apresenta sintomas, ou os sintomas caracterizam uma doença respiratória que vai de leve à moderada e a recuperação, nesses casos, se dá sem a necessidade de tratamento especial (WORLD HEALTH ORGANIZATION, 2021). Em outros casos, a Covid-19 pode levar a um quadro grave, sendo este mais comum em idosos e pessoas com problemas médicos subjacentes, surgindo a necessidade de tratamento e até internação. Não obstante, mais recentemente tem se observado doença grave em indivíduos jovens. A infecção por Covid-19 acontece principalmente por meio de gotículas de saliva ou secreção nasal que são expelidas ao tossir ou espirrar (WORLD HEALTH ORGANIZATION, 2021).

Diante do quadro de pandemia da Covid-19, desde 2020 a estrutura de

notificação de síndrome gripal, que era feita somente por unidades sentinelas, foi alterada, passando as notificações a serem feitas também pelas unidades de atendimento. Com o objetivo de monitorar casos de Covid-19, o Ministério da Saúde incluiu a testagem do vírus SARS-CoV-2, causador da Covid-19, à vigilância de SRAG; além disso, a necessidade de um paciente apresentar dispneia para ser considerado com SRAG foi retirada do protocolo (BRASIL, 2021).

2.4 DADOS

Considerando a grande quantidade de campos de informação referentes à SG e à SRAG disponíveis no SIVEP-Gripe, é necessário, na geração de um modelo preditivo de Aprendizado de Máquina (AM), fazer a redução da dimensionalidade desses dados. Segundo Faceli et al. (2011), muitos problemas que podem ser tratados por AM possuem um número elevado de atributos e a maioria das técnicas de AM não conseguem lidar com números tão grandes de atributos; este problema é denominado maldição da dimensionalidade. Dessa forma, a redução do número de atributos melhora o desempenho dos modelos e reduz o custo de processamento. As técnicas para a redução de dimensionalidade estão divididas em dois grandes grupos: Agregação e Seleção de atributos. As principais técnicas de agregação combinam atributos através de funções lineares ou não lineares. Já para a seleção de atributos deve-se levar em conta os atributos considerados importantes – aqueles que melhoram o desempenho do modelo –, enquanto atributos irrelevantes devem ser eliminados (FACELI et al., 2011).

Uma das preocupações de uma pesquisa que envolve coleta e processamento de dados é justamente a qualidade dos dados. Um dos problemas que impactam a qualidade de dados é a ausência e a inconsistência de dados ou até mesmo o atraso na digitalização dos mesmos. Quanto ao atraso na digitalização de dados, técnicas de *nowcasting* (GÜNTHER et al., 2021; BASTOS et al., 2019), como as utilizadas no Infodengue (CODECO et al., 2018), conseguem contornar o problema de maneira muito satisfatória. Observando o SIVEP-Gripe os registros válidos de internação por SRAG para o período de 2017 a 2020, a mediana do tempo decorrente entre os primeiros sintomas e a internação do paciente é de 3 dias [0 – 16] (95% CI). Já a mediana do tempo entre os primeiros sintomas e a digitação (entrada dos dados no sistema) é de 9 dias [2 – 99] (95% CI) (BRASIL, 2020b). Nota-se, aqui, a necessidade de maior agilidade nos prazos de notificação, o que

pode melhorar a qualidade na gestão da saúde, principalmente quando se fala em modelos preditivos. Olhando os dados do CNES (BRASIL, 2020a), percebe-se que boa parte desses atrasos pode ser devido à falta de infraestrutura em algumas localidades, como a ausência ou insuficiência de rede para a transmissão de dados ou, até mesmo, a falta de equipamentos.

Este trabalho trata prioritariamente de dados advindos do SIVEP-Gripe, mas é possível fazer um paralelo com os dados do SINAN, tendo em vista que este possui muita similaridade com os dados do SIVEP-Gripe, tanto na estrutura de coleta quanto no armazenamento dos dados. Alguns trabalhos mostram a utilidade dos dados do SINAN para outros escopos que não SRAG. Um deles (DELZIOVO et al., 2018) fala da utilização dos dados do SINAN como fonte para a avaliação de ações de combate ao crime de violência sexual contra a mulher. Esse estudo faz uma análise utilizando dados do SINAN referentes a notificações dos casos suspeitos e confirmados de violência sexual contra mulheres e adolescentes residentes em Santa Catarina, no período de 2008 a 2013. Outro trabalho (SILVA et al., 2017) faz a avaliação da qualidade e aceitabilidade dos dados sobre tuberculose, que estão na base de dados do SINAN. O estudo de Silva et al. (2017) realiza um processo híbrido, com uma etapa qualitativa, utilizando-se de consultas a especialistas em vigilância da tuberculose, para seleção dos indicadores e definição dos parâmetros para a avaliação do sistema de vigilância. Ambos os autores avaliam que os dados disponíveis têm qualidade suficiente para a realização de pesquisas e gestão na saúde.

Vale destacar que dados de notificação em sistemas convencionais de dados não são a única fonte para a predição de surtos de SRAG. Trabalhos como o de Volkova et al. (2017) e também o de Santillana et al. (2015) utilizam dados de mídias sociais, como o Google Searches e Twitter, para complementar o conjunto de dados coletados no sistema de saúde. Segundo os autores de ambos os trabalhos, isto traz melhoria na acurácia das predições para a ocorrência de SRAG.

O clima desempenha um papel relevante na transmissão de muitas doenças infecciosas, algumas das quais estão entre as causas mais importantes de mortalidade e morbidade nos países em desenvolvimento. Frequentemente, essas doenças ocorrem como epidemias que podem ser desencadeadas pela variabilidade das condições climáticas que favorecem taxas de transmissão mais altas. No Brasil verifica-se uma alta correlação entre a ocorrência de SRAG e os períodos de chuva

e frio no sul, enquanto que, para o norte, a correlação é moderada para os períodos de chuva (FREITAS, 2013; PSCHEIDT et al., 2021). Quando se observam dados na forma de série temporal, o clima pode influenciar na sazonalidade de um modelo. Por exemplo, em geral, o número de casos de SG no sul do Brasil começa a aumentar no mês de abril, enquanto os casos de SRAG passam a aumentar em junho, apresentando um pico no final de julho e em agosto. Ou seja, os surtos de SG antecedem as epidemias de SRAG.

Quando se fala em melhoria na qualidade dos dados, é importante destacar a recente criação no Brasil da Rede Nacional de Dados em Saúde (RNDS) que é uma plataforma nacional de interoperabilidade (troca de dados) em saúde, instituída pela Portaria GM/MS nº 1.434, de 28 de maio de 2020. A RNDS é um programa do Governo Federal focado na transformação digital da saúde e busca promover a troca de informação na rede de atenção à saúde. A RNDS busca ser uma plataforma informacional de alta disponibilidade, segura e flexível, que possibilitará o uso ético dos dados de saúde. A plataforma irá favorecer a pesquisa e a inovação e, também, o surgimento de novos serviços que beneficiarão a população brasileira (BRASIL, 2021a). A disponibilização de dados dessa plataforma irá facilitar a elaboração de trabalhos como este, que necessitam de dados com a melhor qualidade e disponibilidade possíveis.

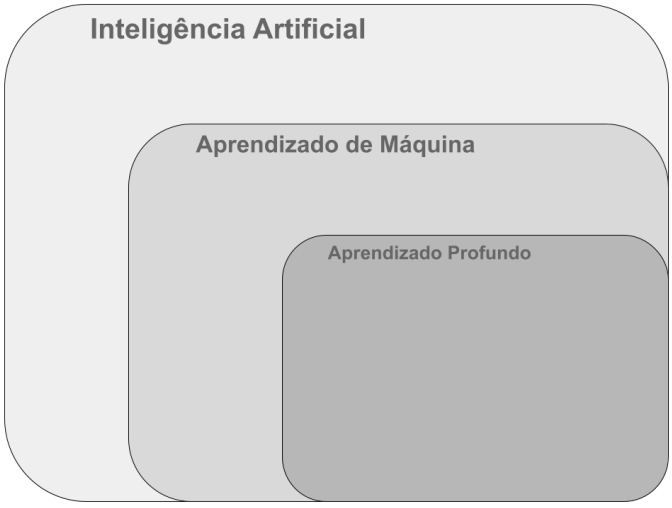
2.5 INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA

A Inteligência Artificial (IA) tem sido amplamente utilizada para a solução de problemas de classificação e predição. Há algum tempo, a área de IA era vista como uma área teórica, aplicada apenas em pequenos problemas curiosos e desafiadores, mas de pouco resultado prático. A partir da década de 70, houve uma maior disseminação do uso de técnicas de computação baseadas em IA para a solução de problemas reais. Muitas vezes esses problemas eram tratados computacionalmente por meio da aquisição do conhecimento de especialistas de um dado domínio, por exemplo do domínio da medicina, que era então transformado em código. Com o tempo, os problemas a serem resolvidos pelas máquinas tornaram-se mais complexos e a quantidade de dados envolvida muito maior. Para isto era necessário um processo que reduzisse a necessidade de especialistas e que, a partir de experiências passadas, pudesse criar uma hipótese ou função, capazes de tratar o problema que se queria resolver. A esse

processo de indução de uma hipótese ou aproximação de uma função, utilizando experiências passadas, dá-se o nome de Aprendizado de Máquina (AM, tradução de *Machine Learning*) (FACELI et al., 2011).

A decomposição de IA em subáreas traz uma visão mais clara dos conceitos, onde temos IA como termo genérico para qualquer programa de computador que tenha similaridade à inteligência humana e inclua AM e Aprendizado Profundo de Máquina (do inglês *Deep Learning*). A Figura 2 apresenta esta relação de forma gráfica, mas vale lembrar que existem ainda muitas outras subáreas dentro da IA. Aprendizado de Máquina inclui todas as abordagens que permitem que as máquinas aprendam a partir dos dados sem que haja uma programação explícita para isso. O aprendizado profundo incorpora modelos computacionais e algoritmos que buscam imitar uma rede neural do cérebro humano (JAKHAR e KAUR, 2019).

Figura 2 – Representação das subáreas de IA.



Fonte: Adaptado de Jakhar e Kaur (2019).

O AM é um subcampo da IA e também uma aplicação da mesma. AM é um método que automatiza a análise de dados através de algoritmos utilizados para identificar padrões de forma iterativa nos dados, aprendendo com eles. As aplicações de AM são, de forma mais ampla, classificadas em três categorias: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. A primeira lida com padrões já identificados de dados (dados rotulados); já a segunda tem como objetivo aprender com os padrões dos dados e não há uma identificação deles. O aprendizado por reforço pode ser considerado uma

extensão do aprendizado supervisionado, baseando-se em “recompensas” e “punições” conforme a iteração da aplicação em um ambiente dinâmico. Uma das aplicações do AM é na mineração de dados, que envolve a identificação de padrões em grandes conjuntos de dados (WAHL et al., 2018).

Existe uma extensa lista de algoritmos que são aplicados frequentemente em AM. Para citar alguns dos mais utilizados temos: *support vector regression* (SVR); *random forest* (RF); *multiple linear regression* (MLR); *multi-layer perceptron* (MLP); *k-nearest neighbor* (KNN) e *convolutional neural networks* (CNNs) (LIAKOS et al., 2018).

2.5.1 IA EXPLICÁVEL

É de conhecimento geral o exponencial crescimento da IA e os benefícios das suas incontáveis aplicações na ciência, educação, medicina, comércio, indústria, entretenimento, etc. Porém, uma grande parte dessas aplicações não têm como explicar suas decisões, tornando-se assim uma caixa preta. Para algumas aplicações, pode até não ser necessária essa explicação e alguns pesquisadores da área até argumentam que dar importância à explicação é algo errado, difícil de se obter e pode ser até desnecessário. Por outro lado, aplicações críticas no campo da defesa, medicina, finanças e direito, por exemplo, trazem a necessidade de explicação para que os seus usuários possam realmente confiar na solução de IA e possam gerenciá-la de forma transparente. À agregação de uma explicação a um processo de IA dá-se o nome de Inteligência Artificial Explicável (IAE) (GUNNING, 2019). A Figura 3 mostra uma compreensão intuitiva do assunto IAE, destacando conceitos relacionados ao tema.

Figura 3 – Nuvem de palavras de IAE.

Fonte: Adadi, Berrada (2018).

O objetivo de um sistema de IAE é de, através de explicações, tornar o seu comportamento mais inteligível. Os princípios gerais utilizados nos sistemas de IA mais compreensíveis e explicáveis são: capacidade de explicar seus recursos e entendimentos; explicar o que fez, o que está fazendo e o que fará a seguir; divulgar as informações salientes sobre as quais está atuando (BELLOTTI e EDWARDS, 2001). Cada explicação, que pode ser completa ou parcial, deve ser definida dentro de um contexto dependente da tarefa, habilidade e expectativas do usuário. A definição de interpretabilidade e explicabilidade dependem do domínio específico, não sendo possível defini-las fora desse domínio (GUNNING, 2019).

Quando avaliamos a completude de um determinado modelo de IA, temos a seguinte visão: modelos totalmente interpretáveis, que permitem explicações completas e transparentes; modelos parcialmente interpretáveis, que fornecem algumas partes importantes do seu processo de raciocínio. As restrições de interpretabilidade para um modelo interpretável são definidas conforme o domínio, enquanto um modelo caixa preta não obedece a essas restrições (GUNNING, 2019). Uma IAE permite melhorar modelos e obter novos *insights* sobre um determinado problema e pode ajudar na verificação de previsões, tornando um sistema de IA mais confiável (ADADI e BERRADA, 2018). A Figura 4 ilustra as principais motivações para o uso de uma IAE.

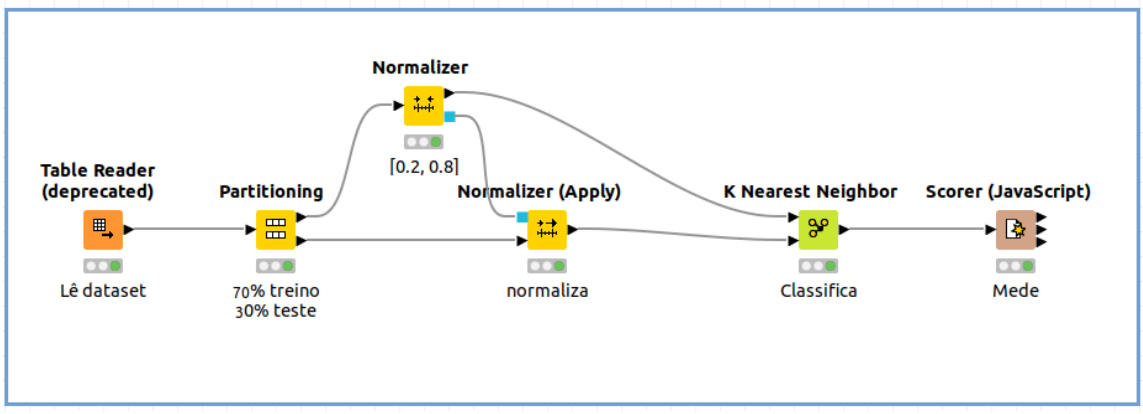
Figura 4 – Motivações para IAE.

Fonte: Adaptado de Adadi, Berrada (2018).

2.5.2 FERRAMENTA KNIME

Bastante utilizada na área de IA, de forma a dispensar a implementação de extensos e complicados códigos, o Knime é um ambiente modular, que permite a fácil montagem visual e execução interativa de uma pipeline de dados (Figura 5). A ferramenta é projetada como uma plataforma de ensino, pesquisa e colaboração, que permite a integração simples de novos algoritmos e ferramentas, bem como métodos de manipulação ou visualização de dados na forma de novos módulos ou nós. A ferramenta disponibiliza diversos algoritmos de AM como o Decision Tree, K KNN, SVM e LSTM (BERTHOLD, 2009; KNIME, 2019). A versatilidade da ferramenta possibilita desde a coleta de dados até o tratamento e aplicação de algoritmos que permitem a geração de modelos preditivos. Knime é uma plataforma de análise de dados de código aberto com um grande conjunto de blocos de construção e ferramentas de terceiros. Pode ser utilizado para carregar os dados em um relatório final ou para prever novos valores usando um modelo encontrado anteriormente. O Knime está disponível em quatro versões: Desktop/Professional, Team Space, Server e Cluster Execution (BAKOS, 2013). Devido à implementação em forma de *pipelines* e utilização de muitos algoritmos clássicos, o Knime permite o entendimento claro da geração dos modelos preditivos.

Figura 5 – Pipeline no Knime.



Fonte: O autor.

2.6 UTILIZAÇÃO DE IA NA PREDIÇÃO DE SURTOS E EPIDEMIAS

Pode-se enumerar diversas aplicações de IA em SRAG, como por exemplo: detecção e diagnóstico precoce da infecção; acompanhamento de tratamentos; rastreamento de contato dos indivíduos; projeção de casos e mortalidade;

desenvolvimento de medicamentos e vacinas; redução da carga de trabalho dos profissionais de saúde; e prevenção da doença (VAISHYA et al., 2020). Exemplos de aplicação de IA na predição de casos de Covid-19 também podem ser encontrados facilmente em alguns trabalhos publicados recentemente (VAISHYA et al., 2020; SHAHID, 2020).

Especificamente na redução da carga de trabalho dos profissionais da saúde, a IA ajuda no diagnóstico precoce e no fornecimento de tratamento em um estágio inicial, utilizando abordagens digitais e ciência de decisão. Nesse aspecto, também pode permitir um melhor treinamento para estudantes e médicos diante de uma nova doença como a Covid-19 (VAISHYA et al., 2020). Uma das aplicações possíveis de IA no desenvolvimento de novos medicamentos está relacionada à predição da afinidade entre ligantes e proteínas, reduzindo, assim, a necessidade de experimentos *in vitro* (DA SILVA et al., 2020). Com a utilização de IA, é possível rastrear e prever a natureza de um vírus a partir dos dados disponíveis em mídias sociais e plataformas de mídia. É possível predizer riscos de uma infecção e sua provável disseminação, bem como o número de casos positivos e morte em qualquer região. Com a IA é possível prever quais regiões, países e pessoas são mais vulneráveis e, com isto, tomar medidas de contingenciamento e mitigação (VAISHYA et al., 2020).

Chimmula e Zhang (2020) realizaram, no Canadá, o primeiro estudo para modelar a transmissão de doenças infecciosas e, assim, prever a gravidade da Covid-19, usando abordagens de aprendizado profundo. Foi utilizada uma rede LSTM para modelagem de séries temporais com o objetivo de desenvolver um modelo para predição do surto de Covid-19 no Canadá. Os autores utilizaram dados públicos fornecidos pela Universidade John Hopkins e pela autoridade de saúde canadense. Na época, eles estimaram o final do surto para meados de junho de 2020. Pela análise dos padrões dos dados disponíveis na época, foi possível observar que as medidas rápidas e eficazes utilizadas pelos gestores da saúde no Canadá para reduzir a exposição da população tiveram um impacto positivo em relação a outros países como os EUA e Itália. A taxa de transmissão verificada no Canadá tinha uma tendência linear, enquanto nos EUA havia um crescimento exponencial nas taxas de transmissão. Na época, com o ajuste nos dados e simulações nos modelos, os autores previram que o pico de Covid-19 no Canadá aconteceria em 2 semanas. Devido à similaridade da pandemia de Covid-19 com a

Gripe Espanhola, que durou dois anos, os autores ponderaram que a epidemia poderia terminar em três meses. Por outro lado, melhorias tecnológicas e cooperações internacionais poderiam melhorar o quadro da pandemia de Covid-19 no Canadá e reduzir sua duração. Nota-se que diversos fatores externos podem alterar os resultados de uma modelagem preditiva, fazendo com que uma predição possa se desviar da realidade – principalmente em se tratando de uma pandemia.

Shahid et al. (2020) compararam modelos de previsão de estatísticas, aprendizado de máquina e aprendizado profundo de máquina aplicados em dados de Covid-19 para dez países, incluindo Brasil, Alemanha, Itália, Espanha, Reino Unido, China, Índia, Israel, Rússia e EUA. Os conjuntos de dados da Covid-19 foram divididos em amostras de treinamento de 110 dias e de 48 dias para teste. Os conjuntos de dados possuem três características: casos confirmados, óbitos e casos recuperados. Os autores aplicaram regressores para a modelagem dos dados de Covid-19, incluindo ARIMA, SVR com *kernels* polinomiais e RBF, LSTM, GRU e Bi-LSTM, para previsões futuras sobre casos confirmados, mortes e recuperados. MAE, RMSE e r2_score foram utilizados como medida de desempenho para comparar os diversos modelos. Os modelos ARIMA e SVR não se mostraram eficientes na predição, apresentando maior erro de previsão e valores negativos de r2_score. Já o LSTM gerou os melhores valores de MAE, RMSE e r2_score, sendo que este último chegou a valores bem próximos de um. Os resultados das medidas de desempenho variaram de país para país, tendo como possíveis fatores para esta variação a população e o comportamento diferenciado da pandemia nos países avaliados. Os autores concluíram que o Bi-LSTM é um preditor adequado para os conjuntos de dados sequenciais utilizados e também em outros conjuntos de dados semelhantes, de forma a possibilitar um melhor planejamento e gestão de surtos, epidemias e pandemias pelos gestores da saúde.

2.7 SÉRIES TEMPORAIS e LSTM

Uma grande parte dos dados disponíveis para análise no mundo real são de natureza temporal, sendo que os dados coletados em intervalos de tempos regulares são conhecidos como dados de série temporal. A previsão de uma série temporal utiliza métodos de previsão de tendências/padrões futuros para um conjunto de dados históricos específicos, com características temporais. Os dados de uma série temporal são normalmente decompostos em tendência, sazonalidade e

erro. Uma tendência numa série temporal é observada quando um determinado padrão se repete em intervalos de tempo regulares, que pode estar relacionado a fatores externos como bloqueio do país devido a uma pandemia, distanciamento social obrigatório, quarentenas, clima, etc. (CHIMMULA e ZHANG, 2020). Uma alternativa para a previsão de séries temporais é o modelo *Autoregressive Integrated Moving Average* (ARIMA), desenvolvido por Box e Jenkins (2015), utilizado para caracterizar e prever observações de séries temporais disponíveis em períodos com espaços iguais, mas as redes tipo *Long Short Term Memories* (LSTM) recorrentes têm se mostrado muito eficientes no tratamento de séries temporais. Elas têm a capacidade de abordar as limitações das técnicas tradicionais de previsão de séries temporais, adaptando a falta de linearidades de alguns conjuntos de dados (CHIMMULA e ZHANG, 2020).

Os dados percebidos em uma série temporal podem ser univariados, quando uma sequência de medições de uma única variável é obtida, ou, ainda, dados multivariados, quando na sequência de medições são observadas mais de uma variável (PRIETO et al., 2015; FACELI et al., 2011). Nas últimas décadas, a classificação de séries temporais multivariadas tem provocado muito interesse em algumas áreas. As classificações multivariadas de séries temporais são frequentemente aplicadas na área da saúde (KANG e CHOI, 2014),

2.8 MÉTRICAS EM MODELOS PREDITIVOS

É importante que os resultados de predição obtidos com o Aprendizado de Máquina sejam medidos. Métricas como RMSE, RMSPE e MAPE são bastante utilizadas em trabalhos realizados na área de infecção respiratória como os de Volkova et al. (2017) e Santillana et al. (2015). O coeficiente de correlação de Pearson também é bastante utilizado nessa área. Nos trabalhos de Cheng et al. (2020) e Santillana et al. (2015) foram apresentados valores para Pearson superiores a 90%, o que pode ser considerado como um ótimo resultado para uma predição.

Coeficiente de determinação (R^2 ou R-quadrado) (1) — O coeficiente de determinação (WRIGHT, 1921) pode ser interpretado como a proporção da variância na variável dependente que é previsível a partir das variáveis independentes (CHICCO et al., 2020).

$$R^2 = 1 - \frac{\sum_{i-1}^{m}\left(X_i - Y_i\right)^2}{\sum_{i-1}^{m}\left(\underline{Y} - Y_i\right)^2} \quad (1)$$

(pior valor = −∞; melhor valor = +1)

Mean square error (MSE) (2) — O MSE pode ser usado quando houver *outliers* que necessitem ser detectados. O MSE é ótimo para atribuir pesos maiores a esses pontos, claramente, se o modelo eventualmente produzir uma única previsão muito ruim, a parte quadrada da função aumenta o erro (CHICCO et al., 2020).

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(X_i - Y_i\right)^2 \quad (2)$$

(melhor valor = 0; pior valor = +∞)

Mean absolute error (MAE) (3) — O MAE pode ser usado quando valores discrepantes representarem partes corrompidas dos dados. O MAE não penaliza muito os *outliers* de treinamento, fornecendo, assim, uma medida de desempenho genérica e limitada para o modelo. Por outro lado, se o conjunto de teste também tiver muitos *outliers*, o desempenho do modelo será medíocre (CHICCO et al., 2020).

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|X_i - Y_i\right| \quad (3)$$

(melhor valor = 0; pior valor = +∞)

Root mean square error (RMSE) (4) — As duas grandezas MSE e RMSE estão relacionadas monotonicamente (através da raiz quadrada). A ordenação de modelos de regressão utilizando MSE será idêntica a uma ordenação de modelos que utilizem o RMSE (CHICCO et al., 2020).

$$RMSE = \sqrt{\frac{1}{m}\sum_{i-1}^{m}\left(X_i - Y_i\right)^2} \quad (4)$$

(melhor valor = 0; pior valor = +∞)

Coeficiente de correlação de Pearson (5) — O coeficiente de correlação de Pearson serve para avaliar a força de uma relação linear entre duas variáveis. Pode assumir um intervalo de valores de -1 a +1. Um valor positivo denota correlação linear positiva e um valor negativo denota uma correlação linear negativa. Quanto

mais próximo o valor estiver de +1 ou -1, mais forte será o valor da correlação (FU et al., 2020).

$$\rho = \frac{\sum_{i=1}^{n}\left(X_i - \underline{X}\right)\left(Y_i - \underline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \underline{X}\right)^2} \cdot \sqrt{\sum_{i=1}^{n}\left(Y_i - \underline{Y}\right)^2}} \quad (5)$$

(melhor valor = +1 ou -1; pior valor = 0)

A multidisciplinaridade dos conceitos vistos acima permite o entendimento do problema a ser resolvido por esta pesquisa e apoiam o atingimento de uma solução. Em outras palavras: surtos e epidemias de SRAG constituem um problema de saúde pública; a possibilidade de se prever surtos de SRAG com base em dados de surtos e epidemias de períodos anteriores contribui para que medidas sejam tomadas para conter e controlar uma epidemia, reduzindo seus impactos na saúde pública. Vale destacar que vários outros algoritmos de aprendizado de máquina, com diferentes abordagens, foram experimentados, mas devido à falta de bons resultados na geração de modelos, não foram citados aqui.

REFERÊNCIAS

ADADI, Amina; BERRADA, Mohammed. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). **IEEE access**, v. 6, p. 52138-52160, 2018.

ALONSO, Wladimir Jimenez et al. Covid-19 em contexto: comparação com a mortalidade mensal por causas respiratórias nos estados brasileiros. **Interamerican Journal of Medicine and Health**, v. 3, p. 1-21, 2020.

BAKOS, Gábor. **KNIME essentials**. Packt Publishing Ltd, 2013.

BASTOS, Leonardo S. et al. A modelling approach for correcting reporting delays in disease surveillance data. **Statistics in medicine**, v. 38, n. 22, p. 4363-4377, 2019.

BELLOTTI, Victoria; EDWARDS, Keith. Intelligibility and accountability: human considerations in context-aware systems. **Human–Computer Interaction**, v. 16, n. 2-4, p. 193-212, 2001.

BERTHOLD, Michael R. et al. KNIME-the Konstanz information miner: version 2.0 and beyond. **AcM SIGKDD explorations Newsletter**, v. 11, n. 1, p. 26-31, 2009.21

BOX, George EP et al. **Time series analysis: forecasting and control**. John Wiley & Sons, 2015.

