



II Semana da Demografia

O QUE SE PERDE E O QUE SE TEM: ANÁLISE DA QUALIDADE DO PREENCHIMENTO DOS DADOS DE ÓBITOS E NASCIMENTOS BRASILEIROS (1994 – 2021)¹

João Gabriel Malaguti²

Mirna Tetzner Ramos³

RESUMO

Os dados de mortalidade e natalidade são fundamentais para os estudos demográficos. Por esta razão é fundamental que se atente para a qualidade das fontes dos mesmos. Nessa perspectiva, este estudo se propõe a inventariar os valores suspeitos presentes no Sistema de Informações sobre Mortalidade (SIM) e no Sistema de Informações sobre Nascidos Vivos (SINASC) a partir dos valores faltantes das diferentes características disponibilizadas pelos sistemas. Foi realizada uma análise descritiva e visual para identificação de valores extremos e destoantes da série histórica (1994-2021 para dados do SINASC, 1996-2021, para dados do SIM) e, posteriormente, proposta e analisada uma métrica da qualidade de preenchimento considerando e desconsiderando os valores suspeitos. Foram encontrados valores suspeitos nos dados de nascimento para os estados de São Paulo (nos anos de 1994 a 1996) e Sergipe (em 1994), assim como os dados para instrução materna no ano de 1999 para vários estados, além de outros pontos específicos. Finalmente, pode-se identificar primeiramente que há uma tendência de queda generalizada dos dados faltantes no período para ambos os registros e, em segundo lugar, que a pandemia de COVID-19 não teve impacto significativo no preenchimento dos documentos.

Palavras-Chaves: Não resposta; Registros administrativos; Sistema de Informações sobre Mortalidade; Sistema de Informações sobre Nascidos Vivos.

INTRODUÇÃO

As estatísticas vitais de óbitos e nascidos vivos são objeto de profunda análise e interesse no campo da demografia. Dois dos três principais elementos utilizados para compreensão das dinâmicas de crescimento populacional são originadas dessas estatísticas: a mortalidade e a fecundidade. Além dessas dinâmicas, podem ser calculados outros indicadores caros à demografia social, como a taxa de mortalidade materna, taxa de mortalidade infantil, taxa de natalidade, entre outras taxas relacionadas a tipos de causa de mortalidade específica.

¹ Trabalho apresentado na II Semana da Demografia da Universidade Estadual de Campinas, evento que ocorreu entre os dias 22 e 26 de abril de 2024.

² Doutorando na Escola Nacional de Ciências Estatísticas.

³ Mestranda na Escola Nacional de Ciências Estatísticas.

Esses indicadores de saúde são essenciais para a vigilância das condições de vida da população, assim como do desempenho do sistema de saúde de um país. Porém, a qualidade de registros administrativos depende de muitos fatores que vão desde o acesso da população aos instrumentos de registro nas unidades notificadoras, do correto preenchimento dos documentos-base dos sistemas de informação ou mesmo de eficiência nos processos de coleta, processamento e consolidação de dados.

O SIM e o SINASC

Os sistemas de informação em saúde de interesse para este trabalho são o Sistema de Informações sobre Mortalidade (SIM) e o Sistema de Informações sobre Nascidos Vivos (SINASC), ambos indispensáveis para cálculos demográficos. Segundo a publicação da RIPSA “Indicadores básicos para a saúde no Brasil: conceitos e aplicações” (2008), o SIM é o sistema mais antigo de informação de saúde no Brasil. Instituído em 1975 pelo Ministério da Saúde (MS) e com consolidação de dados a nível nacional e informatização a partir de 1979 (Brasil, 2009a), o sistema amparou a posterior criação do SINASC, criado em 1990 e consolidado nacionalmente em 1994. A gestão nacional do SIM e do SINASC é competência da Secretaria de Vigilância em Saúde, com as Secretarias de Estado da Saúde e Secretarias Municipais de Saúde suas gestões estaduais e municipais, respectivamente (Brasil, 2009b).

O SIM tem como base de informação a Declaração de Óbito (DO), para registro de óbitos em geral e óbitos fetais, preenchido por médicos (causas naturais) ou peritos legistas (causas não naturais) (Brasil, 2022b). A DO foi padronizada nacionalmente em 1976, e sua última atualização foi em 2014. As DOs assumirão fluxos diferentes conforme a característica ou local de ocorrência do óbito. Os documentos serão distribuídos pelas Secretarias Municipais de Saúde entre os estabelecimentos de saúde, Institutos de Medicina Legal, Serviços de Verificação de Óbitos, entre os profissionais médicos, cartórios de registro civil e outras instituições que tenham permissão legal para sua utilização (Brasil, 2009b).

Quando falamos do SINASC, o documento de registro base do sistema é a Declaração de Nascido Vivo (DNV), padronizada nacionalmente desde 1990 pelo MS (Brasil, 2022a). O sistema possibilita o monitoramento dos nascidos vivos em território brasileiro, das características do pré-natal, da gestação e do parto. Estão autorizadas como unidades notificadoras para a DNV os estabelecimentos e Serviços de Saúde, onde possam ocorrer partos, médicos e enfermeiros, parteiras tradicionais reconhecidas e vinculadas a unidades de saúde e Cartórios de Registro Civil (Brasil, 2009b).

Em 1991 é criado o Departamento de Informática do Sistema Único de Saúde (DATASUS), que passa a possuir o encargo de fornecer soluções para controle e processamento de dados dos sistemas de informação em saúde. A criação do DATASUS possibilita uma transformação no nível de acesso e da facilidade para obtenção de dados de saúde, sendo possível atualmente acessar as informações por meio dos tabuladores como o TABNET, TABWIN e o TABDOS, via download de microdados e através de painéis de monitoramento.

