

Tratamento de dados, vieses e privacidade

Bruna Almeida Osti* e Fabio Fogliarini Brolesi†

Instituto de Computação, Universidade Estadual de Campinas - UNICAMP
Campinas, SP

Email: *b231024@dac.unicamp.br, †brolesi@gmail.com

Resumo—Este trabalho busca abordar o planejamento do projeto da disciplina, explanando como os dados que serão utilizados foram obtidos e estruturados, possíveis vieses e problemas de privacidade e possíveis tratamentos pra evitá-los, além de como será feito os tratamentos dos dados e estruturação do pipeline de baseline em que os dados serão utilizados.

Index Terms—Trustworthy, AI, Ethical

I. INTRODUÇÃO

Os modelos prognósticos usam vários fatores em combinação para prever o risco de resultados clínicos futuros em pacientes. Um bom modelo deve (i) fornecer previsões precisas que informam os pacientes e seus cuidadores, (ii) apoiar a pesquisa clínica e (iii) permitir decisões para melhorar os resultados dos tratamentos aos pacientes, segundo [1]. Um modelo prognóstico tem três fases principais: desenvolvimento do modelo (incluindo validação interna), validação externa e investigações de impacto na prática clínica. Embora muitos modelos prognósticos sejam propostos, poucos são atualmente usados na prática, conforme mostra [1].

Neste trabalho pretende-se estruturar uma automação que seja capaz de prever quais pacientes sairão da unidade de terapia intensiva dadas características presentes nos datasets fornecidos que tem relação com cirurgias ortopédicas e dados de estadia em unidades de terapia intensiva.

II. DADOS

Os dados que serão utilizados foram disponibilizados pelo Professor e são extraídos dos registros hospitalares do Instituto Nacional de Traumatologia e Ortopedia Jamil Haddad (INTO)¹, que possui uma equipe de tecnologia da informação para pré-processar o texto. Eles referem-se a pacientes que foram submetidos a cirurgias ortopédicas e registros de internações em unidades de terapia intensiva de alguns pacientes.

As observações referem-se a dados com uma única manipulação feita para anonimizar as informações do paciente e do médico. Nenhum outro tipo de dados foi agrupado ou agregado, pois o conjunto de dados fornece informações para cada item da linha (paciente que foi internado no hospital ou descrição detalhada de um procedimento específico do paciente - internação na UTI ou cirurgia).

A partir das modalidades de [2], os dados são estruturados, tabulares, relacionados a *event stream*, ou fluxo de eventos, que são registros feitos de tempos em tempos, mas sem uma frequência definida.

¹<http://www.into.saude.gov.br/Default.aspx>

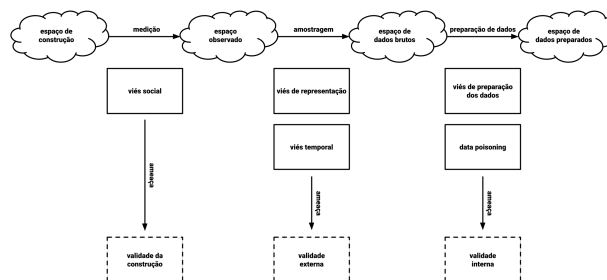


Figura 1. Diagrama de modelo mental de espaços, validades e vieses conforme [2] (adaptado)

O espaço de dados original é bruto, exceto pela transformação de anonimização de pacientes e médicos.

O espaço de dados preparados contempla a união dos conjuntos de dados de pacientes, cirurgias e internações, nesta ordem. Na medida em que o aprendizado de máquina olhará para o prognóstico do paciente, então este é o elemento de interesse, e desta forma, tanto as cirurgias quanto os registros de internações são vinculados a eles. Eventualmente pode ser possível existir pacientes sem cirurgias, ou cirurgias sem internação. Então isto também foi levado em consideração no momento da estruturação do conjunto de dados final.

