



Cleber Vinicius Brito dos Santos<sup>a</sup>  
 <https://orcid.org/0000-0001-5710-2866>

Vanessa de Melo Ferreira<sup>a</sup>  
 <https://orcid.org/0000-0002-7308-0676>

João Roberto Cavalcante Sampaio<sup>a</sup>  
 <https://orcid.org/0000-0003-2070-3822>

Patricia Canto Ribeiro<sup>b</sup>  
 <https://orcid.org/0000-0002-2731-331X>

Hermano Albuquerque de Castro<sup>b</sup>  
 <https://orcid.org/0000-0003-1191-5671>

Adriana Coser Gutierrez<sup>b</sup>  
 <https://orcid.org/0000-0002-7305-5841>

Ingrid D'avilla Freire Pereira<sup>c</sup>  
 <https://orcid.org/0000-0002-2042-2871>

Marcio Fernandes Nehab<sup>d</sup>  
 <https://orcid.org/0000-0002-4659-6247>

Maria Martha Duque de Moura<sup>d</sup>  
 <https://orcid.org/0000-0002-4853-6491>

André Reynaldo Santos Perisse<sup>b</sup>  
 <https://orcid.org/0000-0002-5253-5774>

<sup>a</sup>Universidade do Estado do Rio de Janeiro, Departamento de Epidemiologia. Rio de Janeiro, RJ, Brasil.

<sup>b</sup>Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública Sérgio Arouca. Rio de Janeiro, RJ, Brasil.

<sup>c</sup>Fundação Oswaldo Cruz, Escola Politécnica de Saúde Joaquim Venâncio. Rio de Janeiro, RJ, Brasil.

<sup>d</sup>Fundação Oswaldo Cruz, Instituto Fernandes Figueira. Rio de Janeiro, RJ, Brasil.

#### Contato:

Cleber Vinicius Brito dos Santos  
E-mail:  
[santos.cleber@posgraduacao.uerj.br](mailto:santos.cleber@posgraduacao.uerj.br)

Os autores declaram que o estudo não foi subvencionado e que não há conflitos de interesses.

Os autores informam que este estudo não foi apresentado em evento científico.

## Incompletude da variável profissão/ocupação nos bancos de síndrome gripal, síndrome respiratória aguda grave e mortalidade, Brasil, 2020-2021

*Incompleteness of the profession/occupation variable in the flu-like syndrome, severe acute respiratory syndrome and mortality databases, Brazil, 2020-2021*

### Resumo

**Objetivo:** descrever a incompletude da informação sobre profissão/ocupação nas bases de dados de síndrome respiratória aguda grave (SRAG), síndrome gripal (SG) e no Sistema de Informações sobre Mortalidade (SIM) no Brasil. **Métodos:** estudo descritivo utilizando os bancos de dados de SRAG, SG e SIM. Calcularam-se percentuais de incompletude na variável profissão/ocupação segundo sexo, macrorregiões e unidades da federação, em 2020-2021. **Resultados:** o percentual de incompletude foi de 94,7% no banco de SG; 97,7% no de SRAG; e 17,0% no SIM. Em todas as macrorregiões a incompletude foi superior a 91,0% nos bancos de SG e SRAG; e superior a 13,0% no SIM. Todas as unidades da federação apresentaram percentuais de incompletude acima de 90,0% para SG; de 74,0% para SRAG; e de 6,8% para óbitos. Amapá apresentou maior percentual de incompletude na base de dados de SG (98,1%); Rio Grande do Sul (99,4%) na de SRAG; e Alagoas (45,0%) no SIM. **Conclusões:** observaram-se elevados percentuais de incompletude da variável profissão/ocupação nos sistemas de informação estudados. Recomenda-se uma articulação intersetorial, envolvendo representantes dos governos e dos trabalhadores, para formulação de estratégias que contornem a falta de informação sobre ocupação/profissão nas bases de dados relevantes para a vigilância em saúde.

**Palavras-chave:** vigilância epidemiológica; saúde do trabalhador; epidemiologia descritiva; síndrome respiratória aguda grave; registros de mortalidade.

### Abstract

**Objective:** to describe the incomplete filling out of the profession/occupation variable in the flu-like syndrome, severe acute respiratory syndrome and mortality databases in Brazil. **Methods:** descriptive study with secondary data from flu-like syndrome (FLS), severe acute respiratory syndrome (SARS) and mortality databases (SIM). We calculated the absolute and relative filling frequencies of the profession/occupation variable according to State, gender, regions, and federative units, for 2020 and 2021. **Results:** we found a 94.7% incompleteness on the FLS database, 97.7% of missing profession/occupation data on the SARS, and 17.0% on the SIM database. Incompleteness frequency was above 91.0% in all Brazilian regions for FLS and SARS. Incompleteness on the mortality database was over 13.0%. All federative units presented incompleteness above 90.0% for FLS, 74.0% for SARS, and 6.8% for mortality in all genders. Higher levels of missing data were found in the states of Amapá for FLS (98.1%), Rio Grande do Sul for SARS and Alagoas (45.0%) for mortality (99.4%). **Conclusions:** all databases showed a high incompleteness of the profession/occupation variable. We recommend an articulation between the Ministry of Health, Ministry of Labor and workers' representations to solve this lack of data on occupation/profession in public databases.

**Keywords:** epidemiologic surveillance; occupational health; descriptive epidemiology; severe acute respiratory syndrome; mortality registries.