Esta abertura dos dados do SIM e do SINASC permitiu a utilização das informações destes sistemas por toda sociedade. Ampliou-se o horizonte de usuários de dados sobre informações de saúde, criando um ambiente que promove a discussão e a produção de novas análises e perspectivas acerca dos indicadores. Porém, para que sejam produzidas análises corretas é preciso fomentar constantemente, tanto a abertura de novas informações pelo DATASUS, garantindo-se o anonimato dos usuários, como a qualidade das bases de dados.

Este trabalho apresenta dois objetivos principais: primeiro, fazer uma crítica aos dados existentes no TABNET e identificar possíveis pontos suspeitos a partir de uma análise descritiva visual; seguida de uma proposta de métrica de qualidade da completude dos registros, que também será analisada.

MÉTODOS

Optamos por utilizar a ferramenta de tabulações pela internet (TABNET) ao invés das opções para Windows e DOS (TABWIN e TABDOS, respectivamente) por ser o mais acessível a todos, não dependendo da instalação de qualquer programa para análise. Por esta ferramenta tem-se acesso às informações sobre nascidos vivos a partir de 1994 e sobre mortalidade a partir de 1979, embora haja uma quebra de série histórica devido à mudança entre as versões 9 e 10 da Classificação Internacional de Doenças (CID), dividindo o período em 1979 a 1995 e 1996 em diante. Para uma melhor comparabilidade entre a completude dos dois sistemas, preferiu-se utilizar apenas os dados do bloco da CID-10 para a mortalidade.

As variáveis que podem ser utilizadas para tabulações se encontram nos Quadros 1 e 2, incluindo também o ano a partir dos quais existem dados e quais são as categorias de interesse para este estudo. As variáveis que não apresentam valores na coluna de categorias de interesse não serão utilizadas para análise.

QUADRO 1 – Relação das variáveis presentes para tabulações sobre mortalidade (SIM)

VARIÁVEL	INÍCIO DA VIGÊNCIA	CATEGORIAS DE INTERESSE
Capítulo CID-10	1996	
Grupo CID-10	1996	
Categoria CID-10	1996	
Causa – CID-BR-10	1996	
Causa mal definidas	1996	Sint sinais e achad anorm, exc morte súbita infância
		Doenças do aparelho circulatório, mal definidas
		Doenças do aparelho respiratório, mal definidas
		Afecções original período perinatal, mal definidas
Faixa Etária	1996	Idade ignorada
Faixa Etária OPS	1996	
Faixa Etária det	1996	
Fx. Etária Menor 1a	1996	
Sexo	1996	Ign
Cor/raça	1997	Ignorado
Escolaridade	1996	Ignorado
Estado civil	1996	Ignorado
Local ocorrência	1996	Ignorado
Mês ocorrência	1996	Ignorado

Fonte: Datasus (Tabnet), ago/2023.

Para as variáveis relativas à mortalidade não serão utilizadas Faixa Etária OPS e Faixa Etária det por terem os mesmos valores de não resposta da variável Faixa Etária, enquanto a variável Fx. Etária Menor 1ª não será contemplada por ser caso específico que depende da resposta de Faixa etária. As variáveis relativas à CID-10 são encaixadas, isto é, dependem da variável um nível abaixo, por exemplo, a Categoria é que define o Grupo e assim por diante. Neste caso, optou-se por utilizar apenas a variável mais “interior”, Causas mal definidas.

QUADRO 2 – Relação das variáveis presentes para tabulações sobre nascidos vivos (SINASC)

VARIÁVEL	INÍCIO DA VIGÊNCIA	CATEGORIAS DE INTERESSE
Local ocorrência	1994	Ignorado
Mês ocorrência	1994	Ignorado
Idade da mãe	1994	Idade ignorada
Instrução da mãe	1994	Ignorado
Estado civil mãe	2000	Ignorado
Duração gestação	1994	Ignorado
Tipo de gravidez	1994	Ignorada
Grupos de Robson	2014	
Adeq quant pré-natal	2014	Ignorado
Tipo de parto	1994	Não Classificados Não informado
Consult pré-natal	1995	Ignorado
Sexo	1994	Ign
Cor/raça	1997	Ignorado
Apgar 1º minuto	1994	Ignorado
Apgar 5º minuto	1994	Ignorado
Peso ao nascer	1994	Ignorado
Anomalia congênita	1999	Ignorado
Tipo de anomalia congênita	1999	

Fonte: Datasus (Tabnet), ago/2023.

Tipo de anomalia congênita depende de o nascido vivo ter a anomalia congênita e sua categoria de dado ignorado infelizmente também inclui aqueles que não possuem anomalia, impossibilitando seu uso. Grupos de Robson, por sua vez, não apresentam nenhum nascimento classificado como ignorado em todo o território nacional desde sua criação.

Para uma análise que não seja dominada pelos efeitos da quantidade de nascimentos e óbitos nas unidades da federação que disponham de maior população, em particular São Paulo e Minas Gerais, O número de casos em cada ano, unidade da federação e categoria de interesse será ponderada pela quantidade de todos os casos de mesma configuração, ou seja, trabalharemos com as proporções.

Estas proporções serão submetidas a uma análise descritiva visual para a detecção de valores suspeitos, primeiro a partir de diagramas de caixa com dados de todas as unidades da federação por ano, respeitando o período nas quais as variáveis estão presentes nos dados. Após esta etapa, uma análise mais detalhada dos casos suspeitos será feita pela confecção de gráficos de dispersão das variáveis e unidades da federação específicas. Caso valores suspeitos sejam detectados, estes serão marcados e análises da métrica de qualidade serão produzidas, uma considerando estes valores e outras nas quais estes são excluídos.

