



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA

DISSERTAÇÃO DE MESTRADO

Modelos para automatização de análises de *desfechos clínico-assistenciais*

Bruna Stella Zanotto

Orientador: Prof. Dra. Carisi Anne Polanczyk

Co-orientador: Prof. Dra. Ana Paula da Silva Beck Etges

Porto Alegre, Brasil

2020



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA

DISSERTAÇÃO DE MESTRADO

Modelos para automatização de análises de *desfechos clínico-assistenciais*

Bruna Stella Zanotto

Orientador: Prof. Dra. Carisi Anne Polanczyk

A apresentação desta dissertação é exigência do Programa de Pós-graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul, para obtenção do título de Mestre.

Porto Alegre, Brasil.

2020

CIP - Catalogação na Publicação

Zanotto, Bruna Stella

Modelos para automatização de análises de desfechos clínico-assistenciais / Bruna Stella Zanotto. -- 2020. 160 f.

Orientadora: Carisi Anne Polanczyk.

Coorientadora: Ana Paula Etges.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Medicina, Programa de Pós-Graduação em Epidemiologia, Porto Alegre, BR-RS, 2020.

1. Inteligência artificial. 2. Gestão em saúde. 3. prontuário eletrônico. 4. Valor em saúde. 5. Processamento de linguagem natural. I. Polanczyk, Carisi Anne, orient. II. Etges, Ana Paula, coorient. III. Título.

BANCA EXAMINADORA

Prof Dr Sotero Serrate Mengue, Programa de Pós-Graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul

Prof Dr Antonio Luiz Pinho Ribeiro, Programa de Pós-Graduação em Ciências da Saúde, Universidade Federal de Minas Gerais.

Prof Dra Mariana Recamonde Mendonza, Programa de Pós Graduação em Computação, Universidade Federal do Rio Grande do Sul

DEDICATÓRIA ou MENSAGEM

Dedico este trabalho a minha avó Maria, que se manteve, em meio a chamas, forte, de cabeça erguida e, principalmente resiliente. Ela, que diante das viradas de avesso da vida, nos ensina que felicidade e a garra vêm de dentro pra fora, e que nunca nos falte esperança de dias melhores. És um exemplo de mulher e pilar de toda família.

“Brindemos por tudo de bom que realizamos neste ano, esperando que o próximo
concorra para a nossa paz e prosperidade”

Zulmiro Stella

AGRADECIMENTOS

A realização e o alcance dos objetivos desta pesquisa não teriam sido possíveis sem a contribuição direta de muitas pessoas, pessoas as quais costumo chamar de raras, raras pois compartilharam comigo toda forma de conhecimento, incentivo, e amizade neste ano. Aos meus raros eu dedico enorme agradecimento e reconhecimento da importância de cada um nesse processo.

Agradeço as minhas queridas professoras Carisi e Ana Paula. Quero que saibam que reconheço o privilégio de tê-las conhecido, de trabalhar e aprender com vocês. Carisi, obrigada por confiar na minha capacidade me dando total liberdade para conseguir conduzir esse trabalho. Sou grata por todos os ensinamentos compartilhados de forma admirável, por me guiar desde os primeiros passos no mundo da pesquisa, pós-graduação e ter despertado em mim a vontade de lecionar. Tens uma admirável e virtuosa capacidade de liderança e motivação dentro do nosso grupo de pesquisa.

Ana, obrigada por acreditar em mim e pelos tantos incentivos e desafios propostos de enxergar sempre além. Tenho certeza de que não chegaria aqui sem o seu apoio. Você foi e está sendo muito mais que coorientadora: para mim será sempre mestre e amiga.

Agradeço aos pesquisadores do IATS e NATS, colegas que se tornaram amigos, obrigada pela companhia, e torcida sempre. Aos novos colegas que encontrei nesta caminhada, de outras faculdades e estados do Brasil, sou eternamente grata pelos ensinamentos e pelas diferentes perspectivas que pude enxergar neste ano. Em especial ao professor Marcos, Eduardo e Avner, obrigada pela paciência. Ricardo, obrigada pelo “*team work*” e amizade que se fortaleceram este ano.

Agradeço aos meus amigos que acompanharam toda a trajetória até aqui. Em especial, Ingrid, Bárbara, Lari, Luisa, vocês são maravilhosas, sou muito abençoada em ter pessoas tão queridas que me querem tão bem. Obrigada família, de longe e de perto pelas boas vibrações e mensagens de incentivo. Em especial ao meu padrinho Steffan e Flávia, conversas com vocês sempre leves e repletas de aprendizado, vocês são exemplos de profissionais, pais e amigos.

Obrigada Gustavo pela cumplicidade, chegou tão de repente e de maneira tão sincera me mostrou a virtude que é ter alguém do lado, o significado de companheirismo e admiração. Muito feliz em compartilhar este momento contigo.

Meu maior agradecimento a meus pais, Cesar e Maristela, que me viram crescer e

deixaram com que eu tomasse decisões importantes. Obrigada por terem, com muito carinho e apoio, não medido esforços para que eu conquistasse mais esta etapa da minha vida.

Sou o resultado da força e confiança de cada um de vocês.