## Introdução

A ocorrência de dados faltantes é um grande desafio para os estudos epidemiológicos e para a vigilância em saúde<sup>1</sup>. A falta de informações pode causar vieses nos estudos<sup>2</sup>, bem como pode prejudicar a interpretação dos dados e a formulação de estratégias de prevenção de agravos, inclusive aqueles de natureza ocupacional<sup>3</sup>.

Durante emergências em saúde pública, como a pandemia de covid-19, a disponibilidade de dados se torna ainda mais primordial para o enfrentamento da doença. Em 2015 e 2016, por exemplo, o adequado preenchimento de variáveis sobre malformações congênitas no Sistema de Informações sobre Nascidos Vivos (Sinasc) contribuiu para evidenciar a associação temporal entre o aumento de nascimentos de crianças com microcefalia e/ou alterações no sistema nervoso central e a epidemia pelo vírus Zika<sup>4</sup>. Ou seja, o acesso a dados abertos de boa qualidade em tempo real, durante a ocorrência de surtos, permite que pesquisadores e tomadores de decisão possam utilizar as evidências científicas para subsidiar o controle de doenças e agravos à saúde<sup>5</sup>.

O registro consistente de informações epidemiológicas é essencial para o entendimento dos fatores de risco de uma doença, sua dispersão geográfica e transmissibilidade. Adicionalmente, a integridade do conteúdo de sistemas de informação é imprescindível para a modelagem de dados que podem auxiliar no planejamento de respostas de contenção e controle de doenças, a fim de reduzir sua carga<sup>5</sup>.

Durante a pandemia de covid-19, por sua vez, informações sobre o perfil sociodemográfico e socioeconômico das populações mais atingidas foram fundamentais para a gestão da vigilância em saúde, sobretudo em um cenário de recrudescimento das desigualdades socioeconômicas. Nesse momento, as atividades laborais mobilizaram intenso fluxo de pessoas — a despeito das recomendações de distanciamento físico — e cada indivíduo infectado tornou-se um potencial agente transmissor do vírus SARS-CoV-2, seja nos transportes públicos, no ambiente de trabalho ou até mesmo no próprio domicílio<sup>6,7</sup>.

Nesse contexto, as notificações de síndrome respiratória aguda grave (SRAG), que incluem internações

e óbitos, foram amplamente utilizadas pela vigilância e por estudos como uma medida *proxy* para os casos de covid-19, pois correspondiam a cerca de 98% dos casos notificados de SRAG<sup>8-14</sup>.

Portanto, não saber a profissão ou ocupação dos indivíduos — nos casos leves e graves da doença e nos óbitos — se torna um problema para atuação e intervenção do setor da saúde, principalmente no que diz respeito à saúde do trabalhador, dado o impacto da covid-19 enquanto doença relacionada ao trabalho<sup>7,15</sup>.

Estrutural e historicamente, o campo da Epidemiologia aplicado aos serviços de saúde sofre com baixa qualidade de informações disponíveis acerca de sua força de trabalho em sistemas de informação nacionais<sup>16</sup>.

Este estudo teve por objetivo descrever a incompletude da variável profissão ou ocupação nas bases de dados de síndrome gripal (SG), SRAG e de mortalidade por todas as causas do Sistema de Informação de Mortalidade (SIM), durante os anos de 2020 e 2021, no Brasil.

## Métodos

### Desenho de estudo e contexto

Trata-se de um estudo descritivo, que quantifica a incompletude da variável profissão/ocupação, nos registros de notificação de SG, de SRAG e de mortalidade por todas as causas do SIM.

### Participantes

Foram considerados todos os registros de casos em pessoas com 18 anos de idade ou mais, no período de 2020 a 2021.

### Fontes de dados

Foram utilizados dados secundários provenientes dos sistemas de informação em saúde gerenciados pela Secretaria de Vigilância em Saúde, do Ministério da Saúde do Brasil. Os dados de notificação de SG foram coletados do eSUS VE Notifica<sup>17,18</sup>. As notificações de SRAG foram provenientes do Sistema de Informação da Vigilância Epidemiológica da Gripe

(Sivep-Gripe)<sup>18,19</sup>. Os registros de óbitos por todas as causas foram obtidos no SIM<sup>19,20</sup>. Todos os bancos de dados contêm informações anonimizadas no nível do indivíduo, sem permitir acesso a dados sensíveis e/ou rastreáveis. Informações mais detalhadas sobre os bancos de dados, como extensão e dicionário de dados, podem ser vistas nas referências citadas<sup>17-20</sup>.

### Variáveis

A variável de interesse para o estudo é a profissão/ocupação. Ela deve ser preenchida com um código referente à Classificação Brasileira de Ocupações (CBO)<sup>21</sup>. Considerou-se incompletude no registro a ausência do código CBO no respectivo campo. As demais variáveis de interesse foram sexo, UF e macrorregião.

### Análise dos dados

Foram calculadas as frequências absolutas e relativas de incompletude variável profissão/ocupação, segundo sexo, UF e macrorregião. Os percentuais de incompletude foram obtidos segundo a fórmula:

$$\left(1 - \left(\frac{\text{casos ou óbitos com CBO}}{\text{total de casos e óbitos notificados/declarados}}\right)\right) \times 100.$$

Todas as análises foram realizadas no *software* R, versão 4.1.2<sup>22</sup>, e o código utilizado está disponível online no repositório “Missing\_data\_br” da plataforma Github.<sup>23</sup>

### Considerações éticas

Este estudo foi realizado exclusivamente com dados secundários de acesso aberto e domínio público. Não houve apreciação por Comitê de Ética em Pesquisa.

