

Cristiane Yumi Nakamura¹
Everton Osnei Cesario¹
Lorena Maria Tatim Farhat¹
Simone Tetu Moysés¹
Deborah Ribeiro Carvalho¹

The use of data-mining to support maternal and child health management

ABSTRACT | Introduction: *Data mining (DM) is part of KDD (Knowledge Discovery in Database), which is a computational process focused on discovering new, valid and useful knowledge in databases in order to substantiate decision-making processes. Objective:* *The aim of the current integrative review is to investigate the potential use of DM to support maternal and child health management processes. Methods:* *Search process was adapted from the PRISMA method and applied to BVS, PubMed, Scopus (Elsevier) and IEEE Xplore repositories based on MeSH® terms such as “data mining” AND health AND child* OR maternal OR pregnant*. Results:* *Twenty-nine documents were included in the study: nine of them confirmed previous findings, ten focused on investigating models’ accuracy (lack of discussion with previous studies) and ten reported understandable, valid, new and useful results – associations between new chemical elements and ozone on asthma, between specific Australian ethnicity and stillbirth, between Indian American race and prematurity, between vaccination and prematurity, between geriatric pregnancy and low prenatal adherence, among others. Conclusion:* *Results have indicated DM’s potential to support maternal and child health management processes, mainly in interdisciplinary domains. A gap was identified between DM use and the effective application of its results.*

Keywords | *Maternal and Child Health; Social Determinants of Health; Data mining.*

| Mineração de dados no apoio à gestão em saúde materno-infantil

RESUMO | Introdução: A Mineração de Dados (MD) é uma etapa do Knowledge Discovery in Database (KDD), processo computacional que visa a descobrir conhecimentos novos, válidos e úteis em base de dados, fornecendo subsídios para a tomada de decisão. **Objetivo:** Verificar o potencial do uso da MD para apoio à gestão em saúde materno-infantil. **Métodos:** Busca adaptada do método PRISMA, realizada nos repositórios BVS, PubMed, Scopus (Elsevier) e IEEE Xplore, utilizando os descritores “data mining” AND health AND child* OR maternal OR pregnant*. **Resultados:** Foram incluídos 29 documentos, dos quais nove corroboraram a literatura, dez focaram a acurácia do modelo e dez encontraram resultados compreensíveis, válidos, novos e úteis, como associação de novos elementos químicos ao ozônio e asma, grupo étnico australiano e natimortalidade, raça indígena americana e prematuridade, vacinação e prematuridade, gestantes idosas e baixa adesão ao pré-natal. **Conclusão:** Os resultados demonstraram o potencial do uso da MD para apoio à gestão, especialmente quando a interdisciplinaridade se faz presente. Identificou-se uma lacuna entre o uso da MD e a aplicação efetiva dos seus resultados.

Palavras-chave | Saúde materno-infantil; Determinantes sociais da saúde; Mineração de dados.

¹Pontifícia Universidade Católica do Paraná. Curitiba/PR, Brasil.

INTRODUÇÃO |

Os primeiros 1.000 dias de vida são considerados críticos, pois situações ocorridas nesse período afetam a vida e a saúde da gestante e da criança, com repercussões na fase adulta e terceira idade¹. Essas repercussões podem ser positivas, como aumento da estatura, melhoria do desenvolvimento cognitivo, motor e socioemocional, da performance escolar, da capacidade de aprendizagem, de trabalho e produtividade. Entretanto, implicações negativas suscitam elevação da criminalidade, analfabetismo, obesidade, doenças cardíacas, problemas de saúde mental, morbidade e mortalidade na infância²⁻⁷.

Esse processo saúde-doença é influenciado por condições individuais (idade, sexo e estilo de vida) e coletivas (ambientais e socioeconômicas), representadas pelos Determinantes Sociais da Saúde (DSS)⁸. O modelo compreende um conjunto de cinco camadas concêntricas que demonstram as principais influências na saúde e respectivo nível de intervenção, por meio do qual é possível inferir o grau de governabilidade do Estado.

As camadas mais internas remetem às condições individuais, de menor governabilidade do Estado, e as mais externas, às condições gerais, de maior governabilidade do Estado e que influenciam e modulam as camadas mais individuais. Mudanças focadas nas camadas mais internas requerem ações pontuais, de prevenção ou educação em saúde, enquanto as intervenções voltadas às camadas mais externas são mais efetivas e demandam ações de promoção, com mudanças estruturais por meio de estratégias econômicas e/ou acordos envolvendo outros setores além da saúde⁸.

A despeito do conhecimento existente e ações já implementadas, dada a complexidade dos primeiros 1.000 dias de vida, desfechos indesejáveis, e por vezes evitáveis, perduram, ocasionando problemas de saúde pública^{7,9}. Nesse sentido, a Mineração de Dados (MD) pode auxiliar na descoberta de novos conhecimentos que possam vir a subsidiar a tomada de decisão¹⁰.

A MD é a segunda etapa do processo de Knowledge Discovery in Database (KDD) ou Descoberta de Conhecimento em Bases de Dados. A primeira etapa é o pré-processamento e a última, o pós-processamento. O KDD é um processo computacional que visa extrair informação e conhecimento implícito e escondido em uma base de dados, que podem ser úteis e até então

desconhecidos¹⁰. O algoritmo identifica padrões, relacionando variáveis e valores com a ocorrência ou não do desfecho estudado. Assim, a partir desses resultados (padrões), é possível identificar associações entre fatores de exposição e desfecho. Para gestão em saúde, é importante que essas associações estejam representadas de forma compreensível, de maneira que seres humanos possam identificar as causas (fatores de exposição) e consequências do problema de saúde (desfecho)¹¹⁻¹².