SUMÁRIO

| | |
|---|------------|
| 1. INTRODUÇÃO | 9 |
| 2. REVISÃO DE LITERATURA | 12 |
| 2.1. CUIDADOS EM SAÚDE BASEADOS EM VALOR (VBHC)..... | 12 |
| 2.2. DESFECHOS EM SAÚDE | 14 |
| 2.3. PRONTUÁRIO ELETRÔNICO / <i>ELETRONIC HEALTH RECORDS (EHR)</i> | 18 |
| 2.4. A INTELIGÊNCIA ARTIFICIAL COMO CATALISADOR PARA MELHORAR O VALOR NOS CUIDADOS DE SAÚDE | 21 |
| 2.4.1. <i>Aprendizado de máquina clássico / Machine Learning</i> | 22 |
| 2.4.2. <i>Aprendizado Profundo / Deep Learning</i> | 24 |
| 2.4.3. <i>Processamento de Linguagem Natural / Natural Language Processing</i> | 25 |
| 2.5. O PROBLEMA DE CLASSIFICAÇÃO DE TEXTO | 25 |
| 2.5.1. <i>Representação de texto</i> | 26 |
| 2.5.2. <i>Classificadores do texto</i> | 28 |
| 2.5.3. <i>Ontologias como modelo de classificação</i> | 33 |
| 2.6. AVC..... | 34 |
| 3. OBJETIVOS | 37 |
| 4. REFERÊNCIAS BIBLIOGRÁFICAS | 38 |
| 5. ARTIGO(S)..... | 46 |
| 5.1. ARTIGO 1..... | 47 |
| 5.2. ARTIGO 2 | 89 |
| 6. CONCLUSÕES E CONSIDERAÇÕES FINAIS | 116 |
| 7. LIMITAÇÕES DO ESTUDO | 118 |
| 8. TRABALHOS FUTUROS | 119 |
| APÊNDICE A - DETALHES METODOLÓGICOS ADICIONAIS | 120 |
| ANEXO A. CÓPIA DO PARECER DE APROVAÇÃO DO PROJETO NO COMITÊ DE ÉTICA..... | 126 |
| ANEXO B. MANUAL DE TREINAMENTO PARA OS ALGORITMOS..... | 130 |

ABREVIATURAS E SIGLAS

AVC- Acidente Vascular Cerebral

AVCi- Acidente Vascular Cerebral isquêmico

BERT - *Bidirectional Encoder Representations from Transformers*

BoW – Bag of Words

CNN- *Convolutional Neural Networks*

IA – Inteligência Artificial

KNN- *K-Nearest Neighbor*

MF- *Metafeatures*

NBC- *Naïve Bayes Classifier*

NIHSS- National Institute of Health Stroke Scale

PLN- Processamento de Linguagem Natural

PRO – *patient-reported outcome*

PROM- *patient-reported outcome measure*

SVM- *Support Vector Machine*

TDABC- time-driven activity-based costing

TF- *Term-frequency*

TF-IDF- *Term Frequency-Inverse Document Frequency*

TI – Tecnologia de Informação

UPI - Unidade Práticas Integradas

VBHC- Value-based healthcare

VMO- Value Management Office

APRESENTAÇÃO

Este trabalho consiste na dissertação de mestrado intitulada “Proposição de modelo para a automatização de análises de desfechos de saúde” apresentada ao Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul, em Porto Alegre, 11 de dezembro de 2020. O trabalho é apresentado em sete partes, na ordem que segue:

Introdução, Revisão da Literatura e Objetivos

Artigo(s)

Conclusões e Considerações Finais

Limitações do Estudo

Trabalhos Futuros

Apêndices

Anexos

1. INTRODUÇÃO

A fim de compreender e buscar alternativas ao atual cenário de recursos escassos e crescente demanda por qualidade nos serviços de saúde, Porter *et al.* (2010) definiram valor dos cuidados de saúde como resultados de saúde alcançado por cada dólar gasto para alcançar estes resultados (PORTER, 2010). Entre as premissas do modelo de gestão de cuidados de saúde orientados a valor (VBHC), a capacidade tecnológica de medição é fundamental para viabilizar informações que sustentem decisões voltadas à melhoria contínua dos cuidados em saúde (PORTER; LEE, 2013).

Em pesquisa feita pela Medtronic – The Economist, o Brasil posicionou-se atrás de seus vizinhos territoriais nos avanços em prol à adoção de uma gestão de saúde baseada em valor. O domínio mais crítico foi justamente de mensuração de desfechos reportados pelo paciente e de custos (ATEFA SHAH; BLASZCZYNSKI, 2016). Na maioria dos hospitais a mensuração de desfecho ainda é baseada na coleta e relato de desfechos clínicos como mortalidade e complicações intra-hospitalares (KATZ; FRANKEN; MAKDISSE, 2017), mas que não refletem outros aspectos de cuidado que importam para os pacientes.

Na tentativa de traduzir o conceito à prática médica, Lee *et al.* (2010) sugerem ser necessário que as organizações de saúde desenvolvam formas de capturar os dados que permitam avaliar a qualidade da entrega do serviço aos pacientes e os custos reais para prover esses resultados. Estes dados são essenciais para avaliar valor, mas requerem um avanço em capacidade tecnológica para que possam ser monitorados e controlados (TSAI; PORTER; ADAMS, 2018).

Alguns casos de sucesso de avanços tecnológicos para monitoramento de desfechos, como desenvolvido na Holanda (ATEFA SHAH; BLASZCZYNSKI, 2016), demonstram que é possível definir medidas robustas para resultados, coletar informações comparativas e ajustá-las a risco. Esses casos também demonstram que através de uma nova abordagem institucional melhorias substanciais foram alcançadas na capacidade de entrega de valor aos pacientes. Os esforços de medição de resultados existentes, embora imperfeitos, têm reduzido a mortalidade e melhorando outros resultados de saúde, beneficiando a tomada de decisões clínicas e motivando melhorias nos processos assistenciais e de gestão em saúde (PORTER; TEISBERG, 2006).

De acordo com algumas definições, qualquer programa de computador que exiba um

comportamento humano inteligente é uma forma de IA (MEULEN, 2019; RUSSELL, 1995). Alguns autores (FALCONER; NAND; LIOW *et al.*, 2014; GOPAL; SUTER-CRAZZOLARA; TOLDO *et al.*, 2018), alegam o uso de Inteligência Artificial (IA) como parte da entrega do cuidado em saúde, seja através de softwares para priorização de pacientes (VALUCK; BLAISDELL; DUGAN *et al.*, 2017), e para isso algumas técnicas tem se destacado. Alguns autores alegam que o uso de inteligência artificial também é uma forma de viabilizar a adoção da gestão baseada em valor, e para isso algumas técnicas têm se destacado. O processamento de linguagem natural faz uso de técnicas como aprendizado de máquina e aprendizado profundo para reconhecimento de estruturas a nível da semântica. Seu potencial na área de saúde está em conseguir identificar e fornecer um conjunto de informações confiáveis a partir de dados não estruturados, como os que compõem os registros clínicos eletrônicos dos pacientes (MESKÓ; GÖRÖG, 2020).