### Resultados

As tabelas a seguir mostram as porcentagens de incompletude da variável profissão/ocupação nas três bases de dados em foco, estratificadas por sexo,

segundo UF, macrorregiões e total, na população com 18 anos ou mais.

A **Tabela 1** exibe os resultados das notificações de síndrome gripal. Verificou-se que todas as UFs apresentaram incompletude da profissão/ocupação acima de 90,0% na população com 18 anos ou mais, em ambos os sexos. Ao se considerar o sexo feminino, no Amapá, foi observada a maior porcentagem de incompletude da variável profissão/ocupação (97,8% de dados faltantes). Já no sexo masculino, o maior percentual de incompletude foi encontrado em Santa Catarina (98,9%). A UF com o menor percentual de incompletude foi Espírito Santo (91,0% de dados faltantes); e a UF com o maior foi Amapá (98,1% de ausência). Quanto às macrorregiões, a Centro-Oeste obteve os percentuais mais elevados de incompletude da variável profissão/ocupação para o sexo feminino (93,8%); e a região Sul, para o sexo masculino (97,8%).

Na **Tabela 2** estão presentes os resultados para as notificações de síndrome respiratória aguda grave. O Rio Grande do Sul apresentou os maiores percentuais de incompletude para ambos os sexos, com 99,5% para o sexo feminino e 99,3% para o sexo masculino. A UF com menor percentual de incompletude foi Amapá para o sexo feminino (82,4%) e Roraima (74,2%) para o sexo masculino. Considerando o total de notificações, a UF com menor percentual de incompletude da variável profissão/ocupação no banco SRAG foi Roraima (78,5%); e com o maior, o Rio Grande do Sul (99,4%). A região Sul foi a macrorregião com o maior percentual de incompletude na base de SRAG, em ambos os sexos (98,8% para o sexo feminino e 98,2% para o sexo masculino).

Por fim, a **Tabela 3** mostra os resultados do percentual de incompletude no SIM. Os resultados por sexo não se diferiram significativamente dos resultados agregados. O Paraná apresentou menor incompletude para ambos os sexos, com 6,8% para o sexo feminino e 7,8% para o sexo masculino. A UF com maior percentual de incompletude para o sexo feminino foi Alagoas (46,9%); e para o sexo masculino (46,3%), a Paraíba. Entre as UF, o menor percentual de incompletude foi no Paraná (7,3%); e o maior, em Alagoas (45,0%). A região Nordeste apresentou os maiores percentuais de incompletude tanto no sexo feminino (19,4%) quanto no masculino (22,7%).

**Tabela 1** Percentual de incompletude da variável profissão ou ocupação nos bancos de dados de síndrome gripal\*, segundo sexo, UF, macrorregiões, Brasil, 2020-2021

Região/UF	Síndrome gripal								
	Feminino			Masculino			Total		
	CBO preenchido	Total de casos	Incompletude (%)	CBO preenchido	Total de casos	Incompletude (%)	CBO preenchido	Total de casos	Incompletude (%)
Centro-Oeste	214.866	3.479.338	93,8	112.753	2.868.715	96,1	327.619	6.348.053	94,8
DF	41.100	506.676	91,9	19.112	389.831	95,1	60.212	896.507	93,3
GO	71.552	1.660.367	95,7	30.403	1.348.140	97,7	101.955	3.008.507	96,6
MS	77.167	799.799	90,4	49.496	671.504	92,6	126.663	1.471.303	91,4
MT	25.047	512.496	95,1	13.742	459.240	97,0	38.789	971.736	96,0
Nordeste	771.520	9.459.649	91,8	342.309	7.392.297	95,4	1.113.829	16.851.946	93,4
AL	29.103	496.017	94,1	12.083	387.581	96,9	41.186	883.598	95,3
BA	231.644	2.372.241	90,2	96.919	1.909.846	94,9	328.563	4.282.087	92,3
CE	106.355	1.464.263	92,7	48.833	1.160.978	95,8	155.188	2.625.241	94,1
MA	46.942	726.128	93,5	21.304	550.234	96,1	68.246	1.276.362	94,7
PB	78.813	992.530	92,1	34.201	761.417	95,5	113.014	1.753.947	93,6
PE	134.694	1.633.968	91,8	62.801	1.224.480	94,9	197.495	2.858.448	93,1
PI	36.435	573.150	93,6	14.802	449.425	96,7	51.237	1.022.575	95,0
RN	80.784	850.083	90,5	39.563	665.329	94,1	120.347	1.515.412	92,1
SE	26.750	351.269	92,4	11.803	283.007	95,8	38.553	634.276	93,9
TO	23.141	323.852	92,9	10.614	262.898	96,0	33.755	586.750	94,3
Norte	192.351	2.999.654	93,6	103.000	2.452.226	95,8	295.351	5.451.880	94,6
AC	16.769	179.051	90,6	9.625	146.612	93,4	26.394	325.663	91,9
AM	53.990	765.397	92,9	27.185	591.728	95,4	81.175	1.357.125	94,0
AP	3.755	167.419	97,8	2.171	137.439	98,4	5.926	304.858	98,1
PA	44.835	797.798	94,4	25.608	676.803	96,2	70.443	1.474.601	95,2
RO	37.809	525.978	92,8	20.830	445.250	95,3	58.639	971.228	94,0
RR	12.052	240.159	95,0	6.967	191.496	96,4	19.019	431.655	95,6
Sudeste	1.314.402	21.012.604	93,7	688.325	18.031.377	96,2	2.002.727	39.043.981	94,9
ES	1.211	35.226	96,6	93.286	1.012.384	90,8	94.497	1.047.610	91,0
MG	298.476	4.423.407	93,3	135.854	3.774.568	96,4	434.330	8.197.975	94,7
RJ	214.936	3.768.656	94,3	105.819	2.922.610	96,4	320.755	6.691.266	95,2
SP	799.779	12.785.315	93,7	353.366	10.321.815	96,6	1.153.145	23.107.130	95,0
Sul	418.139	6.254.208	93,3	187.293	8.611.226	97,8	605.432	14.865.434	95,9
PR	6.708	115.477	94,2	3.209	126.421	97,5	9.917	241.898	95,9
RS	292.241	3.759.828	92,2	130.283	3.489.356	96,3	422.524	7.249.184	94,2
SC	119.190	2.378.903	95,0	53.801	4.995.449	98,9	172.991	7.374.352	97,7
<b>Brasil</b>	<b>2.911.278</b>	<b>43.205.453</b>	<b>93,3</b>	<b>1.433.680</b>	<b>39.355.841</b>	<b>96,4</b>	<b>4.344.958</b>	<b>82.561.294</b>	<b>94,7</b>

