



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE COLETIVA
CURSO DE DOUTORADO EM SAÚDE COLETIVA

TAINA BARBIE DO ESPÍRITO SANTO

Casos Confirmados e Óbitos por Covid-19 no Brasil: Análise da Qualidade dos dados, 2020 a 2023

Florianópolis/SC

2025

TAINA BARBIE DO ESPÍRITO SANTO

Casos Confirmados e Óbitos por Covid-19 no Brasil: Análise da Qualidade dos dados, 2020 a 2023

Tese submetida ao Programa de Pós-graduação em Saúde Coletiva, da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Doutor em Saúde Coletiva.

Orientadora: Prof.^a Dr.^a Alexandra Crispim Boing

Temática: Epidemiologia das Doenças Transmissíveis

Florianópolis/SC

2025

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Espírito Santo, Taina Barbie do
Casos Confirmados e Óbitos por Covid-19 no Brasil:
Análise da Qualidade dos dados, 2020 a 2023 / Taina Barbie
do Espírito Santo ; orientador, Alexandra Crispim Boing,
2025.
123 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro de Ciências da Saúde, Programa de Pós
Graduação em Saúde Coletiva, Florianópolis, 2025.

Inclui referências.

1. Saúde Coletiva. 2. Saúde Coletiva. 3. Qualidade de
Dados. 4. Epidemiologia. 5. Covid-19. I. Crispim Boing,
Alexandra. II. Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Saúde Coletiva. III. Título.

Taina Barbie do Espírito Santo

**Casos Confirmados e Óbitos por Covid-19 no Brasil: Análise da Qualidade dos dados,
2020 a 2023**

O presente trabalho em nível de Doutorado foi avaliado e aprovado, em 05 de dezembro de 2025, pela banca examinadora composta pelos seguintes membros:

Ana Cristina Vidor, Dra
Secretaria Municipal de Saúde de Florianópolis

Ana Luiza de Lima Curi Hallal, Dra
Universidade Federal de Santa Catarina

Paulo Roberto Barbato, Dr.
Universidade Federal de Santa Catarina

Rodrigo de Macedo Couto, Dr
Ministério da Saúde

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutora em Saúde coletiva.

 Documento assinado digitalmente
Ana Luiza de Lima Curi Hallal
Data: 10/12/2025 20:51:24-0300
CPF: ***.987.250-**-
Verifique as assinaturas em <https://v.ufsc.br>

Profa. Dra. Ana Luiza de Lima Curi Hallal
Coordenação do Programa de Pós-Graduação

 Documento assinado digitalmente
Alexandra Crispim da Silva Boing
Data: 12/12/2025 09:27:36-0300
CPF: ***.146.819-**-
Verifique as assinaturas em <https://v.ufsc.br>

Prof.^a Dr.^a Alexandra Crispim Boing
Orientadora

Florianópolis, 2025

Dedico esta tese a todos aqueles afetados pelos danos irreparáveis da pandemia, às vítimas, às famílias enlutadas e aos profissionais de saúde que estiveram na linha de frente. Que, como pesquisadores, possamos contribuir para prevenir e mitigar tragédias dessa magnitude.

AGRADECIMENTOS

A Deus, pelo amor incondicional e por me permitir chegar até aqui, mesmo quando o caminho parecia incerto.

À minha mãe, à minha avó e a toda a minha família, pelo amor, abrigo, força e por acreditarem em mim.

Aos amigos que caminharam ao meu lado e transformaram os dias difíceis em esperança e afeto.

À minha orientadora, Dra. Alexandra Crispim Boing, por seu apoio constante, paciência e sensibilidade ao compreender os desafios deste percurso. Sua orientação foi fundamental para que este trabalho se concretizasse.

Aos membros da banca, pela disponibilidade, atenção e contribuições que enriqueceram este trabalho.

Aos professores do Programa de Pós-Graduação em Saúde Coletiva da Universidade Federal de Santa Catarina, pelos ensinamentos, incentivo e pela dedicação em manter viva a chama da ciência.

E a todos que, de alguma forma, estiveram presentes nesta caminhada, meu sincero agradecimento.

RESUMO

A pandemia de COVID-19 expôs, em escala global, as fragilidades dos sistemas de informação em saúde e evidenciou a importância da qualidade dos dados para a vigilância epidemiológica e a tomada de decisão. Esta tese analisou a qualidade dos dados de casos confirmados e óbitos por COVID-19 no Brasil, entre janeiro de 2020 e dezembro de 2023, considerando a confiabilidade, completude e oportunidade dos registros. Trata-se de um estudo descritivo, de abordagem quantitativa, desenvolvido em duas etapas complementares. A análise da confiabilidade avaliou as distorções dos registros à luz da Lei de Benford, em países com elevado número de notificações, como Estados Unidos, Índia, França, Alemanha, Japão e Brasil, utilizando dados secundários da Universidade Johns Hopkins, entre 2020 e 2022. Já a análise da completude e da oportunidade utilizou 2.200.796 registros de Síndrome Respiratória Aguda Grave por COVID-19 notificados no sistema SIVEP-Gripe entre 2020 e 2023, abrangendo 66 variáveis da ficha de investigação. A oportunidade dos dados foi mensurada pelos intervalos entre início dos sintomas e notificação, notificação e digitação, e notificação e encerramento, considerando os parâmetros históricos da vigilância (7, 30 e 60 dias, respectivamente). Nos resultados de confiabilidade, o Brasil foi o único país a não apresentar conformidade com a Lei de Benford nos registros de casos. Para os óbitos, apenas Estados Unidos e Japão apresentaram conformidade estrita ou aceitável, enquanto o Brasil mostrou conformidade marginalmente aceitável. Na análise de completude, observou-se que 59% das variáveis apresentaram grau ruim ou muito ruim, e apenas 24% foram classificadas como excelente ou boa. Em relação à oportunidade, o tempo médio de notificação reduziu de 13 para 7 dias. A digitação superou 90% de registros oportunos na maioria dos estados brasileiros, enquanto o encerramento manteve-se em torno de 86%. Ainda assim, a meta de 90% de notificações oportunas não foi alcançada. Apesar da redução nos tempos médios de notificação e digitação, as metas de oportunidade preconizadas não foram atingidas em sua totalidade. Os resultados indicam avanços pontuais na agilidade do fluxo informacional, mas persistem lacunas relevantes na qualidade e padronização dos registros. Conclui-se que a pandemia expôs limitações dos sistemas de informação em saúde e reforçou a urgência de fortalecer a governança dos dados, a capacitação profissional e a cultura de monitoramento e investimentos financeiros contínuos e direcionados às ações de vigilância epidemiológica, elementos essenciais para respostas mais eficazes em futuras emergências sanitárias.

Palavras-chave: Covid-19; Síndrome respiratória aguda grave; Sistemas de informação em Saúde; Confiabilidade dos dados; Lei de Benford.

ABSTRACT

The COVID-19 pandemic exposed, on a global scale, the fragilities of health information systems and highlighted the importance of data quality for epidemiological surveillance and decision-making. This thesis analyzed the quality of data on confirmed COVID-19 cases and deaths in Brazil, between January 2020 and December 2022, considering the reliability, completeness, and timeliness of records. This was a descriptive study with a quantitative approach, developed in two complementary stages. The reliability analysis assessed distortions in the records based on Benford's Law in countries with high numbers of notifications—such as the United States, India, France, Germany, Japan, and Brazil—using secondary data from Johns Hopkins University between 2020 and 2022. The completeness and timeliness analysis included 2,200,796 records of Severe Acute Respiratory Syndrome (SARS) due to COVID-19 reported in the SIVEP-Gripe system from 2020 to 2023, covering 66 variables from the investigation form. Data timeliness was measured by the intervals between symptom onset and notification, notification and data entry, and notification and case closure, considering historical surveillance parameters (7, 30, and 60 days, respectively). Regarding reliability, Brazil was the only country that did not show conformity with Benford's Law for case records. For deaths, only the United States and Japan showed strict or acceptable conformity, while Brazil demonstrated marginally acceptable conformity. In the completeness analysis, 59% of the variables were classified as poor or very poor, and only 24% as good or excellent. Concerning timeliness, the average notification time decreased from 13 to 7 days. Data entry exceeded 90% of timely records in most Brazilian states, while case closure remained around 86%. Nevertheless, the target of 90% timely notifications was not reached. Despite some improvements in notification and data entry times, the overall timeliness goals were not fully achieved. The results indicate limited progress in data flow efficiency, with persistent gaps in data quality and standardization. It is concluded that the pandemic exposed historical weaknesses in health information systems and reinforced the urgent need to strengthen data governance, professional training, and a culture of continuous monitoring—key elements for more effective responses to future public health emergencies.

Keywords: COVID-19; Severe Acute Respiratory Syndrome; Health Information Systems; Data Reliability; Benford's Law.

LISTA DE FIGURAS

Figura 1 – Fluxograma de Registro dos casos de SG e SRAG.....	34
Figura 2 – Fluxograma apresentando as etapas da revisão de literatura sistemizada.....	51
Artigo I	
Figura 1 – Distribuição dos casos e óbitos por COVID-19 segundo a Lei de Benford. Brasil, EUA e Índia, 2020 a 2022.....	82
Figura 2 – Distribuição dos casos e óbitos por COVID-19 segundo a Lei de Benford. França, Alemanha e Japão, 2020 a 2022.....	83
Figura 4 – Distribuição dos casos e óbitos por COVID-19 segundo a Lei de Benford. Brasil, 2020 a setembro de 2022, por período.....	84
Artigo II	
Figura 1 – Percentual de notificação, digitação e encerramento oportunos dos casos de SRAG por COVID-19 segundo Unidade Federativa do Brasil, 2020–2023.....	107

LISTA DE TABELAS

Tabela 1 – Distribuição de frequência do primeiro dígito de acordo com a Lei de Benford.....	40
Tabela 2 – Análise do grau de conformidade segundo a Lei de Benford.....	43

Artigo I

Tabela 1 – Distribuição dos primeiros dígitos dos casos e óbitos por COVID-19, 2020 a 2022.....	85
Tabela 2 – Distribuição dos primeiros dígitos dos casos e óbitos por COVID-19 por períodos, Brasil, 2020 a 2022.....	86

Artigo II

Tabela 1 – Percentual de variáveis dos casos de SRAG por COVID-19 segundo grau de completude por ano de notificação no Brasil, 2020–2023.....	107
---	-----

LISTA DE QUADROS

Quadro 1 - Estratégias de buscas utilizadas para levantamento da revisão de literatura.	47
Quadro 2 - Variáveis referentes aos Dados do Paciente	63
Quadro 3 - Variáveis referentes aos Dados Clínicos e Epidemiológicos.....	64
Quadro 4 - Variáveis referentes aos Dados de Conclusão do Caso	66
Quadro 5 - Análise da Oportunidade	67

LISTA DE ABREVIATURAS E SIGLAS

ANVISA – Agência Nacional de Vigilância Sanitária
CENEPI – Centro Nacional de Epidemiologia
CDC – Centers for Disease Control and Prevention
CIEVS – Centro de Informações Estratégicas de Vigilância em Saúde
COVID-19 – Coronavirus Disease 2019
DATASUS – Departamento de Informática do Sistema Único de Saúde
DSEI – Distrito Sanitário Especial Indígena
DMA – Desvio Médio Absoluto
ESPII – Emergência em Saúde Pública de Importância Internacional
GAL – Sistema Gerenciador de Ambiente Laboratorial
IC – Intervalo de Confiança
LB – Lei de Benford
MDA – Média dos Desvios Absolutos
NMV/r – Nirmatrelvir/Ritonavir
OPAS – Organização Pan-Americana da Saúde
OMS – Organização Mundial da Saúde
PNVS – Política Nacional de Vigilância em Saúde
Qui-Quadrado (χ^2) – Teste Estatístico de Comparação de Frequências Observadas e Esperadas
REDCap – Research Electronic Data Capture
RSI – Regulamento Sanitário Internacional
SARS-CoV – Coronavírus da Síndrome Respiratória Aguda Grave
SARS-CoV-2 – Severe Acute Respiratory Syndrome Coronavirus 2
SE – Semana Epidemiológica
SIH – Sistema de Informações Hospitalares
SINAN – Sistema de Informação de Agravos de Notificação
SIM – Sistema de Informações sobre Mortalidade
SIS – Sistema de Informação em Saúde
SIVEP-Gripe – Sistema de Informação de Vigilância Epidemiológica da Gripe
SRAG – Síndrome Respiratória Aguda Grave
SUS – Sistema Único de Saúde

SVS – Secretaria de Vigilância em Saúde

SVSA – Secretaria de Vigilância em Saúde e Ambiente

Teste D1 – Teste do Primeiro Dígito

Teste D2 – Teste do Segundo Dígito

Teste D1D2 – Teste dos Dois Primeiros Dígitos

Teste da Soma – Teste da Soma

TR-Ag – Teste Rápido de Antígeno

VOI – Variant of Interest (Variante de Interesse)

VOC – Variant of Concern (Variante de Preocupação)

VUM – Variant Under Monitoring (Variante sob Monitoramento)

WHO – World Health Organization

SUMÁRIO

1	INTRODUÇÃO	16
2	OBJETIVOS	19
2.1	Objetivo Geral	19
2.2	Objetivos Específicos	19
3	REVISÃO DE LITERATURA.....	20
3.1	COVID-19.....	20
3.2	Contexto histórico da Pandemia Covid19	23
3.3	Vigilância em saúde	26
3.4	Sistemas de Informação em Saúde	29
3.5	Desafios nos Registros de Covid-19 no Brasil	35
3.6	Qualidade dos Dados.....	36
3.7	Qualidade de dados na pandemia	38
3.8	Lei de Benford.....	40
3.8.1	Testes associados à Lei de Benford.....	42
3.8.2	Testes estatísticos	43
3.9	Estudos utilizando a Lei de Benford.....	45
3.10	Estudos sobre Análise da Qualidade de dados de Covid-19	46
3.10.1	Estudos relacionados à Lei de Benford	52
3.10.2	Estudos relacionados à completitude dos dados.....	58
4	MÉTODOS.....	60
4.1	Delineamento do Estudo	60
4.2	Análise da Qualidade de Dados de Covid-19.....	60
4.2.1	Confiabilidade de dados	60
4.2.2	Completude e Oportunidade.....	62
4.3	Análise dos Dados da Completude e Oportunidade	68
4.4	ASPECTOS ÉTICOS	68
5	RESULTADOS	68
5.1	ARTIGO I	69
5.2	ARTIGO II	91
6	CONSIDERAÇÕES FINAIS	111
7	REFERÊNCIAS	113

Apêndices 119
ANEXOS 122

1 INTRODUÇÃO

A pandemia da Covid-19, em decorrência de sua rápida disseminação em níveis alarmantes, desencadeou uma crise global com impactos devastadores na saúde e na economia (MOURA et.al., 2021). O governo chinês notificou a Organização Mundial da Saúde (OMS) em 31 de dezembro de 2019 sobre um surto de pneumonia grave em Wuhan, província de Hubei, cujo agente etiológico identificado foi um novo coronavírus, o SARS-CoV-2, causador da COVID-19 (BAQUI et al., 2020).

Nos primeiros 30 dias, a China já havia registrado mais de 11 mil casos e 259 óbitos e a doença rapidamente se espalhou para países da Ásia, Europa e América do Norte. Em 30 de janeiro de 2020, a OMS declarou Emergência em Saúde Pública de Importância Internacional (ESPII), e em 11 de março, decretou a situação como pandemia, em um cenário com mais de 110 mil casos distribuídos em 114 países (CAVALCANTE et al., 2020). No final de março, havia mais de 850 mil casos e 40 mil mortes em todo o mundo (IDROVO; MANRIQUE-HERNÁNDEZ, 2020). Até 31 de dezembro de 2022, três anos após o início da pandemia, o mundo acumulava mais de 660 milhões de casos e 6,6 milhões de óbitos.

No Brasil, os primeiros casos foram confirmados em fevereiro de 2020 e até dezembro de 2022, o país acumulava mais de 36 milhões de casos e 693 mil óbitos, ocupando a quinta posição mundial em número absoluto de casos e a segunda em óbitos (Johns Hopkins University, 2022). A magnitude desses números impressiona não apenas pela dimensão humanitária, mas também pelas mudanças profundas no cotidiano da população e na situação econômica do país (BRESSER-PEREIRA, 2020).

Diferente de outros eventos epidêmicos, a Covid-19 foi acompanhada em tempo real, por veículos de imprensa, instituições acadêmicas e órgãos governamentais. A produção e a disseminação de dados acompanharam o avanço da pandemia, tornando a qualidade dessas informações um elemento crucial para a tomada de decisão e formulação de políticas públicas (SILVA; FIGUEIREDO FILHO, 2021). Contudo, esse volume crescente de informações também expôs fragilidades dos sistemas nacionais de informação em saúde. Em vários países, surgiram

dúvidas sobre a confiabilidade dos dados oficiais, especialmente no Brasil, onde se observaram mudanças nos critérios de registro, indisponibilidade de dados e sobrecarga dos sistemas resultando em registros incompletos e atrasados e, por vezes, de baixa qualidade (BALASHOV; YAN; ZHU, 2021).

No Brasil, episódios como a interrupção do painel Coronavírus em junho de 2020, motivada por razões políticas, e a tentativa de alteração arbitrária de variáveis e padrões de divulgação, contrários às diretrizes internacionais, reforçaram a percepção de fragilidades na gestão da informação. Outro episódio relevante ocorreu em dezembro de 2021, quando o sistema de monitoramento nacional permaneceu fora do ar por cerca de 30 dias, coincidindo com a emergência da variante Ômicron, com alta taxa de transmissibilidade (SILVA; FIGUEIREDO, 2021). Esses “apagões de dados” comprometeram a transparência, dificultaram análises oportunas e prejudicaram a construção de indicadores relevantes para o enfrentamento da doença (VILLELA; GOMES, 2022).

O Brasil dispõe de uma ampla rede de Sistemas de Informação em Saúde (SIS) de âmbito nacional, com grande parte de suas informações disponível pelo Departamento de Informática do SUS (DATASUS). A apropriação dessas informações pelos gestores e pela sociedade deve considerar tanto seus pontos fortes quanto suas limitações, as quais são identificadas a partir de avaliações regulares e sistemáticas dos dados disponibilizados (LIMA et al., 2009).

Durante a pandemia, a resposta operacional exigiu uma rápida adaptação e alavancagem das capacidades dos Sistemas de Informação em Saúde (SIS) para que fossem feitas análises em tempo real que permitissem compreender a situação epidemiológica e elaborar medidas de controle adequadas. Entretanto, a magnitude da crise sobrecarregou tais sistemas em diversos países, expondo suas limitações frente às demandas emergenciais (MAHMOOD et al., 2020).

Apesar da contínua melhoria dos SIS, persistem fragilidades na qualidade dos dados. Estudos apontam que, no Brasil, existem evidências de deficiências na qualidade dos dados, o que indica necessidade de qualificar e potencializar os SIS enquanto ferramenta para a produção de conhecimentos em saúde e utilizá-los, satisfatoriamente, na gestão do Sistema Único de Saúde (SUS) (NERI, 2016). A

informação é essencial para analisar as condições de saúde e subsidiar os processos de tomada de decisão e, por esta razão, é fundamental que os dados dos sistemas de informação sejam confiáveis, completos e oportunos (MASSOUDI; SOBOLEVSKAIA, 2021); (IDROVO; MANRIQUE-HERNÁNDEZ, 2020).

Nesse contexto, torna-se necessário adotar abordagens estatísticas capazes de avaliar a confiabilidade dos dados. A Lei de Benford (LB), que descreve a distribuição de dígitos, tem sido aplicada em diferentes campos para identificar inconsistências em dados, incluindo processos eleitorais, estatísticas econômicas, relatórios contábeis e, mais recentemente, registros de Covid-19 (CAMPOLIETI, 2022). Paralelamente, a completude, entendida como o grau de preenchimento das variáveis em sistemas de informação, constitui outra dimensão essencial, pois a ausência de dados compromete análises epidemiológicas (ROMERO; CUNHA, 2006; SILVA et al., 2020).

Diante do exposto, o interesse e a motivação para desenvolver este estudo surgiu a partir do reconhecimento da potencialidade que as informações têm, enquanto instrumento, para contribuir com a efetivação de mudanças positivas nas condições de vida e saúde da população, por meio do subsídio que pode oferecer aos processos de tomada de decisão da gestão.

Nessa perspectiva, considerando que apenas informações de boa qualidade podem cumprir o papel de ferramenta para a gestão do Sistema Único de Saúde (SUS) e para produção de conhecimentos em saúde, o presente estudo busca responder à questão norteadora: Qual a qualidade dos dados dos casos e óbitos por COVID-19 no Brasil?

2 OBJETIVOS

2.1 Objetivo Geral

Analisar a qualidade de dados dos casos confirmados e óbitos por Covid-19 no Brasil, no período entre janeiro de 2020 e dezembro de 2023.

2.2 Objetivos Específicos

- Verificar a conformidade da Lei de Benford com os registros dos casos e óbitos por Covid-19 no Brasil e nos países com maior número de casos confirmados: Estados Unidos (EUA), Índia, França, Alemanha e Japão.
- Analisar indícios de desvios entre as proporções esperadas e registradas dos casos confirmados e óbitos por Covid-19 no Brasil, EUA, Índia, França, Alemanha e Japão.
- Analisar a completude das notificações de casos de síndrome respiratória aguda grave por Covid-19 registradas no Sistema de Informação de Vigilância Epidemiológica da Gripe (SIVEP-Gripe).
- Analisar a oportunidade de notificação, digitação, investigação e encerramento das notificações de casos de síndrome respiratória aguda grave por Covid-19 registradas no Sistema de Informação de Vigilância Epidemiológica da Gripe (SIVEP-Gripe).

3 REVISÃO DE LITERATURA

Para a construção teórica deste estudo foi realizada uma revisão de literatura que abordam aspectos conceituais e teóricos concernentes a Covid-19, aos Sistemas de Informação em Saúde, Qualidade de Dados, incluindo a Confiabilidade, Completude e Oportunidade de notificação, e a Lei de Benford.

3.1 COVID-19

A COVID-19 é uma doença respiratória aguda causada pelo coronavírus SARS-CoV-2, potencialmente grave, de elevada transmissibilidade e de distribuição global. O SARS-CoV-2 é um vírus pertencente ao subgênero Sarbecovírus, gênero Betacoronavírus, da subfamília Orthocoronavirinae, família Coronaviridae. O Coronavírus da Síndrome Respiratória Aguda Grave (SARS-CoV) ocasionou surtos em 2002 e 2003 na província de Guangdong na China, e o Coronavírus da síndrome respiratória do Oriente Médio (MERS-CoV) foi o patógeno responsável por graves surtos em 2012 (ZHU et al., 2020).

O SARS-CoV-2 dissemina-se principalmente pelo contato próximo entre uma pessoa infectada e outra suscetível, como em situações de conversas cotidianas. O período de incubação pode variar de 1 a 14 dias (CDC, 2024b), com mediana variando de acordo com a variante do vírus (BRASIL, 2024). A transmissibilidade do vírus, entretanto, não é uniforme, variando conforme a carga viral do indivíduo infectado, o tipo e a intensidade do contato, as características do ambiente e a adoção de medidas de prevenção e controle (WHO, 2021b).

As partículas respiratórias eliminadas pela boca ou nariz de indivíduos infectados durante a tosse, espirro, fala, canto ou respiração podem ser classificadas em gotículas e aerossóis. As gotículas, geralmente maiores que 5 µm, tendem a se depositar rapidamente nas mucosas da boca e do nariz, atingindo principalmente as vias aéreas superiores. Já os aerossóis são menores e mais leves, permanecendo suspensos no ar por períodos mais longos e, quando inalados, podem penetrar nas porções mais profundas do trato respiratório (BRASIL, 2023d).

De acordo com a Organização Mundial da Saúde (OMS), a transmissão da COVID-19 pode ocorrer de três formas principais: contato direto ou indireto, por meio

da aproximação física ou de superfícies contaminadas (fômites); transmissão por gotículas, quando partículas maiores contendo o vírus são expelidas e se depositam nas mucosas de pessoas próximas, geralmente a menos de 1 metro; e transmissão por via aérea, caracterizada pela inalação de aerossóis que podem permanecer suspensos no ambiente, seja a curta distância, próximos ao emissor, ou a longa distância, por horas e a mais de 1 metro de alcance (WHO, 2021b; WHO, 2021c; WHO, 2023b).

A infecção por SARS-CoV-2 pode variar de casos assintomáticos até quadros críticos, exigindo atenção aos sinais de agravamento que demandam hospitalização. Os sintomas mais comuns incluem febre, tosse, fadiga, dispneia, mialgia e cefaleia, podendo ocorrer também manifestações como dor de garganta, congestão nasal, náusea, vômito, diarreia, além de anosmia e ageusia (WHO, 2023e; CDC, 2023a). Pode permanecer assintomática nos estágios iniciais, até o surgimento de pneumonia grave, dispneia, disfunção de órgãos e até morte. Embora todas as idades possam ser afetadas, a infecção tem menor impacto clínico em crianças do que em adultos. A maioria das crianças infectadas parece ter um curso mais brando e melhores resultados clínicos (ZHANG, 2023).

De forma geral, os casos podem ser classificados em: assintomáticos, quando não há sintomas; leves, com sintomas inespecíficos sem alteração respiratória significativa; moderados, com acometimento do trato respiratório inferior mas saturação $\geq 94\%$; graves, quando há síndrome respiratória aguda grave (SRAG), dispneia ou saturação $\leq 94\%$; e críticos, caracterizados por síndrome do desconforto respiratório agudo, sepse, choque séptico, trombose ou falência de múltiplos órgãos (BRASIL, 2022c; WHO, 2023e).

Atualmente, os principais fatores de risco associados ao desenvolvimento de formas graves da COVID-19 incluem a idade avançada (acima de 60 anos, com risco progressivo conforme a idade aumenta), a presença de comorbidades como diabetes, hipertensão, doenças cardiovasculares e pulmonares crônicas, doença renal crônica, obesidade, câncer, demência, transtornos mentais e condições de imunossupressão, incluindo infecção pelo HIV. Também se destacam como grupos vulneráveis as gestantes e puérperas, sobretudo aquelas com idade superior a 35

anos, obesidade ou condições crônicas associadas, além de indivíduos tabagistas e não vacinados contra a COVID-19 (WHO, 2023e).

O diagnóstico laboratorial da infecção por SARS-CoV-2 pode ser realizado por meio de testes rápidos de antígeno (TR-Ag) ou por testes de biologia molecular (CDC, 2023c; BRASIL, 2022c, 2022g). Os TR-Ag são imunoenaios point-of-care, baseados na imunocromatografia de fluxo lateral, capazes de detectar antígenos virais em amostras nasais ou de nasofaringe, sendo indicados para identificação da infecção ativa na fase aguda. Já os testes moleculares identificam o RNA viral em secreções respiratórias, utilizando metodologias como a RT-qPCR — considerada o padrão-ouro para o diagnóstico — e a RT-LAMP, que apresenta menor complexidade técnica e maior rapidez na execução. (BRASIL, 2024).

O manejo adequado dos casos suspeitos ou confirmados da covid-19 depende do reconhecimento precoce de sinais e sintomas da doença e do monitoramento contínuo dos pacientes. A maioria dos casos de covid-19 apresentam sintomas leves e podem ser tratados ambulatorialmente. Medicamentos analgésicos e antipiréticos podem ser utilizados para manejo dos sintomas. Para adultos com alto risco de progressão para doença grave, opções terapêuticas antivirais estão disponíveis para reduzir o risco de hospitalização ou morte em nível ambulatorial (WHO, 2023j).

No Brasil, o SUS incorporou a combinação de antivirais nirmatrelvir e ritonavir (NMV/r) para o tratamento da infecção por SARS-CoV-2 em grupos populacionais específicos, com o objetivo de reduzir internações, complicações e óbitos por covid-19 (Brasil, 2022h). A incorporação do medicamento é particularmente relevante, pois, apesar dos avanços da vacinação, pessoas idosas e imunossuprimidas podem apresentar resposta vacinal reduzida, tornando necessária a disponibilização de alternativas terapêuticas. O NMV/r é indicado para pacientes com alto risco de evolução para doença grave, com diagnóstico confirmado de covid-19 por teste rápido de antígeno ou exame molecular, que apresentem sintomas leves ou moderados e não necessitem de oxigênio suplementar, incluindo pessoas imunocomprometidas com 18 anos ou mais e idosos com 65 anos ou mais. O tratamento deve ser iniciado até cinco dias após o início dos sintomas. (BRASIL, 2024).

