



**MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE CIÊNCIAS DA SAÚDE DE PORTO ALEGRE
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO
E GESTÃO EM SAÚDE**

Uilian Loose

**Framework para Análise de Riscos de Modelos para Predição da
Propagação da Covid-19**

Porto Alegre
2021

Uilian Loose

**Framework para Análise de Riscos de Modelos para Predição da
Propagação da Covid-19**

Linha de pesquisa: Sistemas Inteligentes e Aplicações na Saúde

Dissertação

Dissertação submetida ao Mestrado Acadêmico do Programa de Pós-graduação em Tecnologias da Informação e Gestão em Saúde da Universidade Federal de Ciências da Saúde de Porto Alegre, como requisito parcial à obtenção do título de Mestre.

Orientadora: Prof^a. Dr^a. Juliana Silva Herbert

Coorientadores: Prof^a. Dr^a. Ana Beatriz G. da Veiga
Prof. Dr. Marcelo F. da Costa Gomes

Porto Alegre
2021

Uilian Loose

Framework para Análise de Riscos de Modelos para Predição da Propagação da Covid-19

Dissertação submetida ao Mestrado Acadêmico do Programa de Pós-graduação em Tecnologias da Informação e Gestão em Saúde da Universidade Federal de Ciências da Saúde de Porto Alegre, como requisito parcial à obtenção do título de Mestre.

Orientador: Prof^a. Dr^a. Juliana Silva Herbert

Coorientadores: Prof^a. Dr^a. Ana Beatriz G. da Veiga
Prof. Dr. Marcelo F. da Costa Gomes

Aprovada em: _____ de _____ de _____.

BANCA EXAMINADORA:

Prof. Dr. Oswaldo Gonçalves Cruz.
Fundação Oswaldo Cruz

Prof. Dr. Paulo Gilberto Cogo Leivas
Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)

Profa. Dra. Maria Eugênia Bresolin Pinto
Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)

©2021

UNIVERSIDADE FEDERAL DE CIÊNCIAS DA SAÚDE DE PORTO ALEGRE
Praça Sarmiento Leite, 245 – Centro Histórico
Porto Alegre – RS CEP: 90050-170

Os conceitos expressos neste trabalho são de responsabilidade do autor e da orientadora.

Dedico este trabalho aos meus filhos Édson, Elisiane e Thiago e ao meu grande parceiro de vida, Charles. Aos meus filhos, porque, assim como a educação, é neles que ancore minha esperança e, ao meu grande amor, pelo cuidado conosco, pela inspiração, coragem e parceria em todas as horas. Vocês tornaram essa jornada mais leve. Hoje não foi um dia fácil, meu corpo está refletindo a exaustão da mente. Adormeci febril e acordei com ele segurando a minha mão. Nosso silêncio explica nosso amor e cumplicidade e nossos gestos gritam para o mundo que, com dedicação, fé e amor, **a gente pode tudo**. Por Você!

AGRADECIMENTOS

Citarei aqui apenas representantes de tantas pessoas que transformaram minha existência em algum momento dessa etapa trilhada. Já adianto: minha dificuldade com certeza está em ser breve sem perder a intensidade e sinceridade.

À minha família, por estar, em todo o tempo, ao meu lado. Por doar os finais de semana e muitas noites para a concretização dessa jornada que é nossa. Este trabalho amadureceu e criou forma no melhor momento das nossas vidas, quando Charles e eu ficamos completos com a chegada dos nossos amores Édson, Lisie e Thiago.

Obrigado, Dra. Bibiana Boaventura, minha sócia, por todo o suporte, pelo apoio e pela compreensão nas ausências. Nossa trajetória acadêmica é marcada por inspirações, decepções, dúvidas e certezas, e ter ao nosso lado pessoas que acreditem que chegaremos ao fim faz toda a diferença. Aos meus colegas de projeto, Andreas, Aline e Graziela, obrigado pela parceria.

Meu agradecimento especial aos meus coorientadores Ana Veiga e Marcelo Gomes: é paradoxal o sentimento de extrema honra e medo de não suprir toda a responsabilidade.

Feliz daquele que consegue encontrar no seu mestre exatamente a parcela do mundo que se idealiza. Obrigado professora Juliana, pela melhor forma de ser mestre, pelo grande exemplo de pessoa e de modo tão simples, com atitudes grandes, sempre ao lado, ser o melhor exemplo de líder. Desejo ter muito de você, alcançar quem precisa com ternura, com simplicidade, ética, flexibilidade, com paixão e seriedade. Que honra a minha ter o teu nome no meu *lattes*.

Os meus Mestres, muito obrigado.

RESUMO

O surgimento repentino da Covid-19, a transmissão descontrolada da doença e os desafios enfrentados pela falta de acesso de dados em saúde para o desenvolvimento de pesquisas reforçaram os riscos assumidos pelos pesquisadores no desenvolvimento de trabalhos em saúde. Diante desse desafio, este trabalho tem como objetivo o desenvolvimento de um framework para análise de riscos de modelos de predição da propagação da Covid-19 no contexto brasileiro. Para chegar ao resultado, foram realizadas revisões narrativas para entender o estado da arte da doença, da estrutura de vigilância epidemiológica no Brasil, dos modelos preditivos utilizados e dos fundamentos jurídicos para pesquisa e acesso à informação em saúde. Depois disso, uma revisão sistemática selecionou 30 modelos para predição da propagação da Covid-19 que foram avaliados, identificando-se então critérios de risco. Esses critérios foram estruturados em uma estrutura análoga a uma árvore de decisão, a qual foi submetida à avaliação de um grupo focal composto por equipe multidisciplinar. O resultado é um framework para análise de riscos de modelos para predição da propagação da Covid-19 no contexto brasileiro.

Palavras-Chave: Covid-19, SARS-CoV-2, modelo preditivo, análise de riscos, Lei Geral de Proteção de Dados, segurança da informação, base de dados.

ABSTRACT

The sudden appearance of Covid-19, the uncontrolled transmission of the disease and the challenges faced by the lack of access to health data for research development reinforced the risks assumed by researchers in the development of health work. Faced with this challenge, this work aims to develop a framework for risk analysis of Covid-19 propagation prediction models in the Brazilian context. To reach the result, narrative reviews were carried out to understand the state of the art of the disease, the structure of epidemiological surveillance in Brazil, the predictive models used and the legal foundations for research and access to health information. Afterwards, a systematic review selected 30 models to predict the spread of Covid-19 that were evaluated, and risk criteria were then identified. These criteria were structured in a structure analogous to a decision tree, which was submitted to evaluation by a focus group composed of a multidisciplinary team. The result is a framework for risk analysis of models to predict the spread of Covid-19 in the Brazilian context.

Keywords: Covid-19, SARS-CoV-2, predictive model, risk analysis, General Data Protection Law, information security, database.

LISTA DE TABELAS

<i>Tabela 1 - Base de Pesquisa.</i>	56
--	-----------

LISTA DE FIGURAS

Figura 1 - Fluxograma de avaliação de casos e monitoramento de contatos. _____	23
Figura 2 - Painel Monitoramento de Dados Abertos do Brasil. _____	27
Figura 3 - Ecossistema de atendimento e de dados em saúde. _____	29
Figura 4 - Hierarquia das Normas - Desenvolvido pelo autor. _____	35
Figura 5 - Fluxo da metodologia de pesquisa. _____	52
Figura 6 - Áreas de estudo da Revisão Sistemática. _____	55
Figura 7 - Orquestração de atividades ao longo de funções de negócio. _____	68
Figura 8 - Módulo Processo. _____	70
Figura 9 - Manual BPMN desenvolvido para o mapeamento de processos da UFCSPA. _____	71
Figura 10 - Fase 1 do Módulo Processo. _____	73
Figura 11 - Bloco 1 da Fase 1. _____	74
Figura 12 - Fase 2 do Módulo Análise. _____	76
Figura 13 - Bloco 2 da Fase 2 do Módulo de Análise. _____	77
Figura 14 - Bloco 3 da Fase 2 do Módulo de Análise. _____	78
Figura 15 - Bloco 4 da Fase 2 do Módulo de Análise. _____	78
Figura 16 - Fase 3 do Módulo Análise. _____	79
Figura 17 - Bloco único da Fase 3 do Módulo de Análise. _____	81
Figura 18 - Bloco 1 da Fase 1 do Módulo Relatório de Riscos. _____	82
Figura 19 - Capa do Guia do Framework de Análise de Riscos de Modelos Preditivos. _____	83

LISTA DE ABREVIATURAS E SIGLAS

ABRASCO	: Associação Brasileira de Saúde Coletiva
ANAC	: Agência Nacional de Aviação Civil
BPM	: <i>Business Process Management</i>
CDC	: Código de Defesa do Consumidor
CE	: Critérios de Exclusão
CEP	: Comissão de Ética em Pesquisa
CF	: Constituição Federal
CNJ	: Conselho Nacional de Justiça
CNS	: Conselho Nacional de Saúde
CPF	: Cadastro de Pessoa Física
CTI	: Centro de Terapia Intensiva
HIV	: <i>Human Immunodeficiency Virus</i>
LGPD	: Lei Geral de Proteção de Dados
LINDB	: Lei de Introdução às normas do Direito Brasileiro
MS	: Ministério da Saúde
OMS	: Organização Mundial de Saúde
OPAS	: Organização Pan-Americana de Saúde
RAS	: Rede de Atenção à Saúde
RNDS	: Rede Nacional de Dados em Saúde
SIM	: Sistema de Informação sobre Mortalidade
SINASC	: Sistema de Informações sobre Nascidos Vivos
STF	: Supremo Tribunal Federal
STJ	: Superior Tribunal de Justiça
SUS	: Sistema Único de Saúde
TCLE	: Termo de Consentimento Livre e Esclarecido
UFCSPA	: Universidade Federal de Ciências da Saúde de Porto Alegre
UNB	: Universidade de Brasília
USP	: Universidade de São Paulo
UTI	: Unidade de Terapia Intensiva

SUMÁRIO

AGRADECIMENTOS	4
1 INTRODUÇÃO	11
2 PROBLEMA E JUSTIFICATIVA	14
3 OBJETIVO	17
3.1 OBJETIVOS ESPECÍFICOS	17
4 TRABALHOS RELACIONADOS	19
5 REFERENCIAL TEÓRICO	21
5.1 COVID-19	21
5.1.1 <i>Vigilância Epidemiológica</i>	24
5.2 BASES DE DADOS EM SAÚDE	26
5.2.1 <i>RNDS</i>	28
5.2.2 <i>Big Data</i>	30
5.3 MODELOS PREDITIVOS	31
5.4 RISCOS E ASPECTOS LEGAIS	32
5.4.1 <i>Contextualização</i>	32
5.4.2 <i>Aspectos Jurídicos</i>	34
6 MÉTODO	52
6.1 REVISÃO NARRATIVA	53
6.2 REVISÃO SISTEMÁTICA	53
6.3 DESENVOLVIMENTO DO FRAMEWORK	59
6.4 GRUPO FOCAL	60
7 RESULTADOS E DISCUSSÕES	63
7.1 O FRAMEWORK	63
7.1.1 <i>Planejamento do Framework</i>	65
7.1.2 <i>Desenvolvimento do Framework</i>	67
7.2 AVALIAÇÃO DO FRAMEWORK	83
8 CONSIDERAÇÕES FINAIS	85
9 REFERÊNCIAS	87
10 APÊNDICE A - ARTIGO DA REVISÃO SISTEMÁTICA	95
11 APÊNDICE B - CAPÍTULO DE LIVRO	111
12 APÊNDICE C - O FRAMEWORK	116

1 INTRODUÇÃO

O vírus denominado SARS-CoV-2 identificado na cidade de Wuhan, China, no final de 2019 [1] ressaltou os desafios enfrentados pelos cientistas no desenvolvimento de soluções em saúde durante grave crise de saúde global: a falta de informações disponíveis sobre o assunto, a alta transmissibilidade da doença e a instabilidade jurídica com objetivo de defender os pesquisadores no enfrentamento da doença. Estes desafios motivaram o trabalho de pesquisa aqui apresentado.

Pesquisadores das mais diversas áreas do conhecimento, dos quatro cantos do mundo, se reuniram com o mesmo propósito [2]: combater a doença.

A rápida propagação da doença, as suas consequências devastadoras e os limites geográficos existentes em pouco tempo exigiram esforços da comunidade científica para o desenvolvimento de ferramentas capazes de auxiliar na tomada de decisões [2] [3]. Os modelos para predição da propagação da Covid-19 passaram a ser grandes aliados no enfrentamento da doença. No entanto, a velocidade de propagação do vírus é muito maior que a capacidade de produção de conhecimento, o que fez com que os cientistas, em especial no Brasil, se deparassem com dois problemas, a disponibilidade e a qualidade de dados de saúde existentes para análises preditivas [4] e os riscos legais existentes considerando a recente mudança legislativa sobre a proteção de dados pessoais.

Um dos desafios enfrentados pelos pesquisadores desenvolvedores de modelos preditivos na área da saúde é justamente com a população usada para seus estudos e a confiabilidade dos dados existentes em um cenário tão inconstante como se apresentou durante a pandemia da Covid-19 [5]. Esses dados podem ser inexistentes, inconsistentes, não rastreáveis, não acessíveis ou conflitantes com a legislação vigente, fazendo com que sejam desenvolvidos estudos com as informações possíveis e não com as informações ideais, limitando, muitas vezes, os resultados obtidos e atrapalhando no enfrentamento à doença.

A disponibilidade de dados em saúde no Brasil é um problema histórico, sobretudo por causa das informações fragmentadas resultantes dos diferentes órgãos que atuam no setor [6]. Lacunas na concepção dos dados de saúde muitas vezes impedem que estes dados sejam disponibilizados porque, na origem, há problemas na coleta e no preenchimento dos dados, manuseio de formulários e sistemas, capacidade de tomada de decisões e falta de padrões para a notificação de dados essenciais [7].

Os dados já existentes e disponíveis podem comprometer os estudos ou por serem utilizados de forma ou desagregada ou agregada de forma errada pelo processamento dos dados realizados pelos órgãos de gestão competentes por aquela base de dados [8]. Além disso, deve-se considerar que, muitas vezes, os dados sequer estão acessíveis à pesquisa.

Com o advento de normas protetoras de dados pessoais, a insegurança para o desenvolvimento de pesquisas em saúde corrobora com a dificuldade de pesquisa no Brasil. Nos últimos anos, conflitos com a legislação vigente passaram a marcar a vida dos pesquisadores, sobretudo pelas novidades normativas que dizem respeito ao direito de personalidade e ao direito/dever de sigilo.

Há evidente preocupação com o armazenamento dos dados individuais e a garantia dos direitos personalíssimos dos indivíduos. Por outro lado, o interesse coletivo exige acesso às informações para trazer respostas aos problemas comuns enfrentados no decorrer de uma situação de crise, como é o caso da pandemia da Covid-19. Em outras situações, o problema insurge com relação à confiabilidade dos dados existentes. Isto é, os dados existem, mas, por falha de planejamento, não podem ser utilizados por imprecisão, incoerência, incompatibilidade com a realidade ou com impossibilidade de rastreabilidade da origem do dado.

Os estudos e análises realizados no contexto da Covid-19 são exemplos relacionados aos problemas anteriormente apresentados, pois a urgência por respostas e a busca por soluções eficientes esbarram em vários momentos nas limitações dos dados capazes de gerar informações que contribuam para o

enfrentamento dessa pandemia ou na insegurança jurídica a que o pesquisador é exposto.

A pandemia da Covid-19 mostrou que o ideal, muitas vezes, está longe de ser possível, sobretudo no Brasil, considerando as suas características geográficas e econômicas. Entretanto, é necessário olhar para esse problema buscando alternativas possíveis para que, a médio e longo prazo, seja possível utilizar um histórico de dados adequado e segurança jurídica aos pesquisadores que, incansavelmente, trabalham para o desenvolvimento de novos conhecimentos. Por isso, esse estudo propõe o desenvolvimento de um framework para análise de riscos de modelos para predição da propagação da Covid-19.

2 PROBLEMA E JUSTIFICATIVA

A partir da década de 30, o Brasil passou pela quebra das autonomias estaduais e consequente centralização do poder, com a criação de vários órgãos administrativos com caráter regulador. Paradoxalmente, houve a descentralização da gestão de temáticas básicas, como a saúde, com a “criação de órgãos administrativos de caráter regulador atingindo as mais diferentes dimensões da vida social”, conforme destacam Moraes e Santos (1998).

Até a Constituição de 1988, marco inaugural do direito à saúde e obrigação do Estado, era o texto da Constituição de 1934 que tratava de garantir tratamento ao tema da saúde, deliberando essa competência para a União e os Estados federados¹.

O destaque à temática como obrigação do Estado e a descentralização do gerenciamento da saúde desencadearam a falta de planejamento e prioridades por parte da Gestão Pública. As necessidades e prioridades da atenção à saúde, tratando-se de um país com as dimensões do Brasil, não são as mesmas em todo o território, o que demanda estratégias especiais nas diferentes regiões do país considerando as suas respectivas peculiaridades. Dessa forma, o mapeamento, levantamento e gerenciamento de dados em saúde devem seguir uma estratégia unificada com referência as suas peculiaridades.

A década de 90 foi marcada pela ascensão da Internet e pelo desenvolvimento das tecnologias de informação [10], auxiliando na coleta, no armazenamento e gerenciamento de dados em saúde. Há muitos aspectos positivos nesse desenvolvimento, mas o rápido desenvolvimento

¹ Art.10. Compete concorrentemente à União e aos Estados: (...)

II - cuidar da saúde e assistência públicas;

Art. 121. A lei promoverá o amparo da produção e estabelecerá as condições do trabalho, na cidade e nos campos, tendo em vista a proteção social do trabalhador e os interesses econômicos do País.

§1.º A legislação do trabalho observará os seguintes preceitos, além de outros que colimem melhorar as condições do trabalhador: (...)

h) assistência médica e sanitária ao trabalhador e à gestante, assegurando a esta descanso antes e depois do parto, sem prejuízo do salário e do emprego, e instituição de previdência, mediante contribuição igual da União, do empregador e do empregado, a favor da velhice, da invalidez, da maternidade e nos casos de acidentes de trabalho ou de morte.

descentralizado pode ser um problema estruturado que pode estar mascarado. Verifica-se que a captura e o gerenciamento das informações de saúde passaram a ser mais importantes que a qualidade e direcionalidade dos dados armazenados [9]. Ou seja, de nada adianta ter os dados se eles não são confiáveis ou úteis para a gestão da saúde.

Moraes e Santos (1998) dizem que os dados de saúde passaram a ser meras justificativas para a contratação de sistemas cada vez mais robustos, já que essas tecnologias seriam capazes de auxiliar na gestão da saúde, como se os sistemas, por si só, fossem capazes de resolver os problemas de saúde. Ter dados fragmentados, descentralizados e descontextualizados apenas gera custos de armazenamento pois perdem a capacidade de gerar informações de valor.

No Brasil, as unidades de saúde públicas possuem prontuários médicos à sua própria maneira: algumas digitalizam, outras possuem seus prontuários em papel. Identificam-se nesse cenário problemas como a perda de completude das informações para a assistência adequada ao paciente: quando o paciente pode ser atendido por várias unidades de saúde, isso quer dizer que o prontuário médico deixa de ser do paciente e passa a ser da unidade de saúde, prejudicando não só o tratamento daquele paciente como as informações para gestão da saúde daquela região como dados para uso científico [11].

Um dos desafios no desenvolvimento de modelos matemáticos preditivos está relacionado justamente à disponibilidade de dados disponíveis para pesquisa [5] porque os bancos de dados relacionados à saúde possuem limitações que interferem significativamente na utilização dos dados disponíveis para o desenvolvimento científico. Além disso, é um grande desafio desenvolver soluções e tecnologias de maneira rápida, eficiente numa área desconhecida, justamente em meio à crise de saúde, com sérias limitações, como no caso da pandemia da Covid-19, por exemplo.

Em paralelo às limitações citadas, em meio à pandemia, a Lei Geral de Proteção de Dados – LGPD entra em vigor no Brasil com objetivo de regulamentar a proteção de dados pessoais².

O pesquisador, além de, na prática, não ter disponíveis os dados necessários para o desenvolvimento ideal de pesquisa, enfrenta o desafio contínuo de não infringir nenhuma norma legal que coloque seu trabalho ou a si próprio em risco.

Com o objetivo de enfrentar esses desafios, em especial neste momento especial da pandemia da Covid-19, propõe-se o desenvolvimento de um framework que poderá ser utilizado para auxiliar os desenvolvedores de modelos preditivos na análise de risco dos dados, porque, além da dificuldade de produzir ciência durante uma pandemia, por conta da situação de emergência, das incertezas e das ameaças existentes [12], a falta de planejamento para a produção e disponibilização de dados válidos e a dificuldade de acesso aos dados já existentes adicionadas ao risco de infringir eventual regulamentação legal são preocupantes.

Moraes e Santos (1998) dizem que “estamos consumindo sistemas de informação, mas não estamos colhendo resultados”. Ou seja, não adianta termos disponíveis as melhores tecnologias de gestão e armazenamento de dados se elas não podem ser utilizadas para gerar informação adequada. Os autores ainda apontam que, atualmente, a produção de dados de saúde está mais vinculada aos interesses políticos do que aos interesses sociais.

Por isso, um framework que auxiliará na análise de riscos dos dados utilizados nos modelos para predição da propagação da Covid-19 à disposição dos pesquisadores, além de ser uma ferramenta de grande importância nessa corrida contra o tempo, poderá ser utilizado para novos desafios que deverão ser enfrentados pelos pesquisadores brasileiros no futuro.

² Lei nº 13.709, de 14 de agosto de 2018 alterada pela Lei nº 13.853, de 8 de julho de 2019.

3 OBJETIVO

O objetivo desta pesquisa é desenvolver um framework para análise de riscos de modelos preditivos para prevenção da propagação da Covid-19 no contexto brasileiro.

3.1 OBJETIVOS ESPECÍFICOS

Para alcançar o objetivo geral da pesquisa, foram propostos os seguintes objetivos específicos:

- a) selecionar modelos de predição da propagação da Covid-19 já publicados:
 - definir critérios para a seleção desses modelos;
 - desenvolver uma revisão sistemática com base nos critérios estabelecidos para não haver risco de vieses na seleção dos modelos;
- b) analisar os modelos selecionados de forma a definir fatores comuns que contribuam para a avaliação da adequação dos dados de entrada, adequação das bases de dados utilizadas por cada modelo e os riscos envolvidos:
 - identificar dados de entrada dos modelos previamente selecionados;
 - identificar os fatores que precisam ser avaliados nos modelos para minimizar riscos;
 - analisar como os riscos que os dados utilizados nos modelos selecionados podem trazer aos desenvolvedores;
 - estabelecer sugestões, com base em fontes jurídicas para minimizar os riscos que os dados utilizados nos modelos selecionados podem trazer aos desenvolvedores;
- c) desenvolver um framework de análise de riscos:

- estruturar uma árvore de decisão para orientar o pesquisador a analisar cada um dos fatores de avaliação estabelecendo pesos para cada um dos fatores avaliados;
- estruturar uma forma de retornar para o pesquisador as respostas verificadas pelo framework;

d) realizar a avaliação do framework:

- selecionar um grupo de pesquisadores multidisciplinares para o Grupo Focal;
- realizar Grupo Focal para discussão das etapas do framework;
- realizar ajustes sugeridos pelo Grupo Focal.

4 TRABALHOS RELACIONADOS

Harayama (2020) fez uma reflexão sobre as implicações bioéticas relacionadas aos sistemas de *big datas* utilizados em modelos preditivos de vigilância epidemiológica no Brasil. Dos trabalhos identificados até o momento, esse foi o que mais se aproximou da temática proposta neste estudo. O autor faz uma revisão narrativa relacionando os aspectos éticos dispostos na legislação e o uso das tecnologias para gestão de dados, de acordo com a recomendação da WHA 71.7³ [13].

As reflexões feitas por Harayama (2020) a respeito da limitação dos dados disponíveis, em obediência às regras estabelecidas pela legislação brasileira, não são suficientes para estabelecer os impactos gerados no uso desses dados em modelos preditivos.

O trabalho desenvolvido por Jia et al. (2020) aborda a efetividade do planejamento de armazenamento de dados de saúde para diminuir os prejuízos em eventual incidente de saúde pública. Os autores justificam que os dados coletados a partir de diversas fontes, se utilizados com métodos analíticos adequados, podem contribuir para a restauração da normalidade em eventual crise de saúde pública [12].

Em relação à ética no desenvolvimento de tecnologias para o armazenamento de dados de indivíduos contaminados com a Covid-19 a partir da geolocalização, a Organização Mundial da Saúde (OMS) emitiu orientações provisórias com princípios éticos, considerações técnicas e requisitos aderentes aos princípios legais relativos à respectiva prática [14].

Langarizadeh, Moghbeli e Aliabadi (2017) desenvolveram uma revisão sistemática para estudar artigos que discutiam questões éticas envolvendo serviços de saúde associados à tecnologia de informação. O artigo escrito a partir do estudo desenvolvido pelos autores retrata a necessidade de estabelecer diretrizes e padrões nos atendimentos de forma a tornar os profissionais de

³ Relatório da 71ª assembleia da OMS - World Health Assembly (WHA).

saúde responsáveis pelos seus atos [15] e conseqüentemente determina quais e como as informações daquele atendimento serão armazenadas.

Embora os temas abordados em cada um desses trabalhos citados estejam diretamente relacionados ao estudo proposto, não se identificaram, até o momento, relatos de estudos que objetivem a construção de um framework de análise de modelos preditivos para propagação da Covid-19.

5 REFERENCIAL TEÓRICO

A multidisciplinaridade da pesquisa exige que elementos básicos de cada área do conhecimento sejam explorados para fundamentar este estudo. A Covid-19 foi, de fato, a grande responsável por ressaltar o problema enfrentado pelos pesquisadores com a relação entre a falta de dados disponíveis para a pesquisa e os riscos gerados para os pesquisadores e para os consumidores dos resultados obtidos pela pesquisa. A estrutura básica de vigilância epidemiológica do Brasil contextualiza como a doença é identificada e como acontecem as notificações e coleta de dados em meio à pandemia.

A estrutura e gerenciamento das bases de dados em saúde explicam muito das dificuldades enfrentadas pelos pesquisadores. Por isso, identificar o que existe no Brasil, quais as alternativas existentes e as novidades para auxílio à pesquisa, são ações relevantes para fundamentar essa pesquisa. Assim como os dados, os modelos preditivos que utilizam essas bases de dados devem ser explorados e trazidos ao estudo como fundamento.

Por fim, aspectos jurídicos elementares e específicos para pesquisa são explorados, tais como a estrutura de normas brasileiras, os aspectos jurídicos básicos em saúde, os riscos envolvendo a saúde e a responsabilidade atraída pelos pesquisadores. A proposta é que os principais conceitos e desafios dessas grandes áreas do conhecimento sejam definidos de forma a justificar esta pesquisa.

5.1 COVID-19

Desde janeiro de 2020, a Covid-19 está classificada internacionalmente como emergência de saúde pública [16] e é descrita pelo Ministério da Saúde (2020) como uma “infecção respiratória aguda causada pelo coronavírus SARS-CoV-2, potencialmente grave, de elevada transmissibilidade e de distribuição global” [18].

Em 2002, o SARS-CoV, um vírus da mesma família do SARS-CoV-2, desencadeou uma síndrome respiratória aguda grave e, com a experiência progressiva, muitas informações sobre a família do vírus já eram conhecidas,

considerando que ambos possuem 86% de similaridade no seu genoma [16]. Embora essa fosse uma informação animadora, características muito específicas sobre o vírus SARS-CoV-2 foram sendo descobertas posteriormente, resultando em um comportamento da doença de 2019 bastante diferente do observado em 2002.

A contaminação silenciosa, que ocorre quando indivíduos infectados não apresentam sintomas, é apenas uma das dificuldades enfrentadas no combate à doença. De acordo com o Ministério da Saúde (2020), os sintomas são distúrbios olfativos e gustativos, dores de garganta ou de cabeça, tosse, coriza, calafrios e febre. A confirmação da doença pode se dar por (i) critérios clínicos, nos casos de disfunção olfativa ou disfunção gustatória sem causa pregressa ou ainda no contato com indivíduos confirmados, (ii) por critério de imagem ou por (iii) critério laboratorial [17].

De acordo com a Portaria GM/MS N° 1.792/2020, todos os casos com diagnóstico da doença, independentemente dos sintomas, devem ser notificados ao Ministério da Saúde (MS) (i) pelos laboratórios⁴, através da Rede Nacional de Dados em Saúde – RNDS, (ii) pelas unidades de saúde por meio do sistema e-SUS Notifica, (iii) pelas unidades de Vigilância Sentinela através do Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe) e (iv) pelos hospitais, através do SIVEP-Gripe [17].

Considerando a elevada transmissibilidade do vírus [18], estratégias para conter a propagação da doença, conforme o Ministério da Saúde, são o monitoramento dos indivíduos de acordo com o processo apresentado na Figura 1 e a adoção de um conjunto de medidas como cobrir nariz e boca com o antebraço ao tossir ou espirrar, não tocar olhos, nariz e boca, manter distanciamento entre as pessoas e manter os ambientes limpos e bem ventilados, o que se denomina “etiqueta respiratória” para reduzir a contaminação entre os indivíduos [17].

⁴ Art. 8º- A. É obrigatória a notificação ao Ministério da Saúde de todos os resultados de testes diagnósticos para detecção da COVID-19, realizados por laboratórios da rede pública, rede privada, universitários e quaisquer outros, em todo território nacional.

Há duas formas de imunizar a população: após infecção de grande parte da população ou através de vacina [19]. O desafio está, justamente, em diminuir os impactos da doença enquanto a população não é totalmente imunizada.

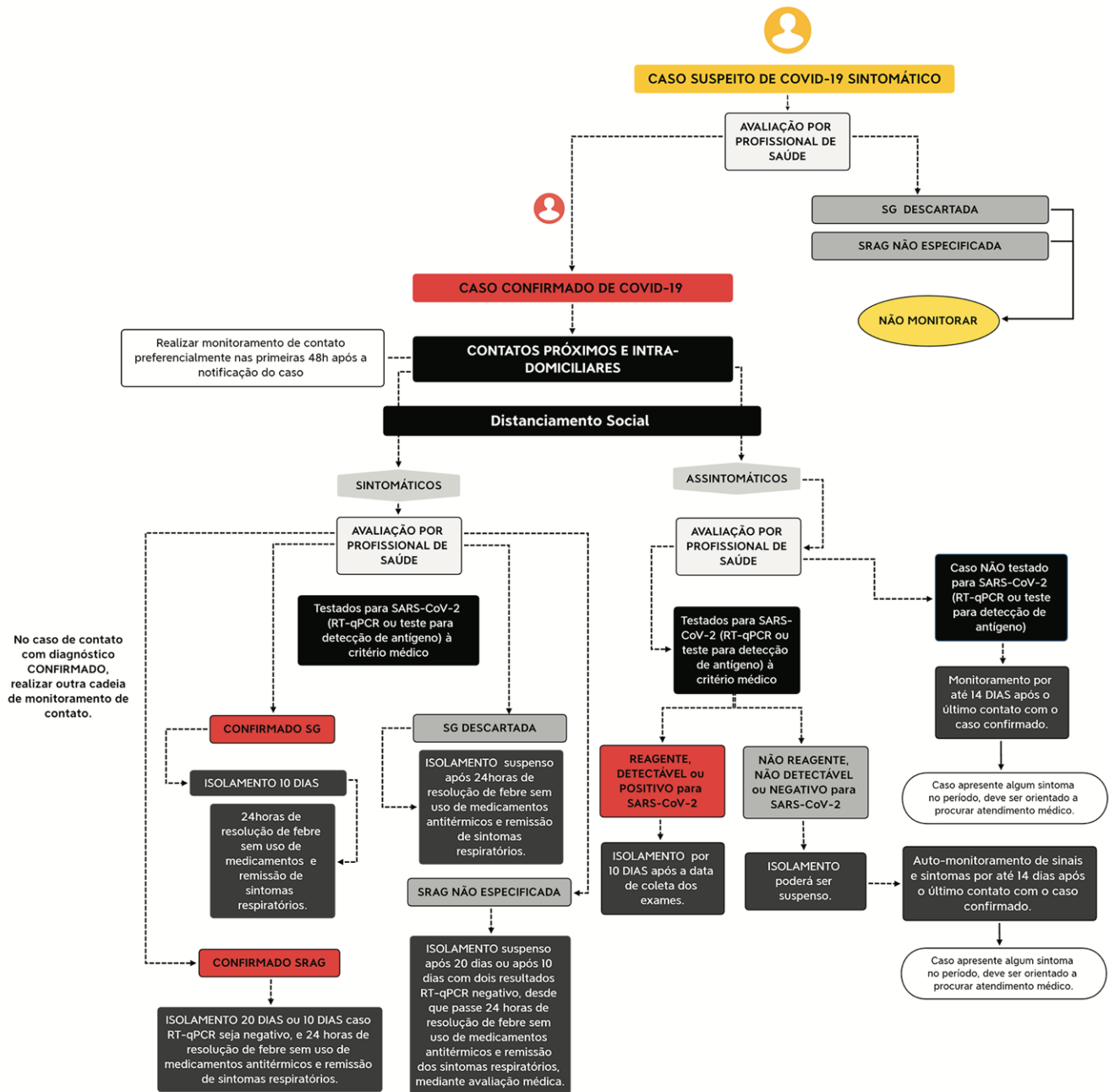


Figura 1 - Fluxograma de avaliação de casos e monitoramento de contatos.

Fonte: Brasil. Ministério da Saúde, 2020.

5.1.1 Vigilância Epidemiológica

Segundo a sétima edição do Guia de Vigilância Epidemiológica, as investigações de surtos e epidemias têm como objetivo principal identificar formas de interromper a transmissão e prevenir a ocorrência de casos novos [20]. Daí a importância dos dados capturados, armazenados e disponibilizados para o desenvolvimento de pesquisa, pois essas investigações, além de trazerem respostas à situação de crise, permitem que os cientistas identifiquem novas questões a serem objeto de pesquisas.

O Sistema Único de Saúde (SUS) está estruturado de forma híbrida, em que a prestação e a produção acontecem tanto pelo setor público quanto pelo privado. Embora o sistema esteja formatado para funcionar de maneira complementar, devido à capacidade e restrições públicas, a responsabilidade primária é pública, conforme estabelece a Constituição da República Federativa do Brasil (CRFB)⁵ [21].

A partir dos anos 90, foi incorporada ao sistema de saúde brasileiro a estrutura de Vigilância Epidemiológica, definida como “um conjunto de ações que proporcionam o conhecimento, a detecção ou prevenção de qualquer mudança nos fatores determinantes e condicionantes de saúde individual ou coletiva, com a finalidade de recomendar e adotar as medidas de prevenção e controle das doenças ou agravos” sob o princípio⁶ da utilização das informações epidemiológicas para as orientações centralizadas, gerenciamento das prioridades e alocação de recursos de saúde [22].

O Ministério da Saúde e a Secretaria de Vigilância em Saúde definem as funções da vigilância epidemiológica como coleta de dados, processamento de dados coletados, análise e interpretação dos dados processados,

⁵ Art. 196. A saúde é direito de todos e dever do Estado, garantido mediante políticas sociais e econômicas que visem à redução do risco de doença e de outros agravos e ao acesso universal e igualitário às ações e serviços para sua promoção, proteção e recuperação.

⁶ Lei nº 8.080/90 - Art. 7º. As ações e serviços públicos de saúde e os serviços privados contratados ou conveniados que integram o Sistema Único de Saúde (SUS), são desenvolvidos de acordo com as diretrizes previstas no art. 198 da Constituição Federal, obedecendo ainda aos seguintes princípios:

VII - utilização da epidemiologia para o estabelecimento de prioridades, a alocação de recursos e a orientação programática;

recomendação das medidas de prevenção e controle apropriadas, promoção das ações de prevenção e controle indicadas, avaliação da eficácia e efetividade das medidas adotadas e divulgação de informações pertinentes.

Essas funções estão distribuídas nos três níveis de competência (municipal, estadual e federal) do sistema brasileiro. Em todos os níveis, os dados devem ser devidamente recolhidos, tratados e ajustados, para que ao serem analisados em conjunto e de acordo com as estratégias de gestão epidemiológica, seja constituída uma grande base de dados, capaz de gerar informações que possam subsidiar um “processo dinâmico de planejamento, avaliação, manutenção e aprimoramento das ações”, conforme estabelece o Guia de Vigilância Epidemiológica desenvolvido pelo Ministério da Saúde do Brasil [20].

No ano de 2000, em apoio à OMS, o Brasil estruturou uma rede de pontos de referência com a finalidade de reunir informações sobre síndromes gripais, com o objetivo de programar ações específicas em casos epidêmicos como é o caso da Covid-19. Esses pontos de referência são chamados de Unidades Sentinelas, de responsabilidade e execução dos governos estaduais. Foi com base nos dados recolhidos por essas unidades que, em março de 2020, a Secretaria de Vigilância em Saúde do Ministério da Saúde (SVS/MS) declarou a transmissão comunitária da doença Covid-19 no Brasil [17].

Para centralizar as informações epidemiológicas, durante a pandemia, o Ministério da Saúde do Brasil, através da Portaria 1.434/2020 instituiu o Programa Conecte SUS com a finalidade de instituir a Rede Nacional de Dados em Saúde - RNDS e dispor sobre a adoção de padrões de interoperabilidade em saúde. O programa objetiva a integração dos estabelecimentos de saúde públicos e privados e dos órgãos de gestão em saúde dos municípios, estados e união de forma a centralizar as informações de saúde da população.