BRASIL. Ministério da Saúde. **Base de dados do CNES.** 2020a. http://cnes.datasus.gov.br. Acesso em: 09 de ago. 2020.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Coordenação-Geral de Desenvolvimento da Epidemiologia em Serviços. **Guia de Vigilância em Saúde**: volume único [recurso eletrônico] / Ministério da Saúde, Secretaria de Vigilância em Saúde, Coordenação-Geral de Desenvolvimento da Epidemiologia em Serviços. – 3ª. ed. – Brasília: Ministério da Saúde, 2019. 740 p.: il.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância das Doenças Transmissíveis. **Guia para a Rede Laboratorial de Vigilância de Influenza no Brasil** [recurso eletrônico] / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância das Doenças Transmissíveis. – Brasília: Ministério da Saúde, **2016.**

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Guia de vigilância epidemiológica** / Ministério da Saúde, Secretaria de Vigilância em Saúde. – 6. ed. – Brasília: Ministério da Saúde, 2005. 816 p. – (Série A. Normas e Manuais Técnicos)

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise em Saúde e Doenças não Transmissíveis. **Guia de vigilância epidemiológica Emergência de saúde pública de Importância nacional pela Doença pelo coronavírus 2019 – covid-19** [recurso eletrônico] / Ministério da Saúde, Secretaria de Vigilância em Saúde. – Brasília: Ministério da Saúde, 2021. 86 p.: il.

BRASIL. Ministério da Saúde. **Rede Nacional de Dados em Saúde - RNDS**. 2021a. Disponível em: https://www.gov.br/saude/pt-br/assuntos/rnds. Acesso em: 02 fev. 2022.

BRASIL. Ministério da Saúde. SINAN – **Sistema de Informação de Agravos de Notificação**. 2021b. Disponível em: http://portalsinan.saude.gov.br, Acesso em: 02 fev. 2021).

BRASIL. Secretaria de Vigilância em Saúde. **SIVEP-Gripe – Sistema de Informação da Vigilância Epidemiológica da Gripe**. 2020b. Disponível em: https://sivepgripe.saude.gov.br/sivepgripe. Acesso em: 22 jan 2022.

CHENG, Hao-Yuan et al. Applying machine learning models with an ensemble approach for accurate real-time influenza forecasting in Taiwan: development and validation study. **Journal of medical Internet research**, v. 22, n. 8, p. e15394, 2020.

CHICCO, Davide; WARRENS, Matthijs J.; JURMAN, Giuseppe. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and SINANRMSE in regression analysis evaluation. **PeerJ Computer Science**, v. 7, p. e623, 2021.

CHIMMULA, Vinay Kumar Reddy; ZHANG, Lei. Time series forecasting of Covid-19 transmission in Canada using LSTM networks. **Chaos, Solitons & Fractals**, v. 135, p. 109864, 2020.

CODECO, Claudia et al. Infodengue: A nowcasting system for the surveillance of arboviruses in Brazil. **Revue d'Épidémiologie et de Santé Publique**, v. 66, p. S386, 2018.

COSTA, L. M. C.; MERCHAN-HAMANN, E. Influenza pandemics and the structure of Brazilian health care system: brief history and characterization of the scenarios. **Rev Pan-Amazônica Saúde**, v. 7, p. 11-25, 2016.

DA SILVA, Amauri Duarte; BITENCOURT-FERREIRA, Gabriela; DE AZEVEDO JR, Walter Filgueira. Taba: A tool to analyze the binding affinity. **Journal of computational chemistry**, v. 41, n. 1, p. 69-73, 2020.

DELZIOVO, Carmem Regina et al. Quality of records on sexual violence against women in the Information System for Notifiable Diseases (Sinan) in Santa Catarina, Brazil, 2008-2013. **Epidemiologia e Serviços de Saúde**, v. 27, 2018.

FACELI, K et al. Inteligência Artificial–uma abordagem de Aprendizado de Máquina. **Rio de Janeiro: LTC**, 2011.

FREITAS, Felipe Teixeira de Mello. Sentinel surveillance of influenza and other respiratory viruses, Brazil, 2000-2010. **Brazilian Journal of Infectious Diseases**, v. 17, n. 1, p. 62-68, 2013.

FU, Ting et al. Correlation research of phase angle variation and coating performance by means of Pearson's correlation coefficient. **Progress in Organic Coatings**, v. 139, p. 105459, 2020.

GUNNING, David et al. XAI—Explainable artificial intelligence. **Science Robotics**, v. 4, n. 37, p. eaay7120, 2019.

GÜNTHER, Felix et al. Nowcasting the Covid-19 pandemic in Bavaria. **Biometrical Journal**, v. 63, n. 3, p. 490-502, 2021.

JAKHAR, Deepack; KAUR, Ishmeet. Artificial intelligence, machine learning and deep learning: definitions and differences. **Clinical and experimental dermatology**, v. 45, n. 1, p. 131-132, 2019.

KANG, Hyohyeong; CHOI, Seungjin. Bayesian common spatial patterns for multi-subject EEG classification. **Neural Networks**, v. 57, p. 39-50, 2014.

KNIME. Open for Innovation Knime. 2019. Disponível em: https://www.knime.com/. Acesso em: 210 jan. 2022.

LIAKOS, Konstantinos G. et al. Machine learning in agriculture: A review. **Sensors**, v. 18, n. 8, p. 2674, 2018.

MARTINS, L. G. et al. A Vigilância da Influenza no Rio Grande do Sul. **Bol Epidemiológico**, v. 13, p. 4-11, 2011.

PRIETO, Oscar J.; ALONSO-GONZÁLEZ, Carlos J.; RODRÍGUEZ, Juan J. Stacking for multivariate time series classification. **Pattern Analysis and Applications**, v. 18, n. 2, p. 297-312, 2015.

PSCHEIDT, Veridiane Maria et al. Epidemiology of human adenovirus associated with respiratory infection in southern Brazil. **Reviews in Medical Virology**, v. 31, n. 4, p. e2189, 2021.

SANTILLANA, Mauricio et al. Combining search, social media, and traditional data sources to improve influenza surveillance. **PLoS computational biology**, v. 11, n. 10, p. e1004513, 2015.

SHAHID, Farah; ZAMEER, Aneela; MUNEEB, Muhammad. Predictions for Covid-19 with deep learning models of LSTM, GRU and Bi-LSTM. **Chaos, Solitons & Fractals**, v. 140, p. 110212, 2020.

SILVA, Gabriela Drummond Marques da et al. Evaluation of data quality, timeliness and acceptability of the tuberculosis surveillance system in Brazil's micro-regions. **Ciência & Saúde Coletiva**, v. 22, p. 3307-3319, 2017.

VAISHYA, Raju et al. Artificial Intelligence (AI) applications for Covid-19 pandemic. **Diabetes & Metabolic Syndrome: Clinical Research & Reviews**, v. 14, n. 4, p. 337-339, 2020.

VOLKOVA, Svitlana et al. Forecasting influenza-like illness dynamics for military populations using neural networks and social media. **PloS one**, v. 12, n. 12, p. e0188941, 2017.

WAHL, Brian et al. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? **BMJ global health**, v. 3, n. 4, p. e000798, 2018.

WORLD HEALTH ORGANIZATION. Coronavirus. 2021. Disponível em: https://www.who.int/health topics/coronavirustab=tab1. Acesso em: 07 de fev 2021.

WORLD HEALTH ORGANIZATION et al. WHO Director-General's opening remarks

at the media briefing on Covid-19-11 March 2020. 2020.

WORLD HEALTH ORGANIZATION. WHO Coronavirus (Covid-19) Dashboard. 2022. Disponível em: https://covid19.who.int/. Acesso em: 11 de fev 2022.

WRIGHT, Sewall. Correlation and causation. 1921.

3 OBJETIVOS

3.1 OBJETIVO GERAL

Gerar modelos preditivos com bom desempenho, utilizando Aprendizado de Máquina, para a identificação anual da ocorrência de surtos de SRAG.

3.2 OBJETIVOS ESPECÍFICOS

1) Analisar uma das bases de dados disponíveis (SIVEP-Gripe) para poder submetê-la ao processo de tratamento de dados, buscando entender o significado desses dados e sua contribuição efetiva em um modelo preditivo.
2) Definir métricas para a avaliação da qualidade dos modelos preditivos gerados.
3) Através da experimentação de diferentes algoritmos de Aprendizado de Máquina e também da identificação dos atributos das bases de dados, obter melhores resultados para as métricas definidas, com base em trabalhos similares já publicados.
4) Verificar se os dados, coletados pelas Unidades Sentinelas e disponíveis no SIVEP-Gripe, têm boa qualidade para contribuir de maneira satisfatória na obtenção de modelos preditivos para o problema em questão. Os dois aspectos de qualidade estão relacionados à ausência de dados e à consistência dos valores apresentados.
5) Desenvolver uma ferramenta que possa auxiliar os gestores da saúde na predição da ocorrência de surtos de SRAG.

4 SITUAÇÃO DO MANUSCRITO

O Manuscrito intitulado “Epidemiological Surveillance: Machine Learning in Predicting Severe Acute Respiratory Infection Outbreaks” será submetido à revista científica Eurosurveillance, uma publicação eletrônica semanal que apresenta trabalhos sobre vigilância de doenças infecciosas, epidemiologia, prevenção e controle. Todo o conteúdo da revista é de acesso aberto e gratuito para leitores e autores. O fator de impacto mais recente, para o ano de 2020, é 6,3 (InCites Journal Citation Reports, Clarivate analytics, 2021). O artigo foi elaborado de acordo com as normas da revista (Anexo B), está em fase final de revisão e será submetido em breve.

5 ARTIGO CIENTÍFICO

Eurosurveillance

Title Page

Title: Epidemiological Surveillance: Machine Learning in Predicting Severe Acute Respiratory Infection Outbreaks

Authors: Amauri Duarte da Silva[1,6], BSc; Marcelo Ferreira da Costa Gomes[2,6],PhD; Tatiana Schäffer Gregianini[3,6], PhD; Leticia Garay Martins[4,6], PhD; Ana Beatriz Gorini da Veiga[5,6], PhD.

Affiliations:

1. Programa de Pós-Graduação em Tecnologias da Informação e Gestão em Saúde, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA). Rua Sarmento Leite, 245 – Porto Alegre, RS – 90050-170, Brazil. ORCID: 0000-0001-6395-458X. E-mail: amauri@ufcspa.edu.br

2. Fiocruz, Programa de Computação Científica, Grupo de Métodos Analíticos em Vigilância Epidemiológica (MAVE). Av Brasil, 4365. Rio de Janeiro, RJ – 21040-900, Brazil. ORCID 0000-0003-4693-5402. E-mail: marfcg@gmail.com

3. Laboratório Central de Saúde Pública, Centro Estadual de Vigilância em Saúde da Secretaria de Saúde do Estado do Rio Grande do Sul – LACEN/CEVS/SES-RS. Av. Ipiranga, 5400. Porto Alegre, RS, Brasil, CEP 90450-190. ORCID: 0000-0002-9912-9060 E-mail: tatiana-gregianini@saude.rs.gov.br ; tatianagregianini@gmail.com

4. Centro Estadual de Vigilância em Saúde da Secretaria de Saúde do Estado do Rio Grande do Sul – CEVS/SES-RS. Av. Ipiranga, 5400. Porto Alegre, RS, Brasil, CEP 90450-190. ORCID: 0000-0002-5614-6952. E-mail: leticia-martins@saude.rs.gov.br

5. Programa de Pós-Graduação em Tecnologias da Informação e Gestão em Saúde, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA). Rua Sarmento Leite, 245. Porto Alegre, RS – 90050-170, Brazil. ORCID: 0000-0003-1462-5506. E-mail: anabgv76@gmail.com; anabgv@ufcspa.edu.br

6. These authors contributed equally to this article and share first authorship

Corresponding author: Ana Beatriz Gorini da Veiga

Corresponding author email: anabgv76@gmail.com; anabgv@ufcspa.edu.br

Abstract

Abstract text: Background: Outbreaks of Severe Acute Respiratory Infection (SARI) occur annually in several geographic regions, with seasonal peaks varying from region to region.

Case notification is important to prepare the healthcare network for patient attendance and hospitalization. Therefore, health managers need adequate resource planning tools for each SARI season. Goal: Generate models for the prediction of SARI outbreaks based on machine learning. Methods: In this study, data from the reporting of SARI hospitalization cases in Brazil from 2013 to 2020 were used, excluding SARI cases caused by COVID-19. These data were prepared in order to feed a neural network configured to generate predictive models for time series. The neural network was implemented with a pipeline tool. Models were generated for the five Brazilian regions and the models were validated for different years of SARI outbreaks. Results: With the use of neural networks, it was possible to generate predictive models for SARI peaks, including the volume of cases per season and the beginning of the pre-epidemic period with good performance (R^2 = 0·97 [0·95 – 0·98], considering the observed season of 2019 in the Southeast region of Brazil).

Keywords: SARI, Machine Learning, Predictive Models, Epidemiological Surveillance, Neural Networks.

Conflict of interest: ABGV holds a Research Fellowship (PQ2) from CNPq.

Funding statement: This work was supported by Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul, Ministry of Health of Brazil, and Conselho Nacional de Desenvolvimento Científico e Tecnológico (FAPERGS/MS/CNPq 08/2020 - PPSUS – Grant process 21/2551-0000059-7). ABGV holds a Research Fellowship (PQ2) from CNPq (Grant process 306369/2019-2)

Manuscript text file

Epidemiological Surveillance: Machine Learning in Predicting Severe Acute Respiratory Infection Outbreaks

Background: Outbreaks of Severe Acute Respiratory Infection (SARI) occur annually in several geographic regions, with seasonal peaks varying from region to region. Case notification is important to prepare the healthcare network for patient attendance and hospitalization. Therefore, health managers need adequate resource planning tools for each SARI season. **Goal:** Generate models for the prediction of SARI outbreaks based on machine learning. **Methods:** In this study, data from the reporting of SARI hospitalization cases in Brazil from 2013 to 2020 were used, excluding SARI cases caused by COVID-19. These data were prepared in order to feed a neural network configured to generate predictive models for time series. The neural network was implemented with a pipeline tool. Models were generated for the five Brazilian regions and the models were validated for different years of SARI outbreaks. **Results:** With the use of neural networks, it was possible to generate predictive models for SARI peaks, including the volume of cases per season and the beginning of the pre-epidemic period with good performance (R^2 = 0·97 [0·95 – 0·98], considering the observed season of 2019 in the Southeast region of Brazil). **Conclusion:** The identification of the period of occurrence of a SARI outbreak is possible using predictive models generated with the use of neural networks and algorithms that apply time series.

Keywords: SARI, Machine Learning, Predictive Models, Epidemiological Surveillance, Neural Networks.

Introduction

Viral respiratory infections are widely distributed and easily spread in the community, affecting millions of individuals annually in all geographic regions; thus, they constitute a serious public health problem worldwide, with high morbidity and mortality, especially in children, the elderly, and in immunocompromised patients [1]. Most acute respiratory infections are caused by viruses, and symptoms range from mild (runny nose and cough), influenza-like illnesses (ILI), to even Severe Acute Respiratory Infection (SARI). ILI is characterized by fever accompanied by cough or sore throat, usually within the first seven days of infection, while SARI includes, in addition to these symptoms, O_2 saturation below 95%, dyspnea and increased respiratory rate, requiring hospitalization of the patient [2, 3].

The seasonality of SARI outbreaks vary between different geographic regions [4]. In Brazil, a continental-size country, a clear seasonal southward wave is seen, starting in late January in the North Region, and reaching the South Region in the middle of the year [5, 6]. Of note, the intensity and duration of outbreaks also vary by region [7].

Influenza A (IAV) and influenza B (IBV) viruses cause human influenza, considered one of the most important infectious diseases of humanity. Seasonal flu epidemics occur every year and, eventually, new viral subtypes with pandemic potential emerge. In 1918–1919 the humanity was ravaged by the Spanish Flu, caused by IAV H1N1; and in 2009, the Influenza A pandemic occurred – the first influenza pandemic of the 21st century, caused by a new H1N1 IAV subtype (H1N1pdm09) [1]. In Brazil, of the 88,464 cases of SARI hospitalization reported in 2009, 50,482 were confirmed as IAV H1N1pdm09, with 2,060 deaths [8].

Besides influenza virus, other respiratory viruses are also associated with SARI epidemics. In this sense, the new coronavirus SARS-CoV-2 emerged in humans in the end of 2019, causing the COVID-19 pandemic [9]. Within 2 years, more than 420 million COVID-19 cases have been reported, approximately 6 million people have died, and we are still facing challenges in trying to contain and control the pandemic.

Prevention and control of SARI outbreaks rely on constant epidemiological surveillance. In addition, information of previous seasonal epidemics of respiratory viral infection, with accumulated data of cases, may be used to predict following epidemic seasons. In this sense, the development of prediction models that take into account variables specific for each geographic region is important to prepare the healthcare network and to help health authorities in decision-making for policies and planning. In this study, we develop and analyze predictive models for SARI outbreaks using data from Brazil, and results show that using data from previous epidemics it is possible to predict SARI outbreaks with good precision.

Methods

Data

In Brazil, SARI became a disease of universal notification in 2009 in response to the influenza A H1N1 pandemic [10]. Until 2018 the notification was carried out in the SINAN system [11], and then migrated to the SIVEP-Gripe system [12], which collects data related to SARI cases, including age, sex, symptoms, comorbidities and other risk factors, date of first symptoms, vaccine information, PCR results for some respiratory viruses, hospitalization date, date of discharge or death, health unit and geographic location of the patient.

The dataset used in this study for the generation and validation of predictive models has 689,797 records for the period between January 2009 and March 2021. The dataset used had 252 variables and, after several simulations, the best results for the predictive model included the variables: number of daily notifications (used as a dependent variable), date of first symptoms, sex (male or female), age group (young, adult and elderly), and positive diagnosis for influenza A or influenza B virus. Data that identify ILI and SARI – such as fever, sore throat, dyspnea, and cough – were not used, as this brings redundancy with the number of notifications, not giving positive effects on the generation of models.

The dataset contains data from the five Brazilian regions (North, Northeast, Midwest, Southeast and South). The analyses were carried out preferably with data from the South region because data collection in this region is more consistent than in other regions and, also, the South region has a better-defined period of SARI occurrence (winter months) in

relation to the other regions. In terms of the volume of annual notifications of SARI hospitalization, there is a significant variation over the years. Taking, for example, the South region, we had the years 2014 and 2015 with the lowest volumes of notifications (4,754 and 4,572, respectively) and, at the other extreme, the years 2016 and 2020 had large volumes of notifications (11,266 and 21,860, respectively).

In Brazil, there is a high correlation between the occurrence of SARI and the rainy and cold periods in the south, while, for the north, the correlation is moderate for the rainy periods [6, 7]. Considering that the SIVEP-Gripe data does not include the temperature variable, and with the objective of evaluating the contribution of this variable in the generation of predictive models for the prediction of SARI, the daily averages of minimum and maximum temperatures and thermal amplitude per region were added to the dataset, as well as the total average of temperatures per region.

The temperature information was calculated based on data obtained from the INMET website [13]. The data retrieved bring values of maximum and minimum temperatures, at different times of the day in different meteorological stations spread in each state of the five Brazilian regions. When using the temperature variables to generate the predictive models, no significant improvements in accuracy were observed, so this variable was not used in the models presented as results in this study. Metrics obtained using temperature data can be seen in the Supplementary Material.

Data processing

Data were treated taking into account the date of the first reported SARI symptoms. Due to the difference in protocols in the first years (2009 to 2012) and also the insufficiency of data for 2021 (only three months), the predictive model was generated using data from the years 2013 to 2020.

After the previous step of filtering and analyzing the data, the dataset containing the SARI notification data was grouped by date of the first symptoms and, thus, the grouping result showed a further reduction in the amount of data available for the generation of models. For most variables, the grouping was done by adding the number of cases of hospitalization for SARI on each date of the first symptom that pointed to a certain category (age group, sex, etc.). For sex, the grouping was done in a similar way, adding up the total number of cases that pointed to the male or female categories on each date of first symptoms. Grouping by age group was done for youth (0–19 years), adults (20–59 years) and elderly (>60 years).

In order to standardize the notifications in the different periods, the records referring to COVID-19 cases were removed. At the end of this first process of filtering and excluding records with problems, a total of 354,249 records were available for generating predictive models. Due to the specificities of each region of Brazil, the data were separated by region and, during the evaluation process, different predictive models were generated for each region. At the end of data processing, temperature data was added. These data, grouped and divided by region, were used to generate predictive models. To reduce notification bias, the 7-day moving average was applied to the amount of daily notifications.

To suit the LSTM (Long Short Term Memories) algorithm, the date of the first symptoms was decomposed into year, month and day. The values of all categories were normalized between 0 and 1. The normalization model was stored to be used in the normalization of the

data to be used in later predictions. The categories, except for the number of notifications, were arranged in a vector way.

In order to simplify the generation of predictive models, the weeks of the year are shown in the analysis instead of the epidemiological weeks.

Generation of predictive models

The generation of predictive models was simulated using machine learning algorithms such as Linear Regression, Random Forest, Naive Bayes, Tree Assemble, RPROP MLP, but the verified accuracies were very low. In a more in-depth analysis of the problem to be solved, a time series approach was shown to be more appropriate. Initially, the ARIMA algorithm was used, but the results were not significant either, so a recurrent neural network (RNN), more specifically the Long Short Term Memory (LSTM) was used, which makes it possible to work with multivariate time series [14] and which has an architecture that works well with seasonality [15]. The flow for predictive model generation with LSTM, implemented in Knime [16], was adapted from Melcher [17]. Data were separated into two sets: 80% for model training and 20% for model testing. The models were trained with data available from the years prior to the year in which the prediction is to be made, with 2013 as the beginning of the period.

Looking for models with better performance, combinations of different categories (independent variables) were used and, therefore, different metrics were obtained for the tests, with the categories sex (male and female), age group (young, adult and elderly), positive diagnosis for influenza A and influenza B were the ones that gave the best performance in the prediction made in the test stage of the model, with $R^2 = 0{\cdot}99$, MAE = 1·28, MSE = 2·89, RMSE = 1·70, MSD = 0·39 and MAPE = 0·06. Metrics for the other combinations of categories can be seen in the Supplementary Material. It is important to note that this is not a rule and that the combination of different variables and number of variables may or may not give a better result. For this reason, several simulations were necessary to find the best combination of independent variables that would give a better performance in the validation of predictions.

Tools

The Knime [16], a tool that uses pipelines, was used for data preparation, generation of Machine Learning models and validation of these models, generating the results to be analyzed. In the analysis of the results, the programming language R was used [18] with RStudio interface [19].

Validation

For validation, models were generated to predict the reporting curves of SARI cases from years for which data is already available. The approximation of the curves was verified, considering the week of the beginning of the pre-epidemic period, volume of cases of notification of the season, and peak week. Accuracy metrics were used in the comparison (Table 1). To generate these metrics, 100 simulations were created for each scenario tested. We report metrics in terms of median and 95% CI. Due to the COVID-19 pandemic, which started in 2020 and remained active in 2022, most of the validations were made for the year

2019, which, in addition to not showing the pandemic bias, also has the largest amount of data for training models, as it accumulates data from previous years, starting in 2013.

Table 1. The table shows the values of prediction validation metrics, using the median and 95% CI, considering the weekly total of notification cases for the South and Southeast regions for the years 2018 and 2019. It also shows the annual total of reported cases (observed and predicted) in the two regions.

Target Year - Region	Pearson	R^2	RMSE	MAPE	Observed Cases	Predicted Cases
2019 - South	0·90 [0·90 - 0·91]	0·82 [0·81 - 0·83]	52 [51 - 53]	11 [10 - 14]	9,936	9,405 [9,105 - 9,738]
2018 - South	0·84 [0·81 - ·86]	0·71 [0·66 - 0·73]	87 [81 - 94]	19 [17 - 24]	8,918	6906 [6,348 -7,550]
2019 - Southeast	0·98 [0·97 - 0·99]	0·97 [0·95 - 0·98]	35 [30 - 46]	10 [7 - 13]	14,332	14399 [13,763 - 15,022]
2018 - Southeast	0·80 [0·75 - 0·83]	0·63 [0·56 - 0·69]	147 [141 - 155]	28 [24 - 32]	13,390	10581 [9,616 - 11,503]

Results

The prediction models generated with neural networks, using time series algorithms, give good results in the prediction of SARI outbreak seasons. As can be seen in Table 1, the generated models had a good performance in the prediction. For both the South and Southeast regions, the performance measures were better for the year 2019 due to the greater amount of data for training the models. This relationship of the volume of data with the performance of the models was observed in the simulations for previous years, where the volume of data reduces according to the period.

Figures 1 and 2 show the prediction curves generated by the models. As already mentioned, the curves generated for the year 2019 show greater coincidence with the observed curve. The previous year's curve (green) is being shown in order to assess whether the simple application of data observed in the previous year would be a way of predicting the following year. The variability of the generated models, obtained from random seeds, is small, with lower and upper limits very close to the median. The metrics collected for this graph showed a good performance prediction, presenting a Pearson coefficient of 0·90 [0·90 – 0·91], R^2 = 0·82 [0·81 – 0·83], RMSE = 52 [51 – 53] and MAPE = 11 [10 – 14].

Figure 1: Prediction of the notification of hospitalization of SARI cases for the South region, in the years 2018 and 2019. The graphs show the curves of cases observed in the previous year (green), cases observed in the year (blue), and the median of the prediction of cases (red) with their variability. It also identifies the volumes of cases that characterize the pre-epidemic period, high intensity, and very high intensity.

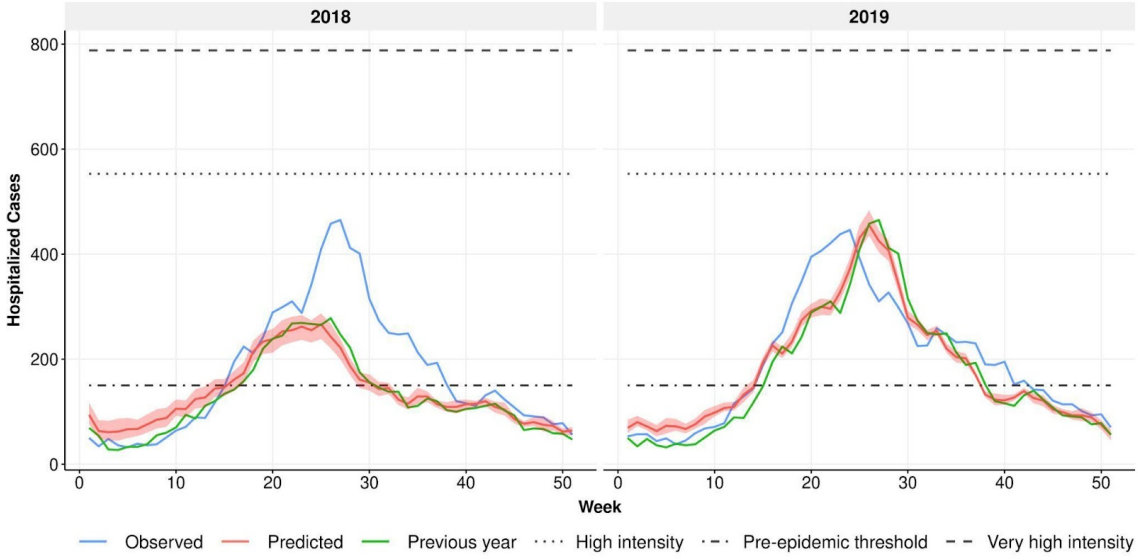
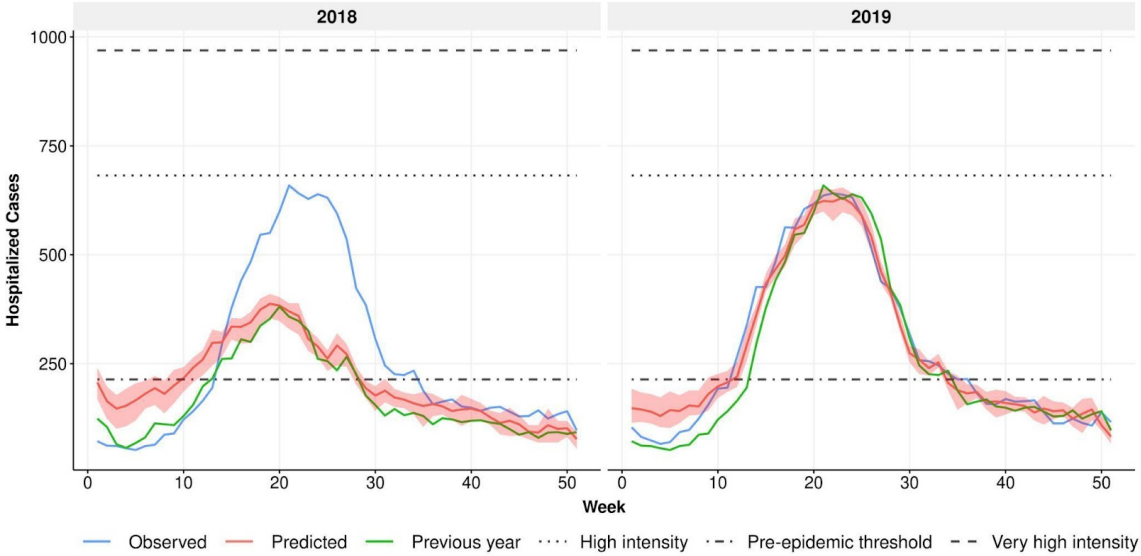


Figure 2: Prediction of the notification of hospitalization of SARI cases for the Southeast region, in the years 2018 and 2019. The graphs show the curves of cases observed in the previous year (green), cases observed in the year (blue), and the median of the prediction of cases (red) with their variability. It also identifies the volumes of cases that characterize the pre-epidemic period, high intensity and very high intensity.