A métrica de qualidade utilizada para avaliar os diferentes estados neste trabalho é uma composição de médias das proporções de grupos de variáveis. Isto é, as variáveis foram divididas em grupos temáticos (3 para os dados de mortalidade e 4 para os de nascimentos) e para cada um destes grupos foi calculada a média das proporções de dados faltantes. Calcula-se a média destas médias dos grupos para obter-se a métrica de comparação.

Os grupos temáticos para os óbitos foram ocorrência (inclui as variáveis local de ocorrência e mês de ocorrência), *causa mortis* (causas mal definidas) e características do falecido (faixa etária, sexo, cor/raça, escolaridade e estado civil). Os grupos temáticos para os nascimentos foram ocorrência (inclui as variáveis local de ocorrência e mês de ocorrência), características da mãe (idade da mãe, instrução da mãe e estado civil da mãe), informações sobre gestação e parto (duração da gestação, tipo de gravidez, adequação quanto ao pré-natal, tipo de parto e consultas pré-natal) e características do nascido vivo (sexo, raça/cor, Apgar 1º minuto, Apgar 5º minuto, peso ao nascer e anomalia congênita).

A interpretação da métrica é simples: quanto mais próxima de zero, mais completos estão os registros, com o valor zero representando preenchimento perfeito, isto é, de todos os campos. Quanto mais distante de zero, maior o uso das categorias “ignorado” e afins.

RESULTADO E DISCUSSÃO

Utilizando o *software* estatístico R (R CORE TEAM, 2023), computamos as proporções e preparamos os gráficos como descritos na seção anterior. A Figura 1 apresenta os diagramas de caixa para as variáveis relacionadas à mortalidade que apresentam algum valor suspeito (em vermelho), a saber, local de ocorrência em 1999, sexo em 1998 e faixa etária em 1996.

Procurando os valores extremos destas variáveis nestes anos específicos encontram-se os estados responsáveis pelos pontos suspeitos: Paraná, Goiás e Sergipe, respectivamente. Criando os gráficos de pontos para estas combinações (Figura 2), podemos notar que, enquanto que os valores para Paraná e Goiás realmente são suspeitos, os de Sergipe parecem seguir um padrão de queda, apesar do primeiro valor ser mais alto.

Os dados relativos aos nascimentos, no entanto, além de apresentarem questões pontuais como os apresentados anteriormente, apresentam também dois outros problemas: um para uma variável em um ano específico, mas para vários estados e outro para estados em períodos específicos, mas para várias variáveis. Além destes fatores, Anomalia congênita apresenta valores iguais a zero para todas as UFs em 1999 e 2000.

Como podemos notar na Figura 3, os valores para a instrução materna em 1999 destoam dos outros anos, tanto os anteriores quanto os posteriores. Ao analisar os gráficos de dispersão das 27 UFs (alguns dos quais podem ser vistos na Figura 4), encontram-se picos inexplicáveis para 16 UFs: Roraima, Pará, Amapá, Tocantins, Ceará, Paraíba, Pernambuco, Sergipe, Bahia, Minas Gerais, Paraná, Rio Grande do Sul, Mato Grosso do Sul, Mato Grosso, Goiás e Distrito Federal. Além destes casos, a variável também apresenta saltos similares para os estados de Alagoas em 2018 e Rondônia em 2017.

Já a questão de valores suspeitos para várias variáveis para estados em períodos específicos apresenta dois casos: Sergipe em 1994 e São Paulo em 1994 a 1996. Sergipe em 1994 apresenta valores suspeitos para todas as variáveis existentes na data, com exceção de Mês de ocorrência. Estes valores suspeitos se comportam como os outros valores já mencionados, picos sem explicação.

