MODELAGEM DE DADOS RELACIONAL DOS MICRODADOS DO SINASC COMO FERRAMENTA DE SUPORTE PARA ANÁLISE DE INDICADORES

Francisco Meneguini[1], Carlos Eduardo Beluzo[2], Luciana Correia Alves[3]

[1]Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), Câmpus Campinas, Campinas, SP, Brasil (fjmeneguini@gmail.com)
[2]Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP), Câmpus Campinas, Campinas, SP, Brasil (beluzo@ifsp.edu.br)
[3]Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brasil (lcalves@unicamp.br)

Resumo: Este estudo propõe uma infraestrutura reprodutível para ingestão, harmonização e modelagem relacional dos microdados do SINASC no recorte 2013-2023. O pipeline processou 30.983.111 registros sem perda de linhas, preservou a rastreabilidade por logs, integrou referência territorial do IBGE e gerou visões mensais de prematuridade e baixo peso ao nascer. Os dados foram organizados em PostgreSQL com tabela fato particionada por ano, dimensões de apoio e índices compostos para consultas epidemiológicas. Na etapa complementar, as séries mensais foram exportadas, validadas e comparadas em modelos ARIMA e LSTM no Google Colab. O ARIMA apresentou melhor desempenho para prematuridade, enquanto o LSTM foi superior para baixo peso, reforçando que o melhor modelo depende da assinatura temporal do indicador e da estrutura da série.

Palavras-chave: SINASC; engenharia de dados; modelagem relacional; séries temporais; prematuridade.

INTRODUÇÃO

O SINASC é uma fonte estratégica para monitoramento da saúde materno-infantil no Brasil, mas sua utilização em pesquisa ainda exige grande esforço de preparação, harmonização e validação dos dados. Em séries históricas longas, alterações de layout e codificação aumentam o risco de inconsistências metodológicas e reduzem a comparabilidade entre estudos.

Ao mesmo tempo, a literatura em saúde pública e ciência de dados tem reforçado que a qualidade do pipeline é parte da própria metodologia de pesquisa. Isso significa que a robustez das análises depende não apenas do modelo estatístico adotado, mas também da rastreabilidade, integridade e padronização da base que alimenta essas análises (Lima et al., 2009; Paiva et al., 2011; Szwarcwald et al., 2019).

Esse cenário é particularmente relevante quando se trabalha com bases administrativas em escala nacional. O SINASC reúne milhões de registros anuais, com cobertura territorial ampla, mas também apresenta desafios de harmonização histórica entre layouts, codificações e campos sensíveis a ausência ou inconsistência. Em fluxos ad hoc, cada grupo de pesquisa tende a reconstruir seu próprio processo de limpeza, o que eleva retrabalho, dificulta auditoria e enfraquece a reprodutibilidade.

Neste contexto, este trabalho organizou uma arquitetura de dados reprodutível para o SINASC, combinando ETL automatizado, modelagem relacional em PostgreSQL e publicação de indicadores mensais. A proposta não se limita à preparação técnica da base: ela estrutura um ambiente que permite pesquisas descritivas, inferenciais e preditivas com menor fricção operacional e maior rastreabilidade metodológica. Dois desfechos foram priorizados por sua relevância epidemiológica e operacional: prematuridade e baixo peso ao nascer.

Como etapa complementar, as séries exportadas foram testadas em modelos ARIMA e LSTM para verificar o uso analítico da infraestrutura produzida. Essa inclusão é importante porque mostra que a base não foi construída apenas para armazenamento; ela foi desenhada para suportar análise temporal real, com consumo direto por ferramentas estatísticas e de aprendizado de máquina.

O objetivo do estudo foi desenvolver e validar uma base analítica escalável e auditável para uso em

pesquisas populacionais e vigilância em saúde. A hipótese de trabalho é que um pipeline auditável, associado a um banco relacional bem modelado, reduz a fricção de preparo dos dados e aumenta a confiabilidade das análises temporais.

Além do valor técnico, a proposta responde a uma necessidade científica concreta. Indicadores como prematuridade e baixo peso ao nascer são amplamente usados em vigilância materno-infantil e funcionam como sinais sintéticos de desigualdade, qualidade da assistência e mudanças territoriais. Ter essas medidas disponíveis em formato padronizado, mensal e rastreável significa reduzir o tempo entre a pergunta científica e a evidência produzida. Em bases públicas do SUS, esse ganho de infraestrutura é, na prática, um ganho metodológico.