A portaria garante acesso às informações para fins de pesquisa:

*Art. 254-A, § 2º As informações constantes da RNDS poderão ser utilizadas para os seguintes fins:
III - estatísticos e de pesquisas;*

A portaria garante aos pesquisadores acesso às informações contidas na RNDS, no entanto é sabido que o programa ainda é muito recente. Além de, ainda, não haver dados históricos suficientes para o desenvolvimento de pesquisa, a utilização do sistema deve ser amplamente divulgada para que a população, de forma plena, utilize o sistema.

Os registros de saúde de uma nação merecem atenção especial, iniciando pelo conhecimento da sua população a respeito da importância de ter informações corretas até legislação própria capaz de proteger os registros de maneira adequada.

O cuidado com esses dados deve se dar a partir das estratégias básicas definidas pela gestão pública, passando pelo treinamento dos agentes de saúde no atendimento aos pacientes, coleta e armazenamento das informações e, por fim, o acesso da gestão pública e da pesquisa às informações corretas, estruturadas em tempo.

É muito importante que os agentes envolvidos nesse processo entendam a importância da identificação, captura, armazenamento, processamento e disponibilidade desses dados em saúde. A saúde de uma nação, não raras vezes, encontra-se nesses dados.

5.2 BASES DE DADOS EM SAÚDE

No Brasil, os sistemas de gestão de saúde são descentralizados e modelados para fins específicos como o Sistema de Informações sobre Nascidos Vivos (SINASC), o Sistema de Informações sobre Mortalidade (SIM) [11], o SIVEP-Gripe, o e-SUS [17] entre outros. Todas essas grandes bases possuem dados importantes para os estudos epidemiológicos. Esses dados relacionados podem revelar informações importantes sobre a saúde.

O Governo Federal do Brasil possui um painel⁷ com a relação das bases de dados públicas disponíveis, demonstrado na Figura 4. Um dos problemas existentes nesse serviço é que, embora o painel reúna as bases de cada órgão público, as estratégias de captura, armazenamento e gestão dos dados

⁷ <http://paineis.cgu.gov.br/>

continuam sendo descentralizadas, nas quais cada órgão trabalha para atender aos interesses específicos. Dessa forma, nem sempre será prioridade dos órgãos públicos disponibilizar as informações em tempo.

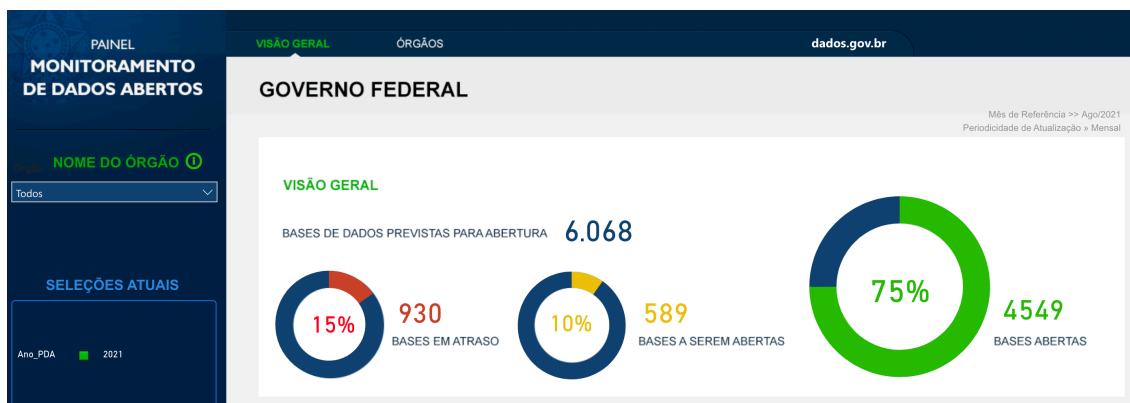


Figura 2 - Painel Monitoramento de Dados Abertos do Brasil.

Fonte: <http://painéis.cgu.gov.br/>

Vale lembrar que, ao tratar de aspectos de saúde de maneira ampla, não são apenas as bases de dados dos órgãos relacionados diretamente à saúde que importam. A título de exemplo, é possível citar o trabalho realizado por Codeço et al. (2020) que utiliza informações de transporte de passageiros através da rede aeroviária brasileira. Importa, nesse caso específico, saber o trajeto que os passageiros costumam fazer entre os aeroportos brasileiros para que se possa mapear o possível trajeto de um vírus ou a sistemática de transmissão da doença.

Em muitos países, assim como no Brasil, as pesquisas diretas, por telefone ou pessoalmente, são os métodos utilizados pelos órgãos públicos para entenderem as mudanças comportamentais dos indivíduos e planejar, a partir daí, estratégias de saúde pública. No entanto, tanto para a coleta de dados quanto para o planejamento de estratégias, são necessários tempo, recursos e instrumentos específicos [24]. É possível que, ao definir novas estratégias para a gestão de saúde pública, a situação que necessitava de intervenção imediata já não exista mais. Daí a preocupação de se desenvolverem bases de dados em saúde bem planejadas, estruturadas e disponíveis, em tempo, à pesquisa.

Tanto os gestores públicos como os cientistas necessitam de respostas rápidas e seguras sobre assuntos relacionados à saúde, uma vez que as

respostas aos problemas de saúde, como é o caso de uma pandemia, precisam ser rápidas, eficientes e eficazes. Além do rápido avanço da pandemia da Covid-19, outro exemplo brasileiro foi o vírus Zika, transmitido pelo mosquito *Aedes Aegypti*, identificado em 2015 [24], que exigiu dos pesquisadores e gestores públicos respostas seguras sobre as doenças transmitidas pelo mosquito e as suas reais consequências. Para a descoberta, é necessário pesquisa e, para pesquisa, inevitavelmente são necessários dados bem estruturados.

5.2.1 RNDS

Em 2017 foi instituída no Brasil a consolidação das normas sobre os direitos e deveres dos usuários da saúde, a organização e o funcionamento do Sistema Único de Saúde⁸. Durante a pandemia, em 2020, a Portaria foi alterada com objetivo de instituir o Programa Conecte-SUS, a fim de integrar estabelecimentos de saúde públicos e privados e dos órgãos de gestão em saúde dos entes federativos, para garantir o acesso à informação em saúde através da Rede Nacional de Dados em Saúde – RNDS⁹.

O programa do Governo Federal objetiva a promoção da troca de informações entre os pontos da Rede de Atenção à Saúde - RAS, permitindo a transição e continuidade do cuidado nos setores público e privado¹⁰. A promessa é que a plataforma se torne, com o tempo, a principal ferramenta informacional de dados de saúde do Brasil, garantindo aos pesquisadores o acesso aos dados de forma ética e segura.

Para auxiliar no combate à Covid-19, a RNDS estruturou os dados em saúde em camadas para garantir a segurança das informações, a interoperabilidade entre os sistemas públicos utilizados, os processos de notificação e vigilância em saúde e canais de acesso, conforme é possível verificar na Figura 3.

⁸ Portaria de Consolidação nº 1/GM/MS, de 28 de setembro de 2017.

⁹ Portaria nº 1.434, de 28 de maio de 2020 - Art. 2º Fica instituído o Programa Conecte SUS, no âmbito do Ministério da Saúde, voltado à informatização da atenção à saúde e à integração dos estabelecimentos de saúde públicos e privados e dos órgãos de gestão em saúde dos entes federativos, para garantir o acesso à informação em saúde necessário à continuidade do cuidado do cidadão.

¹⁰ <https://www.gov.br/saude/pt-br/assuntos/rnds> - acessado em 17 de outubro de 2021 às 09:44.

A preocupação com a segurança das informações, considerando inclusive a Lei Geral de Proteção de Dados – LGPD -, já é um desafio identificado pela equipe de trabalho. Nesse sentido, é necessário fazer duas observações importantes a respeito da segurança de dados: (a) se os dados, antes de disponibilizados, passam pela administração do poder público, é possível que sejam alterados ou disponibilizados de acordo com estratégias políticas e (b) as instituições privadas poderão utilizar a legislação existente atualmente para não disponibilizarem suas bases de dados para o desenvolvimento de pesquisa porque obviamente existe forte interesse comercial nessas informações.

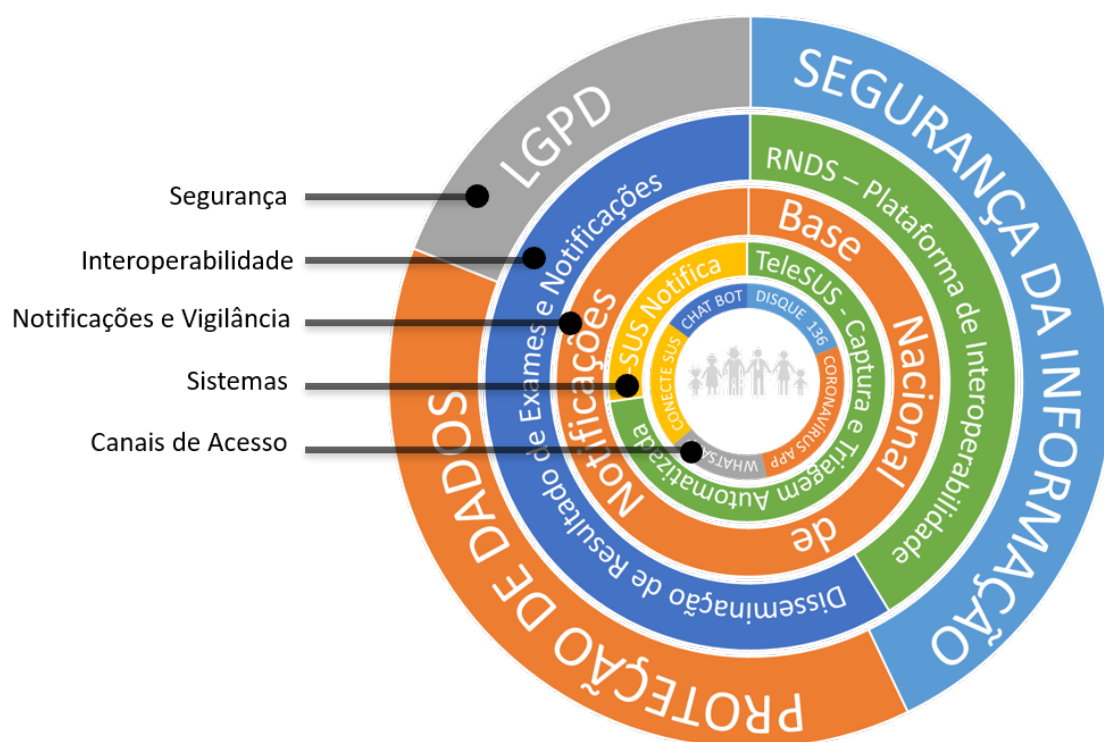


Figura 3 - Ecossistema de atendimento e de dados em saúde.

Fonte: <https://www.gov.br/saude/pt-br/assuntos/rnds>

O ecossistema proposto pela RNDs é, sem dúvida, um diferencial para a disponibilidade de dados em saúde no Brasil, porque um dos grandes desafios é a interoperabilidade dos diferentes sistemas de informação. No entanto, o sistema, em nenhum momento fala a respeito das regras de coleta e gerenciamento de dados.

Dados iguais podem ser disponibilizados por instituições diferentes: o que pode mudar é a regra utilizada para tratar e disponibilizar esses dados. A RNDS não pode levar o pesquisador, que utiliza dados de saúde, a interpretar os dados de forma errada.

O fato de a base de dados ser pública ou ser disponibilizada por um órgão reconhecido não garante confiabilidade aos dados. É fundamental que o pesquisador conheça as regras de coleta e gerenciamento das informações ali dispostas.

Outro ponto fundamental refere-se ao viés político que é possível conter na estratégia de disponibilidade das informações.

5.2.2 Big Data

O *Big Data* dedica-se ao armazenamento de diferentes conjuntos de dados não relacionados, o processamento desse volume de dados e a possibilidade de condução de análises que resultem em informações com valor para a tomada de decisões em tempo hábil [25].

Para a *TechAmerica Foundation*, o “*Big Data* é um fenômeno definido pela rápida aceleração no volume em expansão de dados de alta velocidade, complexos e diversos tipos de dados, geralmente definido em três dimensões - volume, velocidade e variedade” [26].

O potencial no uso de grandes bases de dados, utilizando princípios de *Big Data* pelos órgãos públicos, é amplo. Afinal, explorar dados já disponíveis pelos serviços básicos e encontrar sentido entre eles é um dos caminhos para gerar respostas rápidas aos problemas específicos e ainda mapear informações para o desenvolvimento de estratégias de gestão pública [26].

Os Big Datas são, sem dúvida, alternativas para a obtenção de dados em saúde quando elas não existirem em bases bem estruturadas e planejadas. No entanto, o risco que os dados carregam são enormes pela falta de planejamento, organização e, muitas vezes, lógica.

O risco na utilização de dados a partir dessas bases desestruturadas é potencializado justamente porque o volume de dados é muito grande, portanto, para a utilização dos dados é necessária, a realização de inferência amostral.

5.3 MODELOS PREDITIVOS

Os modelos matemáticos aplicados à epidemiologia estão sendo desenvolvidos desde 1760, iniciados por Daniel Bernoulli com o estudo sobre a varíola [27], sendo aprimorados e utilizados até os dias atuais. Mais recentemente, com a pandemia da Covid-19, é possível, na prática, colher bons resultados desse conhecimento que vem sendo desenvolvido ao longo da história recente.

Araújo (2002) justifica a dificuldade de definir o conceito de modelo matemático e conclui que “modelo é uma representação simplificada de uma situação concreta feita com o objetivo de compreender a situação e prever suas configurações futuras ou de situações semelhantes”. Nessa linha ainda, Goldburg e Luna (2005) definem os modelos como “representações simplificadas da realidade que preservam, para determinadas situações e enfoques, uma equivalência adequada”.

Os modelos matemáticos preditivos devem ser confiáveis e eficientes e para que isso aconteça, é fundamental que a amostra considerada da população, em regra, tenha sido selecionada com critérios bem definidos e conhecimento para determinar a importância relativa das variáveis utilizadas [30]. Um dos desafios de abstração da realidade é justamente definir a importância das variáveis isoladas naquele conjunto utilizado pelo modelo, ainda que elas sejam bem definidas [29].

Como a população selecionada para desenvolver estudos envolvendo predições virais é fundamental, tornam-se desafiadores os estudos preditivos de novas doenças, já que não há informações e conhecimento suficientes para determinar um *dataset* (base de dados) confiável tampouco a importância relativa das informações existentes para que sejam utilizadas como variáveis em um modelo preditivo.

5.4 RISCOS E ASPECTOS LEGAIS

Os riscos podem ser observados por diferentes abordagens, no entanto, nesta pesquisa, eles dizem respeito à segurança jurídica da pesquisa e do pesquisador.

Primeiramente, com intuito de contextualizar o leitor que eventualmente não seja familiarizado com a linguagem jurídica, são apresentados conceitos básicos a respeito da estrutura jurídica e das normas jurídicas brasileiras. Posteriormente são abordados aspectos jurídicos direcionados à pesquisa, à saúde e a dicotomia jurídica entre o acesso e a privacidade de dados em saúde.

5.4.1 Contextualização

Inicialmente, o termo “aspectos jurídicos” passa a impressão de um conjunto de regras e de procedimentos fiscalizatórios, somente. Tais aspectos vão muito além disso, as relações são estabelecidas por critérios objetivos, com regras pré-estabelecidas e por critérios subjetivos, que necessitam da intervenção do Estado para a resolução de um conflito ou para a garantia de um direito ou um dever.

A Constituição de 1988 deu à saúde o status de direito público subjetivo, um protagonismo sem precedentes para o Direito brasileiro. O principal livro de leis estabeleceu que é dever do Estado o controle da saúde da sua população, na oferta tanto de saúde individual quanto de saúde coletiva. O direito à saúde é dito subjetivo justamente porque as necessidades individuais não são as mesmas para todos e, além disso, não é possível estabelecer um rol taxativo de regras e responsabilidades [31].

A ciência em saúde deve ser um dos pilares do desenvolvimento da política de um determinado país, conforme recomenda a I Conferência Nacional de Ciência e Tecnologia em Saúde. Esse pilar deve contemplar as pesquisas direcionadas ao avanço do conhecimento, aos problemas práticos e urgentes e ao desenvolvimento de processos e produtos [32].

O Manifesto da Associação Brasileira de Saúde Coletiva - (ABRASCO) enviado à Conferência Nacional da Ciência, Tecnologia e Inovação do Brasil [32], embora seja de 2001, continua sendo base de referência, agora não com enfoque em tecnologia mas no aspecto legal. Nos últimos anos, o poder legislativo brasileiro evidenciou sua preocupação com a tutela dos direitos individuais [33].

Identifica-se a preocupação dos direitos da personalidade a partir da Constituição Federal publicada em 1988 [34] que diz que “são invioláveis a intimidade, a vida privada, a honra e a imagem das pessoas, assegurado o direito a indenização pelo dano material ou moral decorrente de sua violação”¹¹. Pouco tempo depois, o Código de Defesa do Consumidor - CDC - estabeleceu que “o consumidor, [...], terá acesso às informações existentes em cadastros, fichas, registros e dados pessoais e de consumo arquivados sobre ele, bem como sobre as suas respectivas fontes”¹².

Em 2002, o Código Civil [35] definiu que “a vida privada da pessoa natural é inviolável, e o juiz, a requerimento do interessado, adotará as providências necessárias para impedir ou fazer cessar ato contrário a esta norma”¹³. Com o rápido desenvolvimento da tecnologia e uso da Internet, regras importantes precisaram ser incorporadas à legislação brasileira como a lei 12.737/2012, popularmente conhecida como Lei Carolina Dickmann, que dispõe sobre a tipificação criminal de delitos informáticos, e a lei 12.965/2014, Marco civil da Internet, que “estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil e determina as diretrizes para atuação da União, dos Estados, do Distrito Federal e dos Municípios em relação à matéria”¹⁴. Ainda, em consonância aos movimentos gerados pela União Europeia e pelos Estados Unidos, em 2018 a Lei Geral de Proteção de Dados - LGPD - foi desenvolvida no Brasil, com regras mais específicas para uso dos dados individuais e sanções bem definidas quando utilizadas sem o específico consentimento do indivíduo [36].

¹¹ Art. 5º, inciso x da Constituição Federal.

¹² Art. 43 da Lei 8.078/90.

¹³ Art. 21 do Código Civil.

¹⁴ Art. 1º da Lei 12.965/2014.

Todavia, o rápido desenvolvimento de regras para uso de dados pessoais confronta com o interesse coletivo de produção de ciência e busca de novos conhecimentos. Por certo, é dicotômica a ideia de ter ou não acesso aos dados pessoais dos indivíduos, sendo necessário desenvolver conhecimento científico para o aprimoramento da saúde a partir do macroplanejamento político e estabelecer as regras de acesso às informações pessoais de forma que não inviabilize o progresso científico ou até mesmo a possibilidade de controle de uma pandemia, como é o caso da Covid-19.

5.4.2 Aspectos Jurídicos

Para aplicar as normas adequadamente, faz-se importante a compreensão de conceitos básicos jurídicos como a hierarquia das normas, a validade no tempo e no espaço, os conflitos entre as normas e, por fim, os aspectos fundamentais dos direitos individuais e direitos coletivos.

As normas são criadas de acordo com a evolução e o desenvolvimento da sociedade à medida que surge a necessidade de estabelecer regras e limites entre as relações jurídicas. Embora não exista um consenso na definição de norma jurídica, para este estudo utilizaremos o conceito de Hans Kelsen [37] que diz que algo “deve ser ou acontecer, especialmente que um homem se deve conduzir de determinada maneira”, ou seja, um manual de condutas. Afinal o Direito é, na verdade, a aplicação cotidiana do conjunto de normas existentes.

De acordo com o Conselho Nacional de Justiça – CNJ¹⁵, a Constituição Federal é a maior lei do ordenamento jurídico nacional com objetivo de disciplinar tanto os direitos individuais quanto os direitos coletivos fundamentada nos direitos fundamentais dos direitos humanos, da cultura, da proteção da família, da educação e da saúde. Ou seja, todas as demais normas devem, obrigatoriamente, observar esses direitos fundamentais.

¹⁵ <https://www.cnj.jus.br/cnj-servico-conheca-a-hierarquia-das-leis-brasileiras/>

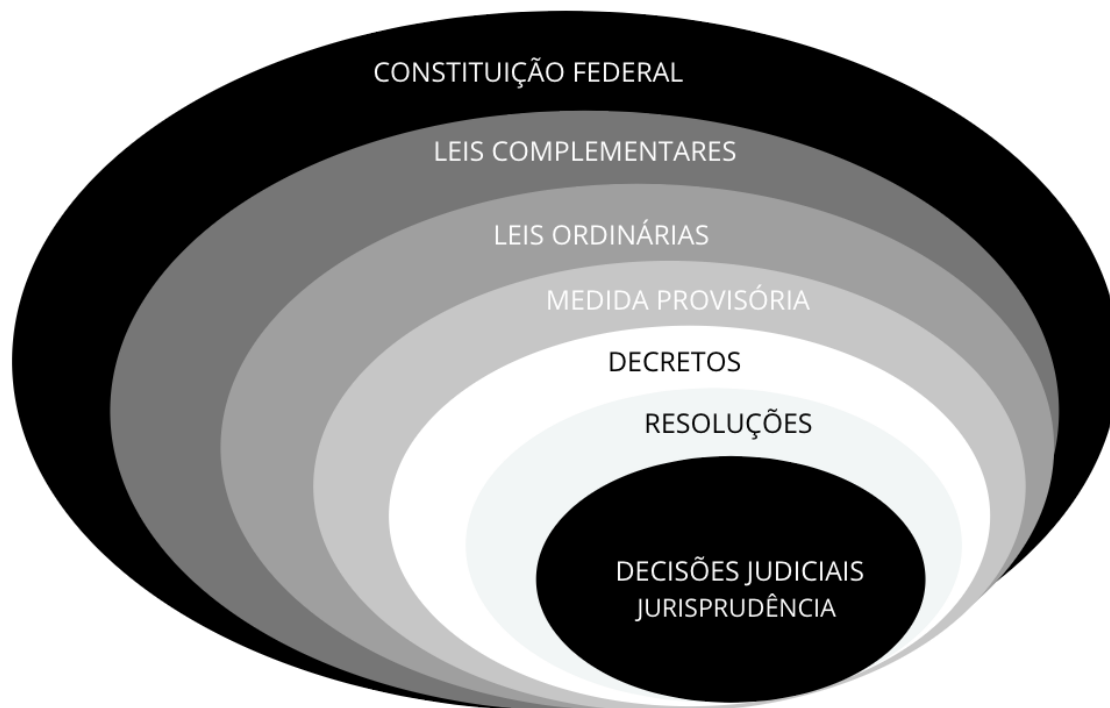


Figura 4 - Hierarquia das Normas - Desenvolvido pelo autor.

Fonte: Autor.

A Figura 4 ilustra a hierarquia das normas, na qual as mais internas devem respeitar as regras mais externas. As leis complementares têm o objetivo de regulamentar lacunas da Constituição Federal – CF, enquanto as leis ordinárias são aquelas criadas pelo poder legislativo (Deputados e Senadores) com aprovação presidencial. As medidas provisórias são as normas desenvolvidas em caráter de urgência que, em princípio, perdem sua validade em sessenta dias. Enquanto os decretos legislativos são atos normativos do Congresso Nacional, como exemplo, a ratificação de tratados internacionais. As resoluções são diretrizes internas de cada órgão e as decisões judiciais, conhecidas como jurisprudências, são decisões reiteradas do Poder Judiciário através dos tribunais, retratando o entendimento e a interpretação sobre as leis nos casos concretos do cotidiano da vida humana. Neste último caso, é importante considerar a hierarquia de decisões entre os órgãos judiciais (justiça de primeiro Grau, Tribunal de Justiça, STJ Superior Tribunal de Justiça e STF Supremo Tribunal Federal).

Considerando a hierarquia das normas brasileiras, quanto ao regramento internacional, cabem ao Presidente da República as relações com os demais países de acordo com o artigo 84, VII e VIII da CF. Em regra, após o poder Executivo concluir as tratativas sobre o tratado internacional, ele é enviado para a aprovação parlamentar que aprova o texto e internaliza a norma através de decreto legislativo.

Com relação à hierarquia de normas, os tratados e convenções internacionais, em regra, são equiparadas às emendas constitucionais quando tratarem de assuntos referentes aos Direitos Humanos (Constituição Federal, art. 5º, §, 3º e art. 47); e às leis ordinárias quando versarem sobre outros temas [38].

Considerando a complexidade e a quantidade de normas existentes no Brasil, é natural que aconteçam conflitos entre elas. O artigo 2º, §1º da LINDB, diz que, diante do conflito de normas, somente uma deve ser aplicada ao caso concreto.

Os critérios para a escolha da norma a ser aplicada, no caso de eventual conflito, devem obedecer nessa ordem à hierarquia, a especialidade e a cronologia. Observado o critério da hierarquia no qual as normas mais internas devem, obrigatoriamente, obedecer aos critérios das mais externas, conforme demonstrado na Figura 4, ainda existindo conflito, passa-se a verificar a especialidade da norma, quando a lei especial derroga a lei geral. Ainda assim existindo o conflito, a norma posterior afasta a anterior. Considerando o artigo 9º da Lei Complementar número 95: a norma mais recente deverá, expressamente, relacionar os dispositivos revogados.

A responsabilidade civil está diretamente relacionada aos atos e aos resultados que apresenta. Diniz (2004) parte do pressuposto de que, qualquer indivíduo que através de um ato (ou a falta de um ato necessário), violar o dever jurídico de não causar dano a terceiro, passa a ter a responsabilidade de reparar o dano causado.

Na realidade, toda a manifestação da atividade do homem traz em si mesma o problema da responsabilidade¹⁶.

Originariamente do latim, o termo responsabilidade é uma transformação do termo *responsus*, o particípio passado do verbo *respondere*, que significa retornar uma garantia ou uma promessa. No latim, o termo é utilizado sempre que um indivíduo causa dano a outro e adquire, a partir daí, a responsabilidade de responder ao prejudicado a reparação dos danos causados pelos seus atos.

A ideia para Gagliano e Pamplona Filho (2014) é que a resposta de reparação reestabeleça à situação original em que se encontrava o sujeito prejudicado como se nada houvesse acontecido. Sabemos que nem sempre é possível reestabelecer o *status quo* (estado original), pois alguns danos são irreparáveis, como o dano à honra e a imagem, por exemplo. A partir daí faz-se necessário avaliar as formas de compensações possíveis, já que a restituição já não é mais possível.

Além de trabalhar a restituição do estado anterior, a responsabilidade civil se presta a reestabelecer a ordem jurídica, no sentido de coibir ações reiteradas que causem danos. Pode-se dizer que é um processo pedagógico de responsabilidade civil. Dantas (1978) diz que o principal objetivo da ordem jurídica é "proteger o lícito e reprimir o ilícito. Vale dizer: ao mesmo tempo em que ela se empenha em tutelar a atividade do homem que se comporta de acordo com o Direito, e reprimi a conduta daquele que contraria".

É dever de todos respeitar e obedecer às leis impostas pelo ordenamento jurídico (obrigação), mantendo um equilíbrio jurídico e social. Mas, assim não sendo, é possível a busca da reparação do direito violado pelo agente (responsabilidade), através da imposição de uma sanção (sentimento de justiça). Nessa linha tênue entre obrigação e responsabilidade, a autora Maria Helena Diniz bem conceitua o direito da vítima em ver seu patrimônio ou direito ressarcido em seu *status quo ante* isto é, anterior à violação pela pessoa do infrator.

O interesse em restabelecer o equilíbrio violado pelo dano é a fonte da responsabilidade civil. Na responsabilidade civil são a perda ou a diminuição verificada no patrimônio do lesado ou o dano moral que geram a reação legal, movida pela ilicitude da ação do autor da lesão ou pelo risco.¹⁷

E, segundo Sergio Cavalieri Filho:

Designa o dever que alguém tem de reparar o prejuízo decorrente da violação de outro dever jurídico. Em apertada síntese, responsabilidade civil é um dever jurídico sucessivo que surge para recompor o dano decorrente da violação de um dever jurídico originário. Só se cogita, destarte, de responsabilidade civil onde houver violação de um dever jurídico e dano.¹⁸

Assim, quem viola uma norma genérica, que impõe esse dever de conduta, ou seja, quem age culposamente, comete um ato ilícito, portanto tem o dever de reparar o dano [39].

Esse entendimento está no Código Civil de 2002 nos artigos 186 e 927, o qual se transcreve abaixo:

Art. 186. Aquele que, por ação ou omissão voluntária, negligência ou imprudência, violar direito e causar dano a outrem, ainda que exclusivamente moral, comete ato ilícito.

Art. 927. Aquele que, por ato ilícito (arts. 186 e 187), causar dano a outrem, fica obrigado a repará-lo.

Parágrafo único. Haverá obrigação de reparar o dano, independentemente de culpa, nos casos especificados em lei, ou quando a atividade normalmente desenvolvida pelo autor do dano implicar, por sua natureza, risco para os direitos de outrem.

A responsabilidade civil é o dever de reparar o dano causado a outrem, na tentativa que tal reparação devolva o estado *a quo* em que a pessoa ou coisa se encontrava, ou que tal reparação seja capaz de satisfazer e/ou substituir aquilo que foi prejudicado.

Portanto, o sujeito só será responsável por indenizar ou reparar determinado dano, quando for comprovada, através do Poder Judiciário, sua culpa na prática do fato ilícito e danoso a outrem. Do contrário, caso não venha a ser comprovada a efetiva culpa, nada será imposto ao suposto indivíduo causador do referido dano [40].

¹⁷ DINIZ, Maria Helena. **Curso de direito civil brasileiro**. São Paulo: Saraiva, 2004. p. 5.

¹⁸ CAVALIERI FILHO, Sergio. **Programa de Responsabilidade Civil**. São Paulo: Atlas, 2010. p. 2.

Ainda, considerando o conflito de normas, é importante explicar a diferença existente entre o direito individual e o direito coletivo. Os direitos coletivos ou difusos são aqueles transindividuais cujos titulares são pessoas indeterminadas conforme, artigo 81, § único, I, do CDC e artigo 1º, I, do CM. Há diferença no conceito de direitos difusos e coletivos. No primeiro não é possível identificar o grupo social que detém o respectivo direito, enquanto, no segundo, o grupo social é identificável.

Os direitos estão intrinsecamente atrelados aos interesses e necessidades coletivos ou individuais. A Supremacia do Interesse Público é fundamento do Direito brasileiro, tornando a garantia desses interesses legítimos dever do Estado [41]. Um exemplo disso diz respeito à Lei nº 13.979/2020 que dispõe sobre as medidas para enfrentamento da emergência de saúde pública de importância internacional decorrente do coronavírus responsável pelo surto de 2019. O que se vê na normativa é uma série de comandos para garantir o interesse coletivo à saúde.

A proposta aqui não é aprofundar a temática dos limites dos direitos coletivos em detrimento dos individuais, mas pontuar que, através de decisão do Supremo Tribunal Federal, os direitos à vida e à saúde coletiva devem prevalecer sobre o direito individual [41].

5.4.2.1 Aspectos Jurídicos da Saúde

A Constituição Federal do Brasil, no artigo 196 traz a saúde como um direito social¹⁹, portanto coletivo, no rol dos direitos fundamentais. Mais adiante o artigo 197 reafirma o interesse público e diz que é responsabilidade do Poder Público regulamentar as ações que dizem respeito ao interesse público relacionado à saúde no Brasil.

*Art. 196. A saúde é **direito de todos e dever do Estado**, garantido mediante **políticas sociais e econômicas** que visem à redução do risco de doença e de outros agravos e ao acesso universal e igualitário às **ações e serviços para sua promoção, proteção e recuperação**.*

¹⁹ Constituição Federal - Art. 6º **São direitos sociais** a educação, **a saúde**, a alimentação, o trabalho, a moradia, o transporte, o lazer, a segurança, a previdência social, a proteção à maternidade e à infância, a assistência aos desamparados, na forma desta Constituição.

*Art. 197. São de **relevância pública as ações** e serviços de saúde, cabendo ao **Poder Público dispor**, nos termos da lei, sobre sua **regulamentação**, fiscalização e controle, devendo sua execução ser feita diretamente ou através de terceiros e, também, por pessoa física ou jurídica de direito privado.*

Isso quer dizer que, além do Estado, também as instituições privadas devem estar atentas aos direitos fundamentais [42].

Os Direitos Fundamentais são, por essência, aqueles necessários à proteção do ser humano devidamente relacionados nas leis basilares de uma nação. Devem ser garantidos através de ferramentas sociais de repercussão geral. Isto é, assim como o direito é garantido, também é um dever de todos.

A VIII Conferência Nacional de Saúde - VIII CNS, ocorrida em 1986, deu origem ao Sistema Único de Saúde (SUS) no Brasil [43] e, já na época, discutiam-se as dificuldades e enfrentamento do desenvolvimento de pesquisas em saúde no Brasil. A partir daí, regulamentações foram desenvolvidas, tais como a Lei 8080/90, que dispõe sobre “as ações e serviços de saúde, executados isolada ou conjuntamente, em caráter permanente ou eventual, por pessoas naturais ou jurídicas de direito Público ou privado.”²⁰

O legislador, à época, já se preocupava com a pesquisa relacionada à saúde, observando a criação de comissões para a integração entre os serviços de saúde e as instituições de pesquisa. Às comissões, a lei estabeleceu a finalidade de propor prioridades, métodos e estratégias para a pesquisa em saúde.²¹

Às unidades político-administrativas (União, Estados, Distrito Federal e Municípios) do Estado cabe a realização de pesquisas e estudos na área da saúde²² estabelecendo inclusive mecanismos para incentivo para participação

²⁰ Lei 8080/90 - Art. 1º Esta lei regula, em todo o território nacional, as ações e serviços de saúde, executados isolada ou conjuntamente, em caráter permanente ou eventual, por pessoas naturais ou jurídicas de direito Público ou privado.

²¹ Lei 8080/90 - Art. 14. Deverão ser criadas **Comissões Permanentes de integração entre os serviços de saúde e as instituições de ensino profissional e superior**. Parágrafo único. Cada uma dessas comissões terá por **finalidade propor prioridades, métodos e estratégias** para a formação e educação continuada dos recursos humanos do Sistema Único de Saúde (SUS), na esfera correspondente, assim como **em relação à pesquisa** e à cooperação técnica entre essas instituições.

²² Lei 8080/90 - Art. 15. A União, os Estados, o Distrito Federal e os Municípios exercerão, em seu âmbito administrativo, as seguintes atribuições: XIX - realizar pesquisas e estudos na área de saúde;

do setor privado com contrapartida da transferência de tecnologias desenvolvidas²³.

Sabe-se que atualmente as instituições privadas são responsáveis por gerar informações sobre o comportamento social. Essas informações muitas vezes são fundamentais para o desenvolvimento de pesquisas científicas. Assim como há a previsão de “*transferência de tecnologia das universidades e institutos de pesquisa ... às empresas nacionais*”, deveria haver a contrapartida, ou seja, a transferência de dados e informações gerados pelo setor privado aos institutos de pesquisa.

5.4.2.2 Aspectos Jurídicos da Informação

Indiscutivelmente, a informação é um Direito Fundamental de toda a pessoa humana: sem informação, um indivíduo não tem possibilidade de se desenvolver e se relacionar. A Constituição Federal, no artigo 5º, aborda o assunto de forma expressa:

*XXXIII – todos têm direito a receber dos órgãos públicos informações de seu **interesse particular**, ou de **interesse coletivo ou geral**, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado”.*

No entanto, o inciso XXXIII do artigo 5º deve ser interpretado de maneira ampla, considerada a relevância de interesse público, inclusive expressamente citado no texto. Ao considerar que os órgãos públicos devem prestar as informações de interesse particular, coletivo ou geral, deve-se lembrar que cabe, antes disso, aos órgãos públicos buscar informações de interesse público justificáveis. Como exemplo, a Receita Federal, na solicitação de acesso ao faturamento e identificação dos seus clientes, às Secretarias de Saúde quando obrigam as empresas privadas de saúde a registrarem seus estabelecimentos, tipos de serviços prestados e assim por diante.

²³ Lei 8080/90 - Art. 46. o Sistema Único de Saúde (SUS), estabelecerá mecanismos de incentivos à participação do setor privado no investimento em ciência e tecnologia estimulará a transferência de tecnologia das universidades e institutos de pesquisa aos serviços de saúde nos Estados, Distrito Federal e Municípios, e às empresas nacionais.

A informação tem relevância fundamental não apenas no Brasil, mas também repercute na esfera internacional através da Carta Africana sobre os Direitos Humanos e dos Povos²⁴ e da Declaração Universal dos Direitos Humanos²⁵ [44], por exemplo.

O Código de Defesa do Consumidor, com a sua promulgação, iniciou a regulamentação a respeito dos dados pessoais.

Lei nº 8.078/1990 - Art. 43. O consumidor, sem prejuízo do disposto no art. 86, terá acesso às informações existentes em cadastros, fichas, registros e dados pessoais e de consumo arquivados sobre ele, bem como sobre as suas respectivas fontes.

Ressalta-se o disposto no parágrafo 4º do artigo 43 da lei:

§ 4º Os bancos de dados e cadastros relativos a consumidores, os serviços de proteção ao crédito e congêneres são considerados entidades de caráter público.

Ou seja, se os dados têm caráter público, significa dizer que a solicitação de acesso às informações pode surgir a partir do interesse público.