Ressalta-se, ainda, que no âmbito do SUS foram incorporados os medicamentos tocilizumabe (Brasil, 2022i) e baricitinibe (Brasil, 2022j) para tratamento de pacientes adultos com covid-19 hospitalizados. A OMS mantém uma lista de tratamentos recomendados, juntamente com as evidências de cada um, disponível em <https://app.magicapp.org/#/guideline/nBkO1E>. O site da Agência Nacional de Vigilância Sanitária (Anvisa) disponibiliza informações atualizadas sobre os medicamentos utilizados no tratamento da covid-19, incluindo orientações sobre uso, indicações e segurança: <https://www.gov.br/anvisa/pt-br/assuntos/campanhas/coronavirus/medicamentos>.

3.2 Contexto histórico da Pandemia Covid19

A pandemia de Covid19 constituiu um evento histórico de grande impacto, com profundas repercussões na vida das pessoas em nível global. Em 31 de dezembro de 2019, o governo chinês notificou a Organização Mundial da Saúde (OMS) sobre um surto de pneumonia grave de etiologia desconhecida em Wuhan, província de Hubei, China, uma cidade de aproximadamente 11 milhões de habitantes. Das 41 pessoas inicialmente hospitalizadas com pneumonia desconhecida até 2 de janeiro de 2020, 27 (66%) tiveram exposição direta ao Mercado Atacadista de Frutos do Mar de Huanan (WOROBEY, 2022). O novo Coronavírus 2019 (2019-nCoV), posteriormente denominado (SARS-CoV-2), foi identificado como o causador do surto e a doença foi denominada Covid-19 (LU; STRATTON; TANG, 2020).

Nos primeiros 30 dias, a China registrou mais de 11 mil casos e 259 óbitos e a doença atingiu outros países da Ásia, Europa e América do Norte. Em 30 de janeiro de 2020, a OMS declarou Emergência em Saúde Pública de Importância Internacional (ESPII). Em um cenário com mais de 110 mil casos distribuídos em 114 países, a OMS decretou pandemia no dia 11 de março de 2020 (CAVALCANTE et al., 2020), diante da rápida disseminação e da gravidade da doença. No final de março de 2020, havia mais de 850 mil casos e 40 mil mortes em todo o mundo (IDROVO; MANRIQUE-HERNÁNDEZ, 2020).

O primeiro caso de covid-19 no Brasil foi confirmado em 26 de fevereiro de 2020, em São Paulo, seguido do primeiro óbito em 12 de março e da declaração de

transmissão comunitária em todo o território nacional em 20 de março. Até dezembro de 2022, o país acumulava mais de 36 milhões de casos e 693 mil óbitos, ocupando a quinta posição mundial em número absoluto de casos e a segunda em óbitos (Johns Hopkins University, 2022).

Passados dois anos, em abril de 2022, o Ministério da Saúde anunciou o fim da ESPIN, considerando a melhora do cenário epidemiológico e o avanço da vacinação. Finalmente, em 5 de maio de 2023, a OMS decretou o fim da ESPII, reconhecendo a redução dos casos graves e óbitos e o controle mais efetivo da covid-19 em nível mundial (WHO, 2023; BRASIL, 2024).

Para além desses marcos cronológicos, o conceito de onda passou a ser aplicado para contextualizar o comportamento da pandemia. O termo não consensual entre especialistas, não tem como embasamento critérios científicos estabelecidos no campo da epidemiologia, mas está relacionado principalmente ao aumento acentuado no número de casos. Alguns estudiosos analisam que o fim de uma onda só seria confirmado com redução significativa de casos (ROCHA, 2021). Contudo, há analistas que consideram a existência de duas ondas, a segunda considerada “anômala”, dado que evoluiu antes que a primeira onda tenha se reduzido de forma significativa, caracterizada também por uma “superposição” das duas ondas (VASCONCELOS et.al, 2021).

A primeira onda epidêmica de COVID-19, no Brasil, ocorreu entre os meses de fevereiro a outubro de 2020, caracterizada pela expansão do número de casos no país, atingindo o pico de casos em julho, com 319 mil registros apenas na semana epidemiológica (SE) 30. Ocorreu uma segunda onda de transmissão, iniciada em novembro de 2020 e associada ao surgimento de novas variantes, permanecendo até novembro de 2021. O pico ocorreu em março, atingindo 539 mil casos na SE 12, seguido de redução gradual a partir de julho (SE 26) (MINISTÉRIO DA SAÚDE, 2025); (FIOCRUZ, 2022).

A terceira onda iniciou em dezembro de 2021, coincidindo com o período de festas, férias, relaxamento de medidas de restrição e a introdução da variante Ômicron. Nesta fase observou-se um acelerado aumento no número de casos, precedido pelo aumento da positividade de testes com velocidade muito superior às

primeiras ondas. Em 2022, o maior aumento de toda a pandemia foi observado no final de janeiro, com mais de 1,3 milhão de casos apenas na SE 4. Em 2023, o aumento mais expressivo de casos ocorreu em janeiro, totalizando 157 mil registros na SE 1, enquanto em 2024 o pico foi registrado no final de fevereiro, com 70 mil casos na SE 9 (MINISTÉRIO DA SAÚDE, 2025); (FIOCRUZ, 2022).

As novas variantes do vírus SARS-CoV-2 são monitoradas em todo o mundo, inclusive no Brasil, para que sejam investigados e relatados seus impactos, já que elas podem alterar as características da doença, da transmissão do vírus, influenciar o impacto da vacina, a terapêutica, gravidade, as metodologias dos testes de diagnóstico ou mesmo a eficácia das medidas de saúde pública aplicadas para prevenção e controle da propagação da covid-19. De acordo com o risco apresentado à saúde pública, a equipe da OMS classifica essas variantes como variantes de preocupação (VOC – do inglês *variant of concern*), variantes de interesse (VOI – do inglês *variant of interest*) ou variantes sob monitoramento (VUM – do inglês *variant under monitoring*) (MINISTERIO DA SAUDE, 2023).

Além da variante Alfa, que predominou no primeiro ano da pandemia no Brasil, tivemos outras, como a variante Gama, detectada em novembro de 2020 (em 6 meses passou de uma participação, entre os casos, de 12%, em dezembro de 2020, para mais de 95%, em maio de 2021) , a variante Delta, detectada em dezembro de 2020 (em 6 meses passou de uma participação de 0,8% em junho de 2021 para mais de 99% em novembro de 2021) e mais recentemente a variante Ômicron, detectada em novembro de 2021 (com participação entre os casos variando de 42,7%, em dezembro de 2021, para 96,4%, em janeiro de 2022) (FIOCRUZ, 2022).

Em meados de dezembro de 2020, ocorreu a aprovação da primeira vacina contra a Covid-19 em uso emergencial – a Comirnaty, desenvolvida pelo consórcio Pfizer/BioNTech – representando um marco mundial no enfrentamento da pandemia. A partir desse momento, autoridades de saúde em diferentes países mobilizaram-se para elaborar estratégias de vacinação e definir grupos prioritários. Apesar da urgência imposta pela pandemia, a campanha de vacinação contra a Covid-19 no Brasil iniciou apenas em 17 de janeiro de 2021, logo após a Agência Nacional de Vigilância Sanitária (Anvisa) aprovar o uso emergencial das vacinas CoronaVac,

produzida pelo Instituto Butantan com o laboratório chinês Sinovac, e AstraZeneca, desenvolvida pela Universidade de Oxford com a Fiocruz.

Inicialmente, com um pequeno número de doses disponível (6,2 milhões), e somente a partir do mês de março o país conseguiu chegar a um volume suficiente para acelerar o processo de vacinação (27,5 milhões). Esse avanço, contudo, não ocorreu com quantidade e velocidade necessárias para impedir o rápido crescimento e grande número de casos, internações e óbitos, bem como a crise e colapso do sistema de saúde, que atingiu o país entre março e junho (Fiocruz, 2022). Até dezembro de 2022, foram aplicadas mais de 503 mil doses de vacina contra a Covid-19 no Brasil (MINISTÉRIO DA SAÚDE, 2023).

A pandemia da Covid-19 impactou e impacta sobremaneira a sociedade global. Especificamente no Brasil, os efeitos devastadores dela foram ainda mais graves, sobretudo, em função das ações do governo federal que desencorajaram medidas de distanciamento, repassando a responsabilidade aos governos dos Estados. Tal conjuntura suscitou uma série de práticas que, por sua vez, facilitaram a maior disseminação do vírus, acarretando um número maior de contaminados e mortes.

3.3 Vigilância em saúde

Entende-se por Vigilância em Saúde o processo contínuo e sistemático de coleta, consolidação, análise de dados e disseminação de informações sobre eventos relacionados à saúde, visando o planejamento e a implementação de medidas de saúde pública, incluindo a regulação, intervenção e atuação em condicionantes e determinantes da saúde, para a proteção e promoção da saúde da população, prevenção e controle de riscos, agravos e doenças. A Vigilância epidemiológica é o conjunto de ações que proporcionam o conhecimento e a detecção de mudanças nos fatores determinantes e condicionantes da saúde individual e coletiva, com a finalidade de recomendar e adotar as medidas de prevenção e controle das doenças, transmissíveis e não-transmissíveis, e agravos à saúde (CONSELHO NACIONAL DE SAÚDE, 2018).

As bases conceituais da Vigilância em Saúde foram estabelecidas a partir da V Conferência Nacional de Saúde, realizada em 1975, que propôs a criação de um

sistema nacional de Vigilância Epidemiológica, cuja implementação ocorreu em 1976 com a estruturação do Sistema Nacional de Vigilância Epidemiológica (SNVE). Esse sistema instituiu a notificação compulsória de casos e óbitos referentes a quatorze doenças em todo o território nacional, configurando-se como o primeiro marco do processo de descentralização das ações de vigilância em saúde para as secretarias estaduais (BRASIL, 1975).

Com a criação do Sistema Único de Saúde (SUS), novos arranjos institucionais foram estabelecidos, entre eles a organização do Centro Nacional de Epidemiologia (CENEPI), no âmbito da Fundação Nacional de Saúde (Funasa); a criação da Agência Nacional de Vigilância Sanitária (Anvisa); e, posteriormente, em 2003, da Secretaria de Vigilância em Saúde (SVS), no Ministério da Saúde, que intensificou o processo normativo de descentralização das ações de vigilância (FRANCO NETTO et al., 2017).

Posteriormente, em 2005, ano da aprovação do Regulamento Sanitário Internacional (RSI), foi instituído o Centro de Informações Estratégicas de Vigilância em Saúde (CIEVS), com a finalidade realizar a detecção, verificação, avaliação, monitoramento e comunicação de rumores e de eventos que possam constituir emergências em saúde pública de relevância nacional. Essa iniciativa consolidou uma rede nacional de alerta e resposta às emergências em saúde pública. A expansão da Rede CIEVS no Brasil tem sido marcada por um aumento significativo no número de unidades em todo o país, refletindo o compromisso em fortalecer a vigilância e a resposta às potenciais emergências em saúde pública. Atualmente, esta Rede é composta por 190 unidades em funcionamento em todo o território nacional, sendo: 01 Nacional, 27 Estaduais (incluindo o DF), 42 Regionais, 26 de Capitais, 46 Municipais, 14 de Fronteiras e 34 dos Distritos Sanitários Especiais Indígenas (DSEI) (BRASIL, 2025).

Por meio da Portaria GM/MS nº 3.252/2009, é fortalecido o papel gestor de estados e municípios e se amplia o escopo das ações de Vigilância em Saúde, compreendendo fortaleceu o papel gestor de estados e municípios e ampliou o escopo das ações de Vigilância em Saúde, que passaram a compreender:

- **Vigilância Epidemiológica:** ações voltadas ao conhecimento, detecção e prevenção de mudanças nos fatores determinantes e condicionantes da saúde individual e coletiva, visando à adoção de medidas de prevenção e controle de doenças e agravos;
- **Vigilância Sanitária:** ações destinadas a eliminar, reduzir ou prevenir riscos à saúde e intervir em problemas decorrentes da produção e circulação de bens e da prestação de serviços de interesse sanitário, abrangendo desde a produção até o consumo;
- **Vigilância em Saúde do Trabalhador:** ações voltadas à promoção da saúde e à redução da morbimortalidade da população trabalhadora, por meio da integração de estratégias que enfrentem agravos e determinantes associados aos processos produtivos;
- **Vigilância em Saúde Ambiental:** ações para o monitoramento e detecção de mudanças ambientais que influenciem a saúde humana, com vistas à prevenção e ao controle de riscos ambientais;

Um marco relevante para a formulação da Política Nacional de Vigilância em Saúde (GT-PNVS) foi a Portaria GM/MS nº 1.378, de 2013, que regulamentou as responsabilidades e estabeleceu diretrizes para a execução e financiamento das ações de vigilância em saúde pela União, Estados, Distrito Federal e Municípios, no âmbito do Sistema Nacional de Vigilância em Saúde e do Sistema Nacional de Vigilância Sanitária. A referida portaria instituiu o Grupo de Trabalho Tripartite, com o objetivo de discutir e elaborar a Política Nacional de Vigilância em Saúde. Essa iniciativa visou fortalecer a capacidade do Estado brasileiro de enfrentar os desafios impostos à vigilância em saúde, considerando as mudanças decorrentes das transições demográfica e epidemiológica, bem como os determinantes sociais da saúde (FRANCO NETTO et al., 2017).

A Política Nacional de Vigilância em Saúde (PNVS) foi instituída em 2018. Essa política é fundamental para assegurar a prevenção e o controle de doenças em todo o território nacional. Atua como um marco orientador das ações de vigilância em saúde, estabelecendo diretrizes para o planejamento, coordenação e execução de atividades que monitoram a situação sanitária, identificam riscos e promovem a prevenção de agravos à saúde. Suas ações abrangem todo o país, desde

municípios de pequeno porte até grandes centros urbanos, garantindo a oferta de medidas de saúde pública adequadas e equitativas à população (CONSELHO NACIONAL DE SAÚDE, 2018).

Em janeiro de 2023, a Secretaria de Vigilância em Saúde, por meio do Decreto nº 11.358/2023, passou a ser denominada Secretaria de Vigilância em Saúde e Ambiente - SVSA. No âmbito da Vigilância em Saúde, a Vigilância Epidemiológica exerce função estratégica na organização, monitoramento e articulação das ações voltadas à detecção e prevenção de alterações nos determinantes e condicionantes da saúde individual e coletiva. Sua atuação é essencial no enfrentamento de emergências de saúde pública, como surtos, epidemias e pandemias, permitindo respostas rápidas e fundamentadas, com o objetivo de minimizar os impactos negativos sobre a sociedade. A efetividade dessas ações depende diretamente da disponibilidade e da qualidade de informações sistematizadas, tornando os Sistemas de Informação em Saúde ferramentas centrais para a coleta, integração e análise de dados, essenciais à tomada de decisão e ao planejamento das medidas de vigilância.

3.4 Sistemas de Informação em Saúde

O Brasil tem historicamente investido na criação de sistemas de informação em saúde para organizar seus dados em meio à complexidade de um país continental. Esse movimento ganhou força nos anos 1970, influenciado pela Organização Mundial da Saúde (OMS) e pela Organização Pan-Americana da Saúde (OPAS), quando foram padronizados os registros de óbito e as guias de internação hospitalar, culminando na criação do Sistema de Informações sobre Mortalidade (SIM) e do Sistema de Informações Hospitalares (SIH) (FARRANHA; RODRIGUES; BATAGLIA, 2022).

Com a consolidação do Sistema Único de Saúde (SUS), previsto na Constituição de 1988 e regulamentado pela Lei nº 8.080/1990, a descentralização da gestão da saúde reforçou o papel dos sistemas de informação como instrumentos de apoio à vigilância, à gestão e à tomada de decisão em todos os níveis federativos. Nesse contexto, o Departamento de Informática do SUS (DATASUS) passou a

centralizar e disponibilizar dados por meio de ferramentas como o TABNET e o TABWIN, que facilitam o acesso e a análise (DANIEL; PEREIRA; MACADAR, 2014).

A estratégia de elaboração de sistemas, em um caráter nacional, é recomendada pelos órgãos internacionais e o Brasil passou a adotá-la. Para a OMS, o Sistema de Informação em Saúde (SIS) é definido como um mecanismo de coleta, processamento, análise e distribuição da informação necessária para se organizar e operar os serviços de saúde e, também, para a investigação e o planejamento com vistas ao controle de doenças. Coletam uma ampla gama de informações de saúde para orientar em todos os níveis de um sistema de saúde, desde a gestão diária até a tomada de decisões políticas e implementação das ações (WHO, 2008).

Foram criados no intuito de superar dificuldades de congregar os diversos dados sobre saúde, possibilitando a geração de informações e sua divulgação para diversas finalidades e áreas, como de vigilância epidemiológica, assistenciais (atenção básica, hospitalares), nascimentos e mortalidades, ambulatoriais, dentre outras. Os SIS desempenham papel fundamental em diferentes frentes, como: apoiar gestores na tomada de decisões; subsidiar diagnósticos que orientem intervenções mais eficazes às necessidades da população; organizar e qualificar a gestão de recursos humanos; orientar a alocação do orçamento disponível; definir prioridades de ação considerando a distribuição territorial e espacial dos problemas de saúde e fomentar a produção científica (FARRANHA; RODRIGUES; BATAGLIA, 2022).

Desta forma, a análise das informações obtidas a partir desses sistemas influencia o processo de trabalho dos profissionais de saúde, o cotidiano da população, orienta a tomada de decisão dos gestores e auxilia a formulação das políticas de saúde (RIBAS, 2022). O SIS pode redirecionar a alocação de recursos conforme a necessidade de cada região ao longo do tempo, de modo que a demanda seja atendida dentro das restrições dos recursos limitados disponíveis (RIBAS, 2022).

Os sistemas de informação em saúde são um componente fundamental de programas eficazes de Saúde Pública. Eles sustentam sistemas de vigilância e resposta eficazes e eficientes (POORANSINGH, et al., 2021). Em ambientes sem

crise, um SIS eficaz pode ser um recurso crítico para melhorar o funcionamento do sistema de saúde e a resiliência às ameaças à saúde coletiva. Em cenários de crise, os SIS de rotina e de emergência fornecem informações vitais em tempo real para orientar as ações de resposta rápida. Apesar da importância de um SIS robusto na resposta a pandemias, muitos países enfrentam desafios na coordenação e expansão do SIS, no compartilhamento de dados e informações, no combate à desconfiança e na desinformação da população (HOTCHKISS; DIANA; FOREIT, 2012).

No Brasil, o Sistema de Informação de Agravos de Notificação (SINAN) foi desenvolvido no início da década de 1990, tendo como objetivo a coleta e o processamento dos dados sobre doenças e agravos de notificação em todo o território nacional, fornecendo informações para a análise do perfil da morbidade da população brasileira. O sistema foi idealizado para ser a principal fonte de informação para estudar a história natural de um agravo ou doença, estimar a sua magnitude como problema de saúde na população, detectar surtos ou epidemias, bem como elaborar hipóteses epidemiológicas a serem testadas em ensaios específicos (BRASIL, 2022).

A concepção do Sinan foi norteadada pela padronização de conceitos de definição de caso, pela transmissão de dados a partir da organização hierárquica das três esferas de governo, pelo acesso à base de dados necessária à análise epidemiológica, e pela possibilidade de disseminação rápida dos dados gerados na rotina da vigilância epidemiológica no âmbito do Sistema Único de Saúde (SUS) (BRASIL, 2022).

Em 2009, foi implantada a vigilância universal de síndrome respiratória aguda grave (SRAG) no Brasil, em decorrência da pandemia de influenza. Na ocasião, o MS incluiu a notificação de casos de influenza A (H1N1) na rotina dos serviços de vigilância em saúde. Para atender à emergência e agilizar a notificação dos casos, foi desenvolvido o aplicativo Sinan Influenza Web. Esse aplicativo possuía um banco de dados único para ser utilizado por todos os usuários com permissão de acesso nas três esferas de governo. As notificações eram incorporadas à base nacional em tempo real, permitindo que as equipes de vigilância tivessem conhecimento imediato dos casos notificados, possibilitando a intervenção oportuna. Porém a versão do

Sinan Influenza Web foi descontinuada em janeiro de 2019, visto que o registro dos casos de SRAG passou a ser realizado no Sistema de Informação da Vigilância Epidemiológica da Gripe (Sivep-Gripe) (BRASIL, 2022).

No Boletim Epidemiológico de 07 de fevereiro de 2020, o MS descreveu o cenário dos casos de Coronavírus na China e instituiu as atividades de planejamento desenvolvidas pela Secretaria de Vigilância em Saúde. Neste primeiro momento, considerando que os casos de síndrome gripal eram notificados apenas por unidades sentinelas, diante da inexistência de um sistema de informação que contemplasse os dados específicos para a investigação dos casos suspeitos da Covid-19, o MS recomendou que todos os casos fossem notificados por meio do REDCap (RESEARCH ELETRONIC DATA CAPTURE), uma plataforma para coleta, gerenciamento e disseminação de dados de pesquisas. Esta plataforma é a versão para mobile do FormSUS, implantado em 2019, porém, não é um sistema de informação.

A ficha de investigação do REDCap contemplava dados de identificação do paciente, sinais e sintomas, morbidades prévias, dados de exposição e viagens. Não existiam campos a respeito do tipo de amostra laboratorial realizada, tampouco de conclusão e encerramento dos casos (confirmação ou descarte), o que dificultava o monitoramento e análise dos casos (MINISTÉRIO DA SAÚDE, 2020).

Com o avanço da pandemia da Covid-19, em 27 de março de 2020, foi lançado o sistema e-SUS Notifica pelo Departamento de Informática do Sistema Único de Saúde (Datasus), que permitiu a notificação imediata de casos leves e moderados de síndrome gripal (SG) suspeitos e confirmados de covid-19 em todo o território nacional. Trata-se de um sistema online a fim de garantir agilidade no processo de notificação e possibilitar que todos os níveis de gestão, profissionais e unidades de saúde tenham acesso em tempo real às notificações realizadas. Em maio de 2021 foi lançado o módulo de 'Monitoramento de contatos' com objetivo de rastrear e monitorar o maior número possível de contatos de pessoas que estiveram com caso suspeito ou confirmado de Covid-19 no período de transmissibilidade da doença, mesmo que ainda estejam assintomáticas (BRASIL, 2022).

Para fins de monitoramento de casos hospitalizados de COVID-19 no Brasil, o Ministério da Saúde incorporou a testagem do vírus SARS-CoV-2, causador da COVID-19, à vigilância de síndrome respiratória aguda grave (SRAG). A notificação de casos é compulsória e os registros são armazenados no banco de dados informatizado SIVEP-Gripe (Sistema de Informação de Vigilância Epidemiológica da Gripe) (ANEXO 1). Tal sistema teve origem na pandemia de influenza H1N1 em 2009, e é mantido desde então para o monitoramento de casos de SRAG e vigilância de eventos inusitados associados a este agravo de saúde no país (NIQUINI, 2020).

Após o encerramento da ESPIN, a covid-19 deixou de ser considerada um evento de saúde pública com potencial ameaça à saúde pública para sua notificação compulsória. No entanto, manteve-se a vigilância universal da covid-19, por conseguinte, em 13 de maio de 2022 foi publicada a Portaria GM/MS nº 1.102, que incluiu a SG suspeita de covid-19, a covid-19, a Srag por SARS-CoV-2, a Síndrome Inflamatória Multissistêmica Pediátrica (SIM-P) e a Síndrome Inflamatória Multissistêmica em Adultos (SIM-A) na Lista Nacional de Notificação Compulsória de doenças, agravos e eventos, exigindo sua notificação compulsória e imediata no sistema de informação correspondente. (BRASIL, 2024).

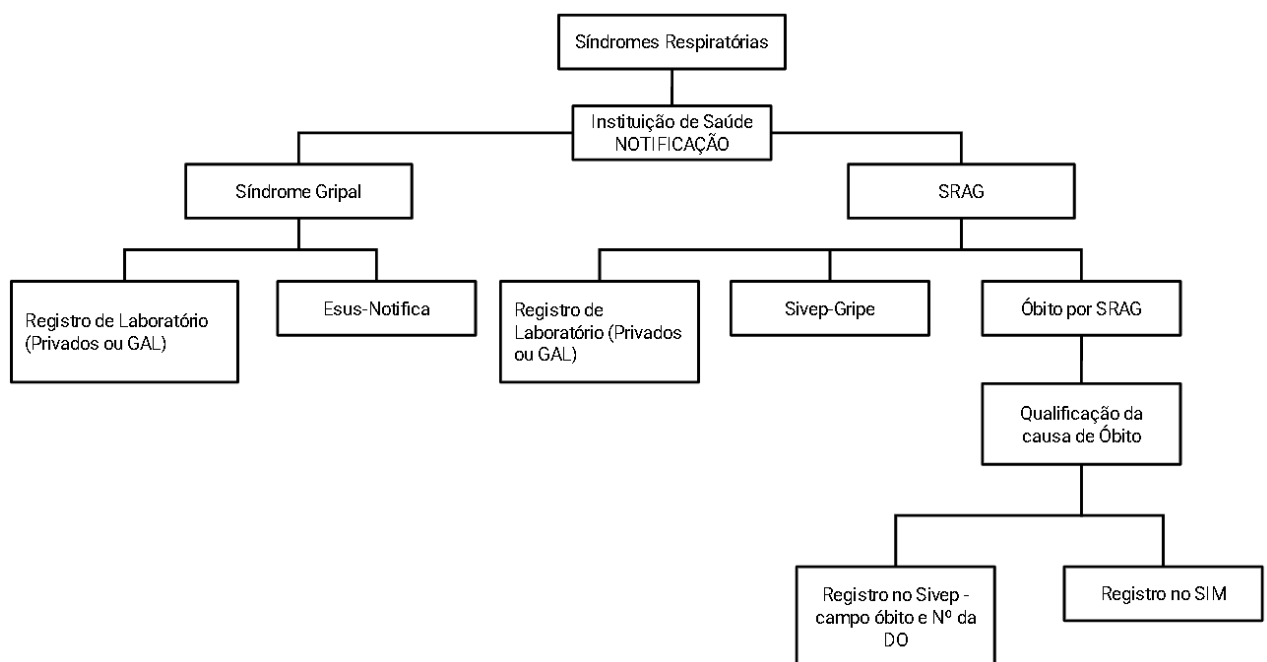
Trajétoria dos dados de Síndrome Respiratória

Com o objetivo de tornar mais clara a dinâmica de circulação das informações, apresenta-se a seguir um fluxograma que descreve a trajetória dos dados de COVID-19 no Sistema de Vigilância em Saúde. O diagrama sintetiza as etapas de registro, os sistemas envolvidos e os pontos de interligação entre as bases, permitindo visualizar de forma integrada o percurso percorrido pelos casos desde a notificação inicial até sua consolidação nos Sistemas de Informação em Saúde.

Todos os casos que atendam à definição de SG e de SRAG hospitalizado (de qualquer etiologia) e óbito por SRAG, independentemente de hospitalização, devem ser notificados e registrados nos sistemas de informação adequados, conforme a Figura 1.

Os casos de SG (leves) são registrados no ESUS-Notifica, enquanto as SRAG são registradas no Sivep-Gripe. Os óbitos por SRAG, independentemente de hospitalização, devem ser registrados no Sivep-Gripe e também deve ocorrer, obrigatoriamente, no Sistema de Informações sobre Mortalidade (SIM).

Figura 1: Fluxograma de Registro dos casos de SG e SRAG



Fonte: Elaborado pelos autores, 2025.

3.5 Desafios nos Registros de Covid-19 no Brasil

No âmbito dos Sistemas de Informação em Saúde (SIS), os registros, processamento e análise de dados, ainda enfrentam desafios. Segundo Melo e Soares (2018), existem dificuldades relacionadas aos SIS, como a ausência de uma política de incentivo ao registro correto dos dados, dificultando a geração de informações corretas e, conseqüentemente, a realização de ações eficazes. Acrescenta-se ainda, a carência de um treinamento sistemático dos profissionais responsáveis por essas atividades.

