

Capítulo

3

Ferramentas OHDSI e a medicina de precisão: Estudos PLE e PLP

Maria Tereza Fernandes Abrahão (USP), Pablo Jorge Madril

Abstract

The Observational Health Data Sciences and Informatics, OHDSI initiative, redefined the area of observational research in health data, making it possible to carry out a systematic analysis on large data masses (big data), coming from various sources, through the definition of a common data model (Common Data Model - CDM OMOP), mechanisms to deal with different vocabularies and the availability of a set of free software tools for analysis. Precision medicine is a dynamic field that integrates the latest research advances in individual treatment. It seeks to provide the most appropriate diagnostic and treatment tools for patients with any type of problem, from heart disease to diabetes and cancer. The term precision medicine represents the work of building a biomedical research network, in order to basically develop a new taxonomic structure for diseases. The OHDSI initiative is part of this network and its tools are those that generate what we call clinical evidence. We will delve into two of these tools: the estimation of effects at the population level (PLE - Estimation at the Population Level) and the forecast at the patient level (PLP - Prediction at the Patient Level). EPP refers to the estimation of the average causal effects of exposures (for example, medical interventions, such as procedures or medications) on specific outcomes of interest. We will approach two types of estimates: Direct effect, which compares the result of the intervention with the non-intervention; Comparative effect, which compares the results of two different interventions. Predictive modeling on a massive patient-specific scale, PLP, has become a reality due to OHDSI, where the common data model (CDM) allows for uniform and transparent analysis on an unprecedented scale. We will cover the use of the PatientLevelPrediction package in R,

which implements the best practices established for the development and validation of models. The chapter covers the PLE and PLP tools provided by OHDSI.

Resumo

A iniciativa OHDSI (Observational Health Data Sciences and Informatics), redefiniu a área de pesquisa observacional em dados de saúde trazendo a possibilidade de realizar análises sistemáticas em grandes massas de dados (big data), providas de diversas fontes, através da definição de um modelo comum de dados (Common Data Model - CDM OMOP), de mecanismos para tratamento de diferentes vocabulários, e da disponibilidade de um conjunto de ferramentas de software livre para análise dos mesmos. A medicina de precisão é um campo dinâmico que integra os últimos avanços da pesquisa no tratamento individual. Ela procura fornecer as ferramentas de diagnóstico e tratamento mais adequados para pacientes com qualquer tipo de problema, desde doenças cardíacas a diabetes e câncer. O termo medicina de precisão representa o trabalho de construção de uma rede de pesquisa biomédica, a fim de desenvolver basicamente uma nova estrutura taxonômica para doenças. A iniciativa OHDSI faz parte desta rede e as suas ferramentas são as que geram aquilo que chamamos de evidência clínica. Nos aprofundaremos em duas destas ferramentas: a estimativa de efeitos no nível populacional (PLE - Population-Level Estimation) e a predição ao nível de paciente (PLP - Patient-Level Prediction). A PLE se refere à estimativa dos efeitos causais médios das exposições (por exemplo, intervenções médicas, como procedimentos ou medicamentos) em resultados específicos de interesse. Abordaremos dois tipos de estimativas: Efeito direto, que compara o resultado da intervenção com a não intervenção; Efeito comparativo, que compara entre si os resultados de duas intervenções diferentes. A modelagem preditiva em escala maciça e específica do paciente, PLP, tornou-se realidade devido a OHDSI, onde o modelo de dados comum (CDM) permite análises uniformes e transparentes em uma escala sem precedentes. Abordaremos o uso do pacote PatientLevelPrediction em R, que implementa as melhores práticas estabelecidas para desenvolvimento e validação de modelos. O capítulo aborda as ferramentas de PLE e PLP disponibilizados pela OHDSI.

3.1 Introdução

Dados de saúde observacionais oferecem oportunidades para gerar evidências reais sobre o efeito de tratamentos que podem melhorar significativamente a vida dos pacientes. Modelos de predição clínica foram desenvolvidos para apoiar o processo de decisão clínica, na qual o médico deve inferir um caminho de diagnóstico ou tratamento com base no histórico disponível do paciente e nas diretrizes clínicas atuais. São também conhecidos como medicina de precisão e utilizados na prática clínica em um amplo espectro de especialidades.

A OHDSI é uma iniciativa global aberta e uma rede de centros de pesquisa baseada em métodos analíticos padronizados, ferramentas de código livre e um modelo comum de dados (CDM - *Common Data Model*) que reúne mais de 2 bilhões de registros clínicos permitindo conduzir uma pesquisa rápida, confiável e economicamente viável.

A crescente rede de bancos de dados padronizados para o CDM permite a validação externa de modelos de predição e estimativa, em diferentes configurações de assistência médica em escala global.

O capítulo está estruturado como se segue. Primeiro, a seção 3.1 oferece uma breve fundamentação sobre a medicina de precisão, a iniciativa OHDSI, iniciativas para o enfrentamento do COVID-19. Em seguida, a seção 3.2 apresenta as estimativas no nível populacional. A seção 3.3, as predições a nível de paciente. A seção 3.4 apresenta as considerações finais e conclusões e a seção 3.5 as referências consultadas.

Na elaboração deste capítulo, seguiu-se o livro da OHDSI, *The book of OHDSI*, de domínio público, sob a licença *Creative Commons Zero v1.0 Universal*, (16/04/2020). O livro é um documento vivo, mantido pela comunidade por meio de ferramentas de desenvolvimento de código aberto e evolui continuamente. A versão online, disponível gratuitamente em¹, sempre representa a versão mais recente. O livro apresenta os três principais casos de uso de estudos observacionais em dados de saúde: caracterização, estimativa no nível da população e previsão no nível do paciente. O texto do livro foi traduzido e complementado na escrita deste capítulo. As figuras e tabelas que ilustram os estudos, foram copiadas do livro. Os exemplos foram elaborados a partir dos conhecimentos adquiridos pelos autores na participação no evento OHDSI COVID-19 *Study-A-Thon*².

3.1.1 A medicina de precisão e a iniciativa OHDSI

¹ The book OHDSI <http://book.ohdsi.org>

² *Study-A-Thon* <https://www.ohdsi.org/88-hours/> e <https://www.ohdsi.org/covid-19-updates/>

A medicina de precisão "foca no entendimento da variabilidade individual na prevenção, cuidado e tratamento das doenças"(Collins et al. 2016). Ela considera as diferenças genéticas, ambientais e de estilo de vida de cada pessoa para determinar o tratamento mais adequado para cada indivíduo. Como exemplos de medicina de precisão temos: prescrição de óculos, específica para cada indivíduo, tratamento de alergias e transfusões sanguíneas. Para todos estes casos e muitos outros, não existem tratamentos genéricos que possam ser aplicados sem levar em conta individualidades.

Ao sustentar estudos de PLP, a iniciativa OHDSI avança a noção de medicina de precisão identificando sub-populações de alto e baixo risco, viabilizando o correto gerenciamento de decisões para cada grupo.

3.1.2 *Study-A-Thon*: Estudo de caso

No fim do ano de 2019 surge e se dissemina pelo mundo um novo coronavírus, designado como Sars-Cov-2 e a doença associada, como COVID-19. Isto impulsiona um esforço sem precedentes da comunidade científica internacional para poder dar respostas rápidas a pandemia que se alastra.

De 26 a 29 de março de 2020, a OHDSI sediou um estudo virtual - *Study-A-Thon* - para auxiliar a tomada de decisões em saúde como resposta à atual pandemia do novo coronavírus.

O evento foi estruturado com dois objetivos principais: (1) gerar evidências imediatas do mundo real sobre perguntas priorizadas, compartilhadas por governos nacionais, agências de saúde pública, instituições relacionadas à saúde e membros da comunidade científica; e (2) projetar estudos específicos do COVID-19 validados e já prontos para execução, assim que os dados estejam disponíveis. Buscou-se:

- Trazer o que é conhecido hoje para descobrir as lacunas de conhecimento e preenchê-las;
- Perceber como a ciência da observação, de dados e a informática mudaram;
- Impor rigor científico na definição de fenótipos que melhorem a eficiência e qualidade da pesquisa;
- Definir um novo padrão de como a ciência aberta e as colaborações podem acontecer e ser um marco de como a ciência deve funcionar no futuro.

Em 88 horas de trabalho, o estudo reuniu 351 participantes de 30 países, envolveu 15 frentes de trabalho simultâneas, foram revisadas mais de 10 mil publicações científicas, elaborados 9 protocolos e lançados 13 pacotes de estudos. Projetou 355 definições de coortes e montou uma rede de dados distribuída para executar os estudos.

O planejamento das atividades acompanhou o seguinte roteiro:

- Revisão da literatura e desenvolvimento do protocolo de estudo;
- Desenvolvimento e avaliação dos fenótipos e coortes;

- Desenvolvimento dos pacotes de estudos disponíveis no GitHub OHDSI³;
- Execução dos pacotes na rede de dados OHDSI;
- Revisão e publicação dos resultados.

Este evento, do qual os autores participaram, gerou publicações e principalmente um exemplo de como conduzir estudos retrospectivos em bases populacionais.

Os exemplos aqui utilizados fazem referência aos citados estudos e as respectivas publicações por serem considerados pelos autores como modelos didáticos e atuais das novas tendências em estudos populacionais.

Dada a abundância de dados, excelência dos estudos e disponibilidade dos resultados, optamos neste capítulo por utilizar diretamente este material para exemplificar cada assunto através das referências adequadas em cada tema, guiando o leitor a navegar diretamente o material no lugar de copiá-lo ou reproduzi-lo aqui.

Uma breve introdução situa o leitor no problema e descreve as perguntas mais urgentes de serem respondidas que foram a motivação da iniciativa, e como aparecem as respostas ao se definirem os estudos.

3.1.2.1 Introdução ao COVID-19

O sequenciamento de genoma completo e a análise filogenética indicaram que o coronavírus que causa o COVID-19 é um beta coronavírus no mesmo subgênero que o vírus da síndrome respiratória aguda grave (SARS). A estrutura da região do gene de ligação ao receptor é muito semelhante à do coronavírus da SARS, e foi demonstrado que o vírus usa o mesmo receptor, a enzima de conversão da angiotensina 2 (ACE2), para entrada de células (Zhou et al. 2020).

Estima-se que o período de incubação do COVID-19 seja dentro de 14 dias após a exposição, com a maioria dos casos ocorrendo aproximadamente quatro a cinco dias após a exposição (Li et al. 2020).

Dentre as manifestações clínicas, a pneumonia parece ser a manifestação grave da infecção mais frequente, caracterizada principalmente por febre, tosse, dispnéia e infiltrados bilaterais na imagem torácica (Guan et al. 2020). No entanto, outras características, incluindo sintomas do trato respiratório superior, mialgias, diarreia e distúrbios do olfato ou do paladar, também são comuns.

3.1.2.2 Estudos priorizados

Com o COVID-19 se espalhando rapidamente e esgotando os sistemas de saúde em todo o mundo, a OHDSI sabia que tinha a oportunidade única de responder a esse desafio. A comunidade projetou e começou a executar estudos em um conjunto internacional de bancos de dados de saúde observacionais (incluindo reivindicações de

³ GitHub OHDSI <https://github.com/ohdsi-studies>

seguro e registros eletrônicos de saúde) para auxiliar na tomada de decisões durante a atual pandemia do COVID-19. O evento conclui com uma apresentação global de vários estudos, o anúncio de projetos e descobertas preliminares. O fim do evento não marcou o fim dos trabalhos. Os grupos continuaram a se reunir para avaliar os resultados, preparar e submeter os artigos para revisão por pares.

Foram priorizados três tipos de estudos:

- **Caracterização:** A maior parte do nosso entendimento da história natural da doença do COVID-19 é limitada a um conjunto discreto de perguntas feitas nos relatos de casos. O objetivo foi elaborar um estudo que resuma descritivamente coortes de pacientes positivos para COVID-19 em toda a comunidade OHDSI à medida que esses dados estivessem disponíveis. Para os pacientes positivos, descreveremos a demografia, condições prévias e medicamentos anteriores na história médica. Quando possível, também realizamos a caracterização estratificada, incluindo: Hipertensão, Diabetes, Cardiopatia, Câncer, Gravidez e outros. Resumimos a incidência de desfechos selecionados (hospitalização com pneumonia, hospitalização que requer cuidados invasivos - UTI, ventilador, intubação, morte) na coorte e subpopulações. Para fornecer um contexto mais amplo para os resultados da caracterização COVID-19, também realizaremos a caracterização em dados históricos de estações virais anteriores para caracterizar influenza, sintomas relacionados a vírus (febre, tosse, mialgia, dispnéia), complicações (pneumonia, dificuldade respiratória aguda) e tratamentos invasivos para problemas respiratórios (ventilação, ECMO, traqueostomia).
- **Estimativas a nível populacional (PLE):** A hidroxicloroquina⁴ é um DMARDs⁵ (*Disease-modifying antirheumatic drugs*) sintético indicado para tratamento da artrite reumatóide que está sendo investigada como potencial profilaxia e tratamento para COVID-19, por suas propriedades antivirais. Propomos realizar análises de séries de casos autocontroladas e comparar a hidroxicloroquina com outras drogas convencionais, para entender melhor seu perfil de segurança e determinar se há riscos diferenciais na incidência de complicações. Também exploraremos a segurança da exposição concomitante à hidroxicloroquina e azitromicina. As análises iniciais usarão a influenza como modelo viral, enquanto as análises subsequentes usarão o COVID-19 quando houver dados disponíveis.

⁴ Hidroxicloroquina <https://pt.wikipedia.org/wiki/Hidroxicloroquina>

⁵ DMARD https://en.wikipedia.org/wiki/Disease-modifying_antirheumatic_drug

- Predição a nível de paciente (PLP): Todos os países do mundo têm como objetivo 'achatar a curva', empregando uma série de medidas de saúde pública destinadas a atrasar a transmissão comunitária do COVID-19, de modo que o número de pacientes infectados espalhe-se ao longo do tempo para que a capacidade do sistema de saúde possa se acomodar às necessidades dos doentes. Em situações em que a demanda por serviços de saúde excede a oferta, as ferramentas de priorização baseadas na gravidade da doença tornam-se valiosas e a educação do público sobre o motivo pelo qual eles não devem procurar cuidados médicos desnecessários, pode fazer a diferença. Estamos propondo três estudos de predição que podem embasar esta discussão: 1) identificar quais pacientes provavelmente precisarão de hospitalização; 2) ajudar a garantir ao público que se forem enviados para casa, a probabilidade de algo ruim acontecer é baixa, de modo que eles sigam o conselho e as recomendações do médico para o autocuidado; 3) usar as informações do histórico médico antes da admissão para ajudar na triagem daqueles que chegam ao hospital e determinar quais casos provavelmente serão mais graves.

Dos descritos anteriormente, um estudo é a primeira caracterização em larga escala de pacientes com COVID-19 hospitalizados nos Estados Unidos e na Ásia, envolvendo seis bancos de dados com pacientes com COVID-19 localizados nos EUA e na Coreia do Sul (Burn et al. 2020).