A agenda de valor apresentar por Porter é composta por seis componentes: unidade práticas integradas (UPI), medida de desfechos e custos para todos os pacientes, pagamento por *bundled* para o ciclo de cuidado, sistema de saúde integrado, alcance geográfico expandido, e desenvolvimento de capacidade tecnológica. Este último, é apresentado de forma transversal a todos os demais, por ser compreendido como essencial para que os outros cinco possam ser implementados. Sistemas de TI consistentes, robustos e integrados podem ajudar as partes de uma UPI a trabalharem juntas, permitir a medição e novas abordagens de reembolso, enlaçar as partes de uma entrega bem estruturada, e permitir alcance geográfico de assistência em pontos remotos (PORTER; LEE, 2013).

Avanços de capacidade computacional e modelos de IA para melhor analisar os dados de saúde têm sido observados como, também, uma forma de viabilizar a adoção de uma gestão baseada em valor para o sistema de saúde (ETGES; RUSCHEL; POLANCZYK *et al.*, 2020). Para isso, algumas técnicas têm se destacado. O processamento de linguagem natural (PLN) faz uso do aprendizado de máquina a fim de envolver a aplicação do conhecimento do léxico, sintático e semântico do idioma (VIEIRA; LIMA, 2001). Seu potencial na área de saúde está em conseguir identificar e fornecer um conjunto de informações confiáveis a partir de dados não estruturados, como os que compõem os registros clínicos eletrônicos dos pacientes (ROSENBLOOM; DENNY; XU *et al.*, 2011).

Por esse motivo, tem-se um investimento na aplicação de técnicas de PLN em diferentes áreas da saúde, entre elas, saúde mental para avaliação de alterações de humor e risco de suicídio (PESTIAN; NASRALLAH; MATYKIEWICZ *et al.*, 2010; VELUPILLAI;

SUOMINEN; LIAKATA *et al.*, 2018), avaliação de preditores e desfechos em grandes amostras da pesquisa clínica (WANG, S. Y.; PERSHING, S.; TRAN, E. *et al.*, 2020), seleção para grupo de pacientes para pesquisa clínica (SILVA, 2018), e identificação de determinantes e comorbidades a nível individual do paciente como uso de cigarro, circunstâncias sociais entre outros (BELL; KILIC; PRABAKARAN *et al.*, 2013; LYNCH; MOORE, 2016).

Na busca por migrar a gestão dos cuidados em saúde para valor, ferramentas que auxiliem na coleta, e análise de dados são avanços fundamentais para proporcionar o monitoramento assertivo e ágil de resultados (ETGES; RUSCHEL; POLANCZYK *et al.*, 2020). Já vem sendo descrito na literatura a lacuna de tecnologia da informação para automatizar a coleta de dados de resultados a partir de registros clínicos eletrônicos (ETGES; RUSCHEL; POLANCZYK *et al.*, 2020; LEE, 2010). A disponibilidade de uma sólida plataforma de tecnologia de informação, que inclui infra-estrutura de *big data*, pode acelerar a adoção de gestão de saúde baseada em valor e ajudar a convencer decisores clínicos e administrativos de que padronização de conjuntos de resultados pode ser alcançado rapidamente para uma crescente gama de condições médicas (KATZ; FRANKEN; MAKDISSE, 2017).

Considerando o cenário brasileiro e a necessidade de se propor novas maneiras de medir desfechos, o tema desta dissertação é o uso de ferramentas tecnológicas informacionais para suportar modelos de avaliação de valor em saúde.

2. REVISÃO DE LITERATURA

Para melhor contextualização dos conceitos abordados neste trabalho, o referencial teórico está dividido em 6 tópicos: cuidados em saúde baseados em valor, desfechos em saúde, prontuários eletrônicos, inteligência artificial como catalisador para melhorar o valor em cuidados em saúde, classificação de texto e, acidente vascular isquêmico (AVC).

2.1. CUIDADOS EM SAÚDE BASEADOS EM VALOR (VBHC)

Historicamente, os serviços e pagamentos de saúde nem sempre estão alinhados com a qualidade dos resultados observados no estado de saúde dos pacientes. Os modelos de pagamento por procedimento (*fee-for-service*) e por capitação limitam a concorrência do setor e a capacidade dos sistemas de saúde de realizar mudanças transformadoras (HERNANDEZ A.; KAPLAN R.S.; WITKOWSKI M.L. *et al.*, 2019).

No livro “*Redefining Health Care*” Porter e Teisberg (2007) debruçam-se sobre o tema da combinação de altos custos, qualidade insatisfatória, acesso limitado na assistência a saúde, e propõem que através do redesenho do sistema de forma orientada a valor, é possível alcançar melhores resultados para a saúde da população sem que os custos se elevem continuamente. A estratégia proposta para transformar o sistema de saúde foi fundamentada no conceito de valor para os pacientes. O valor na assistência à saúde foi definido pelos autores como a razão entre os resultados de saúde que importam para os pacientes e o custo para entregar estes resultados (PORTER, 2010). Merece um destaque importante que a unidade a ser analisado o valor é a condição do paciente: um conjunto inter-relacionado de circunstâncias médicas do paciente que são mais bem endereçados de uma forma integrada, podendo considerar complicações e doenças simultâneas que devem ser tratadas sob a perspectiva do paciente e seu estado de saúde. Os autores ainda ressaltam que a reestruturação da saúde requer a substituição do sistema fragmentado de hoje, baseado no volume e lucratividade (VAN LEIJEN-ZEELLENBERG; ELISSEN; GRUBE *et al.*, 2016) para um sistema centrado no paciente e organizado em torno de suas condições clínicas (PORTER; LARSSON; LEE, 2016).