MATERIAL E MÉTODOS

A pesquisa foi estruturada como estudo aplicado em ciência de dados para saúde pública. A fonte primária foi o acervo público do SINASC em arquivos anuais compactados, cobrindo 2013 a 2023. O volume consolidado no ETL foi de 30.983.111 registros.

O recorte temporal foi escolhido por equilíbrio entre disponibilidade de arquivos, consistência histórica e volume suficiente para análises temporais mensais. Ao reunir onze anos consecutivos, o projeto permitiu comparar tendências, sazonalidade e estabilidade do comportamento dos indicadores em um horizonte longo o bastante para modelagem, mas ainda operacional para uma infraestrutura de pesquisa desenvolvida em um TCC.

O pipeline incluiu: download automatizado, leitura dos arquivos, normalização de ausências, padronização de nomes de colunas, harmonização histórica entre anos, conversão de tipos, regras de validade para campos selecionados, enriquecimento com referência territorial do IBGE e geração de logs de execução. A estratégia adotada preservou os registros originais, marcando campos inconsistentes como ausentes em vez de descartar linhas. Esse desenho foi escolhido para reduzir viés de exclusão e manter a base populacional íntegra mesmo quando alguns campos derivados precisavam ser invalidados.

Na prática, isso significou lidar com layouts heterogêneos entre anos sem perder rastreabilidade. Colunas foram reunidas por união de cabeçalhos, o conteúdo bruto foi limpo e os campos derivados passaram por regras explícitas de tipagem. A ideia foi converter um conjunto de arquivos históricos em uma camada unificada, sem apagar as características originais que pudessem ser úteis em auditorias ou novas interpretações analíticas.

Na camada de ingestão, os arquivos anuais foram obtidos de forma automatizada e organizados por ano, o que permite reexecução controlada. Em seguida, os dados passaram por harmonização de cabeçalhos, padronização de valores faltantes, limpeza de espaços e uniformização de tipos. O objetivo foi transformar um acervo heterogêneo em uma base única e estável, útil para consulta científica sem necessidade de etapas manuais adicionais.

Esse fluxo foi acompanhado por registros de execução e manifesto de arquivos, o que permite identificar em qual rodada cada recurso foi processado. Em bases de saúde pública, esse tipo de controle é particularmente importante porque pequenas decisões de transformação podem alterar resultados agregados. A rastreabilidade reduz esse risco e viabiliza comparação entre versões do pipeline.

A persistência foi implementada em PostgreSQL com tabela fato particionada por ano, dimensões de apoio e índices compostos para consultas frequentes. A partir desse modelo, foram produzidas visões mensais de prematuridade e baixo peso ao nascer, exportadas para análise temporal. O particionamento anual foi adotado para favorecer pruning em consultas por período e melhorar o desempenho em cenários recorrentes de agregação epidemiológica.

A estrutura relacional foi desenhada em torno de consultas científicas recorrentes. Houve uma tabela fato central para os nascimentos e dimensões mínimas para município, estabelecimento e categorias selecionadas. A decisão por um modelo analítico mais enxuto do que um data warehouse clássico foi deliberada: o objetivo era otimizar leitura, reuso e desempenho em consultas epidemiológicas sem introduzir complexidade desnecessária ao projeto.

A verificação de integridade foi executada em múltiplos níveis. Primeiro, houve checagem da transferência e da presença local dos arquivos. Depois, a consistência de linha entre o bruto e o harmonizado foi comparada. Por fim, as ausências em campos-chave foram analisadas antes e depois da transformação. O relatório consolidado mostrou correspondência exata entre linhas lidas e linhas gravadas, sem perda de observações durante o ETL.

Essa etapa é central porque distingue falha real de transformação legítima. Em vez de remover observações quando algum campo não atendia a regra de validação, o fluxo preservou a linha e sinalizou o problema no campo correspondente. Isso evita redução artificial do N e também impede que um valor inválido seja tratado como informação confiável na análise posterior.