A associação entre fatores de exposição e desfecho pode ser testada por métodos estatísticos convencionais. Entretanto, os resultados podem não ser de fácil interpretação, especialmente quando mais de duas variáveis são consideradas simultaneamente¹³. Testes estatísticos apresentam inegável contribuição para delinear o comportamento dos dados¹⁴, porém, quando se dispõe de grande volume de dados, processos oriundos da computação, como a MD, podem vir a contribuir^{10,15}. Os testes estatísticos convencionais requerem um conhecimento prévio de dependência entre as variáveis, pressuposto não exigido para a realização da MD, aumentando a possibilidade de descoberta de novos conhecimentos^{10-11,16}, cuja validade e utilidade precisam ser verificadas por outros meios, como a avaliação de especialistas, a literatura existente e a própria estatística convencional¹⁴.

Considerando a importância dos primeiros 1.000 dias de vida, a permanência de desfechos indesejáveis nesse período e a finalidade da MD de descobrir conhecimentos novos, válidos e úteis, o objetivo deste estudo é identificar, por meio de uma revisão integrativa, os DSS dos primeiros 1.000 dias descobertos pela MD para apoio à gestão em saúde materno-infantil.

MÉTODOS |

As etapas foram adaptadas do método Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)¹⁷. Para a busca dos documentos, foram adotados os descritores “data mining” AND health AND child* OR maternal OR pregnan*, nos repositórios BVS, PubMed, Scopus (Elsevier) e IEEE Xplore, em maio de 2018. O filtro utilizado foi de documentos do tipo artigo, exceto para o repositório IEEE Xplore, em que também foram consideradas conferências, dada a natureza de documentos mais voltados à computação. Não foram aplicados critérios

de temporalidade nem idioma, no intuito de ampliar o escopo dos documentos.

Após a exclusão dos duplicados, foram analisados o título e/ou resumo. Para a leitura integral, foram selecionados documentos que adotassem a MD (KDD, tarefas de mineração, algoritmos, etc.) para o enfrentamento de qualquer situação relacionada à saúde materno-infantil, incluindo gestação, parto, puerpério, nascimento até os dois anos de idade da criança, compreendendo, dessa maneira, todo o período dos primeiros 1.000 dias de vida. Essa etapa de seleção foi realizada por dois pesquisadores independentes e avaliada a partir do teste de concordância Kappa. As calibrações foram realizadas até atingir um valor de Kappa maior que 0,8018. Os documentos com indicativo de seleção divergentes foram avaliados por um terceiro pesquisador independente.

A elegibilidade dos documentos foi baseada na leitura integral com base nos seguintes critérios: idioma em alfabeto romano e representação compreensível dos resultados obtidos pela MD, permitindo identificar os fatores de exposição e desfecho. Foram excluídos documentos disponíveis somente por compra ou comutação bibliográfica internacional, ou que contemplassem população infantil cuja faixa etária não estivesse especificada ou fora dos primeiros mil dias de vida. As características extraídas,

respectiva descrição e finalidade estão relacionadas no Quadro 1.

Para melhor compreensão, as características de temporalidade, quantidade de registros e variáveis das bases de dados, área de publicação, aderência à literatura e DSS serão mais bem detalhadas a seguir.

Para a temporalidade, quantidade de registros e de variáveis das bases de dados, após a extração de cada característica, foi calculada a mediana e verificada a posição dos estudos em relação a ela (inferior ou superior).

Com relação à área de publicação, para as revistas, foi consultado o portal de revistas da BVS (<http://portal.revistas.bvs.br/index.php?lang=pt>) e NLM Catalog/PubMed (<https://www.ncbi.nlm.nih.gov/nlmcatalog?cmd=historysearch&querykey=3>). Consideraram-se como área de publicação em saúde aqueles documentos cuja área temática e assuntos (BVS) ou current subset e broad subject term(s) (PubMed) mencionavam somente a saúde. Quando havia menção à área computacional (por exemplo, informática em saúde), entendeu-se como área de publicação interdisciplinar. A área de publicação em informática foi determinada quando não havia nenhuma menção à saúde (por exemplo, anais de congressos de MD, de modelagem e sis-

Quadro 1 – Características extraídas dos estudos incluídos, respectiva descrição e finalidade

Característica	Descrição	Finalidade
Título/autor/ano de publicação	Identificação do documento.	Inferir a atualidade do tema.
Área de publicação	Área de publicação: saúde, informática ou interdisciplinar.	Identificar a existência de interdisciplinaridade.
Desfecho	Situação ou problema de saúde objeto-alvo do estudo.	Verificar em quais situações tem sido utilizada a MD.
Temporalidade da base de dados	Quantidade de anos da base de dados.	Dimensionar o tamanho da base de dados e inferir se ele influencia a descoberta de novos conhecimentos.
Quantidade de registros	Número de registros da base de dados.	
Quantidade de variáveis	Número de variáveis da base de dados.	
Aderência à literatura	Associações encontradas corroboram, divergem ou são interessantes frente à literatura.	Verificar a contribuição da adoção de MD para a descoberta de conhecimentos novos e válidos.
DSS	Aproximação dos fatores de exposição associados ao desfecho com as camadas do modelo dos DSS.	Identificar a governabilidade e nível de intervenção necessário pelo Estado, de maneira a inferir a utilidade para a gestão, do conhecimento descoberto.

temas inteligentes, de engenharia de software e inteligência artificial etc.).

A aderência das associações à literatura foi considerada conforme o relatado pelos autores de cada estudo: corroboram a literatura, divergem da literatura ou interessantes frente à literatura. Tanto os resultados divergentes quanto os interessantes diferem da literatura, porém os primeiros foram considerados erros, justificados por possíveis vieses e limitações do estudo, enquanto os segundos foram vistos como surpreendentes, por haver plausibilidade para o achado, com sugestão de outros estudos. Um mesmo estudo pode ter encontrado diferentes associações.