O direito de acesso à informação é um dos mais importantes direitos para a promoção da democracia. Durante a pandemia de Covid-19, surgiram evidências concretas do impacto da falta de disponibilidade de dados, como por exemplo, no início de junho de 2020, o painel Coronavírus, que é a principal plataforma oficial de monitoramento de dados da COVID-19 no Brasil, permaneceu indisponível por motivos políticos. O governo nacional sugeriu mudanças arbitrárias nas medições de variáveis e padrões de relatórios que eram contrários às diretrizes epidemiológicas internacionais. Depois de vários dias sem atualizações, o sistema retornou sem estimativas de incidência e mortalidade, além de apresentar inconsistências sistemáticas como casos negativos e erro de medição. O “apagão de dados” brasileiro ganhou visibilidade internacional quando a Universidade John Hopkins ameaçou retirar o Brasil do conjunto de dados mundial. Uma semana depois, o Supremo Tribunal Federal decidiu que o Ministério da Saúde brasileiro deveria se adequar às normas da Organização Mundial da Saúde e divulgar dados completos (SILVA; FIGUEIREDO, 2021).

Novamente, uma interrupção ocorreu no dia 10 de dezembro de 2021, e o restabelecimento dos sistemas e acesso ao público foi anunciado pelo Ministério da Saúde somente em 12 de janeiro de 2022, portanto com indisponibilidade por ao menos 30 dias. Nesse período, o mundo já havia identificado uma variante de preocupação do SARS-CoV-2, a Ômicron, com alta taxa de transmissibilidade. O episódio do “apagão de dados”, em que um ataque cibernético deixou sistemas do Ministério da Saúde fora do ar, impactou na falta de análises e na ausência de indicadores (VILLELA; GOMES, 2022).

Dentre os desafios dos registros de Covid nos SIS, destaca-se a falta de integração entre diferentes sistemas de informação existentes, o que inviabiliza a integração de informações de diferentes fontes, como por exemplo, o Sistema Gerenciador de Ambiente Laboratorial (GAL), o Esus-Notifica e o Sivep-Gripe. Essa fragmentação, impossibilita o acompanhamento dos resultados dos exames em tempo real, atrasa a detecção precoce do caso, além de contribuir para falhas no preenchimento das informações, que muitas vezes precisam ser redigitadas nos diferentes sistemas, gerando duplicidades que demandam um grande trabalho para limpeza dos dados e consequentemente, maiores dificuldades de análise (LANA et al., 2020).

3.6 Qualidade dos Dados

A informação epidemiológica é produzida a partir de dados. Os dados epidemiológicos são produzidos com a finalidade de descrever, acompanhar e comparar características de populações, grupos de indivíduos e coletividades no que afeta a saúde, o bem-estar e a qualidade de vida, bem como determinantes da ocorrência e distribuição dos eventos de saúde. A estrutura do dado em Epidemiologia constitui a base sobre a qual se formulam adequadamente objetivos, hipóteses e metodologia dos estudos (ALMEIDA FILHO; BARRETO, 2011).

Há inúmeros dados que interessam ao raciocínio epidemiológico, dirigidos à descrição e entendimento de problemas e situações de saúde, que resultam da aplicação de amplo espectro de métodos das ciências da saúde, das ciências humanas e sociais e das ciências biológicas, entre outras. Uma vez que sejam analisados, têm seu significado ampliado sob a forma de informação qualificada. A análise e a interpretação agregam valor, contextualizam e referenciam os dados, progressivamente, na extensão em que avançam, e deles se extrai uma representação mais completa e significativa da realidade. É por essa razão que a qualidade dos dados é tão essencial à produção de informações (ALMEIDA FILHO; BARRETO, 2011).

A apropriação dessas informações pelos gestores e pela sociedade organizada para subsidiar suas atividades deve considerar seus pontos fortes e suas limitações, resultado de avaliações regulares e sistemáticas dos dados

disponibilizados. A garantia de uma informação de qualidade é condição essencial para a análise objetiva da situação sanitária, para a tomada de decisões baseadas em evidências e para a programação de ações de saúde (LIMA et al., 2009).

A partir de uma perspectiva científica, a qualidade dos dados vem sendo abordada em diferentes áreas do conhecimento, como por exemplo, a ciência da informação, a computação, a gestão entre outras. A definição teórica para o termo “qualidade dos dados” ainda não obteve consenso, sendo este considerado uma categoria multidimensional, visto como um conceito multifacetado (LIMA et al., 2009). A qualidade dos dados é entendida como “aptidão para uso” (Juran et al. 1974) ou estar “adequado para o uso”. Entretanto, uma informação pode ser considerada de qualidade e adequada para um indivíduo em um dado momento e inadequada e dispensável em outras tarefas. Ainda, um segundo usuário pode considerar a qualidade da mesma informação inapropriada para suas tarefas. Assim, a qualidade da informação é estabelecida a partir do ponto de vista do usuário e a escolha dos indicadores para métricas dependerá da valoração das dimensões da informação em determinado contexto de aplicação (WANG; STRONG, 1996; NAUMANN, 2002; MOHAMMADIET et al., 2017; CICHY; RASS, 2020; HÄRTING; LEWONIEWSKI, 2020).

Para serem realmente valiosos, tanto dados como informações devem possuir várias características essenciais, como ser completos, confiável, oportuno e adequados ao propósito para o qual serão utilizados. Essas características se traduzem em indicadores e dimensões que são utilizadas para avaliar a qualidade de dados e informações (Turban et al. 2004).

Embora não exista consenso sobre definições para qualidade e seus parâmetros no campo da informação, os atributos mais comumente arrolados na avaliação da qualidade dos dados incluem: **confiabilidade, completude ou completitude e oportunidade**. Outros elementos devem ser considerados, ou seja, os dados precisam estar acessíveis, válidos, precisos, coerentes e rastreáveis. A **confiabilidade** significa credibilidade no conteúdo e na fonte da informação (PAIM; NEHMY; GUIMARÃES, 1996). A completude refere-se ao grau em que os registros possuem valores não nulos.

Segundo Lima et al. (2009), a **oportunidade** é o grau em que os dados ou informações estão disponíveis no local e no tempo, para a utilização por quem deles necessita. Na Vigilância em Saúde, o termo se refere à agilidade do sistema em cumprir todas as suas etapas, desde a notificação do caso, passando pela análise dos dados, interpretação da informação e até a difusão da informação, por meio de boletins epidemiológicos, subsidiando ações de prevenção e controle. A oportunidade pode ser analisada segundo os aspectos (ALMEIDA; BARRETO, 2011):

- Oportunidade de notificação: é o intervalo de tempo (em dias) entre a data da notificação e a data dos primeiros sintomas.
- Oportunidade de investigação: intervalo de tempo (em dias) entre a data da investigação e a data da notificação (início da investigação do caso).
- Oportunidade de digitação: intervalo de tempo (em dias) entre a data da digitação e a data da notificação (digitação do caso no sistema).
- Oportunidade de encerramento: intervalo de tempo (em dias) entre a data de encerramento e a data da notificação (encerramento do caso).

3.7 Qualidade de dados na pandemia

A pandemia exacerbou as desigualdades existentes nos sistemas de saúde e as deficiências existentes nos SIS em todo o mundo, tanto dentro dos países quanto entre países (SCHMIDT; ABBOUD; BOGAERT, 2021).

A resposta operacional à pandemia de COVID-19 exigiu a rápida adaptação e alavancagem das capacidades dos SIS existentes para recolher, transmitir e analisar dados em tempo real que permitissem compreender a situação epidemiológica e elaborar medidas de controle adequadas. Devido à natureza sem precedentes e a gravidade da pandemia, as capacidades dos SIS em muitos países foram sobrecarregadas pelas demandas de informações e pelos desafios encontrados. Múltiplas lacunas tecnológicas foram expostas, especialmente em países de baixa e média renda. (MAHMOOD et al., 2020).

Os desafios iniciais variaram de novas demandas aos sistemas de saúde, que já estavam sobrecarregados, até a urgência em determinar como documentar com eficácia os processos contínuos da COVID-19 em bancos de dados (NEGRO-CALDUCH et al., 2021). Entretanto, a pandemia forneceu uma oportunidade para identificar os pontos fortes e fracos dos SIS existentes no contexto de uma emergência de saúde global.

Uma pesquisa realizada com todos os países integrantes da Iniciativa Europeia de Informações sobre Saúde da OMS avaliou as experiências relacionadas ao desempenho dos sistemas de informação em saúde (SIS) e as lições aprendidas durante a pandemia de COVID-19. Enquanto alguns países já dispunham de estruturas mais desenvolvidas de informática em saúde, outros apresentavam limitações significativas de infraestrutura, insuficientes para atender às demandas impostas pela crise. A pandemia não apenas agravou desigualdades preexistentes nos SIS, como também evidenciou a necessidade de combater infodemias e de estabelecer estratégias eficazes de comunicação e engajamento da população, de modo a fortalecer a confiança nas autoridades de saúde pública e nas informações por elas divulgadas (MASSOUDI; SOBOLEVSKAIA, 2021).

O estudo concluiu que as principais lacunas reveladas incluíram serviços de saúde pública com poucos recursos, tecnologias de informação obsoletas e a ausência de interoperabilidade para garantir o fluxo contínuo de dados. Ressalta-se que a existência de bases de dados de saúde de qualidade é um dos indicadores de maturidade dos sistemas de saúde, ao lado da cobertura universal e da efetividade dos serviços sociais e comunitários. Em nível global, as limitações identificadas abrangeram falta de padronização, baixa eficiência na coleta de informações, dificuldades em assegurar a qualidade dos dados, ausência de compartilhamento oportuno, incompletude, indisponibilidade de registros e inexistência de políticas voltadas ao intercâmbio eficiente de dados (MASSOUDI; SOBOLEVSKAIA, 2021).

Informações sólidas, confiáveis e oportunas são a base da tomada de decisões em todos os blocos de construção do sistema de saúde e são essenciais para o desenvolvimento e implementação de políticas do sistema de saúde, governança e regulamentação, pesquisa em saúde, desenvolvimento de recursos humanos, educação e treinamento em saúde e prestação de serviços e

financiamento. As informações têm pouco valor se não estiverem disponíveis em formatos que atendam às necessidades dos formadores de políticas, gestores, profissionais de saúde e comunidade (CHAUHAN; US; SINGH, 2021).

Abordagens estatísticas podem ser aplicadas para avaliar a confiabilidade dos números reportados. Entre elas, destaca-se a Lei de Benford, que descreve a distribuição esperada dos dígitos iniciais em um conjunto de dados numéricos. Essa ferramenta tem sido amplamente utilizada para detectar possíveis inconsistências ou anomalias em diferentes contextos, como processos eleitorais, estatísticas econômicas, relatórios financeiros e contábeis, além de já ter sido aplicada à análise dos registros de casos de COVID-19 (CAMPOLIETI, 2022).

Como os sistemas de informação são particularmente importantes durante emergências de saúde pública, como a pandemia de COVID-19, eles precisam ser constantemente avaliados e ajustados. Nesse sentido, o uso da Lei de Newcomb-Benford (1881, 1938) pode ser uma ferramenta útil devido à sua abordagem rápida e objetiva. Além disso, sua facilidade de aplicação na análise da confiabilidade de dados.

3.8 Lei de Benford

A Lei de Benford (LB) foi originalmente descrita pelo astrônomo e matemático canadense-americano Simon Newcomb, em 1881, quando este percebeu que as primeiras páginas das tabelas de logaritmos (utilizadas para cálculos na época), estavam mais desgastadas do que as últimas, indicando que o valor mais acessado era o número 1, isto é, a frequência de uso diminuía à medida que os números se aproximavam de 9. Assim, Newcomb (1881) foi bem-sucedido em estabelecer uma relação de probabilidade de ocorrência do primeiro e segundo dígitos associados ao uso das tabelas de logaritmos.

Como Newcomb não reuniu dados numéricos ou outras evidências de sua descoberta, permaneceu no esquecimento por vários anos até que o físico e engenheiro eletricitista norte-americano Frank Benford (1938), encontrou a mesma regularidade nos números. Benford publicou um artigo em 1938, denominado *The Law Anomalous Numbers*, em que comprovou a lei usando dados coletados de diversas fontes distintas como comprimento dos rios, tamanho de população,

constantes científicas, números recolhidos nas primeiras páginas de jornais e taxas de mortalidade.

A lei de Benford, também chamada de “lei dos primeiros dígitos”, afirma que os primeiros dígitos dos numerais aderem a um determinado padrão logarítmico, ou seja, para um determinado conjunto de números, aqueles cujo primeiro dígito é 1, aparecerão com mais frequência, seguindo na ordem de 1 a 9 (BENFORD, 1938).

A Lei de Benford é dada por $P(n) = \log_{10} \left(1 + \frac{1}{n} \right)$ em que $n \in [1,9]$ e $P(n)$ referem-se à probabilidade de ocorrência do primeiro dígito n em um universo de números aleatórios, conforme a tabela 1:

Tabela 1: Distribuição de frequência do primeiro dígito de acordo com BL.

Dígito(n)	Probabilidade do Primeiro Dígito P(n)	Percentual
1	0,301	30,1%
2	0,176	17,6%
3	0,125	12,5%
4	0,097	9,7%
5	0,079	7,9%
6	0,067	6,7%
7	0,058	5,8%
8	0,051	5,1%
9	0,046	4,6%

Fonte: Adaptado de Benford (1938,p.556)

A LB é uma prática comum em estudos forenses dentro das ciências sociais e tem sido aplicada em várias disciplinas, por exemplo, finanças e contabilidade, política, medicina e pandemias. A conformidade com a Lei de Benford exige que tenhamos um conjunto de dados com números que tenham pelo menos quatro dígitos. As simulações mostraram que os números devem ter quatro ou mais dígitos para um bom ajuste. No entanto, se não houver muitos números de dois e três dígitos misturados com números maiores, o viés não será suficiente para merecer um ajuste nas frequências esperadas de dígitos (NIGRINI, 2020).

A regra geral é que o conjunto de dados deve ter pelo menos 1.000 registros para esperarmos uma boa conformidade com a Lei de Benford. Para tabelas com menos de 1.000 registros, os testes relacionados a Benford ainda podem ser executados, mas podem apresentar desvios maiores. Outra regra geral é não testar as frequências dos primeiros dois dígitos de conjuntos de dados com menos de 300 registros. O teste do primeiro dígito (com todas as suas falhas) deve ser usado em pequenos conjuntos de dados (NIGRINI, 2020).

Uma propriedade da LB é que os dados manipulados ou fraudulentos se desviam significativamente da distribuição teórica da LB. Devido à facilidade de sua aplicação e abordagem direta, a lei tem sido amplamente utilizada para detectar fraudes e manipulações de dados (BALASHOV; YAN; ZHU, 2021).

3.8.1 Testes associados à Lei de Benford

Diversos testes podem ser aplicados para se avaliar a conformidade de um banco de dados com a Lei de Benford: Teste do Primeiro Dígito, Teste do segundo dígito, Teste dos Dois Primeiros Dígitos e Teste da Soma (NIGRINI, 2012).

Teste do primeiro dígito: compara as frequências dos primeiros dígitos dos itens do banco de dados com as frequências estabelecidas pela LB. Esse teste é uma primeira averiguação geral e serve primariamente para identificar se em um conjunto de dados pode ter havido desvios. É um teste mais geral pois a amostra é dividida em apenas 9 grandes grupos (primeiros dígitos de 1 a 9). Esse teste consiste na análise da conformidade do primeiro dígito dos dados analisados com os resultados esperados pela LB (NIGRINI, 2020).

Teste do segundo dígito: Esse teste consiste na análise do segundo dígito dos dados analisados com a BL. O teste do segundo dígito compara as frequências encontradas para os dígitos de 0 a 9 na posição de segundo dígito. Da mesma forma que o teste do primeiro dígito, seus resultados apontam para uma visão geral dos dados, à diferença do teste anterior, no entanto, permite identificar possíveis vieses e redundâncias nos dados. Conforme Nigrini (2012) a frequência esperada para o segundo dígito se dá como segue: $P(D2 = d2) = \sum_{d1=1}^9 \log 9 (1 + \frac{1}{d1d2})$, $d2 = \{0,1,2,... 9\}$ Onde: $d1$ = primeiro dígito e $d2$ = segundo dígito.

Teste dos dois primeiros dígitos (ou teste de primeira ordem): O teste dos dois primeiros dígitos busca averiguar a distribuição dos dois primeiros dígitos dos números sob análise com a distribuição esperada da LB, sendo mais específico que o primeiro teste, pois agora a amostra é dividida em 90 grupos (compostos pelos dois primeiros dígitos 10 a 99). Esse teste já traz um refinamento mais aprofundado dos dados e permite a identificação de alterações e manipulações nos números, como arredondamentos, reduções e aproximações, com maior confiança estatística, sendo esta a razão pela qual sugere-se que todos os valores do banco de dados sob análise inferiores a 10 sejam eliminados do exame, uma vez que não apresentam um segundo dígito inteiro. Nigrini sugere que esse teste é a melhor alternativa dos testes primários, pois permite ao auditor reduzir a amostra de dados sujeitos a averiguação aprofundada. O teste do segundo dígito é bom para detectar arredondamentos excessivos na base de dados, se for constatada a presença excessiva de dígitos 0 e 5. A distribuição esperada para os dois primeiros dígitos se dá como segue: $P(D1D2 = d1d2) = \log(1 + 1/d1d2)$, $d1d2 \in \{10, 11, 12, \dots, 99\}$. Onde $D1D2$ = dois primeiros dígitos.

Teste da Soma: O teste da soma é um teste avançado proposto por Nigrini (2012) para detectar números excessivamente grandes na base de dados. Trata-se de uma modificação do teste usual para os dois primeiros dígitos, no qual os dados são divididos em 90 categorias (10, 11, ..., 98, 99), e as observações são somadas para cada grupo. O resultado esperado é que cada classe de dígitos some aproximadamente o mesmo montante, isto é, $= 0,011$ ou $1,1\%$ do total da soma de todos os números da base.

3.8.2 Testes estatísticos

Teste Qui-quadrado (χ^2): é usado para comparar um conjunto de resultados reais com os resultados esperados, para testar se existe uma relação entre variáveis distribuídas em categorias (porcentagem por dígito, por exemplo). A hipótese nula (H_0) considerada é quando a frequência observada é igual a frequência esperada e a hipótese alternativa (H_1) refere-se a ocorrência da frequência observadas diferente da frequência esperada. O nível de significância α mensura o risco de se rejeitar uma hipótese verdadeira e vice-versa. Geralmente fixado em 5% ($P=0.05$). O valor

de χ^2 adotando o nível de significância α , é denominado valor crítico. Graus de Liberdade é o número de valores independentes (dimensão da amostra) menos o número de parâmetros matemáticos a serem considerados. Considerando o teste do primeiro dígito, têm-se 9 como número de valores independentes, com isso, chega-se a 8 graus de liberdade. O limite do valor crítico do χ^2 estabelecido é de 15,507 para 8 graus de liberdade e 5% de significância. Portanto, se χ^2 calculado $\geq \chi^2$ crítico, rejeita-se a hipótese nula. Se χ^2 calculado $< \chi^2$ crítico, aceita-se hipótese nula. A hipótese nula é de que os dígitos de uma distribuição se conformam com a Lei de Benford.

Desvio Médio Absoluto (DMA): o Desvio Absoluto Médio, o qual permite avaliar, a partir dos valores críticos definidos TABELA 2, o quanto, na média, essas probabilidades observadas divergiram das esperadas. Sua fórmula é: $DMA = \sum |PR - PE|$ k $i=1$ k Onde k é o número de séries dos dígitos analisados (09 para o teste do primeiro dígito e 90 para o dos dois primeiros dígitos). Segundo Nigrini (2012), a Média dos Desvios Absolutos é, dentre os métodos explicitados, o mais justo na avaliação da conformidade com a Lei de Benford, uma vez que não leva em consideração o número de elementos da amostra. Enquanto o Teste Z, Qui-Quadrado e Kolmogorov-Smirnoff aumentam o rigor à medida que n cresce, o MDA não se altera. Não há, todavia, valor crítico estatisticamente definido para esse índice. Nigrini (2012) propõe o seguinte critério para avaliar a conformidade da base de dados de acordo com o teste DMA:

Tabela 2: Análise do Grau de Conformidade da Lei de Benford

Dígitos	DMA	Grau de Conformidade
Primeiro Dígito	0, 000 a 0,006	Conformidade estrita
	0, 006 a 0,012	Conformidade Aceitável
	0, 012 a 0,015	Conformidade marginalmente aceitável
	Maior 0,015	Não Conformidade
Segundo Dígito	0, 000 a 0,008	Conformidade estrita
	0, 008 a 0,010	Conformidade Aceitável
	0, 010 a 0,012	Conformidade marginalmente aceitável
	Maior 0,012	Não Conformidade

Dois Primeiros Dígitos	0, 000 a 0,0012	Conformidade estrita
	0, 0012 a 0,0018	Conformidade Aceitável
	0, 0018 a 0,0022	Conformidade marginalmente aceitável
	Maior 0,0022	Não Conformidade

Fonte: Adaptado de Nigrini (2012).

3.9 Estudos utilizando a Lei de Benford

Em um contexto mais recente, diversos estudos foram realizados para identificar possíveis desvios dos dígitos em relação à distribuição de Benford. Nigrini (1992, 2000, 2012), partindo do pressuposto de que dados contábeis fidedignos seguissem distribuição de Benford. O autor desenvolveu vários testes para mensurar a conformidade com a Lei de Benford, e foram detectadas fraudes em sete companhias de Nova York pelo escritório da Procuradoria do Brooklyn usando esses testes. Com base nesses sucessos, Nigrini foi chamado a dar consultoria para diversos países. Rauch et. al, (2011) analisaram a conformidade com a Lei de Benford dos primeiros dígitos de dados macroeconômicos reportado ao Gabinete de Estatísticas da União Europeia (Eurostat) pelos países membros da União Europeia (UE). Construíram um ranking dos 27 países membros de acordo com a extensão do desvio encontrado com relação às previsões da Lei. Mebane (2009), um estatístico americano da Universidade de Michigan, estudou dados eleitorais de vários países, incluindo os Estados Unidos, Rússia e México, aplicando a LB.

No contexto de obras públicas, essa ferramenta foi utilizada para analisar as obras relevantes no contexto da Copa do Mundo de 2014, tais como a reforma do Estádio Maracanã, a construção da Arena da Amazônia e a reforma do Aeroporto Internacional de Minas Gerais (Cunha, 2013).

Subnotificação

A utilização dos testes para verificação da conformidade com a Lei de Benford pode indicar um método eficiente para investigar casos de subnotificação e supernotificação de registros da doença COVID-19. A subnotificação pode ser entendida como as várias formas pelas quais os sistemas de vigilância falham ou

não conseguem refletir todas as infecções em uma determinada população. É o número estimado de infecções ocorridas em uma população que não foi captada pelo sistema de vigilância para cada caso notificado em um determinado período de tempo. As subnotificações podem ocorrer em decorrência da não procura ao atendimento médico por sintomas leves e/ou por saber que a doença é autolimitada ou por outro motivo, e uma fração assintomática que não procura atendimento por estar não estão cientes de seu estado de infecção devido à falta de sintomas e, portanto, não podem ser captadas pelos sistemas de vigilância. A subnotificação pode mascarar a verdadeira magnitude da incidência da doença e reduz a eficiência do sistema de notificação e o potencial de vigilância (GIBBONS, et al., 2014).

O número de casos confirmados de Covid-19 é o dado mais importante para que se possa compreender a evolução dessa doença. Contudo, a rápida disseminação da pandemia e o pequeno número de testes realizados tornam difícil estimar o número real de casos e provocam subnotificação em diferentes países. A restrição dos testes compromete o monitoramento da progressão da pandemia, o planejamento de recursos e a avaliação da eficácia das medidas de controle, assim como a comparação com outras regiões e países. Além disso, esse fato pode levar a falsas conclusões de que a doença está sob controle (PRADO, 2020).

Prado (2020), avaliou a subnotificação de casos confirmados de COVID-19 no Brasil e seus estados. Os resultados indicaram que as notificações de casos confirmados no Brasil representaram apenas 9,2% (IC95% 8,8% - 9,5%) dos números reais, muito menos do que o que se observou em outros países (PRADO, 2020).

3.10 Estudos sobre Análise da Qualidade de dados de Covid-19

Com o objetivo de identificar estudos sobre a qualidade de dados de Covid-19, foi realizada uma revisão sistematizada de literatura. Para tal, foi elaborado um protocolo de revisão sistematizada (Apêndice 1), com definição dos objetivos, critérios de seleção dos estudos, bases eletrônicas de busca, descritores, estratégias de busca e período de seleção. O protocolo foi desenvolvido com auxílio de uma bibliotecária da Universidade Federal de Santa Catarina especializada na área da saúde. Como critérios de seleção foram definidos artigos

que apresentassem em seus objetivos analisar qualidade de dados de Covid-19. Não foram incluídos livros, capítulos de livros, monografias, dissertações e teses. A busca foi realizada inicialmente em novembro de 2022. Foram incorporados os artigos originais publicados a partir de 2020. As bases bibliográficas pesquisadas foram National Institutes of Health US National Library of Medicine (PubMed), Medical Literature Analysis and Retrieval System Online (MEDLINE), a Literatura Latino Americana e do Caribe em Ciências da Saúde (LILACS), Scientific Electronic Library Online (SciELO), Web of Science e Scopus. Os descritores e as chaves de busca foram construídos a partir dos termos encontrados nos Descritores em Ciências da Saúde (DeCS) e no Medical Subject Headings (MeSH). Após a seleção dos termos, foram utilizados os operadores booleanos 'AND' e 'OR', para refinamento da busca, conforme disposto no Quadro 1.

Quadro 1 - Estratégias de buscas utilizadas para levantamento da revisão de literatura.