Na etapa complementar, as séries foram avaliadas em Google Colab. Foi adotado split temporal fixo com treino em 2013-2022 e teste em 2023. Os modelos comparados foram ARIMA sazonal, com

configuração SARIMAX (1,1,1)(1,1,1,12), e LSTM com janela de 12 meses. A avaliação utilizou MAE, RMSE e MAPE. Essa comparação foi intencionalmente mantida com o mesmo horizonte temporal para permitir leitura direta dos resultados.

O ARIMA foi usado como baseline estatístico forte em séries sazonais mensais, enquanto o LSTM entrou como alternativa não linear de aprendizado profundo. A janela de 12 meses foi adotada para capturar a sazonalidade anual e permitir comparação mais justa entre os métodos. O split temporal único também evita vazamento de informação do futuro para o passado, algo importante quando o objetivo é simular uso real de previsão.

As métricas foram escolhidas por complementaridade interpretativa. O MAE informa o erro absoluto médio na escala do indicador, o RMSE penaliza mais fortemente erros grandes e o MAPE facilita a comparação relativa entre séries com valores médios distintos. A combinação dessas medidas reduz a chance de uma avaliação parcial do desempenho preditivo.

Na camada de saída, os dois indicadores mensais foram exportados para arquivos CSV específicos e lidos novamente no ambiente analítico. Antes da modelagem, cada série passou por validação estrutural para verificar frequência mensal regular, ausência de lacunas e coerência do intervalo temporal. Esse cuidado é essencial em séries epidemiológicas, porque uma falha de exportação pode gerar uma previsão aparentemente boa, mas metodologicamente inválida.

Do ponto de vista de reprodutibilidade, o fluxo metodológico ficou apoiado em três tipos de artefato: scripts versionados de download, ETL e exportação; SQLs versionados para modelagem e carga incremental; e logs de execução e relatórios de integridade. Essa organização é importante porque o valor científico do trabalho não está apenas nos resultados finais, mas na capacidade de reproduzi-los sem depender de memória operacional do pesquisador.

Por fim, a estratégia metodológica foi pensada para ser transferível. O mesmo desenho pode ser adaptado a outras bases do SUS que apresentem múltiplos arquivos anuais, mudanças de layout e necessidade de consultas periódicas. Essa transferibilidade é relevante porque a contribuição principal do trabalho é o método, não apenas o recorte específico do SINASC.

RESULTADOS E DISCUSSÃO

O principal resultado operacional foi a consolidação de uma infraestrutura íntegra e reprodutível para o SINASC em escala nacional. O ETL processou 30.983.111 registros com correspondência exata entre linhas lidas e linhas gravadas, indicando ausência de perda por exclusão durante a harmonização. Esse achado é relevante porque reduz o risco de distorções silenciosas em bases administrativas de grande volume e preserva a consistência do conjunto populacional ao longo do recorte histórico.

Esse resultado responde diretamente a uma fragilidade comum em fluxos manuais: quando a preparação depende de etapas ad hoc, pequenas diferenças entre rodadas podem produzir conjuntos ligeiramente distintos, o que complica validação e comparação entre estudos. Ao automatizar a ingestão e o ETL, o projeto transforma um trabalho repetitivo em um processo estável, mensurável e auditável.

Além da preservação das linhas, a base passou por harmonização de colunas entre anos, enriquecimento territorial com o IBGE e padronização de códigos relevantes para consultas analíticas. Em bases heterogêneas como o SINASC, esse tipo de tratamento evita que diferenças de layout sejam interpretadas como diferenças substantivas no fenômeno estudado. A infraestrutura, portanto, não é apenas um repositório: ela organiza semanticamente o dado para uso científico.

Na prática, isso significa que o pesquisador deixa de gastar tempo com tarefas de limpeza básicas e passa a trabalhar sobre uma camada já calibrada para consulta. A separação entre dado bruto e dado analítico é uma contribuição metodológica importante, porque reduz o retrabalho e melhora a comparabilidade entre projetos distintos que usem a mesma fonte.