A aproximação dos fatores de exposição às camadas do modelo dos DSS visou a inferir o potencial de utilidade do padrão descoberto para a gestão em saúde, considerando o pressuposto de que quanto mais externa for a camada do modelo dos DSS, maiores serão a governabilidade do Estado e a probabilidade de resultados efetivos⁸.

Assim, os estudos com potencial de contribuição à gestão em saúde materno-infantil incluíram aqueles que estavam alinhados ao objetivo do KDD10 de identificar padrões compreensíveis (critério para elegibilidade), válidos, novos (aderência à literatura) e potencialmente úteis (aproximação às camadas mais externas do modelo dos DSS).

RESULTADOS |

Foram identificados 834 documentos nos repositórios: PubMed (268), BVS (212), IEEE Xplore (111) e Scopus Elsevier (243). Trezentos e sessenta e seis foram excluídos por duplicidade e 468, avaliados por dois pesquisadores independentes na etapa de seleção. O valor de Kappa obtido foi de 0,884, sendo considerado uma concordância quase perfeita¹⁸. Vinte e cinco documentos foram avaliados por um terceiro pesquisador, pela divergência de indicação de seleção pelos dois primeiros.

A partir da Figura 1, é possível identificar o processo de seleção, inclusão e exclusão até a obtenção dos 29 documentos incluídos.

Os desfechos estudados variaram de uma doença específica (oito estudos), um fluxo ou processo de trabalho

na assistência (oito estudos) a situações de saúde mais abrangentes (13 estudos). A prematuridade foi o desfecho mais estudado (10,0%).

Dez estudos encontraram dentre seus achados, associações interessantes, consideradas com potencial de contribuição à gestão em saúde materno-infantil (34,5%). Dos 19 estudos restantes, dez apresentaram alguns resultados, porém não discutiam as associações encontradas à luz da literatura, focando a acurácia do modelo computacional utilizado (34,5%); oito encontraram somente resultados que corroboravam a literatura, não agregando novas descobertas (27,6%); e um estudo encontrou associações que corroboravam e diferiam da literatura (3,4%).

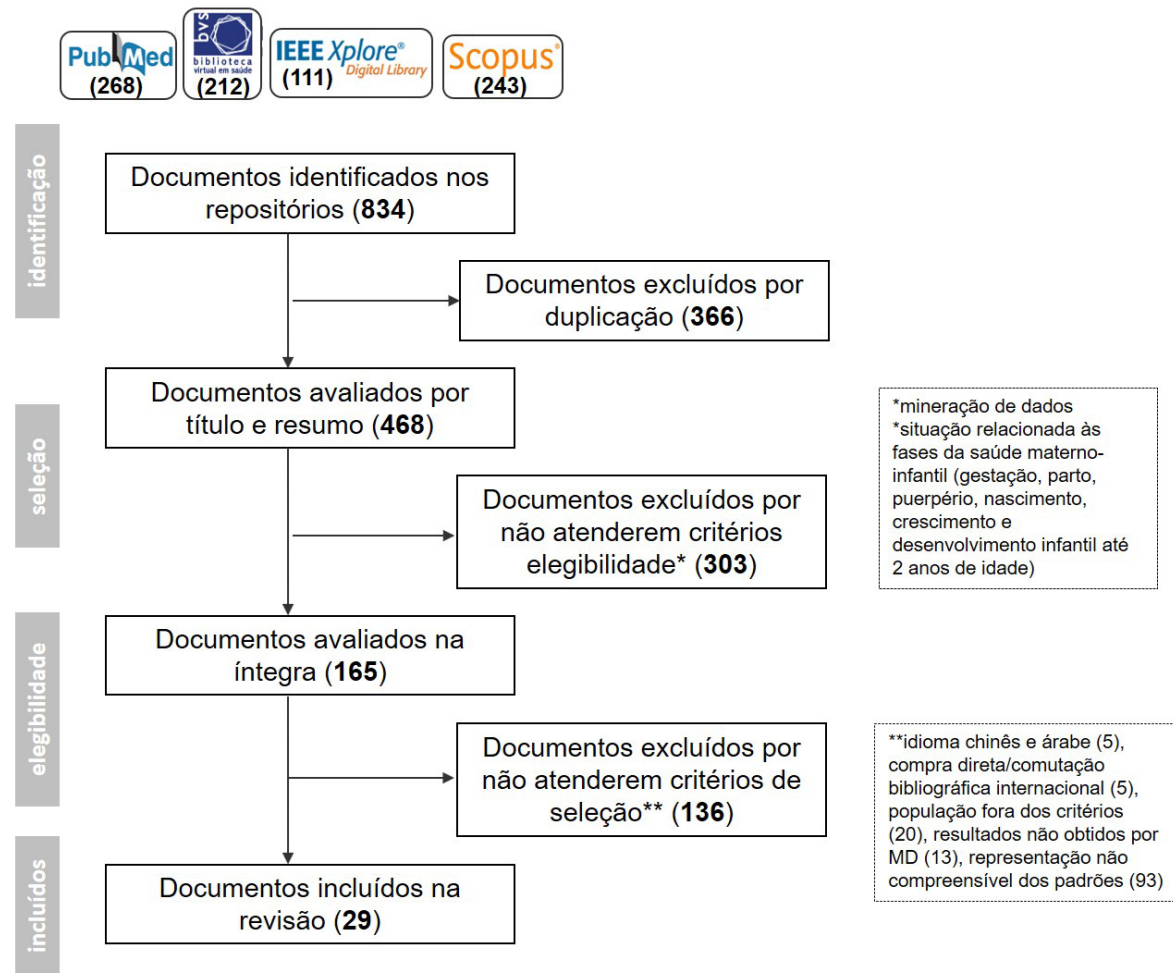
Dos dez estudos com potencial de contribuição à gestão, cinco foram publicados na área interdisciplinar (50,0%), quatro, na saúde (40,0%) e um, na informática (10,0%) – único documento com esse potencial, dentre os 11 publicados na área.