Para orientar a implementação da definição conceitual de VBHC, seis elementos interdependentes foram sugeridos, e juntos constituem a agenda de valor proposta por Porter

et al. (2010) (PORTER, 2010). Desde então, líderes de sistemas de saúde e pesquisadores têm investido esforços em desenvolver ferramentas para implementar a agenda. Algumas organizações e instituições já estão nos estágios de implementar os vários dos componentes estratégicos para tal mudança, como Cleveland Clinic nos Estados Unidos (PORTER, 2009), Martini Klinik na Alemanha (PORTER, 2014) e *Value Management Office* (VMO) no Brasil (MAKDISSE; KATZ; RAMOS *et al.*, 2018). Os seis componentes estratégicos são:

Instituições de saúde devem se organizar em unidades práticas integradas (UPI), onde a equipe dedicada fornece o ciclo completo de cuidados para cada condição de saúde. Para isso, é sugerido o estabelecimento de uma entidade organizacional que concentra múltiplas especialidades e se organiza ao redor de uma condição clínica, com objetivo de entregar o máximo valor para os pacientes ao menor custo possível. O principal caso de sucesso conhecido de IPU é a Martini's Klinik (PORTER, 2014) na Alemanha, fundada em 2005 e concentrou-se unicamente no tratamento do câncer de próstata com o compromisso de melhorar a qualidade do serviço na Alemanha.

As instituições de saúde devem medir os resultados e os custos para todos os pacientes em condições clínicas específicas durante todo o ciclo de cuidados. Para qualquer condição de saúde os resultados são considerados em camadas com três níveis de hierarquia, sendo eles: variação do estado de saúde (nível 1); o processo de recuperação (nível 2), e sustentabilidade da saúde (nível 3). Assim como os desfechos, os custos devem ser medidos a nível de paciente para cada condição clínica e por isso, o método de microcusteio baseado em atividade e tempo (TDABC) tem sido recomendado, uma vez que é um método mais robusto no que diz respeito aos recursos que estão sendo usados para gerenciar a condição de saúde do paciente (ETGES; RUSCHEL; POLANCZYK *et al.*, 2020; PORTER; LARSSON; LEE, 2016; THAKER; PUGH; MAHMOOD *et al.*, 2016; TSAI; PORTER; ADAMS, 2018).

Instituições de saúde devem ser financiadas por meio de pagamento "*bundled*". Os pagamentos devem ser efetuados para uma condição de saúde através do ciclo completo de cuidados. O atual cenário, em que os pagamentos são feitos e incentivados a aumentar seu volume, é um fator limitante para a implementação de uma gestão em saúde centrada no paciente. Se há apenas um valor de pagamento para determinado procedimento, mas a quantidade de tempo, recursos humanos ou materiais necessários para fornecer o serviço variam de paciente para paciente, o prestador de serviços que atender pacientes complexos e de alto risco pode estar sendo penalizado financeiramente. O pacote de pagamentos para todo o ciclo de atendimento obriga os provedores de saúde a aumentar os resultados e diminuir os

custos em todo o ciclo de atendimento. Como tal, os prestadores de cuidados de saúde são responsáveis pelo valor de todo o percurso de atendimento, desde o diagnóstico à reabilitação (VAN VEGHEL; SCHULZ; VAN STRATEN *et al.*, 2018).

As instituições de saúde devem unificar a entrega de cuidados, a fim de eliminar a fragmentação e duplicação de atendimento, e otimizar os tipos de atendimento prestado em cada localização. Esforços que contribuam para que o paciente com sua devida complexidade chegue no local correto, pelo meio correto, no momento correto (PORTER, 2017; 2019).

As instituições de saúde devem expandir geograficamente. Os provedores de serviços devem concentrar-se em aumentar o seu alcance clínico, por meio do direto ou indireto envolvimento com outros provedores, assistência por telemedicina ou clínicas de acesso a atenção básica em saúde (PORTER, 2010).

Além da expansão geográfica, as instituições de saúde devem ter capacidade tecnológica informacional para sustentar a implementação dos demais componentes da agenda de valor. Um requisito importante desses componentes é a disponibilidade de uma plataforma de TI abrangente, que permite a integração dos dados do paciente e o rastreamento contínuo de métricas (MESKÓ; GÖRÖG, 2020).

Embora cada componente da estratégia de valor de Porter e Lee (2010) seja mutualmente importante, esta pesquisa se concentrou nos elementos de medir os resultados e os custos para cada paciente e de uso da tecnologia de informação.

2.2. DESFECHOS EM SAÚDE

Em relação ao numerador da equação de valor, enquanto na indústria e produção de serviços as medidas de desempenho referem-se a função do produto na cadeia de valor (COOPER, 2017), na área da saúde tais medidas são derivadas de diretrizes clínicas baseadas em evidências, que enfatizavam a medição do desfecho (LEUNG; VAN MERODE, 2019; PORTER; LEE, 2013). Tipos de medições de desfecho frequentemente inspiram-se na abordagem *donabedian*, descrita pela primeira vez em 1988, em que as medidas são categorizadas em três categorias: estrutura, processo e resultado (DONABEDIAN, 1988). As medidas estruturais referem-se a estruturas de suporte que permite o atendimento seguro e eficaz. As medidas de processo referem-se a processos de tratamento, cuidado e manejo clínico (DONABEDIAN, 2002). E, finalmente, as medidas de resultado incluem o estado de saúde, medidas, experiência do paciente e qualidade de vida. É importante destacar que os

resultados de saúde devem ser medidos considerando o ciclo completo de cuidados de uma condição médica (KATZ; FRANKEN; MAKDISSE, 2017).

No entanto, mesmo em 2016, resultados não foram medidos com a frequência que deveriam ser. Em uma análise de 1.958 medidas da *US National Quality Measurement Clearinghouse* de um registro de medições de várias organizações de relatórios de qualidade, mostrou que apenas 7% as medidas foram realmente de resultados e menos de 2% foram resultados relatados pelo paciente (PORTER; LARSSON; LEE, 2016). Ainda, a medição de variáveis de processo tem certo efeito limitador sobre o valor, tal medidas recebem pouca atenção de pacientes, que estão interessados nos resultados (PORTER; LARSSON; LEE, 2016).