Na camada relacional, a organização por partições anuais e índices compostos favoreceu consultas analíticas recorrentes com bom desempenho. Isso torna a base mais adequada para estudos epidemiológicos que dependem de agregações por município, UF, período e estabelecimento, sem necessidade de reprocessamento a cada nova pergunta de pesquisa. Os testes de desempenho mostraram que a estrutura responde bem tanto a consultas pontuais quanto a filtros mais amplos, o que é importante em um cenário em que múltiplos usuários podem reutilizar o mesmo banco para diferentes perguntas.

Esse comportamento é coerente com a finalidade do modelo. Em vez de priorizar flexibilidade máxima para qualquer tipo de consulta, a estrutura foi otimizada para os padrões de uso mais prováveis em pesquisa com SINASC: agregações temporais, recortes geográficos e consultas clínicas recorrentes. O uso de partições por ano ajuda o mecanismo do banco a ignorar blocos irrelevantes de dados em consultas específicas, o que melhora eficiência sem sacrificar legibilidade da modelagem.

As visões mensais de prematuridade e baixo peso mostraram cobertura temporal completa de janeiro de 2013 a dezembro de 2023, o que permitiu a modelagem temporal sem lacunas. Na comparação preditiva, o ARIMA apresentou melhor desempenho para prematuridade, com MAE=0,9303, RMSE=1,4852 e MAPE=0,8158. Para baixo peso, o LSTM foi superior, com MAE=1,2233, RMSE=1,5342 e MAPE=1,2989.

Os valores observados são baixos o bastante para mostrar que a base produzida fornece um sinal mensal consistente e útil para modelagem. No caso da prematuridade, a aderência do ARIMA sugere que a série possui comportamento mais compatível com sazonalidade e inércia temporal clássica. Já no baixo peso, a vantagem do LSTM mostra que uma arquitetura mais flexível pode capturar nuances adicionais mesmo com uma janela relativamente curta de treinamento.

Esse contraste sugere que não há modelo universalmente superior para séries epidemiológicas mensais. Em séries mais regulares e com forte componente sazonal, o ARIMA pode ser suficiente e até mais preciso. Em outras condições, a flexibilidade do LSTM pode capturar padrões adicionais e melhorar a previsão. Assim, a infraestrutura produzida não apenas organiza os dados, mas também viabiliza benchmark metodológico por indicador.

Além disso, o contraste entre os modelos reforça uma conclusão importante para o uso em vigilância: a escolha do modelo deve ser orientada pelo indicador e não por preferência metodológica abstrata. Em aplicações reais, isso permite calibrar a estratégia preditiva conforme o comportamento da série, em vez de impor um método único para todos os desfechos.

Em termos de contribuição científica, o estudo mostra que a etapa de engenharia de dados não é apenas operacional: ela condiciona a qualidade da análise temporal e a confiança nos resultados obtidos. A combinação entre rastreabilidade, modelagem relacional e publicação de indicadores cria uma base reutilizável para monitoramento, comparação entre métodos e desenvolvimento de estudos futuros.

O desempenho observado na etapa preditiva ajuda a interpretar o próprio valor da infraestrutura. O ARIMA teve melhor aderência para prematuridade porque essa série pareceu mais alinhada a uma estrutura sazonal clássica, enquanto o LSTM apresentou vantagem em baixo peso, sugerindo que a flexibilidade da rede conseguiu capturar nuances adicionais nesse indicador. Em outras palavras, a mesma base permite discutir o comportamento de modelos distintos sem necessidade de retrabalho estrutural.

A validação das séries mensais exportadas também é um resultado em si. Cobertura completa, frequência mensal regular e ausência de lacunas temporais são pré-condições mínimas para previsões confiáveis. Sem esse cuidado, o melhor modelo pode produzir projeções frágeis ou até enganosas. O fato de as séries exportadas manterem 132 observações contínuas demonstra que a integração entre SQL e análise temporal foi consistente.

Outro ponto importante é a utilidade operacional das projeções. As estimativas para 2024 não foram tratadas como previsão definitiva, mas como prova de que o banco e as visões analíticas podem alimentar rotinas de monitoramento prospectivo. Isso é particularmente valioso em vigilância em saúde, onde a capacidade de antecipação pode orientar priorização de análises mais aprofundadas.

Em síntese, o conjunto de resultados indica que a contribuição central do estudo não é apenas a criação de um banco ou a obtenção de métricas pontuais, mas a entrega de uma infraestrutura científica capaz de sustentar novas perguntas, novos recortes e novas estratégias de modelagem com menor custo de preparação e maior confiança no dado produzido.