O outro estudo, é sobre a segurança da hidroxiquina e foi projetado e executado em um conjunto internacional de bancos de dados de saúde. Este estudo com mais de 950.000 pacientes dos EUA, Inglaterra, Alemanha e Coreia do Sul, enfoca o perfil geral de segurança da hidroxiquina que na época do evento, era considerado como um tratamento potencial para o COVID-19. Uma pré-impressão desta análise está disponível e destaca os riscos envolvidos na terapia combinada de hidroxiquina e azitromicina (Lane et al. 2020).

O terceiro estudo projetou o primeiro modelo de predição validado externamente em pacientes com COVID-19 para apoiar as decisões de triagem. Esse modelo determina se é provável que um paciente precise de hospitalização com base em seus sintomas e foi desenvolvido com dados dos EUA e testado nos dados da Coreia do Sul (Williams et al. 2020).

3.1.3 Coortes utilizadas nos estudos

Tanto os estudos PLE quanto PLP, precisam da definição de populações para comparação e avaliação. Estes grupos de populações são chamados de coortes.

A definição das características dos grupos que formam as coortes é conhecida como fenótipo e ajuda a classificar os indivíduos através de elementos observáveis (tem

febre?). A continuação definimos, fenótipo, coorte, e os tipos de coortes utilizadas pelos estudos.

3.1.4 Fenótipos e Coortes

Fenótipo é o termo criado pelo pesquisador dinamarquês Wilhelm L. Johannsen (1857 – 1912) (Justina et al. 2010), e representa as características (parâmetros) observáveis que definem um indivíduo, sejam elas morfológicas, fisiológicas ou comportamentais.

A fenotipagem é o processo de identificação de pacientes com uma condição ou característica médica por meio de uma consulta de pesquisa a um sistema RES ou repositório de dados clínicos usando um conjunto definido de elementos de dados e expressões lógicas.

No contexto dos estudos, fenótipos representam o conjunto de características observáveis que vão permitir montar coortes. Eles são criados e utilizados nos estudos preditivos, de estimativa e de caracterização.

Para criar fenótipos compostos, primeiro precisamos criar e validar os seus blocos constituintes, por exemplo, pneumonia é um termo usado em 29 fenótipos diferentes. Eles são construídos a partir de pequenos blocos e conjuntos de definições (*concept sets*).

Os fenótipos são direcionados pelo seu uso, por exemplo: Como descobrir casos de uma determinada doença, como a influenza?

Temos duas abordagens:

- Restrito: Procurar pelo diagnóstico ou resultado de teste;
- Abrangente: Muitos dos pacientes não chegam a receber um diagnóstico ou teste. Incluímos então, suspeitos, confirmados e uma lógica com sintomas: (febre E (tosse OU dispneia OU mal-estar OU fadiga OU mialgia)).

Existe um compromisso entre especificidade e sensibilidade sem resposta certa sendo a escolha guiada pelo objetivo definido.

A construção dos fenótipos requer o conhecimento e exploração dos dados subjacentes para detectar problemas e particularidades dos códigos escolhidos, como por exemplo, mudanças nas definições dos termos no decorrer do tempo (CID-9 - CID-10).

A ferramenta Atlas ajuda a explorar os registros disponíveis na rede OHDSI fornecendo a frequência dos conceitos nos diversos conjuntos de dados. Enxergamos assim, o que realmente existe nos dados e quais conceitos podem ser usados para criar conjuntos de conceitos abrangentes.

A geração dos fenótipos para os estudos da COVID-19 seguiu o seguinte processo:

- Pesquisa bibliográfica: Os 36 fenótipos candidatos passaram por uma pesquisa no PubMed, onde foram criadas consultas utilizando os termos como descritores, analisando a literatura mais recente de forma sistemática;
- Definição de Coortes: Foram geradas 355 coortes utilizando estes fenótipos que estão disponíveis em: atlas-covid19.ohdsi.org;
- Validação: Mais de 100 coortes passaram pelo processo de validação e revisão para estudos de predição, estimativa e caracterização.

Definimos “coorte” para significar um conjunto de pacientes que satisfazem um ou mais critérios de inclusão por um período de tempo; "observacional" para significar que não há intervenção ou atribuição de tratamento imposta pelo estudo; "retrospectivo” para significar que o estudo será conduzido usando dados já coletados antes do início do estudo (Abrahão et al. 2019). Os termos coortes e fenótipos são frequentemente usados como sinônimos.

Muitas vezes esta definição de coortes difere da adotada em outras literaturas, onde se assemelha mais com um conjunto de códigos. Enquanto o conjunto de códigos é uma parte importante da definição de uma coorte, ela precisa também da lógica que define como o conjunto de códigos será utilizado para o critério especificado. Uma coorte bem definida inclui como um paciente entra e sai da coorte.

No contexto das ferramentas OHDSI, são criados 3 tipos de coortes: coorte alvo, coorte de comparação e coorte de desfecho. Em estudos de PLE, utilizaremos duas coortes, a coorte alvo e a coorte de desfecho, para poder predizer o fator de risco. A coorte de desfecho inclui tanto os pacientes da coorte alvo quanto os fora dela. As ferramentas de análise fazem a intersecção entre elas, fornecendo os resultados adequados. Assim as coortes ficam independentes dos estudos e podem ser reutilizadas entre os mesmos.

3.1.5 O site <https://data.ohdsi.org>

Os resultados dos conjuntos de cortes definidos e analisados durante o evento do *Study-A-Thon* estão disponibilizados neste site como aplicações Shiny que permitem navegar entre os resultados.

Cada resultado, visualizado como uma página interativa, possui uma ajuda marcada com um (i) que explica a entrada ou gráfico, um conjunto de opções para

escolher a base de dados, a coorte alvo e de comparação. Isto facilita a navegação entre a vasta quantidade de resultados.

Em particular, os resultados da caracterização das coortes corresponde as seguintes entradas na página:

- Aferições Basais (*Baseline*): Define as características básicas dos pacientes hospitalizados comparando duas situações, COVID-19 e Influenza;
 - **Pergunta:** Populações com COVID-19 são comparáveis com populações com Influenza?
 - [Covid19CharacterizationHospitalization](#)⁶
 - Caracterização de adultos hospitalizados com: COVID-19, Influenza. Comparados através de um conjunto de covariáveis agrupadas em faixa etária, gênero, medicamentos e história clínica: geral, cardiovascular e neoplasmas.
 - **Resposta:** Pacientes com COVID-19 costumam ser preferentemente homens, mais jovens e mais saudáveis que os da Influenza.
 - **Publicação:**
<https://www.medrxiv.org/content/10.1101/2020.04.22.20074336v1>
- Validações de coortes: Para diferentes coortes, se apresentam os seguintes indicadores: Contagens (entradas nas coortes, quantidade de indivíduos, etc), taxa de incidência, distribuição temporal, estatística dos critérios de inclusão, esmiuçamento do evento indexador, sobreposição de coortes, etc.
 - Caracterizações
 - [Covid19CohortEvaluationCharacterizations](#)
 - Exposições a medicamentos: DMARDs e outros
 - [Covid19CohortEvaluationDmardsExposures](#)
 - [Covid19CohortEvaluationDmardsExposuresSubset](#)
 - [Covid19CohortEvaluationExposures](#)
 - Desfechos: Eficiência (internações em UTI, intubação) e Segurança (infarto, falha renal, etc)
 - [Covid19CohortEvaluationEfficacyOutcomes](#)
 - [Covid19CohortEvaluationSafetyOutcomes](#)

⁶ Covid19 caracterização <https://data.ohdsi.org/Covid19CharacterizationHospitalization/>

3.1.6 Ferramenta ATLAS

ATLAS⁷ é uma das ferramenta publicamente disponível para pesquisadores conduzir análises científicas em dados observacionais padronizados convertidos para o OMOP CDM V5⁸. Permite criar coortes definindo grupos de pessoas com base em uma exposição a um medicamento ou diagnóstico de uma condição específica usando dados de registros de assistência médica. Os perfis dos pacientes podem ser visualizados dentro de uma coorte específica e análises de estimativa de nível populacional permitem a comparação de duas coortes diferentes. Apresenta as seguintes funcionalidades:

- Fontes de dados: opção de seleção da base de dados a ser analisada. Disponibiliza os gráficos de análise da ferramenta Achilles;
- Vocabulário: seleção e importação dos vocabulários necessários para análise dos dados;
- Conjuntos de conceitos: permite definir um novo conceito e listar/exportar;
- Definição de coortes: preparação das coortes para estudo. Define o grupo de pessoas que satisfazem um ou mais critérios de inclusão por um período de tempo. Como consequência desta definição:
 - uma pessoa pode pertencer a múltiplas coortes,
 - pode pertencer a mesma coorte em diferentes períodos de tempo,
 - uma pessoa não pode pertencer mais de uma vez à mesma coorte durante o mesmo intervalo de tempo,
 - uma coorte pode ter zero ou mais membros;
- Caracterização de coortes: é definida como o processo de geração de estatísticas descritivas da coorte a partir de dados de covariáveis a nível de pessoa. As estatísticas resumidas dessas covariáveis podem ser contagem, média, sd, var, min, max, mediana, intervalo e quantis. Além disso, as covariáveis durante um período podem ser estratificadas em unidades de tempo para análises de séries temporais, como intervalos fixos de tempo relativos a data de início da coorte (por exemplo, a cada 7 dias, a cada 30 dias, etc.) ou em intervalos absolutos do calendário, como semana, mês, trimestre, ano;
- Caminho da coorte: é definido como o processo de gerar uma sequência agregada de transições entre as coortes de eventos e as pessoas nas coortes alvo;
 - Coortes Alvo: cada uma das coortes-alvo serão analisadas em relação às coortes de eventos;
 - Coortes de Eventos: cada coorte de eventos define o passo em um caminho que pode ocorrer para uma pessoa na coorte de tratamento;

⁷ ATLAS <http://www.ohdsi.org/web/atlas/#/home>

⁸ OMOP CDM v5 <http://www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm:single-page>

- Taxas de incidência: as taxas de incidência podem ser geradas incluindo as coortes de meta e resultados. A taxa de incidência é reportada como uma proporção e uma taxa. São fornecidas as contagens brutas de pessoas, casos e tempo em risco;
 - Tempo em risco: define a janela de tempo relativa à data de início ou término da coorte com um deslocamento para considerar a pessoa 'em risco' do desfecho em análise;
 - Critérios de estratificação: fornecer critérios de estratificação opcionais para a análise que dividirá a população em grupos únicos em relação aos critérios definidos;
- Perfis: o Atlas fornece a capacidade de pesquisar e explorar perfis individuais de pacientes em um banco de dados. Essa funcionalidade pode ser acessada clicando no item de menu de perfis, selecionando o banco de dados de interesse e inserindo um número de identificação do paciente. Dentro do perfil apresentado, o menu à esquerda lista os registros individuais que podem ser condições, medicamentos, procedimentos, etc. A tabela no canto inferior direito lista os domínios individuais, IDs de conceitos, nomes de conceitos e dias de início e fim no registro de um determinado paciente. O gráfico pode ser redimensionado para aumentar o zoom em uma determinada janela de tempo, por exemplo, nos primeiros 100 dias. Alterando a janela de tempo de interesse, as tabelas a esquerda e inferior direita mudam para refletir a janela de tempo de interesse;
- Estimativa: o Atlas tem a capacidade de realizar estudos de estimativa usando o design de coorte comparativo. O procedimento de estimativa no Atlas usa a metodologia de escore de propensão⁹. Existem 3 principais modelos de resultados: regressão logística; regressão de poisson; e riscos proporcionais de COX;
- Predição: o Atlas incorporou a capacidade de gerar modelos de predição usando métodos de aprendizado de máquina para medicina de precisão e interceptação de doenças, incluindo:
 - *Regularized regression*
 - *Random forest*
 - *k-nearest neighbors*

Usa um conjunto de covariáveis, incluindo, por exemplo, todas as drogas, diagnósticos, procedimentos, bem como idade, índices de comorbidade, etc. Os modelos de resultados suportados são logísticos, Poisson e sobrevivência (tempo até o evento).

⁹ Um escore de propensão é a probabilidade de uma unidade (por exemplo, pessoa) ser designada para um tratamento particular, dado um conjunto de covariáveis observadas. Os escores de propensão são usados para reduzir o viés de seleção ao equacionar grupos com base nessas covariáveis.

3.1.7 Onde achar: principais referências

As informações a respeito dos componentes da OHDSI podem ser classificadas em:

- Informações gerais: <http://www.ohdsi.org> - É o site principal;
- Código e instalações: <https://github.com/OHDSI/> - Aqui está disponibilizado o código fonte de todas as ferramentas. Em particular destacamos: *Common Data Model* (<https://github.com/OHDSI/CommonDataModel>), com a definição completa do modelo e as implementações para os diversos bancos suportados; Broadsea (<https://github.com/OHDSI/Broadsea>), que disponibiliza uma versão em containers Docker do conjunto de ferramentas, ver também: repositório Docker com os componentes dockerizados (<https://hub.docker.com/u/ohdsi/>); OHDSI-In-a-Box(<https://github.com/OHDSI/OHDSI-in-a-Box>)que disponibiliza uma máquina virtual pronta para testes e demonstrações;
- Tutoriais e vídeos: Procure no Google por YouTube OHDSI (<https://www.google.com/search?q=youtube+ohdsi>), existe muita documentação e tutoriais em vídeo dos eventos anuais do grupo;
- Fórum: Para resolver dúvidas mais frequentes, consulte o fórum (<http://forums.ohdsi.org/>).

3.2 Estimativas no nível da população - PLE

Dados de saúde observacionais, como declarações administrativas e registros eletrônicos de saúde, oferecem oportunidades para gerar evidências reais sobre o efeito de tratamentos que podem melhorar significativamente a vida dos pacientes. Nesta seção, focalizaremos a estimativa do efeito no nível da população, que se refere à estimativa dos efeitos causais médios das exposições (por exemplo, intervenções médicas, como exposições a medicamentos ou procedimentos), sobre resultados específicos de saúde de interesse. A seguir, consideramos duas tarefas diferentes de estimativa:

- Estimativa de efeito direto: estimar o efeito de uma exposição (coorte alvo) no risco de um resultado (coorte desfecho), em comparação com nenhuma exposição;
- Estimativa do efeito comparativo: estimar o efeito de uma exposição (coorte alvo) no risco de um resultado (coorte desfecho), em comparação com outra exposição (coorte de comparação).

Nos dois casos, o efeito causal no nível do paciente contrasta um resultado factual, ou seja, o que aconteceu com o paciente exposto, com um resultado contrafactual, ou seja, o que teria acontecido se a exposição não tivesse ocorrido (direta) ou ocorresse uma exposição diferente (comparativo). Como qualquer paciente revela apenas o resultado factual (o problema fundamental da inferência causal), os vários

modelos de estimativa de efeitos empregam diferentes dispositivos analíticos para esclarecer os resultados contrafactuais.

Os casos de uso para estimativa de efeito no nível populacional incluem seleção de tratamento, vigilância de segurança e eficácia comparativa. Em todos os casos, o objetivo permanece o mesmo: produzir uma estimativa de alta qualidade do efeito causal.

A seguir, descrevemos vários desenhos de estudos de estimativa em nível populacional, todos implementados como pacotes R na Biblioteca de métodos OHDSI¹⁰.