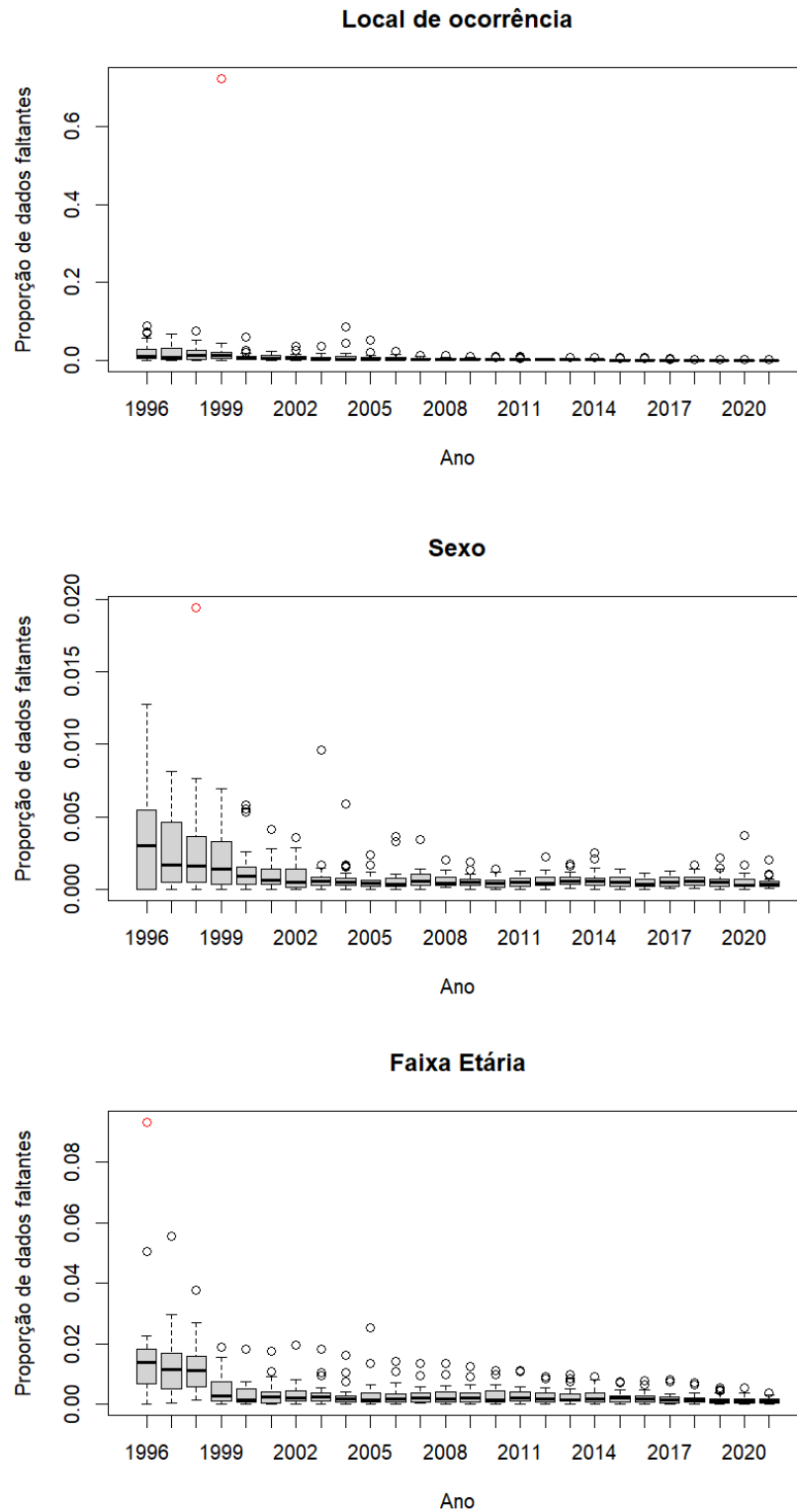
Os valores para São Paulo (Figura 5) são mais interessantes: os pontos para os 3 anos do período possuem o mesmo comportamento para as variáveis Mês da ocorrência, Local da ocorrência, Instrução da mãe, Duração da gestação, Tipo da gravidez, Apgar 1º minuto e Apgar 5º minuto. As variáveis Tipo do parto e Peso ao nascer apresentam picos em 1994, mas não o padrão mencionado, e a variável Sexo picos em 1995 e 1996.

Foram também identificados os seguintes casos pontuais: Roraima para Local da ocorrência em 1994 e para Apgar 1º minuto e Apgar 5º minuto em 1999, Rondônia para Idade da mãe em 94 e Bahia para Sexo em 1996.

Tendo realizado este levantamento de pontos suspeitos, podemos agora calcular as métricas, tanto considerando estes pontos quanto ignorando-os e analisar seus comportamentos ao longo do período. Como para os dados de mortalidade a diferença foi muito sutil (apenas para Paraná e Goiás em anos específicos), será apresentado apenas o gráfico para os dados corrigidos.

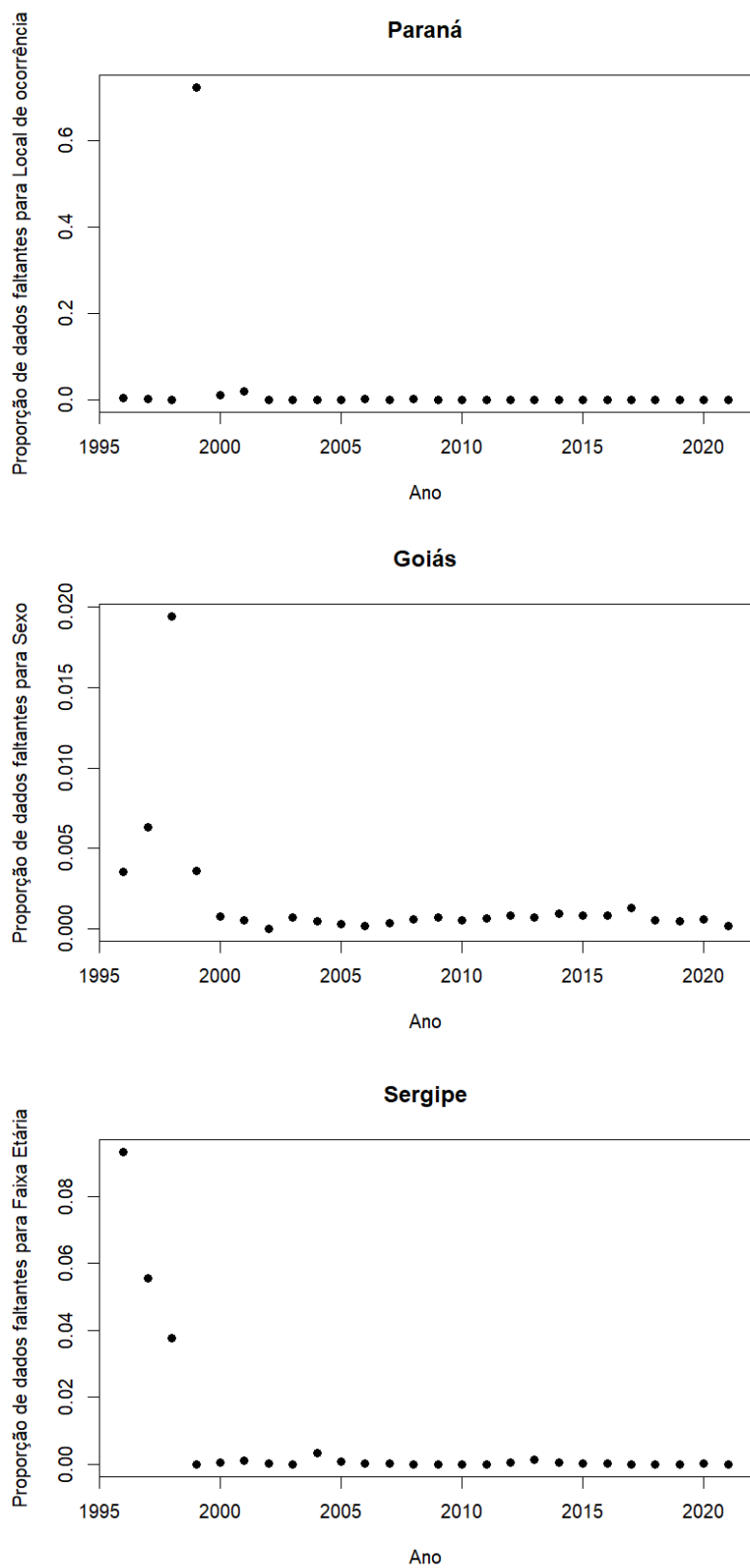
Observando a Figura 6, podemos notar que para todas as Unidades da Federação se teve uma queda no período analisado, isto é, a quantidade de campos não preenchidos caiu desde o início do intervalo em 1996. Interessante notar que para os anos de pandemia para os quais se dispõem de dados (2020 e 2021), não houve um aumento expressivo nos valores, com exceções pontuais como o Acre, que já possuía uma tendência crescente desde 2018, e Espírito Santo, que teve um pico em 2020, mas retornou ao seu padrão pré-pandemia em 2021.

