

Evidencia do mundo real para identificação e tratamento dos pacientes com COVID-19 no Brasil

Coordenador: Fredi Alexander Diaz Quijano – Docente do Departamento de Epidemiologia da Faculdade de Saúde Pública (FSP) da USP e Alexandre Dias Chiavegatto Filho - Docente do Departamento de Epidemiologia da Faculdade de Saúde Pública (FSP) da USP.

Resumo:

O projeto compreende três componentes. O primeiro visa obter e validar um modelo preditivo de diagnóstico de COVID-19 baseado em dados de vigilância em saúde pública do Brasil. O segundo componente foca em pacientes hospitalizados e pretende otimizar o uso de informações rotineiramente coletadas nos serviços de atenção para prever, além do diagnóstico, desfechos de interesse. Finalmente, um terceiro componente envolverá a revisão crítica da literatura sobre ensaios pragmáticos que avaliem intervenções para o tratamento de COVID-19. A avaliação desses estudos envolverá aspectos como a qualidade do reporte, risco de vieses, validação de desfechos compostos e a heterogeneidade dos efeitos.

Componente 1: Predição de COVID-19 entre casos notificados no Brasil. Trata-se de uma pesquisa de avaliação de tecnologias de diagnóstico, aninhada em dados de vigilância epidemiológica obtidos pelo Ministério da Saúde do Brasil. População do estudo: Envolverá todos os casos notificados como suspeitos de COVID-19.

Dados e Análise Estatística Serão analisadas as variáveis incluídas na ficha de notificação da COVID-19 incluindo demográficas, manifestações clínicas e antecedentes de comorbidades e exposições específicas (contato com caso suspeito ou confirmado, viagens). Atualmente já realizamos uma análise com dados dos primeiros meses de vigilância, cujos resultados foram postados como preprint (<https://www.medrxiv.org/content/10.1101/2020.04.05.20047944v1>), onde se descrevem detalhes da metodologia desse componente do projeto. Dessa primeira análise salientamos que obtivemos uma função preditiva com dados de 80% da amostra de São Paulo e Rio de Janeiro. Esse modelo, baseado unicamente em dados de vigilância, teve uma área abaixo da curva ROC de 95% e uma acurácia de 89.47% para prever o diagnóstico de COVID-19 em uma amostra de validação (o outro 20% desses estados) entre pacientes notificados como suspeitos ao sistema de vigilância. Pretende-se ampliar a amostra obter um modelo robusto que seja validado no tempo, incluindo informação das diferentes regiões do Brasil.

Componente 2: Inteligência artificial para decisões clínicas durante a pandemia de COVID-19. Refere-se a um estudo retrospectivo de pacientes atendidos nos hospitais participantes da rede com suspeita de COVID-19. O período analisado será desde a primeira realização de um exame de COVID-19 a partir de 17 de março de 2020 até o dado mais recente disponível no momento da aplicação dos algoritmos. Todos os dados identificadores dos pacientes serão excluídos antes do recebimento dos dados, seguindo as boas práticas adotadas na instituição. As variáveis preditoras para treinar os algoritmos serão todas aquelas coletadas rotineiramente pelos hospitais e disponíveis para análise, principalmente os resultados do hemograma completo (principalmente leucócitos, eosinófilos, basófilos, linfócitos, monócitos, plaquetas, proteína C-reativa, hemácias, hemoglobinas, CHCM, HCM, RDW e VCM), sexo e idade. Caso seja possível, também serão incluídas outras variáveis como sinais vitais e data de início dos sintomas. A performance preditiva dos algoritmos será medida por meio da sensibilidade, especificidade e área abaixo da curva ROC, nos dados de teste. Os hiperparâmetros dos algoritmos serão ajustados por meio de validação cruzada. As

variáveis contínuas serão padronizadas por meio do z-score e as variáveis categóricas serão transformadas em dummies. Serão testadas as performances dos algoritmos mais populares de machine learning para dados estruturados, como redes neurais, random forests support vector machines e gradient boosting trees. Serão incluídos nas análises todos os pacientes que seguiram protocolo de atendimento para casos suspeitos de COVID-19. O pesquisador associado (ADPC) desse projeto já postou um pré-print com os resultados iniciais: <https://www.medrxiv.org/content/10.1101/2020.04.04.20052092v2>. Esse foi produto de uma parceria com o Hospital Israelita Albert Einstein e teve como objetivo utilizar os dados preliminares de pacientes que realizaram o exame RT-PCR de COVID-19 para prever risco de diagnóstico positivo da doença a partir de resultados coletados rotineiramente pelo pronto-socorro do hospital. Todos os algoritmos de machine learning testados apresentaram performance semelhante nos resultados de teste, com área abaixo da curva ROC acima de 0,84. O algoritmo com melhor performance para esses dados foi o support vector machines com área abaixo da curva ROC de 0,85, sensibilidade de 0,68, especificidade de 0,85 e Brier Score de 0,16. A calibração do algoritmo foi boa ao longo de toda a distribuição de probabilidades. Por exemplo, entre os pacientes com probabilidade predita pelo algoritmo de 80 a 100% de terem COVID-19, 82% de fato tinham COVID-19, enquanto entre aqueles com probabilidade entre 0 e 20%, apenas 12% tinham a doença. Dando continuidade, o presente estudo irá ampliar a análise e validação dos algoritmos para outras regiões brasileiras com diferentes contextos socioeconômicos e para outros desfechos relacionados à COVID-19, como a predição de risco de internação em UTI, de uso de ventilação mecânica e de óbito.

Componente 3: Avaliação de ensaios pragmáticos de intervenções para COVID-19. Será realizada uma avaliação prospectiva dos ensaios clínicos pragmáticos de intervenções para COVID-19. Inicialmente uma busca nas bases de dados de protocolos de pesquisa, [clinicaltrials.gov](https://www.clinicaltrials.gov) e base de dados da OMS, International Clinical Trials Registry Platform (ICTRP) será realizada com o objetivo de identificar estudos utilizando dados de mundo real, combinando os termos sinônimos para “pragmatic clinical trial” e “COVID-19”. Os estudos encontrados em um primeiro momento serão avaliados para verificar se a metodologia se confirma como um desenho pragmático. A evolução dos estudos e publicação dos estudos será acompanhada prospectivamente e os dados dos estudos publicados irão compor a principal população deste estudo. A avaliação da qualidade do relato dos estudos será avaliada por meio da ferramenta RECORD-PE. Os domínios serão caracterizados como adequados ou inadequados. A qualidade metodológica dos estudos incluídos será avaliada por meio das ferramentas validadas na literatura para cada tipo de estudo.

Todos os desfechos serão classificados conforme a iniciativa COMET preconiza. Estudos que tiverem desfechos compostos e/ou intermediários como desfechos primários serão avaliados testando o index BACO (<https://www.medrxiv.org/content/10.1101/2020.02.13.20020966v1>). Uma análise descritiva será realizada com os dados obtidos da extração de dados. Análise incluirá avaliação da correlação entre as variações no tamanho do efeito com as limitações metodológicas encontradas. No final, espera-se obter um levantamento das características metodológicas de dois estudos e possíveis causas mais frequentes de viés. Esses resultados apoiarão recomendações para a interpretação de ensaios clínicos pragmáticos aplicados ao COVID-19.

Auxílio financeiro: Ainda não conta com auxílio financeiro. Foi submetido ao edital 09/2020 da CAPES.

Contato: frediazq@msn.com, frediazq@usp.br, e alexdiasporto@usp.br