III. VIÉSES

Considerando os vieses possíveis, [2], conforme Figura 1 e observando os conjuntos de dados utilizados, temos o seguinte:

- **Viés social:** segundo [2], o dado é uma simplificação da realidade. Na medida em que se passa do espaço de construção para o espaço observado, há um julgamento para anotar os dados por parte dos pesquisadores que o fizeram. Este julgamento está sujeito a vieses cognitivos, que podem levar a um viés sociais (mesmo que implícitos) estereotipando categorias de pessoas (mesmo que de maneira inconsciente), gerando desvantagens sistemáticas para pessoas e grupos desprivilegiados. Também os vieses cognitivos no momento de anotar os dados podem fazer com que existam erros sistêmicos. O viés cognitivo citado por [3] também diz que este afeta aspectos da tomada de decisão e raciocínio. Julgamento equivocado, atenção seletiva, crenças de confirmação e preferência por certeza em aspectos médicos em se tratando de ortopedia. Para o presente conjunto de dados, é entendido que pode haver um viés cognitivo que afete os resultados dentro do espaço de construção na medida em que os dados

são originários de resultados de anotações de pessoas que trabalham na área médica e que tem uma visão de mundo condicionada ao estudo na área de saúde (medicina, enfermagem e outros), mas também à realidade socio-histórico-cultural de onde vieram. Neste sentido, é possível que aspectos importantes possam ter sido suprimidos involuntariamente. Como exemplo, desfechos (prognósticos) podem ter sido anotados de forma menos atenta, não trazendo uma precisão para todos os registros. Como outro exemplo, o estudo de [4] fala a respeito de viés de sexo em cirurgias da mão, e para avaliar a presença de viés sexual em pesquisas em condições específicas conhecidas ter maior incidência por sexo, qualquer estudo relacionado à síndrome do túnel do carpo ou à doença de Dupuytren foi revisado separadamente. Estes foram escolhidos a priori porque são condições comuns, amplamente investigados, com conhecida preferência por um sexo específico. A primeira, mais proeminente em mulheres; já a segunda, em homens.

- **Viés de representação:** uma vez o dado coletado, e considerando que o dado é uma simplificação da realidade, ele será representado através de características (*features*). A seleção da amostra, neste caso, pode não ser representativa da população, trazendo a partir do viés de seleção, um recorte que não condiz com a realidade. Também pode ser que, a partir da representação feita, nem todas as características necessárias para construção de um modelo robusto estejam presentes, tornando o resultado mais frágil do que se pretendia. Como exemplo podem existir grupos sub-representados, o que vai afetar o resultado final. No caso de estudos de pacientes que tiveram cirurgias de cunho ortopédico, [5] alerta para um viés a partir da limitação do estudo uma vez que, naquele caso, e também no estudo corrente, as análises são de um único centro clínico. Eventualmente dados de outras localidades trariam maior riqueza para a análise. O estudo de [6] também fala a respeito de dois tipos de vieses a partir do resultado da meta-análise e que representam dois problemas principais: (i) estudos com resultados significativos podem ser super-representados em revisões sistemáticas e meta-análises (chamados falsos positivos) e (ii) estudos sem resultados significativos podem não ser incluídos em revisões sistemáticas e meta-análises (chamados falsos negativos) porque cada estudo, por si só, foi insuficiente, o que significa que algumas opções de tratamento que podem ter efeitos clínicos benéficos não será adotado.
- **Viés temporal:** uma vez que o dado é coletado em algum momento do tempo, este período pode não ser suficiente para representar tudo o que é necessário para apresentação ou modelagem, trazendo apenas um recorte específico e ignorando pontos que poderiam ser cruciais para uma análise com maior acurácia, mas que, dado o corte temporal não foi registrada. Para o caso deste estudo, uma vez que há um recorte de data, pode ser que este recorte temporal não seja representativo para o

fenômeno que se deseja modelar ou que o período seja curto para uma análise mais aprofundada dos dados e posterior construção do modelo.