There is an approximation of the predicted curve of a year (red lines in Figures 1 and 2) and the curve of cases in the previous year (green lines in Figures 1 and 2). Notwithstanding, this approximation is not always observed. In fact, as shown in Figure 3, the generated model shows a clear difference between the predicted curve for the year 2020 and the curve of the previous year (2019). Notably, the volume of cases predicted for 2020 (18,078) was larger than the number of cases notified in the previous year (9,901 cases, 55% difference), but approximate to the volume of observed cases (21,860; 17% difference). This is an interesting result considering that 2020 was an atypical year, during which the

COVID-19 pandemic brought problems, such as inconsistent notifications, especially in the initial peak. Hence, it is clear that the simple use of the previous year's notification curve is not a good way to predict the notification of SARI hospitalization cases for the following year.

Another comparative approach was to use the average notification of hospitalizations due to SARI in the years 2013 to 2018 for the South region. An average volume of 7,357 notifications was obtained against 9,936 observed and 9,327 notification cases predicted by the model (Figure 4). It is important to note that the peak of notifications of the mean curve coincides with the peak of the predicted curve (week 26) and both are 2 weeks apart from the observed curve.

Figure 3: Prediction of the notification of hospitalization of SARI cases for the South region, in the year 2020. The graphs show the curve of SARI hospitalization cases reported in the previous year (green), curve of data observed in the year (blue), and prediction curve of number of cases (Red).

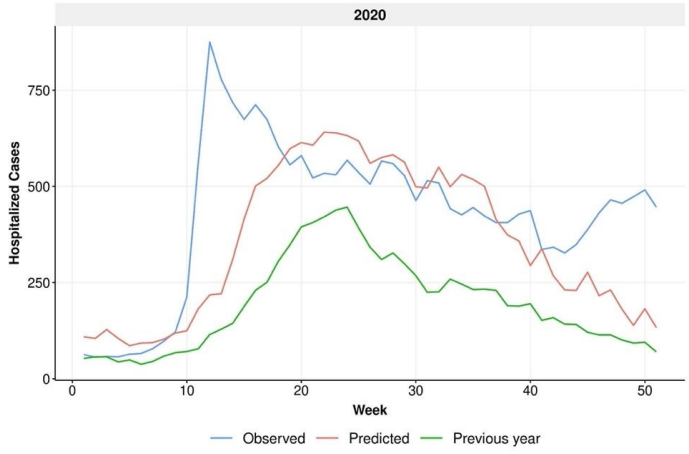
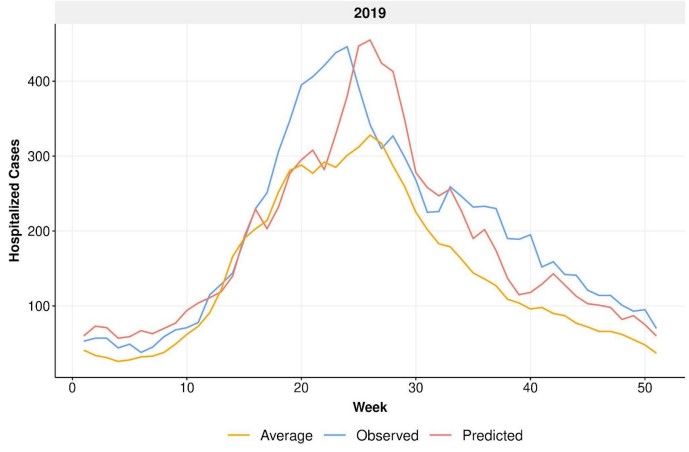


Figura 4: Prediction of the notification of hospitalization of SARI cases for the South region, in the year 2019. The graph shows the curve of the average number of hospitalized cases notified in the period from 2013 to 2018 (orange) and the curve of the data observed in the year 2019 (blue).



Discussion

In this work, we sought to find predictive models for SARI outbreaks using machine learning that presented good performance (shown through metrics). These models can help health managers in the adoption of adequate public health policies, in addition to allocating human and material resources in the necessary quantities and in a timely manner.

As our results show, a prediction with good performance is viable, however exceptional events, such as the emergence of new viruses and pandemics, can be unfavorable to the prediction of the models. Nevertheless, these models provide important information for healthcare management.

Our study also shows that the quality and quantity of data influence the performance of predictive models. As for quality, it is noted that in the South and Southeast regions, which have a more uniform and concise notification, the performance of predictive models is better than for other regions of the country. Table 2 shows the volume of registration of SARI notifications by region and is related to the performance in the prediction of the models (R^2).

Table 2. Amount of SARI data by Brazilian region. The evaluation metric of the predictive models is compared with the number of SARI notification records by region of Brazil. For R^2, validation is being considered for the volume of data from 2013 to 2020, which was applied in the test set.

Region	Total records (1)	Total records treated	Total records year 2020 (2)	Percentage of records(2)/(1)	R^2 test for 2020
North	22,994	2,361	13,129	57·10%	0·95
North East	60,908	2,837	36,335	59·66%	0·84
Midwest	29,712	2,653	12,660	42·61%	0·95
Southeast	164,354	2,897	83,763	50·96%	0·98
South	76,281	2,888	22,140	29·02%	0·96
Sum	**354,249**	**13,636**	**168,027**		

It is important to note that, except for the South region, the other regions had a concentration of records in the year 2020 because of the COVID-19 pandemic. This factor influenced the performance of the models, in addition to the total number of records for each region. One of the explanations for this is the fact that viral respiratory infections are more common in regions with colder winters, such as the southern region of Brazil [5, 6], but SARS-CoV-2 has reached all geographic regions significantly, therefore the increase in case notification in relation to previous years was relatively greater in the other regions. In addition, it is possible that the pandemic alert situation led health institutions to adhere more strictly to SARI notification in the other regions.

Obtaining a metric with great value (R^2 = 0·99) in the model test does not provide a prediction with the same performance when using real data, but it is an indicator that the prediction will be very close to the observed situation. In this sense, as observed in Figure 1, for the year 2019 the model was able to predict the beginning of the pre-epidemic threshold with good accuracy in relation to what happened. The volume of notifications in 2019 was 9,936 cases and the expected volume was 9,327 cases (-6%). If we had used the notification curve for the year 2018 to predict the year 2019, we would have 8,918 cases (-10%) and the pre-epidemic threshold would be off by one week.

In the decomposition of a historical series of SARI notifications for the South region in the year 2019, we can observe that the seasonal factor has an impact on the notification curve and also the presence of noise shows significant variations, bringing difficulties to the approximation of a curve of prediction. Even so, a predictive model based on neural networks and time series can absorb these difficulties (see figure in the Supplementary Material).

The main limitation of this study is related to the quality and quantity of data. With the need to standardize data collection, it was only possible to use data between the years 2013 and 2020, given that previous years used other data collection and recording standards. Some regions of Brazil have fewer cases registered annually, therefore the amount of available records was not favorable to the generation of better models, especially for regions that had fewer cases of notification. Another factor that caused some distortion in the models were the periods where there were sudden changes in the volume of notifications, such as the year 2016 – during which a high number of influenza cases were reported, due to a new strain of IAV H1N1pdm09 [20] – and the year 2020, with the emergence of the COVID-19 pandemic. Despite the difficulties reported here, the notification system in Brazil can be considered very good.

Conclusion

The identification of the period of occurrence of a SARI outbreak is possible using predictive models generated with the use of neural networks and algorithms that apply time series. The prediction offered by these models can help health managers in defining strategies to mitigate and combat these outbreaks, contributing to the safety and quality of the population's health and avoiding unnecessary expenditure on human and financial resources. Information such as the expected volume of SARI notifications, the beginning of the pre-epidemic threshold and the peak week of notifications can be of great value in defining public health policies.

Abnormal events, such as the emergence of pandemics, and the reduced volume of available data, are factors that make it difficult to generate predictive models. With the accumulation of new data from years to come, the tendency is an increase in the quality of these models, allowing the generation of more assertive models.

Even though it is not possible to generate prediction models that perfectly fit the observed curves, the performance metrics obtained in this study are very favorable and show the main points for decision making, such as peak week of an outbreak, case volume and. pre-epidemic threshold.

References

1. World Health Organization (WHO). Global Influenza Strategy 2019–2030. 2019. Https://apps.who.int/iris/bitstream/handle/10665/311184/9789241515320-eng.pdf?ua=1. Acesso em: 28 de set. 2019. Geneva: World Health Organization.

2. Martins L, Menegolla I, Ranieri T, Bercini M, Ueda E, Porto M. A vigilância da

influenza no rio grande do sul. Bol Epidemiológico, v. 13, p. 4–11, 2011.

3. Ministério da Saúde do Brasil. Vigilância sentinela de síndrome gripal (SG) no Brasil. 2017. Secretaria de Vigilância em Saúde.

4. Li Y, Reeves RM, Wang X, Bassat Q, Brooks WA, Cohen C, Zar HJ. Global patterns in monthly activity of influenza virus, respiratory syncytial virus, parainfluenza virus, and metapneumovirus: a systematic analysis. The Lancet Global Health, 7(8), e1031-e1045, 2019.

5. Alonso WJ, Viboud C, Simonsen L, Hirano EW, Daufenbach LZ, Miller MA. Seasonality of influenza in Brazil: a traveling wave from the Amazon to the subtropics. American journal of epidemiology, 165(12), 1434-1442, 2007.

6. Pscheidt VM, Gregianini TS, Martins LG, VEIGA ABG. Epidemiology of human adenovirus associated with respiratory infection in southern brazil. Reviews in Medical Virology, Wiley Online Library, p. e2189, 2021.

7. Freitas FTM. Sentinel surveillance of influenza and other respiratory viruses, brazil, 2000–2010. The Brazilian Journal of Infectious Diseases, Elsevier, v. 17, n. 1, p. 62–68, 2013.

8. Felinto GM, Escosteguy CC, Medronho RA. Fatores associados ao óbito dos casos graves de influenza a (h1n1) pdm09. Cadernos Saúde Coletiva, SciELO Brasil, v. 27, n. 1, p. 11–19, 2019.

9. Wu D, Wu T, Liu Q, Yang Z. The SARS-CoV-2 outbreak: what we know. International Journal of Infectious Diseases, v. 94, p. 44-48, 2020.

10. Costa LMC,; MERCHAN-HAMANN E. Influenza pandemics and the structure of Brazilian health care system: brief history and characterization of the scenarios. Rev Pan-Amazônica Saúde, v. 7, p. 11-25, 2016.

11. Sinan – Sistema de Informação de Agravos de Notificação. 2020. Brasília, Brasil: Ministério da Saúde. [Accessed: 22 Jan 2022]. Available from: http://portalsinan.saude.gov.br

12. SIVEP-Gripe – Sistema de Informação da Vigilância Epidemiológica da Gripe. 2020. Brasília, Brasil: Secretaria de Vigilância em Saúde. [Accessed: 22 Jan 2022]. Available from: https://sivepgripe.saude.gov.br/sivepgripe

13. Instituto Nacional de Meteorologia. Dados Históricos Anuais. 2021. [Accessed: 07 Mar. 2021]. Available from: https://portal.inmet.gov.br/dadoshistoricos. Ministério da Agricultura, Pecuária e Abastecimento.

14. Karim F, Majumdar S, Darabi H, Harford S. Multivariate LSTM-FCNs for time series classification. Neural Networks, v. 116, p. 237-245, 2019.

15. Bandara K, Bergmeir, C; Hewamalage H. LSTM-MSNet: Leveraging forecasts on sets of related time series with multiple seasonal patterns. IEEE transactions on neural networks and learning systems, v. 32, n. 4, p. 1586-1599, 2020.

16. Knime. Open for Innovation Knime. 2019. [Accessed: 22 Jan 2022]. Available from: https://www.knime.com/.

17. Melcher K. Multivariate Time Series Analysis with LSTMs - All Codeless. KNIME Blog. 2021. [Accessed: 22 Jan 2022]. https://www.knime.com/blog/multivariate-time-series-analysis-lstm-codeless.

18. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Accessed: 22 Jan 2022]. Available from: https://www.R-project.org/.

19. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. [Accessed: 22 Jan 2022]. Available from: http://www.rstudio.com/.

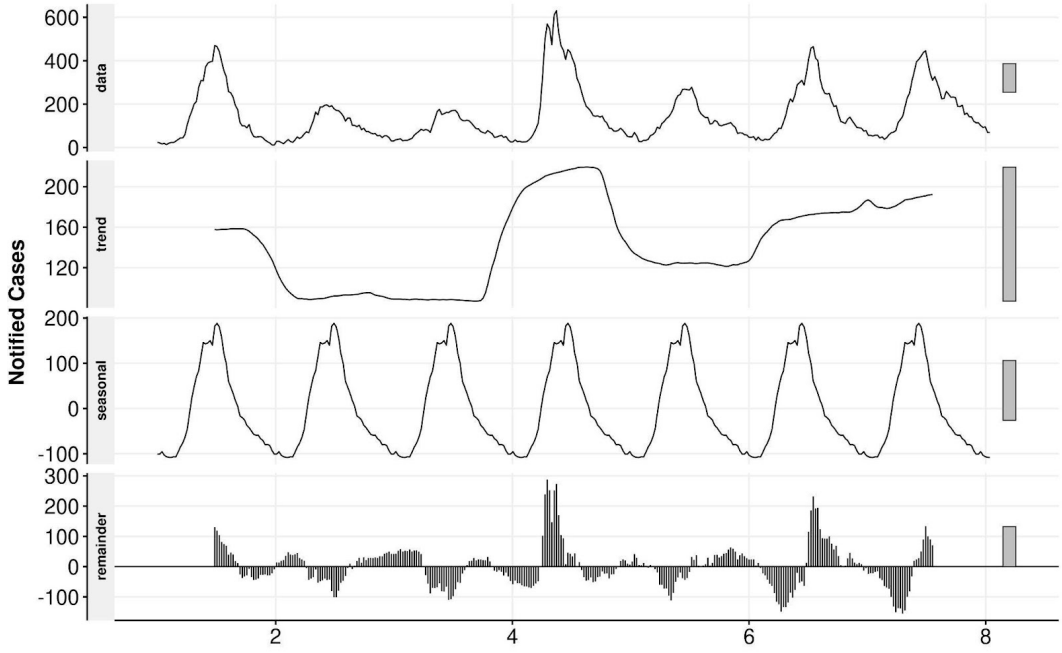
20. Fiocruz (2021). InfoGripe. Available from: http://infogripe.fiocruz.br/.

Supplementary Material

Supplementary Table I: Metrics for the test set to predict the notifications of hospitalizations by SARI in the year 2019 are shown, taking into account a model generated for the South region of Brazil. The set of categories used in the study that brought better performance to the model are shown in bold.

Attributes used	R^2	MAE	MSE	RMSE	MSD	MAPE
sex, age group, ICU use and death	0·99	1·57	3·93	1·98	0·95	0·09
sex, age group, ICU use, death and mean minimum temperature	0·97	2·25	8·01	2·83	2·05	0·12
sex, age group, positive diagnosis for influenza A/influenza B	**0·99**	**1·28**	**2·89**	**1·70**	**0·39**	**0·06**
sex, age group, ICU use, death, positive diagnosis for influenza A/influenza B	0·96	3·02	12·49	3·53	3·01	0·15
sex, age group, positive diagnosis for influenza A/influenza B and mean minimum temperature	0·99	1·39	3·79	1·95	0·64	0·06

Supplementary Figure I: Decomposition of the time series for the period between 2013 and 2019 (sequence 1 to 8) for reported cases of SARI hospitalization in the South region.



6 CONCLUSÃO

Ao longo dos últimos dois anos, curvas epidêmicas dos casos diários de SRAG passaram a fazer parte do nosso cotidiano devido à pandemia de COVID-19, sendo mostradas e compartilhadas pelos diferentes meios de comunicação. Considerando que a SRAG é um problema de saúde pública global mesmo em períodos fora de pandemias, os dados relacionados a casos de infecção respiratória são importantes ao longo de todo o ano, pois informam a situação epidemiológica destas infecções, possibilitando a tomada de medidas de controle de surtos e epidemias de SRAG. Nesse sentido, modelos de predição contribuem para o sistema de saúde.

Conforme visto em Chimmula e Zhang (2020), apesar da boa qualidade dos modelos preditivos gerados, diversos fatores impactam o curso de uma pandemia ou surto de SRAG e a predição de um surto se torna tarefa árdua, com a necessidade de um entendimento adequado dos dados disponíveis e a escolha de um algoritmo apropriado com ajustes finos dos parâmetros. Mesmo que não tenhamos um volume ideal de dados disponíveis e que ocorram algumas anomalias ao longo do tempo, como o surgimento da pandemia de Covid-19, é possível, com a utilização de Inteligência Artificial, obter modelos preditivos para surtos de SRAG com bom desempenho e que podem auxiliar os gestores de saúde na tomada mais acertada de decisão, permitindo que recursos humanos e materiais possam ser alocados na quantidade correta e em momento oportuno. O uso de redes neurais é uma opção eficiente na geração de modelos preditivos utilizando dados de séries temporais.

Este estudo possibilitou gerar diferentes modelos preditivos para SRAG e obter modelos com bom desempenho, contribuindo para a gestão em saúde.

7 RESULTADOS ADICIONAIS

Ao longo dos dois anos de pesquisa foi possível produzir alguns materiais derivados dos objetivos desta pesquisa. Abaixo estão relacionados esses produtos, que são apresentados no apêndice deste trabalho:

1) Artigo submetido à revista Health Policy and Planning (qualis A1) com o título de "Severe acute respiratory infection surveillance in Brazil: the role of public, private, and philanthropic health care units" encaminhado em fevereiro de 2021 e reencaminhado em novembro de 2022 após atender às solicitações de ajustes feitas pelos revisores (Apêndice A).

2) Capítulo de livro da Editora da UFCSPA com o título de “A Importância da Notificação de Casos de Síndrome Respiratória Aguda Grave e da Análise de Dados no Combate a Epidemias e Pandemias”, a ser publicado (Apêndice B).

3) Capítulo do e-Book do Núcleo de Pesquisa Espacial da UFCSPA com o título de “Influência do Tráfego Aéreo na Disseminação de Vírus Respiratórios”, a ser publicado em março de 2022 (Apêndice C).

4) Construção de uma ferramenta web para a visualização dos modelos preditivos gerados (Apêndice D). A ferramenta foi desenvolvida na linguagem R com o uso da biblioteca Shiny e apresenta *dashboards* que mostram gráficos de curva de ocorrência de SRAG para o próximo período, simulações e curvas de anos anteriores acompanhados de métricas de desempenho. Provisoriamente a ferramenta está publicada em https://srag.shinyapps.io/notificacoes_app/. Para a finalização e publicação definitiva da ferramenta, será necessária a validação com usuários, por isso a ferramenta deve ser considerada, no momento, como prova de conceito.

APÊNDICES

APÊNDICE A – Artigo submetido à revista Health Policy and Planning (aguardando segunda revisão).

Assunto: Health Policy and Planning – Decision on Manuscript ID HEAPOL-2021-Feb-0128
De: Health Policy and Planning <onbehalfof@manuscriptcentral.com>
Data: 04/10/2021 14:43
Para: anabgv@ufcspa.edu.br, marfcg@gmail.com

04-Oct-2021

Dear Dr. Veiga,

Your manuscript entitled "Severe acute respiratory infection surveillance in Brazil: the role of public, private, and philanthropic health care units" (HEAPOL-2021-Feb-0128), which you submitted to Health Policy and Planning, has been reviewed. The comments of the reviewer(s) are included at the bottom of this letter.

The reviewer(s) have recommended major revisions to your manuscript. We are therefore unable to accept it as it currently stands and invite you to respond to the reviewer(s)' comments and revise your manuscript.

To revise your manuscript, log into https://mc.manuscriptcentral.com/heapol and enter your Author Center, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript using a word processing program and save it on your computer. Please also highlight the changes to your manuscript within the document by using the track changes mode in MS Word or by using bold or coloured text.

Once the revised manuscript is prepared, you can upload it and submit it through your Author Center.

References-Please check as follows:

Websites-please ensure the date of access is noted-this should be the date you last accessed the website.

Numbering-following revisions the references can move in order-please check that the reference list complies with the references noted within the main document.

Please upload 2 files of the revised manuscript:

File Type "Manuscript" revised manuscript-changes marked/highlighted/underlined/bold

File Type "Manuscript" revised manuscript clean version (black/white)

When submitting your revised manuscript, please respond to the comments made by the reviewers in the space provided. Use this space to document the changes you have made to the original manuscript, on a point-by-point basis. Please be as specific as possible in your response to the reviewers – this will help to expedite our processing of the revised manuscript.

IMPORTANT: Your original files are available to you when you upload your revised manuscript. Please delete any redundant files before completing the submission.

Because we are trying to facilitate timely publication of manuscripts submitted to Health Policy and Planning, your revised manuscript should be uploaded as soon as possible. If it is not possible for you to submit your revision in a reasonable amount of time (preferably within 4-8 weeks), please contact the editorial office for an extension.

Once again, thank you for submitting your manuscript to Health Policy and Planning and I look forward to receiving your revision.

Yours sincerely,

OXFORD UNIVERSITY PRESS | Health Policy and Planning

Severe acute respiratory infection surveillance in Brazil: the role of public, private, and philanthropic health care units

Journal:	*Health Policy and Planning*
Manuscript ID	HEAPOL-2021-Feb-0128.R1
Manuscript Type:	Original Article
Country of Expertise:	Brazil
Keywords:	acute respiratory infection, hospitals, public health, surveillance, infectious diseases, health information system

Severe acute respiratory infection surveillance in Brazil: the role of public, private, and philanthropic health care units.

Abstract

Epidemiological surveillance and notification of respiratory infections is important for management and control of epidemics and pandemics. Fact-based decisions, like social distancing policies and preparation of hospital beds are taken based on several factors, including case numbers, hence health authorities need quick access to reliable and well-analyzed data. We aimed to analyze the role of the Brazilian public health system in the notification and hospitalization of patients with severe acute respiratory infection (SARI). Data of SARI cases in Brazil (2013–2020) were obtained from SIVEP-Gripe platform, and legal status of each health care unit (HCU) responsible for case notification and hospitalization was obtained from CNES database. HCUs that are part of the hospital network were classified as "Public Administration", "Business Entities", "Philanthropic Entities", or "Individuals". SARI notification data from Brazilian macroregions (North, Northeast, Midwest, Southeast, South) were analyzed and compared between administrative spheres. This study reveals that hospitalizations due to SARI increased significantly in Brazil during COVID-19 pandemic, especially in HCUs of Public Administration. In the Southeast and South, where incidence of SARI is high, philanthropic HCUs also contribute in hospitalization of SARI cases, and attend up to 7.4% of the cases notified by the Public Administration. The number of cases is usually lower in other regions, but in 2020 the Northeast showed more hospitalizations than the South. In the South, SARI season occurs later, however in 2020 an early peak was observed because of COVID-19. Notably, the contribution of each administrative sphere that manages hospital networks in Brazil in the control and management of SARI vary between regions. Our approach will allow managers to assess the use of public resources given that there are different profiles of health care in each region of Brazil, and that the public health system has a major role in notifying and attending SARI cases.

Introduction

Epidemiological surveillance of respiratory viruses is key to disease control and outbreak response. In addition, timely notification of cases of respiratory infection can prevent epidemics and pandemics. Despite many initiatives and programmes of the World Health Organization for the global surveillance of influenza and other respiratory viruses (WHO, 2013a; Hay and McCauley, 2018; WHO, 2019), the year 2020 started with an unprecedented pandemic caused by the coronavirus SARS-CoV-2. In a few months, the coronavirus disease (COVID-19) that started in Wuhan, China, at the end of 2019 reached all geographical regions and, as of November 2021, approximately 260 million people have been infected, and 5.2 million have died of COVID-19. The highest number of cases are seen now in the United States, India, and in Brazil (Dong et al., 2020). Though the former has a higher number of cases and deaths, the latter faces additional problems such as one of the worst

economic crises in its history, huge poverty belts, and uncontrolled diseases including dengue, malaria, tuberculosis, among others.

In Brazil, surveillance of respiratory infections is separated into influenza-like illness (ILI) and severe acute respiratory infection (SARI), the latter being of universal notification since 2009 (Brasil, 2010). In the case of ILI, the country adopts a sentinel surveillance network for influenza and other respiratory viruses, which started with a single sentinel unit in 2000, and gradually expanded to reach 60 units in 2010, with at least one unit in 26 of the 27 Federal Units (comprising 26 states and the Federal District). The network of sentinel units includes outpatient clinics, emergency units, and general hospitals that report weekly the aggregate total number of visits and total visits for influenza-like illnesses (ILI) through an online system called SIVEP-Gripe. ILI is defined as any individual presenting fever, accompanied by a cough or sore throat without further diagnosis (Martins et al., 2011).

At the end of 2009, all cases of SARI started to be included on SINAN (Information System on Diseases of Compulsory Declaration), a digital platform used by the Brazilian Health Ministry for reporting notifiable conditions (SINAN, 2020). In 2018, SARI notification migrated to the SIVEP-Gripe database (SIVEP-Gripe, 2020). Health institutions throughout the country notify SARI cases according to a specific notification sheet (Brasil, 2020a), and insert them directly in the system in case they have access to it or, more commonly, send those sheets to health authorities such as the Municipal and State Health Secretariats for insertion into SIVEP-Gripe (Martins et al., 2011).

As proposed by WHO and adopted by several countries, notification of SARI cases is used as syndromic surveillance aimed at quantifying hospitalizations and deaths related to respiratory viral infections (WHO, 2013b), which can be caused by influenza A virus (IAV), influenza B virus (IBV), respiratory syncytial virus (RSV), human adenovirus (HAdV), human parainfluenza virus (hPIV) 1, 2, 3, and 4, rhinovirus, among others (Martins et al., 2011). In Brazil, surveillance at HCUs is coupled with public laboratories that test for these viruses in order to assess viral circulation and seasonality, as well as the emergence of novel respiratory viruses of clinical relevance (Martins et al., 2011; Brasil, 2019).

According to the official protocol, every patient notified as a SARI case must have a suitable biological sample collected for laboratory testing, provided by the Brazilian Unified Health System (SUS) without any additional charge to the patient. In response to the SARS-CoV-2 pandemic, the Brazilian SARI surveillance guideline was updated to monitor COVID-19 hospitalizations as well (Brasil, 2020b). Due to its syndromic nature, SARI surveillance provided a general overview of the impact of COVID-19 hospitalizations in Brazil, even before the guideline update (Niquini et al., 2020).

Brazil is a country of continental size divided into 26 states and a Federal District, organized in five regions: North, Northeast, Midwest, Southeast, and South. Notably, the incidence of respiratory viral infection varies among different regions, with higher incidences in the South and Southeast; in

addition, respiratory viruses may show different circulation patterns and seasonality (Martins et al., 2011; Almeida et al., 2018; Gregianini et al., 2019; Veiga et al., 2020). Therefore, the existence of a uniformly defined national surveillance system with a centralized database is fundamental for situation assessment at the national level and the development of coordinated mitigation efforts.

The hospital network in Brazil is administered primarily by four administrative spheres (CNES, 2020):

• Public Administration: comprises public bodies, autarchies, and public foundations of the Union, the States, the Federal District, and the Municipalities, among others. It is worth mentioning that a subgroup in this category attends only patients linked to the Brazilian Armed Forces;

• Business Entities: made up of public companies, mixed capital companies, open and closed corporations, limited companies, among others;

• Philanthropic Entities: composed of private foundations, autonomous social service, private association, social organization, non-profit entities, and religious organizations;

• Individuals: this is an administrative sphere with little representation in the data collected and consists basically of independent professionals, such as physicians.