Com relação às bases de dados, a menor janela temporal foi de quatro meses e a maior, de 20 anos. Também houve variação com relação à quantidade de registros (menos de cem e acima de um milhão) e variáveis (menos de dez e acima de 1.500). Alguns estudos não informaram tais características – quatro não informaram a janela temporal, três, a quantidade de registros e 11, a quantidade de variáveis. Foram obtidas as seguintes medianas para cada característica da base de dados: cinco anos, 5.903 registros e 401 variáveis. Considerando apenas os dez estudos com potencial de contribuição à gestão, o número inferior e superior à mediana de cada característica foram próximos.

A aproximação dos fatores às camadas do modelo dos DSS8 demonstrou que dos dez estudos com potencial de contribuição à gestão, dois aproximavam-se da camada individual, e oito, das mais externas do modelo, de condições socioambientais gerais ou intermediárias, abrangendo situações de vida e trabalho, incluindo o acesso a serviços de saúde.

Todos os 29 estudos foram experimentos transversais e pontuais, não tendo sido evidenciada em nenhum a aplicação dos resultados obtidos pela MD. Os estudos incluídos nesta revisão e respectivas características extraídas são demonstrados no Quadro 2.

Figura 1 – Etapas da revisão



DISCUSSÃO |

A possibilidade de aplicação da MD em diferentes contextos nos primeiros 1.000 dias de vida foi identificada pela diversidade de desfechos estudados: doenças bucais³⁴, com forte componente ambiental^{23,28}; doenças crônicas transmissíveis³⁹, agudas¹⁴ ou raras⁴¹; falhas na assistência, tanto na atenção primária à saúde^{32-33,38} quanto na secundária ou terciária^{13,29}; situações de saúde como desnutrição infantil^{15,40}, mortalidade materna³⁵⁻³⁶, mortalidade de neonatos até crianças de cinco anos de idade^{22,24,42,43}, entre outros. A prematuridade foi o desfecho mais estudado^{19,21,25}, corroborando sua importância em relação aos óbitos neonatais. Estima-se que, em 2017, ocorreram 2,5 milhões de óbitos neonatais, dos quais dois terços eram prematuros⁴⁵.

O produto final de uma base de dados é o conhecimento⁴⁶. Ressalta-se que sua qualidade interfere no processo KDD e seus resultados¹⁰. Em alguns estudos, para minimizar essas interferências, pesquisadores recorreram a algumas estratégias como a imputação^{26,27,35}, exclusão de registros^{13,22,32,41} e de variáveis consideradas redundantes^{19,40}. Outros indícios da falta de qualidade nos bancos foram associações encontradas, porém relatadas como tendo possibilidade de viés devido à não completude dos registros²⁴.

Alguns estudos não relataram a quantidade de registros^{27,28,36}, de variáveis utilizadas^{13-14,27-29,31,33,38,39,44} ou a janela temporal das bases de dados^{30,34}. A falta dessas informações dificulta o dimensionamento do banco de dados, a reprodutibilidade do experimento e a possibilidade de agregação de novas variáveis em estudos futuros.

Quadro 2 – Estudos incluídos na revisão e características extraídas

Autor	Ano	Área de publicação	Desfecho	Base de dados	Resultados - aderência à literatura e aproximação à camada do modelo dos DSS
Goodwin et al. ¹⁹	2001	Saúde	Prematuridade	20 anos 19.970 registros 1.622 variáveis	CORROBORAÇÃO: individual, intermediário INTERESSANTE: individual
Silva et al. ²⁰	2007	Informática	Eczema infantil	12,5 anos 620 registros Quant. variáveis não informado	Sem análise/discussão frente à literatura, enfoque na acurácia do modelo proposto
Orozova-Bekkevold et al. ²¹	2007	Saúde	Prematuridade	6 anos 92.235 registros 281 variáveis	INTERESSANTE: intermediário
Vianna et al. ²²	2010	Saúde	Mortalidade infantil	5 anos 7.256 registros Quant. variáveis não informado	CORROBORAÇÃO: individual, intermediário
Loucoubar, et al. ²³	2011	Saúde	Malária	18 anos 46.837 registros 34 variáveis	Sem análise/discussão frente à literatura, enfoque na acurácia do modelo proposto
Matharage et al. ²⁴	2011	Informática	Natimortalidade	6 anos 215 registros Quant. variáveis não informado	CORROBORAÇÃO: individual INTERESSANTE: individual
Chen et al. ²⁵	2011	Informática	Prematuridade	4 anos 910 registros Quant. variáveis não informado	CORROBORAÇÃO: individual, estilo de vida, intermediário
Belle et al. ²⁶	2013	Informática	Aborto	Quant. anos não informado 1.435 registros Quant. variáveis não informado	Sem análise/discussão frente à literatura, enfoque na acurácia do modelo proposto
Kolyshkina et al. ²⁷	2013	Informática	Vulnerabilidade infantil	6 anos Quant. registros não informado 45 variáveis	Sem análise/discussão frente à literatura, enfoque na acurácia do modelo proposto
Ariyadasa et al. ¹⁶	2013	Informática	Desnutrição infantil	2 anos 6.714 registros Quant. variáveis não informado	Sem análise/discussão frente à literatura, enfoque na acurácia do modelo proposto
Zacarias et al. ²⁸	2013	Informática	Malária	3 anos Quant. registros não informado Quant. variáveis não informado	Sem análise/discussão frente à literatura, enfoque na acurácia do modelo proposto
Winald et al. ²⁹	2014	Interdisciplinar	Duração de estadia em emergência pediátrica	3 anos 55.183 registros Quant. variáveis não informado	CORROBORAÇÃO: intermediário

*continua.