3.2.1 Delineando o método de coorte

Os indivíduos observados para iniciar o tratamento alvo são comparados com aqueles que iniciam o tratamento comparador. Para ajustar as diferenças entre os dois grupos de tratamento, várias estratégias de ajuste podem ser usadas, como estratificação, correspondência ou ponderação pelo escore de propensão ou adicionando características de linha de base ao modelo de resultado. As características incluídas no modelo de propensão ou modelo de resultado são capturadas antes do início do tratamento.

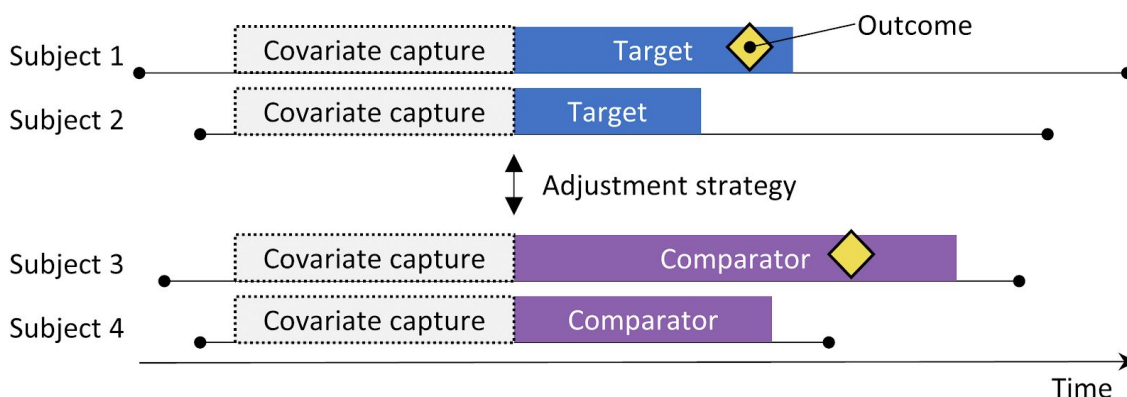


Figura 3.1. O desenho de coorte

O método de coorte tenta emular um ensaio clínico randomizado (Hernán and Robins 2016). Os indivíduos em observação que iniciam um tratamento (o alvo) são comparados aos indivíduos que iniciam outro tratamento (o comparador) e são seguidos por um período específico de tempo após o início do tratamento, por exemplo, o tempo em que permanecem no tratamento. Podemos especificar as perguntas que desejamos responder em um estudo de coorte, fazendo as cinco opções destacadas na Tabela 3.1.

A escolha do modelo especifica, entre outros, o tipo de modelo de desfecho. Por exemplo, poderíamos usar uma regressão logística, que avalia se o desfecho ocorreu ou não, e produz uma razão de chances (odds ratio). Uma regressão logística assume que o

¹⁰ Biblioteca de métodos OHDSI <https://ohdsi.github.io/MethodsLibrary/>

tempo em risco tem a mesma duração para o alvo e o comparador ou é irrelevante. Como alternativa, poderíamos escolher uma regressão de *Poisson* que estima a taxa de incidência, assumindo uma taxa de incidência constante. Frequentemente, é utilizada uma regressão de Cox, que considera o tempo até o primeiro resultado para estimar a taxa de risco, assumindo riscos proporcionais entre o alvo e o comparador.

Tabela 3.1 Principais opções em um projeto de coorte comparativo

Opções	Descrição
Coorte alvo	Uma coorte representando o tratamento alvo
Coorte comparadora	Uma coorte representando o tratamento comparador
Coorte de resultados	Uma coorte representando o resultado do interesse
Tempo em risco	Em que momento (geralmente em relação às datas de início e término da coorte-alvo e comparador) consideramos o risco do resultado?
Modelo	O modelo usado para estimar o efeito enquanto ajusta as diferenças entre o alvo e o comparador

O método de coorte de novos usuários é inerentemente um método para estimativa de efeito comparativo, comparando um tratamento com outro. É difícil usar esse método para comparar um tratamento contra nenhum tratamento, pois é difícil definir um grupo de pessoas não expostas comparável ao grupo exposto. Ao se usar esse desenho para estimativa direta do efeito, a maneira preferida é selecionar um tratamento comparador para a mesma indicação que a exposição de interesse, onde se acredita que o tratamento comparador não tenha efeito sobre o resultado. Infelizmente, esse comparador pode nem sempre estar disponível.

Uma preocupação importante é que os pacientes que recebem o tratamento alvo podem diferir sistematicamente daqueles que recebem o tratamento comparador. Por exemplo, suponha que a coorte alvo tenha em média 60 anos, enquanto a coorte comparadora tenha em média 40 anos. Comparar o alvo com o comparador em relação a qualquer resultado de saúde relacionado à idade (por exemplo, acidente vascular cerebral - AVC) pode mostrar diferenças substanciais entre as coortes. Um investigador desinformado pode chegar à conclusão de que existe uma associação causal entre o tratamento alvo e o AVC em comparação com o grupo de tratamento comparador. De maneira comum, o investigador pode concluir que existem pacientes-alvo que sofreram AVC que não o teriam feito se tivessem recebido o tratamento comparador. Essa conclusão pode estar incorreta. Talvez esses pacientes-alvo tenham sofrido um AVC simplesmente porque são mais velhos; talvez os pacientes-alvo que sofreram AVC pudessem ter feito isso mesmo se tivessem recebido o tratamento comparador. Nesse

contexto, a idade é um "fator de confusão". Um mecanismo para lidar com fatores de confusão em estudos observacionais é por meio de escores de propensão.

3.2.1.1 Escores de propensão

Em um estudo randomizado, um sorteio (virtual) atribui pacientes a seus respectivos grupos. Assim, a probabilidade de um paciente receber o tratamento alvo em relação ao tratamento comparador não se relaciona de forma alguma às características do paciente, como por exemplo, a idade. Não se tem conhecimento do paciente e sabemos com certeza a probabilidade exata de um paciente receber a exposição alvo. Como consequência, e com o aumento da confiança à medida que o número de pacientes no estudo aumenta, os dois grupos de pacientes não podem diferir sistematicamente em relação a qualquer característica do paciente. Esse equilíbrio garantido é válido para as características que o estudo mediu (como a idade), bem como para as características que o estudo não conseguiu medir, como os fatores genéticos do paciente.

Para um determinado paciente, o escore de propensão (*propensity score* - PS) é a probabilidade de que esse paciente tenha recebido o tratamento-alvo em relação ao comparador (Rosenbaum and Rubin 1983). Em um estudo randomizado equilibrado, com dois braços, o escore de propensão é de 0,5 para cada paciente. Em um estudo observacional ajustado ao escore de propensão, estimamos a probabilidade de um paciente receber o tratamento-alvo com base no que podemos observar nos dados antes e no momento do início do tratamento (independentemente do tratamento que eles realmente receberam). Este é uma aplicação simples de modelagem preditiva; ajustamos um modelo (por exemplo, uma regressão logística) que prevê se um sujeito recebe o tratamento-alvo e usamos esse modelo para gerar probabilidades previstas (escore de propensão) para cada sujeito. Ao contrário de um estudo randomizado padrão, pacientes diferentes terão probabilidades diferentes de receber o tratamento-alvo. O escore de propensão pode ser usado de várias maneiras, incluindo a correspondência de assuntos-alvo com assuntos comparadores com escore semelhante, estratificar a população do estudo com base no escore de propensão ou ponderar indivíduos usando a Probabilidade Inversa de Ponderação de Tratamento (*Inverse Probability of Treatment Weighting* - IPTW) derivada do escore de propensão. Ao fazer a correspondência, podemos selecionar apenas um assunto de comparação para cada assunto de destino, ou podemos permitir mais de um assunto comparador por assunto de destino, uma técnica conhecida como correspondência de proporção variável (*variable-ratio matching*) (Rassen et al. 2012).

3.2.1.2 Seleção de variáveis

No passado, os escores de propensão eram calculados com base nas características selecionadas manualmente e, embora as ferramentas OHDSI possam suportar tais

práticas, preferimos usar uma grande quantidade de características genéricas (ou seja, características que não são selecionadas com base nas exposições e resultados específicos do estudo) (Tian et al. 2018). Essas características incluem dados demográficos, bem como todos os diagnósticos, exposições a medicamentos, medições e procedimentos médicos observados antes e no dia do início do tratamento. Um modelo normalmente envolve 10.000 a 100.000 características únicas, as quais ajustamos usando regressão regularizada em larga escala (Suchard et al. 2013) implementada no pacote R Cyclops¹¹. Em essência, deixamos os dados decidirem quais características são preditivas da atribuição do tratamento e devem ser incluídas no modelo.

Normalmente, incluímos o dia do início do tratamento como covariável, porque muitos dados relevantes, como o diagnóstico que leva ao tratamento, são registrados nesta data. Neste dia, o tratamento alvo e o comparador também são registrados, mas eles não devem ser incluídos no modelo de propensão, porque são exatamente isso que estamos tentando prever. Portanto, devemos excluir explicitamente o tratamento alvo e comparador do conjunto de covariáveis

Alguns argumentaram que uma abordagem orientada a dados para a seleção covariável que não depende da experiência clínica para especificar a estrutura causal "correta" corre o risco de incluir erroneamente as chamadas variáveis instrumentais e colisores, aumentando assim a variação e potencialmente introduzindo viés (Hernán et al. 2002). No entanto, é pouco provável que essas preocupações tenham um grande impacto nos cenários do mundo real (Schneeweiss 2018). Além disso, na medicina, a verdadeira estrutura causal é raramente conhecida e, quando se pede a diferentes pesquisadores que identifiquem as covariáveis 'certas' a serem incluídas em uma questão de pesquisa específica, cada pesquisador invariavelmente cria uma lista diferente, tornando o processo irreproduzível. Mais importante, nossos diagnósticos, como inspeção do modelo de propensão, avaliação do equilíbrio de todas as covariáveis e inclusão de controles negativos, identificariam a maioria dos problemas relacionados a colisores e variáveis instrumentais.

3.2.1.3 Calibre

Como o escore de propensão tem valores num continuum de 0 a 1, a correspondência exata raramente é possível. Em vez disso, o processo de correspondência procura pacientes que correspondem à pontuação de propensão de um (s) paciente (s) alvo (s) dentro de alguma tolerância conhecida como "calibre". Seguindo Austin (Austin 2011), usamos um calibre padrão de 0,2 desvios padrão na escala de *logit*¹².

¹¹ Cyclops <https://ohdsi.github.io/Cyclops/>

¹² Logit <https://en.wikipedia.org/wiki/Logit>

3.2.1.4 Overlap: Preference Scores

O método de propensão precisa da existência de pacientes compatíveis. Para isso, um diagnóstico importante mostra a distribuição dos escores de propensão nos dois grupos. Para facilitar a interpretação, as ferramentas OHDSI traçam uma transformação do escore de propensão denominado “escore de preferência”. (Walker et al. 2013). O escore de preferência se ajusta à “participação de mercado” dos dois tratamentos. Por exemplo, se 10% dos pacientes recebem o tratamento alvo (e 90% recebem o tratamento comparador), os pacientes com uma pontuação de preferência de 0,5 têm uma probabilidade de 10% de receber o tratamento alvo. Matematicamente, a pontuação de preferência é:

$$\ln\left(\frac{F}{1-F}\right) = \ln\left(\frac{S}{1-S}\right) - \ln\left(\frac{P}{1-P}\right)$$

Onde F é o *preference score*, S é o *propensity score*, e P é a proporção de pacientes que recebem o tratamento alvo.

Walker et al. (Walker et al. 2013) discutem o conceito de "equilíbrio empírico". Eles aceitam pares de exposição como emergentes de equilíbrio empírico se pelo menos metade das exposições for para pacientes com uma pontuação de preferência entre 0,3 e 0,7.

3.2.2 Coorte auto controlada

A taxa de desfechos durante a exposição à intervenção é comparada com a taxa de desfechos no período pré-exposição. A Figura 3.2 apresenta este desenho.

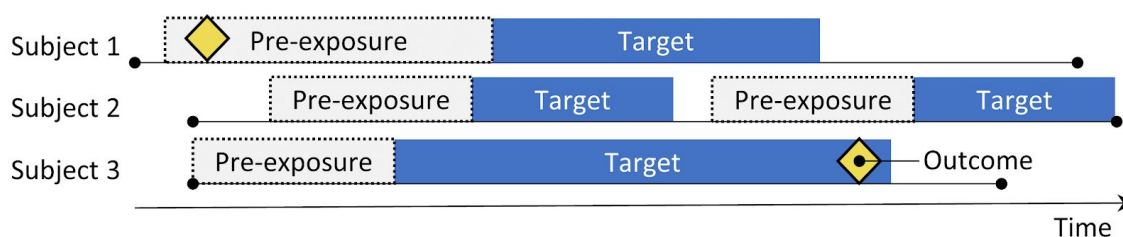


Figura 3.2 Desenho de coorte auto controlado

O desenho de coorte auto controlado (Ryan et al. 2013) compara a taxa de desfecho durante a exposição à taxa de desfecho no período imediatamente anterior à exposição. As quatro opções mostradas na Tabela 3.2 definem uma pergunta de coorte auto controlada.

Tabela 3.2 Principais opções em um projeto de coorte autocontrolado

Opções	Descrição
Coorte alvo	Uma coorte representando o tratamento
Coorte de desfecho	Uma coorte representando o desfecho de interesse
Tempo em risco	Em que momento (geralmente em relação às datas de início e término da coorte alvo) consideramos o risco do resultado?
Tempo de controle	O período usado como o tempo de controle

Como o mesmo sujeito que compõe o grupo exposto também faz parte do grupo de controle, não é necessário fazer ajustes nas diferenças entre as pessoas. No entanto, o método é vulnerável a outras diferenças, como diferenças no risco de linha de base do desfecho entre diferentes períodos de tempo.

3.2.3 Caso-controle

Os indivíduos com o desfecho ('casos') são comparados com indivíduos sem o desfecho ('controles') em termos de status de exposição. Frequentemente, casos e controles são compatíveis com várias características, como idade e gênero. A Figura 3.3 apresenta este desenho.

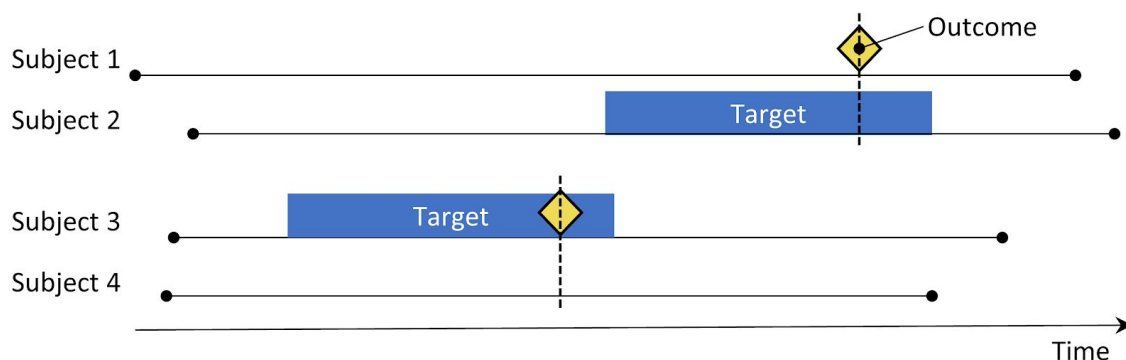


Figura 3.3 Desenho caso-controle

Estudos caso-controle consideram a pergunta "as pessoas com um desfecho específico da doença são expostas com mais frequência a um agente específico do que aquelas sem a doença?" A idéia central é comparar "casos", ou seja, sujeitos que experimentam o resultado de interesse, com "controles", sujeitos que não experimentaram o resultado de interesse. As opções na Tabela 3.3 definem uma questão de caso-controle.