\* Fonte: eSUS VE Notifica.

UF: unidade da federação, CBO: Código Brasileiro de Ocupação, AC: Acre, AL: Alagoas, AM: Amazonas, AP: Amapá, BA: Bahia, CE: Ceará, DF: Distrito Federal, ES: Espírito Santo, GO: Goiás, MA: Maranhão, MG: Minas Gerais, MS: Mato Grosso do Sul, MT: Mato Grosso, PA: Pará, PB: Paraíba, PE: Pernambuco, PI: Piauí, PR: Paraná, RJ: Rio de Janeiro, RN: Rio Grande do Norte, RO: Rondônia, RR: Roraima, RS: Rio Grande do Sul, SC: Santa Catarina, SE: Sergipe, SP: São Paulo, TO: Tocantins.

**Tabela 2** Percentual de incompletude da variável profissão ou ocupação nos bancos de dados de síndrome respiratória aguda grave\*, segundo sexo, UF, macrorregiões, Brasil, 2020-2021

<i>Síndrome respiratória aguda grave</i>									
<i>Região/UF</i>	<i>Feminino</i>			<i>Masculino</i>			<i>Total</i>		
	<i>CBO preenchido</i>	<i>Total de casos</i>	<i>Incompletude (%)</i>	<i>CBO preenchido</i>	<i>Total de casos</i>	<i>Incompletude (%)</i>	<i>CBO preenchido</i>	<i>Total de casos</i>	<i>Incompletude (%)</i>
Centro-Oeste	2.683	117.153	97,7	4.421	142.293	96,9	7.104	259.446	97,2
DF	215	27.824	99,2	333	33.713	99,0	548	61.537	99,1
GO	1.139	42.544	97,3	2.241	52.494	95,7	3.380	95.038	96,4
MS	147	22.670	99,4	252	27.114	99,1	399	49.784	99,2
MT	1.182	24.115	95,1	1.595	28.972	94,5	10.209	53.087	94,8
Nordeste	4.863	249.640	98,1	7.834	294.284	97,3	12.697	543.924	97,6
AL	124	15.498	99,2	175	18.116	99,0	299	33.614	99,1
BA	882	49.261	98,2	1.560	59.431	97,4	2.442	108.692	97,8
CE	1.345	54.167	97,5	1.918	64.684	97,0	3.263	118.851	97,3
MA	560	15.599	96,4	859	20.057	95,7	1.419	35.656	96,0
PB	401	21.451	98,1	758	24.127	96,9	1.159	45.578	97,5
PE	682	54.510	98,7	913	58.948	98,5	704	113.458	98,6
PI	267	13.998	98,1	437	17.636	97,5	3.066	31.634	97,8
RN	374	12.527	97,0	780	15.516	95,0	1.154	28.043	95,9
SE	228	12.629	98,2	434	15.769	97,2	17.544	28.398	97,7
TO	434	8.978	95,2	841	12.101	93,1	1.275	21.079	94,0
Norte	3.488	83.537	95,8	5.806	104.662	94,5	9.294	188.199	96,0
AC	106	3.191	96,7	156	3.842	95,9	262	7.033	96,3
AM	795	24.218	96,7	1.174	28.354	95,9	1.434	52.572	96,3
AP	571	3.244	82,4	863	4.007	78,5	1.969	7.251	80,2
PA	940	31.843	97,0	1.414	40.325	96,5	2.354	72.168	96,7
RO	256	9.651	97,3	547	12.885	95,8	1.197	22.536	96,4
RR	386	2.412	84,0	811	3.148	74,2	994	5.560	78,5
Sudeste	11.753	672.067	98,3	20.052	789.146	97,5	31.805	1.461.213	97,8
ES	216	11.479	98,1	373	13.424	97,2	589	24.903	97,6
MG	3.905	150.818	97,4	6.304	170.514	96,3	2.777	321.332	96,8
RJ	1.151	123.527	99,1	2.312	139.636	98,3	3.463	263.163	98,7
SP	6.481	386.243	98,3	11.063	465.572	97,6	662	851.815	97,9
Sul	2.636	211.881	98,8	4.645	256.145	98,2	7.281	468.026	98,4
PR	1.093	93.685	98,8	1.973	114.207	98,3	1.595	207.892	98,5
RS	367	72.596	99,5	627	84.708	99,3	803	157.304	99,4
SC	1.176	45.600	97,4	2.045	57.230	96,4	3.221	102.830	96,9
<b>Brasil</b>	<b>25.423</b>	<b>1.334.278</b>	<b>98,1</b>	<b>42.758</b>	<b>1.586.530</b>	<b>97,3</b>	<b>68.181</b>	<b>2.920.808</b>	<b>97,7</b>

\*Fonte: Sivep-Gripe.