- **Correlação Expúria:** pelo recorte dos dados não podemos afirmar que as relações do problema são inteiramente observáveis nos dados, o óbito ocorrer depois da cirurgia mas por outros motivos de condições pré-existentes ou não relacionadas com a cirurgia, neste caso, a correlação pode não necessariamente ser uma causalidade. Em [7], é citada a questão da associação percebida mas causada por fator externo relacionado. O exemplo colocado é uma associação positiva ente vida e exercício, mas que uma dieta saudável talvez seja mais importante para aumentar a expectativa de vida, enquanto pessoas com hábitos de exercícios regulares já são mais propensas a uma alimentação saudável. Se a análise incluir exercícios como variável, mas não levou em conta a dieta, pode ser que essa análise leve a uma conclusão errada de que o fator exercício é mais importante do que a dieta.

Sobre vieses, [7] traz alguns pontos relevantes para entendimento do que ele é. Dentre eles o de que o viés é qualquer desvio da verdade no momento da interpretação do dado. Ele também fala sobre o viés de seleção e diz que para minimizar os riscos de aprofundar este viés, ideal é que sejam utilizados certos critérios para selecionar pacientes para um estudo e, em seguida, aleatoriamente atribuir estes pacientes a diferentes grupos. Desta forma, e sem o conhecimento de como a randomização é feita, então este viés acaba sendo diminuído. Outro ponto que [7] para limitar ainda mais o viés de seleção, é ocultar a alocação, onde a indicação de procedimentos para diferentes coortes de estudo são ocultadas para a pesquisa, minimizando o viés.

IV. TRABALHOS RELACIONADOS

Nesta seção iremos abordar alguns trabalhos selecionados que estudaram problemas pós-operatórios de traumas ortopédicos, os trabalhos focam em diferentes faixas etárias e procedimentos realizados. Esses trabalhos são de extrema importância para estabelecer melhores métodos de prevenção, detecção em estágios iniciais e estratégias de intervenção. Existem muitos trabalhos focados em infecção de sítio cirúrgico para cirurgia ortopédica em razão de ser um dos motivos mais comuns de causa de mortalidade nos pacientes.

Em [8] foi descrito a taxa de infecção de sítio cirúrgico em crianças submetidas à cirurgia ortopédica no INTO (Instituto Nacional de Traumatologia e Ortopedia Jamil Haddad) de janeiro de 2012 a dezembro de 2013, e foram incluídos todos os pacientes que possuíam registro de seguimento de até um ano após a data do procedimento, para analisar o perfil desses pacientes. Os pacientes com diagnóstico de infecção de sítio cirúrgico foram pareados com pacientes sem esse diagnóstico com base na idade, na data de internação, na área de atuação ortopédica e no tipo de procedimento cirúrgico.

Foram identificadas 347 cirurgias e 10 infecções de sítio cirúrgico, a análise de correspondência correlacionou a infec-

ção com tempo pré-operatório, peso, escore Z de peso, idade, implante, tipo de procedimento e tempo de internação.

A taxa de infecção foi de 2.88% maior em crianças acima de 24 meses, que realizaram um implante, tiveram mais tempo pré-operatório e de internação.

Por outro lado, em [9] foi investigado a utilidade do preditor MFI para morbidade e mortalidade nos pacientes com trauma ortopédico, o projeto aplicou uma revisão retrospectiva do banco de dados do NSQIP (*National Surgical Quality Improvement Program*) para identificar pacientes com 60 anos ou mais que foram submetidos a cirurgia para fraturas de pelve e membros inferiores entre 2005 e 2014. As fraturas da pelve e das extremidades inferiores em pacientes geriátricos são particularmente incapacitantes devido às restrições de suporte de peso, que requerem dispositivos auxiliares (andador, cadeira de rodas, muletas).

Neste estudo foram identificados mortalidade em 30 dias e complicações pós-operatórias, no qual o desfecho de interesse principal foi a mortalidade em 30 dias. O escore MFI foi comparada à idade, classificação do estado físico e a classificação de feridas usando regressão logística múltipla.