LIMITAÇÕES E IMPLICAÇÕES

Apesar dos resultados positivos, o estudo possui limitações que precisam ser explicitadas. A primeira é que a qualidade final depende da consistência do dado de origem administrativo. Se um campo já nasce incompleto ou irregular na base pública, o pipeline pode preservar a observação e sinalizar ausência, mas não consegue recuperar informação que não existe. A segunda é que a análise temporal foi feita em recorte mensal agregado, sem estratificações subnacionais mais finas nesta versão do artigo.

Há ainda limitações relacionadas à comparação preditiva. O trabalho usou dois modelos e um split temporal único, o que é suficiente para demonstrar utilidade metodológica, mas não esgota o espaço de benchmarking possível. Uma avaliação mais extensa poderia testar outros algoritmos, janelas móveis e variáveis exógenas. Mesmo assim, os resultados obtidos já são adequados ao objetivo do manuscrito, que é mostrar a funcionalidade da infraestrutura e sua utilidade analítica inicial.

Do ponto de vista aplicado, o estudo abre caminho para uso em vigilância em saúde. Uma base mensal padronizada pode ser usada para monitorar tendência, sazonalidade e desvios em relação ao esperado, além de apoiar recortes por território e períodos específicos. Isso é especialmente útil quando o interesse é transformar dado bruto em informação acionável com o menor retrabalho possível.

O desempenho observado na etapa preditiva ajuda a interpretar o próprio valor da infraestrutura. O ARIMA teve melhor aderência para prematuridade porque essa série pareceu mais alinhada a uma estrutura sazonal clássica, enquanto o LSTM apresentou vantagem em baixo peso, sugerindo que a flexibilidade da rede conseguiu capturar nuances adicionais nesse indicador. Em outras palavras, a mesma base permite discutir o comportamento de modelos distintos sem necessidade de retrabalho estrutural.

A validação das séries mensais exportadas também é um resultado em si. Cobertura completa, frequência mensal regular e ausência de lacunas temporais são pré-condições mínimas para previsões confiáveis. Sem esse cuidado, o melhor modelo pode produzir projeções frágeis ou até enganosas. O fato de as séries exportadas manterem 132 observações contínuas demonstra que a integração entre SQL e análise temporal foi consistente.

Outro ponto importante é a utilidade operacional das projeções. As estimativas para 2024 não foram tratadas como previsão definitiva, mas como prova de que o banco e as visões analíticas podem alimentar rotinas de monitoramento prospectivo. Isso é particularmente valioso em vigilância em saúde, onde a capacidade de antecipação pode orientar priorização de análises mais aprofundadas.

Do ponto de vista metodológico, a comparação entre modelos reforça uma lição recorrente em séries epidemiológicas de baixa granularidade temporal: complexidade não garante superioridade. Quando o número de observações é relativamente restrito, modelos mais parcimoniosos podem generalizar melhor em determinados indicadores, enquanto arquiteturas mais flexíveis podem se beneficiar em séries com dinâmica menos linear. O presente trabalho ilustra esse equilíbrio sem inflar a discussão para além do que os dados suportam.

Em síntese, o conjunto de resultados indica que a contribuição central do estudo não é apenas a criação de um banco ou a obtenção de métricas pontuais, mas a entrega de uma infraestrutura científica capaz de sustentar novas perguntas, novos recortes e novas estratégias de modelagem com menor custo de preparação e maior confiança no dado produzido.

CONCLUSÃO

O trabalho demonstrou que é possível transformar microdados do SINASC em uma infraestrutura analítica íntegra, auditável e reutilizável para pesquisa em saúde pública. A combinação de ETL reprodutível, modelagem relacional e indicadores mensais permitiu análises temporais consistentes e comparações entre ARIMA e LSTM.

Como resultado prático, o estudo entrega uma base pronta para investigações epidemiológicas e demográficas, reduzindo retrabalho técnico e ampliando a confiabilidade das análises. Como resultado metodológico, evidencia que a qualidade do pipeline é determinante para a utilidade científica da base.