Com o advento da Internet, foi necessário que regulamentações fossem criadas para limitar o direito e o uso das informações, sobretudo na sua forma digital. Num primeiro momento, com objetivo de proteger os órgãos públicos, regulamentou-se, por de lei especial, o direito de acesso às informações através da ferramenta especial do Habeas Data, com a lei nº 9.507/97. A ferramenta pode ser utilizada para busca de informações constantes nos registros ou bancos de dados de caráter público²⁶.

A lei sobre Direitos Autorais, Lei nº 9.610/1998, começa a regular os direitos autorais sobre as bases de dados privadas, dando ao titular da base de

²⁴ Artigo 9º - Toda pessoa tem direito à informação.

²⁵ Article 19 - *Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.*

²⁶ Art. 1º (VETADO)

Parágrafo único. Considera-se de caráter público todo registro ou banco de dados contendo informações que sejam ou que possam ser transmitidas a terceiros ou que não sejam de uso privativo do órgão ou entidade produtora ou depositária das informações.

Art. 7º Conceder-se-á habeas data: constantes de registro ou banco de dados de entidades governamentais ou de caráter público;

dados a exclusividade de autorizar a distribuição parcial ou total daquele patrimônio²⁷.

Nesse viés privado, surge a lei do cadastro positivo, Lei nº 12.414/2011 para disciplinar a formação e a consulta a bancos de dados com informações de pessoas naturais e jurídicas para formação de histórico de crédito. Aqui, pela segunda vez, o legislador deixa de utilizar o interesse público para ter acesso aos dados particulares, limitando a disponibilidade das informações de acordo com a finalidade de crédito²⁸.

A lei de Acesso à Informação, Lei nº 12.527/2011, estabelece conceitos importantes sobre as informações particulares das regras de acesso das informações. Interessa saber quem deve submeter-se ao respectivo regramento. Além das empresas públicas, submetem-se à lei:

Art. 2º Aplicam-se as disposições desta Lei, no que couber, às entidades privadas sem fins lucrativos que recebam, para realização de ações de interesse público, recursos públicos diretamente do orçamento ou mediante subvenções sociais, contrato de gestão, termo de parceria, convênios, acordo, ajustes ou outros instrumentos congêneres.

Isso quer dizer que somente não se obrigam as empresas de caráter privado com fins lucrativos.

Mais tarde, o Marco Civil da Internet surge a partir da promulgação da Lei nº 12.965/2014, estabelecendo princípios, garantias, direitos e deveres para o uso da Internet no Brasil, além de determinar diretrizes para atuação do Estado

²⁷ Lei 9610/98 - Art. 87. O titular do direito patrimonial sobre uma base de dados terá o direito exclusivo, a respeito da forma de expressão da estrutura da referida base, de autorizar ou proibir:

I - sua reprodução total ou parcial, por qualquer meio ou processo;

II - sua tradução, adaptação, reordenação ou qualquer outra modificação;

III - a distribuição do original ou cópias da base de dados ou a sua comunicação ao público;

IV - a reprodução, distribuição ou comunicação ao público dos resultados das operações mencionadas no inciso II deste artigo.

²⁸ Lei nº 12.414/2011 - Art. 5º São direitos do cadastrado: VII - ter os seus dados pessoais utilizados somente de acordo com a finalidade para a qual eles foram coletados.

Lei nº 12.414/2011 - Art. 7º As informações disponibilizadas nos bancos de dados somente poderão ser utilizadas para: I - realização de análise de risco de crédito do cadastrado; ou II - subsidiar a concessão ou extensão de crédito e a realização de venda a prazo ou outras transações comerciais e empresariais que impliquem risco financeiro ao consulente.

com relação às Informações disponíveis na Rede, prevendo já naquela época a proteção à privacidade de dados pessoais²⁹.

Art. 3º A disciplina do uso da internet no Brasil tem os seguintes princípios:

II - proteção da privacidade;

III - proteção dos dados pessoais, na forma da lei;

O termo “na forma da lei”, no inciso III, deixa claro que a proteção se dará na forma que a lei estipular, no entanto nunca houve tal regulamentação.

Somente em 2018 com a Lei Geral de Proteção de Dados – LGPD, Lei nº 13.709/2018, houve por parte do legislador a estruturação de regras para o efetivo controle e proteção dos dados pessoais das pessoas naturais³⁰. Essas normas ainda não estão totalmente regulamentadas. Isto é, a lei funciona como uma diretriz com objetivos, princípios, consequências e definição de algumas responsabilidades. É necessário que regras definidas nessa grande norma sejam especificadas, ou melhor, regulamentadas de acordo com as necessidades e prioridades estabelecidas pelo grupo de trabalho.

O desenvolvimento da pesquisa está incluído na regulamentação, com peculiaridades. Para a pesquisa, conforme artigos. 4º, 7º e 11 da lei:

Art. 4º Esta Lei não se aplica ao tratamento de dados pessoais:

II - realizado para fins exclusivamente:

b) acadêmicos, aplicando-se a esta hipótese os arts. 7º e 11 desta Lei;

Art. 7º O tratamento de dados pessoais somente poderá ser realizado nas seguintes hipóteses:

IV - para a realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais;

Art. 11. O tratamento de dados pessoais sensíveis somente poderá ocorrer nas seguintes hipóteses:

II - sem fornecimento de consentimento do titular, nas hipóteses em que for indispensável para:

c) realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais sensíveis;

Com relação à saúde pública, objeto deste estudo, a Lei diz que os órgãos de pesquisa, públicos ou privados, poderão ter acesso às bases de dados com

²⁹ Lei nº 12965/2014 - Art. 3º A disciplina do uso da internet no Brasil tem os seguintes princípios: I - garantia da liberdade de expressão, comunicação e manifestação de pensamento, nos termos da Constituição Federal; II - proteção da privacidade; III - proteção dos dados pessoais, na forma da lei;

³⁰ Lei nº 13.709/2018 - Art. 5º Para os fins desta Lei, considera-se: I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável;

observância aos cuidados relativos à Segurança da Informação considerando para as publicações a não identificação e todos os cuidados éticos envolvidos na pesquisa³¹.

Embora o legislador tenha flexibilizado a atenção às regras para não dificultar o desenvolvimento de pesquisa no Brasil, cuidados precisam ser observados sob pena de responsabilização judicial, conforme determina o artigo 42 da mesma lei³².

5.4.2.3 Aspectos Jurídicos da Pesquisa

Muito se fala sobre bioética em estudos clínicos envolvendo seres humanos. No entanto, nem todos os estudos da área da saúde envolvem diretamente o ser humano, como é o caso deste estudo. Nesse caso, é importante falarmos de outros aspectos jurídicos que podem gerar riscos ao resultado da pesquisa.

A Resolução 466/2012 do Conselho Nacional de Saúde – CNS estabelece regras bioéticas com objetivo de assegurar os direitos e deveres que dizem

³¹ Lei nº 13.709/2018 - Art. 13. Na realização de estudos em saúde pública, os órgãos de pesquisa poderão ter acesso a bases de dados pessoais, que serão tratados exclusivamente dentro do órgão e estritamente para a finalidade de realização de estudos e pesquisas e mantidos em ambiente controlado e seguro, conforme práticas de segurança previstas em regulamento específico e que incluam, sempre que possível, a anonimização ou pseudonimização dos dados, bem como considerem os devidos padrões éticos relacionados a estudos e pesquisas.

³² Lei nº 13.709/2018 - Art. 42. O controlador ou o operador que, em razão do exercício de atividade de tratamento de dados pessoais, causar a outrem dano patrimonial, moral, individual ou coletivo, em violação à legislação de proteção de dados pessoais, é obrigado a repará-lo.

§ 1º A fim de assegurar a efetiva indenização ao titular dos dados:

I - o operador responde solidariamente pelos danos causados pelo tratamento quando descumprir as obrigações da legislação de proteção de dados ou quando não tiver seguido as instruções lícitas do controlador, hipótese em que o operador equipara-se ao controlador, salvo nos casos de exclusão previstos no art. 43 desta Lei;

II - os controladores que estiverem diretamente envolvidos no tratamento do qual decorreram danos ao titular dos dados respondem solidariamente, salvo nos casos de exclusão previstos no art. 43 desta Lei.

§ 2º O juiz, no processo civil, poderá inverter o ônus da prova a favor do titular dos dados quando, a seu juízo, for verossímil a alegação, houver hipossuficiência para fins de produção de prova ou quando a produção de prova pelo titular resultar-lhe excessivamente onerosa.

§ 3º As ações de reparação por danos coletivos que tenham por objeto a responsabilização nos termos do caput deste artigo podem ser exercidas coletivamente em juízo, observado o disposto na legislação pertinente.

respeito aos personagens participantes de pesquisas em saúde, mas nenhuma proteção individual deve impactar o desenvolvimento e os direitos de uma nação.

Além da ética, competências políticas e técnicas são também exigidas do pesquisador. Além de conhecimento técnico para conduzir a pesquisa, é exigida do pesquisador prevenção de más condutas científicas como plágio, fabricação e produção de soluções sem a inovação adequada [45].

O *The Leiden Manifesto for research metrics* (Manifesto de Leiden sobre métricas de pesquisa) é um manifesto com 10 princípios para a medição do desempenho de pesquisas publicado na *Nature*³³ com propósito de apoiar pesquisadores a manter a qualidade das suas pesquisas. O manifesto deixa clara a importância dos dados para a ratificação dos resultados encontrados em estudos quando as técnicas e métricas são utilizadas, normalmente bem intencionadas, nem sempre bem informadas, muitas vezes mal aplicadas [45].

Isso quer dizer que a condução da pesquisa, as condições institucionais a responsabilidade social dos pesquisadores e conseqüentemente os resultados obtidos da pesquisa são responsabilidades inerentes do pesquisador [45]. Se, por acaso, no curso da pesquisa ou posteriormente os resultados da pesquisa causarem prejuízos a terceiros, certamente o pesquisador responderá pelos danos causados.

No Brasil, as pesquisas devem seguir as orientações da Comissão Nacional de Ética em Pesquisa - CEPs³⁴ e as resoluções números 240/1997, 466/2012 e 370/2007 do CNS. Nesse sentido, os participantes, grupos ou seus respectivos representantes legais, têm o direito de conhecer todas as etapas e informações da pesquisa em que se envolvem, atraindo para o pesquisador, de forma automática, o dever de informar [46].

De acordo com a Resolução 466/2012 do CNS:

II.23 - Termo de Consentimento Livre e Esclarecido - TCLE - documento no qual é explicitado o consentimento livre e esclarecido do participante e/ou de seu responsável legal, de forma escrita, devendo conter todas as informações necessárias, em linguagem clara e

³³ <https://www.nature.com/> - acessado em 07 de novembro de 2021 às 09:54.

³⁴ <http://conselho.saude.gov.br/comites-de-etica-em-pesquisa-conep>

objetiva, de fácil entendimento, para o mais completo esclarecimento sobre a pesquisa a qual se propõe participar;

II.24 - Termo de Assentimento - documento elaborado em linguagem acessível para os menores ou para os legalmente incapazes, por meio do qual, após os participantes da pesquisa serem devidamente esclarecidos, explicitarão sua anuência em participar da pesquisa, sem prejuízo do consentimento de seus responsáveis legais;

O Termo de Consentimento Livre e Esclarecido – TCLE é, sem dúvida, uma das ferramentas mais importantes para proteção jurídica do pesquisador. A ferramenta trata de assunto relacionado à dignidade do ser humano, o que pode ser visto historicamente desde o Código de Nuremberg e a divulgação pela Declaração Mundial dos Direitos Humanos na Assembleia Geral das Nações Unidas.

O que deve ser observado é que para cada pesquisa deve ser desenvolvido um TCLE específico, com as informações e peculiaridades daquela situação específica sob risco do documento ser juridicamente invalidado. Ora, se o documento se presta para informar o participante da pesquisa, não faz sentido utilizar um documento padrão, por exemplo, porque, ali, não estariam expressas as condições e informações daquela pesquisa em especial.

Conforme Souza et al. (2013), os sujeitos da pesquisa, independentemente da etapa, devem ser amplamente informados, não se sentir pressionado ou coagido. Diante disso, além do processo operacional da pesquisa, os dados utilizados e os resultados obtidos devem, também, constar no rol de informações da pesquisa.

Além do TCLE, os Termos de Responsabilidades devem ser utilizados sobretudo quando houver participação direta de informações ou interferências de instituições. Lembrando que o Termo de Responsabilidade pode ter caminho inverso, isto é, para conceder informações importantes à pesquisa, como base de dados, por exemplo, as empresas através do jurídico podem exigir que os pesquisadores assinem termo de responsabilidade, confidencialidade e não concorrência. Todos esses são termos que funcionam como “acordos” pré-estabelecidos entre as partes que têm, a partir da assinatura, força de lei entre as partes (*pacta sunt servanda*) [47].

5.4.3 Direito à Proteção de Dados e Seus Limites

Um dos grandes problemas atuais diz respeito justamente a instrumentalizar instituições privadas de informações e essas utilizarem a LGPD para negar acesso às informações de interesse público.

Além das informações com caráter legal cadastradas diariamente pelas organizações, informações de comportamento social são geradas através de processos de inteligência. Essas informações poderiam ser utilizadas para o desenvolvimento de pesquisa de interesse público, como é o caso da área da saúde, por exemplo.

Muito se fala em Direito de Acesso à Informação de dados gerados por órgãos públicos, e pouco se fala do acesso de informações geradas pelas instituições privadas. Justamente porque essas informações têm caráter estratégico comercial para as organizações, têm valor econômico. É importante ressaltar aqui que as informações geradas internamente são individuais (da pessoa jurídica), e será que as informações geradas a partir do comportamento social das pessoas possivelmente identificáveis não são informações de interesse social e, portanto, coletivo? Quanto mais quando forem utilizados para o desenvolvimento de pesquisas que interessam ao coletivo.

A título de exemplo podemos citar empresas de telecomunicação, aviação, abastecimento de água e esgoto, iluminação pública e de comércio. Essas empresas existem justamente para suprir necessidades sociais e para isso, utilizam recursos para a pesquisa e desenvolvimento de tecnologias capazes de identificar, segmentar e otimizar a oferta dos serviços ao seu público.

Certamente grandes bases de dados são geradas para o desenvolvimento dessas pesquisas internas, como o consumo de energia elétrica em um determinado bairro, o horário de maior fluxo de pessoas em determinado segmento comercial, o comportamento social nos finais de semana e feriados e assim por diante.

É importante destacar que todos esses dados e as informações geradas por eles já existem nas inúmeras bases de dados espalhadas pelo setor privado.

Ao aproveitar essas informações, sem identificação pessoal, haveria uma contribuição segmentada, estruturada e otimizada para o desenvolvimento de ciência em diversas áreas do conhecimento e, nesse caso, também seria para a área da saúde.

Nenhuma lei, *prima facie*, pode cercear o direito universal do conhecimento e, para se construir o conhecimento é fundamental o acesso às informações. O que se quer apontar aqui é que a LGPD vem para limitar o uso irrestrito de informações pessoais trazendo insegurança às pessoas. No entanto, a pesquisa não pode ser penalizada pela falta de responsabilidade do uso de dados pessoais.

Ainda mais quando o assunto é a saúde. A saúde é um dever do estado, elencada como um direito básico da Constituição Federal, para garantir saúde, disponibilizar meios adequados de acesso e, mais do que isso, garantir informações adequadas à população é fundamental o trabalho da pesquisa.

Os dados, ou primários (aqueles intimamente ligados a um indivíduo) ou os secundários (aqueles gerados a partir da análise de um grupo de pessoas), não podem e não devem permanecer disponíveis somente àquelas instituições que os recolhem e armazenam. O Estado, como fiscalizador e como interessado nas informações para garantir os direitos básicos, precisa ter ferramentas de acesso aos dados, principalmente para que a pesquisa seja desenvolvida sem entraves.

Um dos grandes problemas do Brasil está justamente na disponibilidade de dados e informações aos grupos de pesquisa. É um paradoxo. Enquanto os legisladores estão trabalhando para proteger a individualidade das pessoas, sobretudo no compartilhamento de informações pessoais para uso mercantilista, é necessário olharmos para os anônimos que pesquisam e desenvolvem soluções para os nossos problemas.

A pesquisa acadêmica, afinal, é quem sofre com as limitações impostas por legislações que proíbem a propagação de dados. E aqui precisamos ter um cuidado. É necessário deixar claro que não há crítica ou contrariedade à legislação de proteção de dados pessoais, mas o cuidado de não punir aqueles

que precisam das informações e dos dados para buscar soluções coletivas, de interesse social.

Infelizmente hoje não é possível acessar dados básicos que, em países desenvolvidos, são públicos sem identificar pessoas. Exemplos como quantidade de pessoas que utilizam a malha aeroviária diária, de onde partem e qual é o seu destino são problemas. Conforme estudo de Codeço et al. (2020), hoje no Brasil não é possível saber com precisão o aeroporto de origem e o aeroporto de destino: o que se tem são informações entre cada voo, e sabemos que, como o Brasil tem uma extensão territorial grande, as conexões e escalas fazem parte do cotidiano dos passageiros aeroviários.

As empresas aéreas, por óbvio, têm essas informações, mas a Agência Nacional de Aviação Civil – ANAC não. Por quê?

De maneira prática, a LGPD vem para limitar ainda mais a disponibilidade dos dados que estão em poder dessas empresas particulares. Mas, se essas informações têm relevância para a pesquisa e para soluções públicas, as autarquias não deveriam ter acesso a essas informações, e ainda vou mais adiante: as autarquias, além de receber essas informações, não deveriam disponibilizá-las aos centros de pesquisa?

Até onde as informações pessoais são verdadeiramente particulares?

Um portador de HIV pode não querer ter essa informação publicada. Mas será que essa informação não tem relevância pública quando se fala em saúde pública? As informações particulares não interessam quando não comprometem o direito do outro, como é o caso de informações como CPF, cor do olho, gosto musical, etc.

Assim como o exemplo do portador de HIV, temos aqui os portadores de carga ativa da Covid-19. Embora as recomendações dos órgãos de saúde sejam para que pessoas contaminadas façam quarentena, ainda se sabe pouco sobre a doença, e se sabe menos ainda sobre como ela se manifesta em cada pessoa. Será que, ao expor outra pessoa aos riscos da contaminação, o direito individual

da pessoa contaminada prevalece às garantias de sigilo do indivíduo contaminado?

É uma questão subjetiva de bioética. Mas a ciência precisa de dados para encontrar soluções. Há meios de se garantir proteção à identificação das pessoas e, mesmo assim, utilizar de forma pública as informações relevantes que aquela pessoa gera.

6 MÉTODO

De acordo com as definições de Brasil. Ministério da Saúde (2017) essa é uma pesquisa qualitativa classificada como exploratória quanto aos objetivos e experimental em relação à sua natureza. Como possibilidades metodológicas, a pesquisa qualitativa é aberta a diversas possibilidades, permitindo a construção dinâmica para a coleta e análise dos dados envolvidos [49].

O grande desafio do estudo é: como resolver o problema? Para a busca da solução ideal, é importante o conhecimento sobre as áreas que envolvem o problema.

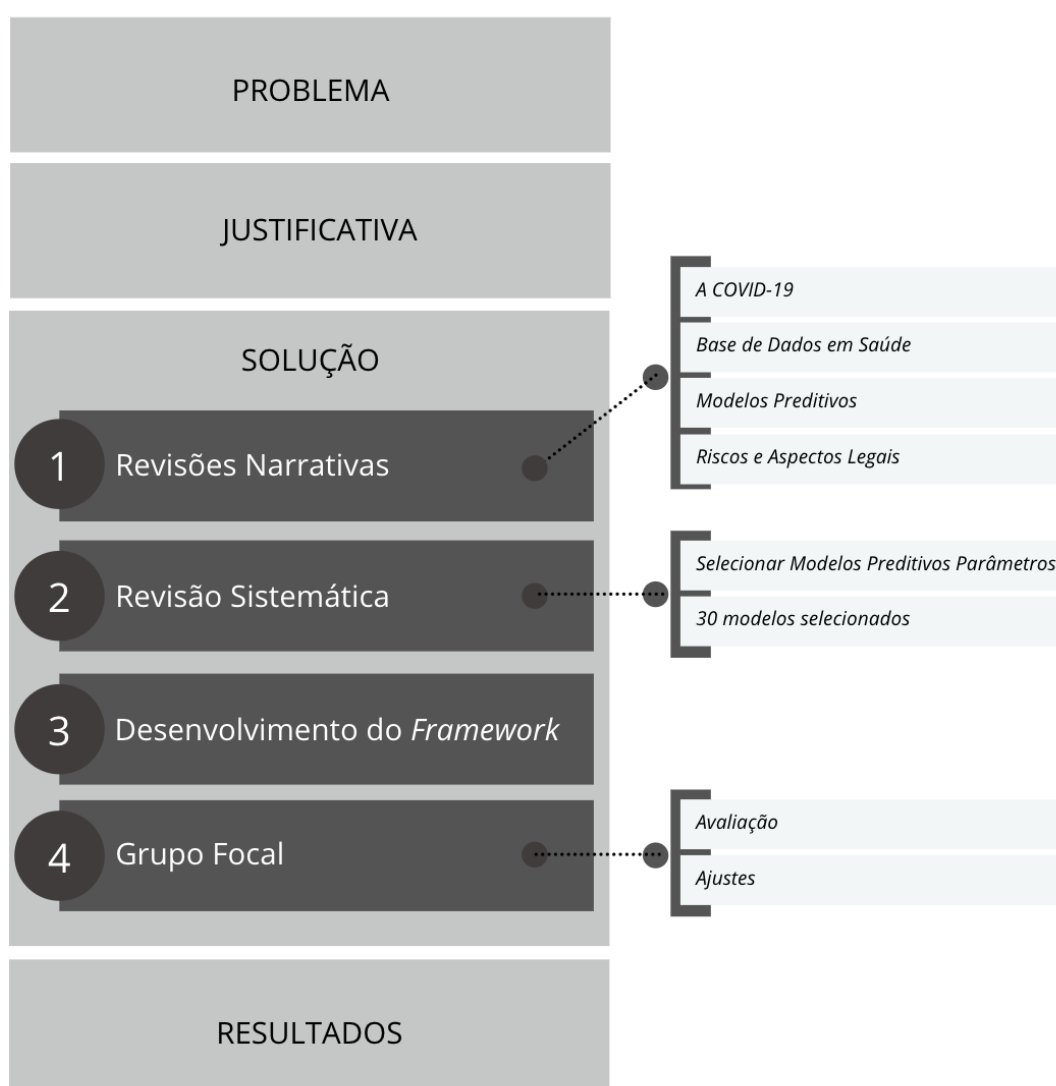


Figura 5 - Fluxo da metodologia de pesquisa.

Fonte: Autor.

6.1 REVISÃO NARRATIVA

A pesquisa iniciou com revisões narrativas, no contexto da pandemia da Covid-19, sobre os principais assuntos abordados no estudo: a doença Covid-19, bases de dados em saúde, modelos preditivos e aspectos jurídicos de pesquisas em saúde.

Diferente das revisões de literatura, as revisões narrativas utilizam fontes bibliográficas para obtenção de resultados de pesquisas de outros autores com propósito de fundamentar o objetivo do estudo [50]. Os textos foram explorados de forma a conceituar, classificar e direcionar o leitor para a compreensão de todos os assuntos abordados na pesquisa.

6.2 REVISÃO SISTEMÁTICA

Uma das formas de identificar as principais características dos modelos preditivos aplicados à propagação da Covid-19 é justamente analisar estudos já publicados. O propósito dessa etapa foi selecionar, com base em critérios bem definidos, modelos preditivos já propostos pela literatura, através de uma revisão sistemática.

A revisão sistemática é uma pesquisa secundária com objetivo de identificar os principais estudos e materiais de um tópico específico a partir de estudos primários [51].

Com base no protocolo estabelecido por Kitchenham (2017), foram identificados, selecionados e analisados modelos matemáticos preditivos para a análise e predição da propagação da Covid-19. No estudo, foram observadas características específicas relacionadas (a) ao local em que o estudo foi realizado para compreender os elementos específicos de localidade, (b) às variáveis de entrada utilizadas para subsidiar o modelo matemático e (c) à forma como os resultados são apresentados para o público a que se destina, objetivando entender se os modelos consideram objetivos específicos.

O protocolo foi desenvolvido e aplicado seguindo as recomendações de Kitchenham (2017) por uma equipe multidisciplinar envolvendo profissionais das principais áreas envolvidas na pesquisa, sendo dois cientistas da computação,

dois médicos e uma profissional da comunicação social, e posteriormente foi verificado por Moher et al. (2009) e Brasil. Ministério da Saúde (2017) para diminuir o risco de viés na seleção dos modelos preditivos analisados.

As seguintes etapas foram desenvolvidas na revisão: definição das questões de pesquisa, elaboração da estratégia de busca, a identificação de critérios de inclusão e exclusão, a definição e aplicação dos critérios de qualidade, a extração dos dados e a análise e síntese dos dados.

As questões de pesquisa objetivaram identificar as principais características dos modelos matemáticos encontrados na literatura para verificar a existência de características específicas a serem consideradas para selecionar o modelo matemático mais adequado para a predição da propagação da Covid-19 em situações específicas. O estudo respondeu, para cada artigo selecionado as perguntas abaixo:

QP1- Quais são os modelos matemáticos utilizados para a predição da propagação do novo coronavírus?

QP2- Quais são as variáveis de entrada utilizadas nos modelos analisados?

QP3- Qual a origem das informações de entrada utilizadas nos modelos matemáticos analisados (*datasets* reais ou simulados)?

QP4- Quais são os aspectos positivos e as limitações encontrados em cada um dos modelos analisados?

QP5- Para quais contextos os modelos analisados foram desenvolvidos?

QP6- Como os resultados de cada modelo são apresentados?

QP7- O modelo analisado tem como base outro modelo já existente?

Para o desenvolvimento da expressão de busca, foram consideradas três grandes áreas do conhecimento e, a partir daí, derivaram-se expressões de cada área, conforme é possível verificar na Figura 3. A expressão final de busca foi

aplicada nas seguintes bases: Pubmed, IEEE, Scopus, Embase e *Web of Science*.

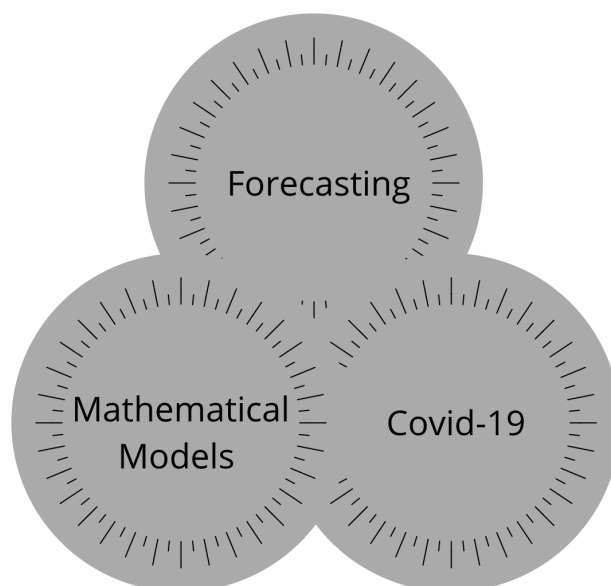


Figura 6 - Áreas de estudo da Revisão Sistemática.

Fonte: Autor.

A partir desse macrodiagrama a seguinte expressão foi criada:

{(COVID-19 **OR** "new coronavirus" **OR** 2019-nCoV **OR** Sars-CoV2 **OR** Sars-CoV-2) **AND** ("mathematical model" **OR** "mathematical modeling" **OR** "mathematical models" **OR** "mathematical modelling") **AND** (prediction **OR** prognosis **OR** spread **OR** prevention **OR** forecasting)}.

A expressão resultou 523 publicações divididas conforme apresentado na Tabela 1:

Tabela 1 - Base de Pesquisa.

Search String application		
Base	Application Date	Output
IEEE	23/08/20	8
EMBASE	23/08/20	84
WEB OF SCIENCE	21/08/20	96
SCOPUS	23/08/20	217
PUBMED	23/08/20	118
		523

Os Critérios de Inclusão (CI) definidos foram: (CI1) artigos publicados de 01/12/2019 até 23/08/2020, (CI2) nos idiomas inglês, espanhol, alemão ou português e (CI3) publicações do tipo artigo. Eles estabelecem as características mínimas para a aceitabilidade do artigo como ponto de partida e utilização dos filtros oferecidos nas próprias bases de busca. Para refinar o filtro, na tentativa de buscar aqueles trabalhos que se adaptam às características desejadas, se definiram os Critérios de Exclusão (CE). Para adequar o protocolo ao método PRISMA, os artigos duplicados foram excluídos antes da aplicação dos critérios de exclusão [52] e, a partir daí, aplicaram-se os seguintes critérios de exclusão: (CE1) são repetidos, (CE2) ainda não foram efetivamente publicados pelo periódico, (CE3) possuam data de publicação futura, (CE4) sejam revisões sistemáticas ou revisões integrativas, (CE5) não especifiquem os dados de entrada, (CE6) não apliquem o modelo em um *dataset* específico, (CE7) não sejam com humanos, (CE8) sejam modelos para um cenário muito específico ou que não contemplem um modelo matemático e (CE9) aqueles que não abordassem modelos específicos para a Covid-19.

Os Critérios de Qualidade (CQ) foram definidos para analisar as validades internas e externas dos artigos incluídos e eliminar possíveis tendências, a partir da escala de aplicação de oito critérios pré-definidos. Para cada critério de qualidade, os artigos foram classificados com 1 (atende o critério) ou 0 (não atende o critério). (CQ1) quando o objetivo do artigo é claro, objetivo e bem definido, (CQ2) para os modelos que explicam individualmente cada variáveis de entrada, (CQ3) modelos do tipo compartimental que utilizem mais de três

grupos, (CQ4) quando o artigo explicar a construção do modelo matemático de forma que o leitor entenda a lógica do script matemático ali proposto, (CQ5) quando os modelos matemáticos detalharem as informações da base de dados, (CQ6) quando o *dataset* for facilmente acessado por qualquer pessoa, (CQ7) quando especifique a data de início e fim, do local específico e das peculiaridades especiais dos dados utilizados para a aplicação no modelo proposto, (CQ8) quando o artigo apresentar os resultados obtidos a um público específico, de maneira direcionada e não de forma genérica.

Após serem aplicados os critérios de qualidade, os trabalhos analisados foram classificados com pontuação entre 1 e 8, sendo selecionados aqueles estudos que somaram pontuação entre 5 e 8, o que resultou em 19 artigos. Dezenove artigos ficaram com pontuação 4. Desses artigos, classificaram-se para a revisão aqueles trabalhos que tiveram qualquer pontuação nos critérios CQ5, CQ6, CQ7 e CQ8, resultando em 6 artigos e alcançando um total de 25 artigos selecionados.

Vinte e cinco modelos matemáticos foram selecionados e analisados de acordo com os critérios estabelecidos pelos autores e um artigo foi escrito com os resultados obtidos (Apêndice A).

Como os resultados obtidos na revisão sistemática não contemplaram modelos desenvolvidos para o cenário brasileiro, modelos preditivos nesse contexto foram somados ao conjunto através de uma busca exploratória. Cinco estudos foram selecionados com base em critérios de importância identificados em conjunto com os orientadores.

O primeiro modelo escolhido foi desenvolvido pelo Observatório Nacional³⁵, pela colaboração de pesquisadores de vinte e oito instituições brasileiras que reuniram esforços para divulgar informações de qualidade baseadas em análises cientificamente embasadas. Um modelo linear ajustado utiliza valores médios de contaminação e probabilidade de uma pessoa saudável contrair a doença dado contato com indivíduo contaminado [54].

³⁵ <https://covid19br.github.io/>

O segundo modelo preditivo foi desenvolvido pela Universidade Federal de Viçosa³⁶ e refere-se a um modelo SEAIR – *Susceptible, Exposed, Asymptomatic, Infectious e Recovered* (modelo preditivo que apresenta cinco grupos de pessoas: suscetíveis, expostos, assintomáticos, infectados e recuperados). O modelo divide a população em sete grupos, analisados, por sua vez, em subgrupos de idade. O modelo é baseado em uma população robusta de valores absolutos o que permite aproximar da realidade as taxas e parâmetros calculados [55].

O terceiro trabalho foi desenvolvido por pesquisadores da Fiocruz³⁷, que abordam a relação entre o contágio e o tráfego aéreo entre as microrregiões brasileiras, determinando a probabilidade da disseminação da Covid-19 pelo Brasil. Essa probabilidade é dada por uma expressão originada do modelo SIR. Foi calculada uma medida associada à probabilidade de importação de casos para microrregiões por via aérea, que mede a proximidade em termos de fluxo de pessoas. Como esperado, quanto menor a distância efetiva entre as microrregiões, menor o tempo até invasão. Para estimar o tempo até disseminação de um vírus para cada destino desde a introdução do mesmo em uma origem *i* se utilizou a distância efetiva calculada, o período de infecção de oito dias e a taxa de reprodução básica, extraída do modelo SIR [56] [23].

Os trabalhos desenvolvidos pela Universidade Federal do Rio Grande do Norte - UFRN³⁸ e pela Universidade Federal de Brasília - UNB³⁹ utilizam modelos compartimentais, com algumas diferenças entre eles. No trabalho apresentado pela UFRN, as populações são divididas entre suscetíveis (S), infectadas (I), recuperadas (R) e mortos (D), e cada um desses grupos é dividido por idade, com um total de 9 subgrupos etários [57]. No trabalho da UNB, a população também foi dividida em (outros) nove grupos etários, além das subpopulações suscetíveis (S), hospitalizadas devido à Covid-19 (H), expostas (em período de incubação e não contagiosas), infectadas (contagiosas, I) e recuperados (R). Nesse modelo, as subpopulações representam a fração do grupo etário [58].

³⁶ <https://www.ufv.br/>

³⁷ <https://portal.fiocruz.br/>

³⁸ <https://www.ufrn.br/>

³⁹ <https://www.unb.br/>

Após a análise minuciosa dos trinta modelos selecionados, uma lista de variáveis e parâmetros utilizados nos trabalhos foi desenvolvida. Esses dados foram devidamente listados e classificados de acordo com o tipo, origem, limitações e observações.

A partir dessa lista de variáveis, foi possível identificar os principais dados utilizados nos modelos matemáticos para predição e propagação da Covid-19 e conseqüentemente os riscos associados à utilização daqueles dados.

6.3 DESENVOLVIMENTO DO FRAMEWORK

Com base nas informações levantadas, um framework de análise de riscos para modelos preditivos foi desenvolvido para auxiliar os pesquisadores no desenvolvimento ou seleção de modelos preditivos (Apêndice C). O desenvolvimento se ancorou na acessibilidade, eficiência e eficácia para que pesquisadores de diferentes áreas do conhecimento possam ser auxiliados nas tomadas de decisões considerando, sobretudo, os potenciais riscos envolvidos nas suas pesquisas.

Framework é um termo que é definido de diversas maneiras pela bibliografia, no entanto uma estrutura de artefatos que constitui um problema abstrato para a solução de problemas definidos, definição dada por Johnson e Foote (1988), é a que melhor representa o termo no contexto deste estudo.

O framework foi estruturado de maneira análoga a uma árvore de decisão associando elementos de técnicas preditivas, estatísticas e estruturais de dados com características dos valores de dados disponíveis para aquele determinado estudo. A estrutura foi dividida em três fases dependentes, que guiam o avaliador na descoberta e análises de riscos do modelo preditivo analisado através de escalas e questionamentos dissertativos. Ao final, um relatório é gerado com observações gerais, potenciais riscos e sugestões que objetivam a mitigação dos riscos mapeados.

6.4 GRUPO FOCAL

De acordo com Trad (2009), o Grupo Focal é uma ferramenta de pesquisa qualitativa baseada na coleta de informações através da interação entre os participantes de pequenos grupos.

A capacidade que a ferramenta tem de fazer com que os participantes interajam e problematizem um assunto foco, sob seus pontos de vista, torna o grupo focal importante técnica de coleta e análise de dados [49].

Nesta pesquisa, o Grupo Focal foi utilizado como técnica para avaliar o framework desenvolvido e complementar as pesquisas explorativas realizadas. Um protocolo foi criado para a realização do Grupo Focal em que foram estabelecidos os critérios para seleção dos participantes, forma de encontro, tempo de duração e questões a serem discutidas.

Os participantes foram definidos de acordo com a respectiva área de atuação e a participação no programa de pós-graduação para não haver necessidade de submissão do projeto ao Comitê de Ética.

Após sua definição, o protocolo foi enviado para os participantes por e-mail.

6.4.1 Protocolo

Quando: dia 30 de setembro de 2021, às 14h.

Participantes:

- Uilian Loose - organizador
- Juliana Herbert – Cientista da Computação
- Cláudia Codeço – Pesquisadora de Modelos Preditivos
- Aline Griza – Comunicadora Social
- Andreas Lutz - Médico
- Graziela – Bióloga e Médica

Como: virtual, através da ferramenta Google Meet.

Condições Gerais: uma mensagem solicitando a autorização para gravação foi enviada antecipadamente para todos os participantes. Caso houvesse objeção de algum dos participantes para a gravação a comunicação deveria ser feita de forma escrita. Todos os participantes concordaram com a respectiva gravação.

- a) Apresentação do framework – apresentação realizada pelo organizador (10 minutos).
- b) Feedback dos participantes a respeito do que entenderam, proposta do framework e contextualização, seguidas de proposições para responder as seguintes perguntas:
 - A linguagem do framework é acessível a todas as áreas de conhecimento?
 - O processo (fluxograma) é contínuo e apresenta todas as possibilidades de análise do modelo matemático?
 - O framework não é recomendado para algum caso específico na área da saúde?
 - Você utilizaria o framework como protocolo de validação de um modelo matemático?
 - Você utilizaria o framework como protocolo de identificação de riscos jurídicos?