*continuação.

Jalali et al. ³⁰	2014	Interdisciplinar	Leucomalácia periventricular após cirurgia cardíaca	Quant. anos não informado 44 registros 36 variáveis	DIVERGÊNCIA: individual INTERESSANTE: intermediário
Huff et al. ³¹	2014	Saúde	Resistência a tetraciclina	3,8 anos 80.241 registros Quant. variáveis não informado	CORROBORAÇÃO: individual, intermediário INTERESSANTE: sociais e ambientais gerais
Jawad et al. ³²	2015	Informática	Parto domiciliar	0,8 anos 5.248 registros 549 variáveis	Sem análise/discussão frente à literatura, enfoque na acurácia do modelo proposto
Le Meur et al. ³³	2015	Saúde	Pré-natal	1 ano 2.518 registros Quant. variáveis não informado	CORROBORAÇÃO: individual, intermediário INTERESSANTE: intermediário
Ivancevic et al. ³⁴	2015	Interdisciplinar	Cárie precoce da infância	Quant. anos não informado 341 registros 35 variáveis	CORROBORAÇÃO: individual, estilo de vida, intermediário INTERESSANTE: sociais e ambientais gerais
Sundararaman et al. ³⁵	2016	Informática	Mortalidade materna	7 anos 2.688 registros 27 variáveis	Sem análise/discussão frente à literatura, enfoque na acurácia do modelo proposto
Gupta et al. ³⁶	2016	Informática	Mortalidade materna	2 anos Quant. registros não informado 9 variáveis	CORROBORAÇÃO: individual, intermediário
Toti et al. ¹⁵	2016	Interdisciplinar	Asma	10 anos 20.959 registros Quant. variáveis não informado	CORROBORAÇÃO: ambientais gerais DIVERGÊNCIA: ambientais gerais INTERESSANTE: ambientais gerais
Mai et al. ³⁷	2016	Interdisciplinar	<i>Overtesting</i>	3 anos 1.685 registros 400 variáveis	Sem análise/discussão frente à literatura, enfoque na acurácia do modelo proposto
Sahle ³⁸	2016	Interdisciplinar	Consulta puerperal	1 ano 6.558 registros Quant. variáveis não informado	INTERESSANTE: intermediário, sociais e ambientais gerais
Elrazek et al. ³⁹	2016	Saúde	Hepatite C	1 ano 3.000 registros Quant. variáveis não informado	CORROBORAÇÃO: individual DIVERGÊNCIA: individual
Anilkumar et al. ⁴⁰	2017	Informática	Desnutrição infantil	2 anos 41.304 registros 401 variáveis	Sem análise/discussão frente à literatura, enfoque na acurácia do modelo proposto
Long et al. ⁴¹	2017	Saúde	Catarata pediátrica	Quant. anos não informado 160 registros 15 variáveis	CORROBORAÇÃO: individual
Tesfaye et al. ⁴²	2017	Interdisciplinar	Mortalidade na infância	1 ano 11.654 registros 21 variáveis	CORROBORAÇÃO: individual, estilo de vida, redes sociais, intermediário

*continua.

*continuação.

Sartorelli et al. ⁴³	2017	Saúde	Mortalidade infantil	5 anos 266 registros 15 variáveis	CORROBORAÇÃO: individual, intermediário
Hu et al. ¹⁴	2017	Interdisciplinar	Retorno à emergência pediátrica	12 anos 125.940 registros Quant. variáveis não informado	CORROBORAÇÃO: individual, intermediário DIVERGÊNCIA: individual INTERESSANTE: intermediário
Chen et al. ⁴⁴	2018	Interdisciplinar	Otimização de cuidados	0,3 anos 1.138.317 registros Quant. variáveis não informado	CORROBORAÇÃO: individual

*conclusão.

A conformação da base de dados e a demonstração dos resultados encontrados estão diretamente relacionadas à finalidade das atividades do KDD: desenvolvimento tecnológico (concepção, desenvolvimento, aprimoramento e otimização de algoritmos), execução (uso de algoritmos desenvolvidos na busca de conhecimentos) e aplicação de resultados (aplicação efetiva dos resultados obtidos)¹¹. Uma das dificuldades desta revisão foi comparar diferentes finalidades de atividades do KDD sob o ponto de vista de contribuição à gestão.

Para a atividade de desenvolvimento tecnológico, por exemplo, uma menor quantidade de variáveis diminui os parâmetros dos algoritmos e aumenta a eficiência computacional⁴⁷. Por exemplo, na previsão de prematuridade, a inclusão de centenas de variáveis às sete mais fortemente associadas aumentou a acurácia em apenas 0,03 na curva ROC¹⁹. Ainda, o estudo com a menor quantidade de registros foi considerado com potencial de contribuição à gestão, por ter encontrado um ponto de inflexão no tempo decorrido de cirurgia, a partir do qual há uma redução significativa na prevenção de leucomalácia periventricular³⁰.

Ao avaliar a quantidade de estudos com potencial de contribuição à gestão em relação às medianas de temporalidade da base de dados, de quantidade de registros e de variáveis, o número de estudos com valores inferiores ou superiores a elas ficou próximo, suscitando a hipótese de que o tamanho da base de dados não é condição indispensável, pois, quando a MD descobre conhecimentos potencialmente úteis para o apoio à tomada de decisão, a base de dados atendeu ao esperado, independentemente de seu tamanho⁴⁸.