FIGURA 1 – Proporções de dados faltantes de diferentes variáveis por ano (SIM)



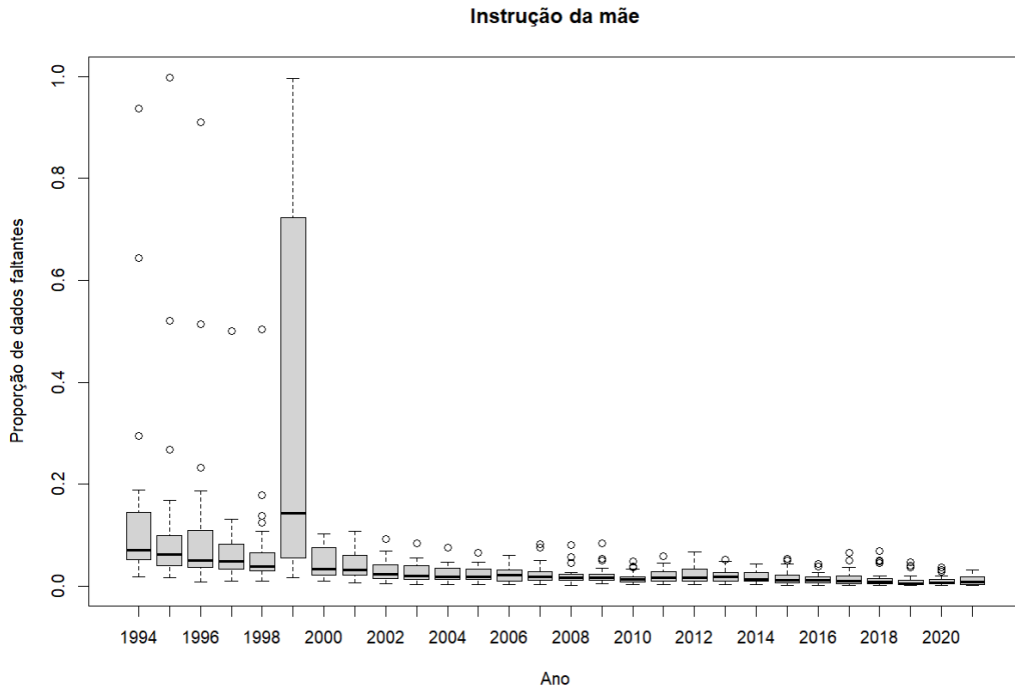
Fonte: Datasus (Tabnet), ago/2023.

FIGURA 2 – Proporções de dados faltantes de variáveis com valores suspeitos por ano (SIM)