Tabela 3.3 Principais opções em um projeto de caso-controle

Opções	Descrição
Coorte de resultados	Uma coorte representando os casos (o resultado do interesse)
Coorte de controle	Uma coorte representando os controles. Normalmente, a coorte de controle é automaticamente derivada da coorte de resultados usando alguma lógica de seleção
Coorte alvo	Uma coorte representando o tratamento
Coorte aninhada	Opcionalmente, uma coorte que define a subpopulação a partir da qual casos e controles são extraídos
Tempo em risco	A que horas (geralmente em relação à data de indexação) consideramos o status da exposição?

Frequentemente, seleciona-se controles para combinar casos com base em características como idade e gênero, para torná-los mais comparáveis. Outra prática generalizada é aninhar a análise em um subgrupo específico de pessoas, por exemplo, pessoas que foram diagnosticadas com uma das indicações da exposição de interesse.

3.2.4 Caso-cruzado

O tempo em torno do resultado é comparado a uma data de controle definida em um intervalo pré-definido anterior à data do resultado. A Figura 3.4 apresenta este desenho.

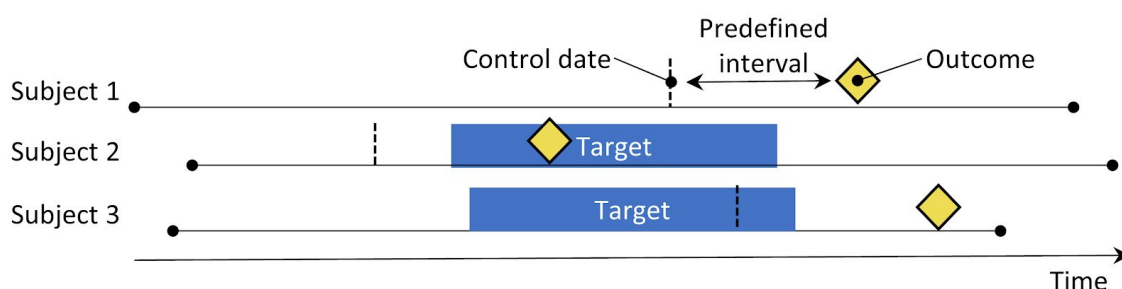


Figura 3.4 Desenho caso-cruzado

O desenho de caso-cruzado avalia se a taxa de exposição é diferente no momento do resultado do que em um número pré-definido de dias antes do resultado. Tenta determinar se há algo de especial no dia em que o resultado ocorreu. A Tabela 3.4 mostra as opções que definem uma pergunta de caso-cruzado.

Tabela 3.4 Principais opções em um desenho caso-cruzado

Opções	Descrição
Coorte de resultados	Uma coorte representando os casos (o resultado do interesse)
Coorte alvo	Uma coorte representando o tratamento
Tempo em risco	A que horas (geralmente em relação à data de indexação) consideramos o status da exposição?
Tempo de controle	O período usado como o tempo de controle

Os casos servem como seus próprios controles. Como projetos autocontrolados, eles devem ser robustos e confusos devido às diferenças entre as pessoas. Uma preocupação é que, como a data do resultado é sempre posterior à data de controle, o método será enviesado positivamente se a frequência geral de exposição aumentar ao longo do tempo (ou enviesado negativamente se houver uma diminuição). Para resolver isso, foi desenvolvido o desenho caso-controle de tempo (Suissa 1995), que adiciona controles, correspondentes por exemplo em idade e gênero, ao desenho caso-cruzado para ajustar as tendências de exposição.

3.2.5 Série de casos auto controlados

A taxa de resultados durante a exposição é comparada à taxa de resultados quando não exposta. A Figura 3.5 apresenta este desenho.

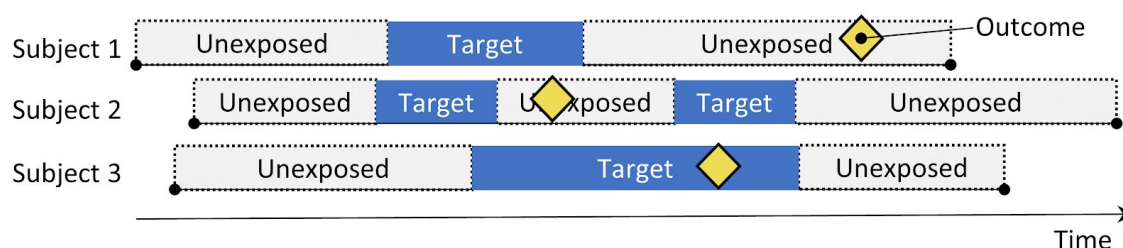


Figura 3.5 O desenho série de casos auto controlados

O desenho Série de Casos Auto controlados (SCA) (Whitaker et al. 2006) compara a taxa de resultados durante a exposição à taxa de resultados durante todo o tempo não exposto, incluindo antes, entre e depois das exposições. É uma regressão de *Poisson* condicionada à pessoa. Assim, ele procura responder à pergunta: “Dado que um paciente tem o resultado, o resultado é mais provável durante o tempo exposto comparado ao tempo não exposto?”. As opções na Tabela 3.5 definem uma pergunta.

Tabela 3.5 Principais opções em um desenho série de casos auto controlado

Opções	Descrição
Coorte alvo	Uma coorte representando o tratamento
Coorte de resultados	Uma coorte representando o resultado do interesse
Tempo em risco	Em que momento (geralmente em relação às datas de início e término da coorte alvo) consideramos o risco do resultado?
Modelo	O modelo para estimar o efeito, incluindo quaisquer ajustes para fatores de confusão que variam no tempo

Como outros desenhos autocontrolados, o SCA é robusto para confundir devido a diferenças entre as pessoas, mas vulnerável a confundir devido a efeitos variáveis no tempo. Vários ajustes são possíveis para tentar contabilizá-los, por exemplo, incluindo idade e tempo. Uma variante especial do SCA inclui não apenas a exposição de interesse, mas todas as outras exposições a medicamentos registrados no banco de dados, potencialmente adicionando milhares de variáveis adicionais ao modelo. A regularização-L1 (*L1-regularization*) usando a validação cruzada para selecionar o hiperparâmetro de regularização é aplicada aos coeficientes de todas as exposições, exceto a exposição de interesse.

Uma suposição importante subjacente ao SCA é que o final do período de observação é independente da data do resultado. Para alguns resultados, especialmente aqueles que podem ser fatais, como derrame, essa suposição pode ser violada. Foi desenvolvida uma extensão para o SCA que corrige qualquer dependência desse tipo (Farrington et al. 2011).

3.2.6 Study-a-thon: Estimativas a nível populacional

Os estudos de estimativa de efeitos em nível populacional examinam os efeitos dos seguintes medicamentos: hidroxicloroquina (HCQ), inibidores da IL6 e JAK, inibidores da protease do HIV e inibidores da protease da hepatite C. Embora as análises de eficácia iniciais sejam feitas usando a influenza como modelo viral, as análises subsequentes usarão os dados dos pacientes COVID-19 quando disponíveis.

Um desses estudos de estimativa de efeitos também se concentrou na segurança do HCQ nos dados dos pacientes com artrite reumatóide, pois o HCQ é um tratamento aprovado para a artrite reumatóide.

3.2.7 Estimativa a nível populacional do uso da Hidroxicloroquina

Risco da hidroxicloroquina, isoladamente e em combinação com azitromicina: um estudo multinacional (Lane et al. 2020).

Introdução: A hidroxicloroquina (HCQ) recebeu recentemente a Autorização de Uso de Emergência pelo FDA e atualmente é prescrita em combinação com azitromicina para pneumonia por COVID-19. Estudamos a segurança da hidroxicloroquina, isoladamente e em combinação com a azitromicina.

Métodos: estudos de coorte de pacientes com artrite reumatóide com idade acima de 18 anos e que iniciaram a hidroxicloroquina foram comparados aos que iniciaram a sulfassalazina e foram acompanhados por mais de 30 dias. As séries de casos autocontroladas (SCCS) foram realizadas para estabelecer ainda mais a segurança em populações mais amplas. A coorte foram incluídos 16 eventos adversos graves (SAEs). Separadamente, foram estudados os SAEs associados à hidroxicloroquina-azitromicina (em comparação com a hidroxicloroquina-amoxicilina). Os dados incluíram 14 fontes de dados de registros médicos eletrônicos da Alemanha, Japão, Holanda, Espanha, Reino Unido e EUA.

Conclusões: foram incluídos 956.374 usuários de hidroxicloroquina e 310.350 de sulfassalazina e 323.122 usuários de hidroxicloroquina-azitromicina e 351.956 de hidroxicloroquina-amoxicilina. Nenhum risco excessivo de SAEs foi identificado quando o uso de hidroxicloroquina e sulfassalazina por 30 dias. O SCCS confirmou esses achados. No entanto, quando a azitromicina foi adicionada à hidroxicloroquina, observamos um risco aumentado de mortalidade cardiovascular em 30 dias, dor no peito / angina e insuficiência cardíaca. A mortalidade cardiovascular também foi aumentada nas análises de tratamento em longo prazo de hidroxicloroquina versus sulfassalazina.

Interpretação: O tratamento com hidroxicloroquina parece seguro durante os 30 dias após o início da artrite reumatóide, mas o uso combinado de hidroxicloroquina e azitromicina pode induzir insuficiência cardíaca e mortalidade cardiovascular. O uso prolongado de hidroxicloroquina pode aumentar o risco de mortalidade na AR. Dado que a hidroxicloroquina, sozinha e em combinação com a azitromicina, está sendo ativamente considerada para o tratamento no COVID-19, pedimos cautela e consideração cuidadosa sobre a relação risco-benefício da eficácia incerta e risco identificado ao gerenciar pacientes com COVID -19.

Até onde sabemos, este é o maior estudo até o momento que analisa a segurança do HCQ. Os resultados preliminares de riscos podem ser vistos em data.ohdsi.org¹³ e os pacotes de estudo podem ser encontrados no OHDSI GitHub¹⁴. A estimativa da eficácia antiviral está aguardando dados dos pacientes COVID-19.

¹³ [data.ohdsi.org https://data.ohdsi.org/Covid19EstimationHydroxychloroquine/](https://data.ohdsi.org/Covid19EstimationHydroxychloroquine/)

¹⁴ OHDSI GitHub <https://github.com/ohdsi-studies/Covid19EstimationHydroxychloroquine>

O estudo foi registrado na Rede Europeia de Centros de Farmacoepidemiologia e Farmacovigilância (*European Network of Centres for Pharmacoepidemiology and Pharmacovigilance* - ENCePP), EU PAS Register number EUPAS34497¹⁵; O protocolo do estudo pode ser avaliado em¹⁶; e o código fonte do estudo em¹⁷.

3.3 Predição ao nível do paciente - PLP

A tomada de decisão clínica é uma tarefa complicada na qual o clínico deve inferir um caminho para diagnóstico ou tratamento com base no histórico médico disponível do paciente e nas diretrizes clínicas atuais. Modelos de predição clínica foram desenvolvidos para apoiar esse processo de tomada de decisão e são utilizados na prática clínica em um amplo espectro de especialidades. Esses modelos preveem um resultado diagnóstico ou prognóstico com base em uma combinação de características do paciente, por exemplo, informações demográficas, histórico da doença e do tratamento.

O número de publicações que descrevem modelos de predição clínica aumentou fortemente nos últimos 10 anos. A maioria dos modelos usados atualmente é estimada usando pequenos conjuntos de dados e considera apenas um pequeno conjunto de características do paciente. Esse baixo tamanho de amostra e, portanto, baixo poder estatístico, força o analista de dados a fazer fortes suposições de modelagem. A seleção do conjunto limitado de características do paciente é fortemente orientada pelo conhecimento especializado disponível. Isso contrasta fortemente com a realidade da medicina moderna, na qual os pacientes geram uma trilha digital rica, que está muito além do poder de qualquer médico para assimilar completamente. Atualmente, os cuidados de saúde estão gerando uma quantidade enorme de informações específicas do paciente armazenadas nos Registros Eletrônicos de Saúde (RES). Isso inclui dados estruturados na forma de diagnóstico, medicação, resultados de testes de laboratório e dados não estruturados contidos nas narrativas clínicas. Não se sabe quanta precisão preditiva pode ser obtida, aproveitando a grande quantidade de dados originários do RES completo de um paciente.

Os avanços no aprendizado de máquina para análises de grandes conjuntos de dados levaram a um interesse crescente em aplicar a predição no nível do paciente a esse tipo de dados. No entanto, muitos esforços publicados na predição no nível do paciente não seguem as diretrizes de desenvolvimento do modelo, não realizam uma validação externa extensa ou fornecem detalhes insuficientes do modelo, o que limita a

¹⁵ ENCePP <http://www.encepp.eu/encepp/viewResource.htm?id=34498>

¹⁶ Protocolo do estudo

<https://github.com/ohdsi-studies/Covid19EstimationHydroxychloroquine/tree/master/documents>

¹⁷ Código fonte <https://github.com/ohdsi-studies/Covid19EstimationHydroxychloroquine>

capacidade de pesquisadores independentes reproduzirem os modelos e realizarem a validação externa. Isso dificulta a avaliação justa do desempenho preditivo dos modelos e reduz a probabilidade de o modelo ser usado adequadamente na prática clínica. Para melhorar os padrões, vários artigos foram escritos detalhando diretrizes para as melhores práticas no desenvolvimento e relatório de modelos de predição. Por exemplo, a rede EQUATOR (*Enhancing the QUALity and Transparency Of health Research*)¹⁸, fornece recomendações claras para relatar o desenvolvimento e validação do modelo de predição e aborda algumas das preocupações relacionadas à transparência.

A modelagem preditiva a nível de paciente em grande escala, tornou-se realidade devido ao OHDSI, onde o modelo de dados comum (*Common Data Model - CDM-OMOP*) permite análises uniformes e transparentes em uma escala sem precedentes. A crescente rede de bancos de dados padronizados para o CDM permite a validação externa de modelos em diferentes configurações de assistência médica em escala global. Acreditamos que isso fornece uma oportunidade imediata para atender grandes comunidades de pacientes que precisam de uma melhor qualidade de atendimento. Tais modelos podem informar cuidados médicos verdadeiramente personalizados, levando a resultados significativamente melhores para os pacientes.