Dado que não se tem um mecanismo de padronização, não é incomum encontramos resultados inconsistentes de medidas e definições utilizadas pelas organizações provedoras, sociedades de especialidade, pagadores e países (DONABEDIAN, 1988; IORIO; CLAIR; INNEH *et al.*, 2016; NORIA; NEEDLEMAN; MIKAMI *et al.*, 2015). A definição de mensuração de desfechos em saúde frequentemente é sobreposta com a de medição de indicadores de qualidade assistencial. No entanto, a mensuração de desfechos visa superar as limitações impostas pelo uso de indicadores, uma vez que, neste contexto, um indicador de qualidade tem pouco valor se o desempenho que está sendo medido não pode ser melhorado (por exemplo, fornecendo feedback do profissional ou a oportunidade de reflexão) ou se nada é feito com o dado observado do indicador (DRIESSEN; VAN ZWET; HAAZEBROEK *et al.*, 2016).

Para entregar valor, os hospitais devem conseguir abranger todas as categorias de desempenho que são mais significativas para linha de cuidado escolhida e condizentes com sua realidade. O uso de medidas de resultado reportadas pelos pacientes (*patient-reported outcomes, PROMs*) deve se tornar cada vez mais frequente entre as formas de avaliação de desempenho adotadas pelos hospitais que almejam estabelecer uma gestão baseada em valor (PORTER; LARSSON; LEE, 2016).

Em meio a este contexto, o Consórcio Internacional de Medição de Desfechos em Saúde (*International Consortium of Health Outcomes Measurement, ICHOM*) convocou grupos de especialistas, juntamente com representantes dos pacientes, para delinear conjuntos de métricas de resultados que importam para os pacientes e comorbidades que devem ser coletadas. Esse grupo de trabalho do ICHOM reforça que o papel de tal parametrização não é conceber novas métricas de resultados, mas concordar sobre qual indicador de medida de

desfecho bem avaliado todos devem usar para uma dada condição médica. A natureza internacional do esforço permitiu aos participantes ver que os pacientes com uma condição específica têm a mesma ou semelhante necessidade em diferentes regiões do mundo (ICHOM; 2012).

Para competir em resultados, é preciso que estes sejam mensurados e amplamente difundidos. O desempenho do sistema de saúde só vai ser significativamente melhorado se os resultados forem mensurados e cada participante do sistema, responsabilizado pelos seus resultados (MICHAEL E. PORTER; KAPLAN, 2016). A possibilidade de medir os resultados e de usar a equidade no controle das circunstâncias iniciais do paciente é necessária para medir resultados e para melhorá-los (PORTER, 2010). Como exemplo de medição de desfechos sem o ajuste de risco, pode ser mencionado o caso do estado de Maine nos Estados Unidos: o sistema de tratamento de dependentes químicos em Maine (*The Maine Addiction Treatment System, MATS*) começou a exigir que as clínicas de reabilitação de drogas publicassem seus resultados. Estudos subsequentes mostraram que a melhoria nos resultados alcançados nos anos que se seguiram à legislação foi quase integralmente atribuível ao fato das clínicas afastarem pacientes considerados problemáticos e de alto risco para aumentar suas taxas de sucesso (SHEN, 2003). Também foram demonstradas melhorias no valor para o paciente quando os resultados são medidos e comparados (VAN DEN BERG; DIJKSMAN; KEUS *et al.*, 2020; VAN VEGHEL; SCHULZ; VAN STRATEN *et al.*, 2018), como compreensão da variabilidade na taxa de cesárias feitas pelos prestadores (PBGH, 2014), eficiência no tempo de atendimento e alta do paciente (IORIO; CLAIR; INNEH *et al.*, 2016) e comparação desses indicadores com médias nacionais (PORTER, 2014).

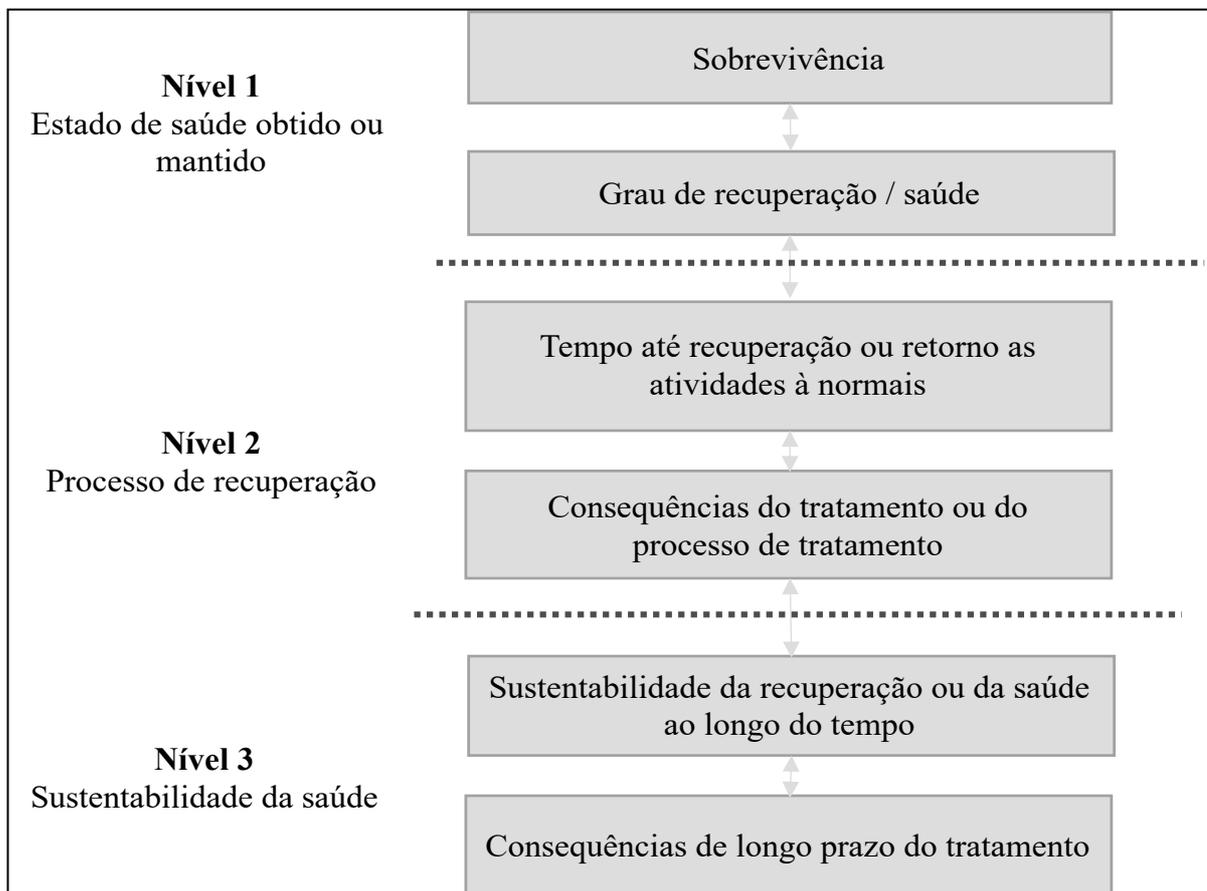
Influenciados pela iniciativa da Martini Klinik, esforços em caráter global de benchmarking em câncer de próstata hoje contém mais de 80 centros em 14 países usando medidas do ICHOM, e tem adicionado centros participantes anualmente (ATEFA SHAH; BLASZCZYNSKI, 2016; PORTER; LARSSON; LEE, 2016). Apesar de iniciativas internacionais, a medição de resultados tem sofrido resistência por parte dos prestadores, que temem vieses e comparações e são facilmente imergidos em inúmeras variáveis de medições que tornam esse trabalho de mensurar a qualidade complexo (PORTER; TEISBERG, 2006). A obrigatoriedade de mensuração e a emissão de boletins de resultados talvez seja o passo mais importante de todos na reforma do sistema de saúde (PORTER; TEISBERG, 2006).