The hospital network managed by the public administration has exclusively public beds, free of charge, whereas hospitals managed by the other entities can have both free public beds and private, paid beds. Since the national surveillance system is administered by the Ministry of Health as part of the freely accessible SUS, and because SARI notification is not mandatory in Brazil despite being universal, there is no guarantee that HCUs outside public administration actually report SARI cases, especially regarding patients under private beds. Therefore, the difference observed in the number of reported cases in March 2020 (incidence of 10.4 cases per 100,000 inhabitants between epidemiological weeks [EW] 10 to 13) not only with respect to January and February (2.1 per 100,000 inhabitants between EW 1 to 9), but also to previous years (median of 1.1 with 95% CI [1.4 − 3.0] between EW 10 to 13 from 2013 to 2019) (Fiocruz, 2020) triggered warranted concerns whether this difference was simply due to sudden adherence of business and philanthropic entities to the notification protocols.

Therefore, to better understand the role of the public health system and investigate the hypothesis of a significant increase in 2020 data solely by a greater contribution from non-publicly administered HCUs, the present study assessed cases of SARI in Brazil between 2013 and 2020, providing the historical level of notifications by each administrative sphere at the national level, as well as in each region. Our results provide a detailed analysis of the potential impact of SARI cases in each administrative sphere measured by notifications, and of potential territorial differences of network coverage. Moreover, this study highlights the importance of the Brazilian public health system in the control and surveillance of respiratory viral infections.

Materials and Methods

The data used in this study were obtained from the SIVEP-Gripe database (SIVEP-Gripe, 2020), and from the CNES system (National Registry of Health Facilities) (CNES, 2020). The SIVEP-Gripe platform provides information about cases of SARI (Severe Acute Respiratory Infection) notified by health institutions in the country, and includes data such as symptoms, date of first symptoms, date of hospital admission, notification unit code, inpatient unit code, establishment's CNES code, etc. Of note, the case definition for SARI was any individual presenting dyspnea (oxygen saturation below 95%), with fever (above 37.5ºC), myalgia, lethargy, cough, sore throat (Martins et al., 2011; Brasil, 2019; Gregianini et al., 2019; Fiocruz, 2020; Veiga et al., 2020). To guarantee comparability over the years and across health care units, we only kept those cases that were notified with explicit mention of the following signs and symptoms: (fever) AND (cough OR sore throat) AND (dyspnea OR oxygen saturation below 95% OR difficulty breathing) AND (hospitalized OR deceased).

The CNES system was accessed to obtain information about each health unit indicated in the SIVEP-Gripe platform, including company name, type of unit, legal nature, location, etc (CNES, 2020). With regards to the legal nature, each health unit was classified based on four administrative spheres: "Public Administration", "Business Entities", "Philanthropic Entities", and "Individuals".

The population data were collected through the projection of the population of the federation units by sex and age groups, found in Tabnet Datasus (Brasil, 2021a). Due to the different population structures in each region of Brazil, the number of hospitalizations and their incidence was standardized according to the direct standardization method (Pan American Health Organization, 2002). The Brazilian population of the year 2020, extracted by age and sex, was considered as the standard population. The calculation of uncertainties associated with observations of cases of hospitalization and notification related to SARI in each administrative sphere for each region and year was performed using a multinomial model, based on the method proposed in Sison and Glaz (1995); Glaz and Sison (1999); May, Johns (2000). To facilitate data visualization, some graphs do not show data from the Individuals administrative sphere because of its low volume of notifications (0.04% of all reported cases).

Information on beds was obtained from the Applied Health Data Science Platform (PCDAS), from the Institute of Communication and Scientific Information and Technology in Health (ICICT) at Fiocruz (Fiocruz, 2021) and complemented with data obtained from DATASUS (Brasil, 2021a). It is important to note that the number of beds reported by each healthcare unit in the available databases does not reflect the actual number of set up and staffed beds, but rather licensed ones. Due to the ineffectiveness of data on available hospital beds in Brazil, it is not possible to carry out a direct analysis considering the volume of hospitalizations and available hospital beds. We included a table (Supplementary Table) as a proxy for the number of hospital beds available in each region, based on the yearly average number of complementary (those reserved for intensive and semi-intensive care) and clinical beds reported monthly by each health unit. Beds reserved for surgeries were left out.

At its inception in 2009, the Brazilian SARI surveillance focused on pandemic influenza strains. In August 2012 (wintertime in the Southern Hemisphere), the notification sheet was updated to include additional respiratory viruses, such as IBV, RSV, hPIV-1, 2, and 3, and HAdV (Brasil, 2012). Due to that change, our analysis included cases of SARI notified from EW 1 of 2013 to EW 53 of 2020, with a total of 615,709 cases. Of these, 2,928 records had invalid notification HCU codes and were excluded, totaling 612,781 cases that were included in further analyses. Of note, 106,315 invalid inpatient unit codes were also found, meaning that the patient was admitted in a health unit, however it is not possible to know in which establishment this admission was made; usually hospitalization occurs in the reporting unit, but this is not always the case, as we show further in this article. Regardless of that, records with such inconsistency were not discarded. As of EW 53 of 2020, 429,448 SARI cases had been notified on SIVEP-Gripe attending the criteria defined here.

For the preparation and generation of data for analysis, the Knime tool was used (KNIME, 2021). Graphics were generated with the R tool (R Core Team, 2020), the ggplot2, dplyr, and stringr packages (Wickham et al., 2019), and R package DescTools' function MultionomCI (Signorell et al. 2021). The environment for executing R codes was Google Collaboratory (Bisong, 2019). The study was approved by the Ethics Committee of UFCSPA (CAAE 75118217.9.3001.5347). All analysis relied on individually based but anonymous SARI notification data, such as those openly provided by the Brazilian Ministry of Health at https://opendatasus.saude.gov.br/dataset/bd-srag-2020. Data regarding the administrative sphere of each healthcare unit, as well as the number of licensed beds are publicly available through the National Registry of Healthcare Units (CNES).

Results

Hospitalizations by Brazilian region

The analysis of the total number of hospital admissions that occurred yearly in each Brazilian health institution showed that there was a great increase in hospitalizations in 2020 compared to previous years (Figure 1A). Notably, in hospitals of Public Administration a more than 10-fold increase was observed, revealing the contribution of this administrative sphere in attending patients during the COVID-19 pandemic.

When evaluating the proportion of notifications of admissions by administrative sphere in each region, our analysis reveals that Philanthropic Entities have historically played a relevant role in the South region whereas for other regions the administrative sphere Public Administration has a greater volume of hospitalization in relation to other spheres (Figure 1B). Results also show that the highest volume of hospitalizations related to SARI in Brazil are usually in the South, except for the year 2020, when the standardized number of hospitalizations in this region was smaller compared to other regions. In contrast, the North and Northeast regions had much higher hospitalization volumes in 2020 than in previous years (Figure 2).

Considering the typical seasonality of SARI in the different Brazilian regions, the analysis

shows that the peak of respiratory infections in the South usually occurs later than in the northern regions (Figure 3).

Hospitalization by administrative sphere

Further analyses of hospitalizations due to SARI in health units of different administrative spheres in each Brazilian region confirmed that most hospitalizations in the Midwest, Northeast, and North regions are reported from institutions of Public Administration, whereas in the South most hospitalizations are reported in Philanthropic Entities, with the exception of the year 2015 (Figure 4A). In the graph, the year 2020 is shown separately to facilitate data visualization, given the much larger volume of notifications for that year.

The contribution of each administrative sphere in the hospitalization of SARI cases in each Brazilian region is maintained along the period, including in 2020 during the COVID-19 pandemic (Figure 4B). Of note, the “Individuals” administrative sphere has little representation in hospital admission notifications, therefore this group is not shown in the graphs.

In some cases, notification of SARI is performed by a certain administrative sphere, but hospitalizations of some patients occur in HCUs of a different administrative sphere (Supplementary Figure). This is evident in the administrative sphere “Public Administration”, for which 7.4% of the SARI cases notified in 2013 were hospitalized in hospitals of the administrative sphere “Philanthropic Entities”. For other administrative spheres, admission to another administrative sphere is less common, except for the “Individuals” administrative sphere, which makes sense considering that this administrative sphere has little contribution in SARI notification, and comprehends mainly physicians in small clinics. Of note, for the years 2018 and 2019 the data reported in the SIVEP-Gripe platform presented inconsistencies for the inpatient unit code, therefore these years were not included in order to avoid misinterpretation.

Distribution by administrative sphere

The distribution of hospitalizations along the period (2013–2020) among administrative spheres for the whole country is shown in Figure 5; the administrative sphere “Individuals” was omitted because the volume of notifications was very low for this group. As can be observed, hospitalizations in institutions of Public Administration increased significantly in 2020, whereas a significant decrease in hospitalizations was observed for the ”Philanthropic Entities” group. For the administrative sphere ”Business Administration”, no significant differences were observed over the years.

Discussion

This study analyzed the notification of SARI cases in Brazil over eight years. Of note, this is the first study about the SARI notification system in Brazil that takes into account differences among institutions according to which administrative sphere they belong to. In a large country such as Brazil,

this type of analysis gives a view of the performance of each administrative sphere in different regions of the country and this allows a better assessment of the implementation of public health policies for the treatment and mitigation of diseases, especially those related to epidemics and pandemics such as respiratory infections.

Our analyses show that in the North, Northeast, and Midwest regions of Brazil, health institutions of the public sector are the main responsible for hospitalization and notification of SARI cases, whereas in the South region hospitals administered by Philanthropic Entities are the main institutions where patients with SARI are hospitalized, and in which confirmed cases are notified to health authorities. As for the Southeast, the most populous region in Brazil, a certain balance can be observed between the notifications made by the Public Administration sphere and Philanthropic Entities. These findings can be a consequence of different levels of participation of philanthropic and private HCUs in SUS, as well as of the socioeconomic inequalities within the country. For instance, a national survey from 2013 showed that families residing in the North and Northeast regions of Brazil were the ones with the lowest percentages of private health plan coverage, with estimates of 13.3% and 15.5%, respectively, whereas the remaining regions had more than 30% of its families with some type of private health plan. The overall percentage of families with a private health plan in the country was estimated at 27.9% (IBGE, 2015). These disparities among different regions have important consequences in terms of the role of the public sector in covering healthcare, be it in the direct administration of HCUs, or in offering beds through partnerships with philanthropic and other not-for-profit institutions.

Although the vast majority of SARI cases notified by a HCU from a given administration group is hospitalized in an institution of the same group, this is not always true; for example, in 2013, approximately 7.4% of the SARI cases notified by HCUs administered by the public sector were admitted in hospitals administered by Philanthropic Entities. This is a clear example of the contribution of the latter to the Brazilian SUS, since those situations usually refer to patients that sought medical attention at a HCU administered by the public sector and were triaged for hospitalization at a philanthropic one that had beds available for SUS.

The fact that most cases are notified and hospitalized within the same administrative sphere allows the use of the notification unit as a proxy for the hospitalization unit in the years 2018 and 2019, for which the latter was unavailable. As shown in Figure 5, it is clear that the dramatic increase in the number of notified SARI cases was not a consequence of greater adherence of the private sector to the national notification system. Although this study was not able to evaluate the level of compliance in each year, the fact that the relative contribution from the Private Sector to the total number of notified cases decreased in 2020 is sufficient to discard that hypothesis. It is indeed possible that the COVID-19 pandemic increased the adherence of HCUs and professionals in general, making a direct comparison with previous years not necessarily accurate, but it is clear that this change in compliance, if true, was not driven by privately ran units. Moreover, as shown in this study, the

proportion of contribution of each administrative sphere for hospitalization of SARI cases during the COVID-19 pandemic in 2020 was very similar to that of pre-pandemic years. Considering that approximately 7.7 million cases of COVID-19 were confirmed in Brazil in 2020 (Brasil, 2021b), these findings reinforce the role of public health institutions in the control and management of the pandemic.

One of the main limitations of the present study is the fact that the SARI notification database (SIVEP-Gripe) does not allow for stratification of cases between public and private healthcare access directly. The use of judicial/legal administrative sphere employed by cross-referencing with the HCU database (CNES) provides a proxy for that, which can have important biases. For instance, units administered by Business Entities or Philanthropic Entities can have beds reserved for public access through SUS, but there is no clear way of separating the percentage of beds reserved for each type of access, let alone for each type of health motivation. Nonetheless, it is known that philanthropic institutions usually share a larger portion of their facilities with SUS in comparison to units administered by Business Entities. Finally, this limitation does not affect the conclusions regarding the hypothesis of whether the dramatic increase in case counts during 2020 could be solely or largely explained by a disproportionately increased adherence in reporting by private units. In fact, this limitation only reinforces the role of public health care access, since there are cases from this particular network being notified by and hospitalized at Business and Philanthropic Entities.

Another limitation is related to the lack of reliable information about the number of hospital beds in the country. For instance, the number of beds reported by each healthcare unit in the available databases does not reflect the actual number of set up and staffed beds, but rather licensed ones. Therefore, the usage of such data as a relevant denominator would be subject to phantom beds that could vary between healthcare units and administrative spheres (Phillip et al., 1984). Hence, an analysis considering the number of beds and the volume of admissions is not reliable.

The policies for notification of health conditions and diseases vary among countries. For example, in New Zealand, the health care system is predominantly public and SARI is monitored by the public hospital network (Huang et al., 2014). In Portugal, hospital administration is either Official (Public Entities) or Private, and the number of private and official hospitals is quite similar; basically, there are no philanthropic hospitals because in 1970, with the emergence of the Portugal SNS (National Health Service), philanthropic health institutions, known as *Misericórdias*, started to be administered by the State (Fernandes and Nunes, 2016). Of note, acute respiratory infections, including influenza-like illness, are not diseases of mandatory notification in Portugal (Portugal, 2017).

Our study allows a comparative analysis between Brazil and these other countries; accordingly, in Brazil, where SARI is of universal notification (Brasil, 2010), notifications are made by outpatient clinics, emergency care departments, or general hospitals, both from public and private networks (Martins et al., 2011). This diversification has proven to be efficient in Brazil, as it represents all social strata of the population, including both sexes and different age groups. Of note, the lowest

income group makes use of the public and philanthropic network, while the highest income group uses primarily the private health network.

The universal coverage of the Brazilian SARI surveillance network has shown to be capable of identifying epidemiological characteristics of cases from novel viruses in contrast with known ones, early detection of invasion (Niquini et al., 2020), as well as genomic characterization (Souza et al., 2020). Moreover, as shown in the present study, data of SARI surveillance provides information about the seasonality of viral infections in different regions of the country, enabling health authorities to better prepare for preventing and controlling outbreaks. Nonetheless, SARI surveillance in Brazil still faces important challenges, especially in terms of timeliness of data insertion, which are paramount for its usage as a data source for analytical methods for up-to-date situation assessment to support decision making for mitigation strategies (Lana et al., 2020).

References

Almeida A, Codeço C, Luz PM. 2018. Seasonal dynamics of influenza in Brazil: the latitude effect. *BMC Infectious Diseases* **18**: 695.

Bisong EO. 2019. Building machine learning and deep learning models on Google cloud platform – a comprehensive guide for beginners. Berkeley, USA: Apress.

Brasil. 2010. Protocolo para o enfrentamento à pandemia de influenza pandêmica (H1N1) *2009*: ações da atenção primária à saúde. Brasília, Brasil: Ministério da Saúde. [Accessed: 25 November 2021]. Available from: https://bvsms.saude.gov.br/bvs/publicacoes/protocolo_enfrentamento_influenza_2009.pdf

Brasil. 2012. Ficha de registro individual destinada para unidades com internação – síndrome respiratória aguda grave (SRAG) – internada ou óbito por SRAG. Brasília, Brasil: SINAN, Ministério da Saúde. [Accessed: 16 February 2021]. Available from: http://portalsinan.saude.gov.br/images/documentos/Agravos/Influenza/Influenza_v5.pdf

Brasil. 2019. Guia de Vigilância em Saúde 3 ed. Brasília, Brasil: Ministério da Saúde, Secretaria de Vigilância em Saúde, Coordenação Geral de Desenvolvimento da Epidemiologia em Serviços. [Accessed: 16 February 2021]. Available from: https://bvsms.saude.gov.br/bvs/publicacoes/guia_vigilancia_saude_3ed.pdf

Brasil. 2020a. Ficha de registro individual – casos de síndrome respiratória aguda grave hospitalizado. Brasília, Brasil: Ministério da Saúde, Secretaria de Vigilância em Saúde. [Accessed: 16 February 2021]. Available from:

https://saude.rs.gov.br/upload/arquivos/carga20190433/05143355-25141516-1-ficha-srag-hospital.pdf

Brasil. 2020b. Guia de Vigilância Epidemiológica – Emergência de Saúde Pública de Importância Nacional pela Doença pelo Coronavírus 2019: Vigilância de Síndromes Respiratórias Agudas – COVID-19. Brasília, Brasil: Ministério da Saúde, Secretaria de Vigilância em Saúde. [Accessed: 16 February 2021]. Available from: https://portalarquivos.saude.gov.br/images/af_gvs_coronavirus_6ago20_ajustes-finais-2.pdf

Brasil. Ministério da Saúde. 2021a. DATASUS. [Homepage]. [Accessed: 25 November 2021]. Available from: http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/projpopuf.def

Brasil. Ministério da Saúde. 2021b. Painel Coronavírus. [Homepage]. [Accessed: 25 November 2021]. Available from: https://covid.saude.gov.br/

CNES – Cadastro Nacional de Estabelecimentos de Saúde. Brasília, Brasil: Ministério da Saúde. [Accessed: 16 February 2021]. Available from: http://cnes.datasus.gov.br

Dong E, Du H, Gardner L. 2020. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infectious Diseases* **20**: P533–4.

Fiocruz. 2020. Grupo de Métodos Analíticos de Vigilância Epidemiológica (MAVE – PROCC/Fiocruz and EMap/FGV), GT-Influenza, Secretaria de Vigilância em Saúde, Ministério da Saúde. InfoGripe – Monitoramento de casos de síndrome respiratória aguda grave (SRAG) notificados no SIVEP-Gripe. Rio de Janeiro, Brasil: Fiocruz. [Accessed: 16 February 2021]. Available from: http://info.gripe.fiocruz.br

Fiocruz – Fundação Oswaldo Cruz. 2021. Applied Health Data Science Platform. [Accessed: 16 November 2021]. Available from: https://pcdas.icict.fiocruz.br.

Fernandes AC, Nunes AM. 2016. Os Hospitais e a Combinação Público-Privado no Sistema de Saúde Português. *Acta Médica Portuguesa* **29**: 217–23.

Glaz J, Sison CP. 1999. Simultaneous confidence intervals for multinomial proportions. Journal of Statistical Planning and Inference 82:251-262.

Gregianini TS, Seadi CF, Zavarize Neto LD, et al. 2019. A 28-year study of human parainfluenza in Rio Grande do Sul, Southern Brazil. *Journal of Medical Virology* **91**: 1423–31.

Hay AJ, McCauley JW. 2018. The WHO global influenza surveillance and response system (GISRS)—a future perspective. *Influenza and Other Respiratory Viruses* **12**: 551–7.

Huang QS, Baker M, McArthur C, et al. 2014. Implementing hospital-based surveillance for severe acute respiratory infections caused by influenza and other respiratory pathogens in New Zealand. *Western Pacific Surveillance and Response Journal (WPSAR)* **5**: 23–30.

IBGE – Instituto Brasileiro de Geografia e Estatística. 2015. Pesquisa Nacional de Saúde, 2013. Rio de Janeiro, Brasil: Instituto Brasileiro de Geografia e Estatística. [Accessed: 16 February 2021]. Available from: https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html?=&t=o-que-e

KNIME. 2021. Available from: https://www.knime.com [accessed 16 February 2021].

Lana RM, Coelho FC, Gomes MFDC, et al. 2020. The novel coronavirus (SARS-CoV-2) emergency and the role of timely and effective national health surveillance. *Cadernos de Saúde Pública* **36**: e00019620. http://dx.doi.org/10.1590/0102-311x00019620.

Martins LG, Menegolla IA, Ranieri T, Bercini M, Ueda ES, Porto MA. 2011. A Vigilância da Influenza no Rio Grande do Sul. *Boletim Epidemiológico* **13**: 4–11.

May W L, Johnson WD. 2000. Constructing two-sided simultaneous confidence intervals for multinomial proportions for small counts in a large number of cells. Journal of Statistical Software, 5(6), 1–24. software, 5(6), 1–24. https://doi.org/10.18637/jss.v005.i06.

Niquini RP, Lana RM, Pacheco AG, et al. 2020. Description and comparison of demographic characteristics and comorbidities in SARI from COVID-19, SARI from influenza, and the Brazilian general population. *Cadernos de Saúde Pública* **36**: e00149420. https://doi.org/10.1590/0102-311x00149420.

Pan American Health Organization. Standardization: a classic epidemiological method for the comparison of rates. Epidemiol Bull 2002, v. 23, n. 3, p. 9-12.

Phillip PJ, Mullner R, Andes S. 1984. Toward a better understanding of hospital occupancy rates. Health Care Financ Rev. 5(4):53-61.

Portugal. 2017. Doenças de Declaração Obrigatória 2013-2016, Volume I. Lisboa, Portugal: Ministério da Saúde, Direção-Geral da Saúde, [Accessed: 16 February 2021]. Available from: https://comum.rcaap.pt/bitstream/10400.26/22529/1/Doenças%20de%20Declaração%20Obrigatória%202013-2016%2C%20Volume%20I%20-%20Portugal.pdf

R Core Team. 2020. R: A language and environment for statistical computing. Austria, Vienna: R Foundation for Statistical Computing. [Accessed: 16 February 2021]. Available from: https://www.R-project.org

Signorell A, et al. 2021. DescTools: Tools for descriptive statistics. R package version 0.99.41.

SINAN – Sistema de Informação de Agravos de Notificação. 2020. Brasília, Brasil: Ministério da Saúde. [Accessed: 16 February 2021]. Available from: http://portalsinan.saude.gov.br

Sison CP, Glaz J. 1995. Simultaneous confidence intervals and sample size determination for multinomial proportions. Journal of the American Statistical Association, 90:366-369.

SIVEP-Gripe – Sistema de Informação da Vigilância Epidemiológica da Gripe. 2020. Brasília, Brasil: Secretaria de Vigilância em Saúde. [Accessed: 16 February 2021]. Available from: https://sivepgripe.saude.gov.br/sivepgripe

Souza WM, Buss LF, Candido DS, et al. 2020. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nature Human Behavior* **4**: 856–65.

Veiga ABG, Martins LG, Riediger I, Mazetto A, Debur MDC, Gregianini TS. 2020. More than just a common cold: Endemic coronaviruses OC43, HKU1, NL63, and 229E associated with severe acute respiratory infection and fatality cases among healthy adults. *Journal of Medical Virology* **93**: 1002–7.

WHO. 2019. Global influenza strategy 2019–2030. Geneva, Switzerland: World Health Organization. [Accessed: 16 February 2021]. Available from: https://apps.who.int/iris/handle/10665/311184

WHO. 2013a. Research needs for the battle against respiratory viruses (BRaVe) – Background document. Geneva, Switzerland: World Health Organization. [Accessed: 16 February 2021]. Available from: https://www.who.int/influenza/patient_care/clinical/BRaVe_Research_Agenda_2013.pdf?ua=1

WHO. 2013b. Global Epidemiological Surveillance Standards for Influenza. Geneva, Switzerland: World Health Organization. [Accessed: 16 February 2021]. Available from:

https://www.who.int/influenza/resources/documents/WHO_Epidemiological_Influenza_Surveillance_Standards_2014.pdf?ua=1

Wickham H, Averick M, Bryan J, et al. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* **4**: 1686. https://doi.org/10.21105/joss.01686.

Figure captions

Figure 1: A) Number of hospitalizations due to SARI in Brazil, 2013–2020. The graph represents the total number of hospital admissions per 1000 inhabitants for each year, stratified by administrative sphere, according to the order in the graph: Public Administration, Business Entities, Philanthropic Entities. The individual administrative sphere is not considered due to a small number of hospitalization cases. B) Proportion of hospitalizations due to SARI in each Brazilian region, by administrative sphere. The five Brazilian regions are shown according to the order in the graph: Midwest, Northeast, North, Southeast, South.

Figure 2: Standardized number of hospitalized cases per 1000 inhabitants, by region in each year.

Figure 3: Seasonality of SARI in Brazil. The total number of patients hospitalized with SARI in each epidemiological week (2017–2020) is shown for each Brazilian region; data for 2020 is shown until epidemiological week 49.

Figure 4: A) Total cases of admissions per region per 1000 inhabitants in each administrative sphere, with a standardized population. To improve the visualization of the graph, data from the administrative sphere “individuals” are not presented, as the number of cases is not significant (only 0.04% of the total cases). The year 2020 is shown separately from the other years because, due to the COVID-19 pandemic, the scale was very different from other years and this would make it difficult to visualize the data in the graphs. B) Percentage of standardized cases of hospitalization due to SARI, as reported by administrative sphere and in each region, in relation to the standardized total number of hospitalized cases due to SARI in each region.

Figure 5: Proportion of hospitalizations notified by each administrative sphere. Percentages were normalized by the total number of hospitalizations notified in each year. The figure shows the percentage of each group along the period 2013–2020 (Public Administration, dotted line; Philanthropic Entities, dashed line; Business Entities, dashed-dotted line), adjusted values, and 95% confidence interval (solid lines and shaded area).

Supplementary Figure: Proportion of hospitalizations in an administrative sphere different from the notification administrative sphere. Percentages were normalized considering the total number of notifications by each administrative sphere in the corresponding year. The four administrative spheres groups were considered according to the order in the graph: Public Administration, Business Entities,

Philanthropic Entities, and Individuals. The years 2018 and 2019 were omitted because of inconsistent data, such as invalid HCU codes.

Supplementary Table: Monthly average number of available beds for the year 2020, by Brazilian region and administrative sphere. Only complementary and clinical beds are being considered. As the updated data is not available in detail, it is presented here as a proxy for the number of available beds in the hospital network.

Figure captions

Figure 1: A) Number of hospitalizations due to SARI in Brazil, 2013–2020. The graph represents the total number of hospital admissions per 1000 inhabitants for each year, stratified by administrative sphere, according to the order in the graph: Public Administration, Business Entities, Philanthropic Entities. The individual administrative sphere is not considered due to a small number of hospitalization cases. B) Proportion of hospitalizations due to SARI in each Brazilian region, by administrative sphere. The five Brazilian regions are shown according to the order in the graph: A) Midwest, B) Northeast, C) North, D) Southeast, E) South.

Figure 2: Standardized number of hospitalized cases per 1000 inhabitants, by region in each year.

Figure 3: Seasonality of SARI in Brazil. The total number of patients hospitalized with SARI in each epidemiological week (2017–2020) is shown for each Brazilian region; data for 2020 is shown until epidemiological week 49.

Figure 4: A) Total cases of admissions per region per 1000 inhabitants in each administrative sphere, with a standardized population. To improve the visualization of the graph, data from the administrative sphere “individuals” are not presented, as the number of cases is not significant (only 0.04% of the total cases). The year 2020 is shown separately from the other years because, due to the COVID-19 pandemic, the scale was very different from other years and this would make it difficult to visualize the data in the graphs. B) Percentage of standardized cases of hospitalization due to SARI, as reported by administrative sphere and in each region, in relation to the standardized total number of hospitalized cases due to SARI in each region.

Figure 5: Proportion of hospitalizations notified by each administrative sphere. Percentages were normalized by the total number of hospitalizations notified in each year. The figure shows the percentage of each group along the period 2013–2020 (Public Administration, dotted line; Philanthropic Entities, dashed line; Business Entities, dashed-dotted line), adjusted values, and 95% confidence interval (solid lines and shaded area).

Supplementary Figure: Proportion of hospitalizations in an administrative sphere different from the notification administrative sphere. Percentages were normalized considering the total number of notifications by each administrative sphere in the corresponding year. The four administrative spheres groups were considered according to the order in the graph: Public Administration, Business Entities,

Philanthropic Entities, and Individuals. The years 2018 and 2019 were omitted because of inconsistent data, such as invalid HCU codes

Supplementary Table: Monthly average number of available beds for the year 2020, by Brazilian region and administrative sphere. Only complementary and clinical beds are being considered. As the updated data is not available in detail, it is presented here as a proxy for the number of available beds in the hospital network.