Nesse sentido, dez documentos trouxeram resultados compreensíveis, válidos e novos (considerados, nesta revisão, como interessantes)^{13,14,19,21,24,30,31,33,34,38} e sua utilidade ficou evidenciada ao identificar que as camadas mais externas do modelo dos DSS foram as que mais tiveram aproximações. São essas camadas que influenciam e modulam as mais internas e onde se encontra a maior possibilidade de resultados efetivos, pois as ações nelas demandam promoção de saúde, políticas públicas e outras estratégias intersetoriais, de governabilidade única e exclusiva do Estado⁸.

Nove estudos encontraram associações que corroboraram e/ou divergiram da literatura, não agregando nenhum novo conhecimento. Considerando que os resultados da MD dependem dos dados¹⁹, a não descoberta de algo novo pode ser devido ao uso de bases de dados com variáveis cuja associação ao desfecho já está bem estabelecida⁴⁹. Para maior probabilidade de descoberta de conhecimento que possa auxiliar no planejamento e ação mais assertiva, é desejável, além das variáveis já conhecidas, outras novas, contemplando todas as camadas possíveis, aumentando a possibilidade de encontrar conhecimentos novos e úteis para o estabelecimento de hipóteses distintas²⁴.

Ressalta-se que, para a aderência à literatura, característica por meio da qual foram inferidas a validade e a novidade dos resultados, foi acatado o relatado pelos autores de cada estudo, o que pode ser considerado um viés importante, visto que os estudos são de áreas diferentes e possuem finalidades distintas, tendo em comum apenas o uso da MD. Contudo, a descoberta de resultados interessantes demonstra o cumprimento da finalidade da MD, de descoberta de conhecimento, fomentando o estabelecimento de novas

hipóteses, que, ao serem confirmadas, poderão auxiliar na tomada de decisão^{10,11}.

Para um resultado mais efetivo, a MD exige a interdisciplinaridade porque os problemas que se propõe a resolver não são da área computacional, sendo ela um meio para contribuir com a solução⁵⁰. Espera-se que os resultados encontrados sejam interpretados por especialistas na área para avaliação de sua utilidade^{14,20}. Detectou-se que os estudos publicados na área de informática foram os que mais pecaram nesse sentido visto que, em sua grande maioria, não foram considerados com potencial de contribuição à gestão^{15,20,25-28,32,35,36,40}, sugerindo que a falta de interdisciplinaridade interfere na utilidade da MD para a gestão.

CONCLUSÃO |

Variáveis redundantes e registros ausentes ou incompletos reforçam a necessidade da sensibilização contínua dos profissionais sobre a importância dos registros corretos, bem como do trabalho interdisciplinar no desenvolvimento dos sistemas, a fim de evitar redundância de variáveis.

Pela identificação da diversidade dos problemas de saúde abordados, obtenção de resultados considerados compreensíveis, novos, interessantes e úteis, com a aproximação das variáveis às camadas mais externas do modelo dos DSS, em que reside a maior governabilidade do Estado, conclui-se que a MD possui potencial de contribuição à gestão em saúde materno-infantil, principalmente quando a interdisciplinaridade se faz presente.

No entanto, notou-se que, dos resultados interessantes, a maioria dos desfechos estudados foi de doenças específicas (asma, cárie precoce na infância, leucomalácia periventricular etc.) ou de processos/serviços de saúde (consulta pré-natal, consulta puerperal, retorno à emergência pediátrica etc.), com poucos desfechos diretamente relacionados à garantia da vida. A ausência de mais resultados interessantes relacionados à mortalidade suscita a necessidade de discussão da inclusão de variáveis nos sistemas de informação em saúde além das já estabelecidas, o que possibilitaria novas descobertas de conhecimento.

Foi possível perceber a ausência de interdisciplinaridade principalmente por parte dos estudos publicados na área de informática, refletindo na obtenção de resultados sem potencial de contribuição à gestão. Ainda, identificou-se uma importante lacuna entre o desenvolvimento e uso da MD e a aplicação dos resultados obtidos, visto que todos os estudos foram ou de desenvolvimento tecnológico ou de execução, com experiências transversais e pontuais. A aproximação dessas áreas distintas gera benefícios tanto para a área-meio (computação) quanto para a área-fim (saúde), uma vez que os resultados produzidos pela computação são úteis e válidos ao agregar ao que já se conhece e ser efetivamente aplicáveis, podem contribuir na melhora da saúde da população.

AGRADECIMENTOS |

Agradecemos ao Programa de Suporte à Pós-Graduação de Instituições Comunitárias de Educação Superior (PROSUC/CAPES), pelo financiamento, modalidade taxa, para a realização.

REFERÊNCIAS |

1. Black RE, Allen LH, Bhutta ZA, Caulfield LE, Onis M, Ezzati M, et al. Maternal and child undernutrition: global and regional exposures and health consequences. *The Lancet*. 2008 [acesso em 26 nov 2018]; 371(9608):243-60. Disponível em: URL: <[https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736\(07\)61690-0.pdf](https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(07)61690-0.pdf)>.
2. Comissão para os Determinantes Sociais da Saúde [Internet]. Redução das desigualdades no período de uma geração: igualdade na saúde através da ação sobre os seus determinantes sociais (relatório final) [acesso em 01 jun 2018]. Disponível em: URL: <http://www.who.int/eportuguese/publications/Reducao_desigualdades_relatorio2010.pdf>.
3. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet*. 2013; 382(9890):427-51.