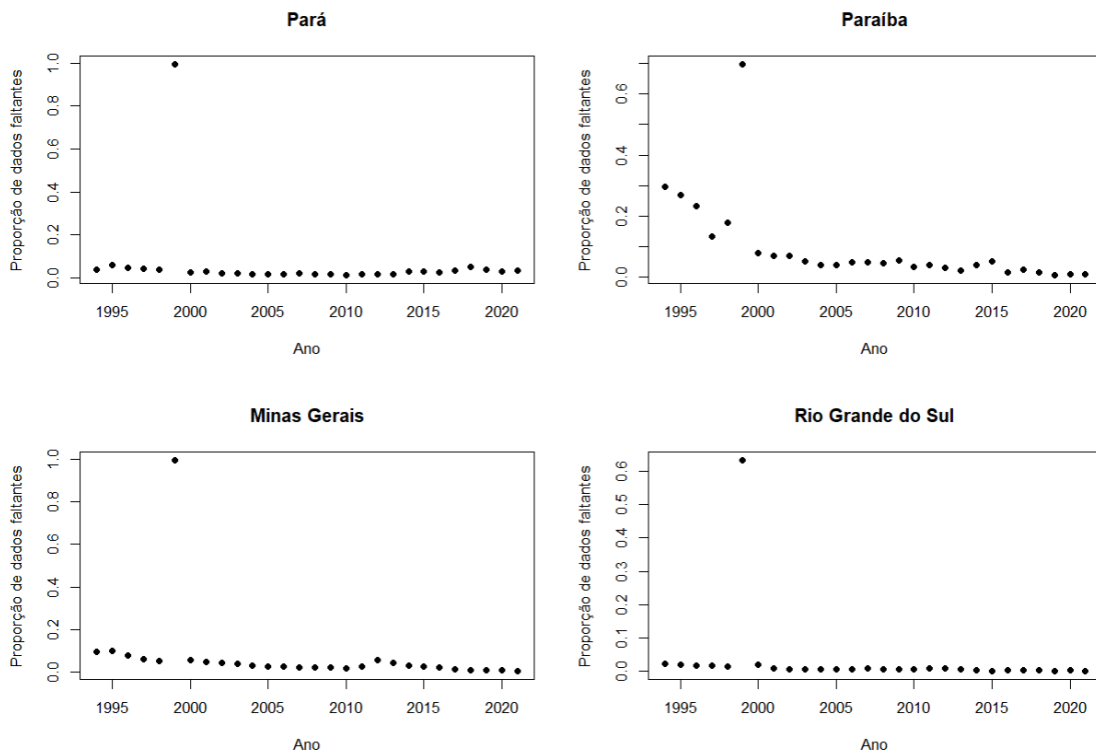
Fonte: Datasus (Tabnet), ago/2023.

FIGURA 3 – Proporções de dados faltantes da instrução materna por ano (SINASC)



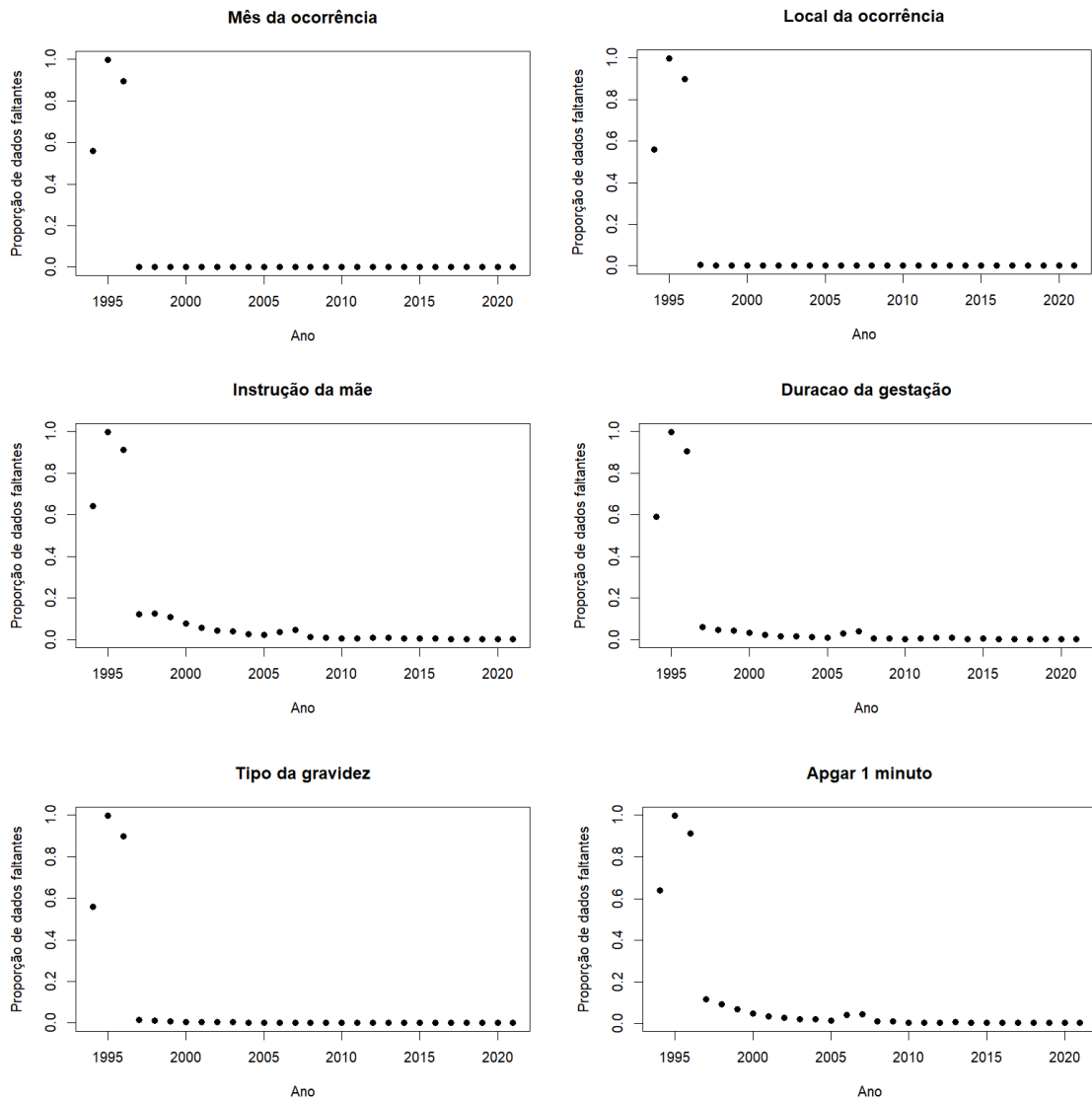
Fonte: Datasus (Tabnet), ago/2023.