Figure 1

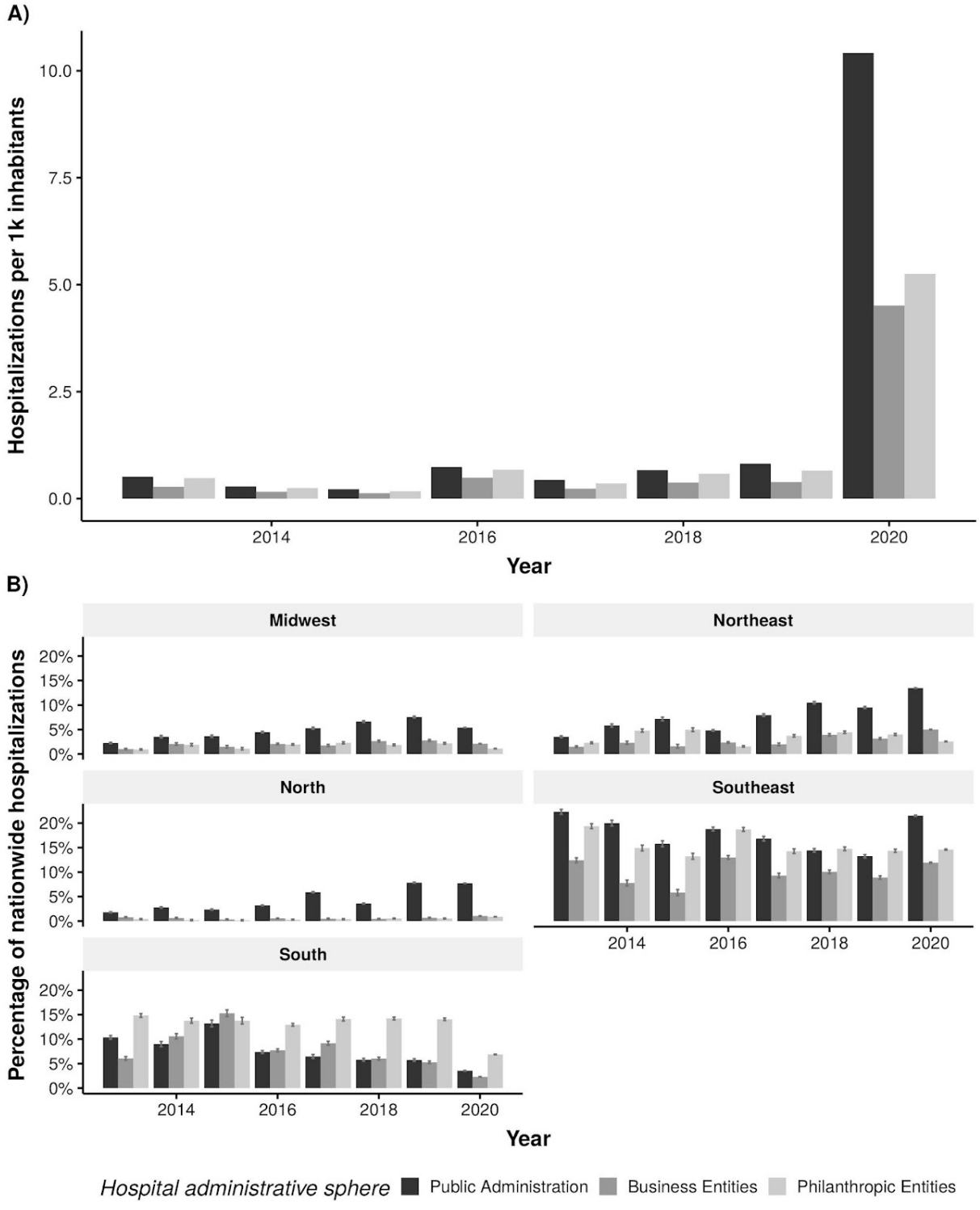


Figure 2

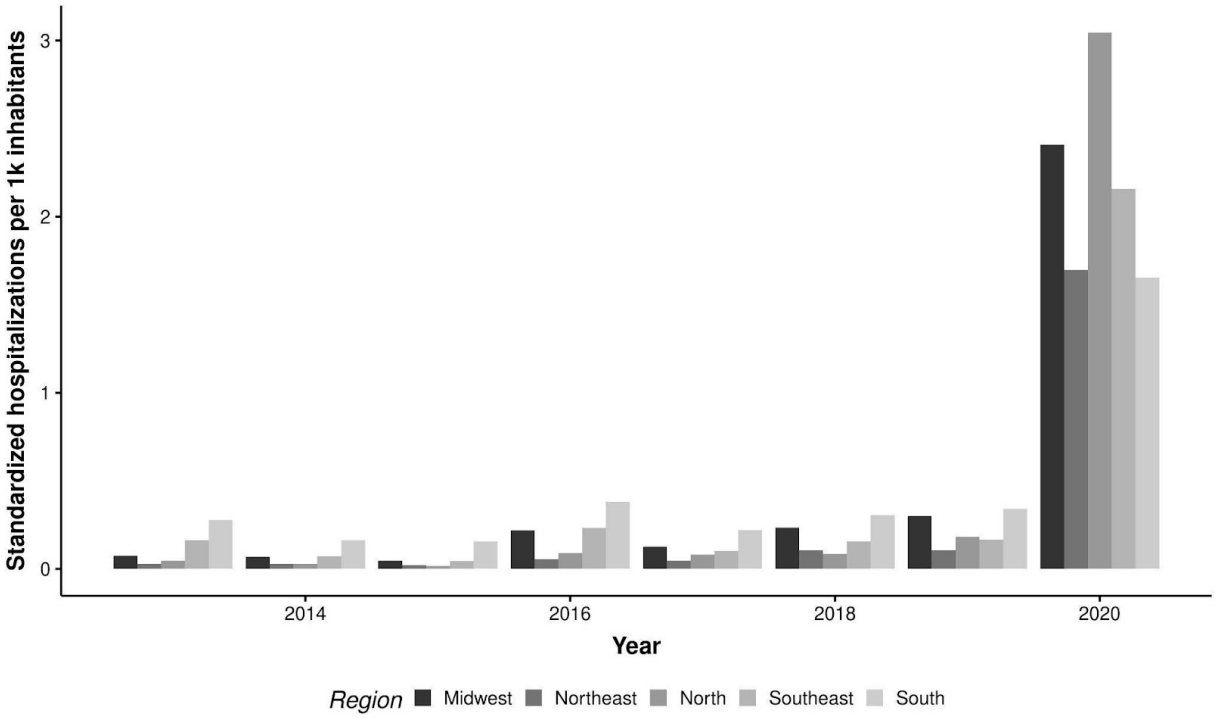


Figure 3

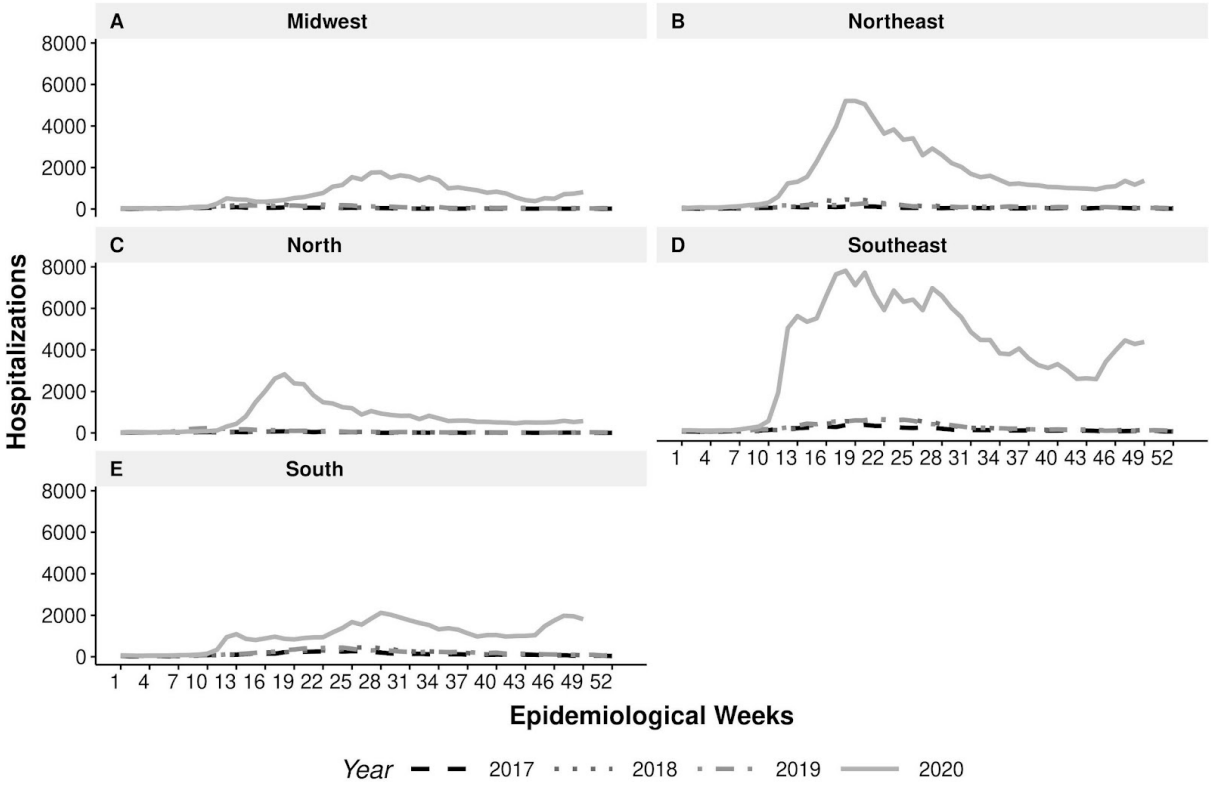


Figure 4

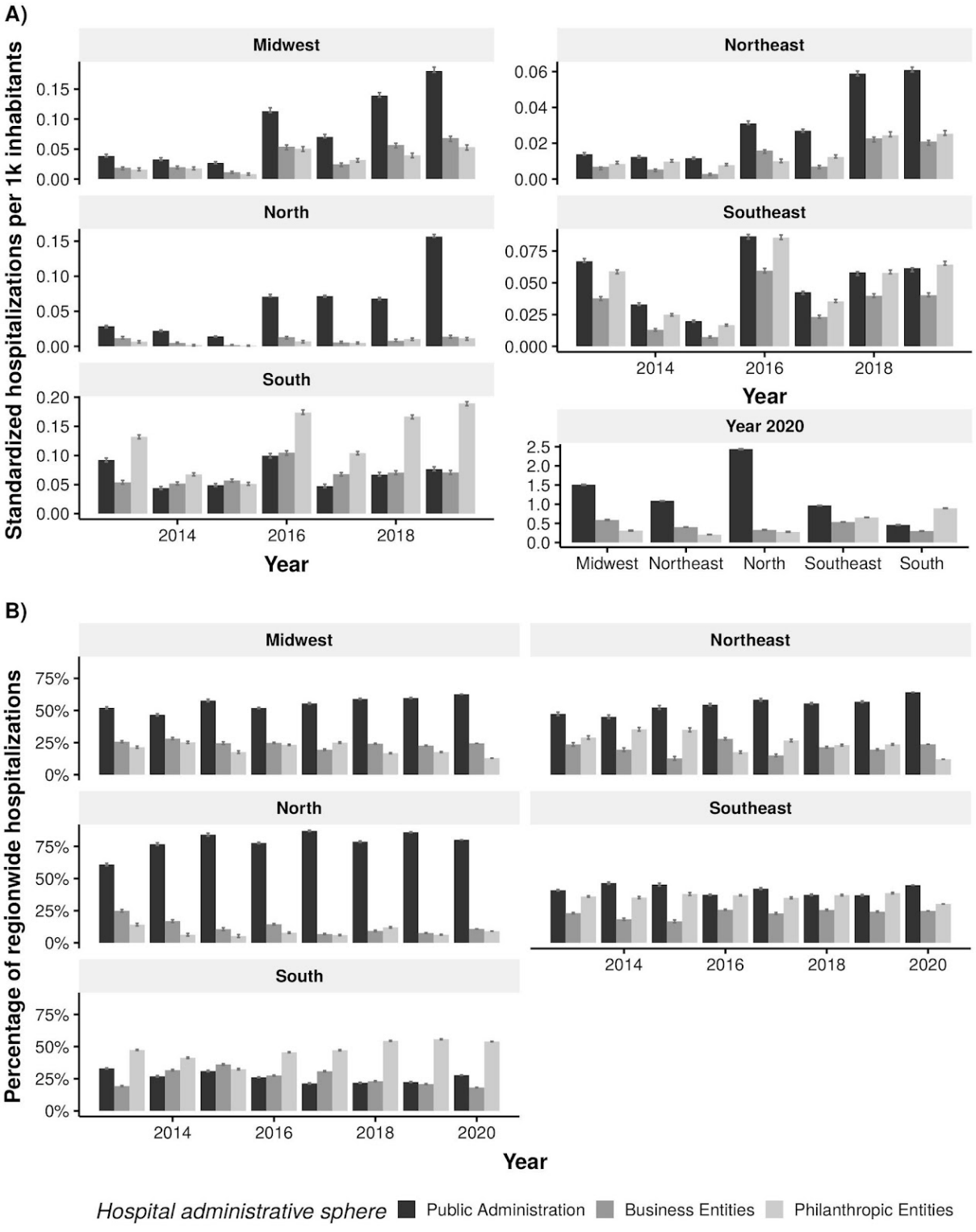
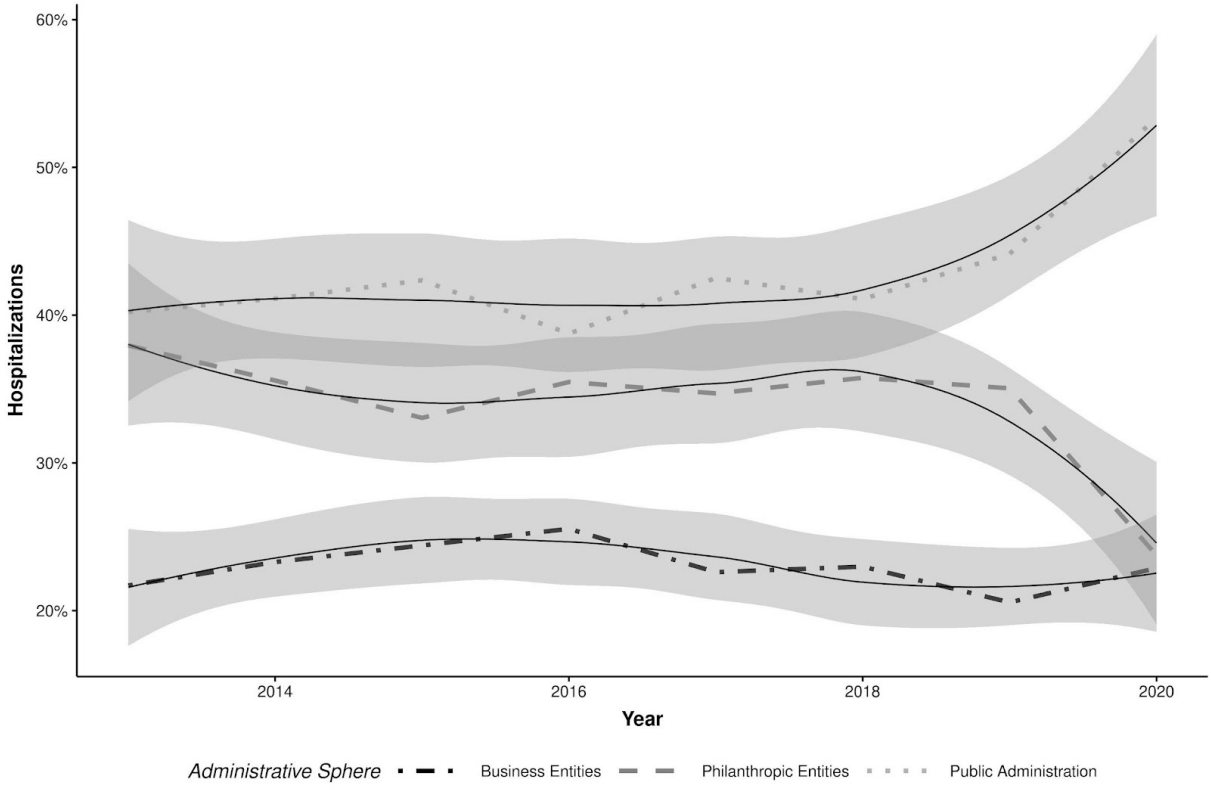
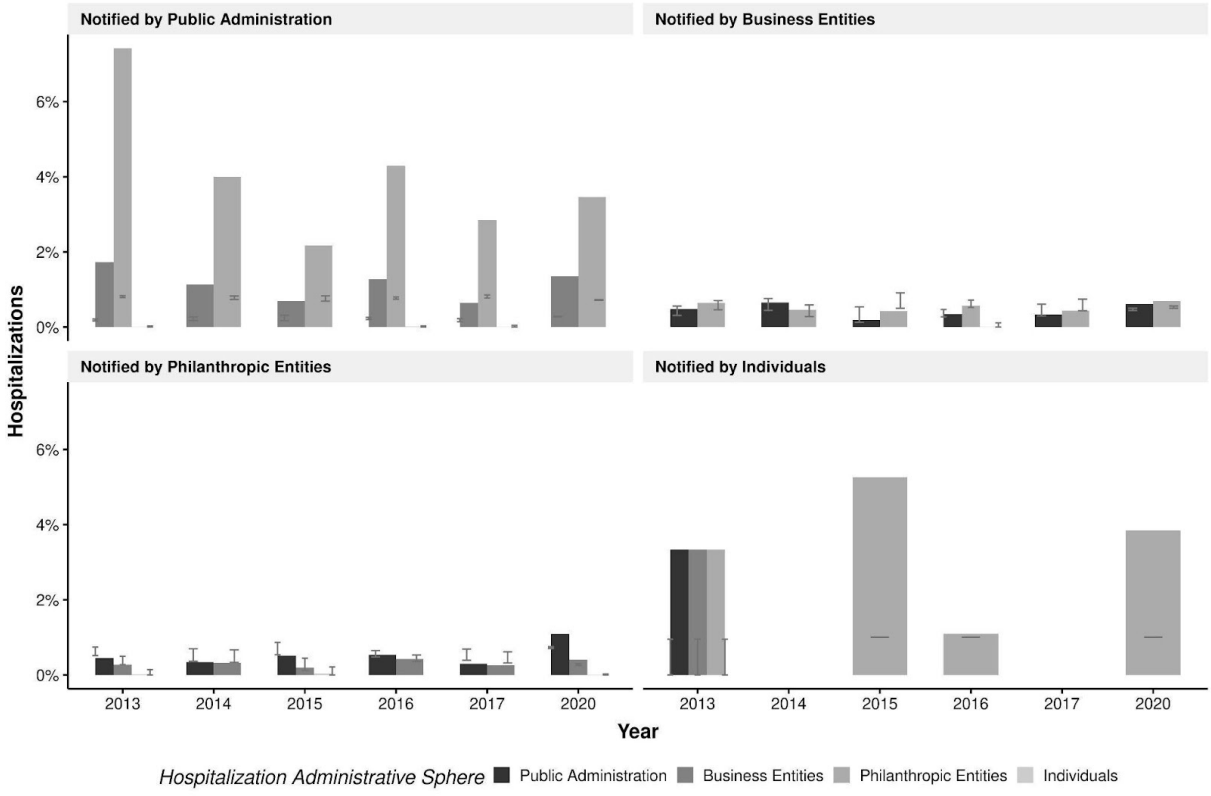

A)
Standardized hospitalizations per 1k inhabitants
Midwest
Northeast
North
Southeast
South
Year 2020
Year
2014
2016
2018
Midwest
Northeast
North
Southeast
South
B)
Percentage of regionwide hospitalizations
2020
Hospital administrative sphere
Public Administration
Business Entities
Philanthropic Entities

Figure 5



Supplementary Figure



Supplementary Table

Average Number of Licensed Beds

	Midwest	Northeast	North	Southeast	South	Sum
Public Administration	9,149 (38.9%)	40,455 (59.4%)	13,152 (64.9%)	40,747 (33.5%)	8,978 (20.3%)	112,481 (40.5%)
Business Entities	9,512 (40.4%)	13,797 (20.2%)	4,872 (24.0%)	33,742 (27.8%)	8,152 (18.5%)	70,075 (25.2%)
Philanthropic Entities	4,873 (20.7%)	13,901 (20.4%)	2,247 (11.1%)	46,988 (38.7%)	27,033 (61.2%)	95,042 (34.2%)
Total	**23,534**	**68,153**	**20,271**	**121,477**	**44,163**	**277,598**

APÊNDICE B – Capítulo do livro da Editora UFCSPA

A Importância da Notificação de Casos de Síndrome Respiratória Aguda Grave e da Análise de Dados no Combate a Epidemias e Pandemias

Amauri Duarte da Silva, Marcelo Ferreira da Costa Gomes e Ana Beatriz Gorini da Veiga

Abstract

Notification of SARS (Severe Acute Respiratory Syndrome) is paramount for the control of outbreaks and epidemics of respiratory infections. The use of notification data becomes effective when predictive models are created, which allow public managers to take early actions to combat and prevent outbreaks and eventually pandemics. Analysis of these data also allows identifying how notification and hospitalization of SARS cases occur in each region of Brazil, serving for important decisions such as the distribution of public funds. However, the notification system presents problems related mainly to inadequate information technology infrastructure in health units, which impact the quality of the data available; techniques such as nowcast can be used to minimize these problems.

Resumo

A notificação de SRAG (Síndrome Respiratória Aguda Grave) é de extrema importância para o controle de surtos e epidemias por infecções respiratórias. A utilização desses dados se torna efetiva quando são criados modelos preditivos que permitam aos gestores públicos a tomada antecipada de ações de combate e prevenção de surtos que podem levar a pandemias. A análise desses dados permite identificar como as notificações e internações de casos de SRAG ocorrem em cada região do Brasil, servindo também para decisões importantes como a distribuição de verbas públicas. Entretanto, o sistema de notificação apresenta problemas, relacionados principalmente à deficiência na infraestrutura de tecnologia de informação das unidades de saúde, comprometendo a qualidade dos dados disponíveis; técnicas como nowcast podem ser utilizadas para minimizar esses problemas.

#.1. Notificação da Síndrome Respiratória Aguda Grave

A Síndrome Respiratória Aguda Grave (SRAG) constitui um problema de saúde pública global, sendo causa importante de hospitalização e mortalidade, principalmente entre crianças, pessoas com mais de 65 anos e indivíduos com comorbidades [Veiga et al. 2020]. A definição de caso de SRAG é qualquer indivíduo que apresente febre (acima de 37,5ºC, mesmo que referida), acompanhada de tosse ou dor de garganta, e que apresente dispneia ou saturação de O_2 <95% ou desconforto respiratório [Ministério da Saúde 2021].

Muitos casos de SRAG estão associados à infecção por vírus respiratórios, alguns com alto potencial pandêmico, como por exemplo o vírus influenza A e alguns coronavírus [Veiga et al. 2020]. No Brasil, a notificação de SRAG se tornou compulsória em 2006 [Ministério da Saúde 2006] e, a partir do final de 2009, os casos de SRAG passaram a ser incluídos no Sinan (Sistema de Informação de Agravos de Declaração Compulsória), plataforma digital utilizada pelo Ministério da Saúde para registro de notificação compulsória[2]. Em 2018, os dados de notificação de SRAG e de Síndrome Gripal (SG)

[2] http://portalsinan.saude.gov.br

passaram a ser incluídos no Sistema de Informação de Vigilância Epidemiológica da Gripe (SIVEP-Gripe)[3,3].

A notificação de casos e disponibilização dos dados de SRAG são extremamente importantes para o controle de surtos e epidemias por infecções respiratórias. Com base em dados epidemiológicos e da circulação dos principais vírus respiratórios na população, é possível realizar análises de predição da ocorrência e da intensidade de surtos de SRAG, utilizando, por exemplo, Aprendizado de Máquina (AM). Quando os dados do SIVEP-Gripe são examinados em combinação com os dados do Cadastro Nacional de Estabelecimentos de Saúde (CNES)[4], torna-se possível realizar análises com diversos enfoques e estratificações. Por exemplo, com os dados de SRAG da base do SIVEP-Gripe, analisando os dados de forma estratificada de acordo com os diversos grupos jurídicos responsáveis pela administração da rede hospitalar, é possível identificar como as notificações de SRAG e internações ocorrem em cada região do Brasil. Uma análise desse tipo, além de mostrar a evolução de uma epidemia ou pandemia ao longo do tempo, também permite aos gestores uma melhor alocação das verbas e recursos disponíveis, conforme o perfil de notificação identificado em cada região. Neste sentido, uma análise dos dados de SRAG entre os anos de 2013 e 2019 no Brasil mostra que os hospitais de administração pública têm um importante papel na notificação de casos e hospitalização dos pacientes, principalmente nas regiões norte, nordeste e centro-oeste; já na região sul, hospitais administrados por entidades filantrópicas são os que mais atendem casos de SRAG. Vale lembrar que muitos dos leitos de UTI das instituições sem fins lucrativos são leitos SUS, o que reforça a importância da rede pública de saúde no atendimento dos casos de SRAG no país.

#.2. Análise de dados epidemiológicos

No atual cenário da pandemia de COVID-19, doença causada pelo novo coronavírus SARS-CoV-2, os dados sobre o número de indivíduos infectados pelo coronavírus e de óbitos por COVID-19 são mostrados e analisados pela imprensa [Roncalli e Lacerda 2020] utilizando, em muitos casos, a média móvel [Paixão et al. 2020], a qual tenta ajustar a falta de registro dos dados de casos que ocorreram nos fins de semana e feriados, considerando normalmente um período de sete dias no cálculo da média. A análise feita com esse recurso estatístico seria adequada, mas essa análise não leva em consideração outros atrasos de notificação.

A Figura #.1 apresenta a quantidade de semanas entre as manifestações dos primeiros sintomas de síndrome respiratória apresentados pelo paciente e a digitação (entrada) dos dados no sistema SIVEP-Gripe, para casos digitados entre as semanas epidemiológicas 1 e 52 de 2020 nas cinco regiões brasileiras. Para as regiões sul e sudeste, o tempo que leva entre os primeiros sintomas apresentados pelo indivíduo até a notificação do caso é de até duas semanas, com 50% dos casos notificados até a primeira semana. Já para a região norte, a defasagem entre os primeiros sintomas e a notificação do caso pode ser de até cinco semanas, com 50% dos casos notificados até a segunda semana. O principal motivo para esses atrasos está relacionado a fatores estruturais; por exemplo, em algumas localidades do território brasileiro, os dados ainda são preenchidos em fichas de papel, que só são registradas no sistema informatizado após o acúmulo de um certo volume de fichas. No momento da digitação desses dados, a falta de uma validação pode também levar a um preenchimento incorreto, que poderia ser evitado com a correção automática caso o processo fosse feito no momento em que os dados foram gerados [Lana et al. 2020]. Outro fator que dificulta o processo é a falta de acesso à internet: de um total de 406.356 unidades de saúde no país, 10% não possuem informação quanto ao acesso à internet; e, das 365.618 que possuem informação, 84% possuem acesso e 16% não possuem acesso à

[3] https://sivepgripe.saude.gov.br/sivepgripe
[4] http://cnes.datasus.gov.br

internet e têm que transportar fisicamente seus dados até um estabelecimento que possua acesso à internet e que possa fazer o registro desses dados[5].

Uma solução para minimizar os efeitos do atraso no registro dos dados é o uso da técnica de *nowcast* [Guenther et al. 2020; Bastos et al. 2019] em que, a partir de dados disponíveis de um período passado, realiza-se ajustes nos dados atuais, utilizando modelos gerados por IA (Inteligência Artificial). A Figura #.2 apresenta um exemplo da utilização de *nowcast* pelo sistema InfoGripe[6] para monitoramento e análise de situação de SRAG para a vigilância nacional [Ministério da Saúde 2019], ilustrando tanto a qualidade da estimativa quanto o baixo volume de casos inseridos no sistema ao final da própria semana de primeiros sintomas dos casos que vêm a ser notificados. Na figura é possível visualizar o número de novos casos de SRAG por semana epidemiológica de primeiros sintomas notificados no Brasil nas temporadas de 2017 e 2018, comparando os dados consolidados, o número de casos notificados até o encerramento da própria semana, e a mediana da estimativa calculada pelo sistema InfoGripe ao término de cada semana epidemiológica.