A continuação, descrevemos a estrutura padronizada do OHDSI para predição em nível de paciente (Reps et al. 2018), que com base nas melhores práticas existentes, propõe cinco etapas para: (1) definição transparente do problema; (2) seleção dos conjuntos de dados adequados; (3) construção de variáveis a partir dos dados observacionais; (4) aprendizado do modelo preditivo; e (5) validação do desempenho do modelo. Essa estrutura foi implementada como software de código aberto, utilizando o CDM OMOP, para permitir o compartilhamento e reprodução da avaliação de modelos em vários conjuntos de dados observacionais. Apresentamos na sequência, o pacote R *PatientLevelPrediction*¹⁹ que implementa as melhores práticas estabelecidas para desenvolvimento e validação.

Utilizamos como exemplo o estudo: "*Seek COVER: Development and validation of a personalized risk calculator for COVID-19 outcomes in an international network*" (Williams et al. 2020). Este estudo tem como objetivo desenvolver e validar externamente as pontuações das covariáveis para cálculo do risco estimado para COVID-19 (COVER - *COVID-19 Estimated Risk*) que quantificam o risco de um paciente ter internação hospitalar (COVER-H), precisar de serviços intensivos (COVER-I) ou sofrer uma fatalidade (COVER-F) nos 30 dias seguintes ao diagnóstico de COVID -19. Os resultados do estudo, a calculadora de risco, o link para o protocolo e os pacotes do estudo, estão disponíveis como uma aplicação shiny²⁰.

¹⁸ EQUATOR: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) <https://www.equator-network.org/reporting-guidelines/tripod-statement/>

¹⁹ PatientLevelPrediction <https://ohdsi.github.io/PatientLevelPrediction/>

²⁰ Estudo COVER <https://data.ohdsi.org/Covid19CoverPrediction/>

Na página principal, temos a descrição do estudo implementado, como mostra a Figura 3.6.

Seek COVER: Development and validation of a personalized risk calculator for COVID-19 outcomes in an international network

Development Status: Under Development

This shiny application contains the results of the internal and external validations of the models developed to predict risk of hospitalisation within 30 days of initial presentation to a healthcare professional with symptoms.

During manuscript development and the subsequent review period, these results are considered under embargo and should not be disclosed without explicit permission and consent from the authors.

Abstract Importance COVID-19 is causing high mortality worldwide. Developing models to quantify the risk of poor outcomes in infected patients could help develop strategies to shield the most vulnerable during de-confinement.

Objective To develop and externally validate COVID-19 Estimated Risk (COVER) scores that quantify a patient's risk of hospital admission (COVER-H), requiring intensive services (COVER-I), or fatality (COVER-F) in the 30-days following COVID-19 diagnosis.

Design Multinational, distributed network cohorts.

Setting We analyzed a federated network of electronic medical records and administrative claims data from 14 data sources and 6 countries, mapped to a common data model.

Participants Model development used a patient population consisting of >2 million patients with a general practice (GP), emergency room (ER), or outpatient (OP) visit with diagnosed influenza or flu-like symptoms any time prior to 2020. The model was validated on patients with a GP, ER, or OP visit in 2020 with a confirmed or suspected COVID-19 diagnosis across five databases from South Korea, Spain and the United States.

Outcomes Age, sex, historical conditions, and drug use prior to index date were considered as candidate predictors. Outcomes included i) hospitalization with pneumonia, ii) hospitalization with pneumonia requiring intensive services or death, and iii) death in the 30 days after index date.

Results Overall, 44,507 COVID-19 patients were included for model validation, after initial model development and validation using 6,869,127 patients with influenza or flu-like symptoms. We identified 7 predictors (history of cancer, chronic obstructive pulmonary disease, diabetes, heart disease, hypertension, hyperlipidemia, and kidney disease) which combined with age and sex could discriminate which patients would experience any of our three outcomes. The models achieved high performance in influenza. When transported to COVID-19 cohorts, the AUC ranges were, COVER-H: 0.69-0.81, COVER-I: 0.73-0.91, and COVER-F: 0.72-0.90. Calibration was overall acceptable, with overestimated risk in the most elderly and highest risk strata...

Conclusions and relevance A 9-predictor model performs well for COVID-19 patients for predicting hospitalization, intensive services and fatality. The models could aid in providing reassurance for low risk patients and shield high risk patients from COVID-19 during de-confinement to reduce the virus' impact on morbidity and mortality.

Figura 3.6 Página principal da aplicação shiny do estudo COVER

Clicando na aba Cálculo de Risco (*Risk Calculator*), Figura 3.7, temos a possibilidade de executar um cálculo definindo os valores das variáveis do paciente.

Use this tool to calculate the risk of COVID outcomes:

Age: 62

Sex: Male

History of Cancer

History of COPD

History of Diabetes

History of Heart disease

History of Hyperlipidemia

History of Hypertension

History of Kidney Disease

Calculate Risk

Predicted Risk (%)

Outcome	Predicted Risk (%)
COVER-H	6.9%
COVER-I	2.4%
COVER-F	1.6%

Description

This calculator presents the results of a prediction study that developed 3 prediction models.

Figura 3.7 Calculadora de Risco do estudo COVER

3.3.1 O problema de predição

Entre uma população em risco, nosso objetivo é prever quais pacientes, a partir de um momento definido ($t = 0$), terão algum desfecho durante um período de risco. A predição é feita usando apenas informações sobre os pacientes em uma janela de observação antes desse momento.

A Figura 3.8 ilustra o problema de predição que abordamos. Ela mostra que para definir um problema de predição, precisamos definir $t = 0$ para uma coorte alvo, o desfecho que gostaríamos de prever e o tempo em risco.

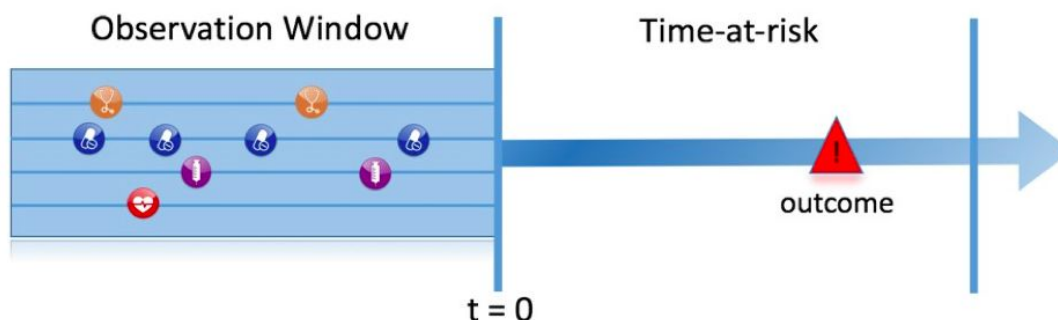


Figura 3.8 O problema da predição

Definimos a pergunta de predição padrão como:

Entre os pacientes [definição da coorte alvo], quem sofrerá um desfecho de [definição da coorte de resultados] durante o próximo [período de tempo em risco]?

Além disso, temos que definir parâmetros para o delineamento do modelo que gostaríamos de desenvolver e determinar os conjuntos de dados observacionais para executar a validação interna e externa. A Tabela 3.6 apresenta as perguntas a serem respondidas a cada parâmetro.

Tabela 3.6 Principais parâmetros de delineamento do projeto de predição

Parâmetros	Descrição
Coorte alvo	Como definimos a coorte de pessoas para quem desejamos prever?
Coorte de desfecho	Como definimos o resultado que queremos prever?
Tempo em risco	Em qual janela de tempo relativa a $t = 0$ queremos fazer a predição?
Modelo	Quais algoritmos queremos usar e quais variáveis preditivas potenciais incluímos?

Resumindo o problema de predição:

- Defina a população-alvo, que são os pacientes aos quais você deseja aplicar o modelo. A população alvo é um conjunto de pacientes que atendem a um ou mais critérios de inclusão por um período de tempo. Por exemplo, uma

população-alvo pode ser pacientes que iniciam o tratamento de depressão observado desde o início do tratamento até a interrupção do tratamento.

- Defina o desfecho para o qual você deseja prever o risco. O desfecho também é uma população definida como um conjunto de pacientes que atendem a um ou mais critérios de inclusão por um período de tempo. Por exemplo, uma população de desfecho podem ser pacientes com AVC observado no momento do primeiro diagnóstico. Se o desfecho ocorrer dentro do período de tempo em risco, o *status do desfecho* será definido como "positivo";
- Defina o período de tempo em risco; este é o intervalo de tempo dentro do qual você deseja prever o desfecho ocorrido. Por exemplo, pacientes com tratamento para depressão podem estar em risco de derrame, desde o dia seguinte ao início do tratamento até 1 ano após o início do tratamento.

Essa estrutura conceitual funciona para todos os tipos de problemas de predição, por exemplo:

- Início e progressão da doença
 - **Estrutura:** Entre os pacientes que foram diagnosticados recentemente com *[uma doença]*, quem passará a ter *[outra doença ou complicação]* dentro de *[horizonte temporal do diagnóstico]*?
 - **Exemplo:** Entre os pacientes com fibrilação atrial recentemente diagnosticados, quem passará a sofrer um AVC isquêmico nos próximos três anos?
- Opções de tratamentos
 - **Estrutura:** Entre os pacientes com *[doença indicada]* que são tratados com *[tratamento 1]* ou *[tratamento 2]*, quais pacientes foram tratados com *[tratamento 1]*?
 - **Exemplo:** Entre os pacientes com fibrilação atrial que tomaram varfarina ou rivaroxaban, quais pacientes recebem varfarina? (por exemplo, para um modelo de propensão)
- Resposta ao tratamento
 - **Estrutura:** Entre os novos usuários de *[um tratamento]*, quem experimentará *[algum efeito]* em *[janela do tempo]*?
 - **Exemplo:** quais pacientes com diabetes que começam com metformina permanecem em uso de metformina por três anos?
- Segurança do tratamento
 - **Estrutura:** Entre os novos usuários de *[um tratamento]*, quem experimentará *[evento adverso]* em *[janela do tempo]*?
 - **Exemplo:** Entre os novos usuários de varfarina, quem terá um sangramento gastrointestinal em um ano?
- Adesão ao tratamento

- **Estrutura:** Entre os novos usuários de [um tratamento], quem alcançará [métrica de adesão] na [janela do tempo]?
- **Exemplo:** quais pacientes com diabetes que começam com metformina atingem $\geq 80\%$ de proporção de dias cobertos em um ano?

3.3.2 Extração dos dados: coortes, covariáveis e valores ausentes

Ao criar um modelo preditivo, usamos um processo conhecido como aprendizado supervisionado - uma forma de aprendizado de máquina - que infere a associação entre as covariáveis e o *status de desfecho* com base em um conjunto rotulado de exemplos.

Um conjunto rotulado de exemplos é construído determinando previamente para cada exemplo qual é o status de desfecho correspondente. Normalmente isto é feito por um grupo de especialistas ou através de revisão bibliográfica. Portanto, precisamos de métodos para extrair as covariáveis do CDM para as pessoas na coorte alvo e precisamos obter seus desfechos rotulados.

As covariáveis (também denominadas "preditores", "características" (*features*) ou "variáveis independentes") descrevem os pacientes. As covariáveis podem incluir idade, gênero, presença de condições específicas, procedimentos, etc. As covariáveis geralmente são construídas usando o pacote *FeatureExtraction*²¹, descrito em mais detalhes no "*The Book of OHDSI*"²². Para predição, só podemos usar dados antes (e na data em que a pessoa entra no grupo alvo (data de indexação).

Também precisamos obter o *status do desfecho* (também conhecido como "*labels*" ou "*classes*") de todos os pacientes durante o período de risco.

Resumindo:

Selecione o conjunto de dados que será usado para desenvolver o modelo:

- Verifique se a população alvo tem tamanho suficiente para o desenvolvimento do modelo;
- Verifique se há um número suficiente de desfechos na população-alvo durante o período de risco.

Selecione entre um conjunto de variáveis preditivas padronizadas (recomendamos selecionar todas as variáveis padronizadas):

- Pode escolher períodos de tempo diferentes para construir variáveis antes da data de início do período de risco;
- Pode escolher entre conceitos demográficos, condições, medicamentos, medições, procedimentos e observações;
- Pode agrupar conceitos com base em uma hierarquia no vocabulário.

²¹ *FeatureExtraction* <https://ohdsi.github.io/FeatureExtraction/>

²² *The book OHDSI*, capítulo 11 <https://ohdsi.github.io/TheBookOfOhdsi/PatientLevelPrediction.html>

No exemplo do COVER, observando o protocolo do estudo²³, temos no item 5 - Métodos, o delineamento do estudo (5.1 - *Study Design*), a descrição das fontes de dados (5.2 - *DataSources*), a definição das populações (5.3 - *Study Populations*) onde definimos: coortes alvo (5.3.1 - *Target Cohort(s)*), coortes de validação (5.3.2 - *Validation cohorts*) e coortes de desfecho (5.3.3 - *Outcome Cohorts*) junto com a definição do período de risco (5.3.4 *Time at Risk*).

Durante o *Study-A-Thon*, foi utilizado o Atlas²⁴ para armazenar as definições das coortes²⁵ e dos estudos²⁶ de predição como referência dos estudos desenvolvidos durante o evento. Como exemplo, temos o estudo do risco de internação²⁷.

3.3.2.1 Exemplo de extração de dados

A Tabela 3.7 mostra um exemplo da tabela COHORT (do modelo CDM). Nela identificamos duas coortes: o COHORT_DEFINITION_ID = 1, é a coorte alvo (por exemplo, “pessoas diagnosticadas com fibrilação atrial”) e o COHORT_DEFINITION_ID = 2, define a coorte de desfecho (por exemplo, "acidente vascular cerebral").

Temos assim pacientes, identificados pelo seu SUBJECT_ID, que pertencem apenas a uma das coortes (exemplo, o paciente com SUBJECT_ID=1) ou as duas (exemplo, o paciente com SUBJECT_ID=2).

Tabela 3.7 Exemplo de tabela COHORT

COHORT_DEFINITION_ID	SUBJECT_ID	COHORT_START_DATE
1	1	2000-06-01
1	2	2001-06-01
2	2	2001-07-01

A Tabela 3.8 fornece um exemplo da tabela CONDITION_OCCURRENCE (CDM). O ID de conceito 320128 refere-se ao diagnóstico de "Hipertensão essencial".

²³ Protocolo do COVER:

https://github.com/ohdsi-studies/Covid19PredictionStudies/blob/master/HospitalizationInSentHomePatients/docs/PLP_protocol_Q1%2BQ2_20200329.docx

²⁴ Atlas <https://atlas.ohdsi.org/#/home>

²⁵ Definição das coortes <https://atlas.ohdsi.org/#/cohortdefinitions>

²⁶ Parâmetros do estudo <https://atlas.ohdsi.org/#/prediction>

²⁷ Estudo do Risco de internação <https://atlas.ohdsi.org/#/prediction/3>

Tabela 3.8 Exemplo da tabela CONDITION_OCCURRENCE

PERSON_ID	CONDITION_CONCEPT_ID	CONDITION_START_DATE
1	320128	2000-10-01
2	320128	2001-05-01

Com base neste exemplo de dados, e assumindo que o período de risco é o ano seguinte à data de indexação (a data de início da coorte alvo), podemos construir as covariáveis e o status do desfecho. Uma covariável indicando “Hipertensão essencial no ano anterior” terá o valor 0 (não presente) para o paciente com ID = 1 (o diagnóstico ocorreu (2000-10-01) após a data de indexação (2000-06-01)) e o valor 1 (presente) para o paciente com ID = 2 (o diagnóstico ocorreu (2001-05-01) antes da data de indexação (2001-06-01)). Da mesma forma, o status do desfecho será 0 para o paciente com ID = 1 (essa pessoa não teve entrada na coorte de desfechos) e 1 para o paciente com ID = 2 (o desfecho ocorreu dentro de um ano após a data de indexação).