Medidas de resultado muitas vezes falham em capturar dimensões que são altamente importantes para os pacientes (NORIA; NEEDLEMAN; MIKAMI *et al.*, 2015; PORTER;

LARSSON; LEE, 2016). O que geralmente importa para os pacientes são resultados que abrangem todo o ciclo de cuidado - incluindo estado de saúde alcançada (por exemplo, sobrevivência, status funcional, qualidade de vida); o tempo, complicações e sofrimento envolvido em obter cuidados; e a sustentabilidade dos benefícios alcançados (por exemplo, tempo até a recorrência) (PORTER; LARSSON; LEE, 2016; THAKER; PUGH; MAHMOOD *et al.*, 2016).

Diante disso, Porter et al. (2016) introduziu a hierarquia dos desfechos que consiste em resultados para qualquer condição de saúde que pode ser organizada em uma hierarquia de três níveis (**Figura 1**).

Figura 1-Hierarquia de Desfechos proposta por Porter.



Fonte: a autora, traduzida e adaptada de Porter (2010).

Construir modelos de monitoramento de desfechos que contemplem medidas dos três níveis requer gerenciamento da quantidade e qualidade de informações coletadas (IORIO; CLAIR; INNEH *et al.*, 2016; NORIA; NEEDLEMAN; MIKAMI *et al.*, 2015). A definição de formas de coleta e avaliação de desfechos na rotina de atendimento da equipe assistencial

ainda é um obstáculo (MEULEN, 2019; PORTER; LARSSON; LEE, 2016; PORTER; TEISBERG, 2006).

A evolução para registros que informam e fornecem assistência centrada no paciente praticável pode ter dois caminhos principais: primeiro, a educação e disseminação de uma cultura de valor, que instantaneamente reforça a equipe para registrar resultados importantes sobre o paciente, seja por meio de questionários validados ou encontros multidisciplinares mais eficazes; e, segundo, a criação de uma plataforma de medição de resultados em tempo real integrada ao sistema de prontuários eletrônico do hospital (BLUMENTHAL; TAVENNER, 2010). Essa reflexão já foi colocada por diversos autores e nos leva a sugerir que o VBHC não é viável sem investimento em tecnologia da informação (BLUMENTHAL; TAVENNER, 2010; ETGES; RUSCHEL; POLANCZYK *et al.*, 2020; LEE, 2010; MEULEN, 2019).

2.3. PRONTUÁRIO ELETRÔNICO / *ELETRONIC HEALTH RECORDS (EHR)*

Ao descrever sobre registros eletrônicos em saúde, Pinto (2006) constatou que este possui diferentes denominações, sendo elas: prontuário do paciente, prontuário médico e registro do paciente.

A Resolução nº 1638, de 09 de agosto de 2002, no Artigo 1º do Conselho Federal de Medicina (CONSELHO FEDERAL DE MEDICINA, 2002), define o prontuário do paciente como “(...) documento único constituído por um conjunto de informações, sinais e imagens registradas, geradas a partir de fatos, acontecimentos e situações sobre a saúde do paciente e a assistência a ele prestada, de caráter legal, sigiloso e científico, utilizado para possibilitar a comunicação entre membros da equipe multiprofissional e a continuidade da assistência prestada ao indivíduo”.

Tal registro, portanto, é caracterizado pela variedade de autores que produzem seus registros (enfermeiros, médicos, fisioterapeutas, nutricionistas, fonoaudiólogos, psicólogos, assistentes sociais, entre outros) e pela variedade de informações nele registradas, as quais podem ser objetivas ou subjetivas (CONSELHO FEDERAL DE MEDICINA, 2002). Dentre esses dados estão aqueles em formato de texto livre, também denominados de narrativas clínicas como, por exemplo, evoluções do estado clínico do paciente, laudos de procedimentos e exames, sumários de alta, anamnese, descrição dos sintomas, evolução do paciente ao longo de internação, entre outras (HYUN; JOHNSON; BAKKEN, 2009; RUIZ,

2016).