O estudo identificou 36.424 pacientes e entre eles 27.8% dos homens com 79.5 de idade em média, a pontuação média do MFI da amostra foi de 0.12 (± 0.09), o que significa que, em média, os participantes tinham entre uma a duas condições em sua medida de fragilidade. As análises sugeriram que a maioria dos pacientes da tinha uma das 11 variáveis que contribuem para a fragilidade.

Finalmente, o estudo concluiu que o MFI é um preditor significativo de morbidade e mortalidade em pacientes de trauma ortopédico, pois o seu uso pode promover uma ferramenta individualizada de gerenciamento de risco que pode ser usada por um time interdisciplinar para melhorar resultados.

Por fim, [10] buscou determinar a prevalência de condições crônicas de saúde física relatadas antes da lesão, no momento da lesão, até 1 ano após a lesão e de 1 a 5 anos após a lesão. Além disso, verificar o risco de condições crônicas de saúde física relatadas de 1 a 5 anos após a lesão em pessoas com trauma ortopédico e outros tipos de trauma maior.

Este estudo utilizou dados vinculados do Victorian State Trauma Registry (VSTR), do Victorian Registry of Births, Deaths and Marriages (BDM), do Victorian Admitted Episodes Dataset (VAED) e do Victorian Emergency Minimum Dataset (VEMD). Incluindo pacientes com trauma maior (N=28.522) com idade igual ou superior a 18 anos, cadastrados pelo VSTR, com datas de lesão de 2007 a 2016, e que sobreviveram até pelo menos 1 ano após a lesão. Os casos de trauma maior foram classificados em 4 grupos: (1) lesão ortopédica, (2) traumatismo crânio-encefálico grave (s-TBI), (3) lesão medular e (4) outro trauma maior.

Não houve diferenças significativas no risco de condições relatadas de 1 a 5 anos após lesão ortopédica em comparação com outros grupos de trauma maior, entretanto a prevalência de condição crônica foi de 69.3% no geral, as condições mais comuns foram artrite e artropatias, câncer e doenças cardiovasculares. A maior prevalência de novas condições após

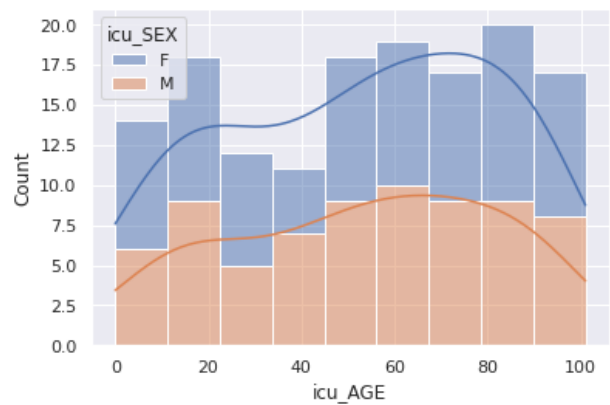


Figura 2. Histograma de idade considerando sexo com o KDE (*kernel density estimation*) associado

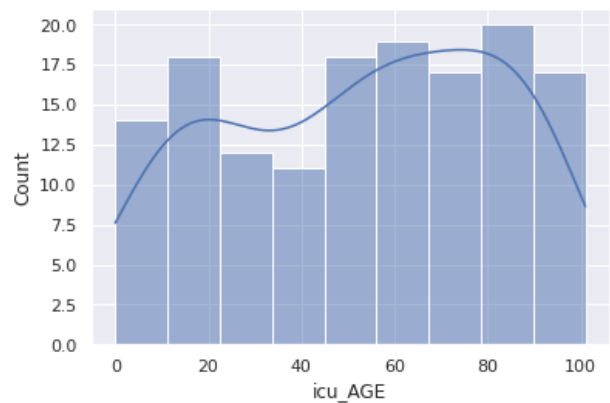


Figura 3. Histograma de idade dos pacientes com o KDE (*kernel density estimation*) associado

a lesão foi encontrada em pessoas com TCE (21.7%) e trauma ortopédico maior (21.4%), enquanto a menor prevalência foi encontrada em pessoas com outros tipos de trauma maior (9.2%).