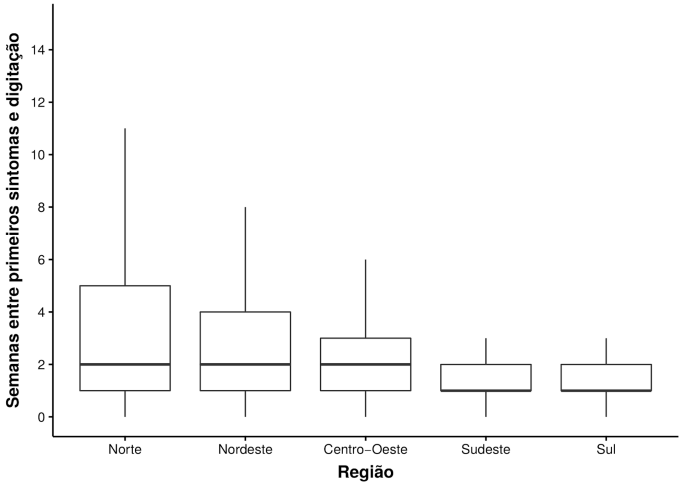


Figura #.1. Atraso nas notificações de SRAG.
Fonte: SIVEP-Gripe

Com a pandemia da COVID-19, em 2020, houve um aumento significativo de notificações de casos de SRAG em todo o Brasil, inclusive em regiões onde a notificação de SRAG costumava ser pouco significativa como, por exemplo, norte e nordeste. Esta mudança de perfil pode ser observada quando verificamos a quantidade de notificações de SRAG por semana epidemiológica nas cinco regiões brasileiras. As regiões Norte e Nordeste tiveram um aumento de notificações proporcionalmente superior às regiões Sudeste e Sul. Nota-se também um deslocamento dos períodos de pico de notificações para todas as regiões. Esse deslocamento se deu por antecipação ou atraso no período de pico ao longo das semanas epidemiológicas. Um aumento substancial de casos de SRAG e um deslocamento no período normalmente esperado de pico demandam uma estratégia de coleta e análise de dados ainda mais elaborada, visando garantir a acurácia dos resultados.

[5] http://cnes.datasus.gov.br
[6] http://info.gripe.fiocruz.br/

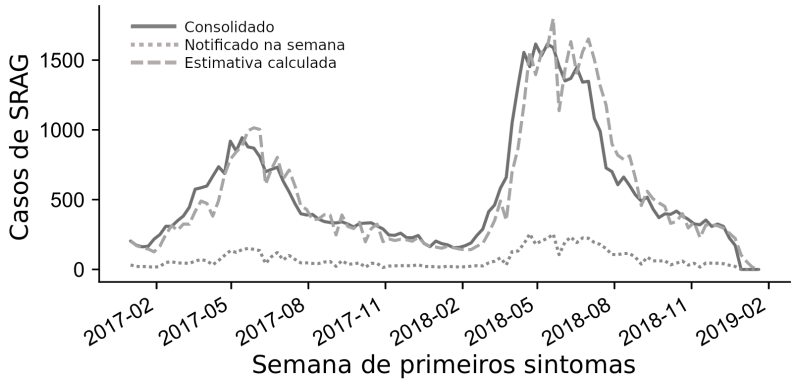


Figura #.2. Estimativa de casos recentes (nowcast) de SRAG.
Fonte: InfoGripe

#.3. Para além dos dados epidemiológicos

Conforme já mencionado, os dados de notificação obtidos do Sinan fornecem informações sobre a circulação de vírus respiratórios. Sabe-se que alguns vírus apresentam uma sazonalidade característica, geralmente associada à temperatura, em que ocorre um aumento da circulação viral – e, portanto, dos casos de SRAG – com a redução da temperatura [Price et al. 2019; Pscheidt et al. 2020]. Desta forma, dados relacionados a fatores ambientais, principalmente temperatura e pluviometria, também devem ser considerados na predição de ocorrência de SRAG, a fim de otimizar a tomada de medidas de controle e prevenção de infecções respiratórias virais. Uma boa aplicação da sazonalidade de um vírus pode ser observada no Info Dengue, que mostra a situação epidemiológica da dengue, chikungunya e zika em 1.827 municípios brasileiros [Codeço et al. 2018].

#.4. Conclusão

A eficácia do sistema de notificação de SRAG é fundamental para que a qualidade dos dados gerados possibilite a análise dos casos com base em informações atualizadas, aplicando metodologia adequada de correção para dados ausentes. Essas medidas, quando realizadas de maneira oportuna, servem de base para tomada de decisão dos gestores de saúde, pois permitem a alocação adequada de recursos financeiros e humanos no enfrentamento às infecções respiratórias. Com isso, evita-se o esgotamento de vagas em hospitais com a aplicação tempestiva de medidas de contenção. Além disso, o custo de operação se torna menor, pois uma tomada de decisão em tempo oportuno permite que um processo licitatório para a aquisição de medicamentos e equipamentos seja adequadamente conduzido, sem a necessidade de contratações emergenciais, as quais tendem a apresentar alto custo. Ademais, essas contratações aumentam o risco de infração à legislação o que, não invariavelmente, resulta em processos judiciais e desgaste à imagem da gestão pública. É importante destacar os possíveis efeitos de uma análise baseada em dados defasados: a população, com conhecimento muitas vezes insuficiente para entender os dados apresentados, pode ser levada a interpretar uma situação grave ou gravíssima – em que a infecção por um vírus respiratório como, por exemplo, o SARS-CoV-2 esteja em alta – como uma situação de normalidade, pois os dados analisados não refletem a realidade daquele momento.

Desta forma, é visível a importância do investimento dos governos para o aperfeiçoamento na coleta, análise e disponibilidade dos dados de SRAG com qualidade. Esse aperfeiçoamento deve vir de forma ampla, com melhoria na infraestrutura das unidades de saúde, melhoria nos protocolos para registros de dados e aplicação de Inteligência Artificial. A pergunta que fica é: se tivéssemos um sistema ideal de notificação de SRAG, qual seria o impacto na redução da carga de doenças respiratórias e,

consequentemente, na economia para o sistema público de saúde? Da mesma forma, podendo prever antecipadamente a dimensão de um surto de SRAG, qual seria o impacto econômico resultante da redução nas contratações emergenciais?

#.5. Leituras recomendadas

Codeço, C., Coelho, F., Cruz, O., Oliveira, S., Castro, T., Bastos, L. (2018) Infodengue: A nowcasting system for the surveillance of arboviruses in Brazil. Revue d'Épidémiologie et de Santé Publique 66: S386.

de Mello Freitas, F. T. (2013). Sentinel surveillance of influenza and other respiratory viruses, Brazil, 2000–2010. The Brazilian Journal of Infectious Diseases, 17(1), 62-68.

Referências

Bastos, L. S., Economou, T., Gomes, M. F., Villela, D. A., Coelho, F. C., Cruz, O. G., ... & Codeço, C. T. (2019) A modelling approach for correcting reporting delays in disease surveillance data. Statistics in Medicine 38: 4363-4377.

Codeço, C., Coelho, F., Cruz, O., Oliveira, S., Castro, T., Bastos, L. (2018) Infodengue: A nowcasting system for the surveillance of arboviruses in Brazil. Revue d'Épidémiologie et de Santé Publique 66: S386.

Guenther, F., Bender, A., Katz, K., Kuechenhoff, H., Hoehle, M. (2020) Nowcasting the Covid-19 pandemic in Bavaria. Biometrical Journal: 1-13, https://doi.org/10.1002/bimj.202000112

Lana, R. M., Coelho, F. C., Gomes, M. F. D. C., Cruz, O. G., Bastos, L. S., Villela, D. A. M., Codeço, C. T. (2020) Emergência do novo coronavírus (SARS-CoV-2) e o papel de uma vigilância nacional em saúde oportuna e efetiva. Cadernos de Saúde Pública 36: e00019620.

Ministério da Saúde, Secretaria de Vigilância em Saúde. Ficha de registro individual: casos de síndrome respiratória aguda grave hospitalizado. Disponível em: https://saude.rs.gov.br/upload/arquivos/carga20190433/05143355-25141516-1-ficha-srag-hospital.pdf, 2021 (acessado em 07/Jan/2021).

Ministério da Saúde, Secretaria de Vigilância em Saúde. Portaria nº 5, de 21 de fevereiro de 2006. Inclui doenças na relação nacional de notificação compulsória, define doenças de notificação imediata, relação dos resultados laboratoriais que devem ser notificados pelos Laboratórios de Referência Nacional ou Regional e normas para notificação de casos. Diário Oficial União. 22 fev 2006; Seção 1, pág. 34.

Ministério da Saúde, Secretaria de Vigilância em Saúde. Plano de Contingência para Resposta às Emergências em Saúde Pública - Influenza Preparação Para a Sazonalidade e Epidemias. Brasília-DF, 2019. Disponível em: https://portalarquivos2.saude.gov.br/images/pdf/2019/marco/20/Plano-de-Conting--ncia-para-Sazonalidade-e-Epidemias-de-Influenza---Final-enviado-19.03.2019.pdf. Acesso em: 12 jan. 2020.

Paixão, B., Baroni, L., Salles, R., Escobar, L., de Sousa, C., Pedroso, M., ... & Ogasawara, E. (2020) Estimation of COVID-19 under-reporting in Brazilian States through SARI. arXiv preprint arXiv:2006.12759.

Price, R. H. M., Graham, C., Ramalingam, S. (2019) Association between viral seasonality and meteorological factors. Scientific Reports 9: 929. https://doi.org/10.1038/s41598-018-37481-y.

Pscheidt, V. M., Gregianini, T. S., Martins, L. G., Veiga, A. B. G. (2020) Epidemiology of human adenovirus associated with respiratory infection in southern Brazil. Reviews in Medical Virology:e2189. doi: https://doi.org/10.1002/rmv.2189

Roncalli, A., Lacerda, J. S. (2020) Jornalismo como forma de conhecimento: a questão da divergência dos dados de tendência da covid-19 divulgados pelo consórcio de imprensa e pela SESAP-RN. Scielo preprints. https://doi.org/10.1590/SciELOPreprints.1141

Veiga, A. B. G., Martins, L. G., Riediger, I., Mazetto, A., Debur, M. C., Gregianini, T. S. (2020) More than just a common cold: Endemic coronaviruses OC43, HKU1, NL63, and 229E associated with severe acute respiratory infection and fatality cases among healthy adults. Journal of Medical Virology; 1-6. doi: 10.1002/jmv.26362.

APÊNDICE C – Capítulo do e-Book do Núcleo de Pesquisa Espacial da UFCSPA

Influência do Tráfego Aéreo na Disseminação de Vírus Respiratórios

Amauri Duarte da Silva - amauriduartes@gmail.com

Ana Beatriz Gorini da Veiga - anabgv@ufcspa.edu.br

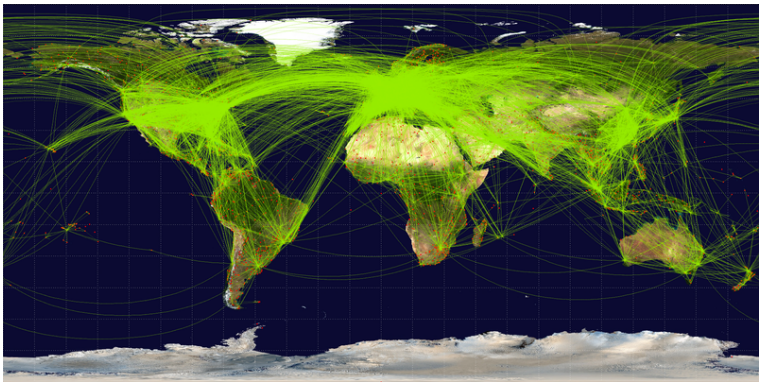
Marcelo Ferreira da Costa Gomes - marfcg@gmail.com

Resumo

A contaminação por vírus respiratórios, dentro das cabines dos aviões, é, numericamente, pouco significativa. Em contrapartida, a disseminação de vírus por passageiros, oriundos de diversos locais do mundo, após o desembarque, tem grande impacto no crescimento de surtos, epidemias e pandemias causadas por doenças respiratórias, pois a quantidade de passageiros circulantes entre países e internamente é muito grande. Assim, voos domésticos e internacionais tornam-se um dos principais agentes na disseminação de vírus entre a população, mas é possível prever a ocorrência de surtos, epidemias e pandemias causadas por vírus respiratórios, como o Sars-CoV-2 ou Influenza A (H1N1), e dessa forma adotar medidas de contenção como a redução de voos comerciais internacionais.

Na segunda metade do século XX, o aumento das viagens por via aérea transformou mundialmente o quadro epidemiológico. Este tipo de viagem faz com que uma pessoa infectada possa percorrer longas distâncias em um curto espaço de tempo, sendo que este tempo é menor do que o próprio período de incubação do vírus e isto facilita a sua disseminação. Dessa forma, os impactos epidemiológicos e a sua gestão, a nível global, tiveram que ser repensados pelas autoridades [1]. A figura 1 ilustra o volume de rotas mundiais, dando uma ideia da possível disseminação de um vírus através dessas rotas.

Figura 1 – Mapa do tráfego regular de companhias aéreas em todo o mundo, por volta de junho de 2009.



Fonte: Jpatokal, CC BY-SA 3.0, via Wikimedia Commons (2021)

A principal preocupação em relação ao deslocamento aéreo não está na contaminação de cabine, pois com base em uma série de estudos da qualidade do ar de cabine, publicados durante 15 anos anteriores ao ano de 2004, poucos casos de suspeita de SARS (*severe acute respiratory syndrome*) podem ter sido transmitidos a bordo (possibilidade de apenas em quatro voos).

Percebeu-se que a transmissão em voos, se ocorrer, é devido ao contato pessoa a pessoa, ou seja, proximidade e não devido à dispersão de organismos através dos sistemas de ventilação das aeronaves, tendo em vista que esses sistemas têm alta eficiência, obtida através dos filtros HEPA (*High Eficiency Particulate Air*) [2]. Ainda, conforme um estudo do MIT (Massachusetts Institute of Technology) [3], somente um a cada 4.300 passageiros, em um voo com duas horas de duração, com todos os assentos ocupados, teria chances de contrair COVID-19 de seu vizinho. Esta probabilidade diminui para um a cada 7.000 passageiros, com espaçamento de um assento entre eles.

Há um volume enorme de pessoas sendo transportadas diariamente por via aérea. Os dados de disponibilidade de assentos, englobando todas as companhias aéreas do mundo, coletados pelo OAG (Official Aviation Guide) [4], podem ser vistos na Tabela 1, onde há uma comparação entre a quantidade de rotas, assentos disponíveis, número de voos e operadoras para uma semana dos anos de 2021 e 2020. Destacam-se, além do enorme volume de assentos disponíveis, também a quantidade de rotas oferecidas e o aumento da disponibilidade em 2021.

Tabela 1 - Disponibilidade mundial de assentos e rotas aéreas

Semana 33	Assentos	Voos	Operadoras	Rotas
09-08-2021	76.000.000	521.000	691	22.640
10-08-2020	59.800.000	413.000	644	18.930

Fonte: OAG. (2021)

Em um estudo realizado na China [5], buscando a associação entre voos internacionais e domésticos com o coronavírus, foi encontrada uma forte correlação linear entre casos de COVID-19 domésticos e o volume de passageiros para regiões dentro da China (R^2 = 0,92, p = 0,19) e uma correlação, maior ainda, entre os casos internacionais de COVID-19 e o volume de passageiros (R^2 = 0,98, p < 0,01). O estudo mostrou que as viagens aéreas são um grande facilitador na distribuição internacional dos casos de COVID-19. Já a disseminação doméstica do vírus sofre influência de outros meios de transporte, como trens e ônibus, e por isso a correlação com voos domésticos é menor.

O México é outro país onde o espalhamento de vírus pela malha aérea foi estudado entre março e abril de 2009. Khan e colegas [7] analisaram a disseminação do vírus de influenza A (H1N1), a partir de voos com origem no México[7]. A análise feita por esse estudo mostrou que um total de 2,35 milhões de passageiros voaram do México para 1.018 cidades em 164 países do mundo. A comparação entre viajantes que partiram do México com importações confirmadas de H1N1, associadas a viagens para o México, tiveram um alto grau de correlação. Considerando os 20 países com os maiores volumes de passageiros vindos do México, 16 países confirmaram importações associadas a viagens para este país. A partir de uma curva ROC (Receiver Operating Characteristic Curve), traçando a relação entre os fluxos de tráfego aéreo internacional e a importação do H1N1, observou-se que os países que recebem mais de 1400 passageiros oriundos do México, corriam um alto risco para importação de H1N1 (área sob a curva = 0,97). A correlação feita entre movimentos internacionais de viajantes e o H1N1 geralmente é intuitiva, mas segundo os autores os resultados do estudo sugerem que uma análise quantitativa de padrões de tráfego aéreo internacional pode antecipar os riscos de importação de doenças infecciosas globais.

Conclusão

[7] Como ainda não havia dados dos voos de 2009 disponíveis, os autores utilizaram os voos do mesmo período do ano de 2008 como *proxy* para 2009.

Além dos estudos citados aqui, existem muitos outros que apontam para o impacto de voos comerciais no crescimento de um surto, epidemia ou pandemia, através da disseminação de vírus. Como as medidas restritivas a voos não têm um impacto imediato, há uma necessidade de ferramentas para a predição do aumento de casos de doenças respiratórias causadas por vírus, de forma que as autoridades possam antecipar suas medidas de contenção e, entre elas, a redução da circulação de voos ou do número de passageiros transportados. Dados de voos comerciais, associados a dados de síndrome respiratória aguda grave, podem ser utilizados como variáveis independentes em modelos preditivos, melhorando a qualidade desses modelos e do acerto de suas predições.

Referências Bibliográficas

1 Tatem AJ, Huang Z, Das A, Qi Q, Roth J, Qiu Y. Air travel and vector-borne disease movement. Parasitology. 2012; 139(14), 1816-1830.

2 Aerospace Medical Association Task Force. Emerging infectious diseases including severe acute respiratory syndrome (SARS): Guidelines for commercial air travel and air medical transport. Space. 2004; 75, 85-6.

3 Barnett A, Fleming K. Covid-19 risk among airline passengers: Should the middle seat stay empty?. MedRxiv. 2020.

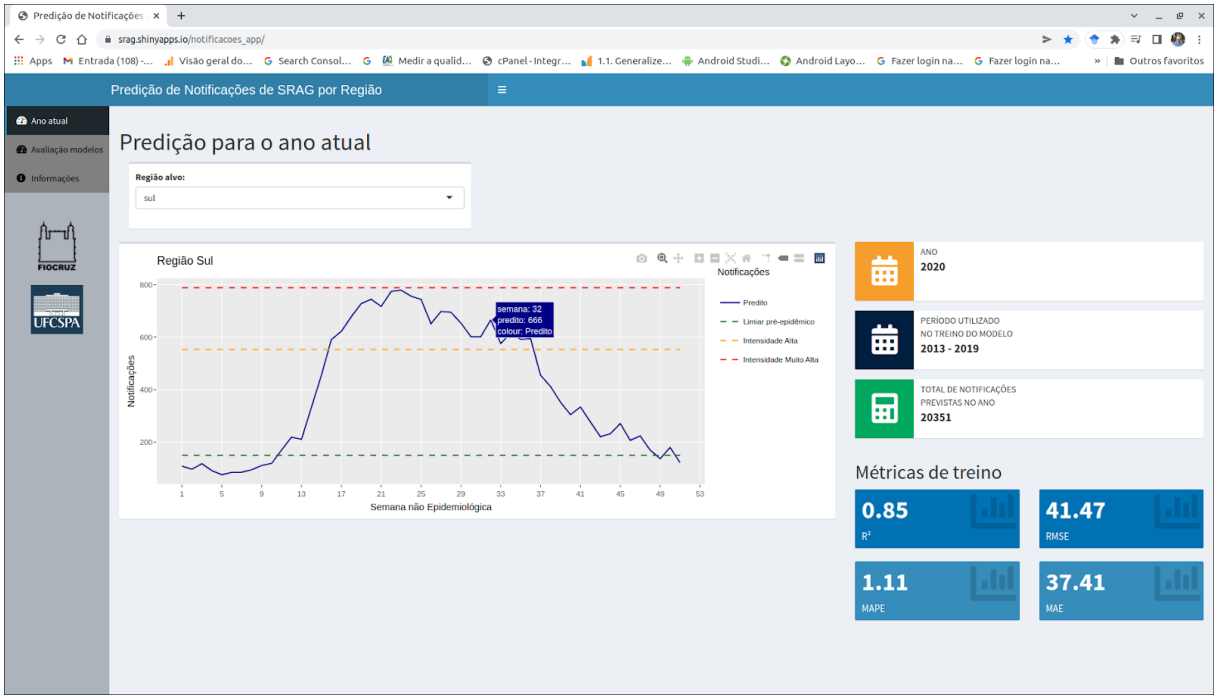
4 Official Aviation Guide - OAG. [acesso em 2021 Aug 24]. Disponível em: https://www.oag.com/

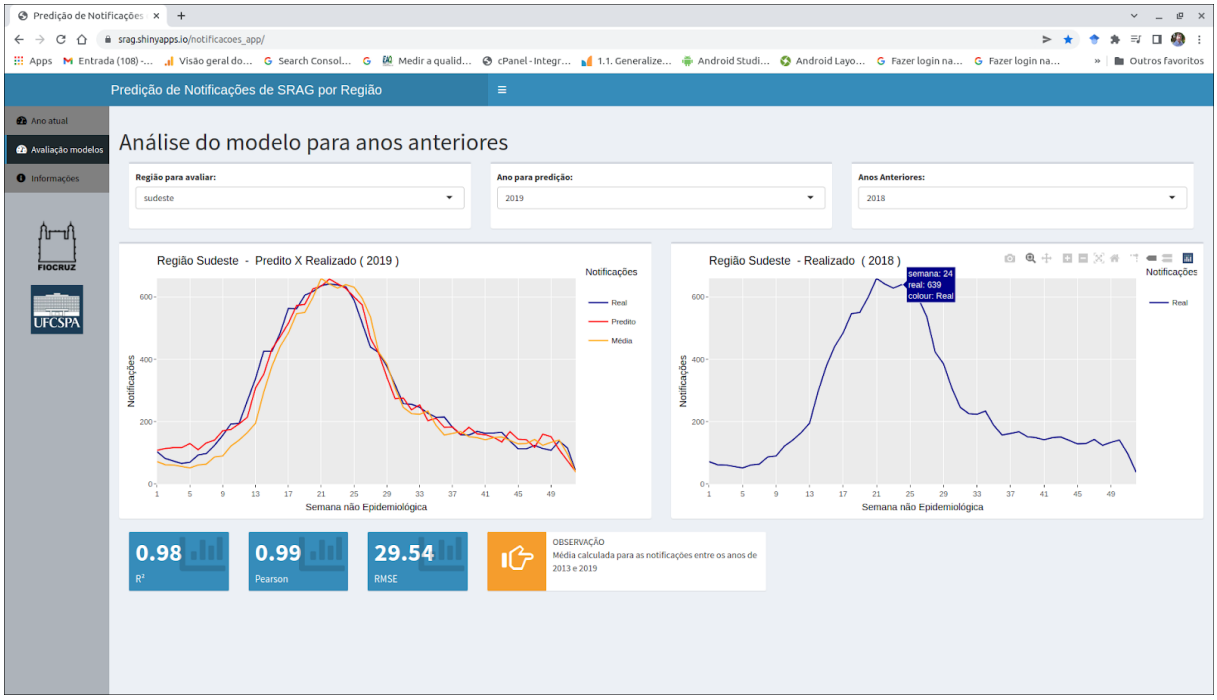
5 Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Ichii H, Zacharski M, et al. The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak. Journal of Microbiology, Immunology and Infection. 2020; 53(3), 467-472.

6 Suzumura T, Kanezashi H, Dholakia M, Ishii E, Napagao SA, Pérez-Arnal R, et al. The impact of COVID-19 on flight networks. In 2020 IEEE International Conference on Big Data (Big Data). 2020, December; 2443-2452.

7 Khan K, Arino J, Hu W, Raposo P, Sears J, Calderon F, et al. Spread of a novel influenza A (H1N1) virus via global airline transportation. New England journal of medicine. 2009; 361(2), 212-214.

APÊNDICE D – Telas da Ferramenta Web para a Visualização dos Modelos Preditivos





ANEXOS

ANEXO A – Registro do projeto no Comitê de Pesquisa (ComPesq) da UFCSPA

Universidade Federal de Ciências da Saúde de Porto Alegre – UFCSPA

Pró-Reitoria de Pesquisa e Pós-Graduação
Comissão de Pesquisa

ATESTADO DE REGISTRO DE PROJETO DE PESQUISA

Atestamos que o projeto de pesquisa intitulado 'Vigilância Epidemiológica: Uso de Aprendizado de Máquina na Predição de Surtos de Síndrome Respiratória Aguda Grave' foi registrado na Comissão de Pesquisa da UFCSPA com o número 100/2021, sob responsabilidade de Ana Beatriz Gorini da Veiga, na forma como foi documentado neste processo 23103.219611/2021-18 até a emissão deste atestado.

O início deste projeto foi marcado em 01 de abril de 2020 e o seu término está previsto para 01 de abril de 2022.

Salientamos que este registro não autoriza o pesquisador a coletar ou analisar dados oriundos de sujeitos de pesquisa. Salientamos também que este registro não garante a concessão de recursos financeiros por parte da UFCSPA a este projeto de pesquisa.

Comissão de Pesquisa (ComPesq)
UFCSPA

Documento assinado eletronicamente por **Renata Padilha Guedes**, **Coordenadora de Pesquisa**, em 17/12/2021, às 12:54, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://sei.ufcspa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1293289** e o código CRC **69A38FD4**.

ANEXO B – Normas da revista Eurosurveillance

(Link para o documento: https://www.eurosurveillance.org/for-authors)

For authors

Eurosurveillance is a weekly electronic publication. It is published online every Thursday (with the exception of e-alerts). The current issue's table of contents is sent by email to all subscribers who sign up to receive the weekly alert. Special issues and topical compilations of selected material from the online issues may be published in printed form with a limited number of copies.

All *Eurosurveillance* content is open access, free of charge for both readers and authors.

- How to submit
- Article types
- Quick reference Table
- Other material
- Corrections/errata
- Supplements
- Tables and Figures
- References
- Authorship
- Ethical issues
- Overlapping publications
- GISAID sequences

How to submit material

All submissions should be sent through the *Eurosurveillance* online submission system. Once you have created an account and logged in you will have access to an online author tutorial to aid you if you have any difficulties during the submission process.

The journal allows pre- and postprint archiving.

Authors who choose to deposit their manuscript in a preprint server should preferably do so ahead or at the point of submission. Authors should consider a preprint server that offers a licence that is compatible with the *Eurosurveillance* CC-BY licence. A directory of available preprint servers and their policies can be found for example on the Sherpa Romeo and ASAPbio websites.

The selected preprint server should clearly state that deposited manuscripts are not peer-reviewed and should not be used for decision making. It should have dedicated sections listing authors' potential conflicts of interest and it should adhere to accepted standards with respect to patient consent and confidentiality and required reviews by ethics committees.

Authors must inform the editors at submission that a preprinted version of their manuscript has been deposited, name the respective preprint server, and provide the DOI for the preprint. During the revision process, the authors are responsible for amending their preprint and pointing readers to subsequent versions. Importantly, they should ensure there is a link from the preprint to the final published article in *Eurosurveillance* and that the full citation including the DOI is given.

All published articles are automatically deposited in PubMedCentral.

In the submission system you will have to provide the following information:

- the article category;
- a declaration that the material is original and has not been submitted elsewhere;
- a declaration that all authors have seen and approved the final manuscript;
- a declaration that the corresponding author, on behalf of all co-authors, has read and agreed to the terms of the *Eurosurveillance* data protection notice;
- a declaration that informed consent has been obtained from persons whose details are described in articles (or from the persons' guardians) that this information may be published;
- a statement on funding and potential competing interests of the authors;
- where appropriate, information on approval of the work by an ethics committee;
- proof of permission to use figures or tables that are adapted or reproduced from other publications.

The following material should be uploaded as separate files:

- a Title page (Word format) including the title, authors (please ensure that the first name is given in full), affiliations (including equal contributions if applicable), corresponding author information, abstract (use the heading 'Abstract' on the previous line); this document will be used to populate the relevant fields;
- a covering letter;
- an anonymised manuscript text in Word format (pdf files cannot be evaluated) without line numbers and with a minimum of eight references for rapid communications and 15 for regular articles. All author-identifiable information – authors' names, affiliations and contributions, as well as any acknowledgements – should NOT be included in the document;
- if a collective author is included (e.g. a working group or disease-specific network) and if the persons comprising the group are to be included at the end of the article, please list them in the relevant field; the contribution of the collective author should be stated in the Authors' contributions section which will be published at the end of the article;
- all figures in an appropriate format (see details in the section Figure formatting); The following file formats are automatically converted into the PDF: Word, RTF, TXT, LaTeX2e, AMSTex, TIFF, GIF, JPEG, EPS, Postscript, PICT, PDF, Excel, and PowerPoint. Other file types are not automatically supported, but can be included as hyperlink items in the PDF file;
- a scan of the agreement with authors signed by the corresponding author on behalf of all authors.

File names of uploaded files must not contain any author-identifiable information that may lead to identification of the author.