FIGURA 4 – Proporções de dados faltantes da instrução materna por ano para UFs selecionadas (SINASC)



Fonte: Datasus (Tabnet), ago/2023.

FIGURA 5 – Proporções de dados faltantes de diferentes variáveis por ano para São Paulo (SINASC)

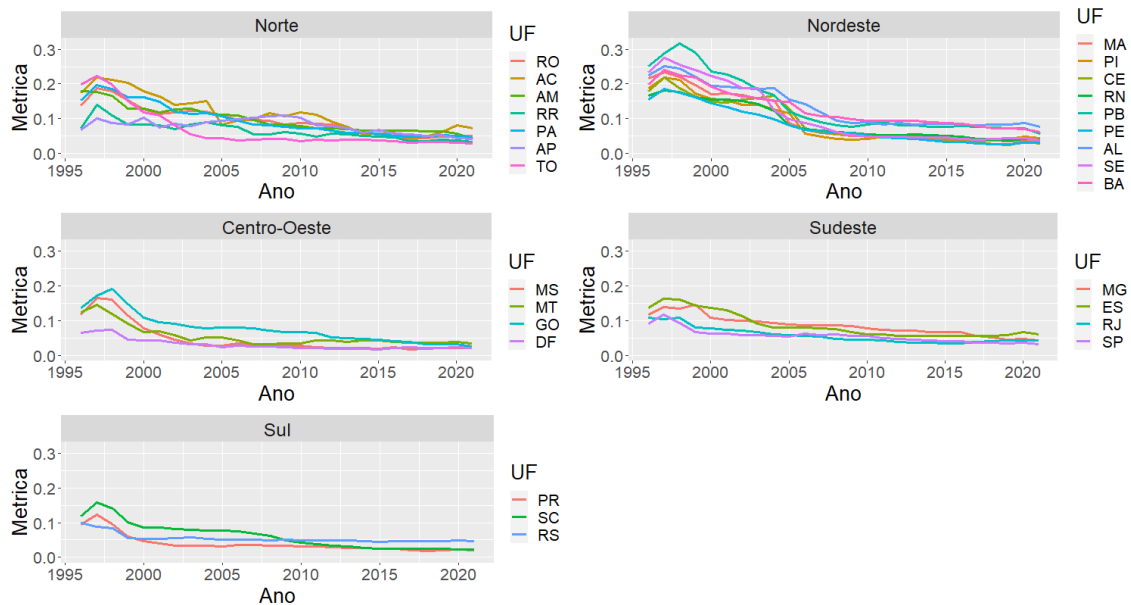


Fonte: Datasus (Tabnet), ago/2023.

Ao olhar os dados para nascidos vivos sem ignorar os valores suspeitos (Figura 7) pode-se notar os valores expressivos para São Paulo de 1994 a 1996, assim como Sergipe em 1994 e os variados picos em 1999. Não se percebem tendências devido ao achatamento do eixo ocasionado pelos altíssimos valores de São Paulo.

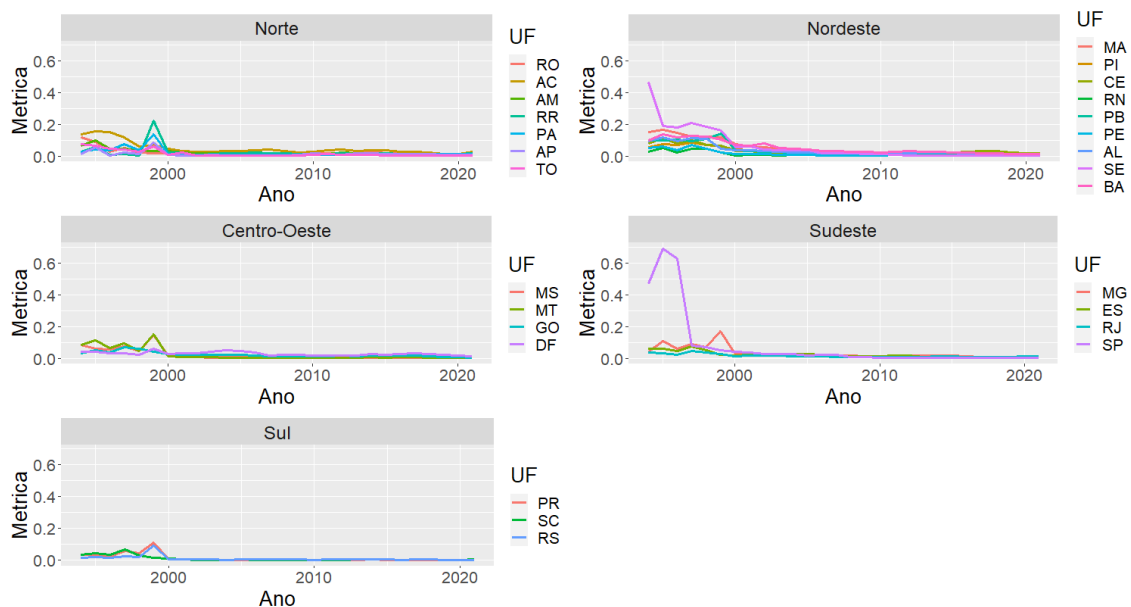
Já na Figura 8, que ignora os valores suspeitos já discutidos, pode-se notar uma tendência de queda, mas com algumas flutuações ao longo dos anos. Os picos existentes em 1995 e 1997/1998 são explicáveis não como suspeitos, mas como resultados de serem mensurações nos anos de adoção das variáveis Consultas pré-natal e Raça/Cor, respectivamente, período no qual algumas UFs já apresentavam valores baixos para seu não preenchimento e outras (as com os picos) valores altos.