6.4.2 Resultados do Grupo Focal

Inicialmente, essa pesquisa considerava apenas modelos matemáticos para predição da propagação no contexto da Covid-19. No entanto, uma das conclusões do Grupo Focal foi que o framework analisado se aplica a qualquer modelo preditivo (computacional, estatístico, matemático, etc.).

Com base nessa conclusão, decidiu-se ajustar o nome do framework e as suas limitações de aplicabilidade. Muito embora os estudos teóricos iniciais tenham utilizado como filtro modelos matemáticos preditivos, como a revisão sistemática realizada para selecionar os modelos utilizados como parâmetro para o estudo, não há nenhum prejuízo para o resultado da pesquisa.

Questões a respeito dos resultados apresentados pelo framework no desenvolvimento prático de pesquisa foram discutidas e dúvidas sobre aspectos legais em pesquisa foram levantadas. Entendeu-se que, de alguma maneira, dúvidas comuns envolvendo responsabilidades do pesquisador deveriam, de alguma maneira, serem contempladas no framework.

Dessa forma, informações comuns sobre responsabilidades do pesquisador, do financiador, de propriedade da base de dados serão incluídas no framework independentemente se aquela informação se aplica ou não ao modelo aplicado pelo avaliador.

7 RESULTADOS E DISCUSSÕES

7.1 O FRAMEWORK

O framework é uma diretriz constituída com base em um conjunto de artefatos com objetivo de avaliar variáveis usadas em modelos matemáticos para predição da propagação da Covid-19.

A pandemia da Covid-19 foi utilizada, nesse projeto, como propulsora do desenvolvimento do framework, quando veio à tona um grande problema enfrentado pelos pesquisadores brasileiros: o acesso aos dados indispensáveis para a produção de informações e pesquisa. O problema é ratificado pelo momento legal brasileiro. Recentemente entrou em vigor no Brasil a Lei Geral de Proteção de Dados – LGPD, que surge nesse cenário de necessidade de desenvolvimento de pesquisas na área da saúde com objetivo de regular as normas de acesso às informações pessoais, fortalecendo no setor privado ainda mais os interesses econômicos vinculados às informações das suas bases de dados.

Esse framework, Apêndice C, tem a função de auxiliar na estrutura e planejamento de diretrizes para o pesquisador que atua em pesquisas envolvendo análises preditivas na área da saúde, trazendo subsídios técnicos multidisciplinares, que, normalmente, são de áreas distintas da área de origem do pesquisador.

O framework contém artefatos estruturados que orientam a avaliação dos modelos matemáticos a partir das informações obtidas em cada etapa da ferramenta, mapeando a proposta do desenvolvedor e os aspectos jurídicos relacionados àquele modelo matemático.

A proposta é que os avaliadores sejam capazes de mapear os riscos envolvidos na utilização do modelo avaliado. Esse processo, além de auxiliar os pesquisadores que desenvolvem os modelos, auxilia o público que deseja utilizar um determinado modelo, mas não possui critérios para avaliá-lo de forma estruturada, transparente e segmentada.

Ressalta-se aqui a importância dos modelos de predição para a ciência e, sobretudo, para o desenvolvimento da saúde. São eles que, muitas vezes, fornecem sínteses do estado do conhecimento (estado da arte), a partir do qual decisões são tomadas, impactando a saúde coletivamente.

Os aspectos legais para o acesso e utilização dos dados em saúde são desafiadores para a pesquisa brasileira. Além das normas colaborarem para a dificuldade de acesso aos dados indispensáveis para a pesquisa, o próprio pesquisador, muitas vezes, pode estar vulnerável na manipulação de determinados dados, inclusive podendo ser responsabilizado.

O que se vê na área científica hoje é uma grande dificuldade de acesso às informações da população que se deseja estudar as quais, por vezes, são permeadas por questões complexas que envolvem estratégia de mercado, economia, interesse político e desinformação, que foram agravadas pelo novo marco regulatório brasileiro referente à proteção de dados pessoais.

Os dados referentes à saúde da população são especialmente afetados, porque a LGPD os classifica como sensíveis, e esses apenas podem ser tratados para fins específicos. Esse pressuposto é levado em consideração pelas grandes organizações, como justificativa para não disponibilizarem dados essenciais para a pesquisa em saúde.

Embora a própria lei regule a disponibilização de dados para que sejam realizados estudos por órgão de pesquisa, sabemos que, economicamente falando, não é interessante para as empresas a divulgação de suas bases de dados, provocando, sob justificativa da LGPD, um retrocesso da pesquisa brasileira.

O que não pode acontecer é que a legislação que vem para proteger a intimidade das pessoas no que diz respeito à divulgação de dados pessoais prejudica ou não dá condições necessárias para o desenvolvimento de pesquisas que são de interesse coletivo, ou melhor, ações que definitivamente podem salvar vidas, como é o caso da Covid-19. Quando esse interesse coletivo é maior do que os interesses individuais, há aí uma questão social a ser discutida. Os interesses coletivos precisam ser levados em consideração e, mais

do que isso, precisam nortear ações de políticas públicas capazes determinar que a pesquisa em saúde, na maioria das vezes, deverá prevalecer aos interesses econômicos.

Também é importante dizer que não é porque um determinado estudo utilizará informações sensíveis de um determinado grupo de pessoas que essas informações não serão protegidas ou que essas informações serão divulgadas sem o interesse direto ou sem o consentimento daquele indivíduo.

Uma das formas de auxiliarmos os cientistas no acesso às informações é justamente subsidiá-los de conhecimentos multidisciplinares para que, além de utilizarem como argumento técnico, tenham meios de indicar mudanças normativas em prol do desenvolvimento científico do país.

Em resumo, a lei não pode inviabilizar o desenvolvimento científico, tampouco deve servir de ferramenta contrária ao desenvolvimento técnico de uma sociedade, de forma a privilegiar interesses particulares em detrimento de interesses coletivos.

7.1.1 Planejamento do Framework

De forma alguma, o framework tem objetivo de ser exauriente. Inclusive, a proposta que se faz é que novas pesquisas e análises direcionadas para cada área do conhecimento sejam propostas por pesquisadores de modo que o framework seja aperfeiçoado e especializado para diferentes linhas de pesquisa.

Para o desenvolvimento do framework, limites e conceitos precisaram ser definidos para não levar o leitor a erro. De forma alguma esta pesquisa tem a intenção de ser uma pesquisa jurídica ou de enfrentamento de teorias de interpretação de normas. A proposta aqui é descortinar, de forma objetiva e estruturada, os desafios enfrentados pelos pesquisadores brasileiros na utilização de dados em saúde e a relação disso com os aspectos legais existentes. Entendimentos devidamente consolidados no universo jurídico serão utilizados para o desenvolvimento das sugestões apresentadas no modelo.

Por isso, coloca-se a relevância de limitar previamente os conceitos utilizados neste estudo para os termos: análise de riscos, modelos preditivos e público-alvo.

O termo análise de riscos utilizado nesta pesquisa se limita a identificar possíveis problemas que, de forma direta, possam atrair responsabilidades indesejáveis aos pesquisadores de modelos preditivos aplicados à saúde, aos financiadores de projetos de pesquisa ou os responsáveis pela disponibilidade dos dados utilizados para a pesquisa e os possíveis impactos negativos que os resultados gerados pelo modelo preditivo podem gerar à sociedade.

Entende-se por modelos preditivos a definição de Araújo (2002), “modelo é uma representação simplificada de uma situação concreta feita com o objetivo de compreender a situação e prever suas configurações futuras ou de situações semelhantes”, e de Goldbarg e Luna (2005), que definem os modelos como “representações simplificadas da realidade que preservam, para determinadas situações e enfoques, uma equivalência adequada”. Os modelos podem ser utilizados para simular situações específicas, considerando um conjunto de parâmetros pré-definidos estatísticos ou convencionados com base em regras ou podem ser utilizados para encontrar valores ideais de parâmetros, a partir do comportamento histórico do evento mapeado.

Por fim, com relação ao público-alvo, o framework foi desenvolvido com foco principal nos pesquisadores que atuam com modelagem preditiva em saúde. Como o contexto da saúde envolve competências multidisciplinares e, nem sempre é possível identificar os riscos envolvidos em pesquisas que envolvem dados sensíveis de saúde, a proposta é que o framework também alerte o pesquisador para aspectos relevantes de áreas do conhecimento que podem não ser do seu domínio.

Além dos pesquisadores, gestores públicos, elaboradores de diretrizes, organizações que apoiam a tomada de decisões e profissionais interessados em análises preditivas na área da saúde são públicos que podem se beneficiar com o framework.

Para fins práticos, neste estudo, toda a pessoa que aplicar o framework desenvolvido para análise de um modelo preditivo será denominado avaliador.

7.1.2 Desenvolvimento do Framework

O framework está dividido em três módulos dependentes entre si. Como uma “espinha dorsal”, o primeiro módulo é um macroprocesso baseado em etapas BPM (*Business Process Management*), elaborado para direcionar corretamente o avaliador por todas as etapas do framework de forma que o fluxo de análise seja simples, contínuo e seguro até o final do processo com a obtenção dos resultados.

Esse primeiro módulo, denominado **Processo**, define todo o processo do framework, que, na visão de Werneck e Teixeira (2009), é proposto como o conjunto de procedimentos organizados, que dita o modo como desenvolvemos atividades, de forma a transformarmos artefatos em um determinado produto que pretensamente tenha alguma utilidade.

O BPM (*Business Process Management*) é traduzido como Gerenciamento de Processos de Negócio, definido como “disciplina gerencial que integra estratégias e objetivos de uma organização com expectativas e necessidades de clientes, por meio do foco em processos ponta a ponta” [62]. É justamente o foco em processos ponta a ponta que se utiliza como estrutura principal do framework, de acordo com a Figura 7:

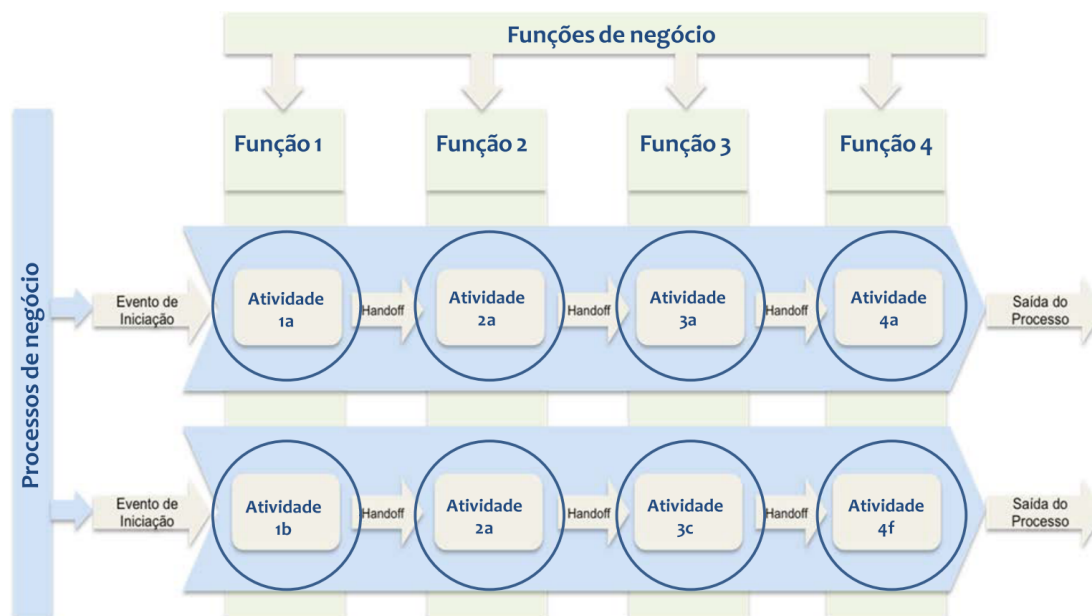


Figura 7 - Orquestração de atividades ao longo de funções de negócio.

Fonte: D. Moher et al., “Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement”, PLoS Med., vol. 6, no 7, 2009.

Neste trabalho, substituímos a palavra negócio por framework para adequar a realidade do projeto, sem qualquer prejuízo à ideia proposta por Association of Business Process Management Professionals (2013).

Os processos, por sua vez, são representados, por modelos de processos representados por símbolos que representam atividades, eventos, decisões e condições do processo [62]. Através dessa representação gráfica, foi possível expor a ordem e o funcionamento do framework.

O segundo módulo do framework é definido como **Análise** e tem o objetivo de provocar o avaliador e convidá-lo a analisar o modelo preditivo analisado de maneira crítica. Esse módulo, assim como o primeiro, é dividido em fases e cada fase é dividida em blocos.

É possível que o avaliador não percorra todos os blocos descritos no módulo **Análise**, e isso será determinado pelas decisões e orientações dispostas no módulo **Processo**.

Por fim, o último módulo denominado **Relatório de Riscos** tem a proposta de trazer os resultados esperados pelo avaliador. Todas as orientações dispostas nesse módulo são sugestivas, relacionando-as sempre com a decisão tomada pelo avaliador nos módulos anteriores.

Um guia foi elaborado para instruir o avaliador na utilização do framework e definir os conceitos, fases e processos utilizados.

7.1.2.1 Módulo Processo

É um macroprocesso, baseado notação BPM que orienta e direciona o avaliador pelo framework conforme Figura 8.

Anexo 1 - MÓDULO PROCESSO

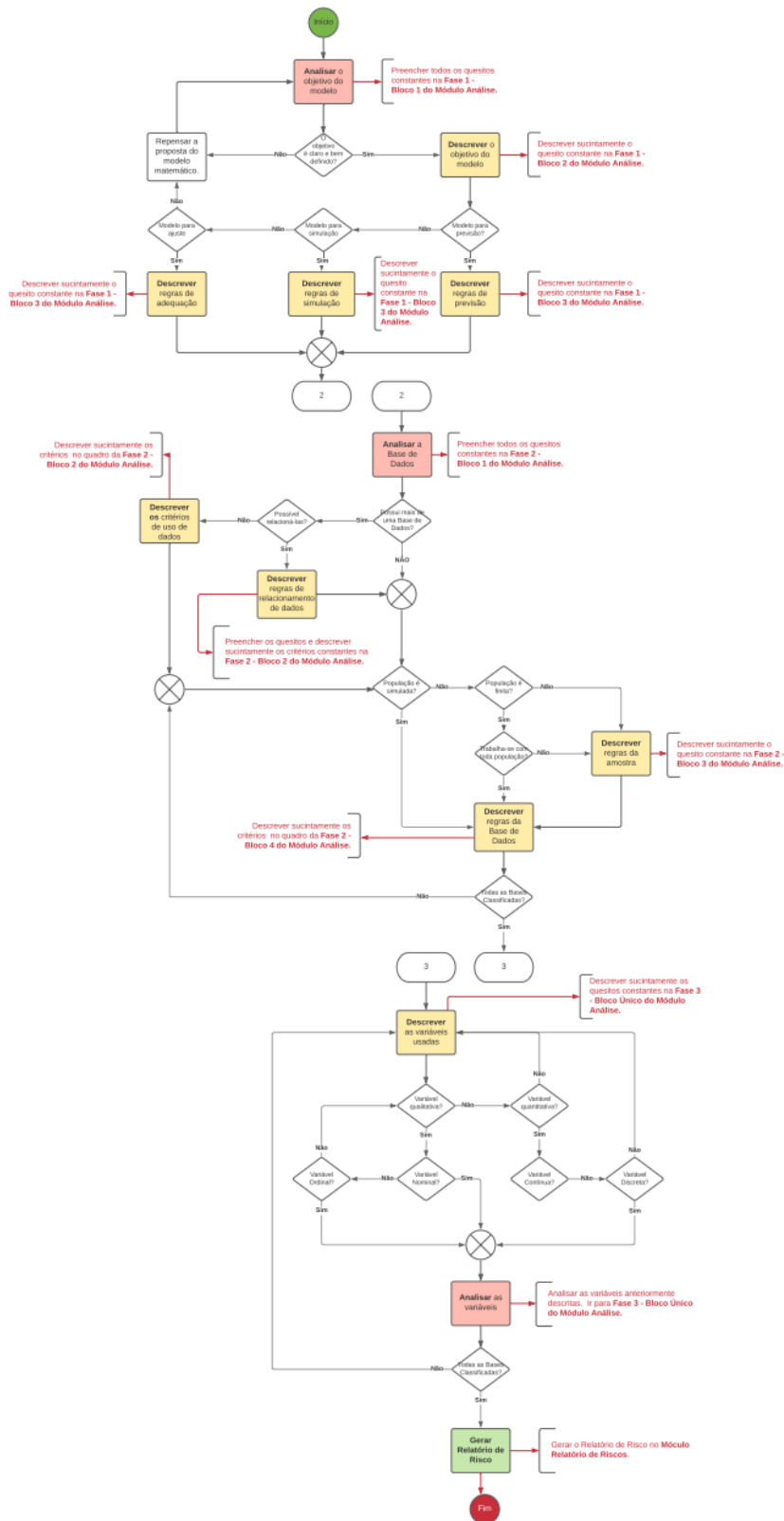


Figura 8 - Módulo Processo.

Para os elementos gráficos utilizados no módulo foram utilizadas as definições BPMN com base em um manual desenvolvido pelo Núcleo de Qualidade Interna (NQI)⁴⁰ da Universidade Federal de Ciências da Saúde de Porto Alegre - UFCSPA:

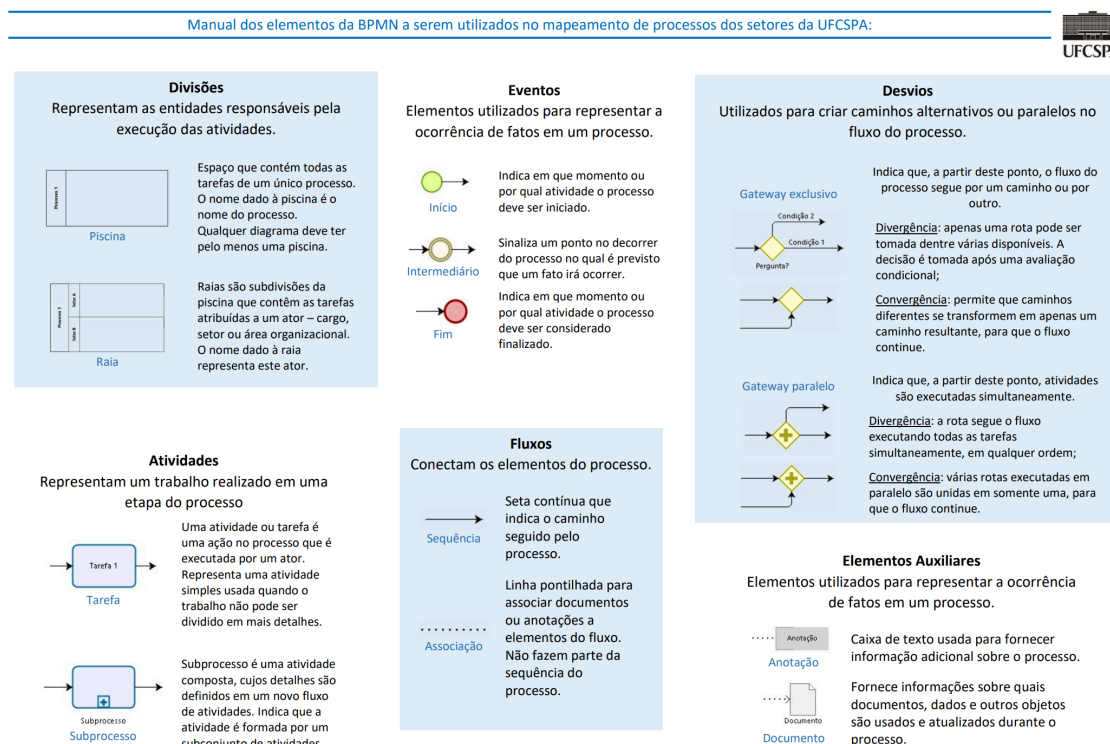


Figura 9 - Manual BPMN desenvolvido para o mapeamento de processos da UFCSPA.

Fonte: NQI da UFCSPA

Com base no elemento que representa uma atividade, apresentado no Manual BPMN desenvolvido pelo Núcleo de Qualidade da UFCSPA, elementos específicos foram desenvolvidos para o framework, para as ações a serem executadas pelo avaliador.

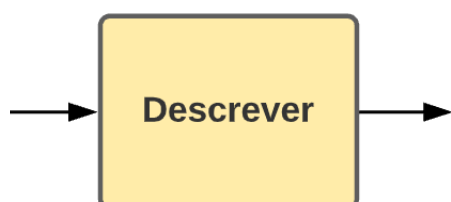
⁴⁰ <https://ufcspa.edu.br/documentos/nqi/guia-m3p-5ed.pdf> - Acessado em 15 de novembro de 2021 às 21:00.



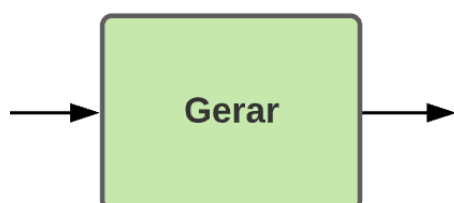
Representação gráfica padrão BPMN



A ação **analisar** redirecionará o avaliador para um bloco do módulo **Análise** sempre com opções de múltipla escolha, normalmente direcionado por uma escala de valores e cores.



A ação **descrever** direcionará o avaliador para blocos específicos do módulo **Análise**, onde precisará responder questionamentos de forma discursiva.



A ação **gerar** direcionará o avaliador para o módulo **Relatório de Riscos** onde será orientado a realizar o cruzamento das informações com o módulo **Análise**. Neste último módulo o avaliador constituirá, portanto, o Relatório de Riscos daquele modelo preditivo avaliado.

O módulo tem três fases e cada uma delas objetiva observar uma determinada especificidade do modelo preditivo analisado, o objetivo, a base de dados utilizada e as variáveis. Cada uma dessas fases será devidamente detalhada na seção módulo Análise.

7.1.2.2 Módulo Análise

Assim como o módulo Processo, o módulo Análise está dividido em três fases. Em cada fase o avaliador será provocado a analisar o modelo preditivo com foco no objetivo, na base de dados utilizada e nas variáveis utilizadas.

O módulo Processo tratará de direcionar o avaliador para blocos específicos do módulo Análise, levando em consideração as decisões tomadas no macroprocesso.

De acordo com a Revisão Sistemática realizada, o objetivo dos modelos preditivos em saúde, em regra, é a busca da solução de um eventual problema. Por isso, a importância de analisar, de forma crítica, se o objetivo está bem definido e se adapta ao problema existente naquele momento. A falta de clareza ou não compatibilidade do objetivo com o problema existente revelam impedimentos para utilização daquele modelo preditivo ao caso concreto. Essa fase garantirá que o avaliador verifique, ao final do macroprocesso, se os resultados encontrados estão em conformidade com as definições do objetivo, além, claro, de identificar os possíveis riscos existentes no objetivo do modelo.

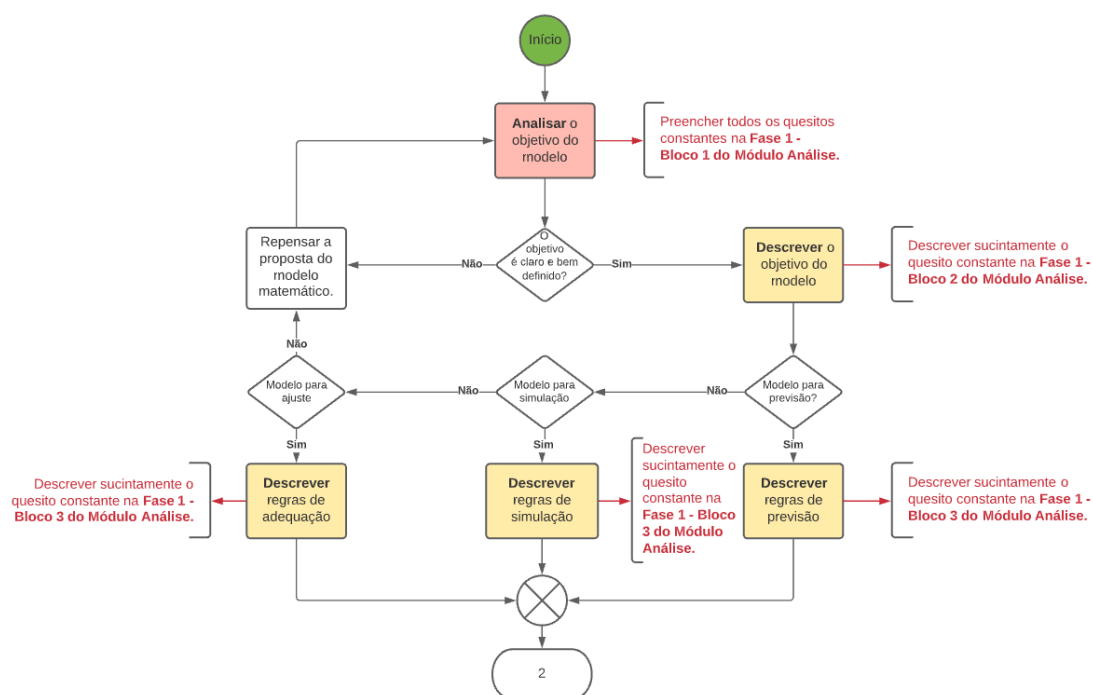


Figura 10 - Fase 1 do módulo Processo.

No início da análise, o avaliador imediatamente será conduzido até o módulo Análise (Anexo 2 do Guia do Framework). Na **Etapa 1** do módulo, o avaliador responderá os quesitos do **Bloco 1**. Nesse bloco, o avaliador analisará se aquele tipo de modelo já existe, se há inovação ou se há releitura de outro modelo já existente.

Sempre que o avaliador passar por uma ação de analisar, os blocos do módulo Análise trarão opções de múltipla escolha, normalmente acompanhada de uma escala de valores, conforme Figura 11.

FASE 1 - OBJETIVOS	
FASE 1 - Bloco 1	
ANALISAR O OBJETIVO DO MODELO	OBSERVAÇÕES
1- Com relação ao problema a ser resolvido pelo modelo preditivo:	
1	não existem modelos preditivos que solucionem o problema. <i>Ao selecionar essa opção, o avaliador está querendo dizer que buscou nas publicações existentes modelos preditivos que pudessem resolver aquele problema específico, mas, não os encontrou.</i>
2	não existem modelos que solucionem o problema com as peculiaridades propostas. <i>Ao selecionar essa opção, o avaliador está querendo dizer que buscou nas publicações existentes modelos matemáticos que pudessem resolver um problema específico. Encontrou modelos mas não com a peculiaridade específica avaliada pela solução proposta.</i>
3	o modelo utiliza outro modelo existente, apenas são feitos ajustes para resolver o problema proposto. <i>Ao selecionar esta opção, o avaliador está querendo dizer que a solução proposta utiliza outro modelo matemático como base mas faz ajustes para melhorá-lo.</i>
4	existem inúmeros modelos que solucionam o problema. <i>Ao selecionar essa opção, o avaliador está querendo dizer que buscou nas publicações existentes modelos matemáticos que pudessem resolver aquele problema específico e encontrou inúmeras soluções para resolver o problema. Esse modelo avaliado será mais uma solução proposta.</i>
2- A resolução do problema:	
<i>Aqui não se deve levar em consideração o impacto da pesquisa na sociedade, mas na aplicabilidade do modelo matemático.</i>	
1	resolverá um problema coletivo, impactará toda a sociedade. <i>Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta impactará toda a sociedade brasileira.</i>
2	resolverá um problema coletivo que não impacta toda a sociedade. <i>Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta resolverá um problema social, que tem interesse coletivo, mas, não impacta necessariamente em toda a sociedade.</i>
3	resolverá o problema de um pequeno grupo de pessoas. <i>Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta impactará em um grupo pequeno de pessoas isoladas. Nesse caso não há interesse coletivo direto.</i>
4	resolverá um problema específico, para uso particular. <i>Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta nesse momento é de interesse do próprio pesquisador ou de uma pessoa específica.</i>

Figura 11 - Bloco 1 da Fase 1.

O avaliador deverá ainda analisar se o modelo preditivo submetido à análise do framework resolverá um problema coletivo ou um problema individual. A partir daqui, faremos, sempre que necessária, uma costura técnica legal, justificando a com os aspectos legais, quanto mais coletivo for o problema resolvido, menor o risco legal, levando em consideração que o interesse coletivo prevalece ao interesse individual.

É necessário lembrar que os direitos coletivos devem prevalecer sobre os direitos individuais, conforme J. L. G. dos Santos et al. (2021) .

Independentemente do caminho que seguir, o avaliador será submetido à ação descrever. O texto incluído aqui tem campo aberto, portanto deve ser utilizado com a maior precisão possível sob pena de prejuízos no resultado obtido. A impossibilidade de identificar, de forma clara e sintética, o objetivo do modelo analisado sugere que aquele modelo deva ser imediatamente revisto.

A Fase 2 do módulo de Análise insurge provocando o avaliador à análise do modelo preditivo em relação ao *dataset* (base de dados) utilizado. A população utilizada no modelo parte, normalmente, das bases de dados já existentes sobre aquela temática. Criticar a origem dos dados, a contextualização e a confiança dessas das informações dispostas é fundamental para a confiança das respostas trazidas pelo modelo.

A falta de histórico, informações coletadas de forma não planejada ou intervenções nas bases as tornam menos confiáveis e muito limitadas. Por vezes, as bases de dados precisam ser relacionadas, e esse próprio processamento gera riscos aos dados lá contidos.

O objetivo dessa fase é provocar o avaliador a criticar o grau de confiança da base de dados, qual o seu propósito, regras de estrutura, estratégias de gerenciamento dos dados, o grau de publicidade/disponibilidade das regras envolvidas e, por fim, através de análise crítica, analisar se os dados contidos nessa base têm qualquer viés político ou comercial. Não há nenhuma pretensão de classificar esse grupo de dados como aplicável ou não, mas orientar o pesquisador com relação aos riscos vinculados às bases utilizadas no estudo.

A Fase 2 está dividida em quatro blocos. O bloco 1 propõe ao avaliador uma análise minuciosa sobre a origem dos dados contidos naquela base de dados, quais as condições existentes para os dados estarem disponíveis e a possibilidade de os dados serem manipulados de acordo com interesses políticos ou comerciais.

Através de uma lente legal, essas condições precisam ser mapeadas para que o avaliador entenda qual é o risco jurídico assumido ao se utilizar uma base de dados que pode não refletir a realidade.

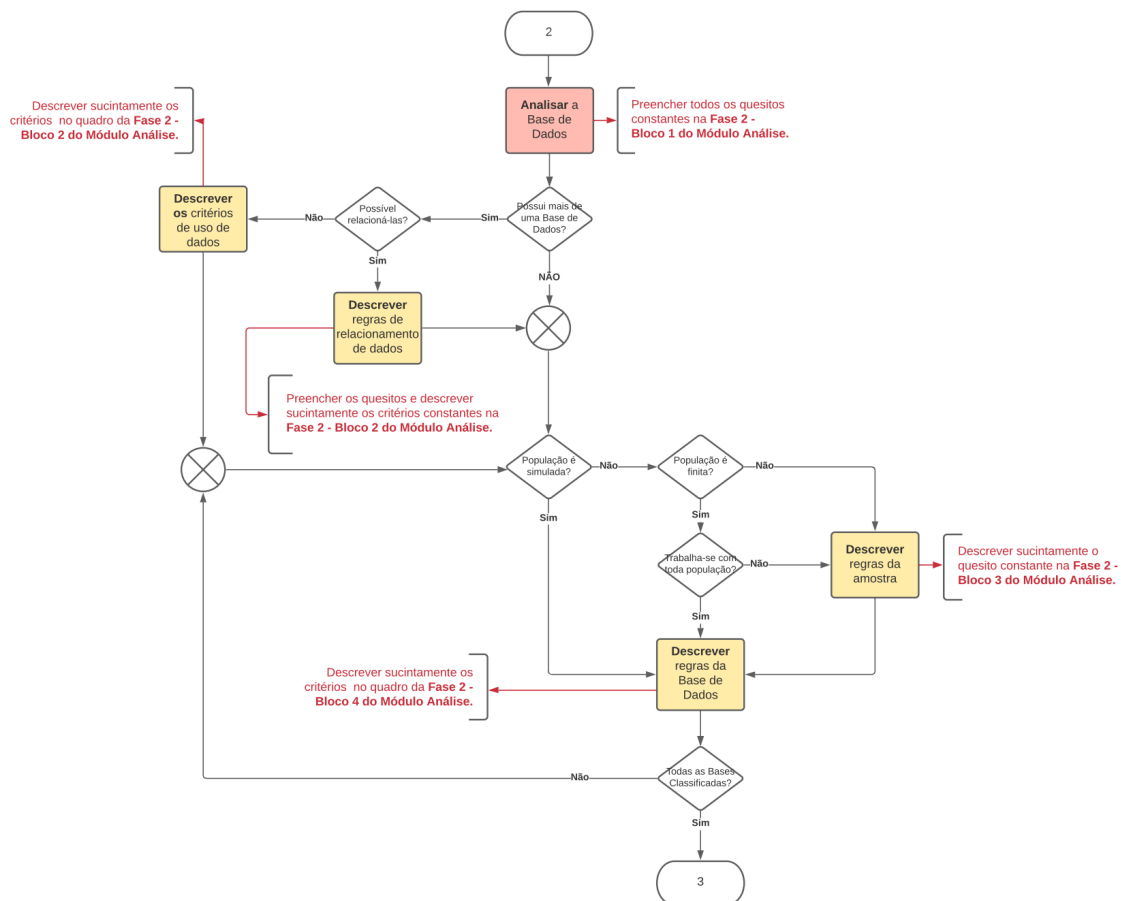


Figura 12 - Fase 2 do módulo Análise.

Um dos grandes desafios dos pesquisadores, com relação às bases de dados é, após o acesso à base, ter condições de relacionar os dados ali contidos. Nem sempre é possível que os dados estejam em um único banco de dados. Quando isso ocorre, é necessário relacionar os dados entre bancos de dados, e esse pode ser um grande desafio.

Em primeiro lugar, é importante que os dados armazenados contenham os mesmos critérios de armazenagem. Por exemplo: uma das bases pode ter a relação dos pacientes atendidos em um determinado hospital que, em um período específico, utilizaram leitos de UTI – Unidade de Terapia Intensiva. Noutra base de dados, há a relação de todos os pacientes que foram diagnosticados com Covid-19. Nesse exemplo simples, fica evidente que não se pode simplesmente relacionar as duas bases de dados para se ter o número de pacientes diagnosticados com Covid-19 que necessitaram internação na UTI.

Uma das formas de trabalhar com os riscos é não os ignorar. Trazer essa informação para o leitor o faz estar atendo às possibilidades de viés. Arriscaria dizer que é um processo de compartilhamento de riscos.

É preciso lembrar que nem sempre os modelos preditivos desenvolvidos serão utilizados para o mesmo contexto que foram desenvolvidos, portanto, é possível que algumas características da base de dados sejam irrelevantes para aquela aplicação específica. Somente quem poderá determinar isso será o usuário do modelo preditivo.

O bloco 2 da fase 2 do módulo de Análise tem a função de descrever as regras de relacionamento entre base de dados, caso exista essa necessidade.

FASE 2 - Bloco 2											
DESCREVER REGRAS - Esse bloco somente deverá ser preenchido se o modelo utilizar mais de uma base de dados!											
É possível relacionar diretamente (entre identificadores) os registros das duas bases de dados?											
SIM	<input type="checkbox"/>										
NÃO	<input type="checkbox"/>										
Descreva de forma clara quais são as regras de relacionamento entre os modelos.											
<table border="1"><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr></table>											

Figura 13 - Bloco 2 da Fase 2 do módulo de Análise.

Outro aspecto relevante diz respeito ao tamanho das bases de dados. É possível que as bases de dados sejam muito grandes, inviabilizando a análise pelo modelo preditivo. Nesse caso, define-se uma amostra dos dados contidos naquela base, o que chamamos de inferência amostral, e, a partir daí, um grupo de dados (amostra) é utilizado como população a ser submetida a análise do modelo preditivo.

Sempre que os dados precisam ser manipulados, geram-se aí riscos de trabalhar com informações que não refletem a realidade. Portanto, o framework provoca o avaliador, no bloco 3 da fase 2 (Figura 14), a analisar as regras utilizadas nas inferências amostrais utilizadas naquele banco de dados estudado.

1- A amostra casual é do tipo:			
1	Equiprobabilística.	Nessa amostra todos os elementos da população original possuem as mesmas chances de estarem na amostra.	<input type="checkbox"/>
2	Randomica.	No critério randomizado não há nenhum tipo de regra. Os elementos são escolhidos aleatoriamente.	<input type="checkbox"/>
3	Sistemática.	A amostra definida por critério sistemático requer regra pré definida que que não esteja diretamente relacionada ao valor do elemento da população.	<input type="checkbox"/>
4	Extratificada.	Nessa opção de amostra extratificada a inferência é feita de forma segmentada utilizando valores definido dos elementos da base.	<input type="checkbox"/>

Descreva de forma clara e objetiva as regras utilizadas para inferência amostral.

Figura 14 - Bloco 3 da Fase 2 do módulo de Análise.