UF: unidade da federação, CBO: Código Brasileiro de Ocupação, AC: Acre, AL: Alagoas, AM: Amazonas, AP: Amapá, BA: Bahia, CE: Ceará, DF: Distrito Federal, ES: Espírito Santo, GO: Goiás, MA: Maranhão, MG: Minas Gerais, MS: Mato Grosso do Sul, MT: Mato Grosso, PA: Pará, PB: Paraíba, PE: Pernambuco, PI: Piauí, PR: Paraná, RJ: Rio de Janeiro, RN: Rio Grande do Norte, RO: Rondônia, RR: Roraima, RS: Rio Grande do Sul, SC: Santa Catarina, SE: Sergipe, SP: São Paulo, TO: Tocantins.

**Tabela 3** Percentual de incompletude da variável profissão ou ocupação nos bancos de dados de mortalidade\* por todas as causas, segundo sexo, UF, macrorregiões, Brasil, 2020-2021

Óbitos por todas as causas									
Região/UF	Feminino			Masculino			Total		
	CBO preenchido	Total de casos	Incompletude (%)	CBO preenchido	Total de casos	Incompletude (%)	CBO preenchido	Total de casos	Incompletude (%)
Centro-Oeste	79.301	90.806	12,7	109.223	128.003	14,7	188.524	218.809	13,8
DF	12.358	13.872	10,9	15.248	17.762	14,2	27.606	31.634	12,7
GO	35.065	41.342	15,2	48.283	57.720	16,3	83.348	99.062	15,9
MS	15.400	16.967	9,2	20.533	23.162	11,4	35.933	40.129	10,5
MT	16.478	18.625	11,5	25.159	29.359	14,3	41.637	47.984	13,2
Nordeste	270.745	336.063	19,4	338.116	437.632	22,7	608.861	773.695	21,3
AL	10.476	19.729	46,9	14.030	24.838	43,5	24.506	44.567	45,0
BA	67.662	84.645	20,1	87.061	114.899	24,2	154.723	199.544	22,5
CE	50.347	56.338	10,6	62.155	72.890	14,7	112.502	129.228	12,9
MA	26.489	32.287	18,0	37.755	47.167	20,0	64.244	79.454	19,1
PB	16.324	26.623	38,7	17.283	32.193	46,3	33.607	58.816	42,9
PE	57.090	64.433	11,4	66.667	77.452	13,9	123.757	141.885	12,8
PI	15.730	19.116	17,7	20.237	25.693	21,2	35.967	44.809	19,7
RN	15.582	20.255	23,1	18.155	25.526	28,9	33.737	45.781	26,3
SE	11.045	12.637	12,6	14.773	16.974	13,0	25.818	29.611	12,8
TO	6.200	7.345	15,6	9.824	11.648	15,7	16.024	18.993	15,6
Norte	68.249	84.261	19,0	99.221	127.391	22,1	167.470	211.652	20,9
AC	2.438	3.751	35,0	3.167	5.700	44,4	5.605	9.451	40,7
AM	17.339	20.100	13,7	25.380	29.538	14,1	42.719	49.638	13,9
AP	2.280	3.355	32,0	3.525	5.364	34,3	5.805	8.719	33,4
PA	30.751	38.366	19,8	43.275	57.489	24,7	74.026	95.855	22,8
RO	7.002	8.521	17,8	10.631	13.293	20,0	17.633	21.814	19,2
RR	2.239	2.823	20,7	3.419	4.359	21,6	5.658	7.182	21,2
Sudeste	550.555	630.309	12,7	618.034	750.388	17,6	1.168.589	1.380.697	15,4
ES	14.359	23.361	38,5	17.874	30.408	41,2	32.233	53.769	40,1
MG	122.394	136.599	10,4	146.858	167.559	12,4	269.252	304.158	11,5
RJ	133.584	154.510	13,5	125.180	169.889	26,3	258.764	324.399	20,2
SP	280.218	315.839	11,3	328.122	382.532	14,2	608.340	698.371	12,9
Sul	175.673	203.602	13,7	215.131	254.211	15,4	390.804	457.813	14,6
PR	70.711	75.874	6,8	93.107	100.933	7,8	163.818	176.807	7,3
RS	68.982	86.391	20,2	76.305	99.983	23,7	145.287	186.374	22,0
SC	35.980	41.337	13,0	45.719	53.295	14,2	81.699	94.632	13,7
<b>Brasil</b>	<b>1.144.523</b>	<b>1.345.041</b>	<b>14,9</b>	<b>1.379.725</b>	<b>1.697.625</b>	<b>18,7</b>	<b>2.524.248</b>	<b>3.042.666</b>	<b>17,0</b>

\* Fonte: SIM.