3.3.2.2 Negativo vs ausente

Os dados de saúde observacional raramente refletem se um valor é negativo ou ausente. No exemplo anterior, simplesmente observamos que a pessoa com ID 1 não apresentava hipertensão essencial antes da data de indexação. Isso pode ocorrer porque a condição não estava presente (negativa) naquele momento ou porque não estava registrada (ausente). É importante perceber que o algoritmo de aprendizado de máquina não pode distinguir entre o negativo e o ausente, e simplesmente avaliará o valor preditivo nos dados disponíveis.

3.3.3 Ajustando o modelo

Ao ajustar um modelo de predição, estamos tentando descobrir a associação entre as covariáveis e o status do desfecho observado a partir de exemplos rotulados. Suponha que tenhamos apenas duas covariáveis, pressão arterial sistólica (PAS) e diastólica (PAD), para que possamos representar cada paciente como um gráfico no espaço bidimensional, como mostra a Figura 3.9. Por exemplo, para um paciente específico com determinados valores de PAS e PAD, a forma (x ou o) do ponto de dados corresponde ao status do desfecho do mesmo (por exemplo, acidente vascular cerebral presente (x) ou ausente (o)).

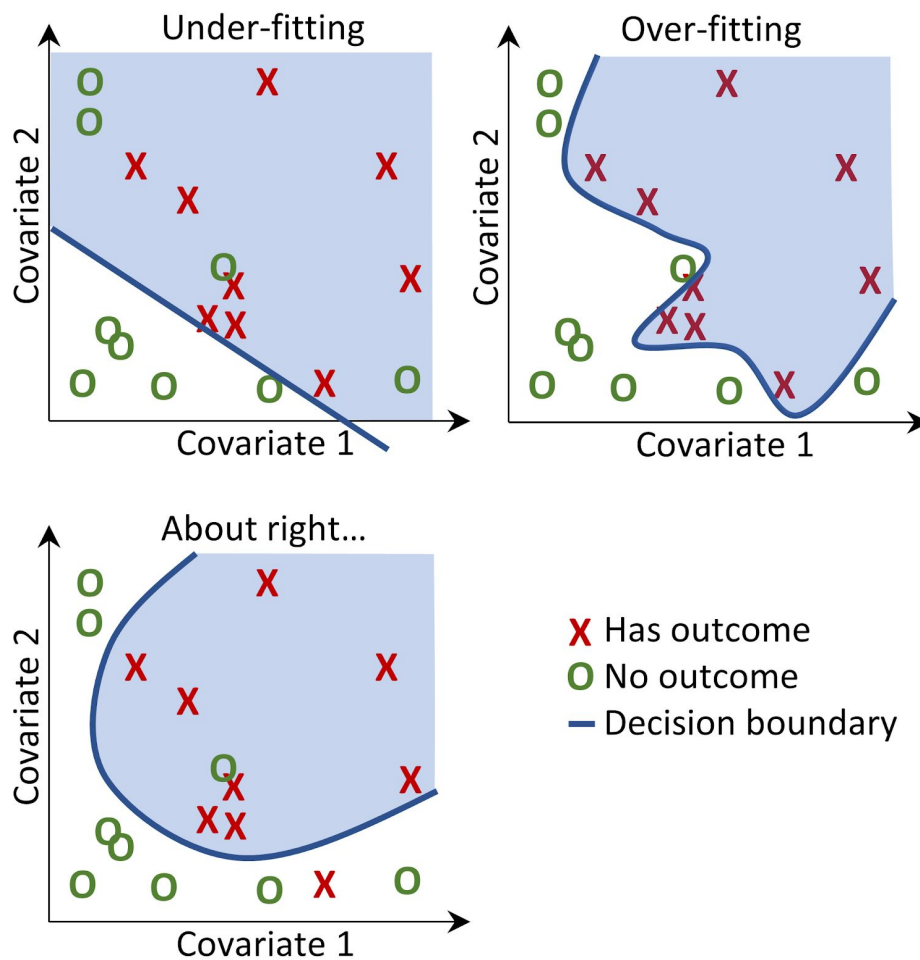


Figura 3.9 Limite de decisão

Na Figura 3.9 podemos ver três limites de decisão diferentes. Os limites são usados para inferir o status do desfecho de qualquer novo ponto de dados. Se um novo ponto de dados cair na área sombreada, o modelo irá prever “desfecho presente”, caso contrário, ele irá prever “desfecho ausente”. Idealmente, um limite de decisão deve particionar perfeitamente as duas classes. No entanto, existe o risco de modelos muito complexos se super ajustarem aos dados. Isso pode afetar negativamente a generalização do modelo para dados não vistos. Por exemplo, se os dados contiverem ruído, com pontos de dados incorretamente rotulados ou posicionados, não desejaríamos ajustar nosso modelo a esse ruído. Portanto, podemos preferir definir um limite de decisão que não discrimine perfeitamente nossos dados de treinamento, mas capture a complexidade "real".

Um modelo de aprendizado supervisionado tentará encontrar um limite de decisão que otimamente separe as duas classes de desfechos. Diferentes técnicas de aprendizado supervisionado levam a diferentes limites de decisão e geralmente existem hiperparâmetros que podem afetar a complexidade do limite de decisão.

Cada algoritmo de aprendizado supervisionado tem uma maneira diferente de aprender o limite de decisão e não é simples descobrir qual algoritmo funcionará melhor em seus dados. Como o teorema *No Free Lunch*²⁸ afirma, nenhum algoritmo sempre superará os outros em todos os problemas de predição. Portanto, recomendamos tentar vários algoritmos de aprendizado supervisionado com várias configurações de hiperparâmetros ao desenvolver modelos de predição no nível do paciente.

Os seguintes algoritmos estão disponíveis no pacote `PatientLevelPrediction`²⁹.

3.3.3.1 Regularized Logistic Regression

A regressão logística LASSO³⁰ (*least absolute shrinkage and selection operator*) pertence à família de modelos lineares generalizados, onde uma combinação linear das variáveis é aprendida e, finalmente, uma função logística mapeia a combinação linear para um valor entre 0 e 1. A regularização do LASSO acrescenta um custo baseado na complexidade do modelo para a função de ganho durante o treinamento do modelo. Esse custo é a soma dos valores absolutos da combinação linear dos coeficientes. O modelo executa automaticamente a seleção de recursos, minimizando esse custo. Utilizamos o pacote `Cyclops`³¹ (*Cyclic coordinate descent for logistic, Poisson and survival analysis*) para realizar regressão logística regularizada em larga escala. A Tabela 3.9 apresenta os hiperparâmetros para a regressão logística.

Tabela 3.9 Hiperparâmetros para Regularized Logistic Regression

<i>Parameter</i>	<i>Description</i>	<i>Typical values</i>
<i>Starting variance</i>	<i>The starting variance of the prior distribution.</i>	0.1

Observe que a variação é otimizada maximizando a probabilidade fora da amostra em uma validação cruzada; portanto, a variação inicial tem pouco impacto no desempenho do modelo resultante. No entanto, escolher uma variação inicial muito distante do valor ideal pode levar a um longo tempo de ajuste.

No exemplo do COVER, podemos ver no Atlas³² os parâmetros definidos para este tipo de modelo para o estudo de risco de hospitalização.

Neste estudo foram analisadas e selecionadas mais de 30.000 covariáveis.

²⁸ *No free lunch theorem* https://en.wikipedia.org/wiki/No_free_lunch_theorem

²⁹ *PatientLevelPrediction* <https://ohdsi.github.io/PatientLevelPrediction/>

³⁰ Lasso [https://en.wikipedia.org/wiki/Lasso_\(statistics\)](https://en.wikipedia.org/wiki/Lasso_(statistics))

³¹ Cyclops <https://ohdsi.github.io/Cyclops/>

³² Parâmetros do LASSO <https://atlas.ohdsi.org/#/prediction/3>

3.3.3.2 Gradient Boosting Machines

O *gradient boosting*³³ é uma técnica de aprendizado de máquina para problemas de regressão e classificação, que produz um modelo de predição na forma de um arranjo de modelos de predição fracos, geralmente árvores de decisão. Ela constrói o modelo em etapas, como outros métodos de *boosting*, e os generaliza, permitindo a otimização de uma função de perda diferenciável arbitrária. O *boosting* funciona adicionando iterativamente árvores de decisão, mas adiciona mais peso aos pontos de dados que são classificados incorretamente por árvores de decisão anteriores na função de custo ao treinar a próxima árvore. Usamos o *Extreme Gradient Boosting*, que é uma implementação eficiente da estrutura de *gradient boosting* implementada no pacote *xgboost* R³⁴ disponível no CRAN. A Tabela 3.10 apresenta os hiperparâmetros para *Gradient Boosting Machines*.

Tabela 3.10 Hiperparâmetros para Gradient Boosting Machines

<i>Parameter</i>	<i>Description</i>	<i>Typical values</i>
<i>earlyStopRound</i>	<i>Stopping after rounds without improvement</i>	25
<i>learningRate</i>	<i>The boosting learn rate</i>	0.005,0.01,0.1
<i>maxDepth</i>	<i>Max levels in a tree</i>	4,6,17
<i>minRows</i>	<i>Min data points in a node</i>	2
<i>ntrees</i>	<i>Number of trees</i>	100,1000

3.3.3.3 Random Forest

*Random forest*³⁵ é uma técnica para combinar múltiplas árvores de decisão em um arranjo que corrige o hábito de as árvores de decisão se ajustarem ao seu conjunto de treinamento. A idéia é reduzir a probabilidade de excesso de ajuste (*overfitting*) usando classificadores fracos e combinando-os em um classificador forte. A *Random forest* realiza isso treinando várias árvores de decisão, mas apenas usando um subconjunto das variáveis em cada árvore e o subconjunto de variáveis diferem entre as árvores. Utilizamos a implementação *sklearn* do pacote Random Forest em Python. A Tabela 3.11 apresenta os hiperparâmetros para *random forest*.

³³ *Gradient Boosting* https://en.wikipedia.org/wiki/Gradient_boosting

³⁴ *xgboost* R <https://cran.r-project.org/web/packages/xgboost/xgboost.pdf>

³⁵ *Random Forest* https://en.wikipedia.org/wiki/Random_forest

Tabela 3.11 Hiperparâmetros para *random forest*

<i>Parameter</i>	<i>Description</i>	<i>Typical values</i>
<i>maxDepth</i>	<i>Max levels in a tree</i>	4,10,17
<i>mtries</i>	<i>Number of features in each tree</i>	-1 = square root of total features,5,20
<i>ntrees</i>	<i>Number of trees</i>	500

3.3.3.4 *K-Nearest Neighbors*

*K-nearest neighbors*³⁶ (KNN) é um algoritmo que usa alguma métrica de distância para encontrar os K-pontos de dados mais próximos de um novo ponto de dados não rotulado. A predição dos novos pontos de dados é a classe mais prevalente dos pontos de dados rotulados K-mais próximos. Há uma limitação de compartilhamento do KNN, pois o modelo requer dados rotulados para realizar a predição de novos dados, e muitas vezes não é possível compartilhar esses dados entre sites de dados. Incluímos o pacote *BigKnn*³⁷ desenvolvido no OHDSI, que é um classificador KNN em larga escala. A Tabela 3.12 apresenta os hiperparâmetros para *K-nearest neighbors*.

Tabela 3.12 Hiperparâmetros para *K-nearest neighbors*

<i>Parameter</i>	<i>Description</i>	<i>Typical values</i>
k	<i>Number of neighbors</i>	1000

3.3.3.5 *Naive Bayes*

O algoritmo *Naive Bayes*³⁸ aplica o teorema de *Bayes* com a suposição simplista de independência condicional entre todos os pares de características, dado o valor da variável de classe. Com base na probabilidade de os dados pertencerem a uma classe e na distribuição anterior da classe, é obtida uma distribuição posterior. *Naive Bayes* não possui hiper-parâmetros.

3.3.3.6 *AdaBoost*

*AdaBoost*³⁹ é uma técnica de realimentação de impulso. O impulso funciona adicionando classificadores iterativamente, mas adiciona mais peso aos pontos de dados

³⁶ *K-nearest neighbors* https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

³⁷ *An R package implementing a large scale k-nearest neighbor classifier using the Lucene search engine* <http://ohdsi.github.io/BigKnn/>

³⁸ *Naive Bayes* https://en.wikipedia.org/wiki/Naive_Bayes_classifier

³⁹ *AdaBoost* <https://en.wikipedia.org/wiki/AdaBoost>

que são classificados incorretamente por classificadores anteriores na função *cost* ao treinar o próximo classificador. Usamos a implementação *sklearn* do pacote *AdaBoostClassifier* em Python. A Tabela 3.13 apresenta os hiperparâmetros para *AdaBoost*.

Tabela 3.13 Hiperparâmetros para AdaBoost

<i>Parameter</i>	<i>Description</i>	<i>Typical values</i>
<i>nEstimators</i>	<i>The maximum number of estimators at which boosting is terminated</i>	4
<i>learningRate</i>	<i>Learning rate shrinks the contribution of each classifier by learning_rate. There is a trade-off between learningRate and nEstimators</i>	1

3.3.3.7 Decision Tree

*Decision Tree*⁴⁰ é um classificador que particiona o espaço variável usando testes individuais selecionados usando uma abordagem gananciosa. O objetivo é encontrar partições que tenham o maior ganho de informações para separar as classes. A árvore de decisão pode cair facilmente no excesso de ajuste (*overfitting*) ao permitir um grande número de partições (profundidade da árvore) e geralmente precisa de alguma regularização (por exemplo, poda ou especificação de hiperparâmetros que limitem a complexidade do modelo). Usamos a implementação *sklearn* do pacote *DecisionTreeClassifier* em Python.

3.3.3.8 Multilayer Perceptron

*Multilayer perceptrons*⁴¹ são redes neurais que contêm várias camadas de nós que pesam suas entradas usando uma função não linear. A primeira camada é a camada de entrada, a última camada é a camada de saída e, no meio, estão as camadas ocultas. As redes neurais geralmente são treinadas usando retropropagação, o que significa que a entrada de treinamento é propagada para frente através da rede para produzir uma saída, o erro entre a saída e o status do resultado é calculado e esse erro é propagado para trás através da rede para atualizar os pesos da função linear. A Tabela 3.14 apresenta os hiperparâmetros para *multilayer perceptrons*.