After all files and information have been uploaded in the submission system, the corresponding author is responsible for checking and approving the pdf. Approval of the pdf is required for the article to be sent to the editorial office.

Submissions should conform to the *Recommendation for the Conduct, Reporting, Editing and Publication of Scholarly work in Medical Journals*, detailed by the International Committee of Medical Journal Editors.

Articles should be written in clear, appropriate and scientific language that is free of jargon. Avoid abbreviations when possible and define them when you first use them. Please use United Kingdom English spelling.

Titles: should be interesting, informative, accurate and as short as possible. They should contain the place/country and the time period covered in the paper.

Keywords: A maximum of eight keywords suitable for indexing should be provided. Please select from the list provided and add others if needed. Separate the keywords by a semi-colon (;).

Main text: An introduction should put the topic into perspective using up-to-date references, and clearly state the objective of the work and its relevance. The relevant methods should be presented at an appropriate level of detail; molecular diagnostic techniques that are published or standard, for example, can be named and referenced and do not need to be described in detail. It is important to make it clear at all times which results are the work of the authors, presented as part of the article, and which are already known and given as background or for comparison. If tables and figures are provided, the text should shortly describe and summarise the content, but not unnecessarily repeat the information; the reader should be able to understand text and illustrations independently of each other. The European and international relevance should be discussed with relevant references, and where appropriate, lessons learnt and recommendations for the future should be presented.

Article types

Summary table for quick reference

Rapid communications

Rapid communications are timely, authoritative short reports on important communicable disease findings and events where rapid dissemination of information could potentially lead to a prompt change in an ongoing public health situation or create awareness for topics we consider to be of timely relevance. These articles are usually published within two weeks from submission, but when necessary, publication can be arranged within hours of submission (see e-alerts). They undergo rapid independent peer review by at least one expert in the field.

To allow for such rapid processing, these articles are short, usually around 1,200 words, and have a minimum of eight and up to about 20 references and four illustrations (figures or tables). The abstract should not exceed 80 words.

Rapid communications should not have an IMRaD (Introduction, Methods, Results and Discussion) structure. Subheadings as appropriate for the content should be used instead and rapid communications should start with a brief description of the current event (a few lines only) and the aim of the study. Necessary background information should be integrated with the discussion where authors put their data into context with other data for their country and Europe, and other relevant literature.

E-alerts are published ad hoc to disseminate information about an important event that should not wait until the next regular publication of *Eurosurveillance*.

Regular articles

Research articles provide original results from studies on any aspect of communicable disease epidemiology, prevention and control. These papers should include new data or insights of public health importance and consist of ca 3,500 words, a minimum of 15 and up to around 30 references and six illustrations (figures or tables). We require authors to follow the CONSORT guidelines for reporting randomised controlled trials and to include the respective filled in CONSORT checklist in the submission. For economic evaluations of health interventions, we require authors to follow the CHEERS guidelines and to include the respective filled in CHEERS checklist in the submission. Research articles should usually follow the IMRaD structure (Introduction, Methods, Results and Discussion) and have a structured abstract that should not exceed 250 words.

Surveillance articles should focus on the analysis and interpretation of epidemiological trends regarding a specific disease, pathogen or health event based on data from a national or international surveillance system. They could also present the evaluation of a surveillance system or the establishment of a new system. These articles may present data from a (regional) national or international surveillance and should contain a meaningful analysis of specific aspects of the data and put these in a wider context. Surveillance articles should have an IMRaD (Introduction, Methods, Results and Discussion) structure and a structured abstract (Background, Aim, Methods, Results and Conclusion).

The **Introduction** should describe the public health importance of the disease/pathogen/health event, the rationale for the surveillance and the purpose and the objectives of the study. The **Methods** should provide a detailed description of the system (can also be brief if described elsewhere and referenced), applied case definitions, data collection (voluntary/mandatory), processing and validation. An ethical statement should be included. The **Results** should be presented in a similar sequence as the methods. Data should span several years to assess trends; however, the time span could be shorter for emerging diseases or new systems. Data should be analysed by age and sex, when relevant. The **Discussion** should put findings into (European) perspective and highlight lessons learnt and good practices. Authors should discuss the impact that implemented prevention and control measures (or the lack of such measures) have had on trends and the possible need for resource allocation. For description of surveillance systems, results should be interpreted taking into account key attributes such as stability, sensitivity, representativeness and usefulness. Limitations and data quality should also be discussed.

Surveillance articles can be up to ca 3,500 words long. They can have a minimum of 15 and up to ca 30 references, and a maximum of six illustrations (figures or tables). The abstract should not exceed 250 words.

Outbreak reports on national or international outbreaks should be submitted once fully investigated and focus on new or unexpected aspects and on lessons learnt. They should have the following structure: A **Background** that describes very briefly the pathogen and its transmission, puts the outbreak into context in terms of incidence in Europe and/or in the country or region. The following **Outbreak detection** passage should describe details of the signal followed by what is presented in the report. The **Methods** should present the data source(s), case definition, epidemiological (case-control study etc.) and environmental/trace-back investigations as well as microbiological investigations. Data should be analysed by sex and where applicable gender. **Results** should be presented in a similar sequence as methods and an epicurve and – where applicable, a graphical timeline for the sequence of events should be added. A specific **Outbreak control measures** section should present measures taken and highlight the difficulties encountered and/or the successes following their implementation. The **Discussion** should put the findings into (European) perspective and highlight lessons learnt.

Where not essential for understanding of the content, locations will be anonymised so as to prevent identifiability of individuals.

The length of the outbreak reports is up to ca 3,500 words, with a minimum of 15 and up to around 30 references and six illustrations (figures or tables) and the abstract should not exceed 200 words. We encourage authors to follow the STROBE guidelines that were set up for observational studies and to include the respective filled in STROBE checklist in the submission.

Euroroundups should provide an analysis of a contemporary specific aspect or function of communicable disease surveillance, prevention or control in at least five European countries, and present an in-depth comparison of systems and/or data. The average length of these articles is 3,500 words with a minimum of 15 and up to around 30 references and six illustrations (figures or tables). The abstract should not exceed 200 words.

Review articles provide a comprehensive state-of-the-art overview of issues of major public health importance within the field of communicable disease surveillance, prevention or control. They usually are about 4,000 words in length, and contain up to 80 references and six illustrations (figures or tables). All review articles should explain the search strategy and selection criteria, justify the inclusion/exclusion of material and state the sources. Review articles should have a structured abstract that should not exceed 250 words.

For systematic reviews, we require authors to follow the PRISMA guidelines and to include the respective filled in PRISMA checklist in the submission.

Perspectives provide an insightful analysis of practices, policies and guidance on communicable disease prevention and control, as well as guidance on developments in the field of vaccines and immunisation. These articles have an average length of ca 2,000 words, and contain a minimum of 10 and up to 20 references and four illustrations (figures or tables). They do not follow an IMRaD structure and are structured by meaningful subheadings. The abstract should not exceed 200 words.

Summary table for quick reference

For more detail see Article types

What article type is appropriate for my content?	
Rapid communication	Timely data that could lead to prompt change in PH practice
Research article	Original results from studies
Surveillance article	Epidemiological trends based on data from a national or international surveillance system
Outbreak report	National or international outbreaks with new or unexpected aspects/lessons learnt, once fully investigated
Euroroundup	Joint/comparative analysis from at least five European countries
Review	State-of-the-art overview of issues of major public health importance, providing added value over previous reviews on the topic
Perspective	Insightful analysis of practices, policies and guidance
Meeting report	Contact the editorial team before submission
Editorial	Only upon invitation by the editorial team
Letter to the editor	Refer to a recent article published in Eurosurveillance
Abstract – formal requirements	
Rapid communication	Unstructured, ≤ 80 words
Research article	Structured, ≤ 250 words

Surveillance article	Structured, ≤ 250 words
Outbreak report	Unstructured, ≤ 200 words
Euroroundup	Unstructured, ≤ 200 words
Review	Structured, ≤ 250 words
Perspective	Unstructured, ≤ 200 words
Text – formal requirements	
Rapid communication	≤ 1,200 words; ≤ 4 illustrations (figures or tables); 8-20 references.No IMRaD. A few lines max introducing the current event; methods are reported together with each result under descriptive subheadings appropriate for the content; Discussion/Conclusion section is required.
Research article	≤ 3,500 words; ≤ 6 illustrations (figures or tables); 15-30 references.IMRaD format. CONSORT and CHEERS guidelines
Surveillance article	≤ 3,500 words; ≤ 6 illustrations (figures or tables); 15-30 references.IMRaD format.
Outbreak report	≤ 3,500 words; ≤ 6 illustrations (figures or tables); 15-30 references.No IMRaD. Instead sections Background, Outbreak detection, Methods, Results, Outbreak control measures, Discussion.STROBE guidelines
Euroroundup	≥ 5 countries - please consult previously published Euroroundups for the format.≤ 3,500 words; ≤ 6 illustrations (figures or tables); 15-30 references.
Review	≤ 4,000 words; ≤ 80 references.PRISMA guidelines for systematic reviews

Perspective	≤ 2,000 words; ≤ 4 illustrations (figures or tables); 10-20 references.No IMRaD. Descriptive subheadings appropriate to the content.
Meeting report	≤ 2,000 words; 10 references.
Editorial	≤ 1,500 words; ≤ 4 illustrations (figures or tables); ≤ 20 references.Max 2 authors.
Letter to the editor	≤ 600 words; ≤ 8 references.No original data.≤ 12 weeks since original article.
Figures and Tables	
Tables	In MS Word inserted in the main manuscript
Figures	Editable (vectorised) files (for details see Tables and Figures)
Elements of the title page at submission	
.	Manuscript title
.	Author names (including full first names)
.	ORCID (if available)
.	Affiliations (including equal contributions if applicable)
.	Corresponding author information
.	Abstract
.	Keywords
.	Conflict of interest statement
.	Funding statement
Other information to be included in all articles	
Ethical statement	Information on ethical approval of the study, including the registration number of the relevant ethics board, or a statement explaining why such approval was not necessary
Authorship	Follows ICMJE authorship criteria

Contributions	To be detailed for each individual author
Informed consent	Even if consent has been obtained, identifiable information about individuals should be kept to the necessary minimum
Acknowledgements	If applicable
References	In Vancouver style

Other material

The following content types are not peer-reviewed. However, we may consult an expert for advice on the content of such items.

Editorials are written by experts invited to comment on articles and special topics covered by *Eurosurveillance* and usually have a maximum of two authors. Editorials are usually 1,500 words long and contain a maximum of 20 references and four illustrations (figures or tables). As the editorial reflects the personal opinion of the author, the sections Funding information and Authors' contributions are not required. A Conflict of interest statement should be included.

Letters to the editor comment on recent *Eurosurveillance* articles and should be submitted within 12 weeks after the publication of the article in question. They are intended to stimulate scientific discussion and are not a format for the publication of original data. Their average length is 600 words, with five or fewer references. As Letters to the editor do not contain original data, a section with an Ethical statement is not required. A section on Authors' contributions and a Conflict of interest statement should be included. Only one letter may be submitted by any single author or group of authors on any one published paper.

Meeting reports should provide a synopsis of the content of the presentations and have up to 2,000 words, 10 references (including, when possible, links to full reports of conference activities) and no illustrations. Before submitting a meeting report, please contact the editorial team.

Addenda provide new information after publication of an article in cases where such information is relevant but has no bearing on the study outcome and conclusions. Moreover, the information should neither have been available at the time of publication of the article nor justify the publication of a second full article. Addenda should be very short and may contain maximum one illustration.

Example

Corrections/errata

The editorial team should be informed immediately of any errata or corrections to be made. Such changes are made immediately in the original article as well as the pdf, together with an editorial note explaining the nature and date of the change.

Supplementary material

Materials that are not essential for the reader to understand the methodology and results of an article, and to follow the logic of the text, can be submitted as supplementary materials. They are intended to host **additional** information that is not really part of the main study results. Their content is not edited or quality-checked, therefore any data on which you base your discussion and conclusions should be presented in the main manuscript.

Such materials may, for example, comprise questionnaires, search strategies for systematic reviews, additional pictures/videos of a time series where only some time points are in the main paper, models or other very detailed experimental information, sensitivity analyses etc. that help increase transparency and reproducibility of presented findings. Large sets of raw data are generally not considered supplementary materials and these should be deposited preferably in publicly funded open access repositories that would allow citing them.

Supplementary material should not be used for additional discussion, analysis, or interpretations of the findings in the article. We wish to emphasise that a supplement is not a means to circumvent restrictions on the number of figures in a manuscript. Rather these restrictions should guide the author to find a focus of what their main message is.

Supplementary materials will be made available on the *Eurosurveillance* website alongside the article, on behalf of the authors who remain responsible for the accuracy of the content. If any of the supplemental material has been previously published, the authors are responsible for obtaining the required permissions and attributing the source material. The supplementary materials will not be edited.

Authors may submit PDFs, Excel files, images and audio-visual materials. Pdf format is preferred. If possible, all material should be combined in one file. They should be referred to in the manuscript text as Supplementary Table S1, S2.., Supplementary Figure S1, S2..., or Supplement S1, S2..., as appropriate for its content.

The supplement files themselves should be headed with a short descriptive title and contain the following disclaimer:

> "This supplementary material is hosted by *Eurosurveillance* as supporting information alongside the article [Title], on behalf of the authors, who remain responsible for the accuracy and appropriateness of the content. The same standards for ethics, copyright, attributions and permissions as for the article apply. Supplements are not edited by *Eurosurveillance* and the journal is not responsible for the maintenance of any links or email addresses provided therein."

Please note: the authors' names should not be displayed upon submission because *Eurosurveillance* uses double-blind review and reviewers will have access to the supplementary material. Supplementary materials should be submitted in English unless otherwise agreed with the editorial team. Authors should format the text and tables etc. using the following instructions: Supplement style and formats. The length of supplementary materials should be reasonable and generally not exceed 20 pages in total. The supplementary materials will not undergo formal peer review. They will be made available to reviewers as supporting information and the editors expect authors to take possible reviewer comments into account where appropriate. Models submitted as supplementary materials will undergo an initial check by board members with respective expertise before they are sent together with the manuscript to the peer-reviewers.

Tables and Figures

Table formatting

Tables must be created in Word. The full table (title, table, notes) should be inserted in the manuscript directly after the first paragraph in which it is mentioned. As tables must be editable, images are not acceptable.

Numbers and percentages should be split into separate columns. To aid readability in both the online and .pdf versions of the article, portrait-oriented tables are preferred whenever possible.

Figure formatting

A figure's title and notes should be inserted in the manuscript directly after the first paragraph in which it is mentioned. Figure files should not be inserted in the manuscript, but should be uploaded as separate, editable files.

Figures should always be provided as vector files (.pdf, .eps, .wmf, .emf, .svg) and should not include bitmap elements (i.e. a map as a picture in the background). Bitmap files (.jpg, .bmp, .gif, etc) cannot be used because they cannot be edited and are not linked to the original data.

For .pdf files, be sure to export the figure as a pdf directly from the programme in which it was created; do not export the figure in another format and then save it as a pdf or it will not be a vector file.

Graphs should be provided in Excel format, whenever possible. If this is not possible, please follow the general guidelines for figure file types.

Photographs should be given as high-resolution bitmap files (.jpg, .tif, etc.). These should be provided as stand-alone original files, not within Word or PowerPoint documents.

Pie charts are generally not used in *Eurosurveillance*. Unless otherwise agreed with the editor, please choose a different type of graph.

Titles, notes and references

In general, table/figure titles should be short and follow the format: disease, stratification, town, country, date (n = #). Abbreviations should be avoided unless necessary in particularly long titles.

Any additional information needed to understand a table/figure should be given below the table/figure in the following order:

(i) all abbreviations used in the table (in alphabetical order)

(ii) all footnotes used in the table (using a lettered list format, alphabetically by row)

(iii) any further notes

(iv) data source (if applicable)

All tables/figures should be able to stand alone outside the context of the manuscript; the reader should be able to understand the information without referring to explanations in the text.

Any references cited within a table/figure should be numbered chronologically, following the last reference number cited in the manuscript.

Previously published material

Tables and figures that have already been published can only be accepted under specific circumstances. When appropriate, authors are required to obtain permission from the copyright holder to reproduce the material in question; this should be done ahead of submission. The original source of the material needs to be clearly acknowledged and/or referenced. Copyright also needs to be observed, for example, for maps used as a background in figures.

References

Citations are numbered in the order of appearance in the text. Reference numbers are placed in square brackets [1] in the text. References cited in a table or figure legend should be numbered after the citations in the text.

Papers that are accepted for publication can be cited as forthcoming. Preprints should be cited in journal style and include the DOI. Other material not yet accepted for publication cannot be cited and is instead indicated in parentheses in the main text, either as data not shown, if the information comes from one of the authors, or as personal communication, if the information comes from someone else. Personal communications must include the name of the person and the date the communication took place.

References should be formatted according to the uniform requirements for manuscripts submitted to biomedical journals'(Vancouver style). Do not use italics, bold or underlining.

If there are more than six authors, list the first six authors followed by et al. For example:

1. Rose ME, Huerbin MB, Melick J, Marion DW, Palmer AM, Schiding JK, et al. Regulation of interstitial excitatory amino acid concentrations after cortical contusion injury. Brain Res. 2002;935(1-2):40-6.

More samples of reference formats can be seen at:
http://www.nlm.nih.gov/bsd/uniform_requirements.html

Citing preprints

Preprints can be cited in submitted articles, however, the citations should be clearly labelled PREPRINT and we encourage authors to ensure that versions published after submission and before the finalisation of an article are cited where applicable.

Authors should not cite their own preprint version in their submitted article, however, they should mention it in the acknowledgments section of the final article.

Authors and acknowledgements

All listed authors should have made substantive intellectual contributions to the article, be aware of its submission to *Eurosurveillance* and able to account for its content. All authors (this includes members of a collectove group author) must fulfil all four ICMJE authorship criteria.

Clearly identify and provide the email address for the corresponding author.

The article needs to detail, for all authors, the individual contributions, Conflicts of interest in relation to this article.

The names of people who have made contributions that are not sufficient to fulfil all authorship criteria can be mentioned in the acknowledgements. You may acknowledge anyone who has helped you with any aspect of the report, but it is always the corresponding author's responsibility to obtain permission from anyone being acknowledged.

Because of Crossref restrictions, we can accommodate a maximum number of 5 affiliations to a single author.

In our submission system, authors can also include an ORCID (Open Researcher and Contributor ID), if they have one. We encourage authors to use this system.

It is possible to provide a collective name as an author (working group, disease-specific network, etc.). The members of such a group can be listed at the end of the article and will appear in PubMed/MEDLINE indexation.

A statement on funding for the work described in the manuscript should be included.

Any changes to the author list after submission must approved by every author and discussed with the editorial team. Please consult our Editorial policy for the formal procedure.

Ethical issues

All articles should contain a section stating whether ethical approval was obtained for the study and if not, explain why this was not necessary.

For articles reporting outcomes from studies involving humans and related data, *Eurosurveillance* requests authors to declare that the planning conduct and reporting of studies was in line with the Declaration of Helsinki, as revised in 2013.

Approval to conduct the study should be obtained from an independent local, regional or national review body (e.g. ethics committee, institutional review board). The name of the board and the number/ID of the approval(s) should be given.

The *Eurosurveillance* editors reserve the right to judge whether the conduct of the study/research was appropriate.

Informed consent should be obtained for individuals who may be identifiable in submitted manuscripts. The patient consent should be archived with the authors. Non-essential details leading to identification of individuals should be omitted from

manuscripts. In exceptional circumstances, such as during a public health emergency, other legal basis than consent for processing personal data might apply. Such circumstances need prior discussion with the journal.

For studies reporting experiments on animals, authors are requested to state whether institutional and national standards for the care and use of animals were followed and that the study has been approved by an ethical review committee. Guidance on animal research ethics is available from the International Association of Veterinary Editors' Consensus Author Guidelines on Animal Ethics and Welfare.

Overlapping publications

Eurosurveillance publishes original material; however, in certain circumstances, the journal may allow authors to submit material that has previously been published or for which publication is forthcoming, if such dissemination would be beneficial from a scientific or public health perspective, such as reaching a wider readership that cannot read the primary publication or has limited access to it. For example, this could be material previously published in a language other than English. Such overlapping publication will mainly be considered for short articles on topics that have been previously covered in the bulletins or websites of the national public health institutes. *Eurosurveillance* aims to add value to these publications and expects authors to widen the discussion so that it reflects the larger European context and additional references should be included. Clear reference to the previous publication needs to be made in the manuscript and cover letter in order to avoid duplicate publication as defined by the ICMJE.

Through its rapid communications, *Eurosurveillance* has a long-standing track record of disseminating information that may have immediate implications for public health. A longer article that expands on preliminary findings previously covered in a rapid communication might be considered as an individual publication. Furthermore, the editors may accept summary articles based on longer reports or guidelines previously published by public health authorities or similar organisations, in order to make them available in English or support the widest possible dissemination of important information to different target audiences.

It should, however, be made clear in the submission that the material has been published elsewhere and, where appropriate, permission from the editor/publisher of the primary publication must be sought and documented in advance. If published, the primary publication should be clearly acknowledged in the manuscript with a reference and, where possible, the web link to the original material.

Prospective authors should consult the guidelines on 'Overlapping Publications' in the *ICMJE Recommendations for the Conduct, reporting, Editing and Publication of Scholarly work in Medical Journals*.

Secondary publications are subject to the normal *Eurosurveillance* review process.

GISAID sequences

Although the database of the Global Initiative on Sharing All Influenza Data (GISAID) is publicly accessible, attention should be paid to correct attribution of the data used. You should acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's EpiFlu database on which the research is based, and to refer to the GISAID website. In addition to an appropriate acknowledgement, we recommend including a table in the Methods section, listing all sequences with the respective background information, unless there is an unmanageable number of them. Examples of how to do this can be found here: Article 1, Article 2

Contacting the editorial team

If you have any questions about *Eurosurveillance*, please contact our editorial team at eurosurveillance@ecdc.europa.eu

ANEXO C – Normas da revista Health Policy and Planning

(Link para o documento: https://academic.oup.com/heapol/pages/General_Instructions#Submission%20process)

Instructions for Authors

Health Policy and Planning will be published online-only from May 2020. All print editions will be discontinued.

Health Policy and Planning improves the design, implementation and evaluation of health policies in low- and middle-income countries through providing a forum for publishing high quality research and original ideas, for an audience of policy and public health researchers and practitioners. *HPP* is published 10 times a year.

HPP has a double-blinded peer-review policy. This means the identity of the authors is anonymous to the reviewers and vice-versa. All types of papers are peer reviewed and all article abstracts from each issue are translated into French, Spanish and Chinese.

Before you submit please make sure you have followed all the relevant instructions. A checklist for authors is available on the *HPP* webpage.

Please submit your paper to the most appropriate section. Read our Section Summaries.

Please note that submission of a paper implies that it reports unpublished work and that it is not under consideration for publication elsewhere. Plagiarism, including duplicate publication of the author's own work, in whole or in part without proper citation is not tolerated by *HPP*. Submitted manuscripts are screened with iThenticate software, as part of the CrossCheck initiative to detect and prevent plagiarism.

- Guidance

Guidance

Improving chances of publication

As well as the high overall quality required for publication in an international journal, authors should take into consideration:

- Addressing *HPP*'s readership: national and international policy makers, practitioners, academics and general readers with a particular interest in health policy issues and debates.
- Manuscripts that fail to set out the international debates to which the paper contributes, and to draw out policy lessons and conclusions, are more likely to be rejected, returned to the authors for redrafting prior to being reviewed, or undergo a slower acceptance process.
- Economists should note that papers accepted for publication in HPP will consider the broad policy implications of an economic analysis rather than focusing primarily on the methodological or theoretical aspects of the study.
- Public health specialists writing about a specific health problem or service should discuss the relevance of the analysis for the broader health system. Those submitting health policy analyses should draw on relevant bodies of theory in their analysis, or justify why they have not, rather than only presenting a narrative based on empirical data.
- Primarily focus on one or more low- or middle-income countries.

The editors cannot enter into correspondence about papers considered unsuitable for publication and their decision is final. Neither the editors nor the publishers accept responsibility for the views of authors expressed in their contributions. The editors reserve the right to make amendments to the papers

submitted although, whenever possible, they will seek the authors' consent to any significant changes made. The manuscript will not be returned to authors following submission unless specifically requested.

Should you require any assistance in submitting your article or have any queries, please do not hesitate to contact the editorial office at hpp.editorialoffice@oup.com.

Should your manuscript require any English language editing, we recommend contacting AuthorAid, a free network which provides free mentoring and English-language editing for researchers in low-and middle-income countries.

Manuscript format and style for all articles

Only articles in English are considered for publication.

The journal follows Oxford SCIMED style. Please refer to these requirements when preparing your manuscript. More information on preparing your manuscript is available. Oxford English spelling style should be used consistently throughout your manuscript. (-ize/-ization), except in quotations and in references.

Prepare your manuscript, including tables, using a word processing program and save it as a *.doc*, *.rtf* or *.ps* file. Use a minimum font size of 11, double-spaced and paginated throughout including references and tables, with margins of at least 2.5 cm. The text should be left justified and not hyphenated.

Authorship: Please note there should be a LMIC (lower-middle-income) named author from the region of the paper included in the paper. If all named authors are from HICs' (high-income countries) please clarify the reason for this on the Title Page. In addition, a contributorship statement is required - please note as follows on the Title Page:

Please clarify how each author has contributed to the paper. You need to create a list assigning a person’s name against the following roles or tasks:

- Conception or design of the work
- Data collection
- Data analysis and interpretation
- Drafting the article
- Critical revision of the article
- Final approval of the version to be submitted - all named authors should approve the paper prior to submission.

The Title Page should be uploaded as a separate file type "Title Page" and contain the following information:

1. Title
2. Corresponding authors name,address,country,e-mail address, ORCID details
3. Each authors affiliation and qualification (BSc, MA, PhD...)
4. Keywords and an abbreviated running title
5. 2-4 Key messages, detailing the main points made in the paper
6. A word count of the full article:Word Limits do not include Abstract, References, Figure/Table legends.
7. Ethical Approval
    - if no ethical approval was required for the research, please note the reason:
        - Example A: Ethical approval for this type of study is not required by our institute.
        - Example B: Ethical approval for this research was waived by the authors institute/s IRB.
    - Ethical Approval Received-Please note the institute/s which approved the research with reference number.

Original Articles-word limit 6000

Review Articles- word limit 10000

Commentaries: Word limit 1200

How to do...or not to do - word limit 3000

Methodological Musings-word limit 3000

Innovation and practice reports: word limit 2000

Funding/Acknowledgements/Conflicts of interest/Ethical approval should be noted on the title page.

In the acknowledgements, all contributors who do not meet the criteria for authorship should be listed. Sources of funding for research must be explicitly stated, including grant numbers if appropriate. Other financial and material support, specifying the nature of the support, should be acknowledged as well.

Figures should be designed using a well-known software package for standard personal computers. If a figure has been published earlier, acknowledge the original source and submit written permission from the copyright holder to reproduce the material.

Please be aware that the requirements for online submission and for reproduction in the journal are different: (i) for online submission and peer review, please upload your figures separately as low-resolution images (.jpg, .tif, .gif or. eps); (ii) for reproduction in the journal, you will be required after acceptance to supply high-resolution .tif files. Minimum resolutions are 300 d.p.i. for colour or tone images, and 600 d.p.i. for line drawings. We advise that you create your high-resolution images first as these can be easily converted into low-resolution images for online submission.

Figures will not be relettered by the publisher. The journal reserves the right to reduce the size of illustrative material. Any photomicrographs, electron micrographs or radiographs must be of high quality. Wherever possible, photographs should fit within the print area or within a column width. Photomicrographs should provide details of staining technique and a scale bar. Patients shown in photographs should have their identity concealed or should have given their written consent to publication. When creating figures, please

make sure any embedded text is large enough to read. Many figures contain miniscule characters such as numbers on a chart or graph. If these characters are not easily readable, they will most likely be illegible in the final version.

Certain image formats such as .jpg and .gif do not have high resolutions, so you may elect to save your figures and insert them as .tif instead.

For useful information on preparing your figures for publication, go to the Digital Art Support page.

All measures should be reported in SI units, followed (where necessary) by the traditional units in parentheses. There are two exceptions: blood pressure should be expressed in mmHg and haemoglobin in g/dl. For general guidance on the International System of Units, and some useful conversion factors, see 'The SI for the Health Professions' (WHO 1977).

Manuscript file must include text body. Title Page, Figures and Tables should be uploaded separately.