UF: unidade da federação, CBO: Código Brasileiro de Ocupação, AC: Acre, AL: Alagoas, AM: Amazonas, AP: Amapá, BA: Bahia, CE: Ceará, DF: Distrito Federal, ES: Espírito Santo, GO: Goiás, MA: Maranhão, MG: Minas Gerais, MS: Mato Grosso do Sul, MT: Mato Grosso, PA: Pará, PB: Paraíba, PE: Pernambuco, PI: Piauí, PR: Paraná, RJ: Rio de Janeiro, RN: Rio Grande do Norte, RO: Rondônia, RR: Roraima, RS: Rio Grande do Sul, SC: Santa Catarina, SE: Sergipe, SP: São Paulo, TO: Tocantins.

## Discussão

Neste estudo, foram observados grandes percentuais de incompletude da variável profissão/ocupação em todas as bases de dados analisadas, destacando-se aqueles dos bancos de SG e SRAG. A elevada incompletude foi observada principalmente nos registros de casos, em indivíduos de ambos os sexos e de forma generalizada nas UF e nas grandes regiões brasileiras.

Apesar da existência de diversos métodos para lidar com informações faltantes em bancos de dados (e.g., regressão linear, *multiple imputation*, *bootstrap* e *bayesian bootstrap*), os elevados percentuais de incompletude em todos os estratos impossibilitam a investigação de quaisquer relações entre a ocupação e o risco à saúde<sup>24-28</sup>.

Dados precisos e confiáveis sobre a saúde do trabalhador e possíveis doenças ocupacionais são considerados o ponto de partida essencial para o desenvolvimento de programas de prevenção<sup>29</sup>. Esses dados nos informam sobre quem ficou doente e como, quando e onde o agravo ocorreu. Portanto, eles têm o potencial de identificar alvos para medidas preventivas e de controle. Adicionalmente, tais informações exibem potencial considerável como ferramentas na avaliação do progresso no cumprimento das metas de prevenção.

Comparando-se as bases de dados estudadas, o SIM, por se basear no registro de óbito e ser a principal fonte de dados sobre mortalidade no Brasil, apresenta menores percentuais de incompletude na variável ocupação. O SIM é o sistema de informação mais antigo do Brasil, havendo maior conhecimento dos profissionais sobre a necessidade de seu adequado preenchimento, além de existir campos específicos para informar caso o óbito seja resultado de uma atividade laboral (ex: acidente de trabalho). Ainda, os óbitos são considerados desfechos duros, ou definitivos (do inglês *hard endpoint*), ou seja, são eventos bem definidos. Dessa forma, naturalmente, espera-se maior ocorrência de preenchimento dos dados individuais na declaração de óbito<sup>30</sup>.

Especificamente no cenário da pandemia de covid-19, o Brasil apresentou uma das menores taxas de testagem do mundo<sup>31-33</sup>. Alguns trabalhos utilizaram dados de taxa de mortalidade geral como forma de quantificar a carga da doença em nível nacional, com uso de dados de séries históricas de mortalidade e cálculo do excesso de mortalidade, porém tais métodos se tornam impraticáveis ao se desagregar por categorias profissionais<sup>34,35</sup>.

Como previamente exposto, para a quantidade de óbitos é possível obter dados precisos sobre o número de casos, ou seja, o numerador da taxa de mortalidade.

No entanto, a interpretação desses dados costuma ser difícil na ausência de dados confiáveis sobre o número de trabalhadores sob risco, quantitativo essencial para compor seu denominador. Normalmente, utilizam-se dados de censo ou pesquisas sobre a força de trabalho para fornecer estimativas do número de trabalhadores em grupos ocupacionais, industriais e sociodemográficos. Contudo, deve-se atentar que os dados advindos do censo, por exemplo, são frequentemente desatualizados e inadequados para alguns grupos ocupacionais, como trabalhadores sazonais, trabalhadores informais ou temporários, assim como imigrantes. De tal forma, os indicadores calculados podem não ser confiáveis<sup>36-38</sup>.

Em algumas situações, dados de várias fontes podem ser combinados para desenvolver denominadores mais apropriados para uma análise particular<sup>39</sup>. No entanto, deve-se reconhecer que, embora seja possível obter dados precisos para alguns resultados, as limitações nos dados do denominador disponíveis podem introduzir vies e erros aleatórios no cálculo das taxas.

Embora as fontes de dados utilizadas forneçam informações valiosas, compreendem apenas um subconjunto de casos. O relacionamento entre base de dados (em inglês, *linkage*) fornece um meio de combinar dados de duas ou mais fontes e, assim, permitir o aproveitamento de mais informações. Ao vincular os registros de um mesmo indivíduo em diferentes bases de dados, as informações dessa pessoa podem ser reunidas em um único arquivo. Isso permitiria que as pessoas representadas em vários conjuntos de dados fossem identificadas, assim como forneceria informações mais abrangentes sobre esses casos. Essa melhoria já foi observada para doenças como a tuberculose<sup>39</sup>. Tal abordagem fornece um mecanismo para que os dados disponíveis sejam mais eficientes. Na prática, no entanto, as oportunidades de vincular bases de dados são severamente limitadas pela falta de um identificador pessoal exclusivo nos arquivos de dados, além de fragilidades em relação à privacidade dos usuários.

A qualidade desses e de outros sistemas de informação nacionais também poderia ser aprimorada, a partir da contínua avaliação. Analogamente, é necessário agir para a capacitação e conscientização dos profissionais de saúde a respeito da importância do adequado preenchimento das informações das fichas de notificação, a fim de melhorar o cuidado aos indivíduos, bem como aprimorar as ações de vigilância em saúde e ambiente.