⁴⁰ *Decision Tree* https://en.wikipedia.org/wiki/Decision_tree

⁴¹ *Multilayer perceptrons* https://en.wikipedia.org/wiki/Multilayer_perceptron

Tabela 3.14 Hiperparâmetros para *multilayer perceptrons*

<i>Parameter</i>	<i>Description</i>	<i>Typical values</i>
<i>alpha</i>	<i>The l2 regularization</i>	0.00001
<i>size</i>	<i>The number of hidden nodes</i>	4

3.3.3.9 Deep Learning

*Deep Learning*⁴² é um ramo de aprendizado de máquina (*Machine Learning*) baseado em um conjunto de algoritmos que tentam modelar abstrações de alto nível de dados usando um grafo profundo com várias camadas de processamento, compostas de várias transformações lineares e não lineares. É semelhante ao *Multilayer perceptrons*, mas possui várias camadas ocultas que visam aprender representações latentes úteis para a predição. Em *BuildingPredictiveModels*⁴³ no pacote *PatientLevelPrediction*, descrevemos esses modelos e hiperparâmetros com mais detalhes.

3.3.3.10 Outros algoritmos

Outros algoritmos podem ser adicionados à estrutura do pacote *PatientLevelPrediction*. Os detalhes podem ser encontrados em “*Adding Custom Patient-Level Prediction Algorithms*⁴⁴” no pacote *PatientLevelPrediction*⁴⁵.

3.3.4 Validação e métricas dos modelos preditivos

3.3.4.1 Tipos de avaliação

Podemos avaliar um modelo de predição medindo a concordância entre a predição do modelo e o status do desfecho observado, o que significa que precisamos de dados nos quais o status do desfecho é conhecido.

Para avaliação, devemos usar um conjunto de dados diferente do usado para desenvolver o modelo; caso contrário, corremos o risco de favorecer modelos com excesso de ajuste (*overfitting*) e que podem não ter um bom desempenho para novos pacientes.

⁴² *Deep Learning* https://en.wikipedia.org/wiki/Deep_learning

⁴³ *BuildingPredictiveModels*

<https://ohdsi.github.io/PatientLevelPrediction/articles/BuildingDeepLearningModels.html>

⁴⁴ <https://ohdsi.github.io/PatientLevelPrediction/articles/AddingCustomAlgorithms.html>

⁴⁵ *PatientLevelPrediction* <https://ohdsi.github.io/PatientLevelPrediction/>

Distinguimos entre:

- **Validação interna:** usando diferentes conjuntos de dados extraídos do mesmo banco de dados para desenvolver e avaliar o modelo;
- **Validação externa:** desenvolvendo o modelo em um banco de dados e avaliando em outro banco de dados.

Existem duas maneiras de executar a validação interna:

- Uma abordagem de **conjunto de validação** divide os dados rotulados em dois conjuntos independentes: um conjunto de treinamento e um conjunto de testes (o conjunto de análises). O conjunto de treinamento é usado para treinar o modelo e o conjunto de testes é usado para avaliá-lo. Podemos simplesmente dividir nossos pacientes aleatoriamente em um conjunto de treinamento e teste, ou podemos optar por:
 - Divida os dados com base no tempo (validação temporal), por exemplo, treinamento em dados antes de uma data específica e avaliação nos dados após essa data. Isso pode nos informar se nosso modelo se generaliza para diferentes períodos de tempo;
 - Divida os dados com base na localização geográfica (validação espacial).
- A **validação cruzada** é útil quando os dados são limitados. Os dados são divididos em **n** conjuntos de tamanhos iguais, onde **n** precisa ser pré-especificado (por exemplo, **n=10**). Para cada um desses conjuntos, um modelo é treinado em todos os dados, exceto os dados desse conjunto, e é usado para gerar previsões para o conjunto de espera. Dessa forma, todos os dados são usados uma vez para avaliar o algoritmo de construção do modelo. Na estrutura de previsão no nível do paciente, usamos a validação cruzada para selecionar os hiperparâmetros ideais.

A validação externa visa avaliar o desempenho do modelo nos dados de outro banco de dados, ou seja, fora das configurações em que foi desenvolvido. Essa medida de transportabilidade do modelo é importante porque queremos aplicar nossos modelos não apenas no banco de dados em que ele foi treinado. Diferentes bancos de dados podem representar diferentes populações de pacientes, diferentes sistemas de saúde e diferentes processos de captura de dados. Acreditamos que a validação externa de modelos de previsão em um grande conjunto de bancos de dados é uma etapa crucial na aceitação e implementação de modelos na prática clínica.

3.3.4.2 Métricas de desempenho

Limiar de predição:

Um modelo de predição atribui um valor entre 0 e 1 para cada paciente correspondendo ao risco do paciente ter o desfecho durante o período de risco. Um valor de 0 significa risco de 0%, um valor de 0,5 significa risco de 50% e um valor de 1 significa risco de 100%. Métricas comuns, como precisão, sensibilidade, especificidade e valor preditivo positivo, podem ser calculadas especificando primeiro um limiar que é usado para classificar os pacientes como tendo o desfecho ou não durante o período de risco. Por exemplo, dada a Tabela 3.15, se definirmos o limiar como 0,5. Os pacientes 1, 3, 7 e 10 tiveram um risco previsto maior ou igual ao limiar de 0,5, de modo que a previsão será de ter o desfecho. Todos os outros pacientes tiveram um risco previsto menor que 0,5, portanto, a previsão para eles será de não ter o desfecho.

Tabela 3.15 Exemplo de uso de um limite na probabilidade prevista

<i>Patient ID</i>	<i>Predicted risk</i>	<i>Predicted class at 0.5 threshold</i>	<i>Has outcome during time-at-risk</i>	<i>Type</i>
1	0.8	1	1	TP
2	0.1	0	0	TN
3	0.7	1	0	FP
4	0	0	0	TN
5	0.05	0	0	TN
6	0.1	0	0	TN
7	0.9	1	1	TP
8	0.2	0	1	FN
9	0.3	0	0	TN
10	0.5	1	0	FP

Se é previsto que um paciente tenha o desfecho, e isso acontece durante o período de risco, será definido como verdadeiro positivo (*true positive* - TP). Se se prevê que um paciente tenha o desfecho mas isso não acontece, é definido como falso positivo (*false positive* - PF). Se se prevê que um paciente não tenha o desfecho e isso acontece, é definido como verdadeiro negativo (*true negative* - TN). Finalmente, se se prevê que um paciente não vai ter o desfecho mas o desfecho acontece, será chamado de falso negativo (*false negative* - NF).

As seguintes métricas baseadas em limite podem ser calculadas:

- precisão: $(TP + TN) / (TP + TN + FP + FN)$
- sensibilidade: $TP / (TP + FN)$
- especificidade: $TN / (TN + FP)$
- valor preditivo positivo: $TP / (TP + FP)$

Observe que esses valores podem diminuir ou aumentar se o limiar for reduzido. A redução do limiar de um classificador pode aumentar o denominador, aumentando o número de resultados retornados. Se o limiar foi definido anteriormente muito alto, os novos resultados podem ser todos verdadeiros positivos, o que aumentará o valor preditivo positivo. Se o limiar anterior for quase certo ou muito baixo, abaixar ainda mais o limiar introduzirá falsos positivos, diminuindo o valor preditivo positivo. Para sensibilidade, o denominador não depende do limiar do classificador ($TP + FN$ é uma constante). Isso significa que a redução do limiar do classificador pode aumentar a sensibilidade, aumentando o número de resultados positivos verdadeiros. Também é possível que a redução do limiar mantenha a sensibilidade inalterada, enquanto o valor preditivo positivo flutua.

Discriminação:

É a capacidade de atribuir um risco maior aos pacientes que experimentaram o desfecho durante o período de risco. A curva ROC (*Receiver Operating Characteristics*) é criada plotando no eixo **x**, o valor $x = 1 - \text{especificidade}$, e no eixo **y**, o valor $y = \text{sensibilidade}$, para todos os limiares possíveis. Um exemplo de gráfico ROC é apresentado na Figura 3.8.

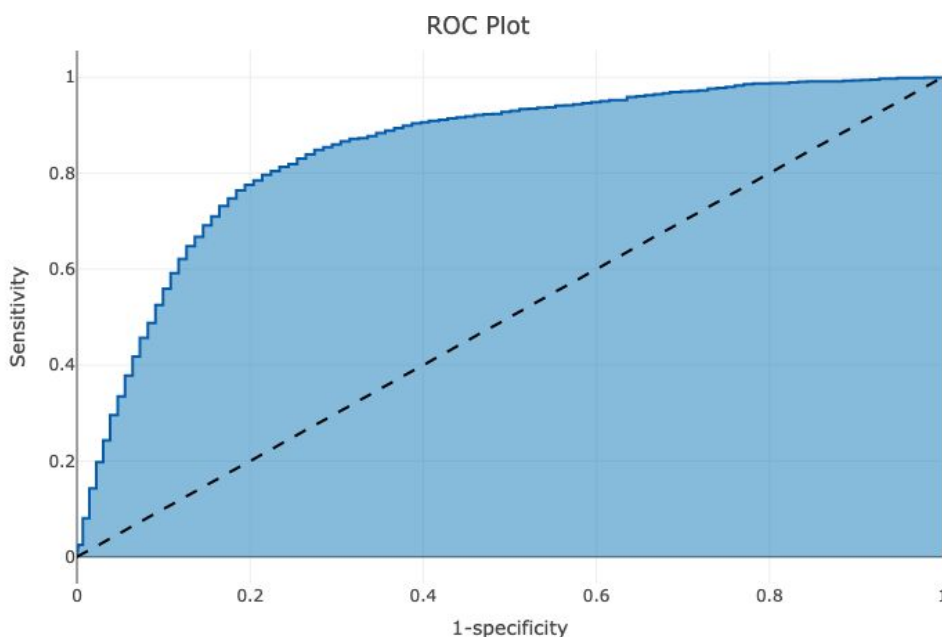


Figura 3.8 Curva ROC

A área sob a curva de característica operacional do receptor (AUC) fornece uma medida geral de discriminação em que um valor de 0,5 corresponde à atribuição aleatória do risco e um valor de 1 significa discriminação perfeita. A maioria dos modelos de predição publicados obtém AUCs entre 0,6-0,8.

A AUC fornece uma maneira de determinar quão diferentes são as distribuições de risco previstas entre os pacientes que experimentam o resultado durante o período de risco e aqueles que não. Se a AUC for alta, as distribuições serão bem separadas, enquanto que quando houver muita sobreposição, a AUC estará mais próxima de 0,5, como mostra a Figura 3.9. Se as duas classes tiverem distribuições semelhantes de risco previsto, o ROC estará próximo da diagonal, com a AUC próxima a 0,5.

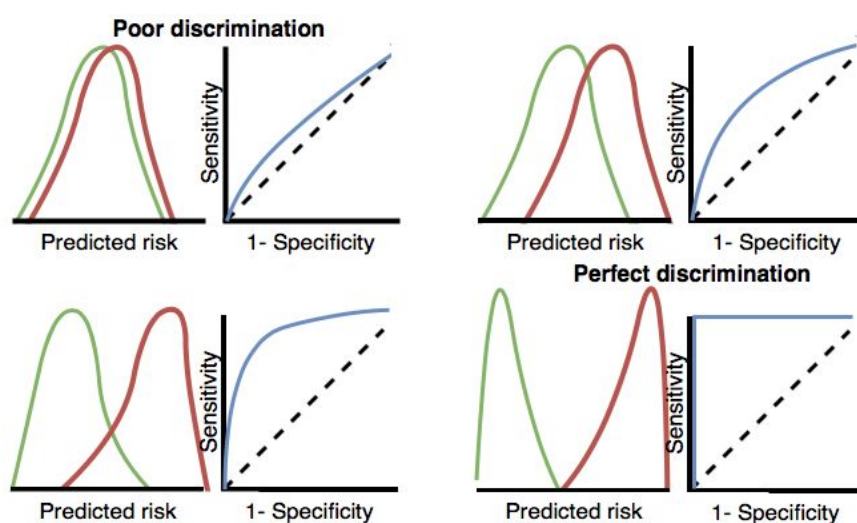


Figura 3.9 Como os gráficos ROC estão relacionados à discriminação

Para desfechos pouco frequentes (casos raros), mesmo um modelo com uma AUC alta pode não ser prático, porque para cada positivo acima de um dado limiar também pode haver muitos negativos (ou seja, o valor preditivo positivo será baixo). Dependendo da gravidade do desfecho e do custo (risco à saúde e/ou monetário) de alguma intervenção, uma alta taxa de falsos positivos pode ser indesejada. Para casos raros, recomenda-se outra medida conhecida como a área sob a curva de precisão-sensibilidade (AUPRC - *area under precision-recall curve*). O AUPRC é a área abaixo da linha gerada, plotando a sensibilidade no eixo x (também conhecida como recall) e o valor preditivo positivo (também conhecido como precisão) no eixo y.

Calibração:

Calibração é a capacidade do modelo de atribuir o risco correto. Por exemplo, se o modelo atribuiu a 100 pacientes um risco de 10%, 10 deles deveriam experimentar o resultado durante o período de risco. Se o modelo designou 100 pacientes com um risco

de 80%, então 80 dos pacientes deveriam experimentar o resultado durante o período de risco. A calibração é geralmente calculada dividindo os pacientes em decis (divisão em 10 partes iguais) do risco previsto e em cada grupo calculando o risco médio previsto e a fração dos pacientes que experimentaram o desfecho durante o período de risco. Em seguida, plotamos esses dez pontos (risco previsto no eixo y e risco observado no eixo x) e verificamos se eles caem na linha $x = y$, indicando que o modelo está bem calibrado. Um exemplo de gráfico de calibração é apresentado mais adiante na Figura 3.10. Também ajustamos um modelo linear usando os pontos para calcular a interceptação (que deve ser próxima de zero) e o gradiente (que deve ser próximo de 1). Se o gradiente for maior que 1, o modelo está atribuindo um risco maior que o risco real e se o gradiente for menor que 1, o modelo está atribuindo um risco menor que o risco real. Observe que implementamos *Smooth Calibration Curves*⁴⁶ no pacote R para capturar melhor a relação não linear entre o risco previsto e o observado.

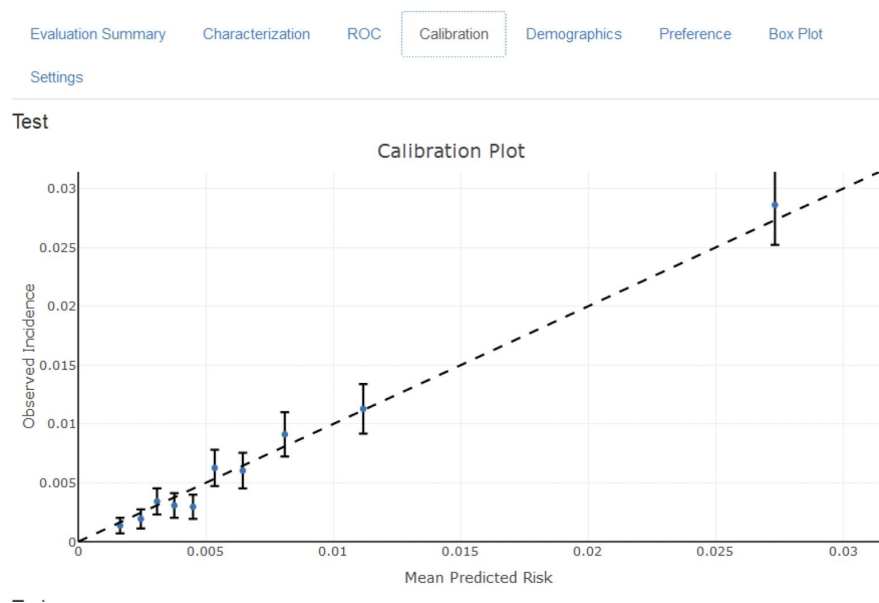


Figura 3.10 Gráfico de calibração

3.3.5 Study-A-Thon PLP - Resultados do COVER

No exemplo do COVER⁴⁷, a aplicação shiny nos permite visualizar os diversos gráficos previamente comentados. Para isso, clicamos no item do menu: *Performance* e visualizamos três abas: Sumário (*Summary*), Discriminação (*Discrimination*) e Calibração (*Calibration*) como mostra a Figura 3.11.