V. TRATAMENTO DOS DADOS

Para o trabalho, será balanceada a quantidade de amostras da variável de sexo por categoria ou possivelmente retirá-lo, afim de minimizar o viés neste tipo de variável, e tornar o modelo mais equilibrado.

A Figura 2 mostra a distribuição de idade a partir dos sexos, enquanto a Figura 3 mostra apenas a distribuição, ignorando esta variável.

Outro ponto, levantado por [7] é o viés de seleção. Para não montar um modelo que possa segregar baseado em localização, removeremos as features de leito, deixando apenas a ala (localização mais geral) de onde o paciente se encontra.

Outro ponto para garantir a anonimização será a realização do shift temporal. Desta forma, faremos com que as datas tenham uma diferença temporal consistente, mas não a real.

Sobre a etapa de preparação de dados, [11] diz que para a preparação, é preciso ser transparente sobre os dados de

treinamento selecionados em relação às características demográficas dos paciente, como quantidades, distribuição da idade dos pacientes, representação de raça e etnia, assim como o gênero. Também deve-se esforçar para compilar conjuntos de dados tão diversos e grandes quanto possível para ter uma melhor representação de todos os grupos de pacientes.

Já na etapa de modelagem [11] afirma que pode-se perpetuar o viés existentes nos dados. Uma aplicação simplista de um sistema inteligente, aprende boas previsões para a população média, mas não necessariamente faz com que o modelo aprenda bons resultados para aqueles que estão sub-representados nos dados devido ao viés de amostragem.

Para [11], várias estratégias reduzem o risco de viés, como deixar transparente os conjuntos de dados de treinamento selecionados, abordagens matemáticas para despolarização, interpretabilidade ou explicabilidade do modelo e monitoramento. Essas estratégias devem se tornar as melhores práticas para qualquer sistema baseado em aprendizado de máquina desenvolvido com a aplicação médica em mente. É crucial que os vieses sejam mitigados ao desenvolver e implantar sistemas para evitar a desigualdade nos cuidados de saúde para grupos específicos de pacientes e garantir uma funcionalidade segura para todos os pacientes.

Cirurgiões ortopédicos, segundo [7], geralmente dão muita importância ao p-valor nas análises feitas. Tradicionalmente ele tem o limiar no valor arbitrário 0,05. Este valor, cabe ressaltar, veio a partir de uso em experimentos da indústria e agricultura, e portanto não tem bases clínicas. Também segundo [7], o p-valor ignora magnitude, não tem uma precisão bem estimada e não leva os vieses em consideração. Portanto, para evitar uma confiança excessiva é necessário garantir um tamanho de amostras apropriado e utilizar intervalos de confiança. Estes intervalos serão responsáveis pela magnitude, direção da associação, e precisão de uma medida de parâmetro.

VI. ESTRUTURA DE PIPELINE

O pipeline de dados para o presente trabalho utilizará a metodologia presente em 4, a partir da organização e separação dos dados, será segregado em treino e teste, e após criação de modelos estatísticos gerados a partir de referências, será selecionado um conjunto deles para verificação de valores a partir da curva ROC e da matriz de confusão com resultados ponderados.

REFERÊNCIAS

- [1] J. C. Patino, Cecilia Maria e Ferreira, “Prognostic studies for health care decision making,” 2019.
- [2] K. R. Varshney, *Trustworthy Machine Learning*. Chappaqua, NY, USA: Independently Published, 2022.
- [3] S. Janssen, T. Teunis, D. Ring, and R. Parisien, “Cognitive biases in orthopaedic surgery,” *Journal of the American Academy of Orthopaedic Surgeons*, vol. Publish Ahead of Print, 11 2020.
- [4] L. K. Kalliaainen, I. Wisecarver, A. Cummings, and J. Stone, “Sex bias in hand surgery research,” *The Journal of Hand Surgery*, vol. 43, pp. 1026–1029, Nov. 2018.
- [5] Y. Tang, H. Li, and Z. Guo, “Prediction of ICU admission after orthopaedic surgery in elderly patients,” *Pakistan Journal of Medical Sciences*, vol. 37, May 2021.