## Conclusão

Em todas as bases de dados analisadas, verificou-se elevado percentual de incompletude da variável profissão/ocupação no Brasil, em 2020 e 2021. Os achados

corroboram o processo de invisibilização dos trabalhadores mais impactados ao longo do período de estudo, durante a pandemia de covid-19. No contexto pandêmico, quando foram recomendadas medidas de distanciamento social com restrições de atividades produtivas, a ausência de informações sobre a saúde dos trabalhadores e o impacto da doença em grupos específicos, como profissionais de saúde e da educação, dificultaram as orientações de políticas públicas, tanto para restrição quanto para retorno das atividades.

A falta de preenchimento de variáveis nas declarações/fichas de notificação afeta a qualidade da

informação necessária para subsidiar a tomada de decisão em todos os níveis de gestão do sistema de saúde. A melhoria na qualidade das bases de dados quanto à completude pode impactar positivamente a saúde pública e as políticas públicas.

Por fim, ressaltamos que é essencial a união de esforços por parte dos diversos atores envolvidos – incluindo Ministério do Trabalho e Previdência, Ministério da Saúde, Ministério da Educação e grupos de trabalhadores, como sindicatos – para articular, em conjunto, a formulação de soluções para o problema da falta de dados no campo da saúde do trabalhador.

## Referências

1. Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc.* 1996;91(434):473-89.
2. Van Der Heijden GJMG, Donders ART, Stijnen T, Moons KGM. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol.* 2006;59(10):1102-9.
3. Batista AG, Santana VS, Ferrite S. Registro de dados sobre acidentes de trabalho fatais em sistemas de informação no Brasil. *Cienc Saude Colet.* 2019;24(3):693-704.
4. Marinho F, Araújo VEM, Porto DL, Ferreira HL, Coelho MRS, Lecca RCR, et al. Microcefalia no Brasil: prevalência e caracterização dos casos a partir do Sistema de Informações sobre Nascidos Vivos (Sinasc), 2000-2015. *Epidemiol Serv Saude.* 2016;25:701-12.
5. Xu B, Kraemer MUG, Xu B, Gutierrez B, Mekaru S, Sewalk K, et al. Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect Dis.* 2020;20(5):534.
6. Demenech LM, Dumith SC, Vieira MECD, Neiva-Silva L. Income inequality and risk of infection and death by COVID-19 in Brazil. *Rev Bras Epidemiol.* 2020;23:e200095.
7. Maeno M. Covid-19 como uma doença relacionada ao trabalho. *Rev Bras Saude Ocup.* 2021;46:e54.
8. Santos CVB, Valiati NCM, Noronha TG, Porto VBG, Pacheco AG, Freitas LP, et al. The effectiveness of COVID-19 vaccines against severe cases and deaths in Brazil from 2021 to 2022: a registry-based study. *Lancet Reg Health Am.* 2023;20:100465.
9. Santos CVB, Noronha TG, Werneck GL, Struchiner CJ, Villela DAM. Estimated COVID-19 severe cases and deaths averted in the first year of the vaccination campaign in Brazil: A retrospective observational study. *Lancet Reg Health Am.* 2023;17:100418.
10. Coelho LE, Luz PM, Pires DC, Jalil EM, Perazzo H, Torres TS, et al. Prevalence and predictors of anti-SARS-CoV-2 serology in a highly vulnerable population of Rio de Janeiro: a population-based serosurvey. *Lancet Reg Health Am.* 2022;15:100338.
11. Santos CVBD, Cavalcante JR, Pungartnik PC, Guimarães RM. Transição da idade de casos, internações e óbitos em internações por covid-19 no município do Rio de Janeiro. *Rev Bras Estud Popul.* 2022;39:e0195.
12. Santos CVBD, Cavalcante JR, Pungartnik PC, Guimarães RM. Análise espaço temporal do primeiro ano da pandemia de COVID-19 no município do Rio de Janeiro. *Rev Bras Epidemiol.* 2021;24:e210046.
13. Cavalcante JR, Xavier DR, Santos CVBD, Pungartnik PC, Guimarães RM. Análise espacial do fluxo origem-destino das internações por síndrome respiratória aguda grave por covid-19 na região metropolitana do Rio de Janeiro. *Rev Bras Epidemiol.* 2021;24:e210054.
14. Fundação Oswaldo Cruz. Recomendações para o planejamento de retorno às atividades escolares presenciais no contexto da pandemia de covid-19: ano 2021. Rio de Janeiro: Fiocruz; 2021.
15. Medeiros LCA, Borges MCAL, Gonsalves MP, Teodózio GC, Santos WM. Perfil sócio demográfico dos pacientes acometidos pela covid-19. *Environmental Smoke.* 2021;4(2):42-8.
16. Gomez CM, Lacaz FAC. Saúde do trabalhador: novas-velhas questões. *Cienc Saude Colet.* 2005;10(4):797-807.
17. Brasil. Ministério da Saúde. SRAG 2020 – Banco de Dados de Síndrome Respiratória Aguda Grave – incluindo dados da covid-19. Open DataSUS [Internet]. 2022 [citado em 14 mar 2022]. Disponível em: <https://opendatasus.saude.gov.br/dataset/srag-2020>
18. Brasil. Ministério da Saúde. SRAG 2021 – Banco de Dados de Síndrome Respiratória Aguda Grave –