4. Cunha AJAD, Leite AJM, Almeida IS. The pediatrician's role in the first thousand days of the child: the pursuit of healthy nutrition and development. *J Pediatr.* 2015; 91(6 S1):S44-S51.
5. Victora GC, Horta BL, Mola CL, Quevedo L, Pinheiro RT, Gigante D, et al. Association between breastfeeding and intelligence, educational attainment, and income at 30 years of age: a prospective birth cohort study from Brazil. *The Lancet.* 2015; 3(4):e199-e205.
6. Mameli C, Mazzantini S, Zuccotti G. Nutrition in the first 1000 days: the origin of childhood obesity. *Int J Environ Res Public Health.* 2016; 13(9):838.
7. Moore TG, Arefadib N, Deery A, West S. The first thousand days: an evidence paper [Internet]. 2017 [acesso em 07 set 2018]. Disponível em: URL: <<https://www.suicideinfo.ca/resource/the-first-thousand-days-an-evidence-paper/>>.
8. Dahlgren G, Whitehead M. Policies and strategies to promote social equity in health. Stockholm: Institute For Future Studies; 1991.
9. Brasil. Ministério da Saúde. Política Nacional de Atenção Integral à Saúde da Criança: orientações para implementação [Internet]. Brasília: Ministério da Saúde; 2018 [acesso em 22 mar 2019]. Disponível em: URL: <http://www.saude.pr.gov.br/arquivos/File/Politica_Nacional_de_Atencao_Integral_a_Saude_da_Crianca_PNAISC.pdf>.
10. Fayyad U, Shapiro G, Smyth P. From data mining to knowledge discovery in data base. *AI Magazines.* 1996; 17(3):37-54.
11. Goldschmidt R, Passos E, Bezerra E. Data mining: conceitos, técnicas, algoritmos, orientações e aplicações. 2. ed. Rio de Janeiro: Elsevier; 2015.
12. Carvalho DR, Escobar LFA, Almeida DT. Pontos de atenção para o uso da mineração de dados da saúde. *Inf Inf.* 2014; 19(1):249-72.
13. Hu YH, Tai CT, Chen SCC, Lee HW, Sung SF. Predicting return visits to the emergency department for pediatric patients: applying supervised learning techniques to the Taiwan National Health Insurance Research Database. *Comput Methods Programs Biomed.* 2017; 144:105-112.
14. Toti G, Vilalta R, Lindner P, Lefer B, Macias C, Price D. Analysis of correlation between pediatric asthma exacerbation and exposure to pollutant mixtures with association rule mining. *Artif Intell Med.* 2016; 74:44-52.
15. Ariyadasa SN, Munasinghe KL, Senanayake HDS, Fernando MGNAS. Knowledge extraction to mitigate child malnutrition in developing countries (Sri Lankan context). In: Anais da 4. Conferência Internacional Sobre Sistemas Inteligentes, Modelagem e Simulação. 2013 jan 29-31; Bangkok, Tailândia. Nova Jersey: IEEE; 2013.
16. Witten IH, Frank E. Data mining practical machine learning tools and techniques. 2. ed. San Francisco: Elsevier; 2005.
17. Moher D, Liberati A, Tetzlaff J, Altman DG. The Prisma Group (2009) preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009; 6(7):e1000097.
18. Landis J, Koch GG. The measurements of agreement for categorical data. *Biometrics.* 1977; 33(1):159-74.
19. Goodwin LK, Iannacchione MA, Hammond WE, Crockett P, Maher S, Schlitz K. Data mining methods find demographic predictors of preterm birth. *Nurs Res.* 2001; 50(6):340-5.
20. Silva D, Alahakoon D, Dharmage S. Cluster analysis using the GSOM: patterns in epidemiology. In: Anais da 3. Conferência Internacional Sobre Informação e Automação para Sustentabilidade. 2007 dez 4-6; Melbourne, Austrália. Nova Jersey: IEEE; 2007.
21. Orozova-Bekkevold I, Jensen H, Stensballe L, Olsen J. Maternal vaccination and preterm birth: using data mining as a screening tool. *Pharm World Sci.* 2007; 29(3):205-12.
22. Vianna RC, Moro CM, Moysés SJ, Carvalho D, Nievola JC. Data mining and characteristics of infant mortality. *Cad Saúde Pública.* 2010; 26(3):535-42.
23. Loucoubar C, Paul R, Bar-Hen A, Huret A, Tall A, Sokhna C, et al. An exhaustive, non-euclidean, non-parametric data mining tool for unraveling the complexity of biological systems--novel insights into malaria. *PLoS One;* 2011; 6(9):e24085.