Prior Publication Policy

[Based on a statement developed by a group of editors of journals that publish articles on health, health services, and health policy. Journals currently using this statement include: Health Affairs, Health Services Research, Inquiry, Journal of Health Politics, Policy and Law, Journal of Health Services Research & Policy, Medical Care, and the Milbank Quarterly.]

Background

The policy of the journals subscribing to this statement is to consider for publication only original work that has not previously been published. Questions about what constitutes previous publication are arising with increasing frequency because of the growth of electronic publishing and the increasing number of reports and papers being produced by organizations and agencies. This statement provides guidance on this issue.

There are legitimate reasons why research may be disseminated before submission to a journal. Active communication among researchers about preliminary findings or the circulation of draft reports for discussion and critique contributes to the eventual quality of published work. In addition, organizations that support or carry our research have an understandable interest in disseminating their work. From the perspective of journals, these reasons for dissemination must be balanced against two considerations. The first is the value of the peer review process. The rules against prior publication are intended to add some assurance of the credibility of published research.

Papers are often improved during the peer review process, with findings, conclusions, and recommendations sometimes changed in response to reviewers' comments. The public and policymakers might be confused or misled if there were multiple versions of a paper in the public domain. Second, from a more parochial viewpoint, journal space is limited, and much time and expense are involved in the evaluation, publication, and distribution of journal articles. Journals must make difficult choices about what to include; there is less value in publishing papers that have already been disseminated to their target audiences.

We discuss here several types of dissemination and provide guidelines with respect to the prior publication question. This discussion is essentially an elaboration of two rules, the first emphasizing previous dissemination of the material, the second stressing disclosure.

- Rule One: If the material in a paper has already been disseminated to a journal's audience, particularly in a format that appears to be a final product, then it is unlikely that a second version will be worth publishing in the journal.
- Rule Two: It is the responsibility of authors to let editors know at the time of submission whether a paper's contents have been previously disseminated in any manner so that the editors can determine whether to proceed with the review process.

Previous Presentations at Meetings

Presentation of a paper at conferences or seminars usually does not jeopardize the possibility of publication.

Working Papers

Dissemination of "working papers" to a limited audience will not ordinarily jeopardize publication. Working paper series are used by many organizations as a means of enabling researchers to obtain critiques from fellow researchers. Working papers covered by this policy are those that are released by the author or an organization rather than by a publisher, are not advertised to the public, and are marked as drafts that are subject to future revision. HPP will not publish papers for which a similar working paper is already available in the public domain.

Internet Postings

Release via the Internet may jeopardize journal publication under some circumstances. Presentation of the work as a final report is a marker of an attempt to reach a wide audience, particularly when combined with efforts to direct traffic to the work (e.g., via links on other sites) and efforts to attract attention (e.g., press releases). In contrast, if a document is posted on the Internet only to facilitate communication among colleagues with the aim of getting feedback, and if there has been no attempt to otherwise attract the attention of journalists, the public, or the broader research community to the document, then this is unlikely to preclude journal publication.

In general, when posting on the Internet serves similar functions as presentation at professional meetings - facilitating the development of papers and the improvement of the research, influencing future revisions, and not constituting a "finished" product - it would not be considered prior publication. On the other hand, when the Web site posting functions as a virtual version of a conventional publication, which may even be copyrighted by the posting

organization, the benefit of an additional publication in the journal will be scrutinized carefully.

In cases where there has been little to no exposure at the time that a paper has been submitted to the journal, but the circumstances surrounding the posting make it likely that a high level of exposure (press coverage, etc.) might occur, then the author should remove a posting as a condition for further consideration of the manuscript.

Authors who post papers on a Web site and do not want it to constitute prior publication should also post a disclosure statement such as: "This draft paper is intended for review and comments only. It is not intended for citation, quotation, or other use in any form." This statement should be kept on the Web site throughout the review process and until the paper is actually accepted for publication in a journal. Once accepted, authors should post a message to the effect that: "A revised final version of this paper will appear in (Journal Name), volume, issue." Authors also should include this statement as a header or footer on every page of the paper.

Formal Reports from Foundations, Academic Institutions, Institutes, Trade Associations, and Government Agencies

The dissemination efforts of foundations, government agencies, research institutes, and other organizations that support or carry out research can complement publication in peer-reviewed journals. If publication in one of our peer-reviewed journals is desired, organizational publications should be timed to coincide with or follow journal publication, with appropriate copyright permissions having been obtained. This sequence ensures that the peer-review process will have an opportunity to correct deficiencies of method or presentation.

Formal, published reports that have gone through an editorial process, that have been intended to reach a wide audience, and that are publicized and available to any interested party (whether free or not) usually will not be

considered for journal publication. A paper that is based on such a report might be considered for publication if it were sufficiently different in emphasis or intent. In such instances, the author should explain at the time of submission (or before) how the paper differs from the previously released report and why its publication would represent a distinct and important contribution beyond that version.

Policy briefs

If the findings of a piece of research have been published locally (i.e. in a specific country) with the aim of influencing policy debates in that country then even if the brief is available on the web we may consider publishing an article so long as (i) the brief has not had wide circulation outside the country and (ii) the brief is clearly targeted at policy-making audiences, and hence does not include the detailed discussion of methods and perhaps findings that one might expect in a journal article.

Media Publicity

If results reported in a working paper have become widely known as a result of media exposure (or even if the potential for widespread exposure remains during review), and that working paper is readily available to interested readers (e.g., through a Web site), an editorial judgment will be made whether journal publication would be appropriate. Authors can help protect their work from unwanted media exposure by making clear on working drafts, copies presented at conferences, and other versions that it is a draft that has not yet undergone peer review for publication and that findings and conclusions are subject to change. Authors also should request that any "stories" derived from interviews with the media be embargoed until the work is published or released by the publisher (see, for example, Fontanarosa, P.B., and C.D. DeAngelis. 2002. The Importance of the Journal Embargo. Journal of the American Medical Association 288: 748-750). Any accepted manuscript released to the media should contain the statement: "A revised final version of this paper will appear in (Journal Name), volume, issue." Journal policies involving author contact

with members of the media may vary, depending on the issue or journal. Thus, authors should check with the editor before speaking with or distributing papers to members of the media.

Importance of Disclosure

In contrast to the editors' decision whether a certain paper has been disseminated too widely to warrant journal publication, there is very little judgment involved in whether an author should disclose previous dissemination. Prior to, or at the time of, submission of a paper that has been disseminated in any of the ways discussed previously, authors should bring this to the attention of the editor so that a determination can be made before the paper goes into the peer-review process. In so doing, authors should describe in what form and how the work was previously disseminated and how the submitted manuscript differs from previously disseminated versions. Editors might be receptive to a modified version of a paper that has been widely disseminated if the submitted version has a different focus (e.g., more emphasis on methods, more sophisticated analytic approach, or discussion of developments that have transpired since the initial dissemination). The key point is to let editors know about any dissemination that will have, or is likely to have, occurred before the journal article is published rather than have it discovered during or after the review or editorial process. As part of the submittal, authors should include copies of other related papers that might be seen as covering the same material.

Failure to disclose could preclude publication in the journal or, if already published, could result in a notice in the journal about the failure and may result in a retraction of the article.

Manuscript Preparation

Page 1: *Title Page* – as above.

Page 2: *Abstract.* The abstract should be prepared in one paragraph, no headings are required. It should describe the purpose, materials and methods, results,

and conclusion in a single paragraph no longer than 300 words without line feeds.

Page 3: *Introduction.* The Introduction should state the purpose of the investigation and give a short review of the pertinent literature, and be followed by:

Materials and methods. The Materials and methods section should follow the Introduction and should provide enough information to permit repetition of the experimental work. For particular chemicals or equipment, the name and location of the supplier should be given in parentheses.

Results. The Results section should describe the outcome of the study. Data should be presented as concisely as possible, if appropriate in the form of tables or figures, although very large tables should be avoided.

Discussion. The Discussion should be an interpretation of the results and their significance with reference to work by other authors.

Abbreviations. Non-standard abbreviations should be defined at the first occurrence and introduced only where multiple use is made. Authors should not use abbreviations in headings.

All *measures* should be reported in SI units, followed (where necessary) by the traditional units in parentheses. There are two exceptions: blood pressure should be expressed in mmHg and haemoglobin in g/dl. For general guidance on the International System of Units, and some useful conversion factors, see 'The SI for the Health Professions' (WHO 1977).

References. References must follow the Harvard system and must be cited as follows:

Baker and Watts (1993) found...

In an earlier study (Baker and Watts 1993), it...

Where works by more than two authors are cited, only the first author is named followed by 'et al.' and the year. The reference list must be typed double-spaced in alphabetical order and include the full title of both paper (or chapter) and journal (or book), thus:

Baker S, Watts P. 1993. Paper/chapter title in normal script. Journal/book title in italics *Volume number in bold* : page numbers.

Baker S, Watts P. 1993. Chapter title in normal script. In: Smith B (ed). *Book title in italics*. 2nd edn. Place of publication: Publisher's name, page numbers.

Tables All tables should be on separate pages and accompanied by a title - and footnotes where necessary. The tables should be numbered consecutively using Arabic numerals. Units in which results are expressed should be given in parentheses at the top of each column and not repeated in each line of the table. Ditto signs are not used. Avoid overcrowding the tables and the excessive use of words. The format of tables should be in keeping with that normally used by the journal; in particular, vertical lines, coloured text and shading should not be used. Please be certain that the data given in tables are correct. Tables should be provided as Word or Excel files.

Availability of Data and Materials

Where ethically feasible, *Health Policy and Planning* strongly encourages authors to make all data and software code on which the conclusions of the paper rely available to readers. Authors are required to include a Data Availability Statement in their article.

We suggest that data be presented in the main manuscript or additional supporting files, or deposited in a public repository whenever possible. For information on general repositories for all data types, and a list of recommended repositories by subject area, please see Choosing where to archive your data.

Data Availability Statement

The inclusion of a Data Availability Statement is a requirement for articles published in *Health Policy and Planning*. Data Availability Statements provide a standardised format for readers to understand the availability of data underlying the research results described in the article. The statement may refer to original data generated in the course of the study or to third-party data analysed in the article. The statement should describe and provide means of access, where possible, by linking to the data or providing the required unique identifier.

The Data Availability Statement should be included in the endmatter of your article under the heading 'Data availability'.

More information and examples of Data Availability Statements.

Data Citation

Family Practice supports the Force 11 Data Citation Principles and requires that all publicly available datasets be fully referenced in the reference list with an accession number or unique identifier such as a digital object identifier (DOI). Data citations should include the minimum information recommended by DataCite:

- [dataset]* Authors, Year, Title, Publisher (repository or archive name), Identifier

*The inclusion of the [dataset] tag at the beginning of the citation helps us to correctly identify and tag the citation. This tag will be removed from the citation published in the reference list.

Preprint policy

Authors retain the right to make an Author's Original Version (preprint) available through various channels, and this does not prevent submission to the journal. For further information see our Online Licensing, Copyright and Permissions policies. If accepted, the authors are required to update the status

of any preprint, including your published paper's DOI, as described on our Author Self-Archiving policy page.

Types of papers

Health Policy and Planning welcomes submissions of the following article types:

- Original research
- Review articles
- Methodological musings
- Innovation and practice reports
- Commentaries
- 'How to do (or not to do)...' [for example, see Hutton & Baltussen, HPP, 20(4): 252-9] and
- '10 best resources' [for example, see David & Haberlen, HPP, 20(4): 260-3].

Original Research

Manuscripts should preferably be a *maximum* of *6,000* words, excluding tables and figures/diagrams.

The manuscript will generally follow through sections: Title page, Abstract (no more than 300 words), Introduction, Methods, Results, Discussion, Conclusion, Acknowledgements, References. However, it may be appropriate to combine the results and discussion sections in some papers. Tables and Figures should not be placed within the text, rather provided in separate file/s.

For the reporting of statistical analyses please consider the following additional points:

- Focus the statistical analysis at the research question.
- Provide information about participation and missing data.

- As much as possible, describe results using meaningful phrases (e.g., do not say "beta" or "regression coefficient", but "mean change in Y per unit of X"). Provide 95% confidence intervals for estimates.
- Report the proportions as *N* (%), not just %.
- Report *P* values with 2 digits after the decimal, 3 if <0.01 or near 0.05 (e.g., 0.54, 0.03, 0.007, <0.001, 0.048). Do not report *P* values greater than 0.05 as "NS".
- Always include a leading zero before the decimal point (e.g., 0.32 not .32).
- Do not report tests statistics (such as chi-2, T, F, etc.)."

For acknowledgements, figures and measures see above.

Review Articles

Manuscripts should preferably be a *maximum of 10,000 words*, excluding tables, figures/diagrams and references.

Reviews may be invited. They generally address recent advances in health policy, health systems and implementation. *Systematic reviews are particularly welcomed,* but may not be appropriate for every topic. If authors are submitting a review article that is not a systematic review then the paper should explain why a systematic review was not feasible/desirable, and the review methods should be described in a way that is as clear and as replicable as possible.

The manuscript will generally follow through sections: Abstract (no more than 300 words), Introduction, Methods, Results, Discussion, Conclusion, References. However, it may be appropriate to combine the results and discussion sections in some papers. Tables and Figures should not be placed within the text, rather provided in separate file/s.

Checklists have been developed for a number of study designs, including randomized controlled trials (CONSORT), systematic reviews (PRISMA), observational studies (STROBE), diagnostic accuracy studies (STARD) and qualitative studies (COREQ, RATS). We recommend authors refer to the EQUATOR Network website for further information on the available reporting

guidelines for health research, and the MIBBI Portal for prescriptive checklists for reporting biological and biomedical research where applicable. Authors are requested to make use of these when drafting their manuscript and peer reviewers will also be asked to refer to these checklists when evaluating these studies.

Commentaries

Short commentaries on topical issues in health systems are welcomed - *please email the editorial office prior to submission*. Most such commentaries are commissioned by the editors, but the journal will also consider unsolicited submissions. Commentaries should of broad interest to readers of *Health Policy and Planning*, and while they are not research papers, they should be well substantiated. Manuscripts should preferably be a *maximum of 1,200 words*, excluding tables, figures/diagrams and references.

The manuscript will generally contain a short set of key take-home messages. Tables and Figures should not be placed within the text, rather provided in separate file/s.

How To Do...Or Not To Do

This series is meant to explain how to use a particular research or analytical method (e.g. social network analysis, discrete choice experiment etc.). The research or analytical methods discussed should be well accepted and clearly defined: this category of paper is not meant to address methodological debates but rather to help disseminate and promote the use of well-accepted methodologies.

Manuscripts should preferably be a *maximum of 3,000 words* excluding tables, figures/diagrams and references.

- The sections must be arranged as follows: i) Title page, ii) Abstract, iii) Introduction, iv) Body of the paper, and v) References. Main sections should be coordinated by the author, and inserted between Introduction

and Reference sessions. Please contact our office before submitting a manuscript in this category.

Tables and Figures should not be placed within the text, rather provided in separate file/s.

10 Best Resources

This 10 best is a series of articles that identify and outline the 10 most useful resources from a range of sources to help facilitate a better understanding of a particular issue in global health.

We often commission these articles but we also hear unsolicited suggestions.

For acknowledgements, figures and measures see Title page.

Methodological Musings

This series is meant to address methodological issues in health policy and systems research, where there is currently a lack of clarity about accepted research methods. This series is intended to support the development of the health policy and systems research field, through supporting methodological discussion.

Manuscripts should preferably be a *maximum of 3,000 words*, excluding tables, figures/diagrams and references.

- The sections must be arranged as follows: i) Title page, ii) Abstract, iii) Introduction, iv) Body of the paper, and v) References. Main sections should be coordinated by the author, and inserted between Introduction and Reference sessions. Please contact our office before submitting a manuscript in this category.
- For acknowledgements, figures and measures see Title page.

Innovation and Practice Reports

These short reports are narratives and/or reflections/experiences from the perspective of health leaders, managers and practitioners operating at the national or sub-national level which focus on innovative approaches to strengthen health systems. They do not need to report a completely new activity or practice but could consider an adaptation or modification to an existing one. Papers should highlight the experience of health system practitioners in taking action to strengthen health systems through innovative activities. These activities might address governance or human resource management approaches, for example, rather than having a health care focus. Other relevant activities include practices to build capacity, develop new partnerships, new approaches to management, or restructuring relationships within health systems implemented at scale with the intention of promoting changes in practice. The innovations should preferably have been implemented for sufficient time to allow authors to demonstrate their potential system benefits, including sustained improvement over time. We encourage authors to think how the experience they report adds to existing work in their own setting, as well as other settings - but this is not essential.

We will not consider clinical and pharmaceutical innovations and practices.

Manuscripts should be a *maximum* of *2,000 words*. The manuscript will generally present the following sections:

- *Key Messages* (2-4 key messages or lessons for consideration in other settings)
- *Abstract* (no more than 300 words)
    - *Introduction* – outline the background to and context of the activity or practice: what is it and why does it matter in your health system? Please also clarify how you generated the reflections presented in this report: who was involved, what did you do, what forms of evidence are used?

- *Implementation* – how was the activity or practice developed and implemented? was it adapted over time and if so, how and why? how was it scaled up, and to what level was it scaled?
- *Achievements/Challenges* – what benefits have been seen in the health system? at what scale? What challenges were faced?
- *Enablers/Constraints* – what factors enabled implementation, scale-up and achievements, and what factors constrained them?
- *Conclusions*: what are the key lessons for other health leaders in other settings concerning this activity/practice
- *References*

If used, Tables and Figures should not be placed within the text, rather provided in separate file/s. In the main body of the paper, sub-headings may be useful to signal key elements of the experience reported.

Ethics approval

Ethics approval is not required for this type of article, which is intended to allow reflections on innovative approaches to strengthen health systems written by health system leaders and managers working at national or sub-national level.

However, in the introduction, please briefly clarify how you generated the reflections presented in this report including who was involved, what you did and what forms of evidence were used.

If you require writing assistance, please do contact us at

hpp.editorialoffice@oup.com.

Submission process

Pre-submission language editing

Authorship

Originality

Online submission

Pre-Submission Language Editing

HPP asks all authors to ensure that their papers are written in as high a standard of English as possible before submission to the journal. If your first language is not English, to ensure that the academic content of your paper is fully understood by journal editors and reviewers, you may want to consider using a language editing service. Language editing does not guarantee that your manuscript will be accepted for publication. For further information on this service, please click here. Several specialist language editing companies offer similar services and you can also use any of these. Authors are liable for all costs associated with such services. If your first language is not English, to ensure that the academic content of your paper is fully understood by journal editors and reviewers is optional. Language editing does not guarantee that your manuscript will be accepted for publication. Further information on the Language services is available. Several specialist language editing companies offer similar services and you can also use any of these. Authors are liable for all costs associated with such services.

Authorship

All persons designated as authors should qualify for authorship. The order of authorship should be a joint decision of the co-authors. Each author should have participated sufficiently in the work to take public responsibility for the content. Authorship credit should be based on substantial contribution to conception and design, execution, or analysis and interpretation of data. All authors should be involved in drafting the article or revising it critically for important intellectual content, must have read and approved the final version of the manuscript and approve of its submission to this journal. An email confirming submission of a manuscript is sent to all authors. Any change in authorship following initial submission would have to be agreed by all authors as would any change in the order of authors.

Originality

Manuscripts containing original material are accepted for consideration with the understanding that neither the article nor any part of its essential substance, tables, or figures has been or will be published or submitted for publication elsewhere. This restriction does not apply to abstracts or short press reports published in connection with scientific meetings. Copies of any closely related manuscripts should be submitted along with the manuscript that is to be considered by HPP. HPP discourages the submission of more than one article dealing with related aspects of the same study. For further information on the prior publication policy see Prior Publication Policy.

During the online submission procedure, authors are asked to provide:

- information on prior or duplicate publication or submission elsewhere of any part of the work;
- a statement of financial or other relationships that might lead to a conflict of interest or a statement that the authors do not have any conflict of interest;
- a statement that the manuscript has been read and approved by all authors (see also section on authorship);
- name, address, telephone and fax number of the corresponding author who is responsible for negotiations concerning the manuscript;
- copies of any permissions to reproduce already published material, or to use illustrations or report sensitive personal information about identifiable persons.

All papers submitted to HPP are checked by the editorial office for conformance to author and other instructions all specified below. Non-conforming manuscripts will be returned to authors.

If authors are unsure about the originality of their manuscript or any part of it, they should contact the editorial office at hpp.editorialoffice@oup.com.

Online Submission

Prior to submission please carefully read instructions on each type of paper and closely follow instructions on word count, abstract, tables and figures and references. This will ensure that the review and publication of your paper is as efficient and quick as possible. The Editorial Office reserve the right to return manuscripts that are not in accordance with these instructions.

All material to be considered for publication in Health Policy and Planning should be submitted in electronic form via the journal's online submission system. Once you have prepared your manuscript according to the instructions below, instructions on how to submit your manuscript online can be found by clicking Manuscript Submission.

The OUP LaTeX template produces manuscripts matching the formatting requirements of the journals listed here. The template is available on Overleaf.com, at CTAN, and as a direct download by clicking this link.

Conflict of Interest

Authors must declare any conflicts of interest during the online submissions process. The lead author is responsible for confirming with the co-authors whether they also have any conflicts to declare.

Ethical Approval

A requirement of publication is that research involving human subjects was conducted with the ethical approval of the appropriate bodies in the country where the research was conducted and of the ethical approval committees of affiliated research institutions elsewhere. Furthermore, subjects' consent must have been obtained according to the Declaration of Helsinki. A clear statement addressing all these points must be made in any submitted manuscript presenting such research. In original articles, this information must also be included in the methods section of the submitted manuscript. Please note that it is the responsibility of the corresponding author to ensure that the relevant ethical approval described above is provided. The Editors-in-Chief reserve the

right to refuse publication where the required ethical approval/patient consent is lacking, or where the approval/consent provided is deemed incomplete or ambiguous.

Funding

Details of funding sources should be listed in a separate section entitled "Funding". This should appear before the acknowledgements section.

The following rules should be followed:

- The sentence should begin: ‘This work was supported by …’
- The full official funding agency name should be given, i.e. ‘the National Cancer Institute at the National Institutes of Health’ or simply 'National Institutes of Health' not ‘NCI' (one of the 27 subinstitutions) or 'NCI at NIH’ - see the full RIN-approved list of UK funding agencies for details
- Grant numbers should be complete and accurate and provided in brackets as follows: ‘[grant number ABX CDXXXXXX]’
- Multiple grant numbers should be separated by a comma as follows: ‘[grant numbers ABX CDXXXXXX, EFX GHXXXXXX]’
- Agencies should be separated by a semi-colon (plus ‘and’ before the last funding agency)
- Where individuals need to be specified for certain sources of funding the following text should be added after the relevant agency or grant number 'to [author initials]'.

An example is given here: ‘This work was supported by the National Institutes of Health [P50 CA098252 and CA118790 to R.B.S.R.] and the Alcohol & Education Research Council [HFY GR667789].

Oxford Journals will deposit all NIH-funded articles in PubMed Central. See Author self-archiving policy for details. Authors must ensure that manuscripts are clearly indicated as NIH-funded using the guidelines above.

Permissions

Authors are reminded that it is their responsibility to comply with copyright laws. It is essential to ensure that no parts of the submission have or are due to appear in other publications without prior permission from the copyright holder and the original author. Materials, e.g. tables, taken from other sources must be accompanied by a written statement from both author and publisher giving permission to HPP for reproduction.

Copyright

Upon receipt of accepted manuscripts at Oxford Journals authors will be invited to complete an online copyright licence to publish form.

Please note that by submitting an article for publication you confirm that you are the corresponding/submitting author and that Oxford University Press ("OUP") may retain your email address for the purpose of communicating with you about the article. You agree to notify OUP immediately if your details change. If your article is accepted for publication OUP will contact you using the email address you have used in the registration process. Please note that OUP does not retain copies of rejected articles

It is a condition of publication in Health Policy and Planning that authors assign licence to publish to Oxford University Press. This ensures that requests from third parties to reproduce articles are handled efficiently and consistently and will also allow the article to be as widely disseminated as possible. In assigning licence to publish, authors may use their own material in other publications provided that the Journal is acknowledged as the original place of publication, and Oxford University Press is acknowledged as the original Publisher.

Third-Party Content in Open Access Papers

If you will be publishing your paper under an Open Access licence but it contains material for which you do not have Open Access re-use permissions, please state this clearly by supplying the following credit line alongside the material:

Title of content

Author, Original publication, year of original publication, by permission of [rights holder]

This image/content is not covered by the terms of the Creative Commons licence of this publication. For permission to reuse, please contact the rights holder.

Prior Publication Policy

Please review our prior publication policy. We expect authors to disclose any prior dissemination including via a website or at national meetings.

Change of Address

Please notify the editors of any change of address. After manuscript acceptance, please also notify the publishers: Journals Production Department, Oxford University Press, Great Clarendon Street, Oxford, OX2 6DP, UK. Telephone +44 (0) 1865 556767 , Fax +44 (0) 1865 267773.

Important Notes to Authors

The manuscripts will not be returned to authors following submission unless specifically requested.

Proofs

Authors are sent page proofs by email. These should be checked immediately and corrections, as well as answers to any queries, returned to the publishers as an annotated PDF via email or fax within 3 working days (further details are supplied with the proof). It is the author's responsibility to check proofs thoroughly.

Permission to Reproduce Figures and Extracts

Permission to reproduce copyright material, for online publication in perpetuity, must be cleared and if necessary paid for by the author; this includes applications and payments to DACS, ARS and similar licensing agencies where appropriate. Evidence in writing that such permissions have been secured from the rights-holder must be made available to the editors.

It is also the author's responsibility to include acknowledgements as stipulated by the particular institutions. Please note that obtaining copyright permission could take some time. Oxford Journals can offer information and documentation to assist authors in securing online permissions: please see the Guidelines for Authors section on the Rights & permissions page.

Should you require copies of this then please contact the editorial office of the journal in question or the Oxford Journals Rights department on journals.permissions@oup.com.

For a copyright prose work, it is recommended that permission is obtained for the use of extracts longer than 400 words; a series of extracts totalling more than 800 words, of which any one extract is more than 300 words; or an extract or series of extracts comprising one-quarter of the work or more. For poetry: an extract of more than 40 lines; series of extracts totalling more than 40 lines; an extract comprising one-quarter or more of a complete poem.

Supplementary Data

Supporting material that is not essential for inclusion in the full text of the manuscript, but would nevertheless benefit the reader, can be made available by the publisher as online-only content, linked to the online manuscript. The material should not be essential to understanding the conclusions of the paper, but should contain data that is additional or complementary and directly relevant to the article content. Such information might include more detailed methods, extended data sets/data analysis, or additional figures.

It is standard practice for appendices to be made available online-only as supplementary data. All text and figures must be provided in suitable electronic

formats. All material to be considered as supplementary data must be submitted at the same time as the main manuscript for peer review. It cannot be altered or replaced after the paper has been accepted for publication, and will not be edited. Please indicate clearly all material intended as supplementary data upon submission and name the files e.g. 'Supplementary Figure 1', 'Supplementary Data', etc. Also ensure that the supplementary data is referred to in the main manuscript where necessary, for example as '(see Supplementary data)' or '(see Supplementary Figure 1)'.

Open Access

HPP offers the option of publishing under either a standard licence or an open access licence. Please note that some funders require open access publication as a condition of funding. If you are unsure whether you are required to publish open access, please do clarify any such requirements with your funder or institution.

Should you wish to publish your article open access, you should select your choice of open access licence in our online system after your article has been accepted for publication. You will need to pay an open access charge to publish under an open access licence.

Details of the open access licences and open access charges.

OUP has a growing number of Read and Publish agreements with institutions and consortia which provide funding for open access publishing. This means authors from participating institutions can publish open access, and the institution may pay the charge. Find out if your institution is participating.

Please note that you may be eligible for a discount to the open access charge based on society membership. Authors may be asked to prove eligibility for the member discount.

Health Policy and Planning is committed to transitioning to full Open Access as soon as possible and in any event no later than when 75% of its research content is published Open Access.

Financial support from cOAlition S funders will end at the latest by 31 December 2024.

Publication Ethics

HPP and Oxford University Press are members of the Committee on Publication Ethics (COPE). This journal follows the guidance provided in COPE's Core Practices. The journal also subscribes to the International Committee of Medical Journal Editors (ICJME) Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly work in Medical Journals. The journal expects all parties involved in the publication of content in HPP (the publisher, editors, authors, and reviewers) to follow these guidelines on best practice and publication ethics.

CrossRef Funding Data Registry

In order to meet your funding requirements authors are required to name their funding sources, or state if there are none, during the submission process. For further information on this process or to find out more about the CHORUS initiative please see CHORUS at Oxford University Press page.