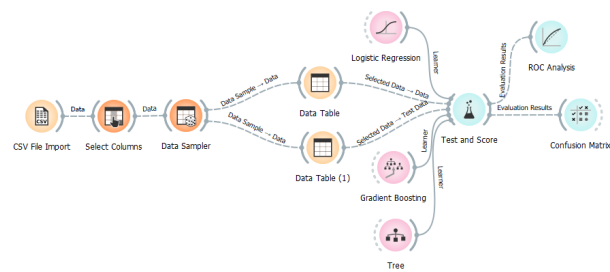


Figura 4. Estrutura esquemática do pipeline de dados: a partir dos dados originais, realizamos as transformações e seleções necessárias; em seguida, separamos em conjunto de treino e teste; a seguir, aplicamos os modelos para avaliarmos os mais adequados, avaliando a partir da curva ROC e matriz de confusão, com valores ponderados.

- [6] J. Scott, J. X. Checketts, C. M. Cooper, M. Boose, C. Wayant, and M. Vassar, “An evaluation of publication bias in high-impact orthopaedic literature,” *JBJS Open Access*, vol. 4, p. e0055, Apr. 2019.
- [7] J. F. Stenson and C. K. Kepler, “Bias in prospective research and how to avoid it,” *Clinical Spine Surgery: A Spine Publication*, vol. 32, pp. 254–255, July 2019.
- [8] M. de Queiroz Leite Chagas, A. M. M. Costa, P. H. B. Mendes, and S. C. G. Júnior, “ANÁLISE DAS INFECÇÕES DE SÍTIO CIRÚRGICO EM PACIENTES PEDIÁTRICOS APÓS CIRURGIA ORTOPÉDICA: UM ESTUDO CASO-CONTROLE,” *Revista Paulista de Pediatria*, vol. 35, pp. 18–24, Mar. 2017.
- [9] C. C. L. Vu, R. P. Runner, W. M. Reisman, and M. L. Schenker, “The frail fail: Increased mortality and post-operative complications in orthopaedic trauma patients,” *Injury*, vol. 48, pp. 2443–2450, Nov. 2017.
- [10] A. Y. Gelaw, B. J. Gabbe, and C. L. Ekegren, “Chronic physical health conditions after injury: A comparison of prevalence and risk in people with orthopedic major trauma and other types of injury,” *Archives of Physical Medicine and Rehabilitation*, vol. 103, no. 9, pp. 1738–1748, 2022.
- [11] K. N. Vokinger, S. Feuerriegel, and A. S. Kesselheim, “Mitigating bias in machine learning for medicine,” *Communications Medicine*, vol. 1, Aug. 2021.