⁴⁶ <https://ohdsi.github.io/PatientLevelPrediction/reference/plotSmoothCalibration.html>

⁴⁷ Visualizando os resultados do estudo COVER <http://evidence.ohdsi.org:3838/Covid19CoverPrediction/>

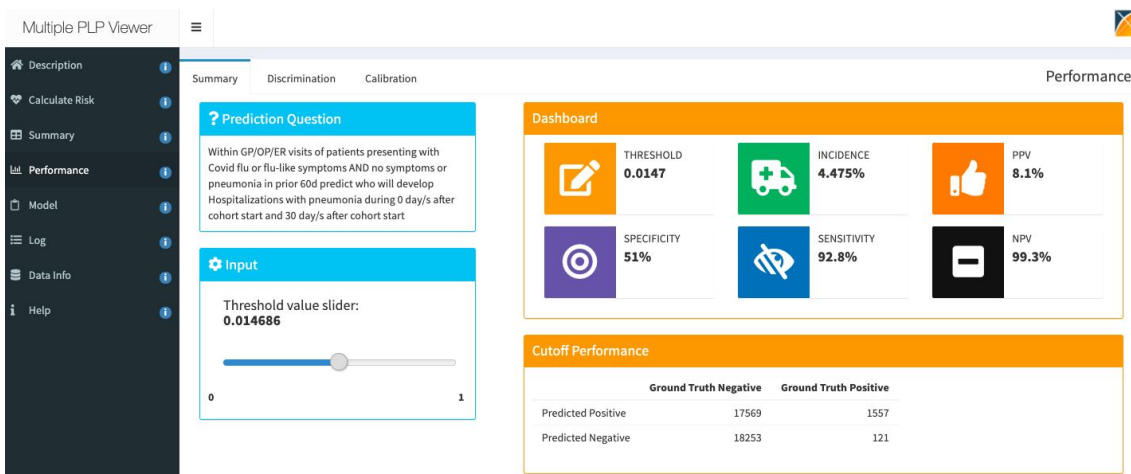


Figura 3.11 Visualização dos resultados do estudo COVER: Performance

Na aba de Sumário, podemos ver os valores principais: Limiar (*threshold*), Incidência, Valor Preditivo Positivo (PPV), Especificidade, Sensibilidade e Valor Preditivo Negativo (NPV) e um controle horizontal que nos permite controlar o valor do limiar e verificar as alterações dos valores principais.

Em particular, no canto inferior direito, temos o quadro do Desempenho do Limiar (*Cutoff Performance*). Nele estão representados como colunas, os valores dos negativos (*Ground Truth Negative*), que são a soma dos valores previstos positivos e negativos (Falsos Positivos + Verdadeiros Negativos), e dos positivos (*Ground Truth Positive*) que são a soma dos Verdadeiros Positivos e dos Falsos Positivos.

Ao alterar o valor do limiar, podemos observar como muda o desempenho e os valores principais.

Na aba de Discriminação (*Discrimination*), apresentada nas Figura 3.12 e Figura 3.13, podemos ver os diversos gráficos de Curva ROC, Precisão-Sensibilidade (PR *Precision-Recall*), Distribuição de Score de Predição e de Preferência e outros.

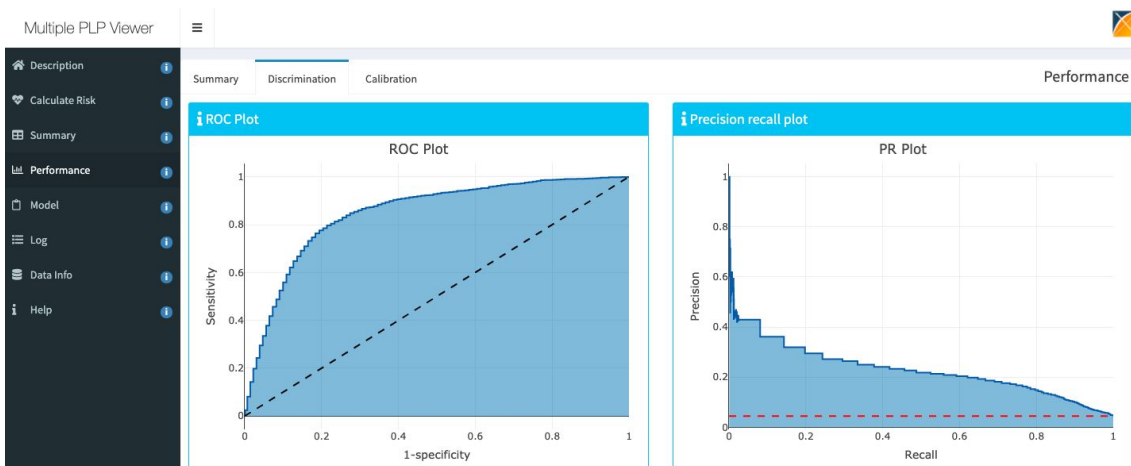


Figura 3.12 Gráficos de Curva ROC, Precisão-Sensibilidade (PR *Precision-Recall*)

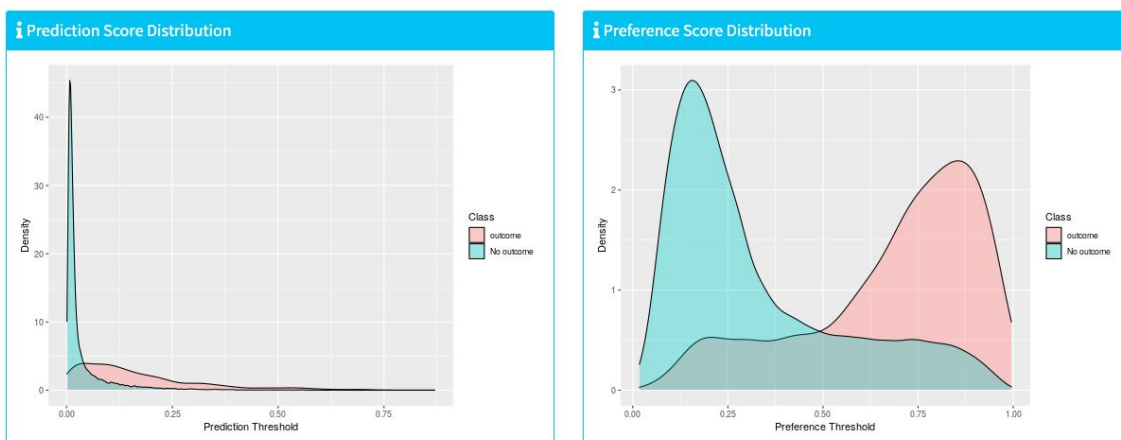


Figura 3.13 Gráficos de Curva ROC, Distribuição de Score de Predição e Preferência

Por último, na aba Calibração, Figura 3.14, podemos ver as curvas de calibração.

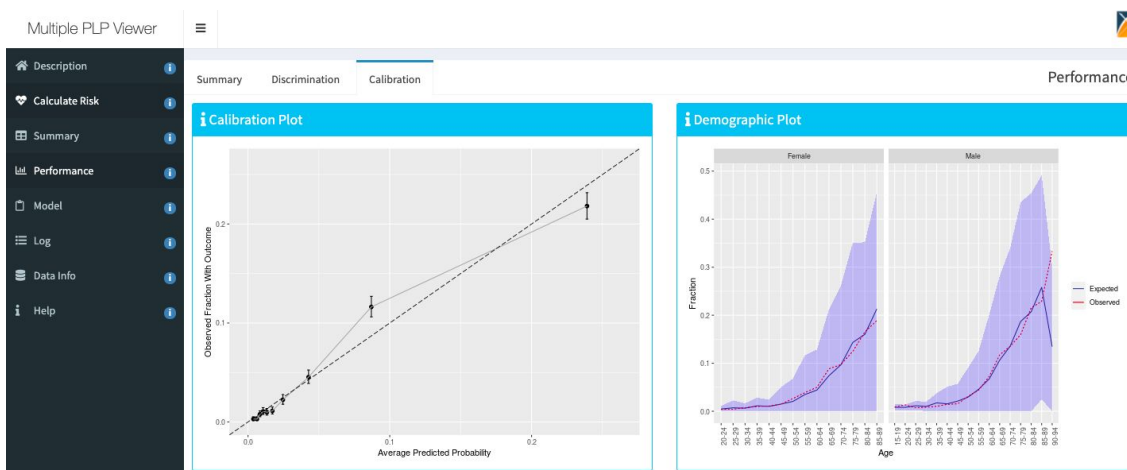


Figura 3.14 Gráficos de Curva ROC, curvas de calibração

3.4. Considerações finais e conclusões

A iniciativa OHDSI é um marco na definição e implementação de estudos observacionais trazendo o conceito de ciência aberta para a realidade da prática clínica através da padronização de um modelo de dados, na definição de fenótipos, coortes, estudos, algoritmos e métodos.

O evento do *Study-A-Thon* a respeito da COVID-19, mostrou esta efetividade numa hora crítica em que foi necessária uma resposta rápida para uma pandemia, mostrando de forma prática os novos caminhos da pesquisa observacional e produzindo evidência clínica relevante numa agilidade antes impensável.

As ferramentas OHDSI ajudam a melhor compreender as nuances dos estudos de efeitos populacionais e dos estudos preditivos além de compilar e deixar disponíveis todas os dados e processos que resultaram nos estudos publicados, desenhando um caminho de aprendizado que auxiliará as próximas gerações de epidemiologistas.

3.5. Referências bibliográfica

Abrahão MTF, Nobre MRC, Madril PJ. O estado da arte em pesquisa observacional de dados de saúde: A iniciativa OHDSI. *Soc Bras Comput* [Internet]. 2019 Jun 11 [cited 2020 Jun 3]; Available from: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/29/98/248-1>

Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat* [Internet]. 2011 Mar [cited 2020 Jun 8];10(2):150–61. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3120982/>

Burn E, You SC, Sena A, Kostka K, Abedtash H, Abrahao MTF, et al. An international characterisation of patients hospitalised with COVID-19 and a comparison with those previously hospitalised with influenza. *medRxiv* [Internet]. 2020 Apr 25 [cited 2020 May 29];2020.04.22.20074336. Available from: <https://www.medrxiv.org/content/10.1101/2020.04.22.20074336v1>

Collins H, Calvo S, Greenberg K, Forman Neall L, Morrison S. Information Needs in the Precision Medicine Era: How Genetics Home Reference Can Help. *Interact J Med Res* [Internet]. 2016 Apr 27 [cited 2020 Jun 3];5(2). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4917728/>

Farrington CP, Anaya-Izquierdo K, Whitaker HJ, Hocine MN, Douglas I, Smeeth L. Self-Controlled Case Series Analysis With Event-Dependent Observation Periods. *J Am Stat Assoc* [Internet]. 2011 Jun 1 [cited 2020 Jun 4];106(494):417–26. Available from: <https://doi.org/10.1198/jasa.2011.ap10108>

Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* [Internet]. 2020 Apr 30 [cited 2020 Jun 3];382(18):1708–20. Available from: <http://www.nejm.org/doi/10.1056/NEJMoa2002032>

Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a

prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*. 2002 Jan 15;155(2):176–84.

Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol* [Internet]. 2016 Apr 15 [cited 2020 Jun 3];183(8):758–64. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4832051/>

Justina LAD, Caluzi JJ, Meghioratti FA. A herança genotípica proposta por Wilhelm Ludwig Johannsen. 2010;5(1):17.

Lane JCE, Weaver J, Kostka K, Duarte-Salles T, Abrahao MTF, Alghoul H, et al. Safety of hydroxychloroquine, alone and in combination with azithromycin, in light of rapid wide-spread use for COVID-19: a multinational, network cohort and self-controlled case series study. *medRxiv* [Internet]. 2020 Apr 10 [cited 2020 May 29];2020.04.08.20054551. Available from: <https://www.medrxiv.org/content/10.1101/2020.04.08.20054551v1>

Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med* [Internet]. 2020 Mar 26 [cited 2020 Jun 3];382(13):1199–207. Available from: <http://www.nejm.org/doi/10.1056/NEJMoa2001316>

Rassen JA, Shelat AA, Myers J, Glynn RJ, Rothman KJ, Schneeweiss S. One-to-many propensity score matching in cohort studies. *Pharmacoepidemiol Drug Saf*. 2012 May;21 Suppl 2:69–80.

Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *J Am Med Inform Assoc* [Internet]. 2018 Aug 1 [cited 2020 Jun 5];25(8):969–75. Available from: <https://academic.oup.com/jamia/article/25/8/969/4989437>

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. 1983;15.

Ryan PB, Schuemie MJ, Madigan D. Empirical performance of a self-controlled cohort method: lessons for developing a risk identification and analysis system. *Drug Saf*. 2013 Oct;36 Suppl 1:S95-106.

Schneeweiss S. Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. *Clin Epidemiol* [Internet]. 2018 Jul 6 [cited 2020 Jun 4];10:771–88. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6039060/>

Suchard MA, Simpson SE, Zorych I, Ryan P, Madigan D. Massive Parallelization of Serial Inference Algorithms for a Complex Generalized Linear Model. *ACM Trans Model Comput Simul* [Internet]. 2013 Jan 1 [cited 2020 Jun 4];23(1):10:1–10:17. Available from: <https://doi.org/10.1145/2414416.2414791>

Suissa S. The case-time-control design. *Epidemiol Camb Mass*. 1995 May;6(3):248–53.

Tian Y, Schuemie MJ, Suchard MA. Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *Int J Epidemiol*. 2018 01;47(6):2005–14.

Walker AM, Patrick AR, Lauer MS, Hornbrook MC, Marin MG, Platt R, et al. A tool for assessing the feasibility of comparative effectiveness research. In 2013.

Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Stat Med*. 2006 May 30;25(10):1768–97.

Williams RD, Markus AF, Yang C, Salles TD, Falconer T, Jonnagaddala J, et al. Seek COVER: Development and validation of a personalized risk calculator for COVID-19 outcomes in an international network. *medRxiv* [Internet]. 2020 May 27 [cited 2020 May 29];2020.05.26.20112649. Available from: <https://www.medrxiv.org/content/10.1101/2020.05.26.20112649v1>

Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* [Internet]. 2020 Mar [cited 2020 Jun 3];579(7798):270–3. Available from: <https://www.nature.com/articles/s41586-020-2012-7>

OHDSI. *The book of OHDSI* <https://github.com/OHDSI/TheBookOfOhdsi> Este livro está licenciado sob a licença *Creative Commons Zero v1.0 Universal*.