Bases Bibliográficas eletrônicas	Estratégia de busca	Referências Obtidas
PubMed/ Medline	("Coronavirus Infections"[Mesh] OR "Coronavirus Infections" OR "COVID-19"[Mesh] OR "Covid-19" OR "SARS-CoV-2"[Mesh] OR "SARS-CoV-2" OR "SARS Virus"[Mesh] OR "SARS Virus" OR "SARSCoV2" OR "SARS2" OR "COVID19" OR "COVID-2019" OR "COVID 2019" OR "SARS COV 2" OR "2019-nCoV" OR "2019ncov" OR "coronavírus" OR "nCoV 2019") AND ("Data Analysis"[Mesh] OR "Data Analysis" OR "Data Analyses" OR "Data Accuracy"[Mesh] OR "Data Accuracy" OR "Data Accuracies" OR "Data Qualities" OR "Data Quality" OR "Data Management"[Mesh] OR "Data	347

	Management" OR "Data Administration" OR "Newcomb-Benford" OR "Benford")	
Scopus	("Coronavirus Infections" OR "Covid-19" OR "SARS-CoV-2" OR "SARS Virus" OR "SARSCoV2" OR "SARS2" OR "COVID19" OR "COVID-2019" OR "COVID 2019" OR "SARS COV 2" OR "2019-nCoV" OR "2019ncov" OR "coronavírus" OR "nCoV 2019") AND ("Data Analysis" OR "Data Analyses" OR "Data Accuracy" OR "Data Accuracies" OR "Data Qualities" OR "Data Quality" OR "Data Management" OR "Data Administration" OR "Newcomb- Benford" OR "Benford")	26

SciELO	("Data Analysis" OR "Data Analyses" OR "Data Accuracy" OR "Data Accuracies" OR "Data Qualities" OR "Data Quality" OR "Data Management" OR "Data Administration" OR "Newcomb-Benford" OR "Benford" OR "Análise de Dados" OR "Análise dos Dados" OR "Análises de Dados" OR "Análises dos Dados" OR "Confiabilidade dos Dados" OR "Acurácia dos Dados" OR "Exatidão dos Dados" OR "Precisão dos Dados" OR "Qualidade dos Dados" OR "Gerenciamento de Dados" OR "Administração de Dados" OR "Gestão de Dados" OR "Análisis de Datos" OR "Análisis de los Datos" OR "Exactitud de los Datos" OR "Calidad de los Datos" OR "Manejo de Datos" OR "gestión de datos" OR "administración de datos") AND ("Coronavirus Infections" OR "Covid-19" OR "SARS-CoV-2" OR "SARS Virus" OR "SARSCoV2" OR "SARS2" OR "COVID19" OR "COVID-2019" OR "COVID 2019" OR "SARS COV 2" OR "2019-nCoV" OR "2019ncov" OR "coronavírus" OR "nCoV 2019" OR "Infecções por Coronavirus" OR "Vírus da SARS" OR "Infecciones por Coronavirus" OR "Virus del SRAS")	176
--------	---	-----

Web of Science	("Coronavirus Infections"[Mesh] OR "Coronavirus Infections" OR "COVID- 19"[Mesh] OR "Covid-19" OR "SARS-CoV- 2"[Mesh] OR "SARS-CoV-2" OR "SARS Virus"[Mesh] OR "SARS Virus" OR "SARSCoV2" OR "SARS2" OR "COVID19" OR "COVID-2019" OR "COVID 2019" OR "SARS COV 2" OR "2019-nCoV" OR "2019ncov" OR "coronavírus" OR "nCoV 2019") AND ("Data Analysis"[Mesh] OR "Data Analysis" OR "Data Analyses" OR "Data Accuracy"[Mesh] OR "Data Accuracy" OR "Data Accuracies" OR "Data Qualities" OR "Data Quality" OR "Data Management"[Mesh] OR "Data Management" OR "Data Administration" OR "Newcomb-Benford" OR "Benford") AND ("Pandemics")	258
----------------	--	-----

LILACS	("Data Analysis" OR "Data Analyses" OR "Data Accuracy" OR "Data Accuracies" OR "Data Qualities" OR "Data Quality" OR "Data Management" OR "Data Administration" OR "Newcomb-Benford" OR "Análise de Dados" OR "Análise dos Dados" OR "Análises de Dados" OR "Análises dos Dados" OR "Confiabilidade dos Dados" OR "Acurácia dos Dados" OR "Exatidão dos Dados" OR "Precisão dos Dados" OR "Qualidade dos Dados" OR "Gerenciamento de Dados" OR "Administração de Dados" OR "Gestão de Dados" OR "Análisis de Datos" OR "Análisis de los Datos" OR "Exactitud de los Datos" OR "Calidad de los Datos" OR "Manejo de Datos" OR "gestión de datos" OR "administración de datos") AND ("Coronavirus Infections" OR "Covid-19" OR "SARS-CoV-2" OR "SARS Virus" OR "SARSCoV2" OR "SARS2" OR "COVID19" OR "COVID-2019" OR "COVID 2019" OR "SARS COV 2" OR "2019-nCoV" OR "2019ncov" OR "coronavírus" OR "nCoV 2019" OR "Infecções por Coronavirus" OR "Vírus da SARS" OR "Infecciones por Coronavirus" OR "Virus del SRAS")	31
--------	--	----

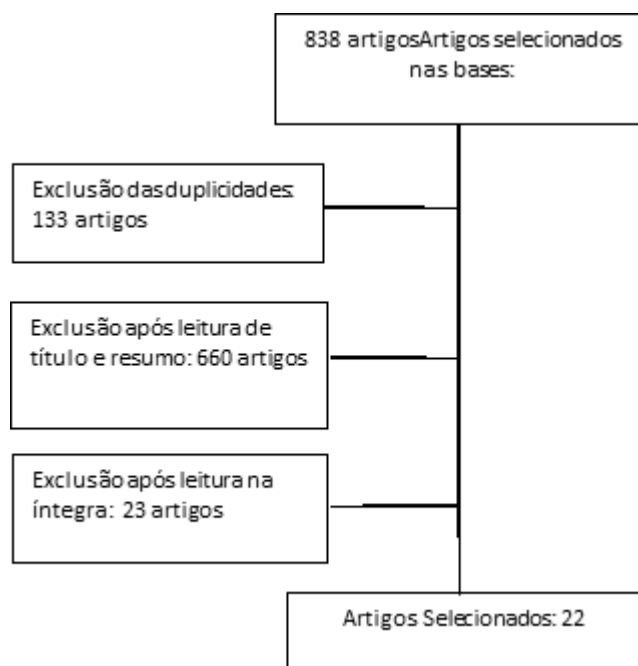
Fonte: Elaborado pela autora (2022).

Os artigos encontrados na busca foram exportados para o gerenciador de referências *Mendeley* para as etapas de exclusão das duplicidades, leitura dos títulos, resumos e leitura na íntegra dos artigos selecionados. Após a seleção final

dos artigos realizou-se a leitura de todas as suas referências a fim de incluir eventuais novos estudos não identificados primeiramente nas bases bibliográficas.

Foram identificados 838 artigos, dos quais foram removidos 133 por duplicidade e por fim, selecionados 22 artigos, conforme figura 1.

Figura 2 - Fluxograma apresentando as etapas da revisão de literatura sistematizada.



Fonte: Elaborado pela autora (2022).

3.10.1 Estudos relacionados à Lei de Benford

Uma das pesquisas realizadas no âmbito mundial, desenvolvida por Balashov et al. (2021), analisou a relação entre a precisão dos dados e os indicadores macroeconômicos para um conjunto de 185 países afetados pela pandemia de COVID-19. Foram utilizados os dados dos casos confirmados e óbitos disponibilizados pela John Hopkins University, entre janeiro e junho de 2020. Para análise, foi utilizada a conformidade com a LB e a análise de regressão para testar associação entre as medidas de adequação e um índice construído a partir de quatro indicadores de desenvolvimento: o *Economist Intelligence Unit Democracy Index*, o produto interno bruto per capita, os gastos com saúde e o Índice de

Cobertura Universal de Saúde. No geral, o estudo concluiu que aproximadamente um terço dos países apresentaram desvios da LB, especialmente para o número de óbitos, e que países com regimes democráticos e economicamente mais desenvolvidos, fornecem dados mais precisos durante as pandemias. Entretanto, os autores alertam sobre a importância de analisar os dados com cautela pois a LB não fornece uma evidência conclusiva, mas normalmente deve ser utilizada como primeiro filtro para detecção de fraudes e os desvios demonstram necessidade de mais investigações (BALASHOV; YAN; ZHU, 2021).

Resultados semelhantes foram encontrados na pesquisa realizada por Moreau (2021), que utilizou a LB para rastrear e detectar inconsistências nos dados de novos casos diários de COVID-19 relatados por 80 países. Dados de 26 países apresentam não conformidade com a lei de Benford (MOREAU, 2021). Outro estudo no âmbito internacional, visou avaliar a confiabilidade dos dados globais do COVID-19 divulgados pelas autoridades locais em 202 países, de dezembro de 2019 a setembro de 2020. A conclusão foi de 69% dos países em todo o mundo apresentam alguns desvios da lei de Benford e cerca de 28% dos países aderiram bem à frequência prevista dos primeiros dígitos. A Alemanha publicou os registros mais confiáveis e o Irã os dados com maiores desvios (FARHADI, 2021).

Ainda em âmbito internacional, uma pesquisa americana utilizou a LB para comparar os dados de Covid dos 10 países com maior taxa de mortalidade, conforme os dados da Universidade Johns Hopkins, no período de janeiro a abril de 2020. O objetivo do estudo foi utilizar um modelo de crescimento epidêmico na presença de intervenções para descrever o potencial de uma curva achatada e, em seguida, investigar se o modelo de crescimento epidêmico seguiu a LB. Entre esses países, a Coreia do Sul mostrou um grau particularmente alto de intervenção de controle. Embora todos os países tenham lutado agressivamente contra a epidemia, a análise mostrou que todos os países, exceto o Japão, satisfizeram a LB, indicando que as taxas de crescimento do COVID-19 estavam próximas de uma tendência exponencial (LEE; HAN; JEONG, 2020).

Com objetivo de avaliar os casos de Covid-19 da China, Itália e EUA, um estudo aplicou a Lei de Benford e os testes qui-quadrado e Kuiper (um teste modificado de Kolmogorov-Smirnov). A pesquisa utilizou os dados do Centro de

Pesquisa de Vírus Corona da Universidade Johns Hopkins para a China, do Centros de Controle de Doenças para os EUA e dos boletins diários do *Dipartimento della Protezione Civile* para a Itália, compreendendo os períodos de janeiro a março de 2020 para China, fevereiro a junho de 2020 para EUA e fevereiro a abril de 2020 para Itália. O estudo evidenciou que os casos da China correspondem à distribuição esperada na Lei de Benford e são semelhantes às observadas nos EUA e na Itália (KOCH; OKAMURA, 2020) .

Outro estudo utilizou a LB para avaliar a qualidade dos dados do sistema de vigilância epidemiológica chinês durante a pandemia, com dados dos relatórios da Organização Mundial da Saúde (OMS) de janeiro a março de 2020. Este estudo corroborou com o estudo de Koch e Okamura (2020), concluindo que a China apresentou qualidade aceitável em seu sistema de vigilância (IDROVO; MANRIQUE-HERNANDEZ, 2020).

Um estudo realizado por Koliass (2022), aplicou a Lei de Benford (LB) aos casos e óbitos de COVID-19 dos países da União Europeia (UE) para detectar possíveis desvios e investigar se possuem associação aos índices de desenvolvimento PIB, gastos com saúde (% do PIB), índice de Cobertura de Serviço (UHC) e taxa de vacinação completa. Os dados analisados foram provenientes do Centro Europeu de Prevenção e Controle de Doenças registrados de março a dezembro de 2021. O estudo concluiu que a Bulgária, Croácia, Lituânia e Romênia estavam de acordo com a distribuição de Benford, entretanto, a Dinamarca, Irlanda e Grécia apresentaram maior desvio. Com relação aos índices de desenvolvimento, os países com alto percentual de vacinação tendem a se desviar da lei de Benford, que pode significar que os países com maior taxa de vacinação também foram associados a menos mortes. As descobertas sugeriram que, em geral, os dados oficiais fornecidos pelas autoridades não confirmam a lei de Benford, mas o autor pontua como limitação que isto não implica que os dados sejam fraudulentos, mas que essa abordagem atua como uma ferramenta preliminar para análise de dados. Sugere estudos com investigação mais aprofundada dos países que apresentaram os maiores desvios (KOLIAS, 2022) .

Na Espanha, o estudo de Morilas-Jurado et al. (2022), utilizou a LB como ferramenta de avaliação de confiabilidade para detectar se os números de mortes

diárias relacionadas à COVID-19 registradas pelo Ministério da Saúde entre março e junho de 2020, pelas diferentes Regiões do país, seguiam a lei. Como qualidade de ajuste da análise, foi utilizado o teste χ^2 (qui-quadrado) e uma análise de sensibilidade para validar os resultados. O estudo identificou desvios e erros nos registros das diferentes regiões, sobretudo nos primeiros meses da pandemia, que podem ter ocorrido por atrasos na comunicação e no registro da informação, erro humano, diferenças nos critérios de registro, ou devido a pandemia seguir uma evolução positiva e o número de óbitos por dia ser muito baixo. Dessa forma, concluíram que a LB pode ser utilizado como ferramenta de auditoria no registro dos dados da COVID-19, e pode ajudar a fornecer dados confiáveis às administrações de saúde em seus diferentes níveis de gestão (MORILLAS-JURADO; CABALLER-TARAZONA; CABALLER-TARAZONA, 2022).

Estudo realizado nos Estados Unidos da América (EUA) objetivou avaliar os relatórios de mortes por COVID-19 durante a primeira onda da pandemia, utilizando o período de março a agosto de 2020, registrados no site do *Center for Disease Control* (CDC). Para avaliação, foi aplicada a LB e as estatísticas DMA (Desvio Médio absoluto) e FD (Fator de distorção) onde valores de DMA > 0,015 indicam não conformidade com a lei e valores < 0,015 mostram algum tipo de conformidade. A pesquisa ressalta que a compatibilidade com a lei de Benford depende da forma como ela é testada, com a evidência mais forte e clara baseada nos critérios DMA. O estudo identificou não conformidade com LB em 47 estados, com apenas 2 estados mostrando algum tipo de conformidade e apontou a subnotificação de mortes por COVID-19 em alguns estados dos EUA. Além disso, concluiu que a lei de Benford pode ser usada mostrou-se útil para avaliar a qualidade dos dados epidemiológicos, como uma ferramenta de diagnóstico preliminar antes de um estudo mais detalhado. A falta de conformidade com a LB pode ser explicada pelo pequeno tamanho da amostra (CAMPOLIETI, 2022).

Outro estudo, desenvolvido por Natashekara (2022), utilizou a Lei de Benford para comparar a distribuição do primeiro dígito dos casos e óbitos de Covid-19 do estado de Kerala e da Índia como um todo, utilizando dados provenientes do governo de Kerala e do PRS Legislative Reforms, entre janeiro de 2020 a maio 2021. Os testes estatísticos desvio absoluto médio (DMA), estatística Z, teste qui-quadrado e Kolmogorov–Smirnov foram utilizados para verificar a conformidade com

a LB. O estudo identificou que a distribuição do primeiro dígito dos casos confirmados está dentro do limite aceitável da LB, no entanto, no que diz respeito às mortes, tanto os dados nacionais como os dados de Kerala não obedecem à distribuição logarítmica de Benford. (NATASHEKARA, 2022).

Na Colômbia, Moreno-Montoya (2020) publicou uma nota metodológica propondo um novo teste para avaliar o cumprimento da lei de Benford para dados com amostras pequenas, pois os métodos convencionais para a avaliação (χ^2 e razão de log-verossimilhança) são controversos quando usados com pequenas mostras. Neste modelo, o estudo assumiu que os dados vêm de uma sequência independente de eventos (ou seja, a ocorrência de qualquer dígito em particular não depende da ocorrência de nenhum dígito anterior) e realizou comparações com testes de Kuiper. O novo teste estatístico possibilita estender o uso da lei de Benford para áreas biomédicas, médicas e de saúde pública com amostras pequenas (MORENO-MONTOYA, 2020).

Outro estudo da Colômbia, avaliou o desempenho do sistema de vigilância colombiano durante os primeiros 50 dias da pandemia de COVID-19. Foram analisados dados publicados entre março e abril de 2020 publicados pelo Institutos Nacionais de Saúde e pela Organização Mundial da Saúde (OMS). A avaliação considerou a qualidade dos dados de acordo com a lei de Benford e aplicação dos testes de verossimilhança logarítmica, qui-quadrado ou testes exatos de Moreno. Na maior parte dos primeiros 50 dias, ocorreu conformidade com a LB, seguida, exceto nos primeiros dias da epidemia. Em geral, o sistema de vigilância em saúde pública colombiano cumpriu a lei de Benford, o que sugere que houve qualidade nos dados (MANRIQUE-HERNÁNDEZ et al., 2020) .

No Brasil, a LB também foi utilizada em uma pesquisa que objetivou avaliar a confiabilidade dos casos e óbitos de COVID-19, provenientes da *John Hopkins University* e *Our World in Data*, registrados entre fevereiro e setembro de 2020. Para as análises, foram aplicados os testes de ajustes (qui-quadrado, desvio médio absoluto e fator de distorção). O estudo encontrou fortes evidências de que os relatórios brasileiros não estão de acordo com as expectativas teóricas da LB. Neste estudo, os autores discutiram sobre o episódio do apagão de dados da Covid-19 do painel Coronavírus, que saiu do ar por motivos políticos e depois de vários dias

retornou com inconsistências. Nesse contexto, salientaram a importância de dados serem confiáveis, completos e disponíveis para o desenvolvimento de políticas de combate à doença (SILVA; FIGUEIREDO FILHO, 2021).

De maneira semelhante, outro estudo buscou evidências sobre a qualidade dos dados do sistema de vigilância em saúde brasileiro, utilizando os dados do site da OMS no período entre fevereiro e agosto de 2020, para avaliar o cumprimento da lei de Benford. Os dados de vigilância brasileira da COVID-19 (casos e mortes cumulativos) foram avaliados de acordo com o quanto eles seguiram a distribuição da lei de Benford usando um teste de razão de verossimilhança logarítmica. Essa rápida avaliação da qualidade dos dados sugeriu que o sistema de vigilância em saúde pública brasileiro apresentou oscilação no cumprimento à LB, tendo desempenho aceitável no início da pandemia e posteriormente uma redução na qualidade dos dados. Os autores discutem que a explicação desses achados corresponde ao período de crise durante a gestão da pandemia, sendo a evidência política mais marcante a mudança de ministros no país (IDROVO; MANRIQUEHERNÁNDEZ; FERNÁNDEZ NIÑO, 2021).

Novamente a LB foi utilizada na avaliação da conformidade das informações sobre o número de casos e óbitos por COVID-19 no Brasil, com auxílio de estatística descritiva e uso de métricas relacionadas ao teste Z e ao desvio médio absoluto, para análise de dados até setembro de 2020. A conclusão foi semelhante aos estudos anteriores, apontando a não conformidade com a LB (CARMO; CANEPPELE; NUNES, 2021). Outro estudo aplicou a LB para analisar os dados de fevereiro a setembro de 2020 do Brasil, nas escalas municipal, estadual e nacional e também obtiveram a não conformidade dos dados (GALVÊAS; BARROS; FUZO, 2021).

Kennedy e Phillip Yam (2020) investigaram a aplicabilidade da lei de Benford e da lei de Zipf aos números nacionais de casos de COVID-19, com o propósito de estabelecer diretrizes que possam subsidiar métodos de detecção de possíveis fraudes em epidemiologia a partir de análises estatísticas formais. Os autores destacam que tais abordagens também podem ser utilizadas na avaliação do desempenho dos sistemas de vigilância em saúde pública. O estudo apresenta argumentos teóricos que justificam a validade dessas leis empíricas nos estágios

iniciais de uma epidemia, apoiados por evidências empíricas preliminares. A partir da análise de dados publicados pela Organização Mundial da Saúde e por diversos governos nacionais, os autores encontraram evidências de que tanto a lei de Benford quanto a lei de Zipf se mostraram amplamente aplicáveis em diferentes países, sendo que os desvios identificados puderam ser explicados a partir do contexto epidemiológico de cada localidade.

3.10.2 Estudos relacionados à completitude dos dados

Outros estudos avaliaram a qualidade de dados utilizando a análise da completitude das informações. Em Portugal, um estudo avaliou a qualidade dos dados do COVID-19 através da análise da integridade e consistência dos dados. Para avaliação da completude dos dados, as informações ausentes foram classificadas como em branco ou ignoradas. Diversas variáveis apresentaram baixo grau de completude. O estudo concluiu que importantes problemas de qualidade dos conjuntos de dados de vigilância podem limitar a usabilidade dos dados para informar boas decisões. Melhorias nos conjuntos de dados foram sugeridas, como por exemplo, simplificação dos processos de entrada de dados, monitoramento constante dos dados e maior treinamento e conscientização dos profissionais de saúde (COSTA-SANTOS et al., 2021).

No Brasil, Maciel et al. (2021) realizaram um estudo com o objetivo de descrever a completude dos dados e avaliar a qualidade do Banco de Dados do Painel COVID-19 no estado do Espírito Santo, no ano de 2020. Foram analisados 27.351 registros de crianças, adolescentes e jovens de 0 a 19 anos, considerando a completude das variáveis, a confirmação da doença e sua evolução. A completude foi entendida como o registro adequado de todas as informações necessárias à inclusão dos casos nos sistemas de informação em saúde. Os resultados evidenciaram fragilidades importantes na qualidade dos dados, que variaram entre classificação ruim e regular para diversas variáveis. Observou-se qualidade regular para o critério de confirmação (16%), ruim para a classificação da doença (44%) e status de notificação (30%) e muito ruim para a evolução dos casos (79%). Entre as variáveis epidemiológicas, destacou-se a variável raça/cor da pele, que apresentou completude regular (17%). Esses achados reforçam os desafios enfrentados pelos

sistemas de informação em saúde durante a pandemia, em especial no que se refere à qualidade e à padronização dos registros.

Liang; Dai; Restaino (2022) realizaram um estudo propondo um modelo de truncamento para estimar as probabilidades de subnotificação de casos infectados em diferentes países. Identificaram que os idosos tinham uma probabilidade significativamente maior de subnotificação do que os mais jovens, que pode ser justificada pois muitos idosos podem ter morrido da doença sem sequer serem diagnosticados. Cingapura tinha a menor probabilidade de subnotificação (apenas cerca de 25%), seguida pelo Japão, Hong Kong e Taiwan. A China continental teve a maior probabilidade de subnotificação (quase 75%). (LIANG; DAI; RESTAINO, 2022).

4 MÉTODOS

4.1 Delineamento do Estudo

Trata-se de um estudo de abordagem quantitativa, analítico, retrospectivo, desenvolvido com dados secundários dos casos confirmados e óbitos por COVID-19.

4.2 Análise da Qualidade de Dados de Covid-19

A análise da qualidade dos dados de Covid-19 foi realizada por meio da análise da **confiabilidade**, da **completude** e da **oportunidade** de notificação, digitação e de encerramento dos casos.

4.2.1 Confiabilidade de dados

Para a análise da confiabilidade dos dados, utilizou-se a conformidade com a Lei de Benford, uma ferramenta útil por sua abordagem rápida e objetiva, além da facilidade de aplicação na análise da confiabilidade de dados (CARMO; CANEPPELE; NUNES, 2021).

População do Estudo

O estudo foi desenvolvido com dados dos casos confirmados e óbitos por COVID-19, disponíveis no painel da Universidade Johns Hopkins, no período entre janeiro de 2020 e dezembro de 2022. O painel interativo on-line foi criado pelo Centro de Ciência e Engenharia de Sistemas (CSSE) da Universidade Johns Hopkins, Baltimore, EUA, para visualizar e rastrear os casos de Covid-19 em tempo real. Os dados exibidos no painel estão disponíveis em: https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv e as fontes referenciadas em uma lista completa em um arquivo “*README*”, no repositório da Universidade, disponível em: <https://github.com/CSSEGISandData/COVID-19/blob/master/README.md> (DONG, Ensheng et al., 2022).

Os dados do Brasil que alimentam o painel da Universidade Johns Hopkins, são provenientes do Painel Coronavírus do Ministério da Saúde, disponíveis em:

https://infoms.saude.gov.br/extensions/covid-19_html/covid-19_html.html. Estes dados são oriundos do sistema e-SUS NOTIFICA, que foi desenvolvido para o registro de casos de Síndrome Gripal suspeitos de COVID-19, e do Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe), que registra os casos da Síndrome Respiratória Aguda Grave (SRAG).

Aplicação Lei de Benford

Primeiramente foi realizada a verificação da distribuição de números diários dos casos e de óbitos por COVID-19, segundo os dados disponíveis no painel da Universidade Johns Hopkins. Conforme Nigrini (2020), a aplicação da LB é recomendada para bases com grande volume de dados, variedade de grandezas e, idealmente com valores compostos por ao menos quatro dígitos. Dessa forma, com o intuito de viabilizar a execução das análises de conformidade, foram selecionados países que apresentaram números expressivos de casos confirmados de Covid-19. Além do Brasil, incluíram-se os cinco países com maior número de registros da doença: Estados Unidos, Índia, França, Alemanha e Japão.

A análise de conformidade com a BL foi conduzida segundo roteiro sugerido por Nigrini (2012). Foram aplicados os Testes do primeiro dígito; teste do segundo dígito, teste dos dois primeiros dígitos (ou teste de primeira ordem) e teste de soma. Adicionalmente, aplicaram-se os testes estatísticos Qui-Quadrado e Média dos Desvios Absolutos (DMA). Adotou-se valor crítico de 15,507 proveniente da tabela estatística da distribuição do qui-quadrado, com 8 graus de liberdade e 5% de significância. O resultados do DMA foram classificados segundo Nigrini (2012): valores entre 0 e 0,006 indicam conformidade estrita; entre 0,006 e 0,012, conformidade aceitável; de 0,012 a 0,015, conformidade marginalmente aceitável e, por fim, acima de 0,015 tem-se não conformidade. A análise considerou o primeiro dígito significativo (ignorado o zero à esquerda).

Para o Brasil, a análise foi desagregada por ondas epidêmicas: primeira (fevereiro a outubro de 2020), segunda (novembro de 2020 até novembro de 2021) e a terceira (a partir de dezembro de 2021), conforme definições do Ministério da Saúde e da FIOCRUZ.

As análises foram conduzidas utilizando o software estatístico R, versão 4.1.2. Para estimar as funções da LB, foi utilizado o pacote `benford.analysis` desenvolvido por Cinelli.

4.2.2 Completude e Oportunidade

População do estudo

A população do estudo foi composta pelos registros de SRAG com classificação final COVID-19, notificados no Brasil, no Sivep-gripe entre 1º de março de 2020 a 31 de dezembro de 2023, segundo calendário epidemiológico. Os dados foram obtidos a partir de fonte de acesso público, não nominal, disponível no site <https://opendatasus.saude.gov.br/dataset/> com atualização até 29 de janeiro de 2024.

Análise da completude

A completude é definida como o grau em que os registros do sistema de informação apresentam valores não nulos. Para esta análise, os campos considerados incompletos são os preenchidos como ignorados e em branco. Define-se **campo obrigatório** como aquele cuja ausência de dado impede a inclusão do registro no sistema, enquanto **campo essencial** é aquele que, embora não seja obrigatório, registra dados necessários para a investigação do caso ou para o cálculo de indicadores epidemiológicos ou operacionais.

Foi analisada a completude das variáveis das fichas de SRAG por Covid-19 no Brasil, notificadas no SIVEP-Gripe. Cada ficha de notificação inclui dados relativos à caracterização sociodemográfica, clínica e de investigação diagnóstica dos casos. Ao longo de 2020, foram realizadas quatro atualizações no sistema, nas quais foram modificadas as regras de preenchimento e campos foram acrescentados ou removidos. Campos preenchidos com categoria 'ignorado', valor zero, data ignorada ou ausência de informação foram considerados incompletos.

Variáveis analisadas:

Foram analisadas a completude das variáveis relativas aos dados do paciente (Quadro 2), Dados Clínicos e Epidemiológicos (Quadro 3) e Dados de Conclusão (Quadro 4).

Quadro 2 - Variáveis referentes aos Dados do Paciente

DADOS DO PACIENTE	
Variável	Características
Sexo	Campo Obrigatório
Idade	Campo Obrigatório
Gestante	Campo Obrigatório
Raça/Cor	Campo Obrigatório
Se indígena, qual etnia?	Campo Essencial
É membro de povo ou comunidade tradicional?	Campo Obrigatório
Se sim, qual?	Campo Obrigatório
Escolaridade	Campo Essencial
Ocupação	Campo Essencial

Fonte: Elaborado pela autora.

Quadro 3 - Variáveis referentes aos Dados Clínicos e Epidemiológicos

DADOS CLÍNICOS E EPIDEMIOLÓGICOS	
Variável	Características
Trata-se de caso nosocomial (infecção adquirida no hospital)?	Campo Essencial
Paciente trabalha ou tem contato direto com aves, suínos, ou outro animal?	Campo Essencial
Paciente trabalha ou tem contato direto com aves, suínos/Outro animal (especificar)	Campo Essencial
Sinais e Sintomas/Febre/ Tosse/ Dor de Garganta/ Dispneia/ Desconforto Respiratório/ Saturação O ₂ < 95% /Vômito/ Dor abdominal/Fadiga/ Perda do Olfato/ Perda do Paladar/ Outros	Campo Essencial
Fatores de risco	Campo Essencial
Fatores de risco/ Puérpera/ Doença Cardiovascular Crônica/ Doença Hematológica Crônica/ Síndrome de Down/ Doença Hepática/ Crônica Asma/ Diabetes mellitus/ Doença Neurológica Crônica/ Outra/ Pneumopatia/ Crônica Imunodeficiência ou Imunodepressão/ Doença Renal Crônica/ Obesidade/ Obesidade (Descrição IMC) Outros (Descrição)	Campo Essencial
Recebeu vacina COVID-19?	Campo Obrigatório
Recebeu vacina contra Gripe na última campanha? Data da vacinação	Campo Essencial
Usou antiviral para gripe?	Campo Essencial
Qual antiviral? Qual antiviral /Outro, especifique/ Data do início	Campo Essencial

do tratamento	
Recebeu tratamento antiviral para covid-19? Qual antiviral? Data do início do tratamento	Campo Essencial
Houve internação? Data da internação por SRAG UF de internação/ Município de internação	Campo Essencial
Internado em UTI? Data da entrada na UTI Data da saída da UTI	Campo Essencial
Uso de suporte ventilatório?	Campo Essencial
Raio X de Tórax/ Data do Raio X/ Aspecto Tomografia/ Data da Tomografia	Campo Essencial
Coletou amostra? Data da Coleta/ Tipo de amostra	Campo Essencial
Resultado da RT-PCR/outro método Data do Resultado	Campo Essencial

Fonte: Elaborado pela autora.

Quadro 4 - Variáveis referentes aos Dados de Conclusão do Caso

DADOS DE CONCLUSÃO	
Variável	Características
Classificação final do caso	Campo Obrigatório
Critério de Encerramento	Campo Essencial
Evolução do caso	Campo Essencial
Data da alta ou óbito	Campo Essencial
Data do Encerramento	Campo Obrigatório

Fonte: Elaborado pela autora.