O bloco 4 da Fase 2 instigará o avaliador a descrever as características da base de dados utilizada. A proposta é que essa análise esteja evidenciada para o leitor que fizer uso dos resultados obtidos a partir daquele modelo preditivo. Exemplos de informações que devem ser descritas: data de captura, tamanho da base, unidades de medida utilizadas, critérios de atualização dos dados etc.

FASE 2 - Bloco 4	
DESCREVER REGRAS DA BASE DE DADOS - não descrever aqui as regras de inferência amostral, se existir.	
Descreva de forma clara quais são as regras da base de dados submetida ao modelo. Observação: Não descrever as regras de amostragem.	

Figura 15 - Bloco 4 da Fase 2 do módulo de Análise.

A última fase do módulo Análise requer uma avaliação pormenorizada das variáveis submetidas ao modelo. As características de cada variável assumem um papel importante para que o avaliador entenda a relação existente entre elas e possa estabelecer eventuais conflitos ou riscos por conta das suas características.

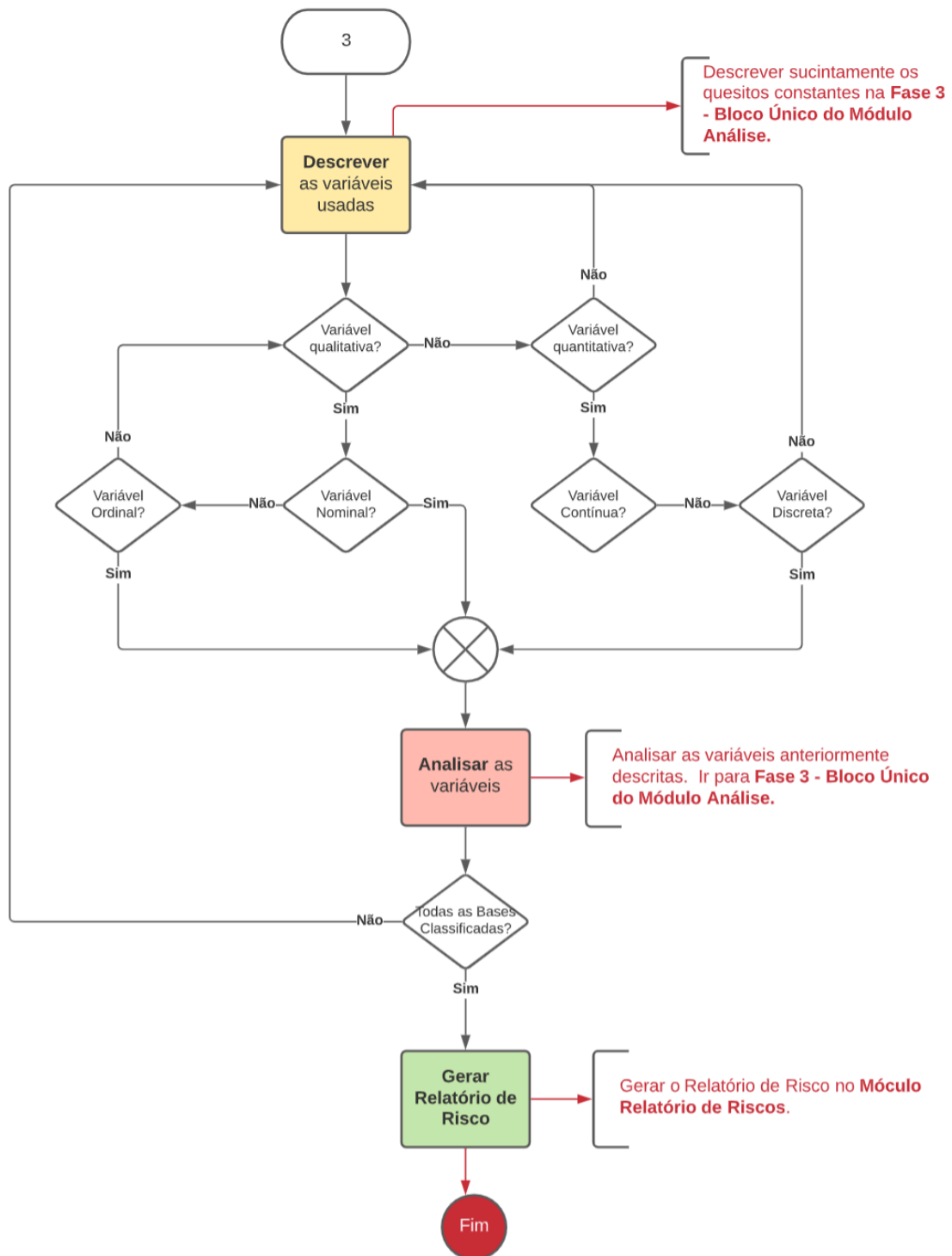


Figura 16 - Fase 3 do módulo Análise.

A proposta é que o Bloco Único seja preenchido seguindo rigorosamente essas etapas:

- 1- Listar, no campo específico todas as variáveis utilizadas no modelo e sua respectiva notação.

2- Classificar cada uma das variáveis listadas como qualitativa ou quantitativa.

As variáveis podem assumir diferentes valores, que podem ser classificados em **quantitativas/numéricas** ou **qualitativos/categóricos**.

3- Classificar cada uma das variáveis listadas na sua respectiva subclassificação.

Além da classificação, aqueles grupos ainda podem ser subdivididos em outras duas categorias:

- **quantitativas/numéricas – discretas:** quando assumem apenas valores inteiros. Ex.: número de irmãos, número de passageiros.
- **quantitativas/numéricas – contínuas:** assume qualquer valor no intervalo dos números reais. Ex.: peso, altura.
- **qualitativos/categóricos – nominais:** quando as categorias não têm uma ordem natural. Ex.: nomes, cores, sexo.
- **qualitativos/categóricos – ordinais:** quando as categorias podem ser ordenadas. Ex.: tamanho (pequeno, médio, grande), classe social (baixa, média, alta), grau de instrução (básico, médio, graduação, pós-graduação).

O bloco está organizado de forma que o avaliador consiga identificar todas as variáveis no mesmo espaço.

FASE 1 - OBJETIVOS	
FASE 1 - Bloco 1	
ANALISAR O OBJETIVO DO MODELO	OBSERVAÇÕES
1- Com relação ao problema a ser resolvido pelo modelo preditivo:	
1	<p>RISCOS: ALTO RISCO - risco da não existência de dados seguros para subsidiar o modelo. A novidade de uma doença normalmente vem acompanhada da falta de informações históricas.</p> <p>PARECER: Incluir, de forma clara, de onde a base de dados foi extraída para o desenvolvimento do trabalho e acrescentar o alto risco de discrepância que pode existir entre os resultados obtidos pelo modelo e os dados apresentados no cenário estudado. Justificar o risco assumido utilizado a novidade do problema e a predominância do interesse coletivo em detrimento do interesse individual.</p>
2	<p>RISCOS: ALTO RISCO - risco da não existência de dados seguros para subsidiar o modelo e risco de não fazer referências aos trabalhos desenvolvidos previamente. O aspecto específico adotado no modelo não exige o pesquisador de fazer referências aos trabalhos realizados anteriormente que utilizam um modelo base já proposto.</p> <p>PARECER: Incluir, de forma clara, de onde a base de dados foi extraída para o desenvolvimento do trabalho e acrescentar o alto risco de discrepância que pode existir entre os resultados obtidos pelo modelo e os dados apresentados no cenário estudado. Justificar o risco assumido utilizado a novidade do problema e a predominância do interesse coletivo em detrimento do interesse individual. Incluir uma sessão que descreva os modelos bases identificados previamente.</p>
3	<p>RISCOS: BAIXO RISCO - como o trabalho utiliza como referência outro modelo já desenvolvido, a responsabilidade é compartilhada.</p> <p>PARECER: Incluir, de forma clara, o modelo utilizado como base para resolução do problema específico. Para não haver infração de propriedade, é importante incluir, precisamente a referência do modelo utilizado e nome dos autores. Caso o modelo não tenha sido publicado, um termo de aceite deve ser devidamente assinado pelos autores.</p>
4	<p>RISCOS: BAIXO RISCO - não é esperado que o modelo apresente resultados discrepantes à realidade. Há risco de apresentar resultados muito próximos ou idênticos a resultados apresentados em trabalhos anteriores. No entanto, o interesse individual é mais evidente que o interesse coletivo.</p> <p>PARECER: Se for utilizado um modelo já conhecido como base de desenvolvimento, deve se indicar objetivamente qual é esse modelo, por quem foi desenvolvido e quais são as principais diferenças entre aquele parâmetro e o desenvolvido.</p>

Figura 18 - Bloco 1 da Fase 1 do módulo Relatório de Riscos.

Os blocos dispostos no módulo Relatório de Riscos são exatamente iguais aos blocos do módulo de Análise. Dessa forma o avaliador fará a relação entre os dados respondidos no módulo Análise e os dados pré-mapeados no módulo Relatório de Riscos.

7.1.2.4 Guia do Framework de Análise de Riscos de Modelos Preditivos

O guia do Framework de Análise de Riscos de Modelos Preditivos é parte integrante do produto desenvolvido e tem o objetivo de nivelar informações para leitores de diferentes áreas do conhecimento.

As informações descritas no guia devem ser analisadas em conjunto com o Framework de Análise de Riscos de Modelos Preditivos, ferramenta de acesso aberto ao público. O documento foi desenvolvido com base na primeira versão publicada do framework.

O guia não substitui a leitura dos textos originais das normas citadas. Além disso, é importante lembrar que a legislação brasileira é flexível e se adapta rapidamente às necessidades, portanto é possível que o documento não acompanhe, em tempo, a evolução das normas.

- 1) Em princípio o framework foi considerado apenas para modelos matemáticos para predição da propagação da Covid-19. Como esta não era uma condição estabelecida para o desenvolvimento, a indicação de uso foi ajustada para modelos preditivos.
- 2) Clareza nos nomes dos processos. Os participantes apresentaram dificuldades em transitar entre os módulos do framework e, após avaliado pelo grupo, se identificou que o nome dos processos gerava dúvidas durante o processo. A nomenclatura dos artefatos do framework foi alterada e devidamente explicada no guia do framework.

8 CONSIDERAÇÕES FINAIS

Um dos desafios enfrentados pelos pesquisadores no desenvolvimento de soluções em saúde durante a grave crise de saúde da pandemia da Covid-19 foi a falta de informações disponíveis sobre o assunto e a alta transmissibilidade da doença em concorrência com a instabilidade jurídica brasileira.

O grande desafio enfrentado nessa pesquisa foi justamente buscar soluções que auxiliem os pesquisadores durante um estado de emergência. A conclusão a que se chega é que não há o devido planejamento para a produção de soluções para problemas em saúde e, quando surgem as necessidades, as dificuldades e desafios são potencializados.

O framework proposto auxiliará os pesquisadores que utilizam modelos preditivos na análise de riscos dos dados utilizados nos modelos para predição da propagação de doenças transmissíveis e funcionará como importante instrumento de análise de modelos preditivos já existentes atualmente.

Para o bojo de conhecimentos necessários para o desenvolvimento da pesquisa foram selecionados modelos preditivos para a predição da propagação da Covid-19 mundialmente e no Brasil através de uma Revisão Sistemática de Literatura, que gerou um artigo científico que se encontra no Apêndice A.

Após selecionados os modelos preditivos, a análise desses modelos possibilitou a definição de um conjunto de fatores que merecem atenção na análise de desenvolvimento de modelos preditivos utilizados para a predição da propagação de doenças transmissíveis.

Um framework de análise de riscos foi desenvolvido com base nesse conjunto de fatores definidos. A ferramenta foi estruturada de forma análoga a uma árvore de decisão com objetivo de provocar a crítica do avaliador para aspectos relacionados ao objetivo, base de dados e variáveis de modelos preditivos.

Embora o estudo tenha sido elaborado no contexto da Covid-19 no Brasil, o framework pode ser adaptado e utilizado para qualquer tipo de doença, com maior ou menor nível de especificidade técnica da doença.

Essa avaliação e outras considerações para aperfeiçoamento da ferramenta foram realizadas após discussão multidisciplinar, por Grupo Focal conforme narrado no capítulo específico.

O framework não tem a pretensão de exaurir as possibilidades de análises tampouco de ser fechado à pesquisa e contribuições apresentadas até aqui. A proposta que se faz é que novas pesquisas e análises direcionadas para cada área do conhecimento sejam propostas por pesquisadores de forma que o framework seja, além de ferramenta para o desenvolvimento de novas pesquisas, aperfeiçoado e especializado em diferentes linhas de pesquisa. Nesse contexto, é precioso lembrar que esta pesquisa nunca teve o propósito de ser uma pesquisa jurídica, mas nada impede que novos estudos sejam desenvolvidos com esse propósito a partir daqui.

A proposta de descortinar, de forma objetiva e estruturada, os desafios enfrentados pelos pesquisadores brasileiros na utilização de dados em saúde e os riscos existentes a respeito dos aspectos jurídicos foi satisfeita.

9 REFERÊNCIAS

- [1] B. Ivorra, M. R. Ferrández, M. Vela-pérez, e A. M. Ramos, “Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China”, n^o January, 2020.
- [2] S. Khajanchi e K. Sarkar, “Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India”, *arXiv*, p. 1–18, 2020.
- [3] K. Abdulmajeed, M. Adeleke, e L. Popoola, “Online Forecasting of Covid-19 Cases in Nigeria Using Limited Data”, *Data Br.*, vol. 30, 2020, doi: 10.1016/j.dib.2020.105683.
- [4] N. Linton *et al.*, “Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data”, *J. Clin. Med.*, vol. 9, n^o 2, p. 538, 2020, doi: 10.3390/jcm9020538.
- [5] S. Banyal, R. Dwivedi, K. D. Gupta, D. K. Sharma, F. Al-Turjman, e L. Mostarda, “Technology landscape for epidemiological prediction and diagnosis of covid-19”, *Comput. Mater. Contin.*, vol. 67, n^o 2, p. 1679–1696, 2021, doi: 10.32604/cmc.2021.014387.
- [6] Ministério da Saúde, Pan American Health Organization, e Fundação Oswaldo Cruz, *A experiência brasileira em sistemas de informação em saúde.*, vol. 1. 2009.
- [7] T. C. de C. L. Neves, L. A. A. Montenegro, e S. D. de A. Bittencourt, “Produção e registro de informações em saúde no Brasil: panorama descritivo através do PMAQ-AB”, *Saúde em Debate*, vol. 38, n^o 103, p. 756–770, 2014, doi: 10.5935/0103-1104.20140069.
- [8] D. C. M. Barbosa e A. C. Forster, “Sistemas de Informação em Saúde: a perspectiva e a avaliação dos profissionais envolvidos na Atenção

- Primária à Saúde de Ribeirão Preto, São Paulo.”, *Cad Saúde Coletiva*, vol. 18, nº 3, p. 424–33, 2010.
- [9] I. H. S. de Moraes e S. R. F. R. dos Santos, “Informação em Saúde: Os Desafios Continuum”, *Cien. Saude Colet.*, vol. 3, nº 1, p. 37–51, 1998, doi: 10.1590/1413-812319983102772014.
- [10] L. V. R. Rezende e M. R. De Lima, “Governance of the internet: A study of the Brazilian Civil Framework”, *Palavra Chave*, vol. 19, nº 1, p. 133–155, 2016, doi: 10.5294/pacla.2016.19.1.6.
- [11] A. D. P. Chiavegatto Filho, “Uso de big data em saúde no Brasil: perspectivas para um futuro próximo”, *Epidemiol. e Serviços Saúde*, vol. 24, nº 2, p. 325–332, 2015, doi: 10.5123/s1679-49742015000200015.
- [12] Q. Jia, Y. Guo, G. Wang, e S. J. Barnes, “Big data analytics in the fight against major public health incidents (Including COVID-19): A conceptual framework”, *Int. J. Environ. Res. Public Health*, vol. 17, nº 17, p. 1–21, 2020, doi: 10.3390/ijerph17176161.
- [13] R. Massato Harayama, “Reflexões sobre o uso do big data em modelos preditivos de vigilância epidemiológica no Brasil”, *Cad. Ibero-Americanos Direito Sanitário*, vol. 9, nº 3, p. 153–165, 2020, doi: 10.17566/ciads.v9i3.702.
- [14] World Health Organization, “Ethical considerations to guide the use of digital proximity tracking technologies for COVID-19 contact tracing”, nº May, p. 6, 2020.
- [15] M. Langarizadeh, F. Moghbeli, e A. Aliabadi, “Application of Ethics for Providing Telemedicine Services and Information Technology”, *Med. Arch. (Sarajevo, Bosnia Herzegovina)*, vol. 71, nº 5, p. 351–355, 2017, doi: 10.5455/medarh.2017.71.351-355.
- [16] A. Wilder-Smith, C. J. Chiew, e V. J. Lee, “Can we contain the COVID-19

- outbreak with the same measures as for SARS?”, *Lancet Infect. Dis.*, vol. 20, nº 5, p. e102–e107, 2020, doi: 10.1016/S1473-3099(20)30129-8.
- [17] Brasil. Ministério da Saúde, “Guia de Vigilância Epidemiológica - emergência de saúde pública de importância nacional pela doença COVID-19”, *Gov. Fed.*, 2020.
- [18] E. Dong, H. Du, e L. Gardner, “An interactive web-based dashboard to track COVID-19 in real time”, *Lancet Infect. Dis.*, vol. 20, nº 5, p. 533–534, 2020, doi: 10.1016/S1473-3099(20)30120-1.
- [19] OPAS, “Perguntas frequentes sobre as vacinas candidatas contra a COVID-19 e os mecanismos de acesso. Versão 2, 27 de agosto de 2020”, *OPAS*, 2021.
- [20] D. de V. E. Ministério da Saúde, Secretaria de Vigilância em Saúde, *Guia de Vigilância Epidemiológica*, vol. 4, nº 3. 1392.
- [21] I. S. Santos, M. A. B. dos Santos, e D. da C. L. Borges, “Mix Público-Privado no Sistema de Saúde Brasileiro : realidade e futuro do SUS”, *Fundação Oswaldo Cruz*, p. 73–131, 2013.
- [22] Brasil, “Lei nº 8.080, de 19 de setembro de 1990.”, *Regulam. do SUS*, 1990.
- [23] N. De Métodos e S. Pública, “Estimativa de risco de espalhamento da COVID-19 no Brasil e avaliação da vulnerabilidade socioeconômica nas microrregiões brasileiras Resumo Os principais achados são : Introdução”, vol. 2020, 2020.
- [24] G. H. Klein, P. G. Neto, e R. Tezza, “Big data e mídias sociais: Monitoramento das redes como ferramenta de gestão”, *Saude e Soc.*, vol. 26, nº 1, p. 208–217, 2017, doi: 10.1590/S0104-12902017164943.
- [25] T. Erl, W. Khattak, e P. Buhler, *Big Data Fundamentals: Concepts*,

Drivers & Techniques. Crawfordsville, 2016.

- [26] TechAmerica Foundation's Federal Big Data Commission, "Demystifying Big Data: A Practical Guide To Transforming The Business of Government Listing of Leadership and Commissioners Global Executive Vice President and General Manager", *UNICOM Gov.*, p. 1–40, 2012.
- [27] W. H. Hamer, *The Milroy Lectures on epidemic disease in England—the evidence of variability and of persistency of type*. London, 2016.
- [28] J. Araújo, "Cálculo, Tecnologias e Modelagem Matemática: ass discussões dos alunos", *Univ. Estadual Paul.*, p. 180, 2002.
- [29] M. C. Goldberg e H. P. L. Luna, *Otimização combinatória e programação linear: modelos e algoritmos*. 2005.
- [30] E. F. Costalonga, A. A. L. Jorge, B. B. Mendonça, e I. J. P. Arnhold, "Modelos matemáticos para previsão de resposta ao tratamento com hormônio de crescimento", *Arq. Bras. Endocrinol. Metabol.*, vol. 52, nº 5, p. 839–849, 2008, doi: 10.1590/s0004-27302008000500016.
- [31] J. C. Júnior, *Comentários à Constituição de 1988*. 2009.
- [32] S. Coletiva e C. Nacional, "Manifesto da Associação Brasileira de Pós-graduação em Saúde Coletiva (ABRASCO) enviado ao Plenário da Conferência Nacional da Ciência, Tecnologia e Inovação", *Cien. Saude Colet.*, vol. 6, nº 2, p. 501–502, 2001, doi: 10.1590/s1413-81232001000200019.
- [33] M. Vieira de Lorenzi Cancelier, "O direito à privacidade hoje: perspectiva histórica e o cenário brasileiro", *Seqüência Estud. Jurídicos e Políticos*, vol. 38, nº 76, p. 213, 2017, doi: 10.5007/2177-7055.2017v38n76p213.
- [34] Brasil, "Constituição (1988). Constituição da República Federativa do Brasil.", *Constituição Federal*, 1988.

- [35] Brasil, “Lei nº 10.406, de 10 de janeiro de 2002.”, *Código Civ.*, 2002.
- [36] Brasil, “Lei nº 13.709, de 14 de agosto de 2018.”, *Lei Geral Proteção Dados*, 2018.
- [37] H. Kelsen, “Teoria Pura do Direito”, *Teor. Pura Do Direito*, p. 271, 1998.
- [38] DANTAS, “Programa de Direito Civil”, vol. 148, p. 148–162.
- [39] P. S. G. ALONSO, *Pressupostos da Responsabilidade Civil*. São Paulo: Saraiva, 2000.
- [40] S. CAVALIERI FILHO, *Programa de Responsabilidade Civil*. São Paulo: Malheiros Editores, 2002.
- [41] R. G. Alves, “Princípio da Supremacia do Interesse Público versus direito à liberdade de consciência e de crença: reflexões à luz das atuais decisões da Suprema Corte sobre vacinação compulsória”, p. 174–203, doi: <http://dx.doi.org/10.11606/issn.2319-0558.v8n2p174-203>.
- [42] Marcelo Schenk Duque, *Direito privado e Constituição: drittwirkung dos direitos fundamentais. Construção de um modelo de convergência à luz dos contratos de consumo*. São Paulo: Revista dos Tribunais, 2013.
- [43] A. de O. Santos, F. P. C. de Barros, e M. C. Delduque, “A pesquisa em saúde no Brasil: desafios a enfrentar”, *Saúde em Debate*, vol. 43, nº spe5, p. 126–136, 2019, doi: 10.1590/0103-11042019s511.
- [44] Semper C 1883, “This is a reproduction of a library book that was digitized by Google as part of an ongoing effort to preserve the information in books and make it universally accessible. <https://books.google.com>”, *Oxford Univ.*, vol. XXX, p. 60, 1994.
- [45] F. J. G. de Lima e A. Bavaresco, “A RESPONSABILIDADE ÉTICO-SOCIAL DO PESQUISADOR NO BRASIL: IMPACTOS DOS DESVIOS

ÉTICOS NA CONDUÇÃO DE PESQUISAS FINANCIADAS COM RECURSOS PÚBLICOS”, *Griot Rev. Filos.*, vol. 14, nº 2, p. 410–426, 2016, doi: 10.31977/grirfi.v14i2.705.

- [46] Barrios, “Resolucao 466”, *Экономика Региона*, vol. 10, nº 9, p. 32, 2012.
- [47] M. K. Souza, C. E. Jacob, J. Gama-Rodrigues, B. Zilberstein, I. Cecconello, e A. Habr-Gama, “Termo de consentimento livre e esclarecido (TCLE): fatores que interferem na adesão”, *ABCD. Arq. Bras. Cir. Dig. (São Paulo)*, vol. 26, nº 3, p. 200–205, 2013, doi: 10.1590/s0102-67202013000300009.
- [48] M. B. Structures, *Métodos de Pesquisa*. Porto Alegre: Editora da UFRGS, 2009.
- [49] D. S. Backes, J. S. Colomé, R. H. Erdmann, e V. L. Lunardi, “The focal group as a technique for data collection and analysis in qualitative research”, *Mundo da Saude*, vol. 35, nº 4, p. 438–442, 2011, doi: 10.15343/0104-7809.2011354438442.
- [50] E. T. Rother, “Revisão sistemática X revisão narrativa”, *Acta Paul. Enferm.*, vol. 20, nº 2, p. v–vi, 2007, doi: 10.1590/s0103-21002007000200001.
- [51] B. Kitchenham, “Procedures for Performing Systematic Reviews”, *Ann. Saudi Med.*, vol. 37, nº 1, p. 79–83, 2017, doi: 10.5144/0256-4947.2017.79.
- [52] D. Moher *et al.*, “Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement”, *PLoS Med.*, vol. 6, nº 7, 2009, doi: 10.1371/journal.pmed.1000097.
- [53] M. B. Structures, “ROBIS: Ferramenta para avaliar risco de viés em revisões sistemáticas”.

- [54] L. F. S. Scabini, L. C. Ribas, M. B. Neiva, A. G. B. Junior, A. J. F. Farfán, e O. M. Bruno, “Social interaction layers in complex networks for the dynamical epidemic modeling of COVID-19 in Brazil”, *Phys. A Stat. Mech. its Appl.*, vol. 564, p. 1–16, 2021, doi: 10.1016/j.physa.2020.125498.
- [55] A. Arenas *et al.*, “A mathematical model for the spatiotemporal epidemic spreading of COVID19”, *Phys. Rev. X*, p. 1–13, 2020, doi: 10.1101/2020.03.21.20040022.
- [56] F. C. Coelho *et al.*, “Assessing the spread of COVID-19 in Brazil: Mobility, morbidity and social vulnerability”, *PLoS One*, vol. 15, n° 9 September, p. 1–11, 2020, doi: 10.1371/journal.pone.0238214.
- [57] T. M. Rocha Filho *et al.*, “Expected impact of COVID-19 outbreak in a major metropolitan area in Brazil”, *medRxiv*, p. 1–11, 2020, doi: 10.1101/2020.03.14.20035873.
- [58] A. Canabarro, E. Tenório, R. Martins, L. Martins, S. Brito, e R. Chaves, “Data-driven study of the COVID-19 pandemic via age-structured modelling and prediction of the health system failure in Brazil amid diverse intervention strategies”, *PLoS One*, vol. 15, n° 7 July, p. 1–8, 2020, doi: 10.1371/journal.pone.0236310.
- [59] R. E. Johnson e B. Foote, “Designing reusable classes”, *J. Object-Oriented Program.*, vol. 1, n° 2, p. 22–30, 35, 1988.
- [60] L. A. B. Trad, “Focal groups: Concepts, procedures and reflections based on practical experiences of research works in the health area”, *Physis*, vol. 19, n° 3, p. 777–796, 2009, doi: 10.1590/s0103-73312009000300013.
- [61] M. A. F. Werneck e P. F. Teixeira, *Processo de trabalho em saúde*, 2ª Edição. Belo Horizonte: Coopmed, 2009.
- [62] T. C. Morphology, “BPM CBOK”.

- [63] “Manual dos elementos da BPMN a serem utilizados no mapeamento de processos dos setores da UFCSPA : Divisões Representam as entidades responsáveis pela execução das atividades . Eventos Elementos utilizados para representar a ocorrência de fatos em um proc”.
- [64] J. L. G. dos Santos *et al.*, “Collision of Fundamental Human Rights and the Right to Health Access During the Novel Coronavirus Pandemic”, *Front. Public Heal.*, vol. 8, nº January, 2021, doi: 10.3389/fpubh.2020.570243.
- [65] M. J. Ryan e L. S. Wheatcraft, “On the Use of the Terms Verification and Validation”, *INCOSE Int. Symp.*, vol. 27, nº 1, p. 1277–1290, 2017, doi: 10.1002/j.2334-5837.2017.00427.x.

10 APÊNDICE A – ARTIGO DA REVISÃO SISTEMÁTICA

Mathematical Models for Prediction of COVID-19 Propagation: Systematic Review

Uilian Loose, Aline Griza, Andreas Timóteo Lutz, Graziella Moraes Machado, Juliana Silva Herbert

*Graduate Program in Information Technologies and Health Management
Federal University of Health Sciences of Porto Alegre (UFCSPA)*

uilianl@ufcspa.edu.br, alineg@ufcspa.edu.br, lutz.onco@gmail.com, grazigijesen@gmail.com,
julianash@ufcspa.edu.br

ABSTRACT

The sudden emergence of Covid-19, the uncontrolled spread of the disease, and the chaos in healthcare systems worldwide led to challenges for public administrators and society in general, forcing them to make quick decisions. One of the tools used to assist in decision-making is the mathematical prediction model. In this systematic review, we selected recent articles on mathematical models used to predict the propagation of Covid-19. From 1901 identified articles, 65 publications were classified to screen titles and abstracts. Then, 25 studies were classified according to the mathematical method developed, input data, results obtained, and place of publication. Most of the analyzed mathematical models are compartmental, and they add new compartments or specialized numerical methods to solve specific gaps in simpler models like SIR and SEIR. It was identified that these models should not be used broadly but rather to solve particular problems.

Palavras-Chave: Covid-19, SARS-CoV-2, mathematical model, predictive model, virus transmission.

1. Introduction

Scientists from institutions worldwide are committed to understanding the spread of the new coronavirus (SARS-CoV-2) to fight the Covid-19 disease [1]. The virus was identified in Wuhan, China, where the first reported fatalities occurred in late 2019 [2]. The spread of the virus is a growing public health concern worldwide, which enforces public managers' need for strategic actions [1]. However, it is challenging to predict the future of the virus and its spread, especially the number of people who will manifest Covid-19 in

its most severe forms. Direct difficulties are linked to the virus itself, such as the degree of lethality, the incubation time of the virus in the human body, and its transmissibility. In addition, there are indirect difficulties, such as the average time to obtain test results; the randomness of cases; different vaccine technologies; treatment adherence; disseminated information without scientific basis; distinct local epidemiological realities; public health structures without sufficient structure to provide adequate care, and

other factors that contribute to the difficulty of predictions [3].

Therefore, the scientific community has a considerable effort to search for information to help understand the current scenario and the disease behavior. Mathematical modeling for virus propagation predictions is one of the tools used by public management to fight against the spread of the virus. This situation is explained by the success of the adopted strategies depends on the set of actions taken based on the information provided by the prediction models.

Given this scenario, it is crucial to identify the numerical methods and the peculiarities of the predictive mathematical models already proposed so far. More than just presenting new models, it is essential for Covid-19's new predictions to understand if specific characteristics should be considered to select the most appropriate mathematical model for each particular situation.

It is possible that these specific characteristics previously identified contribute to increase the rate of assertiveness in the results and, consequently, in the actions to fight the spread of the SARS-CoV-2 virus.

Linton et al. (2020) state that the average time of infection by the new coronavirus from onset to hospitalization is ten days. The average period of viral incubation is between 2 to 14 days. Also, according to the Italian Higher Institute of Health (ISS), the average time from the onset of symptoms to death is 12 days, and the time between hospitalization and death is seven days [5]. Therefore, these data should be considered when using mathematical prediction models. However, indirect variables, which depend on human interference, rely on preliminary analysis by the developing

model. This is addressed by Khoshnaw et al. (2020), by whom importance is given to the numbers provided to avoid a collapse of hospital beds.

It needs to be emphasized how challenging it is to add behavioral information about the population in mathematical models, such as avoiding contact between people who are not from the same family, using a mask, cleaning hands, and establishing routines that minimize the flow of people [7]. In addition, it is essential to clarify that, although effective vaccines already exist, strategies to control the flow of people continue to be an effective option for fighting the disease and the spread of the virus [8]. Therefore, public managers must have available tools capable of assisting in complex decisions, which involve many variables and data for processing.

The systematic literature review presented in this article aims to identify, select and analyze: (1) the mathematical models for the analysis and prediction of the propagation of the new coronavirus - Covid-19 and (2) ways of communicating the results of these models with society. Furthermore, through the protocol proposed by Kitchenham et al. (2015), were observed specific characteristics related to (a) the location where the study is carried out to understand location-specific elements; (b) the input data that will support the processing to verify if there are different variables used between the different proposed models and (c) whether or not how the result of the models is presented to the target audience is taken into account, aiming to understand if the models consider specific objectives.

2. Method

A systematic review is a secondary study that aims to select publications in a specific area, according to pre-defined criteria, based on primary studies [10] to answer particular questions. The study was developed based on the protocol proposed by Kitchenham et al. (2015) and revised using the checklist presented by PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses [11]. In addition, using the ROBIS framework - Risk of Bias in Systematic Reviews [12], it was also assessed to identify possible bias problems in applying the protocol.

The planning and review of the protocol used were developed in pairs by a multidisciplinary team from the main areas involved in this research: 2 computer scientists, two physicians, and one social communication professional, considering the following steps: (1) definition of research questions, (2) elaboration of the search strategy, (3) identification of inclusion and exclusion criteria, (4) definition and application of quality criteria, (5) data extraction and (6) data analysis and synthesis.

These steps are presented in the following subsections.

2.1. Research Questions

The objective of the research questions is to characterize the mathematical models found in the literature. Thus, it is possible to verify if there are characteristics that should be considered to select the adequate mathematical model to predict the propagation of Covid-19 in specific situations. To this end, the following research questions were defined:

RQ1- What are the mathematical models used to predict the spread of the new coronavirus?

RQ2- What are the input variables used in the analyzed models?

RQ3- What is the source of the input information used in the analyzed mathematical models (are they actual or simulated datasets)?

RQ4- What are the positive aspects and limitations found in each of the analyzed models?

RQ5- For which contexts were the analyzed models developed?

RQ6- How are the results of each model presented?

RQ7- Is the analyzed model based on an existing model?

Before the research process was started, there was an assumption that perhaps not all articles analyzed could correctly answer all research questions. In fact, what is proposed is to carry out a critical analysis to identify the perspectives used by the authors to answer the presented questions.

2.2. Search Strategy

Since the objective was to focus the study on mathematical models related to the prediction of Covid-19 propagation, the multidisciplinary of studies in mathematical and medical sciences was considered. Therefore, five bases were chosen to apply the search expressions: Pubmed, IEEE, Scopus, Embase, and Web of Science.

Articles related to the Covid-19 theme had their publications encouraged by journals to disseminate the knowledge acquired through research and contribute to the scientific area. This situation naturally led to a more flexible review protocol and acceptance of these articles. Therefore, as a limiting criterion, in the

protocol proposed in this study, only articles that were accepted and already published on the cutoff date of the searches were considered.

The analysis of articles published during the pandemic period is weighed as a challenge due to both quantity and quality. However, there is an opportunity to contribute to the classification of a part of this published material. Therefore, the three most significant areas of knowledge applied in the review were considered to develop the search expression, as presented in Figure 1.

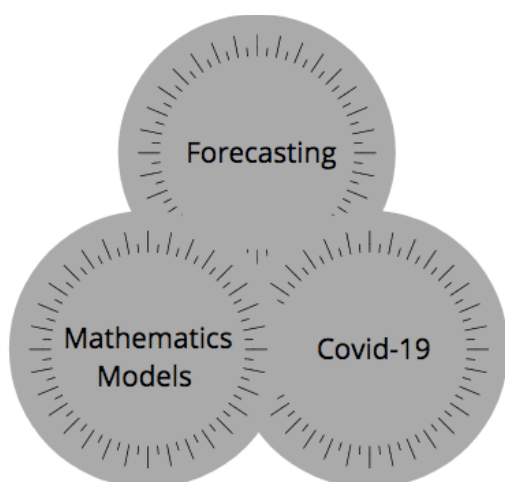


Figure 1 - Area chart.

Based on the chart, it was possible to perform the first tests with the following string:

("COVID-19" OR "Coronavirus" OR "2019-nCoV" OR "SARS-CoV" OR "MERS-CoV" OR "Severe Acute Respiratory Syndrome" OR "Middle East Respiratory Syndrome") AND ("mathematical" OR "statistical" OR "statistic" OR "computational" OR "mathematics" OR "prediction" OR "prognosis") AND ("model" OR "models" OR "modeling")

The results were not satisfactory since the search found 1,901 publications, most of which had only mentions of the expression mathematical or statistical model. Then the keywords were limited, being included in the search string in a combined way, resulting in the following expression:

{(COVID-19 OR "new coronavirus" OR 2019-nCoV OR Sars-CoV2 OR Sars-CoV-2) AND ("mathematical model" OR "mathematical modeling" OR "mathematical models" OR "mathematical modelling") AND (prediction OR prognosis OR spread OR prevention OR forecasting)}.

The expression resulted in 523 publications, divided according to Table 1:

Tabela 1 - Research Base.

Search String application		
Base	Application Date	Output
IEEE	23/08/20	8
EMBASE	23/08/20	84
WEB OF SCIENCE	21/08/20	96
SCOPUS	23/08/20	217
PUBMED	23/08/20	118
		523

2.3. Inclusion and Exclusion Criteria

As the new virus was only identified at the end of 2019, publications from that date were sought. The Inclusion Criteria (IC) defined were: (IC1) articles published from 12/01/2019 to 08/23/2020, (IC2) in English, Spanish, German or Portuguese, and (IC3) article-type publications. The application of the Inclusion Criteria classified 411 articles, as shown in Table 2.

Tabela 2 - Classification of articles.