- incluindo dados da covid-19. Open DataSUS [Internet]. 2022 [citado em 16 nov 2023]. Disponível em: <https://opendatasus.saude.gov.br/dataset/srag-2021-a-2023>
19. Brasil. Ministério da Saúde. Sistema de Informação sobre Mortalidade (SIM) 2020- Dados Abertos – Centrais de Conteúdos – DASNT. SVS/MS [Internet]. 2022 [citado em 14 mar 2022]. Disponível em: <https://opendatasus.saude.gov.br/dataset/sim>
  20. Brasil. Ministério da Saúde. Sistema de Informação sobre Mortalidade (SIM) 2021- Dados Abertos – Centrais de Conteúdos – DASNT. SVS/MS [Internet]. 2022 [citado em 14 mar 2022]. Disponível em: <https://opendatasus.saude.gov.br/dataset/sim>
  21. Brasil. Ministério do Trabalho. CBO – Classificação Brasileira de Ocupações. CBO [Internet]. [citado em 16 nov 2023]. Disponível em: <http://www.mtecho.gov.br/cbsite/pages/home.jsf>
  22. R Core Team. R: A Language and Environment for Statistical R: version 4.1.2 (2021-11-01) -- “Bird Hippie” [Internet]. Vienna: R Foundation for Statistical Computing; 2021 [citado em 22 mar 2022]. Disponível em: <https://cran.r-project.org/src/base/R-4/>
  23. Santos CVB. Repositório Missing\_data\_br. GitHub [Internet]. 2022 [citado em 21 mar 2022]. Disponível em: [https://github.com/epicleber/missing\\_data\\_br](https://github.com/epicleber/missing_data_br)
  24. Santin G, Geoffroy B, Bénézet L, Delézire P, Chatelot J, Sitta R, et al. In an occupational health surveillance study, auxiliary data from administrative health and occupational databases effectively corrected for nonresponse. *J Clin Epidemiol*. 2014;67(6):722-30.
  25. Bang H, Robins JM. Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics*. 2005;61(4):962-73.
  26. Baumgarten M, Oseasohn R. Studies on occupational health: a critique. *J Occup Med*. março de 1980;22(3):171-6.
  27. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2013;22(3):278-95.
  28. 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*. 2021;11(12):e047623.
  29. Feyer A-M, Williamson AM. The role of work practices in occupational accidents. *1991;35(15):1100-4*.
  30. Jorge MHPM, Laurenti R, Gotlieb SLD. Análise da qualidade das estatísticas vitais brasileiras: a experiência de implantação do SIM e do Sinasc. *Cienc Saude Colet*. 2007;12(3):643-54.
  31. Mathieu E, Ritchie H, Rodés-Guirao L, Appel C, Gavrilo D, Giattino C, et al. Coronavirus pandemic (COVID-19). *Our World in Data* [Internet]. 5 mar 2020 [citado em 12 nov 2021]; Disponível em: <https://ourworldindata.org/coronavirus>
  32. Covid-19 no Brasil | Painéis | Twitter: @CoronavirusBra1 [Internet]. [date unknown] [citado em 23 jan 2023]. Disponível em: <https://coronavirusbra1.github.io/>
  33. Silva GA, Jardim BC, Santos CVB. Excesso de mortalidade no Brasil em tempos de covid-19. *Cienc Saude Colet*. 2020;25(9):3345-54.
  34. Orellana JDY, Cunha GM, Marrero L, Moreira RI, Leite IC, Horta BL. Excesso de mortes durante a pandemia de covid-19: subnotificação e desigualdades regionais no Brasil. *Cad Saude Publica*. 2021;37(1):e00259120.
  35. Macaskill P, Driscoll TR. National occupational injury statistics: what can the data tell us? Boca Raton: CRC Press; 1998.
  36. Macaskill P, Mandryk JA, Leigh J. Improved coverage and relevance of national occupational health and safety surveillance in Australia. *Epidemiology and Surveillance Unit. Worksafe Australia*; 1995.
  37. Corvalan CF, Driscoll TR, Harrison JE. Role of migrant factors in work-related fatalities in Australia. *Scand J Work Environ Health*. 1994;20(5):364-70.
  38. Driscoll TR, Ansari G, Harrison JE, Frommer MS, Ruck EA. Traumatic work related fatalities in commercial fishermen in Australia. *Occup Environ Med*. 1994;51(9):612-6.
  39. Bartholomay P, Oliveira GP, Pinheiro RS, Vasconcelos AMN. Melhoria da qualidade das informações sobre tuberculose a partir do relacionamento entre bases de dados. *Cad Saude Publica*. 2014;30(11):2459-70.

## **Contribuição de autoria**

Santos CVB, Melo-Ferreira V, Sampaio JRC, Ribeiro PC, Castro HA, Gutierrez AC, Pereira IDF, Nehab MF, Moura MMD e Perissé ARS contribuíram nas diversas fases da investigação, análise dos dados, redação, revisão e aprovação do texto para publicação. Todos assumem responsabilidade integral pelo estudo e conteúdo publicado.

## **Disponibilidade de dados**

Todo o conjunto de dados anonimizados que dá suporte aos resultados deste estudo está disponível no repositório OPENDATASUS: <https://opendatasus.saude.gov.br/dataset>. Acessado em 14/03/2022.

Recebido: 12/07/2022

Revisado: 02/10/2023

Aprovado: 12/10/2023

*Editor-chefe:*

Eduardo Algranti