Para o cálculo do percentual da completude de cada variável, considerou-se como numerador a quantidade de notificações de SRAG por COVID-19 nas quais o campo foi devidamente preenchido, ou seja, diferente de “ignorado” ou “nulo”, e como denominador, o total de notificações de SRAG por COVID-19. O resultado foi multiplicado por 100 para obter as frequências relativas.

Para medir a qualidade da completude, foram utilizados os critérios estabelecidos por Romero e Cunha (2006), os quais considera a completude como percentual de preenchimento excelente ($\geq 95\%$), bom (90-95%), regular (70-90%), ruim (50-70%) e muito ruim ($< 50\%$). Posteriormente, foi calculada a proporção de cada classificação de grau de completude.

Análise da Oportunidade

Foi realizada análise da oportunidade de notificação, digitação e encerramento dos casos de SRAG por Covid-19 no Brasil, registrados no SIVEP-Gripe. Adotaram-se, como parâmetros de oportunidade, aqueles historicamente utilizados pela vigilância epidemiológica, conforme o Manual de Normas e Rotinas do Sistema de Informação de Agravos de Notificação (SINAN) (BRASIL, 2007), por meio do cálculo dos intervalos em dias entre as respectivas datas, conforme quadro 5:

Quadro 5 - Análise da Oportunidade

ANÁLISE DA OPORTUNIDADE		
Etapas de Oportunidade	Intervalo de tempo entre as datas (em dias)	Tempo Oportuno
Notificação	Notificação e início dos sintomas	Até 7 dias do início dos sintomas
Digitação	Digitação e notificação	Até 30 dias da notificação
Encerramento	Encerramento e notificação	Até 60 dias da notificação

Fonte: Elaborado pela autora.

Para o cálculo do percentual de notificações, digitações e encerramentos oportunos, considerou-se como numerador a quantidade de notificações, digitações e encerramentos com prazo oportuno, e como denominador o total de casos de SRAG por Covid-19.

O Ministério da Saúde recomenda, como parâmetro de referência, que a proporção de notificações e encerramentos oportunos seja superior a 90% (BRASIL, 2008).

4.3 Análise dos Dados da Completude e Oportunidade

A extração do banco de dados, a descrição das frequências e proporções foram conduzidas utilizando o software estatístico R versão 4.1.2. As análises foram estratificadas segundo as Unidades Federativas do Brasil.

4.4 ASPECTOS ÉTICOS

O presente estudo foi desenvolvido a partir de banco de dados secundário, cujas informações encontram-se agregadas e anonimizadas, sem necessidade de contato com o indivíduo, garantindo o sigilo e a privacidade dos dados e dispensando-se aprovação do estudo em Comitê de Ética em Pesquisa.

5 RESULTADOS

Os resultados desta tese estão organizados e apresentados sob a forma de artigos científicos, elaborados a partir das análises desenvolvidas ao longo da pesquisa. O primeiro artigo, intitulado **“Distorções nos dados de COVID-19 à luz da Lei de Benford: comparação internacional e evidências sobre o Brasil”**, foi submetido à **Revista Ciência & Saúde Coletiva** e aborda a análise de confiabilidade dos dados de COVID-19 em perspectiva comparada, com foco no contexto brasileiro.

O segundo artigo, intitulado **“Completude e oportunidade dos registros de SRAG por COVID-19 no SIVEP-Gripe, Brasil, 2020-2023”**, foi submetido à **Revista Brasileira de Epidemiologia** e analisa a qualidade dos registros nacionais de Síndrome Respiratória Aguda Grave, considerando os aspectos de completude e oportunidade das informações.

5.1 ARTIGO I

DISTORÇÕES NOS DADOS DE COVID-19 À LUZ DA LEI DE BENDFORD: COMPARAÇÃO INTERNACIONAL E EVIDÊNCIAS SOBRE O BRASIL

¹ Taina Barbie do Espírito Santo

² Alexandra Crispim Boing

¹ Programa de Pós-Graduação em Saúde Coletiva, Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil. Orcid: <https://orcid.org/0000-0001-5142-483X>. E-mail: tainabarbie7@gmail.com.

² Programa de Pós-Graduação em Saúde Coletiva, Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil. Orcid: <https://orcid.org/0000-0001-7792-4824>. E-mail:

Resumo: O estudo analisou a confiabilidade dos dados de casos e óbitos por Covid-19 nos países com maior número de registros: EUA, Índia, França, Alemanha, Brasil e Japão, com base na conformidade com a Lei de Bendford (LB) entre 2020 e 2022. Trata-se de um estudo descritivo com abordagem quantitativa, com dados secundários da Universidade Johns Hopkins. Aplicaram-se testes estatísticos propostos por Nigrini, como o do primeiro segundo dígitos, teste da soma, Qui-quadrado e desvio médio absoluto. O Brasil foi o único país a não apresentar conformidade com a LB nos registros de casos. Para os óbitos, apenas EUA e Japão tiveram conformidade estrita ou aceitável, o Brasil apresentou conformidade marginalmente aceitável, enquanto França, Alemanha e Índia apresentaram desvios. Conclui-se que a LB é útil como triagem, mas seus resultados devem ser interpretados com cautela, dada a influência das dinâmicas epidêmicas e limitações metodológicas. Reforça-se a necessidade de sistemas de informação mais robustos, oportunos e integrados para garantir maior confiabilidade dos dados epidemiológicos.

Palavras-chave: Lei de Benford; Covid-19; confiabilidade dos dados; vigilância epidemiológica; sistemas de informação.

Distortions in COVID-19 Data in Light of Benford's Law: International Comparison and Evidence on Brazil

Abstract: The study evaluated the reliability of Covid-19 case and death data in the countries with the highest number of records: the United States, India, France, Germany, Brazil and Japan, based on conformity with Benford's Law (BL) from 2020 to 2022. This descriptive quantitative study used secondary data from Johns Hopkins University, applying Nigrini's statistical tests including first- and second-digit tests, the summation test, Chi-square and mean absolute deviation. Brazil was the only country that did not conform to BL in case data. For deaths, only the United States and Japan showed strict or acceptable conformity, Brazil showed marginally acceptable conformity while France, Germany and India presented deviations. It is concluded that BL is useful as a screening tool, but its results should be interpreted with caution given the influence of epidemic dynamics and methodological limitations. The need for more robust, timely and integrated information systems is reinforced to ensure greater reliability of epidemiological data.

Keywords: Benford's Law; Covid-19; data reliability; epidemiological surveillance; information systems.

Distorsiones en los datos de COVID-19 a la luz de la Ley de Benford: comparación internacional y evidencias sobre Brasil

Resumen: El estudio analizó la confiabilidad de los datos de casos y muertes por Covid-19 en los países con mayor número de registros: Estados Unidos, India,

Francia, Alemania, Brasil y Japón, con base en la conformidad con la Ley de Benford (LB) entre 2020 y 2022. Se trata de un estudio descriptivo con enfoque cuantitativo, utilizando datos secundarios de la Universidad Johns Hopkins y pruebas estadísticas propuestas por Nigrini, como los tests del primer y segundo dígito, el test de suma, Chi-cuadrado y desviación media absoluta. Brasil fue el único país que no presentó conformidad con la LB en los casos. Para las muertes, solo Estados Unidos y Japón mostraron conformidad estricta o aceptable, Brasil presentó conformidad marginalmente aceptable mientras que Francia, Alemania e India presentaron desviaciones. Se concluye que la LB es útil como herramienta de cribado, pero sus resultados deben interpretarse con cautela debido a la influencia de las dinámicas epidémicas y las limitaciones metodológicas. Se refuerza la necesidad de sistemas de información más robustos, oportunos e integrados para garantizar mayor confiabilidad de los datos epidemiológicos.

Palabras clave: Ley de Benford; COVID-19; confiabilidad de los datos; vigilancia epidemiológica; sistemas de información.

INTRODUÇÃO

A pandemia da Covid-19, com sua rápida disseminação e elevada letalidade, desencadeou uma crise sanitária e econômica sem precedentes em escala global¹. Até maio de 2023, haviam sido registrados mais de 767 milhões de casos e 6,9 milhões de mortes em todo o mundo. Os países com maior número de casos confirmados foram os EUA, Índia, França, Alemanha, Brasil e Japão². A magnitude desses dados impressiona não apenas pelos impactos humanitários, mas também pelas transformações sociais e econômicas decorrentes³.

Diferente de outros eventos epidêmicos, a Covid-19 foi acompanhada em tempo real, por veículos de imprensa, instituições acadêmicas e órgãos governamentais. A velocidade de produção e disseminação de dados acompanhou o avanço da pandemia, tornando a qualidade dessas informações um elemento crucial para a formulação de políticas públicas baseadas em evidências⁴. Contudo, esse volume crescente de informações também expôs fragilidades dos sistemas nacionais de informação em saúde. Em vários países, surgiram dúvidas sobre a confiabilidade dos dados oficiais⁵, especialmente no Brasil, onde se observaram mudanças nos critérios de registro, indisponibilidade de dados e sobrecarga dos sistemas resultando em registros incompletos e atrasados e, por vezes, de baixa qualidade^{6,4}.

Essas limitações revelam barreiras estruturais à coleta e consolidação de dados epidemiológicos e destacam a importância de sistemas interoperáveis, oportunos e confiáveis para a resposta a emergências sanitárias^{7,8}. Apesar disso, são escassos os estudos que avaliam a qualidade dos dados da pandemia com base em métricas estatísticas consolidadas.

A Lei de Bendford (LB), que descreve uma distribuição esperada dos dígitos em grandes conjuntos numéricos, tem sido utilizada internacionalmente para detectar possíveis distorções ou inconsistências em dados de Covid-19. Estudos identificaram inconformidades em diversos países e regiões^{9,10,11,12,13}. No Brasil, análises também apontaram não conformidade com a LB^{14,15,16,17,4}. Tais pesquisas, contudo, enfrentam limitações quanto ao tamanho da amostra, curta duração da análise e inconsistência nos dados utilizados.

Diante disso, este artigo amplia o escopo desses estudos, ao analisar os países com maior número absoluto de casos de Covid-19, ao longo de um período

contínuo de três anos. Assim, o objetivo do estudo foi avaliar a confiabilidade dos dados de casos e óbitos por Covid-19 nos países com maior número de registros entre 2020 e 2022, com base na Lei de Benford e testes estatísticos complementares.

MÉTODOS

Trata-se de um estudo descritivo, baseado em dados secundários dos casos confirmados e óbitos por Covid-19, obtidos do painel da Universidade Johns Hopkins, abrangendo o período de janeiro de 2020 e dezembro de 2022. Foram selecionados os seis países com maior número absoluto de casos confirmados: Estados Unidos, Índia, França, Alemanha, Brasil e Japão.

A confiabilidade dos dados foi avaliada por meio da aplicação da Lei de Benford (LB), que estabelece a distribuição esperada dos primeiros dígitos em grandes conjuntos de números, conforme a fórmula em que $n \in [1,9]$ e $P(n)$ referem-se à probabilidade de ocorrência do primeiro dígito n em um universo de números aleatórios¹⁸. Segundo essa lei, o dígito 1 deve ocorrer em aproximadamente 30,1% das observações, seguido dos demais até o dígito 9, com frequência decrescente. A aplicação da LB é utilizada para bases com grande volume de dados, variedade de grandezas e, idealmente com valores compostos por ao menos quatro dígitos¹⁹. Os registros diários de casos e óbitos foram organizados por país, e os testes estatísticos seguiram o protocolo proposto Nigrini¹⁹ por teste do primeiro dígito, segundo dígito e teste da soma. O teste do primeiro dígito verifica desvios gerais da distribuição esperada. O teste do segundo dígito aprofunda a análise permitindo identificar vieses sutis nos dados. O teste da soma, por sua vez, avalia se os pares de dígitos iniciais (10 a 99) contribuem proporcionalmente para o total da soma da

base, sendo esperada uma proporção próxima a 1,1% para cada par. Adicionalmente, aplicaram-se o teste Qui-Quadrado e média dos desvios absolutos (DMA). Adotou-se valor crítico de 15,507 proveniente da tabela estatística da distribuição do qui-quadrado, com 8 graus de liberdade e 5% de significância. O resultados do DMA foram classificados segundo Nigrini¹⁹: conformidade estrita ($\leq 0,006$), aceitável (0,006-0,012), marginal (0,012-0,015) e não conformidade ($> 0,015$). A análise considerou o primeiro dígito significativo (ignorado o zero à esquerda).

Para Brasil, a análise foi desagregada por ondas epidêmicas: primeira (fevereiro a outubro de 2020), segunda (novembro de 2020 até novembro de 2021) e a terceira (a partir de dezembro de 2021), conforme definições do Ministério da Saúde²⁰ e da FIOCRUZ²¹.

As análises foram realizadas no software estatístico R (versão 4.1.2), utilizando o pacote estatísticos `benford.analysis`, desenvolvido por Cinelli²².

RESULTADOS

Os países com maior número de casos confirmados de Covid-19 até dezembro de 2022 foram os EUA, Índia, França, Alemanha, Brasil e Japão. No caso do Brasil, foram analisados 36.331.281 registros de casos confirmados e 693.853 registros de óbitos. A aplicação da Lei de Benford revelou que, nos casos confirmados, o número 1 apareceu como primeiro dígito em 22,7% dos registros diários, valor inferior aos 30,1% esperados pela LB. Já nos registros de óbitos, esse mesmo dígito foi observado em 33,9% dos casos, ou seja, acima do padrão esperado pela LB.

Os testes estatísticos de qualidade de ajuste apresentados na Tabela 1, indicaram discrepâncias entre os dados observados e os valores esperados pela LB. Para os registros de casos no Brasil, o teste de χ^2 resultou em 43,804 e o DMA foi de 0,0203, indicando não conformidade com a LB. Para os óbitos, o valor do χ^2 foi de 20,909 e o DMA de 0,0141, classificando-se como conformidade marginalmente aceitável segundo critérios estabelecidos por Nigrini 19. Esses resultados diferem do observado nos demais países analisados.

Nos EUA, os dados de casos confirmados apresentaram conformidade aceitável com a LB, e os dados de óbitos apresentaram conformidade estrita, ou seja, o mais próximo da distribuição teórica esperada. A Índia, França e Alemanha apresentaram resultados semelhantes entre si: os casos confirmados mostraram aceitável com a LB, enquanto os dados de óbitos não demonstraram conformidade. Já os dados do Japão evidenciaram conformidade aceitável tanto para os casos quanto para os óbitos, sendo, junto com os EUA, o único país a apresentar esse padrão (Figuras 1 e 2).

Em termos de distribuição dos primeiros dígitos nos óbitos, a Índia apresentou frequência abaixo do valor esperado pela LB, ao passo que a França e a Alemanha apresentaram frequência superior ao padrão teórico, sugerindo diferentes padrões de distorção entre os países. Essas variações podem estar relacionadas a aspectos distintos de notificação, agregação e publicação dos dados nos respectivos sistemas de vigilância.

No caso do Brasil, também foi realizada uma análise por ondas epidêmicas, conforme apresentado na Figura 3. Em todos os períodos analisados, os registros de casos confirmados apresentaram não conformidade com a LB. Durante a primeira e segunda onda, a distribuição dos primeiros dígitos ficou abaixo da

frequência teórica da LB, enquanto na terceira onda, os valores ficaram acima do esperado. Em relação aos óbitos, os dados da primeira e segunda onda apresentaram frequências nos primeiros dígitos com maior distorção em relação a LB. Contudo, observou-se distribuição mais semelhante à esperada pela LB, resultando em classificação de conformidade marginal aceitável (Tabela 2).

DISCUSSÃO

O presente estudo avaliou a confiabilidade dos dados de casos e óbitos por Covid-19 nos seis países com maior número absoluto de registros, utilizando a Lei de Benford (LB) como ferramenta estatística para identificar distorções numéricas potencialmente indicativas de falhas na integridade dos dados. Os resultados revelam que apenas o Brasil apresentou não conformidade com a LB nos registros de casos confirmados. No que se refere aos óbitos, somente Estados Unidos e Japão demonstraram conformidade estrita ou aceitável, enquanto o Brasil apresentou conformidade marginalmente aceitável, e os demais países não atenderam aos critérios da distribuição esperada pela LB. Esses achados, porém, precisam ser interpretados à luz das limitações da LB. França e Alemanha apresentaram desvios que podem refletir limitações estatísticas da LB em séries temporais e características naturais da curva epidêmica, sem necessariamente indicar falhas no sistema de vigilância^{5,10,12}.

A aplicação da LB, reconhecida como método estatístico eficaz para detecção de anomalias em grandes conjuntos numéricos, evidenciou padrões consistentes com estudos anteriores conduzidos em diferentes contextos geográficos^{15,9,4}. No caso brasileiro, a ausência de conformidade nos dados de casos pode refletir um conjunto de fatores estruturais, operacionais e estatísticos, como subnotificação decorrente de baixa testagem, limitada detecção de casos leves ou assintomáticos,

lacunas na integração dos sistemas de informação, ausência de um sistema de notificação no início da pandemia e variabilidade dos números ao longo da epidemia.

A análise dos óbitos, embora tenha apresentado maior proximidade com os padrões esperados, também evidenciou variações temporais relevantes. A conformidade marginalmente aceitável observada na terceira onda epidêmica no Brasil, pode refletir o aumento abrupto do volume e da dispersão dos dados durante a disseminação da variante Ômicron, uma vez que a LB tende a apresentar melhor desempenho em contextos numéricos caracterizados por alta variabilidade e crescimento acelerado. Dessa forma, parte da aparente melhora pode decorrer de propriedades estatísticas inerentes à dinâmica epidêmica, e não exclusivamente à melhoria dos sistemas de informação ou avanços tecnológicos²³.

É relevante destacar que a LB, embora útil para avaliar padrões numéricos, não permite inferir diretamente as causas das distorções identificadas. Desvios podem estar relacionados a fatores epidemiológicos, como curvas de crescimento com platôs e picos, ou a práticas de notificação que variam ao longo do tempo e entre países (VÂLSAN et al., 2024). Nesse sentido, as evidências apontam que sistemas de mais robustez institucional e desenvolvimento econômico tendem a produzir dados mais confiáveis⁵, o que pode explicar parcialmente os melhores resultados observados em países como Estados Unidos e Japão.

Ainda assim, o fato de França e Alemanha apresentarem não conformidade nos registros de óbitos indica que fatores como o estágio da curva epidêmica, mudanças de testagem, intervenções governamentais ou períodos de estabilização prolongada podem afetar a conformidade com a LB independentemente da qualidade do sistema. Estudos como os de Sambridge & Jackson e Balashov et al. demonstram que a LB tende a ser seguida principalmente em fases de crescimento

exponencial, perdendo aderência em períodos de platô ou flutuações decorrentes de intervenções externas.

No caso do Brasil, as falhas de interoperabilidade, a ausência de padronização entre os sistemas municipais e estaduais, e os episódios de apagão de dados, como aquele ocorrido em meio à emergência da variante Ômicron, prejudicaram a continuidade da vigilância e a qualidade da informação⁶. Soma-se a isso o uso inicial de plataformas não concebidas como Sistemas de Informação em Saúde (SIS), como o REDCap, que apresentava limitações funcionais, ausência de campos obrigatórios para encerramento e dificuldade para extração e análise dos dados levando à criação de sistemas paralelos pelos municípios e estados e à fragmentação da informação²⁰.

Esses entraves comprometem a capacidade analítica do sistema de saúde e fragilizam a formulação de políticas públicas baseadas em evidências. Dados confiáveis são fundamentais para adoção eficiente de recursos, a definição de estratégias de mitigação e manutenção da confiança pública nas ações estatais. Como apontado por Prado²⁴, estima-se que as notificações oficiais no Brasil tenham representado 9,2% (IC95% 8,8% - 9,5%) dos casos reais no início da pandemia, percentual substancialmente inferior ao observado em outros países.

A confiabilidade dos dados é essencial para o planejamento e a implementação de políticas públicas eficazes. Em situações de crise sanitária, como a pandemia de Covid-19, decisões baseadas em dados inconsistentes podem comprometer a resposta governamental, dificultar a alocação de recursos e enfraquecer a confiança da população nas ações de saúde pública.

O presente estudo apontou diferentes questões na qualidade dos dados de Covid-19, desde a verificação da conformidade dos casos e óbitos com a LB, até as

lacunas existentes nos registros dos dados e nos sistemas de informação. Sistemas de Informação robustos que sejam adequadamente desenvolvidos antes de futuros surtos, podem fortalecer a capacidade de resposta dos sistemas de saúde e contribuir para uma preparação mais eficaz frente a emergências sanitárias.

Os SIS de alto desempenho e efetivamente alavancados são catalisadores para informar a governança de saúde global coordenada e garantir o compartilhamento de dados transparentes e oportunos²⁵.

Como limitação, ressalta-se que, embora a Lei de Benford seja um instrumento amplamente utilizado na auditoria de dados, seus resultados devem ser interpretados com cautela. Fatores como tamanho da amostra, granularidade temporal dos dados e os testes estatísticos selecionados podem influenciar os resultados. Assim, os achados apresentados devem ser considerados como indicadores preliminares de inconsistência, a serem complementados por investigações qualitativas e análises do funcionamento dos sistemas de informação.

CONCLUSÃO

Os achados sugerem possíveis limitações nos dados brasileiros, mas tais interpretações devem ser consideradas à luz das restrições metodológicas da Lei de Benford e das dinâmicas epidêmicas. Os resultados deste estudo indicam que os dados de casos da COVID-19 no Brasil apresentaram desvios em relação à Lei de Benford, enquanto outros países analisados mostraram maior aderência. No entanto, variações também foram observadas na métrica de óbitos, inclusive em sistemas de saúde considerados robustos, como Alemanha e França, que não alcançaram plena conformidade. Esses achados sugerem que a não conformidade com a Lei de Benford não deve ser interpretada automaticamente como falha no

sistema de saúde ou nos processos de notificação, uma vez que a própria metodologia possui limitações, especialmente quando aplicada a séries temporais finitas. A análise aponta possíveis fatores que podem influenciar a integridade dos dados, mas a interpretação deve ser cautelosa, considerando que desvios podem ocorrer por múltiplos motivos, incluindo variações naturais na notificação e características estatísticas dos conjuntos de dados.

Nesse contexto, a Lei de Benford se apresenta como uma ferramenta útil de triagem para detectar possíveis distorções numéricas, mas sua interpretação deve ser complementada, sempre que possível, por indicadores internos de qualidade dos dados, como oportunidade, consistência e completude. A aplicação integrada desses elementos, que dependem de microdados não disponíveis em bases internacionais agregadas, permitiria uma avaliação mais abrangente da confiabilidade das informações. Recomenda-se que estudos futuros combinem a LB com análises de oportunidade e completude quando tais dados estiverem acessíveis, o que não foi possível no presente estudo.

A análise indica padrões que podem refletir limitações na qualidade dos dados epidemiológicos, especialmente no contexto brasileiro e reforça, a necessidade de um aprimoramento estrutural e contínuo dos sistemas de informação, tornando-os mais robustos, oportunos e confiáveis. A existência de um SIS eficiente, por si só, não é suficiente para assegurar a integridade da informação em contextos de emergência sanitária. A produção de dados confiáveis exige um sistema de saúde forte e bem estruturado, com vigilância em saúde fortalecida, recursos humanos suficientes, qualificados e valorizados e financiamento público suficiente e contínuo, que permita o funcionamento coordenado e sustentável desses componentes.

A experiência da pandemia de Covid-19 evidencia que a gestão da informação em saúde deve ser tratada como um eixo estratégico da capacidade de preparação e resposta. É necessário construir um modelo resiliente e sistêmico, que articule infraestrutura tecnológica, competências institucionais e estabilidade política e financeira para enfrentar eventos inesperados com agilidade e equidade. O uso de dados válidos e oportunos é indispensável para orientar intervenções, sustentar decisões de governo e assegurar a confiança pública nas ações de vigilância epidemiológica. Trata-se, portanto, de um imperativo não apenas técnico, mas estrutural e político.

REFERÊNCIAS

1. Moura EC, others. Disponibilidade de dados públicos em tempo oportuno para a gestão: análise das ondas da COVID-19. 2021.
2. Johns Hopkins University. Coronavirus COVID-19 Global Cases by Johns Hopkins CSSE [Internet]. 2023. Available from: <https://coronavirus.jhu.edu/data/cumulative-cases>
3. Bresser-Pereira LC. Financiamento da Covid-19, inflação e restrição fiscal. Brazilian J Polit Econ. 2020;40(4):604–21.
4. Silva L, Figueiredo Filho D. Using Benford's law to assess the quality of COVID-19 register data in Brazil. J Public Health (Bangkok). 2021 Mar;43(1):107–10.
5. Balashov VS, Yan Y, Zhu X. Using the Newcomb-Benford law to study the association between a country's COVID-19 reporting accuracy and its development. Sci Rep. 2021 Nov;11(1).

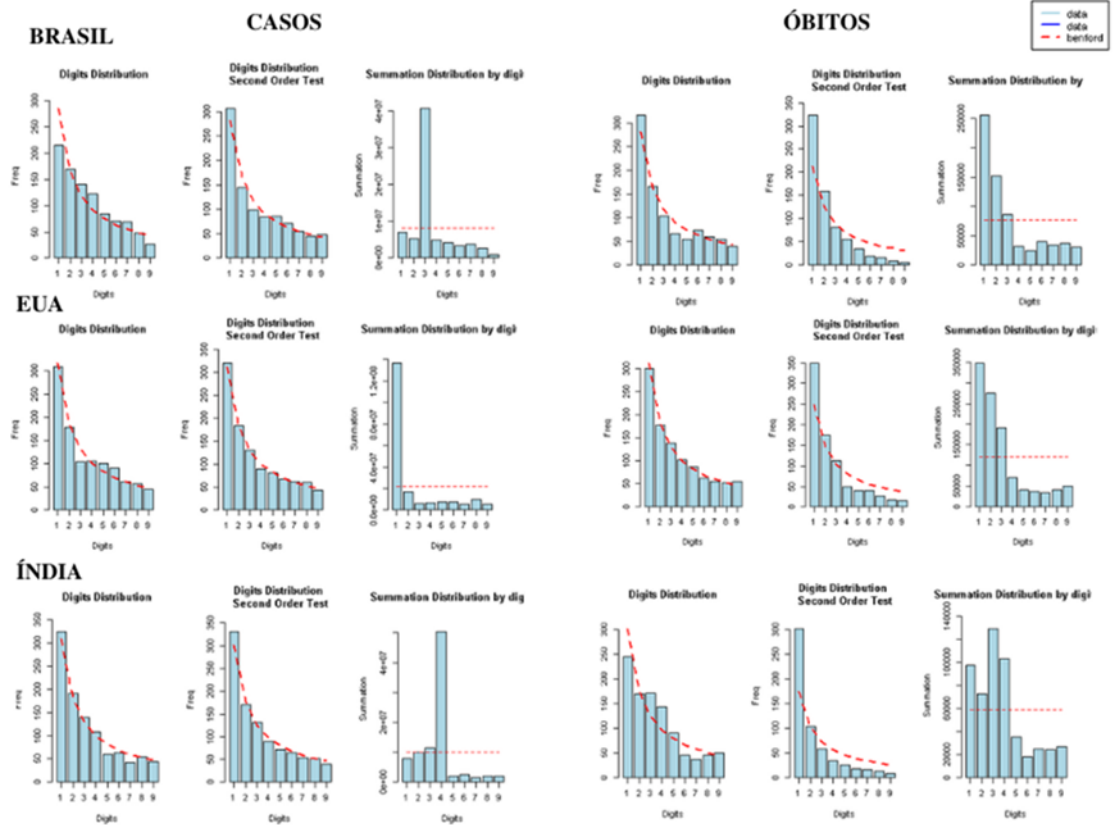
6. Villela DAM, Gomes MF da C. O impacto da disponibilidade de dados e informação oportuna para a vigilância epidemiológica. *Cad Saude Publica*. 2022;38:e00115122.
7. Natário I. Algumas considerações sobre a COVID-19 [Internet]. 2021. Available from:
<https://spestatistica.pt/storage/app/uploads/public/608/6b0/4fb/6086b04fb6d0d924468858.pdf>
8. Neri SCC. A qualidade dos dados dos sistemas de informação em saúde aplicados na atenção à saúde materno-infantil. Universidade não especificada; 2016.
9. Moreau VH. Inconsistencies in countries COVID-19 data revealed by Benford's law. *Model Assist Stat Appl*. 2021;16(1):73–9.
10. Farhadi N. Can we rely on COVID-19 data? An assessment of data from over 200 countries worldwide. *Sci Prog*. 2021;104(2).
11. Koliás P. Applying Benford's law to COVID-19 data: the case of the European Union. *J Public Health (Bangkok)*.
12. Morillas-Jurado FG, Caballer-Tarazona M, Caballer-Tarazona V. Applying Benford's Law to Monitor Death Registration Data: A Management Tool for the COVID-19 Pandemic. *MATHEMATICS*. 2022 Jan;10(1).
13. Campolieti M. Mortes por COVID-19 nos EUA: Lei de Benford e subnotificação. *Rev Saude Publica*. 2022;44(2):e268--e271.
14. Carmo CRS, Canepelle FL, Nunes FC. Analysis of COVID-19 contamination and deaths cases in Brazil according to the newcomb-benford law. *Rev Bras Biometria*. 2021;39(4):522–35.