TITLE	AUTHORS	QC1	QC2	QC3	QC4	QC5	QC6	QC7	QC8	SCORE
<i>COVID-19: Development of a robust mathematical model and simulation package with consideration for ageing population and time delay for control action and resusceptibility</i>	Ng & Gui, 2020	1	1	1	1	1	0	0	1	6
<i>Daily Forecasting of New Cases for Regional Epidemics of Coronavirus Disease 2019 with Bayesian Uncertainty Quantification</i>	Lin et al., 2020	1	1	1	1	0	0	0	1	5
<i>Data-driven modelling and prediction of COVID-19 infection in India and correlation analysis of the virus transmission with socio-economic factors</i>	Kumar et al., 2020	1	1	1	0	0	0	0	1	4
<i>Dynamic model of COVID-19 disease with exploratory data analysis</i>	Adeniyi et al., 2020	1	1	1	1	0	1	0	0	5
<i>Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China Based on Simple Mathematical Model</i>	Zhong et al., 2020	1	1	0	0	1	0	0	1	4
<i>Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020</i>	Ganyani et al., 2020	1	1	1	1	0	0	1	1	6
<i>Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India</i>	Khajanchi & Sarkar, 2020	1	1	1	1	0	0	0	1	5
<i>Forecasting the spread of COVID-19 in Kuwait using compartmental and logistic regression models</i>	Almeshal et al., 2020	1	1	1	1	1	1	1	1	8
<i>Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan</i>	Ndaïrou et al., 2020	1	1	1	1	0	0	0	1	5
<i>Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China</i>	Ivorra et al., 2020	1	1	1	1	1	0	1	0	6
<i>Mathematical Modelling for Coronavirus Disease (COVID-19) in Predicting Future Behaviours and Sensitivity Analysis</i>	Khoshnaw et al., 2020	1	1	1	1	1	1	1	0	7
<i>Mathematical modelling to assess the impact of lockdown on COVID-19 transmission in India: Model development and validation</i>	Ambikapathy & Krishnamurthy, 2020	1	1	1	0	0	0	0	1	4
<i>Modeling and forecasting the COVID-19 pandemic in India</i>	Sarkar et al., 2020	1	1	1	1	1	1	1	0	7
<i>Modeling and forecasting the covid-19 temporal spread in Greece: An exploratory approach based on complex network defined splines</i>	Demertzis et al., 2020	1	1	1	1	1	1	1	1	8
<i>Modeling the impact of mass influenza vaccination and public health interventions on COVID-19 epidemics with limited detection capability</i>	Li et al., 2020	1	1	1	1	1	0	0	1	6
<i>Modeling the role of asymptomatics in infection spread with application to SARS-CoV-2</i>	Dobrovolny, 2020	1	1	1	0	1	1	1	0	6
<i>ONLINE FORECASTING OF COVID-19 CASES IN NIGERIA USING LIMITED DATA</i>	Abdulmajeed et al., 2020	1	1	1	0	0	1	0	0	4
<i>Predicting the epidemiological outbreak of the coronavirus disease 2019 (COVID-19) in Saudi Arabia</i>	Alboaneen et al., 2020	1	1	0	1	1	1	1	0	6
<i>Predicting the Trajectory of Any COVID19 Epidemic From the Best Straight Line</i>	Levitt et al., 2020	0	1	1	0	1	1	1	0	5
<i>Prediction of COVID-19 transmission dynamics using a mathematical model considering behavior changes in Korea</i>	Kim et al., 2020	1	1	1	0	0	0	0	1	4
<i>Rare and extreme events: the case of COVID-19 pandemic</i>	Machado & Lopes, 2020	0	1	1	0	1	1	1	0	5
<i>Resetting the Initial Conditions for Calculating Epidemic Spread: COVID-19 Outbreak in Italy</i>	Babac & Momar, 2020	1	1	0	0	1	1	1	0	5
<i>Short-term forecasting of daily COVID-19 cases in Brazil by using the holt's model</i>	Martinez et al., 2020	1	1	1	0	1	1	1	1	7
<i>Staggered release policies for COVID-19 control: Costs and benefits of relaxing restrictions by age and risk</i>	Zhao & Feng, 2020	1	1	1	1	0	1	1	0	6
<i>SutteARIMA: Short-term forecasting method, a case: Covid-19 and stock market in Spain</i>	Ahmar & del Val, 2020	0	1	1	1	0	0	0	1	4

The Inclusion Criteria, and the use of filters offered in the search bases, established the minimum characteristics for the acceptability of the article as a starting point. To refine the filter and search for articles that fit the desired characteristics, Exclusion Criteria (EC) were defined. Following the PRISMA checklist, duplicated articles were excluded before applying the exclusion criteria [11]. After excluding the duplicated articles, the following exclusion criteria were defined: articles that (EC1) are repeated, (EC2) have not yet been effectively published by

periodical literature, (EC3) have a future publication date, (EC4) are systematic or integrative reviews, (EC5) do not specify the input data, (EC6) do not apply the model to a specific dataset, (EC7) are not made with human beings, (EC8) are models for a particular scenario or not include a mathematical model and (EC9) do not address specific models for Covid-19.

After applying the inclusion and exclusion criteria, 65 articles were classified, as shown in Figure 2.

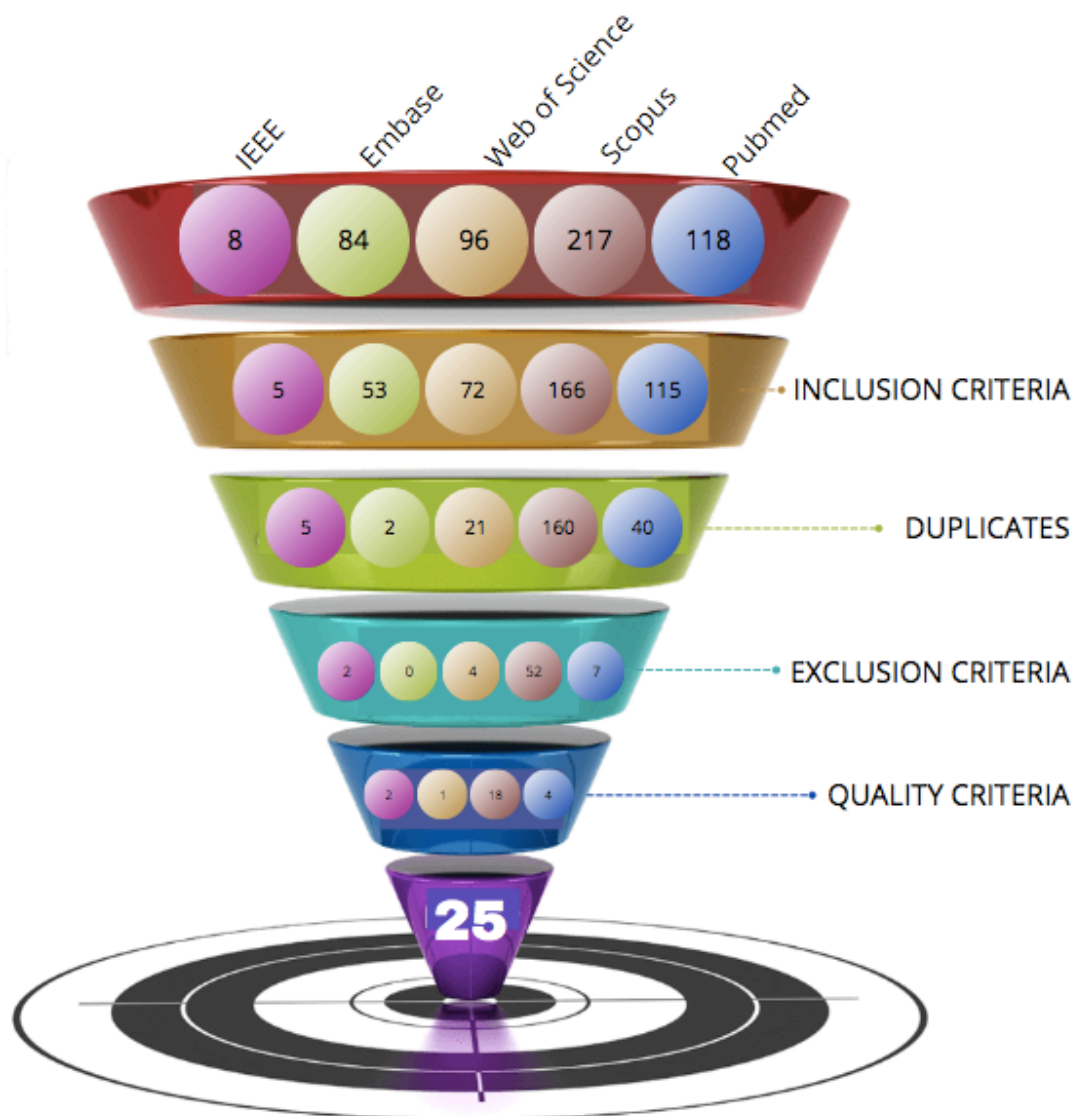


Figure 2 - Selection funnel.

2.4. Quality Criteria

The Quality Criteria (QC) were defined to analyze the articles' internal and external validity and eliminate possible trends, based on the application scale of eight pre-defined criteria. The articles were classified as 1 (meets the criterion) or 0 (does not meet the criterion) for each quality criterion.

QC1 - Is the purpose of the article clear, objective, and well defined? (1 if it fully meets the criterion or 0 if it does not fully meet the criterion).

QC2 - Are the input variables of the mathematical model explained individually? (1 if it fully meets the criterion or 0 if it does not fully meet the criterion). Mathematical models have variables that reflect real measurable data. Therefore, the article that specifies each input variable so that the reader can apply it in an actual situation will meet the criterion.

QC3 - Compartmental models with more than three groups meet the criterion.

QC4 - If the script is explained step by step. The article should explain the construction of the mathematical model so

that the reader understands the logic of the mathematical script proposed there.

QC5 - The mathematical models that detail the information in the database used in the research meet this criterion.

QC6 - If anyone easily accesses the reference dataset for the model, it meets the criterion.

QC7 - The article that determines the start and end date, the specific location, and the particular peculiarities of the data used for application in the proposed model (dataset) meet the criterion.

QC8 - The article that presents results obtained to a specific audience, targeting and not in a generic way, meets the criterion.

After applying the quality criteria, the analyzed articles were classified with scores between 1 and 8, as shown in Table 2.

Studies with scores between 5 and 8 were then selected, resulting in 19 articles. Nineteen articles received a score of 4. Of these articles, those with any score in the QC5, QC6, QC7, and QC8 criteria were classified for the review, resulting in 6 articles and reaching a total of 25 selected articles.

3. Results

In the first stage of the study, the articles were classified by the mathematical methods used. Subsequently, the places of publication of the selected articles were analyzed, and the location of the respective datasets was used. This is called geographic context and will be dealt with at a specific point in this study.

The mathematical methods used by each model were divided into two large groups: individual models and compartmental models. Individual models are characterized by using one or a set of numerical methods to solve the problem. Compartmental models consider that the physical world is divided into related compartments.

Of the 25 models analyzed, 15 use the compartmental method, as shown in Table 3. That is, they are based on existing mathematical models. The difference of each one of them is precisely the number of defined groups.

Table 3 - Compartmental method.

Title	Autors	Method	Numeric Method	Number of groups
Dynamic model of COVID-19 disease with exploratory data analysis	Adeniyi et al., 2020	compartmental	SQUIRES	5 groups
Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China Based on Simple Mathematical Model	Zhong et al., 2020	compartmental	SIR	3 groups
Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India	Dr; Sarkar, 2020	compartmental	SAIUQR	6 groups
Forecasting the spread of COVID-19 in kuwait using compartmental and logistic regression models	Almeshal et al., 2020	compartmental	SIR	3 groups
Mathematical Modelling for Coronavirus Disease (COVID-19) in Predicting Future Behaviours and Sensitivity Analysis	Khoshnaw; Salih; Sulaimany, 2020	compartmental	SEIR	8 groups
Modeling and forecasting the COVID-19 pandemic in India	Science; Phenomena; Sarkar; et al., 2020	compartmental	SEIR	6 groups
Modeling the role of asymptomatics in infection spread with application to SARS-CoV-2	Dobrovolny, 2020	compartmental	Próprio	5 groups
Predicting the epidemiological outbreak of the coronavirus disease 2019 (COVID-19) in Saudi Arabia	Alboaneen et al., 2020	compartmental	SIR	3 groups
Prediction of COVID-19 transmission dynamics using a mathematical model considering behavior changes in Korea	Kim; Seo; Jung, 2020	compartmental	SEIR	6 groups
Resetting the Initial Conditions for Calculating Epidemic Spread: COVID-19 Outbreak in Italy	Babac; Mornar, 2020	compartmental	SIR	3 groups
Staggered release policies for COVID-19 control: Costs and benefits of relaxing restrictions by age and risk	Zhao; Feng, 2020	compartmental	SEIR	3 groups
COVID-19: Development of a robust mathematical model and simulation package with consideration for ageing population and time delay for control action and resusceptibility	Yew; Mei, 2020	compartmental	SEIRS	5 groups
Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan	Science; Phenomena; Ndaïrou; et al., 2020	compartmental	SIR	8 groups
Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China	Ivorra et al., 2020	compartmental	SEIR	9 groups
Modeling the impact of mass influenza vaccination and public health interventions on COVID-19 epidemics with limited detection capability	Li et al., 2020	compartmental	SEIR	10 groups

The SIR is one of the simplest compartmental models. It divides the total population into three groups: (S) Susceptible, (I) Infectious, and (R) Recovered. The greater the number of groups the population is divided into, the greater the complexity of the model. This is because, in addition to determining the characteristics of each group, it is necessary to decide on the transition rates between these groups. Therefore, labels usually define the number of groups in the model. The SEIR model, for example, divides the population into (S) Susceptible, (E) Exposed, (I) Infectious, and (R) Recovered.

The most important characteristic of these compartmental models is not the number of defined groups but the dynamics with which the groups are related and how the transition rates are determined. For this definition, usually ordinary differential equations are used, always considering a limited population. In this type of analysis, the greater the number of compartments used by the author, the greater the complexity of the study, since each compartment, in theory, should determine the rate of information transfer to the other compartments.

The model proposed by Dobrovolny (2020), unlike the others, has its compartments differing from more traditional ones such as SIR and SEIR,

which generally serve as a starting point for the definition and inclusion of new groups in new models.

The studies analyzed use a simpler compartmental model as a starting point

and add new groups and new transition rates between them, as observed in the model proposed by Ndairou F, et al., (2020), in Figure 3.

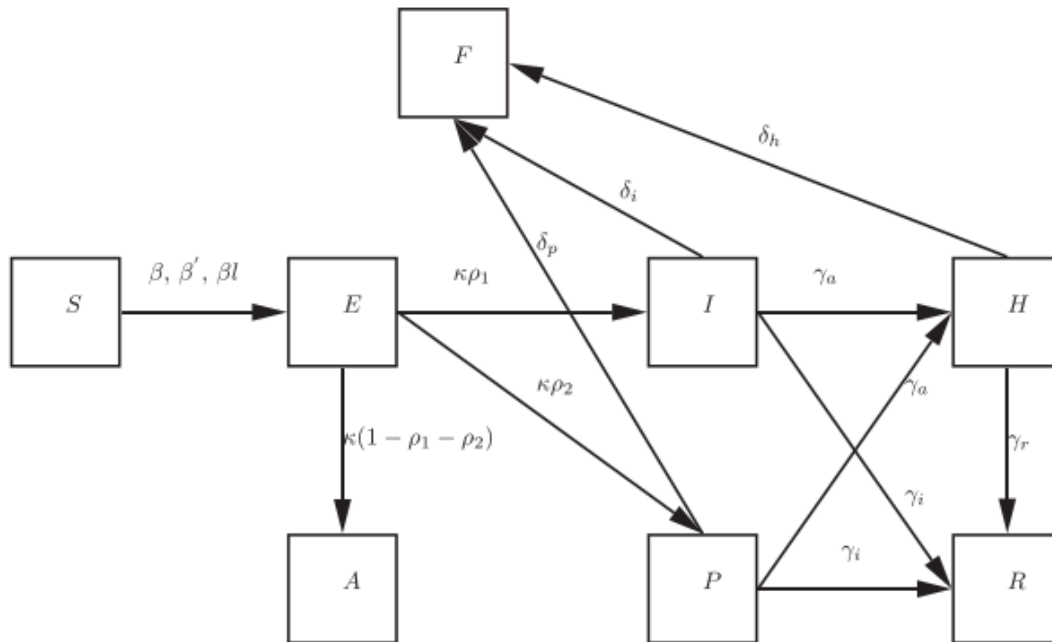


Figure 3 - Example of a compartment model.

Khajanchi S, Sarkar K, (2020) segmented the population assessed into six compartments, namely Susceptible (S), Asymptomatic (A), Reported Symptomatic (I), Unreported Symptomatic (U), Quarantine (Q), and Recovered (R), collectively termed SAIUQR [15].

Because of the new proposals brought to compartmental models, it is crucial to consider the model by Adeniyi et al. (2020), which presumes that recovered people can lose their immunity over time.

Another good proposal considering compartmental models is to define them using original mathematical methods. For example, Zhao & Feng (2020) divide the population analyzed into three major risk

groups and submit each group to a 7-group compartmental model.

Another ten models had their mathematical methodologies. Finally, the authors used numerical methods to reach a natural solution to problems defined in their studies, as shown in table 4.

Lin et al. (2020) developed a compartmental model based on the SEIR model, with 25 subpopulations using Bayesian inference to adjust the variables used in the compartmental model. The Bayesian method considers the researcher's knowledge through random elements, in addition to the information established from the evaluated data.

The linear regression method, which consists of the dependence relationship between the variables defined by each author, was used by Kumar et al. (2020),

Demertzis, Tsiotes, Magafas (2020), and Abdulmajeed; Adeleke; Popoola (2020).

Time series was the system chosen by Levitt, Scaiewicz; Zonta, (2020), Saleh; Boj, (2020), and Tapiwa et al. (2020) to observe disease behavior in a small unit of analysis over time. That is, the authors decided to analyze disease behavior for a longer time in a small population.

Population dynamics can be studied through exponential or logistical growth. The first uses first-order differential equations to calculate the growth rate proportionally to the population verified

at that moment observed, as it is possible to confirm in the study by Martinez, Aragon, Nunes (2020). In addition to the population growth rate being proportional to the population, it must be proportional to the difference between the maximum population and the present population, as is the case described by Machado; Lopes (2020).

Ambikapathy; Krishnamurthy (2020) chose to use 14 first-order differential equations, one for each country studied, relating the rate of variation of standard variables among the populations studied.

Table 4 - Articles with original methods.

TITLE	AUTORS	Method	Numeric Method
Daily Forecasting of New Cases for Regional Epidemics of Coronavirus Disease 2019 with Bayesian Uncertainty Quantification	Lin et al., 2020	Próprio	Modelo Probabilístico + Compartmental
Data-driven modelling and prediction of COVID-19 infection in India and correlation analysis of the virus transmission with socio-economic factors	Kumar et al., 2020	Próprio	Regressão Linear
Modeling and forecasting the covid-19 temporal spread in Greece: An exploratory approach based on complex network defined splines	Demertzis; Tsiotas; Magafas, 2020	Próprio	Regressão linear
ONLINE FORECASTING OF COVID-19 CASES IN NIGERIA USING LIMITED DATA	Abdulmajeed; Adeleke; Popoola, 2020	Próprio	Modelo Auto-regressivo + Modelo de regressão aditiva + Modelo de série temporal
Predicting the Trajectory of Any COVID19 Epidemic From the Best Straight Line	Levitt; Scaiewicz; Zonta, 2020	Próprio	Especializado
Rare and extreme events: the case of COVID-19 pandemic	Machado; Lopes, 2020	Próprio	Molelo logístico e Richard
Short-term forecasting of daily COVID-19 cases in Brazil by using the holt's model	Martinez; Aragon; Nunes, 2020	Próprio	Exponencial
SutteARIMA: Short-term forecasting method, a case: Covid-19 and stock market in Spain	Saleh; Boj, 2020	Próprio	ARIMA
Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020	Tapiwa et al., 2020	Próprio	Probabilístico
Mathematical modelling to assess the impact of lockdown on COVID-19 transmission in India: Model development and validation	Ambikapathy; Krishnamurthy, 2020	Próprio	equações diferenciais

Two articles were rated with the highest score (8 points). In the first article, Almeshal et al. (2020) present information on maximum growth, initially assuming an exponential growth, followed by stable development and ending with a decrease in the contagion rate. This process is done through a simple SIR model and a logistic model to adjust transition rates between compartments. The authors innovate in the proposal when they inform the reader of the data obtained in each phase. In the second article, Demertzis, Tsiotes, Magafas (2020) analyze the population of Greece, building methods for curve fitting within linear, quadratic, cubic, power function, and logarithmic ones. Using least squares, the study shows that the function that best fits the concrete case is a third-degree polynomial. The authors also do spline fittings – a function defined in parts, using second-degree polynomials in each defined interval. The selection of intervals was performed by complex network analysis of time series. This selection had a better fit when compared with a random choice of intervals and cubic curve fitting.

As resulting information, the two studies presented tables with chronological information, which facilitates the use of the respective data for future decision-making actions.

3.1. Input Data

One of the criteria that indicate the accuracy of the results generated by the proposed models is the datasets used since the projections are based on the factual reality of each specific situation. Therefore, most selected articles use public data from the study country or the World Health Organization (WHO) website, if available.

The geographic context of the population analyzed for the development of the mathematical model is a crucial criterion for the model analysis because the characteristics of this population will be stratified. Therefore, the input data processed by the model will be identified from the model analysis there.

3.2. Geographical context of the application

Epidemiological studies are predominantly carried out in regions with the most significant amount of information available, either by temporal history or by public organizations regarding the data structure and availability. Therefore, 25% of the articles analyzed are based on studies from China for the development of their models, mainly because it was where the disease started [26][27][2][21], making it the country with the oldest case history. In addition, China now provides the data in an organized way and in real-time, facilitating the learning process of the models developed and the strategy for measuring these studies.

Of the nine articles that used the China population as a reference for their models, only the studies by Ivorra et al. (2020), Li et al. (2020), and Zhong et al. (2020) were developed in China or with the participation of researchers in Chinese territory. This situation can be evidence of what was verified before: the choice of the statistical dataset depends on the accuracy of the available database.

In contrast, research-based on statistics from the United States was developed by researchers in the country. [17][21][13][18].

4. Discussion

Unlike what was assumed when designing this study, compartmental models are used, in most cases analyzed, as a tool to test input variables and not to obtain output data. Given an ideal contagion curve over time, input variables are tested with different values, using transition rates between compartments obtained from the observed dataset.

However, the major problem identified in this process is that variables become discrete when tested in the models, which can cause discrepancies in the results obtained. For the simulation of a physical (real) scenario, the ideal would be continuous variables, thus accurately representing reality.

A study by Adeniyi et al. (2020), for example, assumes that immune individuals can lose their immunity over time. It's an exciting proposal. However, there is no scientific proof about it, nor about under what conditions it occurs. Furthermore, the rates used to transport an individual from the (I) Immune compartment to the (S) Susceptible compartment are random without memory. That is, they are not linked to the studied dataset.

Due to the characteristic of the numerical method, studies that apply time series use a small sample of the studied population (unit of analysis). The unit is observed sequentially over time, creating relations between the verified instant and the instant immediately before. Briefly, the unit of analysis is observed repeatedly, and the behavior obtained is reflected in the population. One of the problems with this type of analysis is that applying the model to a geographically heterogeneous population can result in substantial errors. Levitt, Scaiewicz, Zonta (2020), for example, used data from the United States and Italy and, given that the US has very distinct internal regional characteristics, the authors adjust the variables by curves with asymptotic behavior, determining a linear function to predict the plateau value. N. Already Saleh; Boj, (2020), and Tapiwa et al. (2020), who also work with time series, use datasets from Spain and China, respectively, so this brief analysis indicates caution in using the last model for a large population.

Models that use regression as a numerical method assume dependency relationships between the variables used. However, the order of observation over time is irrelevant. In practice, one variable will adjust in the function of others. Thus, the great challenge of these models is precisely to define the dependence and the actual relationship between the variables that make up that particular model. For teaching purposes, it is the same as saying that the number of deaths is related to the number of infected individuals or non-immunized individuals. This relationship may even be genuine. However, in the real world, the same model should not be used for different scenarios.

More than half of the articles analyzed presented proposals for compartmental models, the most classic ones, to solve biological problems. However, relating traditional models with more specialist numerical methods, such as time series and regression, can be an excellent alternative to adjust the more classical models. For example, these differentials were observed in the model proposed by Alboaneen et al. (2020), which combines the traditional SIR model with a logistic model to sharpen the growth rate of each population group.

The other compartmental models differ only in the number of groups into which the population is divided and the strategy used to calculate the balance between the respective groups using differential equations.

The models developed by Zhong et al. (2020) and Kumar et al. (2020) stand out since they do not follow the more classical behavior patterns of SARS-CoV-2.

Furthermore, regarding the fields of study of the authors of the articles studied, of the 32 concentration areas identified in the 25 articles evaluated, only 15% of the authors are from the health sector, while 60% are from the field of mathematics, computing or engineering, and 25% from specialized areas such as physics, chemistry, geology, geography, and oceanography. Considering that the study and development of mathematical models for the prediction of Covid-19 propagation concentrate two significant areas of knowledge, it was expected that these large sectors would unite knowledge to obtain more assertive answers.

A new aspect is how each model can be accurate and adjusted to be applied differently. For example, perhaps pure mathematical models are not sufficiently adjustable or configurable for a reality with different statistical histories than those already used.

Verifica-se que um dos fatores determinísticos para os resultados encontrados é especialmente o foco que se deseja dar aos resultados – economia, distanciamento, quarentena, rede hospitalar etc. It is known that one thing depends on the other. However, to achieve more assertive results, the variables must be stratified in a compartmentalized way, making visible only the characteristics and specialties of that respective variable.

5. Conclusion

When talking about a mathematical model, one of the main aspects to be observed is the purpose of building that particular model. Although the background predicts the contagion of a new disease, what determines the results found are the specific objectives, such as using specific techniques, how the database is used, and how the statistical data will be processed. In the field of predicting biological contagions, as is the case with SARS-CoV2, mathematical modeling should not aim to find exact answers to a physical problem. After all, they are overly complex and depend on several unpredictable factors. One of the recommended strategies is to simplify the real problem and, from there, start the search for a natural solution to that simplified problem. That way, there will be a real solution to a streamlined problem that can bring a possible solution to that complex real-world problem.

The future consequences of Covid-19 are difficult to predict in any real-life scenario, considering that the actual results of large-scale vaccine application are currently not known, nor are issues such as variants, strains, and levels of immunity. Furthermore, a series of operational difficulties need to be faced, such as the number of contagion testing, randomness, interventions, complacency in isolation, epidemiological realities, access to information, communication with public managers, in addition to many other factors that contribute to the difficulty in determining criteria for predicting virus contagion [3].

A primary concern is related to the validation of the models that are being developed. As can be seen, the models use specific criteria to solve particular problems of a

predetermined population. Furthermore, the results must be compared with actual data obtained over time.

After analyzing the proposed models and considering that the articles were written at least six months ago, it is identified that, in the physical world, all analyzed characteristics continue to change, requiring the models to adapt to each newly identified feature. For example, we are currently experiencing new pandemic phases, such as reinfection of individuals, new variants, and immunization provided by vaccines with different technologies for example. This further information can contribute to future mathematical models, but what is essentially expected of a predictive model is that it can identify essential characteristics that are still unknown and project the numbers over time. This situation is one of the main reasons that confirms our opinion regarding the multidisciplinary nature of studies involving predictive models for biological systems, as is the case with Covid-19.

In practice, decision-makers depend on a team of scientists to develop a specific mathematical model for their scenario or even on the adherence of one or a few mathematical models capable of guaranteeing the required specificities. It is concluded that it is challenging to have a model that responds in an organized way, in time, to how the pandemic will behave. What currently exists are models dedicated to answering specific aspects that can, in chains, be used in other models and so on until a set of values (answers) for a given scenario is obtained.

6. Conflict of interests

No conflict of interest was identified in the study developed.

References

- [1] A. M. Almeshal, A. I. Almazrouee, M. R. Alenizi, e S. N. Alhajeri, "Forecasting the spread of COVID-19 in kuwait using compartmental and logistic regression models", *Appl. Sci.*, vol. 10, n° 10, 2020, doi: 10.3390/APP10103402.
- [2] B. Ivorra, M. R. Ferrández, M. Vela-pérez, e A. M. Ramos, "Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China", n° January, 2020.
- [3] K. Abdulmajeed, M. Adeleke, e L. Popoola, "Online Forecasting of Covid-19 Cases in Nigeria Using Limited Data", *Data Br.*, vol. 30, 2020, doi: 10.1016/j.dib.2020.105683.
- [4] N. Linton *et al.*, "Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data", *J. Clin. Med.*, vol. 9, n° 2, p. 538, 2020, doi: 10.3390/jcm9020538.
- [5] Istituto Superiore di Sanità, "Report about the characteristics of death patients positive to COVID-19 in Italy (based on data updated on 17th March 2020)", p. 0–5, 2020.
- [6] S. H. A. Khoshnaw, R. H. Salih, e S. Sulaimany, "Mathematical modelling for coronavirus disease (COVID-19) in predicting future behaviours and sensitivity analysis", *Math. Model. Nat. Phenom.*, vol. 15, 2020, doi: 10.1051/mmnp/2020020.
- [7] S. Kim, Y. Bin Seo, e E. Jung, "Prediction of COVID-19 transmission dynamics using a mathematical model considering behavior changes in Korea", *Epidemiol. Health*, vol. 42, p. 1–6, 2020, doi: 10.4178/epih.e2020026.
- [8] B. Ambikapathy e K. Krishnamurthy, "Mathematical modelling to assess the impact of lockdown on COVID-19 transmission in India: Model development and validation", *J. Med. Internet Res.*, vol. 22, n° 5, p. 1–8, 2020, doi: 10.2196/19368.
- [9] Barbara Ann Kitchenham, D. B. Durham, e P. Brereton, *EVIDENCE-BASED SOFTWARE ENGINEERING AND SYSTEMATIC REVIEWS*. Taylor & Francis Group 6000, 2015.
- [10] R. F. Binyousef, A. M. Al-Gahmi, Z. R. Khan, e E. Rawah, "A rare case of Erdheim-Chester

- disease in the breast”, *Ann. Saudi Med.*, vol. 37, n° 1, p. 79–83, 2017, doi: 10.5144/0256-4947.2017.79.
- [11] D. Moher *et al.*, “Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement”, *PLoS Med.*, vol. 6, n° 7, 2009, doi: 10.1371/journal.pmed.1000097.
- [12] *MINISTÉRIO DA SAÚDE Brasília — DF 2017*. 2017.
- [13] H. M. Dobrovolny, “Modeling the role of asymptomatics in infection spread with application to SARS-CoV-2”, *PLoS One*, vol. 15, n° 8 August, p. 1–14, 2020, doi: 10.1371/journal.pone.0236976.
- [14] F. Ndairou, I. Area, J. J. Nieto, e D. F. M. Torres, “Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan”, n° January, 2020.
- [15] S. K. Dr e K. Sarkar, “Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India”, *arXiv*, p. 1–18, 2020.
- [16] M. O. Adeniyi *et al.*, “Dynamic model of {COVID}-19 disease with exploratory data analysis”, *Sci. African*, vol. 9, p. e00477, 2020, doi: 10.1016/j.sciaf.2020.e00477.
- [17] H. Zhao e Z. Feng, “SStaggered release policies for COVID-19 control: Costs and benefits of relaxing restrictions by age and risk”, n° January, 2020.
- [18] Y. T. Lin *et al.*, “Daily forecasting of new cases for regional epidemics of coronavirus disease 2019 with bayesian uncertainty quantification”, *medRxiv*, vol. 2019, 2020, doi: 10.1101/2020.07.20.20151506.
- [19] A. Kumar, P. Rani, R. Kumar, e V. Sharma, “Data-driven modelling and prediction of COVID-19 infection in India and correlation analysis of the virus transmission with socio-economic factors”, n° January, 2020.
- [20] K. Demertzis, D. Tsiotas, e L. Magafas, “Modeling and forecasting the COVID-19 temporal spread in Greece: an exploratory approach based on complex network defined splines.”, *arXiv*, 2020.
- [21] M. Levitt, A. Scaiewicz, e F. Zonta, “Predicting the trajectory of any COVID19 epidemic from the best straight line”, *medRxiv*, 2020, doi: 10.1101/2020.06.26.20140814.
- [22] A. Saleh e E. Boj, “SutteARIMA: Short-term forecasting method, a case: Covid-19 and stock market in Spain”, n° January, 2020.
- [23] G. Tapiwa *et al.*, “Estimating the generation interval for COVID-19 based on symptom onset data”, *medRxiv*, n° March, p. 1–8, 2020, doi: 10.1101/2020.03.05.20031815.
- [24] E. Z. Martinez, D. C. Aragon, e A. A. Nunes, “Short-term forecasting of daily COVID-19 cases in Brazil by using the holt’s model”, *Rev. Soc. Bras. Med. Trop.*, vol. 53, n° May, p. 1–5, 2020, doi: 10.1590/0037-8682-0283-2020.
- [25] J. A. T. Machado e A. M. Lopes, “Rare and extreme events: the case of COVID-19 pandemic”, *Nonlinear Dyn.*, vol. 100, n° 3, p. 2953–2972, 2020, doi: 10.1007/s11071-020-05680-w.
- [26] Q. Li, B. Tang, N. Luigi, Y. Xiao, e J. Wu, “Modeling the impact of mass influenza vaccination and public health interventions on COVID-19 epidemics with limited detection capability”, n° January, 2020.
- [27] L. Zhong, L. Mu, J. Li, J. Wang, Z. Yin, e D. Liu, “Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China Based on Simple Mathematical Model”, *IEEE Access*, vol. 8, p. 51761–51769, 2020, doi: 10.1109/ACCESS.2020.2979599.
- [28] D. Alboaneen, B. Pranggono, D. Alshammari, N. Alqahtani, e R. Alyaffer, “Predicting the epidemiological outbreak of the coronavirus disease 2019 (COVID-19) in Saudi Arabia”, *Int. J. Environ. Res. Public Health*, vol. 17, n° 12, p. 1–10, 2020, doi: 10.3390/ijerph17124568.

11 APÊNDICE B – CAPÍTULO DE LIVRO

Capítulo

X

Os Desafios na Utilização de Modelos Computacionais para Predição da Propagação da Covid-19

Uilian Loose, Juliana Herbert

Abstract

This chapter focuses on the challenges of choosing and using computational models to predict the spread of Covid-19 in times of a pandemic. It takes into account the main characteristics of computational models and how they have the ability to assist in decision making considering constant change. The difficulties faced in choosing the appropriate mathematical class for the prediction of each model take into account the history of the disease's behavior, biological information of the virus and the immunization capacity. In addition, in parallel with the current scenario of disease development, there is the challenge of relying on the information found, since certainty exists only after comparing the data generated by the predictive models with future real data.

Resumo

O presente capítulo foca nos desafios da escolha e utilização dos modelos computacionais para a predição da propagação da Covid-19 em tempos de pandemia. Leva-se em consideração as principais características dos modelos computacionais e como eles têm a capacidade de auxiliar na tomada de decisões em um momento de mudanças constantes. As dificuldades enfrentadas para a escolha da classe matemática adequada para a predição de cada modelo levam em consideração o histórico do comportamento da doença, informações biológicas do vírus e a capacidade de imunização. Além disso, em paralelo ao cenário atual de desenvolvimento da doença, há o desafio de confiar nas informações encontradas, uma vez que a certeza existe somente após comparar os dados gerados pelos modelos preditivos com dados reais futuros.

1.1. Introdução

As ferramentas computacionais são utilizadas para resolver problemas como a capacidade operacional que nós humanos não temos para processar a quantidade de dados existentes para tomadas de decisões. Há um grande esforço da comunidade científica na busca de informações capazes de auxiliar a compreensão do cenário atual, devido à Covid-19 [Almeshal et al., 2020] e uma das ferramentas utilizadas para a tomada de decisões é o modelo computacional de previsão da propagação do SARS-CoV-2. Informações de

transmissibilidade, letalidade e imunidade da Covid-19 em um determinado intervalo de tempo é essencial na definição de estratégias de gestão pública para o enfrentamento do combate à disseminação da doença, por que o sucesso dessas estratégias depende do conjunto de informações de comportamento obtidas através dos modelos computacionais.

Os desafios para a construção desses modelos preditivos são tão complexos quanto identificar as características e o comportamento de uma doença ainda desconhecida. Ainda, além dos diferentes modelos existentes, que utilizam técnicas preditivas distintas, os dados de entrada precisam ser criteriosamente classificados e sua confiabilidade deve ser verificada [De Veaux; Hand, 2005] a fim de se garantir o resultado mais adequado para situações específicas.

1.2. Os modelos computacionais

O grande desafio dos modelos computacionais preditivos é a capacidade de prever adequadamente novos casos de contaminação da Covid-19. Para que a previsão se aproxime da realidade fática, antes mesmo de decidir o tipo de modelo a ser utilizado para a predição, é necessário determinar quais são as respostas esperadas a partir das informações existentes, para que o algoritmo que gerará o modelo escolhido possa ser treinado adequadamente.

As respostas desejáveis se relacionam diretamente às informações já conhecidas, como por exemplo a taxa de transmissibilidade, taxa de letalidade, mapeamento da vulnerabilidade da população suscetível e a taxa de cura. Novas informações como mutações e variantes virais e eficácia imunológica também podem ser determinantes para a escolha do tipo do modelo a ser utilizado. Em um modelo computacional, essas informações são denominadas dados de entrada.

Conhecido o cenário epidemiológico, os dados históricos e os resultados desejáveis, é necessário definir qual a classe matemática ou estatística a ser utilizada para simular hipóteses que serão estabelecidas pelo próprio sistema através de padrões e critérios de sensibilização.

Os dados que subsidiarão a predição precisam ser pré-processados e coerentes entre si para que o modelo identifique padrões e comportamentos repetitivos gerando informações sensibilizadas para as próximas etapas da doença. Sensibilizar o sistema com novos comportamentos da doença como variantes do vírus, mudança estratégica na forma de imunização pela vacina aproximam os resultados obtidos da realidade temporal da doença.