Os registros eletrônicos em saúde, portanto, podem conter campos não estruturados e estruturados. Campos estruturados são dados organizados de tal maneira a rapidamente fornecer a informação, possuem campos específicos de preenchimento e podem ser extraídos em formato de tabelas ou relatórios de indicadores. Geralmente este tipo de dados são os mais utilizados pela facilidade de identificação, uma vez que o conteúdo do campo estruturado representa exatamente a informação como, por exemplo, idade ou uma codificação referente à sexo (NI; WRIGHT; PERENTESIS *et al.*, 2015). Por outro lado, campos não estruturados são aqueles que apresentam a informação em formato de texto. Ao passo que possuem um maior detalhamento, é mais difícil extrair informação deste tipo de dado justamente por estar dependente ao tipo do relato ou escrita de quem o fez. Nesta ótica, os registros produzidos em formato de texto livre apresentam informações com maior riqueza de detalhes se comparados ao conteúdo dos dados estruturados (BEAM, ANDREW L.; KOHANE, ISAAC S., 2018; SHIVADE; HEBERT; LOPETEGUI *et al.*, 2015). Por esta razão, o tratamento adequado dessas informações representa um passo importante para apoiar o julgamento clínico e a tomada de decisões dos profissionais de saúde (NI; WRIGHT; PERENTESIS *et al.*, 2015).

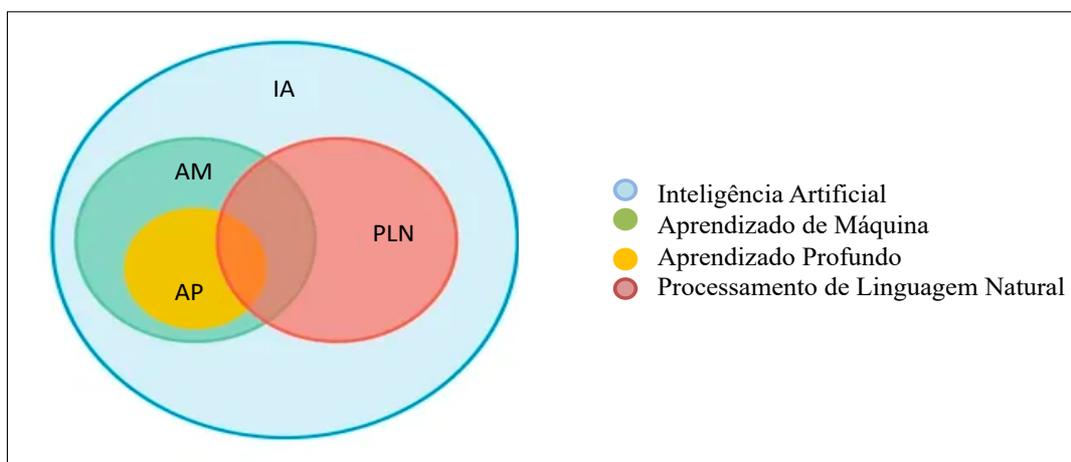
Um exemplo dessa diferença é percebido quando se escolhe utilizar apenas dados estruturados para recuperação de informações. Essa escolha pode resultar na perda de informações clínicas relevantes (HYUN; JOHNSON; BAKKEN, 2009). Por exemplo, a lista de medicamentos pode ser facilmente capturada no formato estruturado, entretanto as nuances relacionadas à administração da medicação, como história dos esquemas anteriores e reações adversas durante o tratamento, são registradas e obtidas nas narrativas clínicas (HYUN; JOHNSON; BAKKEN, 2009). Isto também se aplica às informações colhidas durante a anamnese e transcritas no prontuário como, por exemplo, queixa principal, história da doença atual e pregressa, história familiar e social (GROSSMAN; CARDOSO, 2006). No **Quadro 1** são apresentados dois exemplos de narrativas clínicas.

livre; e *classificação*, que aplica modelos de aprendizado de máquina para identificar resultados derivados dos termos extraídos.

2.4. A INTELIGÊNCIA ARTIFICIAL COMO CATALISADOR PARA MELHORAR O VALOR NOS CUIDADOS DE SAÚDE

Conceitualmente, IA pode ser definida como um conjunto de técnicas analíticas que alavancam grandes conjuntos de dados, estruturados e não estruturados, para oferecer previsões precisas ao encontrar padrões ou classificar dados de interesse (AGRAWAL; GOLDFARB; GANS, 2018; MESKÓ; GÖRÖG, 2020). Técnicas de IA podem ser categorizadas em três grandes grupos (**Figura 2**): i) Técnicas clássicas de aprendizado de máquina, ii) técnicas de aprendizado profundo mais recentes, e iii) métodos de processamento de linguagem natural (PLN) (AGRAWAL; GOLDFARB; GANS, 2018; BEAM, A. L.; KOHANE, I. S., 2018; LING; KURIAN; CASWELL-JIN *et al.*, 2019; MEULEN, 2019; SUN; CAI; LI *et al.*, 2018; XIAO, C.; CHOI, E.; SUN, J., 2018). As três subáreas serão abordadas, porém a ênfase maior será na última, visto seu papel no tema da dissertação.

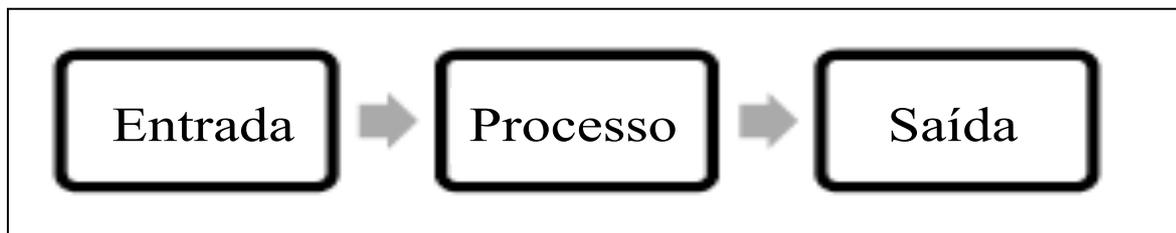
Figura 2- Grande área da Inteligência artificial e como suas subáreas se conectam



Fonte: a autora, 2020.