FIGURA 6 – Métrica de qualidade do preenchimento dos dados de óbitos por UF e ano (SIM)



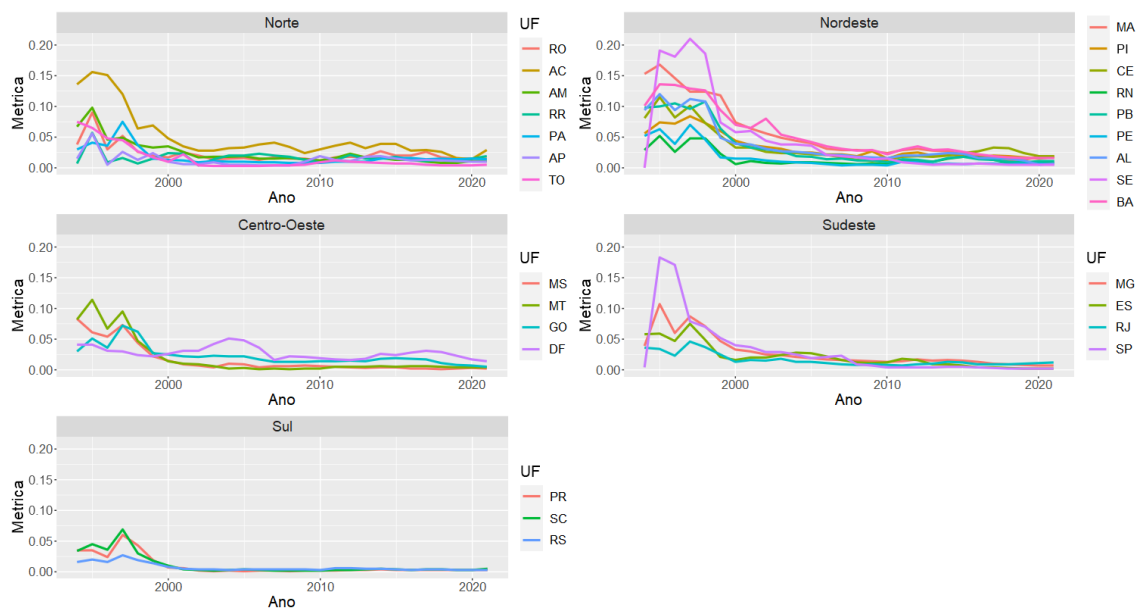
Fonte: TABNET.

FIGURA 7 – Métrica de qualidade do preenchimento dos dados originais de nascimentos por UF e ano (SINASC)



Fonte: Datasus (Tabnet), ago/2023.

FIGURA 8 – Métrica de qualidade do preenchimento dos dados de nascimentos por UF e ano (SINASC)



Fonte: Datasus (Tabnet), ago/2023.

CONSIDERAÇÕES FINAIS

Este artigo objetivou a criação de um inventário de valores suspeitos quanto ao uso das categorias “ignorado” e afins utilizadas pelos dois principais sistemas de informação utilizados por demógrafos com o intuito de apontar possíveis falhas em criações de séries históricas e também prover uma listagem para que os responsáveis pela base possam realizar verificações nos registros originais.

O segundo objetivo deste trabalho foi a proposta e análise de uma métrica para avaliar a qualidade do preenchimento dos registros para todas as Unidades da Federação ao longo dos anos existentes para consulta (1994 a 2021 para os nascimentos e 1996 a 2021 para os óbitos). Pode-se notar uma tendência de queda generalizada no período, para ambos os tipos de registro. Percebeu-se também que a pandemia de COVID-19 não teve impacto no preenchimento dos documentos, embora seja necessário comentar que este trabalho não avalia a cobertura ou validade dos documentos, apenas se os existentes foram preenchidos.

REFERÊNCIAS

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Declaração de nascido vivo:** manual de instruções para preenchimento. Brasília, DF, 2022a. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/svsa/vigilancia/declaracao-de-nascido-vivo-manual-de-instrucoes-para-preenchimento>. Acesso em: 20 set. 2023.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Declaração de óbito:** manual de instruções para preenchimento. Brasília, DF, 2022b. Disponível em:

http://bvsmms.saude.gov.br/bvs/publicacoes/declaracao_obito_manual_preenchimento.pdf. Acesso em: 20 set. 2023.

BRASIL. Ministério da Saúde. **Sistemas de informação em saúde**. Brasília, DF, 28/07/2021. Disponível em: <https://www.gov.br/saude/pt-br/composicao/svsa/vigilancia-de-doencas-chronicas-nao-transmissiveis/sistemas-de-informacao-em-saude>. Acesso em: 18 set. 2023.

BRASIL. Ministério da Saúde. **A experiência brasileira em sistemas de informação em saúde**. Brasília, DF, 2009a.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Portaria nº 116, de 11 de fevereiro de 2009. Regulamenta a coleta de dados, fluxo e periodicidade de envio das informações sobre óbitos e nascidos vivos para os Sistemas de Informações em Saúde sob gestão da Secretaria de Vigilância em Saúde. **Diário Oficial da República Federativa do Brasil**, Brasília, DF, 2009b.

R CORE TEAM. **R: a language and environment for statistical computing** 4.3.0. Vienna, Austria, 2023.

RIPSA – REDE INTERAGENCIAL DE INFORMAÇÃO PARA A SAÚDE. **Indicadores básicos para a Saúde: conceitos e aplicações**. Brasília, DF: OPAS, 2008. Disponível em: <http://tabnet.datasus.gov.br/tabdata/livroidb/2ed/indicadores.pdf>. Acesso em: 19 set. 2023.