Ou seja, o elemento central dos modelos computacionais é composto por um algoritmo preditivo baseado em estatística ou, os mais complexos, a partir de modelos híbridos formados a partir de várias técnicas como Regressão Logística, Redes Neurais Artificiais, Árvore de Decisão, Regras de Decisão, Aprendizagem Baseada em Casos e Classificadores Bayesianos e assim por diante. Cada um com uma proposta diferente que precisa ser considerada na escolha do modelo a ser utilizado, afinal, cada algoritmo possui suas próprias características, exigências e resultados.

1.3. Os dados de entrada e os resultados

Os dados de entrada determinam como o modelo deverá ser parametrizado através de padrões e sensibilizado para as mudanças nas características da doença que ocorrem constantemente.

A qualidade dos dados de entrada está relacionada à qualidade dos resultados obtidos pelos modelos, por isso, qualificar esses dados de entrada é uma ciência própria e engloba definir quais dados são importantes, classificá-los de acordo com as suas características, eliminar erros existentes e relacioná-los de maneira lógica. Esse trabalho

de limpeza e classificação exige que a pessoa que manipula os dados tenha conhecimento prévio do vírus, da situação epidemiológica e da doença causada pelo vírus.

As informações existentes sobre a Covid-19 são muito recentes e muitas dúvidas existem sobre a doença, o que aumenta o número de hipóteses dos modelos computacionais. Recentemente, os cientistas encontraram variantes do SARS-CoV-2, vírus causador da Covid-19, e não se sabe ainda quantas variantes poderão ser desenvolvidas, como elas reagem às vacinas já existentes e qual é o grau de letalidade dessas novas variantes. É importante que o modelo computacional possa ser sensibilizado por novos dados gerados a partir de informações obtidas no curso da doença. Sem conhecer completamente os novos desafios e rumos no desenvolvimento da doença, os modelos computacionais preditivos ficam mais complexos à medida que necessitam, paralelamente ao desenvolvimento da doença, se adaptarem aos novos comportamentos.

Do ponto de vista teórico, quanto mais flexível for o modelo computacional, maior a assertividade na previsão realizada, pois é maior o número de informações. No entanto, na prática, quanto mais variáveis, menor a chance do modelo se aproximar dos números ideais, pois mais deve ser a capacidade do modelo entender e fazer as relações corretas entre as variáveis existentes.

1.4. Os desafios para predição da propagação da Covid-19

Se não existissem os desafios, certamente teríamos o modelo ideal para prever a propagação do vírus responsável pela Covid-19. Fato é que o histórico da doença ainda é muito recente. Embora verifiquemos avanços significativos pela ciência na compreensão desse assunto, ainda estamos longe de garantir que o vírus não proporcionará novas surpresas a qualquer momento.

O comportamento computacional é binário, o processamento de um algoritmo preditivo nada mais é que a sequência de tomadas de decisões binárias e a cada tomada de decisão, o código direciona para trilhas distintas de decisão. A complexidade dos algoritmos preditivos encontra-se justamente na falta de conhecimento do comportamento do vírus e da doença. É importante lembrar que esses modelos são sensibilizados a partir da identificação de novos comportamentos do mundo real e como esse é um vírus novo, a quantidade de dados sobre a doença é pequena para identificar padrões comportamentais num ritmo pandêmico em que novas ações e decisões são tomadas todos os dias para cessar a propagação no mundo. Além da quantidade, os dados existentes não seguem padrões mínimos de armazenamento, são dimensionados sem critérios científicos, muitas vezes pelos critérios estabelecidos pela gestão pública com viés político. O cenário, portanto, já prevê erros uma vez que os dados existentes não demonstram a realidade fática enfrentada nesse mesmo período de extração das informações.

A gravidade da doença, considerando o binômio transmissibilidade e letalidade, ainda atrapalha o avanço de importantes estudos científicos, uma vez que é necessário observar as restrições físicas e comportamentais para o enfrentamento da pandemia pelos próprios cientistas. Como a taxa de transmissibilidade do vírus é bastante alta, foi gerado um alerta especial no comportamento social e esse comportamento gera consequências e reações espontâneas na economia, educação, desenvolvimento interno e na própria saúde. Esses comportamentos são, muitas vezes, imprevisíveis e tornam-se complexos para serem considerados nos modelos computacionais.

Outra complexidade já conhecida diz respeito às variantes do vírus. O SARS-CoV-2 é uma variante do coronavírus, e a partir dessa já foram identificadas novas variantes. Embora a vacinação já tenha iniciado, elas podem não imunizar para as novas variantes SARS-CoV-2 [Ng; Gui, 2020]. Adicionado a isso, as vacinas com eficácia

distintas desafiam ainda mais o desenvolvimento de modelos computacionais preditivos, pois, não há ainda, dados suficientes dos vacinados e os que existem ainda não são suficientes para determinar comportamentos.

Ainda que seja possível utilizar modelos desenvolvidos com outros tipos de coronavírus, a falta de experiências prévias específicas sobre o novo coronavírus dificulta a validação dos modelos preditivos. Somente a validação com os dados futuros revelará a capacidade de um determinado modelo de antecipar informações de forma segura para a tomada de decisão.

1.5. Conclusão

O processo de conhecimento da doença enfrenta muitos desafios. É difícil prever o futuro para as consequências da Covid-19, considerando-se qualquer cenário da vida real, ainda que atualmente não se conheça os resultados reais da aplicação da vacina em larga escala. Uma série de dificuldades operacionais precisam ser enfrentadas como o número de testagem de contágio, aleatoriedade, intervenções, complacência em casa, realidades epidemiológicas, acesso à informação, comunicação com os gestores públicos e muitos outros fatores que contribuem para a dificuldade de determinação de critérios para a previsão de contágio do novo coronavírus [Abdulmajeed; Adeleke; Popoola, 2020].

Os modelos computacionais para predição da propagação do novo coronavírus desempenham uma função fundamental, tanto para entender o curso da pandemia quanto para a concepção de estratégias públicas para o combate da Covid-19 [Sarkar; Khajanchi; Nieto, 2020], no entanto, conclui-se que o maior desafio existente na utilização de modelos computacionais diz respeito à validação dos modelos que estão sendo desenvolvidos. Os modelos utilizam critérios específicos para resolverem problemas pontuais de uma população pré-determinada. Além disso, é necessário que os resultados sejam comparados e confrontados com dados reais obtidos ao longo do tempo e essa informação só teremos no futuro.

1.6. Referências

- Abdulmajeed, K.; Adeleke, M.; Popoola, I. Online forecasting of covid-19 cases in Nigeria using limited data. *Data in Brief*, v. 30, p. 105683, 2020. DOI 10.1016/j.dib.2020.105683.
- Almeshal, A. M.; Almazrouee, A. I.; Alenizi, M. R.; Alhajeri, S. N. Forecasting the Spread of COVID-19 in Kuwait Using Compartmental and Logistic Regression Models. *Applied Sciences*, v. 10, n. 10, p. 3402, 2020. DOI 10.3390/app10103402.
- De Veaux, R. D.; Hand, D. J. How to lie with bad data. *Statistical Science*, v. 20, n. 3, p. 231–238, 2005. <https://doi.org/10.1214/088342305000000269>.
- Ng, K. Y.; Gui, M. M. COVID-19: Development of a robust mathematical model and simulation package with consideration for ageing population and time delay for control action and resusceptibility. *Physica D: Nonlinear Phenomena*, v. 411, p. 132599, 2020. DOI 10.1016/j.physd.2020.132599.
- Sarkar, K.; Khajanchi, S.; Nieto, J. J. Modeling and forecasting the COVID-19 pandemic in India. *Chaos, Solitons & Fractals*, v. 139, p. 110049, 2020. DOI 10.1016/j.chaos.2020.110049.

12 APÊNDICE C – O FRAMEWORK



GUIA

FRAMEWORK DE ANÁLISE DE RISCOS DE MODELOS PREDITIVOS

“Os modelos preditivos têm significativa importância para a ciência, sobretudo para o desenvolvimento da saúde. São eles que, muitas vezes, fornecem sínteses do estado do conhecimento (estado da arte), a partir do qual decisões são tomadas impactando a saúde coletivamente.”

Uilian Loose



SUMÁRIO

1	O FRAMEWORK	2
1.1	DEFINIÇÃO E ANÁLISE DE RISCOS	5
1.2	DEFINIÇÃO DE MODELOS PREDITIVOS	5
1.3	PÚBLICO-ALVO	6
1.4	ESTRUTURA	6
2	ORIENTAÇÕES	8
3	APLICAÇÃO DO <i>FRAMEWORK</i>	9
3.1	FASE 1 – OBJETIVO DO MODELO	9
3.2	FASE 2 – BASE DE DADOS DO MODELO	11
3.3	FASE 3 – VARIÁVEIS DO MODELO	14
4	REFERÊNCIAS	19

FRAMEWORK PARA ANÁLISE DE RISCOS DE MODELOS PREDITIVOS

A falta de histórico e informações sobre determinada doença, aliado à falta de planejamento de informações em saúde para o desenvolvimento de pesquisa pode comprometer fatalmente toda uma população.

O FRAMEWORK

Um dos maiores desafios enfrentados no desenvolvimento de modelos preditivos está relacionado justamente à disponibilidade e à classificação de dados disponíveis para o desenvolvimento de pesquisa, porque os bancos de dados relacionados à saúde, pelo menos no Brasil, possuem limitações importantes acarretando impactos significativos para o desenvolvimento científico. Outro desafio a ser considerado se refere à novidade que uma nova doença pode apresentar, quanto a sua forma, comportamento e todas as suas características. A falta de histórico e informações sobre determinada doença, aliado à falta de planejamento de informações em saúde para o desenvolvimento de pesquisa pode comprometer fatalmente toda uma população.

Com objetivo de enfrentar esses desafios, em especial no momento especial da pandemia da Covid-19, um framework foi desenvolvido para auxiliar os pesquisadores, desenvolvedores de modelos preditivos, na avaliação dos seus modelos e nos dados disponíveis para análise, considerando as condições legais que impactam na disponibilização de dados essenciais para o desenvolvimento de pesquisas emergenciais. Além da dificuldade logística, das incertezas e ameaças em produzir ciência durante a pandemia, a falta de planejamento para a produção e disponibilização de dados válidos e a dificuldade de acesso aos dados existentes é preocupante.

O *framework* é uma diretriz onde o elemento principal é um processo que orienta a realização de etapas, dependentes ou não, para tomada de decisões e ações através de artefatos associados.

A pandemia da Covid-19 foi utilizada como propulsora do desenvolvimento desse framework, quando veio à tona um grande problema enfrentado pelos pesquisadores brasileiros, o acesso aos dados indispensáveis para a produção de informações e pesquisa. O problema é ratificado pelo momento legal brasileiro. Recentemente entrou em vigor no Brasil a Lei Geral de Proteção de Dados – LGPD, que surge nesse cenário de necessidade de desenvolvimento de pesquisas na área da saúde com objetivo de regular as normas de acesso às informações pessoais, fortalecendo no setor privado ainda mais os interesses econômicos vinculados às informações das suas bases de dados.

Este framework tem a função de auxiliar na estrutura e planejamento de diretrizes para o pesquisador que atua em estudos envolvendo análises preditivas na área da saúde, trazendo subsídios técnicos multidisciplinares, que normalmente, são de áreas distintas da área de origem do pesquisador.

O framework é composto por fases bem definidas e cada uma dessas fases é composta por artefatos estruturados que orientam o avaliador na tomada de decisões e ações. De acordo com as informações do avaliador, em cada fase, aspectos jurídicos são mapeados gerando, ao final, um relatório com sugestões para mitigação de riscos envolvendo aquele modelo preditivo desenvolvido.

A proposta é que os avaliadores sejam capazes de mapear os riscos envolvidos em um determinado modelo preditivo aplicado à saúde. Além de auxiliar os desenvolvedores de modelos preditivos, este framework também pode auxiliar o público que deseja utilizar um determinado modelo matemático já desenvolvido anteriormente.

LGPD – Lei Geral de Proteção de Dados: Lei nº 13.709/2018.

Covid-19: é uma doença causada por infecção respiratória aguda causada pelo coronavírus SARS-CoV-2, potencialmente grave, de elevada transmissibilidade e de distribuição global¹.

1- www.gov.br/saude

Os modelos preditivos têm significativa importância para a ciência, sobretudo para o desenvolvimento da saúde. São eles que, muitas vezes, fornecem sínteses do estado do conhecimento (estado da arte), a partir do qual decisões são tomadas impactando a saúde coletivamente.

Os aspectos legais para o acesso e utilização dos dados em saúde são desafiadores para a pesquisa brasileira. Além das normas colaborarem para a dificuldade de acesso aos dados indispensáveis para a pesquisa, muitas vezes, o próprio pesquisador pode estar vulnerável na manipulação de determinados dados, inclusive podendo ser responsabilizado.

O que se vê na área científica hoje é uma grande dificuldade de acesso às informações da população que se deseja estudar que, por vezes, são permeadas por questões complexas que envolvem estratégia de mercado, economia, interesse político e desinformação que foram agravadas pelo novo marco regulatório brasileiro referente à proteção de dados pessoais.

Os dados referentes à saúde da população são especialmente afetados, porque a LGPD os classifica como sensíveis e esses apenas podem ser tratados para fins específicos. Esse pressuposto é levado em consideração pelas grandes organizações, como justificativa, para não disponibilizarem dados essenciais para a pesquisa em saúde.

Embora a própria lei regulamente a disponibilização de dados para que sejam realizados estudos por órgão de pesquisa, sabemos que, economicamente falando, não é interessante para as empresas a divulgação de suas bases de dados, provocando, sob justificativa da LGPD, um retrocesso da pesquisa brasileira.

O que não pode acontecer é que a legislação que vem para proteger a intimidade das pessoas no que diz respeito à divulgação de dados pessoais, prejudicar ou não dar condições necessárias para o desenvolvimento de pesquisas que são de interesse coletivo, ou melhor, ações que definitivamente podem salvar vidas, como é o caso da Covid-19. Quando esse interesse coletivo é maior do que os interesses individuais, há aí uma questão social a ser discutida, os interesses coletivos precisam ser levados em consideração e, mais do que isso, precisam nortear ações de políticas públicas capazes determinar

que a pesquisa em saúde, na maioria das vezes, deverá prevalecer aos interesses econômicos.

Também é importante dizer que não é porque um determinado estudo utilizará informações sensíveis de um determinado grupo de pessoas que essas informações não serão protegidas ou que essas informações serão divulgadas sem o interesse direto ou sem o consentimento daquele indivíduo.

Uma das formas de auxiliarmos os cientistas no acesso às informações é justamente subsidiá-los de conhecimentos multidisciplinares para que, além de utilizarem como argumento técnico, tenham meios de indicar mudanças normativas em prol do desenvolvimento científico do país.

Em resumo a lei não pode inviabilizar o desenvolvimento científico, tampouco deve servir de ferramenta contrária ao desenvolvimento técnico de uma sociedade, de forma a privilegiar interesses particulares em detrimento de interesses coletivos.

O framework passa por quatro fases, uma dependente da outra e cada fase composta por artefatos que direcionam o pesquisador na análise dos riscos em cada etapa.

Definição de Análise de Riscos

Identificar possíveis problemas que podem atrair responsabilidades indesejáveis aos pesquisadores, financiadores ou responsáveis pelos dados utilizados e possíveis impactos negativos que os resultados obtidos pelo modelo podem gerar na sociedade.

Definição de Modelos Preditivos

Conforme a definição de Araújo (2002) “modelo é uma representação simplificada de uma situação concreta feita com o objetivo de compreender a situação e prever suas configurações futuras ou de situações semelhantes”, ainda, nesse sentido, Goldbarg e Luna (2005) definem os modelos como “representações simplificadas da realidade que preservam, para determinadas situações e enfoques, uma equivalência adequada”.

Os modelos podem ser utilizados para simular situações, considerando um conjunto de parâmetros pré-definidos estatísticos ou convencionados com base em determinadas regras. Os modelos ainda podem ser utilizados para se chegar a valores ideais, a partir do comportamento histórico do evento mapeado.

Público-Alvo

O framework foi desenvolvido para auxiliar pesquisadores que trabalham com modelos preditivos em saúde sobretudo com relação aos riscos envolvidos na utilização de dados.

O contexto da saúde envolve competências multidisciplinares e, nem sempre é fácil para o desenvolvedor identificar os riscos envolvidos em pesquisas que envolvem dados sensíveis de saúde. Portanto, a proposta é que a ferramenta também alerte o pesquisador para aspectos relevantes de áreas do conhecimento que, podem, não ser do seu domínio.

Além dos pesquisadores, gestores públicos, elaboradores de diretrizes, organizações que apoiam a tomada de decisões e profissionais interessados em análises preditivas na área da saúde são possíveis usuários interessados na ferramenta.

ESTRUTURA

O framework está dividido em três grandes fases: (1) Processo, (2) Fase de Análise e (3) Relatório de Riscos. Cada fase possui um conjunto de artefatos divididos em etapas para facilitar o trabalho a ser desenvolvido pelo avaliador.

1 – Processo: objetiva direcionar o avaliador pelo framework. Um fluxograma direcionará o avaliador para as demais fases e blocos, de forma que o fluxo de trabalho do avaliador seja contínuo até o final da sua análise.

2 – Fase de Análise: objetiva provocar o avaliador para analisar de maneira crítica aspectos importantes, previamente mapeadas. Essa fase é

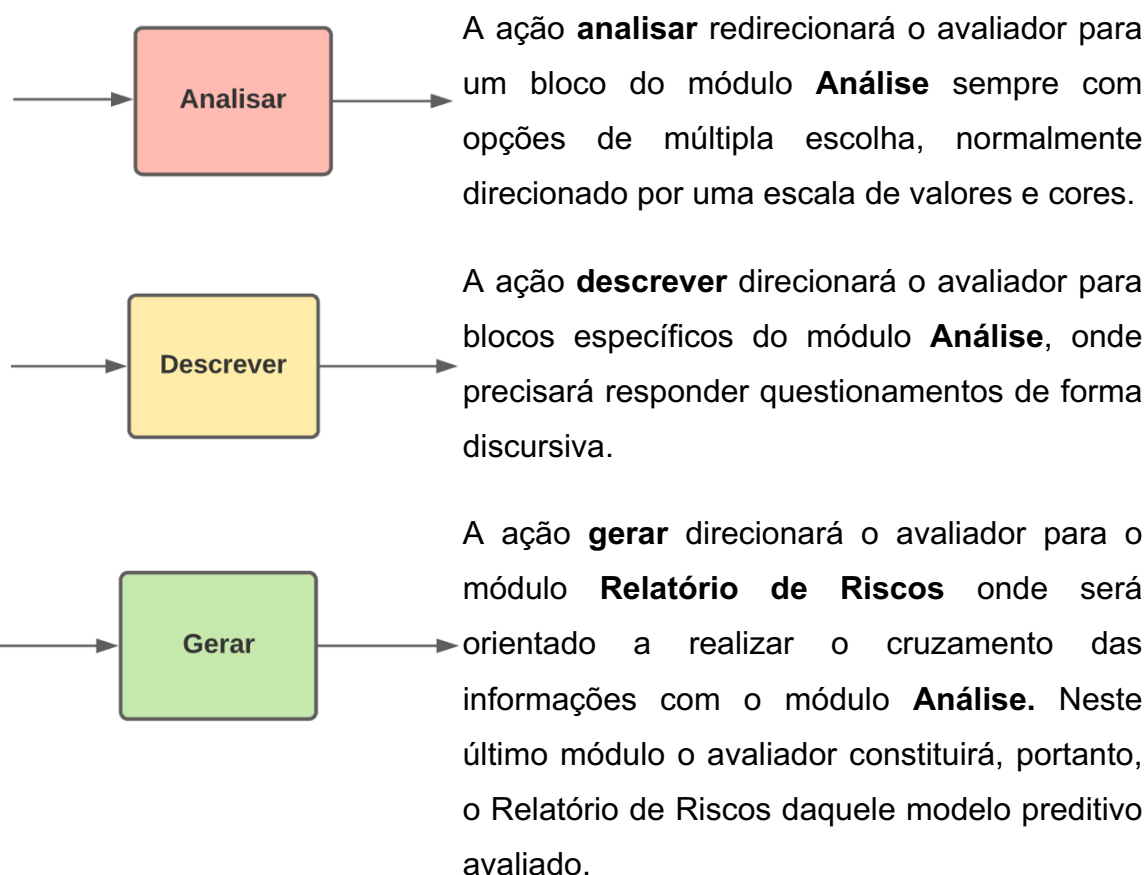
dividida em três grandes sessões e cada sessão, por sua vez, dividida em blocos temáticos. O avaliador não deverá analisar os blocos da fase sem ter o devido redirecionamento da Fase 1.

3 – Relatório de Riscos: a última fase do framework é a fase de compatibilidade das proposições realizadas na Fase 2 com as indicações de riscos e sugestões para o modelo analisado.

O contexto da COVID-19 no Brasil foi utilizado para o desenvolvimento desse trabalho com objetivo de limitar o escopo e direcionar a ferramenta para o problema enfrentado atualmente, no entanto, o framework foi desenvolvido para avaliar qualquer tipo de modelo preditivo no contexto da saúde.

Avaliador – é a pessoa que aplicará o framework para análise de um determinado modelo preditivo.

Um macroprocesso (fluxograma) norteia o framework direcionando o avaliador para as ações que deverá tomar nas diferentes fases do fluxo. O fluxograma apresentará ao avaliador três tipos de processos:

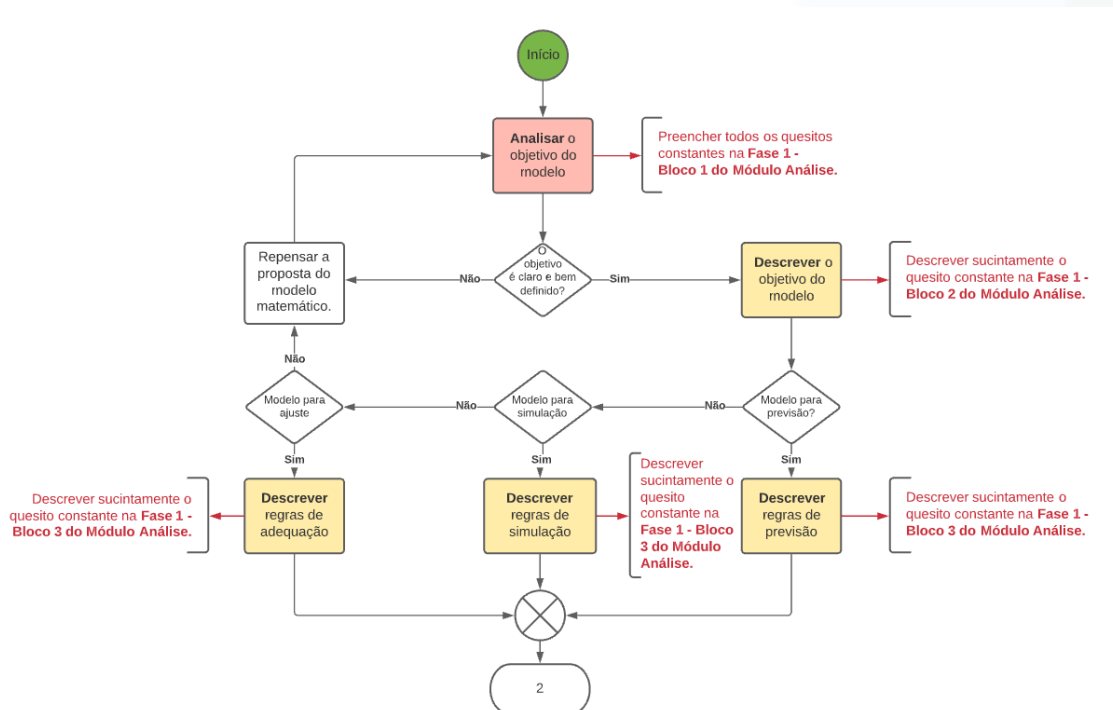


3 APLICAÇÃO DO FRAMEWORK

Para iniciar a aplicação do framework, o avaliador deverá iniciar pelo Anexo 1, seguindo os passos definidos no quadro de orientações.

3.1 FASE 1 – OBJETIVO DO MODELO

Na maioria das vezes, o objetivo do modelo será solucionar um determinado problema. Para isso, é fundamental identificar qual é o objetivo do modelo matemático e se ele está bem definido e claramente descrito. Essa fase é fundamental porque a falta de clareza e de definição impedem a continuidade de avaliação. O objetivo dessa fase é que ao final da avaliação, o pesquisador verifique se os resultados encontrados estão em conformidade com as definições do objetivo, além, claro, de identificar os possíveis riscos existentes no objetivo do modelo.



Avaliar e classificar o objetivo do modelo

Ao chegar nesse ponto, o avaliador deverá ir até o **Anexo 2** (Relatório de Análise). Na **Etapa 1** do anexo, o avaliador deverá responder os dois quesitos

do **Bloco 1**. Nesse bloco, o avaliador deverá analisar se aquele tipo de modelo já existe, se há inovação ou se há uma releitura de outro modelo já existente. Considerando uma escala de riscos de 1 a 4.

Além disso, o avaliador deverá analisar se o modelo submetido resolverá um problema coletivo ou um problema individual. Importante observar que, quanto mais coletivo for o problema resolvido, menor o risco legal, levando em consideração que, o interesse coletivo prevalece ao interesse individual.

Descrever o objetivo do modelo

Na **Fase 1 - Bloco 2** do Anexo 2, o avaliador deverá descrever, de maneira objetiva e direta qual é o principal objetivo do modelo.

FASE 1 - Bloco 2

DESCREVER O OBJETIVO DO MODELO
Descreva de forma clara e objetiva em, no máximo, duas linhas o objetivo do modelo matemático.
<input type="text"/>

Caso não seja possível identificar de forma clara o objetivo do modelo analisado, é recomendável que o modelo seja repensado.

Descrever as regras do modelo

A **Fase 1 - Bloco 3** é dividida em dois momentos: (1) tipo do modelo preditivo selecionando se o modelo é puramente preditivo, se ele serve para fazer simulações ou faz ajustes de função.

FASE 1 - Bloco 3											
DESCREVER REGRAS											
1	Modelo para previsão de dados futuros: marque essa opção se o modelo avaliado utiliza uma base de dados com valores bem definidos e valores reais.										
2	Modelo para simulação de cenários: marque essa opção se a base de dados utilizada possui valores simulados ou fictícios.										
3	Modelo para ajuste de valores: marque essa opção se o autor desenvolveu o modelo para encontrar valores ideais de variáveis. Nesses casos, o que se pretende, na verdade, é ir ajustando as variáveis para que a função se ajuste aos valores ideais.										
Descreva de forma clara quais são as regras básicas do modelo.											
<table border="1" style="width: 100%; height: 100px;"> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td> </td></tr> <tr><td> </td></tr> </table>											

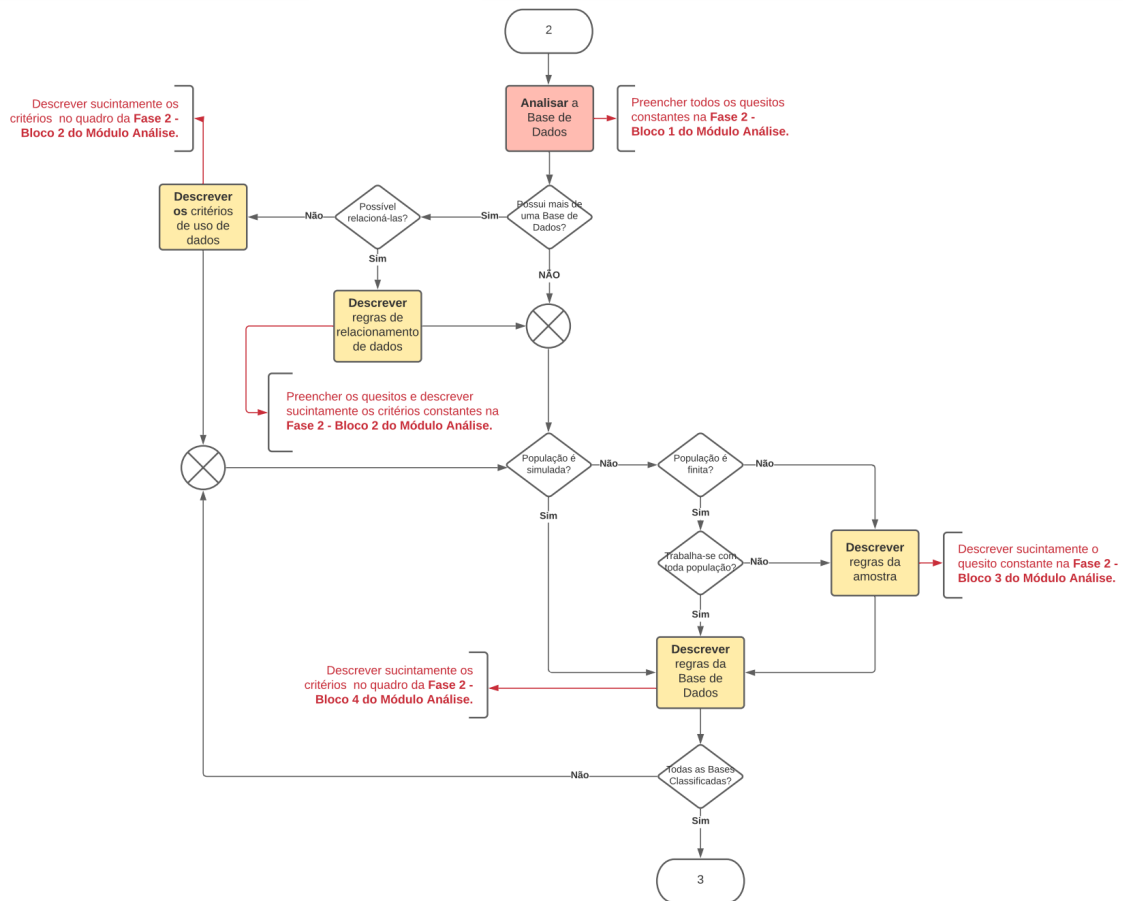
No campo seguinte, imediatamente o avaliador deverá incluir, de forma clara, as regras do modelo preditivo.

FASE 2 – BASE DE DADOS DO MODELO

A população utilizada no modelo preditivo, normalmente, parte das bases de dados existentes sobre o assunto. Entender a origem dos dados, a contextualização e a confiança dessas bases é fundamental para a confiança das respostas trazidas pelo modelo.

A falta de histórico, informações coletadas de forma não planejada ou intervenções nas bases as tornam menos confiáveis e muito limitadas. Por vezes as bases de dados precisam ser relacionadas e esse próprio processamento gera riscos aos dados lá contidos.

O objetivo dessa fase é provocar o avaliador com relação ao grau de confiança da base de dados utilizada como base para o processamento do modelo. Não há nenhuma pretensão de classificar a base de dados como aplicável ou não, mas tão somente orientar o pesquisador com relação aos riscos diretamente vinculados às bases utilizadas no estudo.



Descrever as regras do modelo

A **Fase 2**, é dividida em quatro blocos e objetiva analisar a base de dados de acordo com o seu propósito inicial, regras de estrutura, estratégias de gerenciamento dos dados e o grau de publicidade/disponibilidade dessas regras. O avaliador terá condições de avaliar se os dados contidos nessa base possuem qualquer viés político ou comercial.

A **Fase 2** está dividida em quatro blocos:

Bloco 1 que está dividido em quatro momentos distintos. Após identificar a origem da base de dados, quesitos sobre regras de ajustes de dados, disponibilidade dos dados e possíveis manipulações são realizados para que possíveis vieses sejam considerados na avaliação de risco daquela Base de Dados.

FASE 1 - Bloco 1		
ANALISAR O OBJETIVO DO MODELO		OBSERVAÇÕES
1- Com relação ao problema a ser resolvido pelo modelo preditivo:		
1	não existem modelos preditivos que solucionem o problema.	Ao selecionar essa opção, o avaliador está querendo dizer que buscou nas publicações existentes modelos matemáticos que pudessem resolver aquele problema específico e não encontrou.
2	não existem modelos que solucionem o problema com as peculiaridades propostas.	Ao selecionar essa opção, o avaliador está querendo dizer que buscou nas publicações existentes modelos matemáticos que pudessem resolver um problema específico. Encontrou modelos mas não com a peculiaridade específica avaliada pela solução proposta.
3	o modelo utiliza outro modelo existente, apenas são feitos ajustes para resolver o problema proposto.	Ao selecionar esta opção, o avaliador está querendo dizer que a solução proposta utiliza outro modelo matemático como base mas faz ajustes para melhorá-lo.
4	existem inúmeros modelos que solucionam o problema.	Ao selecionar essa opção, o avaliador está querendo dizer que buscou nas publicações existentes modelos matemáticos que pudessem resolver aquele problema específico e encontrou inúmeras soluções para resolver o problema. Esse modelo avaliado será mais uma solução proposta.
2- A resolução do problema:		<i>Aqui não se deve levar em consideração o impacto da pesquisa na sociedade, mas na aplicabilidade do modelo matemático.</i>
1	resolverá um problema coletivo, impactará toda a sociedade.	Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta impactará toda a sociedade brasileira.
2	resolverá um problema coletivo que não impacta toda a sociedade.	Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta resolverá um problema social, que tem interesse coletivo mas não impacta necessariamente em toda a sociedade.
3	resolverá o problema de um pequeno grupo de pessoas.	Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta impactará em um grupo pequeno de pessoas isoladas. Nesse caso não há interesse coletivo direto.
4	resolverá um problema específico, para uso particular.	Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta nesse momento é de interesse do próprio pesquisador ou de uma pessoa específica.

Descrever as regras de relacionamento entre bases de dados

Bloco 2, objetiva descrever as regras de relacionamento entre base de dados, caso, no modelo analisado, sejam utilizadas mais de uma base de dados.

As relações entre base de dados são formas de combinar os dados entre várias bases de dados para análises, onde a relação define como duas bases de dados se vinculam entre si através de campos ou identificadores comuns.

IMPORTANTE

se o modelo utilizar apenas uma base de dados, os blocos 2 e 3 não deverão ser preenchidos.

Quando é estabelecida uma relação entre os bancos de dados, eles permanecem separados, mantendo, cada um, com o seu nível individual de detalhe e domínios.

Descrever as regras da amostra

Nem sempre todos os dados da base de dados são utilizados para o estudo, nesse caso, define-se uma amostra de dados (inferência amostral), e, a partir daí a população estudada é definida por essa amostra específica. No **Bloco 3** o avaliador deverá descrever as regras dessa amostra, caso exista.

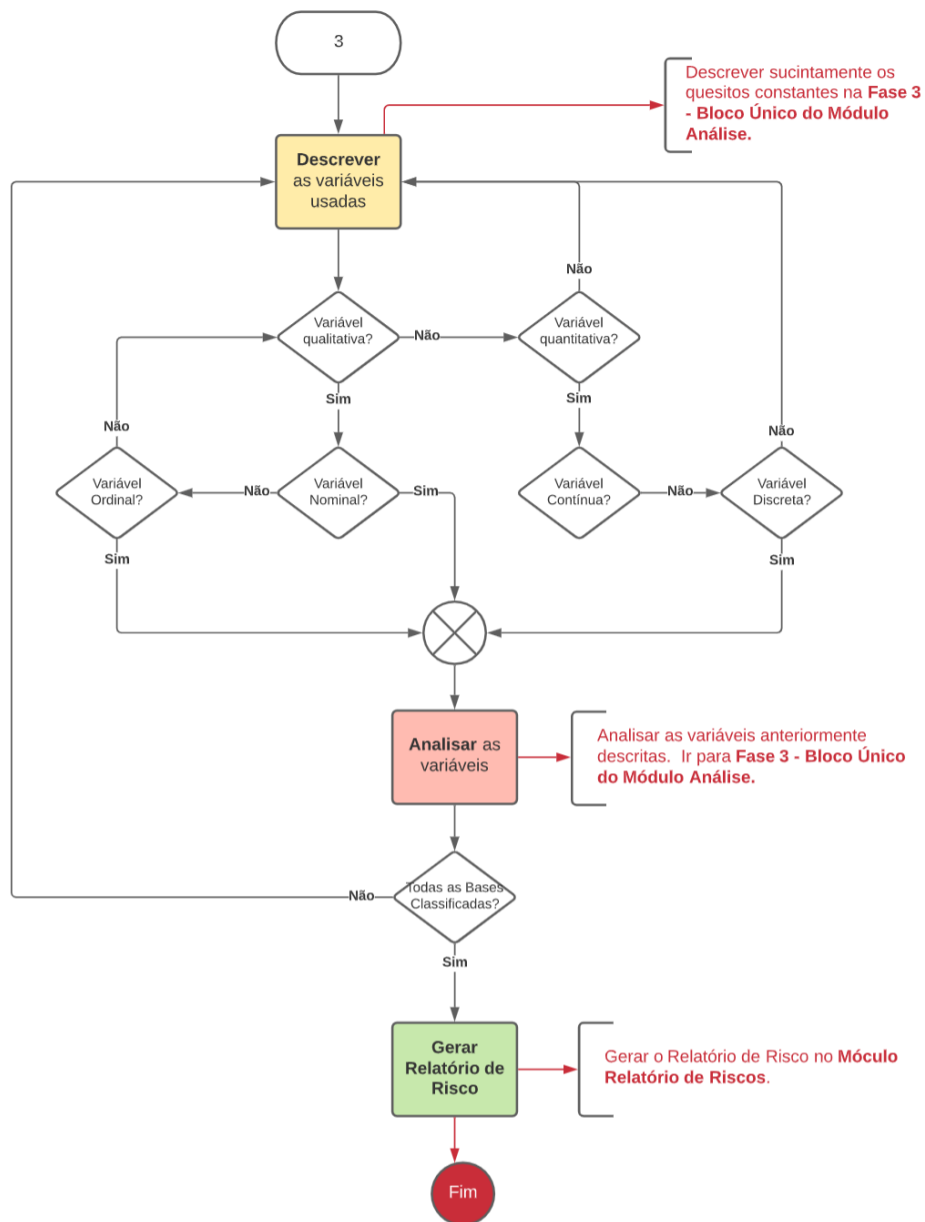
FASE 2 - Bloco 3			
DESCREVER REGRAS DA AMOSTRA - Esse bloco somente deverá ser preenchido se houver inferência amostral!			
1- A amostra casual é do tipo:			
1	Equiprobabilística.	<i>Nessa amostra todos os elementos da população original possuem as mesmas chances de estarem na amostra.</i>	<input type="checkbox"/>
2	Randomica.	<i>No critério randomizado não há nenhum tipo de regra. Os elementos são escolhidos aleatoriamente.</i>	<input type="checkbox"/>
3	Sistemática.	<i>A amostra definida por critério sistemático requer regra pré definida que que não esteja diretamente relacionada ao valor do elemento da população.</i>	<input type="checkbox"/>
4	Extratificada.	<i>Nessa opção de amostra extratificada a inferência é feita de forma segmentada utilizando valores definidos dos elementos da base.</i>	<input type="checkbox"/>
Descreva de forma clara e objetiva as regras utilizadas para inferência amostral.			
<input type="text"/>			
<input type="text"/>			
<input type="text"/>			
<input type="text"/>			
<input type="text"/>			
<input type="text"/>			
<input type="text"/>			

Descrever as regras da base de dados

Por fim, no **bloco 4** o avaliador descreverá, objetivamente, as regras da base de dados submetida ao modelo, exemplo: data de captura, tamanho da base, unidades de medida utilizadas, critérios de atualização dos dados etc.