15. Galvêas D, Barros F, Fuzo CA. A forensic analysis of SARS-CoV-2 cases and COVID-19 mortality misreporting in the Brazilian population. *Public Health*. 2021;196:114–6.
16. Idrovo AJ, Manrique-Hernández EF, Fernández Niño JA. Report From Bolsonaro's Brazil: The Consequences of Ignoring Science. *Int J Heal Serv*. 2021;51(1):31–6.
17. Manrique-Hernández EF, Moreno-Montoya J, Hurtado-Ortiz A, Prieto-Alvarado FE, Idrovo ÁJ. Desempeño del sistema de vigilancia colombiano durante la pandemia de COVID-19: evaluación rápida de los primeros 50 días TT - Performance of the Colombian surveillance system during the COVID-19 pandemic: A rapid evaluation of the first 50 days. *Biomédica* [Internet]. 2020;40:96–103. Available from: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-41572020000600096&lang=pt
18. Benford F. The law of anomalous numbers. *Proc Am Philos Soc* [Internet]. 1938;78:551–72. Available from: <https://www.jstor.org/stable/984802>
19. Nigrini MJ. *Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations*. John Wiley & Sons; 2020.
20. Ministério da Saúde. *Boletim Epidemiológico Especial: Covid-19*. 2023.
21. Fundação Oswaldo Cruz. *Observatório Covid-19 - Boletim Especial: Balanço de dois anos da pandemia Covid-19* [Internet]. 2022. Available from: <https://portal.fiocruz.br/documentos-produzidos-pelo-observatorio-covid-19>
22. Cinelli C. *benford.analysis: análise Benford para validação de dados e análise forense*. 2018.

23. Carvalho CA de, Carvalho VA de, Campos MAG, Oliveira BLCA de, Diniz EM, Santos AM dos, et al. Delay in death reporting affects timely monitoring and modeling of the COVID-19 pandemic TT - O atraso na notificação de óbitos afeta o monitoramento oportuno e modelagem da pandemia da COVID-19 TT - La demora en la notificación de muertes afecta el seg. Cad Saude Publica [Internet]. 2021;37(7). Available from: http://www.scielo.org/scielo.php?script=sci_arttext&pid=S0102-311X2021000705018&lang=pt
24. Prado MF do, others. Análise da subnotificação de COVID-19 no Brasil. Rev Bras Ter Intensiva. 2020;32:224–8.
25. Ribas FV, others. Completeness of notifications of severe acute respiratory syndrome at the national level and in a regional health care in the state of Minas Gerais, during the COVID-19 pandemic, 2020. Epidemiol e Serviços Saúde. 2022;31.

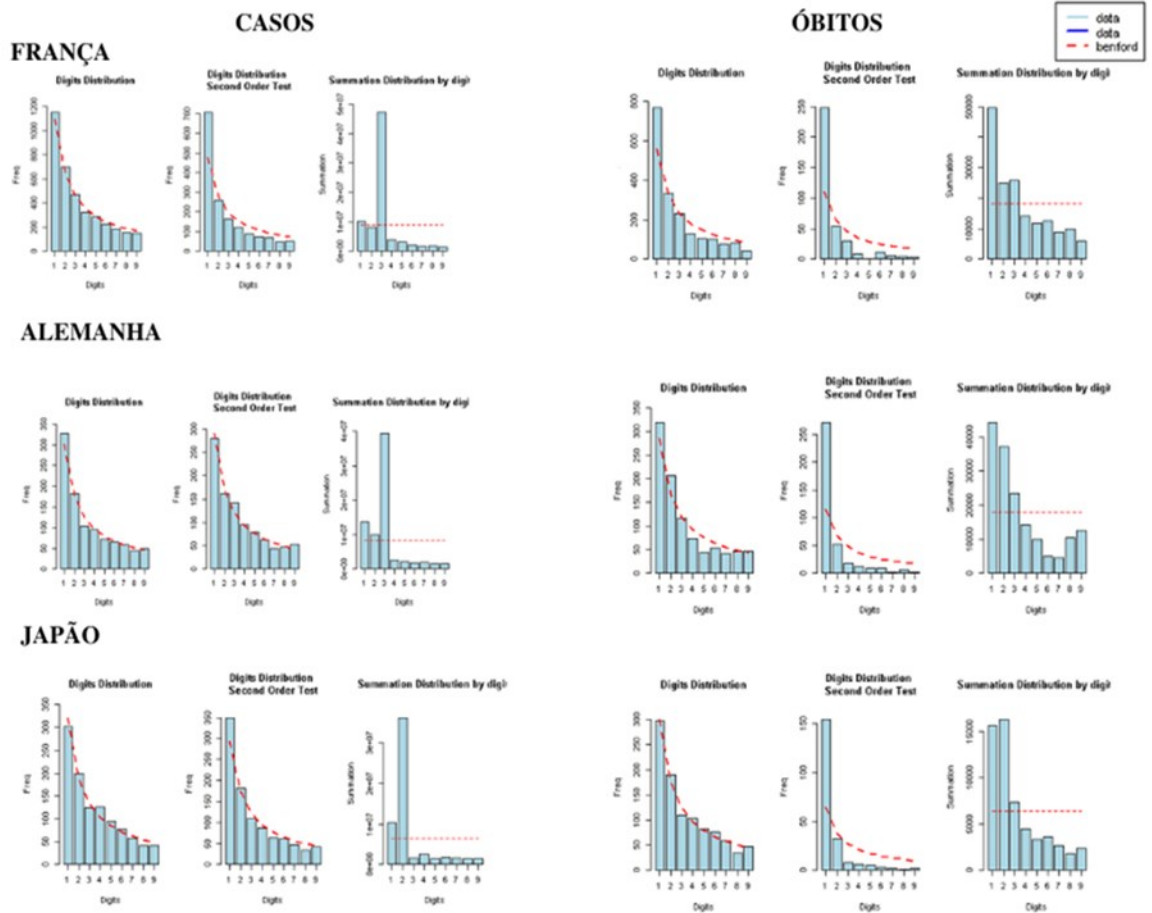
ANEXO – Artigo I

Figura 1: Distribuição dos Casos e óbitos por Covid-19 segundo a Lei de Benford. Brasil, EUA e Índia. 2020 a 2022.



Fonte: Paineis da Johns Hopkins University, 2023.

Figura 2: Distribuição dos Casos e óbitos por Covid-19 segundo a Lei de Benford. França, Alemanha e Japão. 2020 a 2022.



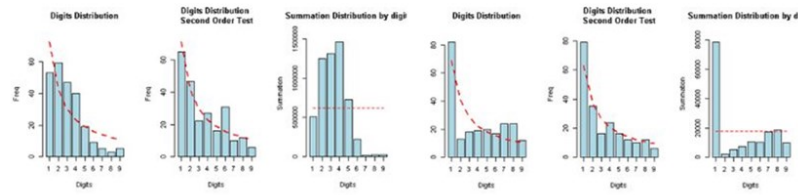
Fonte: Painel da Johns Hopkins University, 2023.

Figura 3: Distribuição dos Casos e Óbitos por Covid-19 segundo a Lei de Benford. Brasil, de 2020 a setembro de 2022, por período.

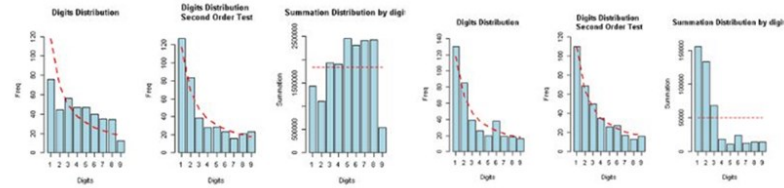
BRASIL

CASOS

Primeira Onda (Fevereiro a Outubro de 2020)



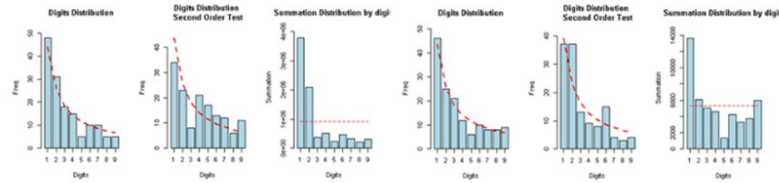
Segunda Onda (Novembro de 2020 a Novembro de 2021)



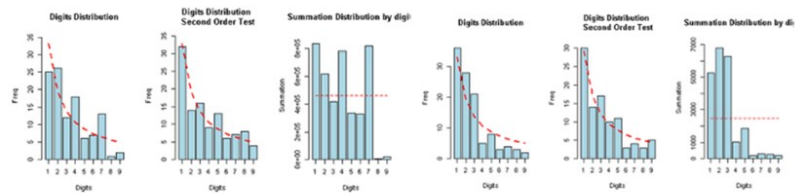
BRASIL

CASOS

Terceira Onda (Dezembro de 2021 a Abril de 2022)



Período de Maio a Setembro 2022



Fonte: Painel da Johns Hopkins University, 2023.

Tabela 1: Distribuição dos primeiros dígitos dos Casos e óbitos por Covid-19 – 2020 a 2022.

Países	Dígitos									N	X ²	Mantissa	DF	MAD	MAD Conformity - Nigrini (2012):
	1	2	3	4	5	6	7	8	9						
LB	30,1	17,6	12,5	9,7	7,9	6,7	5,8	5,1	4,6						
BRASIL															
Casos	22,7	18,0	14,9	13,0	8,9	7,4	7,3	5,1	2,9	947	43,804	0,534	4,325821	0,02037996	Não conformidade
Óbitos	33,9	17,8	11,0	7,1	5,9	7,8	6,3	5,9	4,3	933	20,909	0,010	-2,624009	0,01418429	Marginalmente aceitável
EUA															
Casos	29,4	16,9	9,9	10,1	9,6	8,7	5,8	5,4	4,3	1052	16,522	0,512	3,761769	0,009613756	Aceitável
Óbitos	29,1	17,1	13,4	9,9	8,5	6,1	5,4	5,1	5,4	1036	4,1667	0,503	0,6588172	0,005734481	Conformidade Estrita
ÍNDIA															
Casos	31,6	18,6	13,5	10,5	5,8	6,1	4,2	5,4	4,3	1026	14,424	0,484	-4,162778	0,01026084	Aceitável
Óbitos	24,6	16,9	17,2	14,4	9,1	4,6	3,8	4,6	5,0	1002	66,938	0,511	-1,147386	0,02428896	Não conformidade
FRANÇA															
Casos	31,6	19,2	13,0	8,8	7,8	6,2	5,0	4,2	4,1	3634	24,676	0,464	-7,392553	0,008078884	Aceitável
Óbitos	41,2	17,8	12,4	6,9	5,6	5,3	4,1	4,5	2,3	1861	142,34	0,378	-12,45038	0,02520477	Não conformidade
ALEMANHA															
Casos	32,6	18,4	10,4	9,5	7,3	6,6	6,0	4,5	5,0	1003	7,725	0,482	-13,5668	0,00828330	Aceitável
Óbitos	33,6	21,7	12,3	7,8	4,7	5,7	4,4	4,8	4,9	950	33,469	0,455	-13,10599	0,01760589	Não conformidade

JAPÃO

Casos	28,4	18,7	11,6	11,8	8,9	7,3	5,3	3,9	3,9	1066	13,438	0,501	-0,928032	0,0167504	Aceitável
Óbitos	29,8	19,1	10,9	10,3	8,2	7,6	5,9	3,4	4,7	997	10,719	0,472	-9,587989	0,00795495	Aceitável

Fonte: Painel da Johns Hopkins University, 2023.

Tabela 2: Distribuição dos primeiros dígitos dos Casos e óbitos por Covid-19 por períodos. Brasil, 2020 a 2022.

Países	Dígitos									N	X ²	Mantissa	DF	MAD	MAD Conformity - Nigrini (2012):
	1	2	3	4	5	6	7	8	9						
LB	30,1	17,6	12,5	9,7	7,9	6,7	5,8	5,1	4,6						
BRASIL															
Primeira Onda (Fevereiro a Outubro de 2020)															
Casos	22,1	24,6	19,6	16,7	7,9	3,8	2,1	1,2	2,1	240	52,553	0,480	-12,44324	0,04675407	Não conformidade
Óbitos	35,8	5,7	7,9	8,3	8,7	7,4	10,5	10,5	5,2	229	47,52	0,51	11,63738	0,03991098	Não conformidade
Segunda Onda (Novembro de 2020 a Novembro de 2021)															
Casos	19,4	11,3	14,3	12,0	12,0	10,2	9,0	8,7	3,1	391	61,03	0,593	18.69648	0.04117396	Não conformidade
Óbitos	33.2	21.7	10,0	6,6	5,1	9,7	4,9	4,6	4,1	391	21,007	0,467	-5.508533	0.02288684	Não conformidade

Terceira Onda (Dezembro de 2021 a Abril de 2022)

Casos	32,7	21,1	12,2	10,2	3,4	6,8	6,8	3,4	3,4	147	6,7079	0,471	-6,883467	0,01700893	Não conformidade
Óbitos	31,7	17,2	14,5	8,3	4,1	6,9	5,5	5,5	6,2	145	4,4309	0,491	-0,414170	0,01298896	Marginalmente aceitável

Período de Maio a Setembro 2022

Casos	22,7	23,6	10,9	16,3	5,45	6,3	11,8	0,9	1,8	110	22,898	0,511	-2,126221	0,04159747	Não conformidade
Óbitos	32,7	25,5	19,1	4,5	7,3	2,7	3,6	2,7	1,8	110	17,519	0,44	-18,06821	0,03792606	Não conformidade

Fonte: Painel da Johns Hopkins University, 2023.

5.2 ARTIGO II

¹ Taina Barbie do Espírito Santo

² Alexandra Crispim Boing

¹ Programa de Pós-Graduação em Saúde Coletiva, Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil. Orcid: <https://orcid.org/0000-0001-5142-483X>. E-mail: tainabarbie7@gmail.com.

² Programa de Pós-Graduação em Saúde Coletiva, Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil. Orcid: <https://orcid.org/0000-0001-7792-4824>. E-mail:

Completeness and Opportunity of SRAG records for Covid-19 in SIVEP-Gripe, Brazil 2020-2023

RESUMO

Objetivo: Avaliar a completude e a oportunidade dos registros de Síndrome Respiratória Aguda Grave (SRAG) por COVID-19 no sistema SIVEP-Gripe, no Brasil, entre 2020 e 2023.

Métodos: Estudo descritivo, baseado em dados secundários públicos e anonimizados do SIVEP-Gripe, referentes a 2.200.796 registros de SRAG por COVID-19. Foram analisadas 66 variáveis sociodemográficas, clínicas, de hospitalização, evolução e tratamento. A completude foi classificada segundo critérios de Romero e Cunha (2006). A oportunidade foi medida pelos intervalos entre início dos sintomas e notificação, notificação e digitação, e notificação e encerramento, considerando parâmetros históricos da vigilância (7, 30 e 60 dias, respectivamente).

Resultados: No período, 59% das variáveis apresentaram completude ruim ou muito ruim, e apenas 24% foram classificadas como excelente ou boa. Houve heterogeneidade regional: Amapá, Mato Grosso do Sul, Santa Catarina e Paraná obtiveram maiores proporções de preenchimento satisfatório, enquanto Tocantins,

Pernambuco, Mato Grosso, Rondônia e Alagoas tiveram pior desempenho. Entre 2020 e 2023, o tempo médio de notificação reduziu de 13 para 7 dias. A digitação superou 90% de registros oportunos na maioria dos estados, enquanto o encerramento permaneceu em torno de 86%. Ainda assim, a meta de 90% de notificações oportunas não foi alcançada.

Conclusões: Apesar dos avanços na agilidade dos registros, a completude manteve-se insatisfatória e desigual entre estados. Esses achados evidenciam a necessidade de aprimorar o preenchimento das fichas e investir em capacitação e monitoramento sistemático, a fim de assegurar dados mais consistentes para a vigilância e o planejamento em saúde pública.

Palavras-chave: Covid-19; Síndrome respiratória aguda grave; Sistemas de informação em Saúde; Confiabilidade dos dados.

Abstract

Objective: To evaluate the completeness and timeliness of Severe Acute Respiratory Syndrome (SARS) due to COVID-19 records in the SIVEP-Gripe system in Brazil, from 2020 to 2023.

Methods: This was a descriptive study based on secondary, public, and anonymized data from SIVEP-Gripe, comprising 2,200,796 SARS due to COVID-19 records. Sixty-six sociodemographic, clinical, hospitalization, outcome, and treatment variables were analyzed. Completeness was classified according to Romero and Cunha (2006). Timeliness was assessed through the intervals between symptom onset and notification, notification and data entry, and notification and case closure, considering historical surveillance parameters (7, 30, and 60 days, respectively).

Results: Overall, 59% of the variables showed poor or very poor completeness, while only 24% were classified as excellent or good. Regional heterogeneity was observed: Amapá, Mato Grosso do Sul, Santa Catarina, and Paraná achieved higher proportions of satisfactory completion, whereas Tocantins, Pernambuco, Mato Grosso, Rondônia, and Alagoas had the worst performance. Between 2020 and 2023, the average notification time decreased from 13 to 7 days. Data entry was performed on time in over 90% of records in most states, while case closure remained around 86%. Nevertheless, the 90% target for timely notification was not achieved.

Conclusions: Despite improvements in the timeliness of records, completeness remained unsatisfactory and uneven across states. These findings highlight the need to strengthen data entry practices and invest in continuous training and systematic monitoring to ensure more consistent data for surveillance and public health planning.

Keywords: COVID-19; Severe Acute Respiratory Syndrome; Health Information Systems; Data Accuracy.

INTRODUÇÃO

A pandemia da Covid-19 gerou impactos globais na saúde e na economia (MOURA et.al., 2021), evidenciando a necessidade de sistemas de informação capazes de subsidiar decisões oportunas. Segundo os dados oficiais do Ministério da Saúde, até dezembro de 2023, foram contabilizados mais de 39 milhões de casos da doença. Entre esses casos, 2 milhões foram classificados como graves, resultando em 714 mil óbitos¹.

No Brasil, a vigilância da Síndrome Respiratória Aguda Grave (SRAG), implantada em 2009 após a pandemia de influenza, tornou-se estratégica durante a emergência da Covid-19, pois permite monitorar hospitalizações e óbitos relacionados aos SARS-COV2².

Entretanto, para assegurar um acompanhamento efetivo e contínuo dos casos, é fundamental dispor de sistemas de informação em saúde (SIS) que sejam de qualidade, confiáveis e devidamente preenchidos. Esses sistemas devem conter dados completos, atualizados e registrados com precisão e oportunidade, de modo a oferecer suporte adequado ao planejamento, à tomada de decisões e à execução de ações por parte dos gestores³⁻⁵.

Nesse sentido, a completude e a oportunidade configuram atributos essenciais dos sistemas de informação, estando diretamente relacionadas ao seu desempenho^{6,7}. A completude refere-se à proporção com que cada registro de um sistema apresenta valores distintos das categorias “ignorado” ou “nulo”. Já a oportunidade indica a velocidade com que as informações são disponibilizadas para a tomada de decisão e torna-se fundamental para ações imediatas, como a interrupção da cadeia de transmissão de doenças⁸.

A importância da completude e da oportunidade dos dados torna-se ainda mais evidente em contextos de emergência em saúde pública, como a pandemia de COVID-19, em que a qualidade da informação impacta diretamente a tomada de decisão. Estudos nacionais e internacionais tem apontado fragilidades nos sistemas de informação em contexto de emergências sanitárias, incluindo baixa completude e atrasos na disponibilização de dados⁹. No Brasil, pesquisas durante a pandemia avaliaram a qualidade dos registros em recortes estaduais ou limitadas a um único

ano^{2,10,11}, mas ainda são escassas as análises que contemplam períodos mais longos e abrangência nacional.

Diante desta lacuna, o presente estudo tem como objetivo avaliar a completude e a oportunidade dos registros de Síndrome Respiratória Aguda Grave (SRAG) por Covid-19 no SIVEP-Gripe, no Brasil, entre 2020 a 2023.

MÉTODOS

Trata-se de um estudo descritivo sobre a qualidade dos registros de SRAG por Covid-19. No Brasil, a notificação de SRAG é compulsória e realizada no sistema de informação SIVEP-Gripe, o qual armazena dados clínicos, epidemiológicos e de evolução, preenchidos pela instituição hospitalar notificadora. Desde 2020, o sistema passou por atualizações para incorporar o avanço do conhecimento sobre a Covid-19 e atender às demandas da vigilância epidemiológica¹².

A população do estudo foi composta pelos registros de SRAG com classificação final COVID-19, notificados no Brasil entre 1º de março de 2020 a 31 de dezembro de 2023, segundo calendário epidemiológico. Os dados foram obtidos a partir de base pública, anonimizada, disponível no site <https://opendatasus.saude.gov.br/dataset/> com atualização em 29 de janeiro de 2024.

A qualidade dos registros foi avaliada por meio da análise da completude e da oportunidade de notificação, digitação e encerramento dos registros de SRAG por COVID-19 no Brasil. Foram incluídas 66 variáveis da ficha de notificação de SRAG, selecionadas por representarem as principais de cada categoria, abrangendo

informações sociodemográficas (sexo, idade, gestação, raça/cor, escolaridade, ocupação), características clínicas e sintomas (caso nosocomial, surto de SG, febre, tosse, dor de garganta, dispneia, desconforto respiratório, saturação abaixo de 95%, diarreia, vômito, dor abdominal, fadiga, perda de olfato e de paladar, outros sintomas), condições preexistentes (fatores de risco, doença cardiovascular, hematológica, Síndrome de Down, doença hepática, asma, diabetes, doença neurológica, pneumopatia, imunodeficiência, doença renal, obesidade, descrição do IMC), imunização e tratamento (vacina contra covid-19, data da 1 dose, data da 2 dose, vacina da gripe, antiviral, qual antiviral, outro, início do tratamento), hospitalização e evolução (data da internação, UF de internação, município de internação, internação em UTI, data de entrada e saída da UTI, uso de suporte ventilatório, raio-X, tomografia, coleta e resultado, critério de encerramento, evolução do caso, data da alta ou óbito e data do encerramento).

Para o cálculo da completude de cada variável, considerou-se como numerador o número de registros de SRAG por COVID-19 em que o campo foi devidamente preenchido (diferente de “ignorado” ou “nulo”), e como denominador, o total de registros de SRAG por COVID-19. O resultado foi multiplicado por 100 para obtenção da frequência relativa. A qualidade da completude foi classificada segundo os critérios de Romero e Cunha (2006), os quais considera a completude como percentual de preenchimento excelente ($\geq 95\%$), bom (90-95%), regular (70-90%), ruim (50-70%) e muito ruim ($< 50\%$). Em seguida, calculou-se a proporção de variáveis em cada categoria de classificação.

A oportunidade de notificação foi avaliada pelo intervalo de tempo, em dias, entre a data dos primeiros sintomas e a data da notificação. Para a oportunidade de digitação, considerou-se a diferença entre as datas de notificação e de digitação,

enquanto a oportunidade de encerramento foi medida pela diferença entre a data da notificação e a data de encerramento dos registros.

Adotaram-se, como parâmetros de oportunidade, aqueles historicamente utilizados pela vigilância epidemiológica, conforme o Manual de Normas e Rotinas do Sistema de Informação de Agravos de Notificação (SINAN). Dessa forma, considerou-se oportuna a notificação realizada em até sete dias após o início dos sintomas, a digitação em até 30 dias após a notificação e o encerramento oportuno em até 60 dias a partir da data da notificação¹³.

Para o cálculo do percentual de notificações, digitações e encerramentos oportunos, considerou-se como numerador o número de registros realizados dentro do prazo definido como oportuno, e como denominador, o total de registros de SRAG por Covid-19. O Ministério da Saúde estabelece como parâmetro que mais de 90% dos registros sejam notificados e encerrados oportunamente¹⁴.

A extração do banco de dados, bem como o cálculo das frequências e proporções, foi realizada no software R, versão 4.1.2. Por se tratar de dados secundários, públicos e anonimizados, garantindo o sigilo e a privacidade dos dados, não houve necessidade de submissão do estudo ao Comitê de Ética em Pesquisa.

RESULTADOS

No período de 2020 a 2023, foram registradas 2.200.796 fichas de SRAG por COVID-19 no Brasil. Entre as 66 variáveis analisadas, 59% apresentaram grau de completude considerado muito ruim e ruim, enquanto apenas 24% apresentaram completude excelente ou boa. Entre 2020 e 2021, o número de notificações

aumentou aproximadamente 69%, atingindo o pico da pandemia. Em 2022, houve redução drástica no volume de casos, superior a cinco vezes. Apesar dessas variações, a proporção de variáveis com preenchimento “ruim” ou “muito ruim” manteve-se relativamente estável, indicando persistência de lacunas na completude dos registros (Tabela 1).

Os estados do Amapá, Mato Grosso do Sul, Santa Catarina e Paraná apresentaram as maiores proporções de variáveis classificadas como bom e excelente, alcançando em conjunto aproximadamente 40 a 45% de preenchimento satisfatório. Em contrapartida, Tocantins, Pernambuco, Mato Grosso, Rondônia e Alagoas exibiram os piores desempenhos, com predomínio das categorias ruim e muito ruim, superando 50%. No panorama nacional, observou-se que cerca de 30% das variáveis permaneceram em níveis insatisfatórios (ruim e muito ruim) (Gráfico 1).

Na análise da oportunidade (Gráfico 2), verificou-se redução significativa nos intervalos médios de processamento dos registros. O tempo entre o início dos sintomas e a notificação diminuiu de 13 para 7 dias. De forma semelhante, os intervalos, os intervalos médios para digitação e encerramento reduziram-se em mais de 50%, evidenciando maior agilidade no fluxo do registro e finalização das fichas.

A Figura 1 apresenta o percentual de registros notificados, digitados e encerrados oportunamente segundo as UF do Brasil. Em 2020, apenas 50% dos registros de SRAG por COVID-19 foram notificados no país dentro de 7 dias a partir da data do início dos sintomas. Em 2023, esse percentual aumentou para 76,9%. De forma semelhante ao padrão nacional, todos os estados apresentaram melhora na

notificação oportuna. Apesar dos avanços, nenhuma UF atingiu a meta de 90% dos casos notificados dentro do prazo recomendado.

A avaliação da oportunidade de digitação no Brasil revelou que em 2020, mais de 87% das fichas foram digitadas dentro de 30 dias após a data de notificação, passando para 93% em 2023. A maioria dos estados aumentou a proporção de digitação oportuna ao longo dos anos, alcançando a meta de 90% estabelecida pelo Ministério da Saúde. No entanto, Espírito Santo, Mato Grosso, Pará, Pernambuco e Rio de Janeiro foram os estados que não atingiram esse parâmetro (Figura 1).

Quanto ao encerramento oportuno em até 60 dias da data da notificação, observou-se durante o período analisado, que cerca de 86% dos casos foram encerrados oportunamente no Brasil, com pouca alteração entre os anos. Os estados que encerraram mais de 90% dos casos oportunamente foram Acre, Amapá, Mato Grosso do Sul, Paraíba, Piauí, Roraima, Santa Catarina e Sergipe. Em relação ao tempo de oportunidade, somente a meta de digitação oportuna foi alcançada nos anos de 2022 e 2023, enquanto os demais permaneceram abaixo dos parâmetros estabelecidos (Figura 1).

DISCUSSÃO

O presente estudo analisou a completude e o tempo de oportunidade dos registros de SRAG por Covid-19 do SIVEP-Gripe. Constatou-se que a completude foi avaliada como insatisfatória, sendo a maioria das variáveis classificadas como ruim e muito ruim. Em contrapartida, ocorreu uma significativa redução no tempo de notificação, digitação e encerramento das fichas, indicando uma maior agilidade no processo de alimentação do sistema. Esses achados corroboram estudos anteriores

^{2,11}, que também identificaram baixa completude nos registros de SRAG. Embora a redução no número de registros em 2022 e 2023 tenha acelerado os processos de notificação e encerramento, não se refletiu em avanços expressivos na qualidade do preenchimento das informações.