APÊNDICE

A. Dicionário de dados

Tabela	Campo	Tipo	Descrição
surgeries	COD. SURGERY	numérico inteiro	Código que identifica a cirurgia
surgeries	SURGERY DATE	data e hora	Data e hora da cirurgia
surgeries	MEDICAL RECORD	numérico inteiro	Código de registro médico
surgeries	PATIENT NAME	numérico inteiro	Código referenciando o nome do paciente
surgeries	PROTOCOL NUMBER	numérico inteiro	Número de protocolo
surgeries	SERVICE	texto	Tipo de serviço efetuado
surgeries	MAIN PROVIDER	numérico inteiro	Código do fornecedor
surgeries	ANESTHESIST	numérico inteiro	Código numérico referente à pessoa anestesista
surgeries	SURGERY	texto	Descritivo alfanumérico com código e texto do tipo de cirurgia
surgeries	VALUE	numérico decimal	Valor da cirurgia
surgeries	CID	texto	Descritivo com código CID
surgeries	SURGICAL ROOM	texto	Localização da sala cirúrgica no hospital
surgeries	ENTRY ROOM	data e hora	Momento da entrada na sala
surgeries	EXIT ROOM	data e hora	Momento da saída da sala
surgeries	START ANESTHESIA	data e hora	Momento do início da anestesia
surgeries	END ANESTHESIA	data e hora	Momento do fim da anestesia
surgeries	START SURGERY	data e hora	Momento do início da cirurgia
surgeries	SURGERY END	data e hora	Momento do fim da cirurgia
surgeries	CLEANING START	data e hora	Momento do início da limpeza
surgeries	CLEANING END	data e hora	Momento do término da limpeza
surgeries	DURATION	numérico decimal	Duração
surgeries	DURATION TIME	hora	Tempo de duração em horas
surgeries	SUB_QUEUE ID	numérico inteiro	Código de especificidade de tipo de problema ortopédico
surgeries	SUB_QUEUE	texto	Descrição da especificidade de tipo de problema ortopédico
surgeries	DATE_QUEUE	data	Data de inclusão do tipo de problema ortopédico
surgeries	ADMISSION TYPE	texto	Tipo de admissão do paciente
icu	PROTOCOL NUMBER	numérico inteiro	Código que identifica a internação
icu	OPENING MEDICAL RECORD DATE	data e hora	Momento da internação
icu	ADMISSION DATE	data e hora	Momento de admissão na UTI
icu	DISCHARGE FORECAST	data e hora	Momento de saída da UTI
icu	DISCHARGE DATE	data e hora	Data de saída da UTI
icu	LENGTH OF STAY	numérico inteiro	Dias de internação
icu	MEDICAL RECORD	numérico inteiro	Identificador de registro médico
icu	PATIENT NAME	numérico inteiro	Código referenciando o nome do paciente
icu	AGE	numérico inteiro	Idade do paciente
icu	SEX	texto	Sexo do paciente
icu	HOSPITALIZATION TYPE	texto	Tipo de hospitalização
icu	SERVICE	texto	Serviço executado na hospitalização
icu	1ST SURGERY DATE	data e hora	Momento da primeira cirurgia
icu	TIME UNTIL 1ST SURGERY	numérico inteiro	Tempo até a primeira cirurgia
icu	NUMBER OF SURGERIES	numérico inteiro	Quantidade de cirurgias
icu	TRANSFER PLACE	texto	Destinação final após alta na UTI
icu	DOCTOR	numérico inteiro	Código que identifica o médico
icu	CID	texto	Descritivo com código CID
icu	REASON FOR DISCHARGE	texto	Justificativa da alta na UTI
icu	COMORBIDITIES	numérico booleano	Marcação de comorbidade do paciente (0 para não, 1 para sim)
icu	DATE OF ACCIDENT	data e hora	Data do acidente do paciente
patients	PROTOCOL NUMBER	numérico inteiro	Número de protocolo
patients	ADMISSION DATE	data	Data de admissão
patients	DISCHARGE DATE	data	Data de liberação
patients	MEDICAL RECORD	numérico inteiro	Número do registro médico
patients	PATIENT NAME	numérico inteiro	Código referenciando o nome do paciente
patients	ORIGIN BED	numérico inteiro	Leito de origem
patients	ORIGIN UNIT	texto	Local de origem
patients	DESTINATION BED	texto	Leito de destino
patients	DESTINATION UNIT	texto	Local de destino
patients	ADMISSION ICU	data	Data da entrada
patients	ADMISSION TIME	hora	Hora da entrada
patients	DISCHARGE ICU	data	Data da liberação
patients	DISCHARGE TIME	hora	Hora da liberação
patients	ICU STAY	numérico inteiro	Quantidade de dias de internação
patients	SERVICE	texto	Serviço executado na UTI