FASE 3 – VARIÁVEIS DO MODELO

A terceira e última fase do framework requer uma análise pormenorizada das variáveis submetidas ao modelo. As características de cada variável assumem um papel importante para que o avaliador entenda a relação existente entre elas no modelo e ainda possa estabelecer eventuais conflitos ou riscos por conta das características.



Descrever as características das variáveis

A **Fase 3**, possui **Bloco Único** e, por sugestão, deverá ser preenchido seguindo essas etapas, para melhor observação do avaliador:

- 1- Listar, no campo específico todas as variáveis utilizadas no modelo e sua respectiva notação.
- 2- Classificar cada uma das variáveis listadas como qualitativa ou quantitativa.

O Relatório de Riscos, **Anexo 3**, objetiva, ao final da análise, fazer um panorama geral dos riscos a serem observados pelo pesquisador naquele modelo preditivo específico.

Os riscos considerados dizem respeito aos resultados obtidos pelos modelos analisados. No parecer, pode haver indicação de riscos relacionados: (1) a propriedade intelectual de pesquisa, (2) consentimento informado livre e esclarecido, (3) termo de responsabilidade, (4) assentimento e (5) parecer do comitê de ética.

Os blocos dispostos no Anexo 3 são exatamente iguais aos blocos do Anexo 2, a fim de que o avaliador possa fazer a relação dos anexos, e alcançar os resultados buscados.

Exemplo: se o avaliador, na **Fase 1 - Bloco 1** selecionar a opção dois, no Relatório de Risco, **Anexo 3**, deverá selecionar a mesma opção e considerar somente as orientações e riscos considerados naquela respectiva opção.

Módulo Análise

FASE 1 - Bloco 1	
ANALISAR O OBJETIVO DO MODELO	OBSERVAÇÕES
1- Com relação ao problema a ser resolvido pelo modelo preditivo:	
2	<p>não existem modelos que solucionem o problema com as peculiaridades propostas.</p> <p><i>Ao selecionar essa opção, o avaliador está querendo dizer que buscou nas publicações existentes modelos matemáticos que pudessem resolver um problema específico. Encontrou modelos mas não com a peculiaridade específica avaliada pela solução proposta.</i></p>
	<p>peculiaridades propostas.</p> <p><i>Encontrou modelos mas não com a peculiaridade específica avaliada pela solução proposta.</i></p>
	<p>o modelo utiliza outro modelo existente, apenas são feitos ajustes para resolver o problema proposto.</p> <p><i>Ao selecionar esta opção, o avaliador está querendo dizer que a solução proposta utiliza outro modelo matemático como base mas faz ajustes para melhorá-lo.</i></p>
	<p>existem inúmeros modelos que solucionam o problema.</p> <p><i>Ao selecionar essa opção, o avaliador está querendo dizer que buscou nas publicações existentes modelos matemáticos que pudessem resolver um problema específico e encontrou inúmeras soluções para resolver o problema. Esse modelo avaliado será mais uma solução proposta.</i></p>

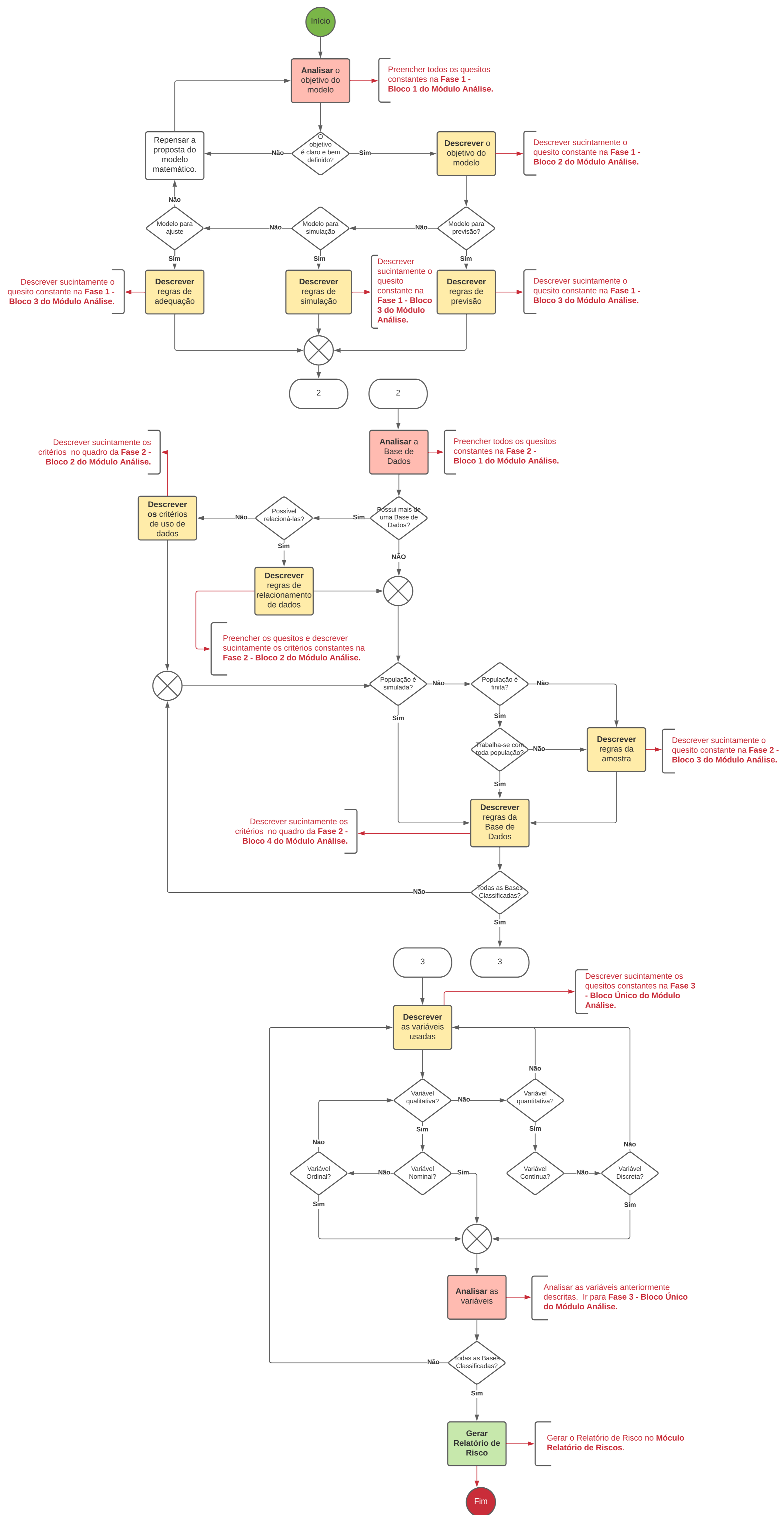
Módulo Relatório de Risco

FASE 1 - Bloco 1	
ANALISAR O OBJETIVO DO MODELO	OBSERVAÇÕES
1- Com relação ao problema a ser resolvido pelo modelo preditivo:	
1	<p>RISCOS: ALTO RISCO - risco da não existência de dados seguros para subsidiar o modelo. A novidade de uma doença normalmente vem acompanhada da falta de informações históricas.</p> <p>PARECER: Incluir, de forma clara, de onde a base de dados foi extraída para o desenvolvimento do trabalho e acrescentar o alto risco de discrepância que pode existir entre os resultados obtidos pelo modelo e os dados apresentados no cenário estudado. Justificar o risco assumido utilizado a novidade do problema e a predominância do interesse coletivo em detrimento do interesse individual.</p>
2	<p>RISCOS: ALTO RISCO - risco da não existência de dados seguros para subsidiar o modelo e risco de não fazer referências aos trabalhos desenvolvidos previamente. O aspecto específico adotado no modelo não exige o pesquisador de fazer referências aos trabalhos realizados anteriormente que utilizam um modelo base já proposto.</p> <p>PARECER: Incluir, de forma clara, de onde a base de dados foi extraída para o desenvolvimento do trabalho e acrescentar o alto risco de discrepância que pode existir entre os resultados obtidos pelo modelo e os dados apresentados no cenário estudado. Justificar o risco assumido utilizado a novidade do problema e a predominância do interesse coletivo em detrimento do interesse individual. Incluir uma sessão que descreva os modelos bases identificados previamente.</p>
3	<p>RISCOS: BAIXO RISCO - como o trabalho utiliza como referência outro modelo já desenvolvido, a responsabilidade é compartilhada.</p> <p>PARECER: Incluir, de forma clara, o modelo utilizado como base para resolução do problema específico. Para não haver infração de propriedade, é importante incluir, precisamente a referência do modelo utilizado e nome dos autores. Caso o modelo não tenha sido publicado, um termo de aceite deve ser devidamente assinado pelos autores.</p>
4	<p>RISCOS: BAIXO RISCO - não é esperado que o modelo apresente resultados discrepantes à realidade. Há risco de apresentar resultados muito próximos ou idênticos a resultados apresentados em trabalhos anteriores. No entanto, o interesse individual é mais evidente que o interesse coletivo.</p> <p>PARECER: Se for utilizado um modelo já conhecido como base de desenvolvimento, deve se indicar objetivamente qual é esse modelo, por quem foi desenvolvido e quais são as principais diferenças entre aquele parâmetro e o desenvolvido.</p>

REFERÊNCIAS

- [1] J. Araújo, “Cálculo, Tecnologias e Modelagem Matemática: as discussões dos alunos”, *Univ. Estadual Paul.*, p. 180, 2002.
- [2] M. C. Goldberg e H. P. L. Luna, *Otimização combinatória e programação linear: modelos e algoritmos*. 2005.

Anexo 1 - MÓDULO PROCESSO



Anexo 2 - MÓDULO ANÁLISE

Avaliador: _____

Data: ____ / ____ / ____

FASE 1 - OBJETIVOS

FASE 1 - Bloco 1

ANALISAR O OBJETIVO DO MODELO

OBSERVAÇÕES

1- Com relação ao problema a ser resolvido pelo modelo preditivo:

1	não existem modelos preditivos que solucionem o problema.	<i>Ao selecionar essa opção, o avaliador está querendo dizer que buscou nas publicações existentes modelos preditivos que pudessem resolver aquele problema específico, mas, não os encontrou.</i>
2	não existem modelos que solucionem o problema com as peculiaridades propostas.	<i>Ao selecionar essa opção, o avaliador está querendo dizer que buscou nas publicações existentes modelos matemáticos que pudessem resolver um problema específico. Encontrou modelos mas não com a peculiaridade específica avaliada pela solução proposta.</i>
3	o modelo utiliza outro modelo existente, apenas são feitos ajustes para resolver o problema proposto.	<i>Ao selecionar esta opção, o avaliador está querendo dizer que a solução proposta utiliza outro modelo matemático como base mas faz ajustes para melhorá-lo.</i>
4	existem inúmero modelos que solucionam o problema.	<i>Ao selecionar essa opção, o avaliador está querendo dizer que buscou nas publicações existentes modelos matemáticos que pudessem resolver aquele problema específico e encontrou inúmeras soluções para resolver o problema. Esse modelo avaliado será mais uma solução proposta.</i>

2- A resolução do problema:

Aqui não se deve levar em consideração o impacto da pesquisa na sociedade, mas na aplicabilidade do modelo matemático.

1	resolverá um problema coletivo, impactará toda a sociedade.	<i>Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta impactará toda a sociedade brasileira.</i>
2	resolverá um problema coletivo que não impacta toda a sociedade.	<i>Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta resolverá um problema social, que tem interesse coletivo, mas, não impacta necessariamente em toda a sociedade.</i>
3	resolverá o problema de um pequeno grupo de pessoas.	<i>Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta impactará em um grupo pequeno de pessoas isoladas. Nesse caso não há interesse coletivo direto.</i>
4	resolverá um problema específico, para uso particular.	<i>Ao selecionar essa opção o avaliador está querendo dizer que a solução proposta nesse momento é de interesse do próprio pesquisador ou de uma pessoa específica.</i>

FASE 1 - Bloco 2

DESCREVER O OBJETIVO DO MODELO

Descreva, em no máximo duas linhas, o objetivo do modelo preditivo.

FASE 1 - Bloco 3

DESCREVER REGRAS

1	Modelo para previsão de dados futuros: marque essa opção se o modelo avaliado utiliza uma base de dados com valores bem definidos e valores reais.
2	Modelo para simulação de cenários: marque essa opção se a base de dados utilizada possui valores simulados ou fictícios.
3	Modelo para ajuste de valores: marque essa opção se o autor desenvolveu o modelo para encontrar valores ideais de variáveis. Nesses casos, o que se pretende, na verdade, é ir ajustando as variáveis para que a função se ajuste aos valores ideais.

Descreva de forma clara quais são as regras básicas do modelo.

FASE 2 - BASE DE DADOS

FASE 2 - Bloco 1

BASE DE DADOS

OBSERVAÇÕES

1- A base de dados é pública?

As bases de dados públicas podem possuir interesses políticos. Um exemplo clássico disso refere-se ao consórcio nacional de dados, após a mudança do governo em relação a disponibilidade de informações referente à Covid-19.

As bases privadas podem deixar de trazer informações estratégicas de interesse comercial.

SIM	Numa escala de 1 - 4, qual a chance dos dados terem interferência política?	1	2	3	4	1- Nenhuma chance do modelo ter interferência de decisões políticas. 2- É possível que o modelo tenha alguma interferência por razões de decisões políticas. 3- É grande a possibilidade do modelo sofrer interferências por razões de decisões políticas. 4- O modelo tem interferência por razões de decisões políticas.	<input type="checkbox"/>
NÃO	Numa escala de 1 - 4, qual a chance dos dados serem influenciados por interesses comerciais?	1	2	3	4	1- Nenhuma chance do modelo ter influência por interesses comerciais. 2- É possível que o modelo tenha alguma interferência por razões comerciais. 3- É grande a possibilidade do modelo sofrer interferências por razões comerciais. 4- O modelo tem interferência por razões comerciais.	<input type="checkbox"/>

2- É possível que dados históricos que já estão na base de dados sejam alterados?

Os dados disponibilizados em canais web, atualizados em tempo real podem sofrer mudanças estratégicas na coleta e armazenamento de dados sem a devida informação.

Em regra, as bases públicas são disponibilizadas em intervalos temporais. Os dados, nesse caso, costumam ser estocásticos o que dificulta a mudança dos dados históricos.

SIM						4- Base de dados atualizada em tempo real, é possível que os dados sufram alterações nas regras de captura e armazenamento ao longo do tempo.	<input type="checkbox"/>
NÃO						1- Base de dados arquivada em intervalos temporais. Os intervalos podem sofrer alterações ao longo do tempo.	<input type="checkbox"/>

3- É possível conhecer e acompanhar toda a regra de coleta e armazenamento das informações na base de dados?

Quando se conhece a regra de captura dos dados e a forma que esses dados são armazenados, gerenciados e distribuídos, em regra, se consegue determinar com maior precisão os resultados do modelo.

As bases privadas podem deixar de trazer informações estratégicas de interesse comercial.

SIM						1- Nesses casos é possível saber da onde veio o dado, como foi capturado e fazer, inclusive, a validação do dado.	<input type="checkbox"/>
NÃO						4- Nesses casos as informações de captura não são armazenadas e os dados passam por vários processos de cadastro e processamento até o banco de dados final (o que é utilizado pelo modelo)	<input type="checkbox"/>

4- É possível que as regras de cadastro dos dados sejam alteradas?

As bases descentralizadas têm grandes chances de terem as regras de armazenamento alteradas.

Em contrapartida, as bases de dados que trabalham com tabulação de " dado atribuído ou dado selecionado ", dificilmente alteram a regra de cadastro.

SIM	Numa escala de 1 - 4, qual a chance dos dados terem a regra de captura ou armazenamento alterada?	1	2	3	4	1- A captura do dado e o processamento acontece no mesmo órgão. A regra de captura dificilmente será alterada. 2- A captura do dado e o processamento acontece no mesmo órgão por pessoas/áreas distintas, mas há treinamento e regra para o cadastro das informações. A regra de captura não será alterada. 3- A captura do dado e o processamento acontece no mesmo órgão mas por pessoas/áreas distintas, sem haver regra de preenchimento e treinamento. A regra de captura possivelmente será alterada. 4- Captura de dados descentralizada, os dados finais são processados por outro órgão - chance grande de ter a regra alterada.	<input type="checkbox"/>
NÃO						1- O processo de captura, armazenamento e gestão da informação é bem definido desde o treinamento dos agentes, sistemas utilizados e regras para o cadastro das informações.	<input type="checkbox"/>

FASE 2 - Bloco 2

DESCREVER REGRAS - Esse bloco somente deverá ser preenchido se o modelo utilizar mais de uma base de dados!

É possível relacionar diretamente (entre identificadores) os registros das duas bases de dados?

SIM NÃO

Descreva de forma clara quais são as regras de relacionamento entre os modelos.

FASE 2 - Bloco 3

DESCREVER REGRAS DA AMOSTRA - Esse bloco somente deverá ser preenchido se houver inferência amostral!

1- A amostra casual é do tipo:

1	Equiprobabilística.	Nessa amostra todos os elementos da população original possuem as mesmas chances de estarem na amostra.	<input type="checkbox"/>
2	Randomica.	No critério randomizado não há nenhum tipo de regra. Os elementos são escolhidos aleatoriamente.	<input type="checkbox"/>
3	Sistemática.	A amostra definida por critério sistemático requer regra pré definida que que não esteja diretamente relacionada ao valor do elemento da população.	<input type="checkbox"/>
4	Extratificada.	Nessa opção de amostra extratificada a inferência é feita de forma segmentada utilizando valores definido dos elementos da base.	<input type="checkbox"/>

Descreva de forma clara e objetiva as regras utilizadas para inferência amostral.

FASE 2 - Bloco 4

DESCREVER REGRAS DA BASE DE DADOS - não descrever aqui as regras de inferência amostral, se existir.

Descreva de forma clara quais são as regras da base de dados submetida ao modelo. **Observação:** Não descrever as regras de amostragem.

FASE 3 - VARIÁVEIS

FASE 3 - Bloco único

DESCREVER CARACTERÍSTICAS DAS VARIÁVEIS

Notação	Descrição	Qualitativa		Quantitativa		
		Nominal	Ordinal	Contínua	Discreta	
						<p>Nominal - são as variáveis que assumem alguns valores pré-definidos e normalmente não são numéricos.</p> <p>Ordinal - variável com valores relativos para denotar ordem.</p> <p>Contínuas - as variáveis contínuas são aquelas que podem assumir qualquer valor dentro de um intervalo. Normalmente são variáveis que carregam taxas, valores médios ou aquelas associadas a alguma unidade de medida.</p> <p>Discretas - são as variáveis que assumem somente valores inteiros.</p>

Anexo 3 - MÓDULO RELATÓRIO DE RISCOS

Avaliador: _____ Data: ____ / ____ / ____

Os riscos considerados dizem respeito aos resultados obtidos pelos modelos analisados. No parecer, pode haver indicação de riscos relacionados: (1) a propriedade intelectual de pesquisa, (2) consentimento informado livre e esclarecido, (3) termo de responsabilidade, (4) assentimento e (5) parecer do comitê de ética.

FASE 1 - OBJETIVOS	
FASE 1 - Bloco 1	
ANALISAR O OBJETIVO DO MODELO	OBSERVAÇÕES
1- Com relação ao problema a ser resolvido pelo modelo preditivo:	
1	<p>RISCOS: ALTO RISCO - risco da não existência de dados seguros para subsidiar o modelo. A novidade de uma doença normalmente vem acompanhada da falta de informações históricas.</p> <p>PARECER: Incluir, de forma clara, de onde a base de dados foi extraída para o desenvolvimento do trabalho e acrescentar o alto risco de discrepância que pode existir entre os resultados obtidos pelo modelo e os dados apresentados no cenário estudado. Justificar o risco assumido utilizado a novidade do problema e a predominância do interesse coletivo em detrimento do interesse individual.</p>
2	<p>RISCOS: ALTO RISCO - risco da não existência de dados seguros para subsidiar o modelo e risco de não fazer referências aos trabalhos desenvolvidos previamente. O aspecto específico adotado no modelo não exige o pesquisador de fazer referências aos trabalhos realizados anteriormente que utilizam um modelo base já proposto.</p> <p>PARECER: Incluir, de forma clara, de onde a base de dados foi extraída para o desenvolvimento do trabalho e acrescentar o alto risco de discrepância que pode existir entre os resultados obtidos pelo modelo e os dados apresentados no cenário estudado. Justificar o risco assumido utilizado a novidade do problema e a predominância do interesse coletivo em detrimento do interesse individual. Incluir uma sessão que descreva os modelos bases identificados previamente.</p>
3	<p>RISCOS: BAIXO RISCO - como o trabalho utiliza como referência outro modelo já desenvolvido, a responsabilidade é compartilhada.</p> <p>PARECER: Incluir, de forma clara, o modelo utilizado como base para resolução do problema específico. Para não haver infração de propriedade, é importante incluir, precisamente a referência do modelo utilizado e nome dos autores. Caso o modelo não tenha sido publicado, um termo de aceite deve ser devidamente assinado pelos autores.</p>
4	<p>RISCOS: BAIXO RISCO - não é esperado que o modelo apresente resultados discrepantes à realidade. Há risco de apresentar resultados muito próximos ou idênticos a resultados apresentados em trabalhos anteriores. No entanto, o interesse individual é mais evidente que o interesse coletivo.</p> <p>PARECER: Se for utilizado um modelo já conhecido como base de desenvolvimento, deve se indicar objetivamente qual é esse modelo, por quem foi desenvolvido e quais são as principais diferenças entre aquele parâmetro e o desenvolvido.</p>
2- A resolução do problema:	
1	<p>RISCOS: ALTO RISCO - a resposta do modelo pode impactar um grupo grande de pessoas. Além de não ser possível prever a forma como a sociedade utilizará os resultados obtidos, a discrepância entre os resultados obtidos e o cenário real atrai responsabilidade ao desenvolvedor.</p> <p>PARECER: Um problema que impacta um grupo de pessoas deve ser observado sob duas óticas: (1) pode resolver um problema social ou auxiliar na resolução do problema; (2) pode repercutir negativamente, trazendo retrocesso, erro e prejuízos sociais. Incluir de forma clara os riscos existentes e instruir o leitor de como os dados obtidos pelo modelo devem ser utilizados.</p>
2	<p>RISCOS: ALTO RISCO - a resposta do modelo pode impactar um grupo grande de pessoas. Além de não ser possível prever a forma como a sociedade utilizará os resultados obtidos, a discrepância entre os resultados obtidos e o cenário real atrai responsabilidade ao desenvolvedor.</p> <p>PARECER: O pesquisador deve deixar claro que a pretensão do modelo não é resolver um problema universal. Deve deixar bem definido no objetivo a qual grupo se destina.</p>
3	<p>RISCOS: BAIXO RISCO - como o modelo impactará somente o interesse de um pequeno grupo, não haverá repercussão social.</p> <p>PARECER: Deve ficar claro qual é o grupo que se destina a resposta do problema.</p>
4	<p>RISCOS: BAIXO RISCO - como o modelo impactará somente o interesse do pesquisador, não haverá repercussão na sociedade.</p> <p>PARECER: Importante deixar claro que o modelo é utilizado para resolver um questionamento/objetivo pessoal.</p>
FASE 1 - Bloco 2	
DESCREVER O OBJETIVO DO MODELO	
FASE 1 - Bloco 3	
DESCREVER REGRAS	
1	<p>RISCOS: ALTO RISCO - o modelo que se propõe a prever dados futuros atrai riscos justamente pela expectativa gerada.</p> <p>PARECER: As informações sobre os riscos de utilizar o modelo em um cenário incompatível ou de utilizar os dados projetados devem ser claras e objetivas. Neste caso, quem utilizar o modelo de previsão futura deverá assumir o risco de utilizar tais informações para tomada de decisões.</p>
2	<p>RISCOS: BAIXO RISCO - o modelo que se propõe a simular alguma situação não tem a pretensão de gerar expectativas. A proposta é testar processos do mundo real com valores aproximados.</p> <p>PARECER: É importante que o pesquisador deixe claro que o modelo se destina a realizar simulações, que são testes e que não devem gerar expectativa de realidade.</p>
3	<p>RISCOS: ALTO RISCO - o modelo que se propõe a encontrar "valores ideais" atrai riscos justamente pela expectativa gerada.</p> <p>PARECER: As informações geradas, para ajustar as variáveis a respostas ideais, possuem riscos sob duas óticas: (1) risco de utilizar o modelo com a expectativa de alcançar resultados inviáveis e (2) gerar expectativa de se alcançar, necessariamente a resposta ideal.</p>

O Relatório final deverá descrever as regras básicas do Modelo analisado.

FASE 2 - BASE DE DADOS

FASE 2 - Bloco 1

BASE DE DADOS

OBSERVAÇÕES

1- A base de dados é pública?

SIM

- | | |
|---|--|
| 1 | RISCOS: BAIXO RISCO - quando a base de dados é pública e as regras de coleta, armazenamento e gerenciamento de dados são bem especificadas os riscos são baixos.
PARECER: Normalmente as bases públicas são menos organizadas e normalizadas que as privadas. Importante que o pesquisador inclua a referência dos dados e anexar ao estudo exatamente a base utilizada. |
| 2 | RISCOS: ALTO RISCO - ao utilizar uma base de dados pública onde o poder público tem poderes e facilidades para definir as regras de gerenciamento dos dados ou que as regras de captura e armazenamento não são disponíveis e claras, o autor do modelo assume a responsabilidade por discrepância dos resultados obtidos em relação a realidade fática.
PARECER: Além de endereçar o dataset utilizado, o risco é mitigado ao anexar ao estudo a a base analítica utilizada. Quando os critérios de gerenciamento dos dados públicos não forem claros ou disponíveis importante o contato com o órgão público responsável pelos dados de forma expressa, questionando tais regras. |
| 3 | RISCOS: ALTO RISCO - ao utilizar uma base de dados pública onde o poder público tem poderes e facilidades para definir as regras de gerenciamento dos dados ou que as regras de captura e armazenamento não são disponíveis e claras, o autor do modelo assume a responsabilidade por discrepância dos resultados obtidos em relação a realidade fática.
PARECER: Além de endereçar o dataset utilizado, o risco é mitigado ao anexar ao estudo a a base analítica utilizada. Quando os critérios de gerenciamento dos dados públicos não forem claros ou disponíveis importante o contato com o órgão público responsável pelos dados de forma expressa, questionando tais regras. |
| 4 | RISCOS: ALTO RISCO - ao utilizar uma base de dados pública onde o poder público tem poderes e facilidades para definir as regras de gerenciamento dos dados ou que as regras de captura e armazenamento não são disponíveis e claras, o autor do modelo assume a responsabilidade por discrepância dos resultados obtidos em relação a realidade fática.
PARECER: Além de endereçar o dataset utilizado, o risco é mitigado ao anexar ao estudo a a base analítica utilizada. Quando os critérios de gerenciamento dos dados públicos não forem claros ou disponíveis importante o contato com o órgão público responsável pelos dados de forma expressa, questionando tais regras. |

NÃO

- | | |
|---|--|
| 1 | RISCOS: BAIXO RISCO - se a base de dados privada informar com clareza a origem dos dados e regras de tratamento os riscos são baixos.
PARECER: As bases privadas são mais organizadas e normalizadas. Além do Termo de Consentimento Informado, o pesquisador deverá incluir a referência da base de dados utilizada no estudo. |
| 2 | RISCOS: BAIXO RISCO - se a base de dados privada informar com clareza a origem dos dados e regras de tratamento os riscos são baixos. Os ajustes podem acontecer desde que sejam conhecidos.
PARECER: As bases privadas são mais organizadas e normalizadas. Além do Termo de Consentimento Informado, o pesquisador deverá incluir a referência da base de dados utilizada no estudo. O TCLE deverá coobrigar a Instituição Privada informar o pesquisador em caso de alteração de regras. |
| 3 | RISCOS: ALTO RISCO - ao utilizar uma base de dados privada, a instituição tem poderes alterar as regras de gerenciamento dos dados ou as regras de coleta e tratamento não são disponíveis, o pesquisador, assume a responsabilidade por discrepâncias dos resultados obtidos em relação a realidade fática.
PARECER: As bases privadas são mais organizadas e normalizadas. Além do Termo de Consentimento Informado, o pesquisador deverá incluir a referência da base de dados utilizada no estudo. O TCLE deverá coobrigar a Instituição Privada informar o pesquisador em caso de alteração de regras. |
| 4 | RISCOS: ALTO RISCO - ao utilizar uma base de dados privada, não é possível garantir que os dados não serão alterados. Nesse caso, o pesquisador assume a responsabilidade por discrepância dos resultados obtidos em relação a realidade fática caso os dados sejam alterados.
PARECER: Além de endereçar o dataset utilizado, o risco é mitigado ao anexar ao estudo a base analítica utilizada. O TCLE deverá coobrigar a Instituição privada caso as informações sejam alteradas sem comunicação. |

2- É possível que dados históricos que já estão na base de dados sejam alterados?

SIM

RISCOS: ALTO RISCO - A possibilidade de alteração de dados que já constam na Base de Dados utilizada coloca em risco os resultados obtidos.
PARECER: Além de indicar a fonte da base de dados, ideal anexar a base de dados analítica e o TCLE assinado pelo responsável da Base de Dados. OBS: o TCLE nem sempre é possível. Nesses casos informar o motivo da impossibilidade.

NÃO

RISCOS: BAIXO RISCO - A impossibilidade de alteração dos dados já disponibilizados na Base de Dados utilizada assegura os resultados encontrados.
PARECER: Indicar a fonte da base de dados, anexar a base de dados analítica e o TCLE assinado pelo responsável da Base de Dados. OBS: o TCLE nem sempre é possível. Nesses casos informar o motivo da impossibilidade.

3- É possível conhecer e acompanhar toda a regra de coleta e armazenamento das informações na base de dados?

SIM

RISCOS: BAIXO RISCO - O fato do pesquisador conhecer todo o processo de coleta e tratamento dos dados garante que o pesquisador avalie os riscos e identifique possíveis problemas na base de dados.

NÃO

RISCOS: ALTO RISCO - O fato do pesquisador não conhecer o processo de coleta e tratamento dos dados, impossibilita que testes de validade sejam realizados. O pesquisador deve incluir essa limitação em sessão especial no estudo.

4- É possível que as regras de cadastro dos dados sejam alteradas?

SIM

- | | |
|---|---|
| 1 | RISCOS: BAIXO RISCO - Quando a captura e o gerenciamento dos dados acontecem no mesmo órgão, o órgão se torna corresponsável pela veracidade dos dados.
PARECER: Além de deixar claro o órgão responsável pela coleta e gerenciamento dos dados, é indicado contextualizar o órgão do estudo realizado e solicitar assinatura do Termo de Consentimento. |
|---|---|

2	RISCOS: BAIXO RISCO - O treinamento da equipe e o processo bem definido no processamento dos dados diminui as chances de problemas com os dados de pesquisa. PARECER: Deixar claro qual o processo utilizado para a coleta e tratamento de dados para cada base de dados.
3	RISCOS: ALTO RISCO - Quando a captura e o gerenciamento dos dados acontecem na mesma instituição por áreas diferentes, é difícil, muitas vezes, identificar os autores que participaram de todas as etapas do processo de gerenciamento. PARECER: O trabalho deve deixar claro quais as instituições envolvidas na captura, gerenciamento e disponibilidade de dados. Alertar os usuários dos resultados obtidos e dos riscos de possíveis alterações de dados é fundamental.
4	RISCOS: ALTO RISCO - Quando a captura e o gerenciamento dos dados acontecem em instituições diferentes, é difícil, muitas vezes, identificar os autores que participaram de todas as etapas do processo de gerenciamento. PARECER: O trabalho deve deixar claro quais as instituições envolvidas na captura, gerenciamento e disponibilidade de dados. Alertar os usuários dos resultados obtidos e dos riscos de possíveis alterações de dados é fundamental.
NÃO	RISCOS: BAIXO RISCO - Quando a captura e o gerenciamento dos dados acontecem na mesma instituição, ela se torna corresponsável pelos dados entregues. PARECER: Além de deixar claro a instituição responsável pelo tratamento de dados, é indicado comunicar à instituição e solicitar assinatura do Termo de Consentimento.

FASE 2 - Bloco 2

DESCREVER REGRAS - Esse bloco somente deverá ser preenchido se o modelo utilizar mais de uma base de dados!

É possível relacionar diretamente (entre identificadores) os registros das duas bases de dados?

SIM	RISCOS: BAIXO RISCO - a regra de relacionamento entre bases de dados permite que o pesquisador realize testes de relação entre dados com objetivo de testar a população utilizada no modelo. PARECER: Importante esclarecer que a base de dados final é resultado do relacionamento entre diferentes estruturas. Quando o interessado pelo modelo consegue identificar as regras dessa relação, ele se torna co-responsável pelos resultados obtidos.
NÃO	RISCOS: ALTO RISCO - não é possível testar a relação entre as bases de dados utilizadas. O autor assume o risco de eventuais erros existentes na relação de dados. O fator crítico, nesse caso, é que possivelmente o autor terá conhecimento sobre eventuais erros de relacionamento entre bases de dados. PARECER: Importante esclarecer que a base de dados final é resultado do relacionamento entre diferentes estruturas e que a regra de relacionamento é desconhecida. Não se recomenda utilizar modelos preditivos para esses casos.

Descreva de forma clara quais são as regras de relacionamento entre os modelos.

FASE 2 - Bloco 3

DESCREVER REGRAS DA AMOSTRA - Esse bloco somente deverá ser preenchido se houver inferência amostral!

1- A amostra casual é do tipo:

1	ALTO RISCO: o critério randomizado não possui nenhum tipo de regra, a probabilidade de escolha dos elementos, embora seja igual para todos os elementos possíveis, pode deixar de trazer para a população elementos específicos.
2	ALTO RISCO: o critério randomizado não possui nenhum tipo de regra, a aleatoriedade dos elementos escolhidos poderá tornar a população estudada tendenciosa.
3	BAIXO RISCO: A amostra é definida por critérios bem definidos não relacionados aos valores do elemento da população. A regra deve estar clara no estudo.
4	BAIXO RISCO: Nessa opção de amostra estratificada a inferência é realizada de forma segmentada utilizando valores definido dos elementos da base. Embora o risco exista ele é bem definido e deve estar claro no estudo.

Descreva de forma clara e objetiva as regras utilizadas para inferência amostral.

FASE 2 - Bloco 4

DESCREVER REGRAS DA BASE DE DADOS - não descrever aqui as regras de inferência amostral, se existir.

Descrição clara quais são as regras da base de dados submetida ao modelo. **Observação:** não descrever as regras de amostragem.

FASE 3 - VARIÁVEIS

FASE 3 - Bloco único

DESCREVER REGRAS DA AMOSTRA - Esse bloco somente deverá ser preenchido se houver inferência amostral!

QUALITATIVA	NOMINAL	ALTO RISCO - as variáveis nominais assumem valores pré-definidos em um conjunto, há chances de interpretação da variável pelo pesquisador ou por quem utilizará os resultados obtidos.
	ORDINAL	BAIXO RISCO - as variáveis ordinais assumem valores pré-definidos, ou seja, as chances de interpretação da variável pelo pesquisador ou por quem utilizará os resultados obtidos são baixos.
QUANTITATIVA	CONTÍNUA	ALTO RISCO - as variáveis contínuas podem assumir qualquer valor dentro de um intervalo. São variáveis que carregam taxas, valores médios ou aquelas associadas a alguma unidade de medida. Há chances de erro interpretativo.
	DISCRETA	BAIXO RISCO - variáveis que assumem somente valores inteiros sem viés interpretativo.