A disparidade da qualidade de dados entre os estados pode ser atribuída à heterogeneidade da estrutura e nos fluxos de trabalho de cada localidade, como o preenchimento dos dados manualmente em fichas de papel, para posterior digitação. Além disso, fatores como diagnósticos imprecisos, escassez de recursos para coleta e registro abrangente das informações, diferenças nos sistemas de saúde, infraestrutura tecnológica e capacitação dos profissionais também contribuem para a variabilidade na qualidade dos dados⁶.

Diversos fatores podem impactar negativamente o preenchimento das fichas, como a falta de compreensão sobre a importância das informações coletadas, a percepção da notificação como tarefa burocrática, a sobrecarga de trabalho, desmotivação entre os profissionais e a definição de outras prioridades pelos gestores locais¹⁵.

No início da crise sanitária, a gestão da pandemia exigia respostas rápidas e a adaptação dos SIS para coleta, análise e disseminação de dados em tempo real, essenciais à compreensão do cenário epidemiológico e à tomada de decisão. Para isso, os dados deveriam ser completos, confiáveis e oportunos^{16,17}.

Durante os primeiros anos da pandemia, o atraso nos registros pode ter sido agravado por limitações do SIVEP-Gripe, que inicialmente não contemplava campos específicos para COVID-19, além de instabilidades frequentes e ausência de interoperabilidade com outros sistemas, como o Sistema Gerenciador de Ambiente

Laboratorial. Soma-se a isso a carência de capacitação sistemática dos profissionais responsáveis pelo preenchimento das fichas⁶.

A baixa qualidade de dados evidenciada pode estar relacionada às instabilidades e indisponibilidades ocorridas no acesso aos dados do Sivep-Gripe. Ocorreram evidências de falta de disponibilidade de dados, como por exemplo, no início de junho de 2020, quando o painel Coronavírus, que é a principal plataforma oficial de monitoramento de dados da COVID-19, permaneceu indisponível por motivos políticos.

Nesta mesma época, o governo federal propôs alterações arbitrárias nas medições de variáveis e padrões de relatórios, em desacordo com as diretrizes epidemiológicas internacionais¹⁸.

Em dezembro de 2021 ocorreu novamente uma interrupção que durou ao menos 30 dias. Nesse mesmo período, o mundo já havia identificado uma variante de preocupação do SARS-CoV-2, a Ômicron, com alta taxa de transmissibilidade.

Esses episódios, conhecidos como “apagões de dados”, impactaram diretamente na indisponibilidade de dados e inviabilizaram registros e análises de indicadores importantes para o controle da doença. Portanto, qualquer interrupção na disponibilidade dos dados e dos sistemas compromete a capacidade de resposta rápida, dificulta a identificação de eventos relevantes e prejudica a formulação de recomendações adequadas¹⁹.

Este estudo permitiu uma análise abrangente da qualidade de dados de COVID19 no Brasil ao longo de quatro anos de pandemia, proporcionando subsídios para o aprimoramento dos registros no SIVEP-Gripe. Embora este estudo revele achados importantes para a vigilância e avaliação da qualidade da informação em

saúde, ele apresenta limitações como, a análise restrita aos casos confirmados, e a ausência de limpeza do banco para eliminação de duplicidades e inconsistências , devido à falta de acesso aos dados individuais.

Como ponto forte, destaca-se a originalidade da análise conjunta da completude e da oportunidade dos registros, fornecendo evidências que podem subsidiar propostas de melhoria no preenchimento das fichas e no funcionamento do sistema. Um SIVEP-Gripe mais qualificado e oportuno contribui para o monitoramento eficaz do território e para decisões mais assertivas em saúde pública.

No âmbito dos Sistemas de Informação em Saúde (SIS), os registros, processamento e análise de dados, ainda enfrentam desafios. Segundo Melo e Soares⁵, existem dificuldades relacionadas aos SIS, como a ausência de uma política de incentivo ao registro correto dos dados, dificultando a geração de informações corretas e, conseqüentemente, a realização de ações eficazes²⁰.

Esses achados sugerem que os esforços realizados durante a pandemia foram eficazes para tornar os processos mais ágeis, mas insuficientes para elevar de forma consistente a qualidade da completude dos registros. A combinação de avanços na oportunidade e persistência de falhas na completude reforça a necessidade de que os municípios e estados, nos serviços de vigilância epidemiológica, implementem rotinas periódicas de avaliação dos registros, com o objetivo de mitigar os problemas de baixa completude.

Além disso, é fundamental instituir processos contínuos de capacitação dos profissionais envolvidos na notificação, visando garantir a padronização e a confiabilidade das informações. Políticas de capacitação contínua, avaliações

periódicas, supervisão ativa e monitoramento sistemático da qualidade da informação são fundamentais para reduzir as disparidades regionais e assegurar dados confiáveis, indispensáveis ao planejamento de ações em saúde coletiva.

REFERÊNCIAS

1. Brasil. Ministério da Saúde. Painel Coronavírus [Internet]. Brasília: Ministério da Saúde; 2024. Available from: <https://covid.saude.gov.br/>.
2. Maciel EL, Jabor PM, Goncalves Jr E, Siqueira PC, Prado TN do, Zandonade E. Estudo da qualidade dos Dados do Painel COVID-19 para crianças, adolescente e jovens, Espírito Santo – Brasil, 2020. Esc Anna Nery [Internet]. 2021;25(spe). Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1414-81452021000500204&lang=pt
3. Siqueira PC, Maciel ELN, Catão R de C, Brioschi AP, Silva TCC da, Prado TN do. Completude das fichas de notificação de febre amarela no estado do Espírito Santo, 2017 *. Epidemiol e Serviços Saúde [Internet]. 2020 Jun;29(3):2019402. Available from: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S2237-96222020000300309&lng=pt&nrm=iso&tlng=pt
4. Massoudi BL, Sobolevskaia D. Keep Moving Forward: Health Informatics and Information Management beyond the COVID-19 Pandemic. Yearb Med Inform [Internet]. 2021 Aug 3;30(01):075–83. Available from: <http://www.thieme-connect.de/DOI/DOI?10.1055/s-0041-1726499>
5. Melo, Heline Maica Coelho de; Soares DJ. A importância do Sistema de Informação na gestão do Sistema Único de Saúde (SUS) [Internet]. Universidade da Integração

- Internacional da Lusofonia Afro-Brasileira (UNILAB); 2018. Available from: https://repositorio.unilab.edu.br/jspui/bitstream/123456789/702/3/2018_arti_hmelo.pdf
6. Lana RM, Coelho FC, Gomes MF da C, Cruz OG, Bastos LS, Villela DAM, et al. Emergência do novo coronavírus (SARS-CoV-2) e o papel de uma vigilância nacional em saúde oportuna e efetiva. *Cad Saude Publica* [Internet]. 2020;36(3):19620. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2020000300301&tlng=pt
 7. World Health Organization. Communicable disease surveillance and response systems: guide to monitoring and evaluating. World Health Organization; 2006.
 8. Doyle TJ. Completeness of Notifiable Infectious Disease Reporting in the United States: An Analytical Literature Review. *Am J Epidemiol* [Internet]. 2002 May 1;155(9):866–74. Available from: <https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/155.9.866>
 9. Costa-Santos C, Neves AL, Correia R, Santos P, Monteiro-Soares M, Freitas A, et al. COVID-19 surveillance data quality issues: a national consecutive case series. *BMJ Open* [Internet]. 2021 Dec;11(12):e047623. Available from: <https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2020-047623>
 10. Corrêa PRL, Ishitani LH, Abreu DMX de, Teixeira RA, Marinho F, França EB. A importância da vigilância de casos e óbitos e a epidemia da COVID-19 em Belo Horizonte, 2020. *Rev Bras Epidemiol* [Internet]. 2020;23:200061. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-790X2020000100206&tlng=pt

11. Ribas FV, Custódio ACD, Toledo LV, Henriques BD, Sedyama CMN de O, Freitas BAC de. Completude das notificações de síndrome respiratória aguda grave no âmbito nacional e em uma regional de saúde de Minas Gerais, durante a pandemia de COVID-19, 2020. *Epidemiol e Serviços Saúde* [Internet]. 2022;31(2). Available from: <https://doi.org/10.1590/S1679-49742022000200004>
12. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise Epidemiológica e Vigilância de Doenças não Transmissíveis. e-SUS Sinan : manual de instruções [Internet]. Brasília: Ministério da Saúde; 2022. Available from: https://bvsms.saude.gov.br/bvs/publicacoes/esus_sinan_manual_instrucoes.pdf
13. Vigilância BM da SS de V em SD de. Sistema de Informação de Agravos de Notificação – Sinan: manual de normas e rotinas. [Internet]. 2006. Available from: https://portalsinan.saude.gov.br/images/documentos/Aplicativos/sinan_net/Manual_Normas_e_Rotinas_2_edicao.pdf
14. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. Roteiro para uso do Sinan net, análise da qualidade da base de dados e cálculo de indicadores epidemiológicos e operacionais. Brasília. 2008.
15. Xavier DR, Albuquerque MP de, Sousa-Carmo SVT de, Pinter A. Evaluation of completeness and timeliness of data in the National Information System for Notifiable Diseases for spotted fever in the state of São Paulo, Brazil, 2007-2017. *Epidemiol e Serviços Saúde* [Internet]. 2023;32(2):2022416. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2237-96222023000200306&tlng=en
16. Neri SCC. A qualidade dos dados dos sistemas de informação em saúde aplicados na atenção à saúde materno-infantil [Internet]. UNIVERSIDADE FEDERAL DA

- BAHIA; 2016. Available from: [https://repositorio.ufba.br/bitstream/ri/21636/1/Tese Suzana Costa Carvalho. 2016.pdf](https://repositorio.ufba.br/bitstream/ri/21636/1/Tese%20Suzana%20Costa%20Carvalho.2016.pdf)
17. Clare T, Twohig KA, O'Connell AM, Dabrera G. Timeliness and completeness of laboratory-based surveillance of COVID-19 cases in England. Public Health [Internet]. 2021 May;194:163–6. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0033350621001219>
 18. Carvalho CA de, Carvalho VA de, Campos MAG, Oliveira BLCA de, Diniz EM, Santos AM dos, et al. Delay in death reporting affects timely monitoring and modeling of the COVID-19 pandemic. Cad Saude Publica [Internet]. 2021;37(7). Available from: http://www.scielo.org/scielo.php?script=sci_arttext&pid=S0102-311X2021000705018&lang=pt
 19. Villela DAM, Gomes MF da C. O impacto da disponibilidade de dados e informação oportuna para a vigilância epidemiológica. Cad Saude Publica [Internet]. 2022;38(7):e00115122. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2022000700101&tlng=pt
 20. Moreira Braz R, Luiz Tauil P, Carolina Faria e Silva Santelli A, Jesus Fernandes Fontes C. Avaliação da completude e da oportunidade das notificações de malária na Amazônia Brasileira, 2003-2012. Epidemiol e Serviços Saúde [Internet]. 2016 Jan;25(1):10–1. Available from: http://www.iec.pa.gov.br/template_doi_ess.php?doi=10.5123/S1679-49742016000100003&scielo=S2237-96222016000100021

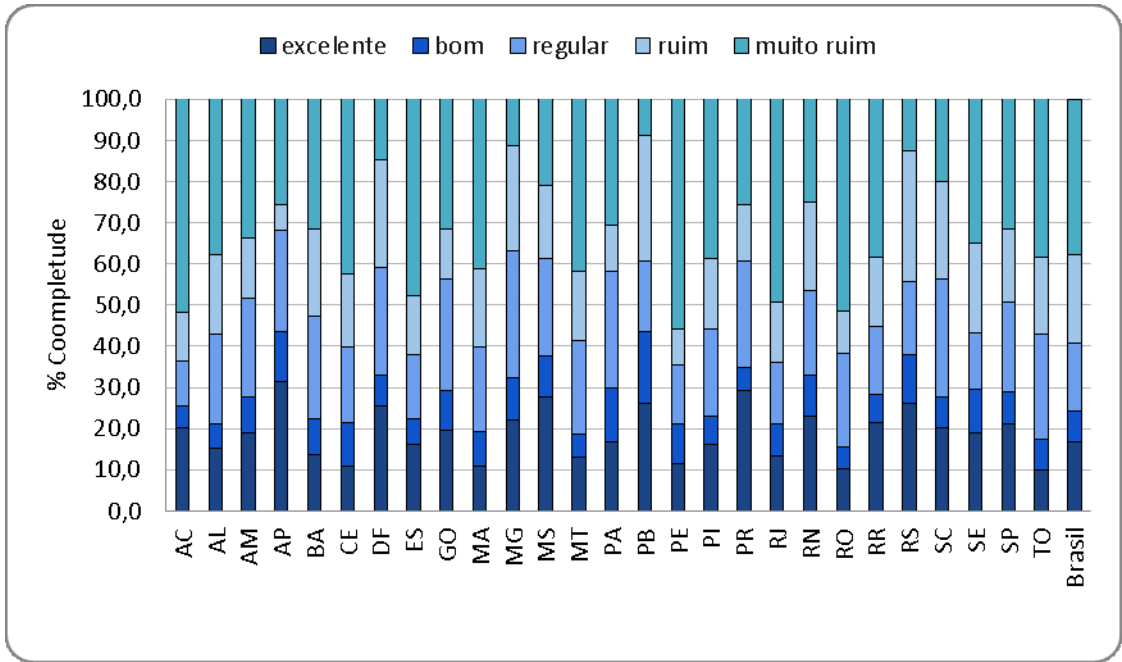
Figuras e tabelas

Tabela 1: Percentual de variáveis dos casos de SRAG por COVID-19 segundo Grau de Completude por ano de notificação no Brasil, 2020-2023.

Grau de Completude (66 Variáveis)	2020 (N=717 615)	2021 (N=1211609)	2022 (N=236884 88)	2023 (N=346 88)
Excelente	17,5%	18,2%	16,7%	15,2%
Bom	7,9%	3,0%	9,1%	12,1%
Regular	25,4%	19,7%	19,7%	18,2%
Ruim	11,1%	22,7%	22,7%	24,2%
Muito ruim	38,1%	36,4%	31,8%	30,3%

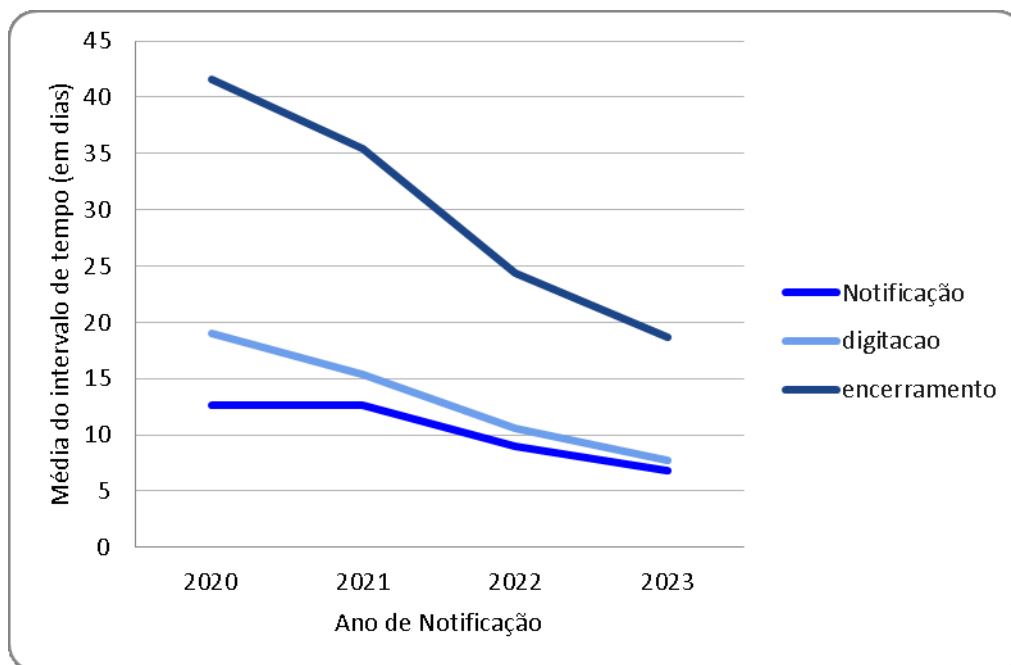
Fonte: Sivep-Gripe, 29/01/2024

Gráfico 1: Percentual de variáveis dos casos de SRAG por COVID-19 segundo grau de completude e UF de residência, Brasil, 2020-2023.



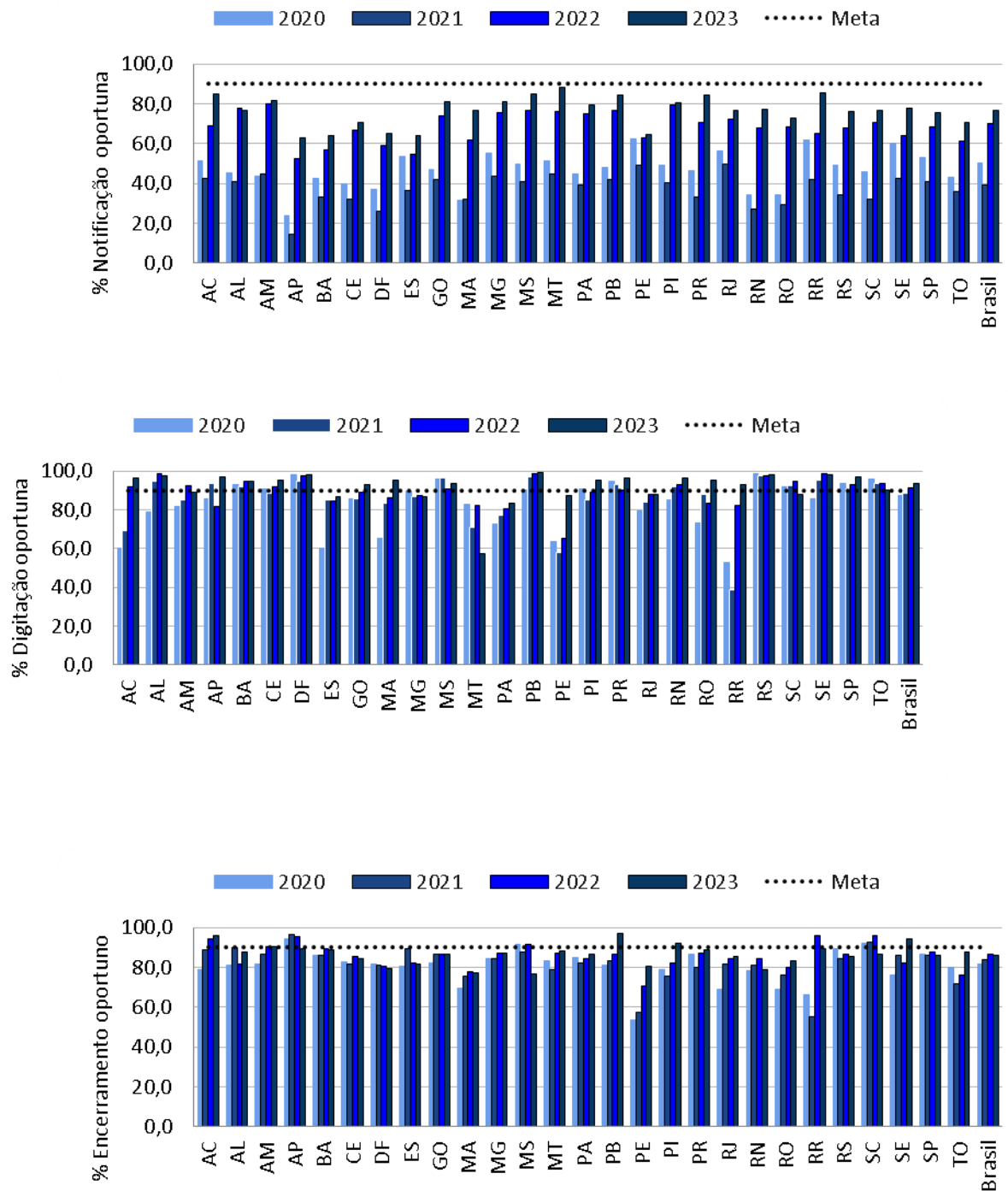
Fonte: Sivep-Gripe, 29/01/2024

Gráfico 2: Média do intervalo de tempo (em dias) das notificações, digitações e encerramento das SRAG por COVID-19 no Brasil, 2020-2023.



Fonte: Sivep-Gripe, 29/01/2024

Figura 1: Percentual de notificação, digitação e encerramento oportunos dos casos de SRAG por COVID19 segundo UF do Brasil, 2020-2023



Fonte: Sivep-Gripe, 29/01/2024

6 CONSIDERAÇÕES FINAIS

Esta tese permitiu uma análise abrangente da qualidade de dados de COVID-19 no Brasil, com foco na completude e na oportunidade dos registros da Síndrome Respiratória Aguda Grave (SRAG) no SIVEP-Gripe, entre 2020 e 2023, e na análise de confiabilidade dos dados nacionais e internacionais à luz da Lei de Benford.

A análise da qualidade dos dados de Covid-19 sugere possíveis fragilidades estruturais e operacionais que comprometeram a confiabilidade, a completude e a oportunidade das informações disponibilizadas durante a pandemia. Essas limitações reforçam a urgência de investimentos financeiros contínuos e direcionados às ações de vigilância epidemiológica para garantir e apoiar sistemas de informação mais robustos, oportunos e transparentes.

Ainda assim, foi possível identificar avanços ao longo do tempo, com redução progressiva do atraso nos registros e melhorias graduais na qualidade da informação, resultado de esforços institucionais, maior pressão social por transparência e adaptação dos sistemas às exigências impostas pela crise sanitária. Esses achados demonstram tanto a vulnerabilidade inicial quanto a capacidade de ajuste dos Sistemas de Informação em Saúde diante de uma situação emergencial.

Todavia, a qualidade da informação depende não apenas da tecnologia disponível, mas de condições estruturais como financiamento contínuo, integração entre sistemas, padronização de processos, estabilidade institucional e valorização dos profissionais de saúde. Nesse sentido, o fortalecimento da vigilância epidemiológica, especialmente por meio de investimentos financeiros estáveis e adequados, torna-se um componente central para garantir a sustentabilidade e melhorias nos registros de dados.

A experiência da pandemia de Covid-19 evidencia que a gestão da informação deve ser tratada como eixo estratégico da preparação da capacidade de preparação e resposta para emergências, demandando interoperabilidade, capacitação permanente, auditoria da qualidade dos dados e mecanismos de transparência ativa. É necessário construir um modelo resiliente e sistêmico, que articule infraestrutura tecnológica, competências institucionais e estabilidade política e financeira para enfrentar eventos inesperados com agilidade e equidade.

Conclui-se que garantir dados confiáveis, completos e oportunos é um imperativo técnico, político e social para sustentar decisões baseadas em evidências, fortalecer a vigilância epidemiológica e proteger a saúde da população em futuras crises.

Por fim, esta tese contribui para a literatura nacional ao oferecer evidências sobre a qualidade dos registros de COVID-19 e ao propor reflexões sobre a governança dos dados em contextos de emergência sanitária. Espera-se que os resultados aqui apresentados subsidiem estratégias de aprimoramento dos sistemas de informação em saúde, orientem práticas mais qualificadas de registro e fomentem uma cultura de informação sólida, transparente e orientada à tomada de decisão no âmbito do Sistema Único de Saúde.

7 REFERÊNCIAS

ALMEIDA FILHO, Naomar de; BARRETO, Mauricio L. Epidemiologia & Saúde: fundamentos, métodos, aplicações. In: Epidemiologia & saúde: fundamentos, métodos, aplicações. 2011. p. 699-699.

AMANI, Behnam; AMANI, Bahman. Efficacy and safety of nirmatrelvir/ritonavir (Paxlovid) for COVID-19: A rapid review and meta-analysis. *Journal of Medical Virology*, v. 95, n. 2, p. e28441, 2023

ARAGÃO, Júlio et al. Introdução aos estudos quantitativos utilizados em pesquisas científicas. *Revista práxis*, v. 3, n. 6, 2011.

BALASHOV, Vadim S.; YAN, Yuxing; ZHU, Xiaodi. Using the Newcomb–Benford law to study the association between a country's COVID-19 reporting accuracy and its development. *Scientific reports*, v. 11, n. 1, p. 1-11, 2021.

Balsari, S, Buckee, C, Khanna, T. Which COVID-19 data can you trust? *Harvard Business Review*, <https://hbr.org/2020/05/which-COVID-19-data-can-you-trust> (2020, accessed 1 September 2020).

BAQUI, Pedro et al. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *The Lancet Global Health*, v. 8, n. 8, p. e1018-e1026, 2020.

BENFORD, F. The law of anomalous numbers. *Proceedings of The American Philosophical Society*, v. 78, p. 551-572, 1938. Disponível em: <https://www.jstor.org/stable/984802>. Acesso em 24/10/22

BRASIL. Ministério da Saúde. CIEVS – Centro de Informações Estratégicas em Vigilância em Saúde. Disponível em: <https://www.gov.br/saude/pt-br/composicao/svsa/cievs>. Acesso em: 18 set. 2025.

BRASIL. Lei nº 6.259, de 30 de outubro de 1975. Dispõe sobre a organização das ações de Vigilância Epidemiológica, sobre o Programa Nacional de Imunizações, estabelece normas relativas à notificação compulsória de doenças, e dá outras providências. **Diário Oficial da União**, 1975.

70

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. Coordenação Geral de Doenças Transmissíveis. Gerência Técnica do Sinan. Roteiro para uso do Sinan net, análise da qualidade da base de dados e cálculo de indicadores epidemiológicos e operacionais. Brasília; 2008.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise Epidemiológica e Vigilância de Doenças não Transmissíveis. e-SUS Sinan : manual de instruções [recurso eletrônico] / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Análise Epidemiológica e Vigilância de Doenças não Transmissíveis. – Brasília : Ministério da Saúde, 2022.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Guia de vigilância epidemiológica : emergência de saúde pública de importância nacional pela doença pelo coronavírus 2019 – covid-19 / Ministério da Saúde, Secretaria de Vigilância em Saúde. – Brasília : Ministério da Saúde, 2022. 131 p. : il.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde e Ambiente. *Guia de vigilância integrada da covid-19, influenza e outros vírus respiratórios de importância em saúde pública* [recurso eletrônico]. Brasília: Ministério da Saúde, 2024. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/guias-e-manuais/2024/guia-vigilancia-integrada-da-covid-19-influenza-e-outros-virus-respiratorios-de-importancia-em-saude-publica>. Acesso em: 6 set. 2025.

BRESSER-PEREIRA, Luiz Carlos. Financiamento da Covid-19, inflação e restrição fiscal. *Brazil. J. Polit. Econ.*, São Paulo, v. 40, n. 4, p. 604-621, 2020.

CAMPOLIETI, Michele. Mortes por COVID-19 nos EUA: Lei de Benford e subnotificação. *Revista de Saúde Pública*, v. 44, n. 2, pág. e268-e271, 2022.

CARMO, Carlos Roberto Souza; DE LIMA CANEPPELE, Fernando; NUNES, Fábio Caixeta. ANALYSIS OF COVID-19 CONTAMINATION AND DEATHS CASES IN BRAZIL ACCORDING TO THE NEWCOMB-BENFORD. *Brazilian Journal of Biometrics*, v. 39, n. 4, p. 522-535, 2021.

CAVALCANTE, João Roberto et al. COVID-19 en Brasil: evolución de la epidemia hasta la semana epidemiológica 20 de 2020. *Epidemiologia e Serviços de Saúde*, v. 29, n. 4, 2020.

CHAUHAN, Himanshu; US, Saurabha; SINGH, Sujeet Kumar. Health Information and Its Crucial Role in Policy Formulation and Implementation. *Journal of Health Management*, v. 23, n. 1, p. 54-62, 2021.

CICHY, Corinna; RASS, Stefan. Fuzzy Expert Systems for Automated Data Quality Assessment and Improvement Processes. In: *EKAU (Posters & Demos)*. 2020. p. 7-11.

CONSELHO NACIONAL DE SAÚDE. Resolução nº 588, de 12 de julho de 2018. Institui a Política Nacional de Vigilância em Saúde. 2018.