24. Matharage S, Alahakoon O, Alahakoon D, Kapurubandara S, Nayyar R, Mukherji M, et al. Analysing stillbirth data using dynamic self-organizing maps. In: Anais do 22. International Workshop on Database and Expert Systems Applications; 2011 29 ago/sep 2. Toulouse, França. Berlin: Springer; 2011.
25. Chen HY, Chuang CH, Yang YJ, Wu TP. Exploring the risk factors of preterm birth using data mining. *Expert Syst Appl.* 2011; 38(5):5384-7.
26. Belle V, Lisboa P. Automated selection of interaction effects in sparse kernel methods to predict pregnancy viability. In: Anais do 2013 IEEE Symposium on Computational Intelligence and Data Mining; 2013 abr 16-19; Singapura, Singapura. New Jersey: IEEE; 2013.
27. Kolyshkina I, Brownlow M, Taylor J. Improving every child's chance in life. In: Anais do 13. International Conference on Data Mining Workshops; 2013 dez 7-10; Dallas, Estados Unidos da América. New Jersey: IEEE; 2013.
28. Zacarias O, Bostrom H. Strengthening the health information system in Mozambique through malaria incidence prediction. In: Anais do 2013 IST-Africa Conference & Exhibition; 2013 maio 29-31; Nairóbi, Quênia. New Jersey: IEEE; 2013.
29. Windal F, Jeribi K, Ficheur G, Degoul S, Martinot A, Beuscart R, et al. Pediatric emergency department crowding: survival tree clustering for length of patient stay. *Stud Health Technol Inform.* 2014; 205:1095-9.
30. Jalali A, Buckley EM, Lynch JM, Schwab PJ, Licht DJ, Nataraj C. Prediction of periventricular leukomalacia occurrence in neonates after heart surgery. *IEEE J Biomed Health Inform.* 2014; 18(4):1453-60.
31. Huff MD, Weisman D, Adams J, Li S, Green J, Malone LL, et al. The frequency of tetracycline resistance genes co-detected with respiratory pathogens: a database mining study uncovering descriptive trends throughout the United States. *BMC Infect Dis.* 2014; 14:460.
32. Jawad F, Choudhury TUR, Najeeb A, Faisal M, Nusrat F, Shamita RC, et al. Data mining techniques to analyze the reason for home birth in Bangladesh. In: Anais do 16. International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing; 2016 junho 1-3; Takamatsu, Japan. New Jersey: IEEE; 2016.
33. Le Meur N, Gao F, Bayat S. Mining care trajectories using health administrative information systems: the use of state sequence analysis to assess disparities in prenatal care consumption. *BMC Health Serv Res;* 2015; 15:200.
34. Ivancevic V, Tusek I, Tusek J, Knezevic M, Elheshk S, Lukovic I. Using association rule mining to identify risk factors for early childhood caries. *Comput Methods Programs Biomed;* 2015; 122(2):175-81.
35. Sundararaman A, Ramanathan SV. A novel approach to estimate maternal mortality rate and its determinants by statistical modelling. In: Anais do 32. International Conference on Data Engineering Workshops; 2016 maio 16-20; Helsinki, Finlândia. New Jersey: IEEE; 2016.
36. Gupta S, Singh SN, Kumar D. An empirical analysis of maternal health data: a case study of India. In: Anais do 2. International Conference on Next Generation Computing Technologies; 2016 outubro 14-16; Dehradun, Índia. New Jersey: IEEE; 2016.
37. Mai MV, Krauthammer M. Controlling testing volume for respiratory viruses using machine learning and text mining. *AMIA Annu Symp Proc.* 2017; 2016:1910-9.
38. Sahle G. Ethiopian maternal care data mining: discovering the factors that affect postnatal care visit in Ethiopia. *Health Inf Sci Syst.* 2016; 4:4.
39. Elrazek A, Amer M, El-Hawary B, Salah A, Bhagavathula AS, Alborai M, et al. Prediction of HCV vertical transmission: what factors should be optimized using data mining/computational analysis. *Liver Int.* 2017; 37(4):529-33.
40. Anilkumar NA, Gupta D, Khare S, Gopalkrishna DM, Jyotishi A. Characteristics and causes of malnutrition across Indian states: a cluster analysis based on Indian demographic and health survey data. In: Anais do 2017 International Conference on Advances in Computing, Communications and Informatics. 2017 set 13-16. Udupi, India. New Jersey: IEEE; 2017.

41. Long E, Xu S, Liu Z, Wu X, Zhang X, Wang J, et al. Construction and implications of structural equation modeling network for pediatric cataract: a data mining research of rare diseases. *BMC Ophthalmol.* 2017; 17(1):74. Disponível em: URL: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/26431/pdf>>.
42. Tesfaye B, Atique S, Elias N, Dibaba L, Shabbir SA, Kebede M. Determinants and development of a web-based child mortality prediction model in resource-limited settings: a data mining approach. *Comput Methods Programs Biomed.* 2017; 140:45-51.
43. Sartorelli AP, Gomes DC, Cubas MR, Carvalho DR. Fatores que contribuem para a mortalidade infantil utilizando a mineração de dados. *Saúde e Pesqui.* 2017; 10(1):33-41.
44. Chen Y, Kho AN, Liebovitz D, Ivory C, Osmundson S, Bian J, et al. Learning bundled care opportunities from electronic medical records. *J Biomed Inform.* 2018; 77:1-10.
45. Organização Mundial da Saúde. Survive and thrive: transforming care for every small and sick newborn (key findings). Genebra: OMS; 2018 [acesso em 30 jan 2019]. Disponível em: URL: <<https://apps.who.int/iris/bitstream/handle/10665/276655/WHO-FWC-MCA-18.11-eng.pdf?ua=1>>.
46. Piatetsky-Shapiro G. Knowledge discovery in real databases: a report on the IJCAI-89 Workshop. *AI Magazines*; 1990 [acesso em 20 dez 2018]; 11(5):37-54. Disponível em: URL: <<https://www.aaai.org/ojs/index.php/aimagazine/article/view/873/791>>.
47. Liu H, Motoda H. Feature selection for knowledge discovery and data mining. Berlin: Springer; 2012.
48. Carvalho DR, Dallagassa M. Mineração de dados: aplicações, ferramentas, tipos de aprendizado e outros subtemas. *AtoZ.* 2014 [acesso em 19 mar 2019]; 3(2):82-86. Disponível em: URL: <<http://www.atoz.ufpr.br>>.
49. Sartorelli AP, Gomes DC, Cubas MR, Carvalho DR. Fatores que contribuem para mortalidade infantil utilizando a mineração de dados. *Rev Saude Pesq.* 2017; 10(1):33-41.
50. Neves BC, Braz MI. Interlocução entre saúde e ciência da informação: proposta para o diagrama multidisciplinar da CI. *Inf Inf.* 2018 [acesso em 28 mar 2019]; 23(3):100-21.

Correspondência para/Reprint request to:

Deborah Ribeiro Carvalho

Programa de Pós-Graduação em Tecnologia em Saúde,

Centro de Ciências da Saúde,

Pontifícia Universidade Católica do Paraná,

Rua Imaculada Conceição, 1155,

Prado Velho, Curitiba/PR, Brasil

CEP: 80215-901

E-mail: ribeiro.carvalho@pucpr.br

Recebido em: 26/09/2019

Aceito em: 24/10/2019