De forma mais ampla, o trabalho mostra que bases administrativas de saúde podem ser tratadas como infraestrutura científica quando recebem tratamento de engenharia compatível com seu volume, variabilidade histórica e potencial analítico. Nesse sentido, o estudo entrega não apenas um banco de dados, mas uma forma de organizar o conhecimento extraído do SINASC para uso contínuo em pesquisas futuras.

Para o contexto do CoBICET, essa entrega é suficiente para demonstrar relevância aplicada, rigor metodológico e possibilidade de reuso em outras investigações. O resultado final é um artigo que discute o valor da estrutura de dados como componente central da produção científica em saúde pública.

AGRADECIMENTOS

Quero agradecer, em primeiro lugar, à minha namorada, Karen Cristina, pelo apoio, pela paciência e por estar comigo em todos os momentos dessa caminhada. Agradeço também aos meus pais, André Meneguini e Patrícia Meneguini, e ao restante da minha família, pelo incentivo, pela compreensão e por acreditarem em mim mesmo nos momentos mais difíceis.

Aos coautores deste projeto, deixo meu agradecimento pela parceria, pelas contribuições e pela troca de conhecimento ao longo do desenvolvimento do trabalho. Por fim, agradeço ao Instituto Federal de São Paulo e à Universidade Estadual de Campinas, pelo suporte institucional e pela formação acadêmica que tornaram este estudo possível.

REFERÊNCIAS

BARRETO, M. L. et al. The Centre for Data and Knowledge Integration for Health (CIDACS): linking health and social data in Brazil. ***International Journal of Population Data Science***, 2019.

BLENCOWE, H. et al. National, regional, and worldwide estimates of preterm birth rates in 2010 with time trends since 1990. ***The Lancet***, 2012.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. ***Time series analysis: forecasting and control***. 5. ed. Hoboken: Wiley, 2015.

DOMBROWSKI, J. G. et al. Effectiveness of the Live Births Information System in the far-western Brazilian Amazon. ***Ciência e Saúde Coletiva***, 2015.

DOMINGUES, R. M. et al. Process of decision-making regarding the mode of birth in Brazil. ***Cadernos de Saúde Pública***, 2014.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. ***Deep learning***. Cambridge: MIT Press, 2016.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. ***Neural Computation***, v. 9, n. 8, p. 1735-1780, 1997.

HYNDMAN, R. J.; ATHANASOPOULOS, G. ***Forecasting: principles and practice***. 3. ed. Melbourne: OTexts, 2021. Disponível em: https://otexts.com/fpp3/. Acesso em: 24 abr. 2026.

LIMA, C. A. et al. Revisão das dimensões de qualidade dos dados e métodos aplicados na avaliação dos sistemas de informação em saúde. ***Cadernos de Saúde Pública***, 2009.

MARTINELLI, K. G. et al. Prematuridade no Brasil entre 2012 e 2019: dados do SINASC. ***Revista Brasileira de Estudos de População***, 2021.

MITCHELL, S. et al. FAIR data pipeline: provenance-driven data management for traceable scientific workflows. ***Philosophical Transactions of the Royal Society A***, 2021.

NAMLI, T. et al. A scalable and transparent data pipeline for AI-enabled health data ecosystems. ***Frontiers in Medicine***, 2024.

PAIVA, N. S. et al. Sistema de Informações de Nascidos Vivos: uma revisão sobre seus usos na pesquisa em saúde. ***Ciência e Saúde Coletiva***, 2011.

ROMERO, D. E.; CUNHA, C. B. Avaliação da qualidade das variáveis epidemiológicas e demográficas do Sistema de Informações sobre Nascidos Vivos, 2002. ***Cadernos de Saúde Pública***, 2007.

SILVA, L. M. S. et al. Avaliação da qualidade dos dados do SINASC e do SIM no período neonatal, Espírito Santo, 2007-2009. ***Ciência e Saúde Coletiva***, 2014.

SZWARCWALD, C. L. et al. Avaliação dos dados do Sistema de Informações sobre Nascidos Vivos (SINASC), Brasil. ***Cadernos de Saúde Pública***, 2019.

WORLD HEALTH ORGANIZATION. ***Preterm birth***. 2022. Disponível em: https://www.who.int/news-room/fact-sheets/detail/preterm-birth. Acesso em: 20 mar. 2026.