CUNHA, F.C.R. Aplicações da Lei Newcomb-Benford à Auditoria de Obras Públicas. Dissertação (Mestrado em Regulação e Gestão de Negócios) - Universidade de Brasília, Brasília, 2013.

DONG, Ensheng et al. The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned. *The Lancet Infectious Diseases*, 2022.

FAGUNDES, Priscila Basto; DE MACEDO, Douglas Dyllon Jeronimo; FREUND, Gislaine Parra. A produção científica sobre qualidade de dados em big data: um estudo na base de dados Web of Science. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, v. 16, n. 1, p. 194-210, 2018.

FARHADI, Noah. Can we rely on COVID-19 data? An assessment of data from over 200 countries worldwide. *Science Progress*, v. 104, n. 2, p. 00368504211021232, 2021.

FARHADI, Noah; LAHOOTI, Hooshang. Forensic Analysis of COVID-19 Data from 198 Countries Two Years after the Pandemic Outbreak. *COVID*, v. 2, n. 4, p. 472-484, 2022.

FERNÁNDEZ-DE-LAS-PEÑAS, César et al. Prevalence of post-COVID-19 symptoms in hospitalized and non-hospitalized COVID-19 survivors: A systematic review and meta-analysis. *European journal of internal medicine*, v. 92, p. 55-70, 2021.

Freire, M.C.M.; Pattussi M.P. Tipos de estudos. IN: ESTRELA, C. Metodologia científica. Ciência, ensino e pesquisa. 3ª ed. Porto Alegre: Artes Médicas, 2018. p.109-127.

FIOCRUZ. Fundação Oswaldo Cruz. Observatório Covid-19 - Boletim Especial: Balanço de dois anos da pandemia Covid-19. 2022. Disponível em: <https://portal.fiocruz.br/documentos-produzidos-pelo-observatorio-covid-19>.

Fundação Oswaldo Cruz (Fiocruz). Observatório Covid-19 - Boletim Especial: Balanço de dois anos da pandemia Covid-19. 2022. Disponível em: <https://portal.fiocruz.br/documentos-produzidos-pelo-observatorio-covid-19>.

FRANCO NETTO, Guilherme et al. Vigilância em Saúde brasileira: reflexões e contribuição ao debate da 1a Conferência Nacional de Vigilância em Saúde. *Ciência & Saúde Coletiva*, v. 22, p. 3137-3148, 2017.

GIBBONS, Cheryl L. et al. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC public health*, v. 14, n. 1, p. 1-17, 2014.

GÓMEZ-CAMPONOV, Mariana et al. Monitoring the Paraguayan epidemiological dengue surveillance system (2009-2011) using Benford's law. *Biomédica*, v. 36, n. 4, p. 583-592, 2016.

GUERRISI, Caroline et al. The potential value of crowdsourced surveillance systems in supplementing sentinel influenza networks: the case of France. *Eurosurveillance*, v. 23, n. 25, p. 1700337, 2018.

HÄRTING, Ralf-Christian; LEWONIEWSKI, Włodzimierz. Main influencing factors of quality determination of collaborative open data pages. *Information*, v. 11, n. 6, p. 283, 2020.

HOTCHKISS, David R.; DIANA, Mark L.; FOREIT, Karen G. Fleischman. How can routine health information systems improve health systems functioning in low-and middle-income countries? Assessing the evidence base. *Health information technology in the International context*, v. 12, p. 25-58, 2012.

HUI, David S. et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases*, v. 91, p. 264-266, 2020.

IDROVO, Alvaro Javier; MANRIQUE-HERNÁNDEZ, Edgar Fabián. <? covid19?> Data Quality of Chinese Surveillance of COVID-19: Objective Analysis Based on WHO's Situation Reports. *Asia Pacific Journal of Public Health*, v. 32, n. 4, p. 165-167, 2020.

ISEA, R. How Valid are the Reported Cases of People Infected with COVID-19 in the World? *Int. J. Coronaviruses* 2020, 1, 53–56

Johns Hopkins University. Coronavirus COVID-19 Global Cases by Johns Hopkins CSSE [Internet]. Disponível em: <https://coronavirus.jhu.edu/data/cumulative-cases>. Acesso em 20/01/2023.

LAL, Arush et al. Optimizing pandemic preparedness and response through health information systems: lessons learned from Ebola to COVID-19. *Disaster medicine and public health preparedness*, v. 16, n. 1, p. 333-340, 2022.

LANA, Raquel Martins et al. Emergência do novo coronavírus (SARS-CoV-2) e o papel de uma vigilância nacional em saúde oportuna e efetiva. *Cadernos de Saúde Pública*, v. 36, p. e00019620, 2020.

LEE, KB.; HAN, S.; JEONG, Y. COVID-19, achatamento da curva e lei de Benford *Physica A: Statistical Mechanics and its Applications*, v. 559, artigo 125090, 2020.

LIMA-COSTA, Maria Fernanda; BARRETO, Sandhi Maria. Tipos de estudos epidemiológicos: conceitos básicos e aplicações na área do envelhecimento. *Epidemiologia e serviços de saúde*, v. 12, n. 4, p. 189-201, 2003.

LU, Hongzhou; STRATTON, Charles W.; TANG, Yi-Wei. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *Journal of medical virology*, v. 92, n. 4, p. 401-402, 2020.

MAHMOOD, Sultan et al. Global preparedness against COVID-19: we must leverage the power of digital health. *JMIR Public Health and Surveillance*, v. 6, n. 2, p. e18980, 2020.

MANRIQUE-HERNÁNDEZ, EF; FERNÁNDEZ-NIÑO, JA; IDROVO, AJ Desempenho global da vigilância epidemiológica do vírus Zika: avaliação rápida de uma epidemia em andamento. *Saúde Pública*, v. 143, p. 14-16, 2017.

MASSOUDI, Barbara L.; SOBOLEVSKAIA, Diana. Keep moving forward: health informatics and information management beyond the COVID-19 pandemic. *Yearbook of Medical Informatics*, v. 30, n. 01, p. 075-083, 2021.

MEBANE. W. R. Election Forensics: vote counts and Benford's Law. *The Society for Political Methodology. Papers, Posters and Syllabi*, N° 620. 2006.

MELO, Heline Maica Coelho de; SOARES, Denise Josino. A importância do Sistema de Informação na gestão do Sistema Único de Saúde (SUS). 2018.

MINISTÉRIO DA SAÚDE. Infecção humana pelo novo coronavírus (2019-nCoV). *Boletim Epidemiológico* 2020.

<https://portalarquivos2.saude.gov.br/images/pdf/2020/fevereiro/07/BE-COE-Coronavirus-n020702.pdf>.

MINISTERIO DA SAUDE. Boletim Epidemiológico Especial: Doença pelo Novo Coronavírus – COVID-19. 2023.

MOHAMMADI, Faegheh et al. Is the information fit for use? Exploring teachers perceived information quality indicators for Farsi web-based learning resources. *Malaysian Journal of Library & Information Science*, v. 20, n. 1, 2015.

NAUMANN, Felix (Ed.). *Quality-driven query answering for integrated information systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.

NEGRO-CALDUCH, Elsa et al. Health Information Systems in the COVID-19 Pandemic: A Short Survey of Experiences and Lessons Learned From the European Region. *Frontiers in public health*, v. 9, 2021.

NERI, Suzana Costa Carvalho. *A qualidade dos dados dos sistemas de informação em saúde aplicados na atenção à saúde materno-infantil*. 2016.

MINISTÉRIO DA SAÚDE. Secretaria de Atenção Especializada à Saúde (SAES). *Protocolo de Manejo Clínico para o Novo Coronavírus (2019-nCoV)*. 2020.

NIGRINI, Mark J. *Forensic analytics: Methods and techniques for forensic accounting investigations*. John Wiley & Sons, 2020.

POORANSINGH, Shalini et al. COVID-19 Highlights a Critical Need for Efficient Health Information Systems for Managing Epidemics of Emerging Infectious Diseases. *Frontiers in Public Health*, v. 9, 2021.

NIQUINI, Roberta Pereira et al. SRAG por COVID-19 no Brasil: descrição e comparação de características demográficas e comorbidades com SRAG por influenza e com a população geral. *Cadernos de Saúde Pública*, v. 36, 2020.

OMS. Organização Mundial da Saúde. World Health Organization (WHO). *Coronavirus disease (COVID-19)*. Disponível em: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19>

PAIM, Isis; NEHMY, Rosa Maria Quadros, GUIMARÃES, César Geraldo. Problematização do conceito "Qualidade" da Informação. *Perspectivas em Ciência da Informação*, v. 1, n. 1, 1996, p. 111–119.

PRADO, Marcelo Freitas do et al. Análise da subnotificação de COVID-19 no Brasil. *Revista Brasileira de Terapia Intensiva*, v. 32, p. 224-228, 2020.

RAHMANDAD, Hazhir; LIM, Tse Yang; STERMAN, John. Behavioral dynamics of COVID-19: Estimating underreporting, multiple waves, and adherence fatigue across 92 nations. *System Dynamics Review*, v. 37, n. 1, p. 5-31, 2021.

RAUCH, Bernhard et al. Fact and fiction in EU-governmental economic data. *German Economic Review*, v. 12, n. 3, p. 243-255, 2011.

RIBAS, Fábio Vieira et al. Completeness of notifications of severe acute respiratory syndrome at the national level and in a regional health care in the state of Minas Gerais, during the COVID-19 pandemic, 2020. *Epidemiologia e Serviços de Saúde*, v. 31, 2022.

ROMERO, Dalia E.; CUNHA, Cynthia Braga da. Avaliação da qualidade das variáveis sócio-econômicas e demográficas dos óbitos de crianças menores de um ano registrados no Sistema de Informações sobre Mortalidade do Brasil (1996/2001). *Cadernos de Saúde Pública*, v. 22, p. 673-681, 2006.

SCHMIDT, Andrea E.; ABBOUD, Linda A.; BOGAERT, Petronille. Making the case for strong health information systems during a pandemic and beyond. *Archives of Public Health*, v. 79, n. 1, p. 1-6, 2021.

SOWE, Alieu; GARIBOLDI, Maria Isabella. An assessment of the quality of vaccination data produced through smart paper technology in The Gambia. *Vaccine*, v. 38, n. 42, p. 6618-6626, 2020

VASCONCELOS, Giovani L.; BRUM, Arthur A.; ALMEIDA, Francisco A. G; MACÊDO, Antônio M. S.; DUARTE-FILHO, Gerson C.; OSPINA, Raydonal. Standard and anomalous waves of COVID-19: A multiple-wave growth model for epidemics. *medRxiv*, 2021. Disponível em: <https://www.medrxiv.org/content/10.1101/2021.01.31.21250867v4>.

VILLELA, Daniel Antunes Maciel; GOMES, Marcelo Ferreira da Costa. O impacto da disponibilidade de dados e informação oportuna para a vigilância epidemiológica. *Cadernos de Saúde Pública*, v. 38, p. e00115122, 2022.

WANG, Richard Y.; STRONG, Diane M. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information System*, v.12, n.4, 1996, p.5-34.

WOROBAY, Michael et al. The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science*, v. 377, n. 6609, p. 951-959, 2022

World Health Organization. Transmission of SARS-CoV-2: implications for infection prevention precautions. Disponível em: <https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions>. Acesso em 04/09/20.

World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. Disponível em: <https://covid19.who.int/table>. Acesso em 02/09/2020.

WHO Health Information Systems. Toolkit on monitoring health systems strengthening, 2008.

ZHANG, Jin-jin et al. Fatores de risco e proteção para morbidade, gravidade e mortalidade por COVID-19. *Clinical Reviews in Allergy & Immunology*, v. 64, n. 1, pág. 90-107, 2023.

Apêndices

Apêndice 1 – Protocolo de Revisão Sistemática



UNIVERSIDADE FEDERAL DE SANTA CATARINA

CENTRO DE CIÊNCIAS DA SAÚDE

PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE COLETIVA

CURSO DE DOUTORADO EM SAÚDE COLETIVA

PROTOCOLO DE REVISÃO SISTEMÁTICA
Taina Barbie do Espírito Santo
OBJETIVO
Identificar estudos sobre qualidade e confiabilidade de dados de Covid-19
CRITÉRIOS DE INCLUSÃO
Publicações até novembro de 2022
Estudos com objetivo de análise de qualidade de dados de Covid-19
BASES ELETRÔNICAS DE BUSCA
PubMed/ Medline
Scopus
LILACS
SciELO
Web of Science
DESCRITORES E ESTRATÉGIAS DE BUSCA

1. **PubMed/ Medline**

("Coronavirus Infections"[Mesh] OR "Coronavirus Infections" OR "COVID-19"[Mesh] OR "Covid-19" OR "SARS-CoV-2"[Mesh] OR "SARS-CoV-2" OR "SARS Virus"[Mesh] OR "SARS Virus" OR "SARSCoV2" OR "SARS2" OR "COVID19" OR "COVID-2019" OR "COVID 2019" OR "SARS COV 2" OR "2019-nCoV" OR "2019ncov" OR "coronavírus" OR "nCoV 2019") AND ("Data Analysis"[Mesh] OR "Data Analysis" OR "Data Analyses" OR "Data Accuracy"[Mesh] OR "Data Accuracy" OR "Data Accuracies" OR "Data Qualities" OR "Data Quality" OR "Data Management"[Mesh] OR "Data Management" OR "Data Administration" OR "Newcomb-Benford" OR "Benford")

2. **Scopus**

("Coronavirus Infections" OR "Covid-19" OR "SARS-CoV-2" OR "SARS Virus" OR "SARSCoV2" OR "SARS2" OR "COVID19" OR "COVID-2019" OR "COVID 2019" OR "SARS COV 2" OR "2019-nCoV" OR "2019ncov" OR "coronavírus" OR "nCoV 2019") AND ("Data Analysis" OR "Data Analyses" OR "Data Accuracy" OR "Data Accuracies" OR "Data Qualities" OR "Data Quality" OR "Data Management" OR "Data Administration" OR "Newcomb-Benford" OR "Benford")

3. **LILACS**

("Data Analysis" OR "Data Analyses" OR "Data Accuracy" OR "Data Accuracies" OR "Data Qualities" OR "Data Quality" OR "Data Management" OR "Data Administration" OR "Newcomb-Benford" OR "Análise de Dados" OR "Análise dos Dados" OR "Análises de Dados" OR "Análises dos Dados" OR "Confiabilidade dos Dados" OR "Acurácia dos Dados" OR "Exatidão dos Dados" OR "Precisão dos Dados" OR "Qualidade dos Dados" OR "Gerenciamento de Dados" OR "Administração de Dados" OR "Gestão de Dados" OR "Análisis de Datos" OR "Análisis de los Datos" OR "Exactitud de los Datos" OR "Calidad de los Datos" OR "Manejo de Datos" OR "gestión de datos" OR "administración de datos") AND ("Coronavirus Infections" OR "Covid-19" OR "SARS-CoV-2" OR "SARS Virus" OR "SARSCoV2" OR "SARS2" OR "COVID19" OR "COVID-2019" OR "COVID 2019" OR "SARS COV 2" OR "2019-nCoV" OR "2019ncov" OR "coronavírus" OR "nCoV 2019" OR "Infecções por Coronavirus" OR "Vírus da SARS" OR "Infecciones por Coronavirus" OR "Virus del SRAS")

4. **SciELO**

("Data Analysis" OR "Data Analyses" OR "Data Accuracy" OR "Data Accuracies" OR "Data Qualities" OR "Data Quality" OR "Data Management" OR "Data

Administration" OR "Newcomb-Benford" OR "Benford" OR "Análise de Dados" OR "Análise dos Dados" OR "Análises de Dados" OR "Análises dos Dados" OR "Confiabilidade dos Dados" OR "Acurácia dos Dados" OR "Exatidão dos Dados" OR "Precisão dos Dados" OR "Qualidade dos Dados" OR "Gerenciamento de Dados" OR "Administração de Dados" OR "Gestão de Dados" OR "Análisis de Datos" OR "Análisis de los Datos" OR "Exactitud de los Datos" OR "Calidad de los Datos" OR "Manejo de Datos" OR "gestión de datos" OR "administración de datos") AND ("Coronavirus Infections" OR "Covid-19" OR "SARS-CoV-2" OR "SARS Virus" OR "SARSCoV2" OR "SARS2" OR "COVID19" OR "COVID-2019" OR "COVID 2019" OR "SARS COV 2" OR "2019-nCoV" OR "2019ncov" OR "coronavírus" OR "nCoV 2019" OR "Infecções por Coronavirus" OR "Vírus da SARS" OR "Infecciones por Coronavirus" OR "Virus del SRAS")

5. Web of Science

("Coronavirus Infections"[Mesh] OR "Coronavirus Infections" OR "COVID-19"[Mesh] OR "Covid-19" OR "SARS-CoV-2"[Mesh] OR "SARS-CoV-2" OR "SARS Virus"[Mesh] OR "SARS Virus" OR "SARSCoV2" OR "SARS2" OR "COVID19" OR "COVID-2019" OR "COVID 2019" OR "SARS COV 2" OR "2019-nCoV" OR "2019ncov" OR "coronavírus" OR "nCoV 2019") AND ("Data Analysis"[Mesh] OR "Data Analysis" OR "Data Analyses" OR "Data Accuracy"[Mesh] OR "Data Accuracy" OR "Data Accuracies" OR "Data Qualities" OR "Data Quality" OR "Data Management"[Mesh] OR "Data Management" OR "Data Administration" OR "Newcomb-Benford" OR "Benford")

ANEXOS

ANEXO 1: Ficha de Investigação da Síndrome Respiratória aguda Grave

Nº	
 MINISTÉRIO DA SAÚDE SISTEMA DE VIGILÂNCIA EM SAÚDE SIVIP-Gripe - SISTEMA DE INFORMAÇÃO DA VIGILÂNCIA EPIDEMIOLÓGICA DA GRIPE FICHA DE REGISTRO INDIVIDUAL - CASOS DE SÍNDROME RESPIRATÓRIA AGUDA GRAVE HOSPITALIZADO 28/09/2023	
CASO DE SÍNDROME RESPIRATÓRIA AGUDA GRAVE (SRAG-HOSPITALIZADO): Indivíduo com *SG que apresenta dispnéia/desconforto respiratório OU pressão persistente no tórax OU saturação de O₂ menor que 96% em ar ambiente OU coloração azulada dos lábios ou rosto. (*SG: indivíduo com quadro respiratório agudo, caracterizado por pelo menos dois (2) dos seguintes sinais e sintomas: febre (mesmo que referida), calafrios, dor de garganta, dor de cabeça, tosse, coriza, distúrbios olfativos ou gustativos). Para efeito de notificação no SIVIP-Gripe, devem ser considerados os casos de SRAG hospitalizados ou os óbitos por SRAG independente de hospitalização.	
1	Data do preenchimento da ficha de notificação:
2	Data de 1º sintomas
3	UF:
4	Município:
5	Código (IBGE):
6	Unidade de Saúde:
7	Código (CNES):
8	Tem CPF? ☐ 1-Sim 2-Não
9	Estrangeiro ☐ 1-Sim 2-Não
10	Cartão Nacional de Saúde (CNS):
11	Nome:
12	Data de nascimento:
13	(Ou) Idade:
14	Sexo: ☐ 1-Masc 2-Fem 9-Ign
15	Raça/Cor: ☐ 1-Branca 2-Preta 3-Amarela 4-Parda 5-Indígena 9-Ignorado
16	Gestante: ☐
17	Se indígena, qual etnia?
18	Se sim, qual?
19	Se membro de povo ou comunidade tradicional? ☐ 1-Sim 2-Não
20	Escolaridade: ☐ 0-Sem escolaridade/Analfabeto 1-Fundamental 1º ciclo (1ª a 5ª série) 2-Fundamental 2º ciclo (6ª a 8ª série) 3-Médio (1ª ao 3ª ano) 4-Superior 5-Não se aplica 9-Ignorado
21	Ocupação:
22	Nome da mãe:
23	CEP:
24	Município:
25	Código (IBGE):
26	Bairro:
27	Logradouro (Rua, Avenida, etc.):
28	Complemento (apto, casa, etc.):
29	(DDD) Telefone:
30	Zona: ☐ 1-Urbana 2-Rural 3-Periurbana 9-Ignorado
31	País: (se residente fora do Brasil)
32	Trata-se de caso nosocomial (infecção adquirida no hospital)? ☐ 1-Sim 2-Não 9-Ignorado
33	Paciente trabalha ou tem contato direto com aves, suínos, ou outro animal? ☐ 1-Sim 2-Não 9-Ignorado
34	Sinais e Sintomas: 1-Sim 2-Não 9-Ignorado ☐ Febre ☐ Tosse ☐ Dor de Garganta ☐ Dispneia ☐ Desconforto Respiratório ☐ Saturação O ₂ < 96% ☐ Diarreia ☐ Vômito ☐ Dor abdominal ☐ Fadiga ☐ Perda do olfato ☐ Perda do paladar ☐ Outros:
35	Possui fatores de risco/comorbidades? ☐ 1-Sim 2-Não 9-Ignorado Se sim, qual(s)? (Marcar X) ☐ Puérpera (até 45 dias do parto) ☐ Doença Cardiovascular Crônica ☐ Doença Hematológica Crônica ☐ Síndrome de Down ☐ Doença Hepática Crônica ☐ Asma ☐ Diabetes mellitus ☐ Doença Neurológica Crônica ☐ Outra Pneumopatia Crônica ☐ Imunodeficiência/Imunodepressão ☐ Doença Renal Crônica ☐ Obesidade, IMC ☐ Outros:
36	Recebeu vacina COVID-19? ☐ 1-Sim 2-Não 9-Ignorado
37	Data da dose vacina COVID-19: Data da 1ª dose: Data da 2ª dose: Data da dose reforço: Data da 3ª dose reforço:
38	Fabricante vacina COVID-19: Fabricante 1ª dose: Fabricante 2ª dose: Fabricante dose reforço: Fabricante 3ª dose reforço:
39	Lote da vacina COVID-19: Lote 1ª dose: Lote 2ª dose: Lote dose reforço: Lote 3ª dose reforço:
40	Recebeu vacina contra Gripe na última campanha? ☐ 1-Sim 2-Não 9-Ignorado
41	Data da vacinação:
Se < 6 meses: a mãe recebeu a vacina? ☐ 1-Sim 2-Não 9-Ignorado Se sim, data: a mãe amamenta a criança? ☐ 1-Sim 2-Não 9-Ignorado Se >= 6 meses e <= 8 anos: Data da dose única 1/1: (dose única para crianças vacinadas em campanhas de anos anteriores) Data da 1ª dose: (1ª dose para crianças vacinadas pela primeira vez) Data da 2ª dose: (2ª dose para crianças vacinadas pela primeira vez)	

Atendimento	42	Usou antiviral para gripe? ☐ 1-Sim 2-Não 9-Ignorado	43	Qual antiviral? ☐ 1-Osetamivir 2-Zanamivir 3-Outro, especifique: _____	44	Data início do tratamento: _____
	45	Recebeu tratamento antiviral para covid-19? ☐ 1-Sim 2-Não 9-Ignorado	46	Qual antiviral? ☐ 1-Nirmatrevir/ritonavir (Paxlovid®) 2-Baricitinibe (Jilumiant®) 3-Outro, especifique: _____	47	Data início do tratamento: _____
	48	Houve internação? ☐ 1-Sim 2-Não 9-Ignorado	49	Data da internação por SRAG: _____	50	UF de internação: _____
	51	Município de internação: _____	Código (IBGE): _____			
	52	Unidade de Saúde de internação: _____	Código (CNS): _____			
	53	Internado em UTI? ☐ 1-Sim 2-Não 9-Ignorado	54	Data da entrada na UTI: _____	55	Data da saída da UTI: _____
	56	Uso de suporte ventilatório: ☐ 1-Sim, invasivo 2-Sim, não invasivo 3-Não 9-Ignorado	57	Raio X de Tórax: ☐ 1-Normal 2-Infiltrado intersticial 3-Consolidação 4-Misto 5-Outro: _____ 6-Não realizado 9-Ignorado	58	Data do Raio X: _____
	59	Aspecto Tomografia: ☐ 1-Típico covid-19 2-Indeterminado covid-19 3-Atípico covid-19 4-Negativo para Pneumonia 5-Outro: _____ 6-Não realizado 9-Ignorado	60	Data da tomografia: _____		
61	Coletou amostra ☐ 1-Sim 2-Não 9-Ignorado	62	Data da coleta: _____	63	Tipo de amostra: ☐ 1-Secreção de Nariz-orofaringe 2-Lavado bronco-alveolar 3-Tecido post-mortem 4-Outra, qual? _____ 5-LCR 9-Ignorado	
Dados Laboratoriais	64	Nº Requisição do GAL: _____	65	Tipo do teste para pesquisa de antígenos virais: ☐ 1-Imunofluorescência (IF) 2-Teste rápido antigênico		
	66	Data do resultado da pesquisa de antígenos: _____	67	Resultado do Teste antigênico: ☐ 1-positivo 2-Negativo 3-Inconclusivo 4-Não realizado 5-Aguardando resultado 9-Ignorado		
	68	Laboratório que realizou o Teste antigênico: _____	Código (CNS): _____			
	69	Agente Etiológico - Teste antigênico: Positivo para Influenza? ☐ 1-Sim 2-Não 9-Ignorado Se sim, qual influenza? ☐ 1-Influenza A 2-Influenza B Positivo para outros vírus? ☐ 1-Sim 2-Não 9-Ignorado Se outros vírus respiratórios qual(is)? (marcar X) ☐ SARS-CoV-2 ☐ Vírus Sincial Respiratório ☐ Parainfluenza 1 ☐ Parainfluenza 2 ☐ Parainfluenza 3 ☐ Adenovírus ☐ Outro vírus respiratório, especifique: _____				
	70	Resultado da RT-PCR/outra método por Biologia Molecular: ☐ 1-Detectível 2-Não Detectível 3-Inconclusivo 4-Não realizado 5-Aguardando resultado 9-Ignorado		71	Data do resultado RT-PCR/outra método por Biologia Molecular: _____	
	72	Agente Etiológico - RT-PCR/outra método por Biologia Molecular: Positivo para Influenza? ☐ 1-Sim 2-Não 9-Ignorado Se sim, qual influenza? ☐ 1-Influenza A 2-Influenza B Influenza A, qual subtipo? ☐ 1-Influenza A(H1N1)pdm09 2-Influenza A(H3N2) 3-Influenza A não subtipado 4-Influenza A não subtipável 5-Inconclusivo 6-Outro, especifique: _____ Influenza B, qual linhagem? ☐ 1-Victoria 2-Yamagata 3-Não realizado 4-Inconclusivo 5-Outro, especifique: _____ Positivo para outros vírus? ☐ 1-Sim 2-Não 9-Ignorado Se outros vírus respiratórios, qual(is)? (marcar X) ☐ SARS-CoV-2 ☐ Vírus Sincial Respiratório ☐ Parainfluenza 1 ☐ Parainfluenza 2 ☐ Parainfluenza 3 ☐ Parainfluenza 4 ☐ Adenovírus ☐ Metapneumovírus ☐ Rotavírus ☐ Rinovírus ☐ Outro vírus respiratório, especifique: _____				
	73	Laboratório que realizou RT-PCR/outra método por Biologia Molecular: _____				
	74	Tipo de amostra sorológica para SARS-CoV-2: ☐ 1-Sangue/plasma/soro 2-Outra, qual? _____ 9-Ignorado		75	Data da coleta: _____	
	76	Tipo de Sorologia para SARS-CoV-2: ☐ 1-Teste rápido 2-Elisa 3-Quimioluminescência 4-Outro, qual? _____ Resultado do Teste Sorológico para SARS-CoV-2: ☐ IgG ☐ IgM ☐ IgA 1-Positivo 2-Negativo 3-Inconclusivo 4-Não realizado 5-Aguardando resultado 9-Ignorado		77	Data do resultado: _____	
	Conclusão	78	Classificação final do caso: ☐ 1-SRAG por Influenza 2-SRAG por outro vírus respiratório 3-SRAG por outro agente etiológico, qual _____ 4-SRAG não especificado 5-SRAG por covid-19			79
80		Evolução do Caso: ☐ 1-Cura 2-Óbito 3-Óbito por outras causas 9-Ignorado		81	Data da alta ou óbito: _____	
83		Número D.O.: _____				
84 OBSERVAÇÕES: _____						
85 Profissional de Saúde Responsável: _____				86 Registro Conselho/Matricula: _____		