Algoritmo é uma sequência finita de ações executáveis que visam obter uma solução para um determinado tipo de problema (MESKÓ; GÖRÖG, 2020). Os algoritmos são os blocos de construção da IA, e são divididos em 3 partes, conforme **Figura 3**:

Figura 3-Representação de um algoritmo.



Fonte: a autora, 2020

Um ponto importante é que algoritmos não aceitam dados de texto, então no caso de documentos e textos livres, o dado precisa assumir uma representação numérica para que possam ser submetidos a métodos formais de classificação (KOWSARI; JAFARI; HEIDARYSAFA *et al.*, 2019). A representação numérica é comumente chamada de *feature*. A *feature* compõe a entrada do dado no algoritmo e pode representar, por exemplo, idade, histórico médico e sexo (MEULEN, 2019).

2.4.1. Aprendizado de máquina clássico / Machine Learning

Os algoritmos de aprendizado de máquina aplicam métodos como estatísticas Bayesianas e regressão linear a conjuntos de dados com objetivo de reconhecimento de padrões (MESKÓ; GÖRÖG, 2020; MEULEN, 2019). Os modelos de aprendizado de máquina podem ser categorizados em três grandes grupos: modelos supervisionados, modelos não supervisionados e modelos de reforço (BARBA, 2019; MEULEN, 2019).

Em um modelo de aprendizado supervisionado, o algoritmo aprende com pares específicos de entrada-saída, fornecendo uma resposta-chave que o algoritmo pode usar para avaliar sua precisão nos dados de treinamento (KOWSARI; JAFARI; HEIDARYSAFA *et al.*, 2019; MESKÓ; GÖRÖG, 2020; MEULEN, 2019). Modelos mais avançados precisam de dados anotados para garantir que eles possam aprender a tarefa a qual foram desenhados, de modo que a contribuição dedicada de anotadores de dados familiarizados com formato do banco de dados é de importância crucial para o benefício de implementação de IA no ambiente de saúde (MESKÓ; RADÓ, 2019). A

Figura 4 ilustra o desenvolvimento de modelos supervisionados.

Figura 4 – Ilustração do aprendizado supervisionado.



Traduzido e adaptado pela autora, 2020. Fonte: Ronald van Loon, Big Data Made Simple.

A marcação e a anotação de texto desempenham um papel importante na integração de dados estruturados e não estruturados (LOPES; VIEIRA; FINATTO et al., 2009). Por sua natureza, os dados não estruturados dificultam a extração direta informações significativas e por isso a marcação e a anotação de texto desempenham um papel importante na integração desses dados (KNOX; RITA; EID et al., 2005).

Um modelo não supervisionado, ao contrário, fornece dados não rotulados que o algoritmo tenta entender extraíndo recursos e padrões por conta própria. Problemas de classificação e problemas de regressão são duas grandes áreas onde a aprendizagem supervisionada é útil; enquanto que, quando se tem uma coleção de exemplos sem um resultado desejado específico ou resposta correta, a abordagem não supervisionada pode ser uma alternativa, uma vez que o algoritmo tenta encontrar automaticamente a estrutura nos dados, extraíndo recursos úteis e analisando sua estrutura, como por exemplo, separando a base de dados em grupos de características semelhantes (KOWSARI; JAFARI; HEIDARYSAFA *et al.*, 2019; MESKÓ; GÖRÖG, 2020; MEULEN, 2019). A **Figura 5** ilustra o desenvolvimento de modelos não supervisionados.

Figura 5- Ilustração do aprendizado não supervisionado



Traduzido e adaptado pela autora, 2020. Fonte: Ronald van Loon, Big Data Made Simple.

Por isso, a aprendizagem supervisionada é mais adequada para problemas em que existe um conjunto de pontos de referência disponíveis ou uma verdade fundamental com a qual treinar o algoritmo (KOWSARI; JAFARI; HEIDARYSAFA *et al.*, 2019; MEULEN, 2019; SUN; CAI; LI *et al.*, 2018). Dado que isso nem sempre está disponível, uma abordagem frequente que tem sido adotada é o uso de métodos mistos, denominados modelos semi-supervisionados. A aprendizagem semi-supervisionada consiste no conjunto de dados de treinamento com dados rotulados e não rotulados (MESKÓ; GÖRÖG, 2020; MEULEN, 2019).

2.4.2. *Aprendizado Profundo / Deep Learning*

O aprendizado profundo é um método relativamente novo de aprendizado de máquina que usa “camadas de neurônios” em uma rede para encontrar padrões ocultos em conjuntos de dados complicados, como imagens (MESKÓ; GÖRÖG, 2020). O aprendizado profundo é visto como a próxima etapa do aprendizado de máquina e tem a promessa de criar algoritmos que resolvam alguns dos desafios mais difíceis em IA, como visão computacional e reconhecimento de fala, permitindo o treinamento de algoritmos que funcionam em tempo real com altíssima precisão. Algoritmos de aprendizado profundo usam as chamadas "Redes

Neurais Artificiais" para aprender padrões, inspirados na relação e sinapses dos neurônios do nosso cérebro (XIAO, C.; CHOI, E.; SUN, J., 2018).

2.4.3. Processamento de Linguagem Natural / Natural Language Processing

Em paralelo, estudos de PLN desenvolvidos com o propósito de extração de informação vêm progredindo com o passar dos anos e revelando um caminho que pode melhorar a precisão da combinação paciente e desfecho (NI; WRIGHT; PERENTESIS *et al.*, 2015). Informações clínicas importantes geralmente são registradas em texto livre não estruturado e convertê-lo em um formato estruturado pode ser uma tarefa demorada que pode não capturar com sucesso todas as facetas das informações (KREIMEYER; FOSTER; PANDEY *et al.*, 2017).

As técnicas de PLN aproveitam a IA para analisar grandes volumes de dados textuais para interpretar a linguagem. No PLN, os métodos de análise sintática e semântica analisam o texto de acordo com as regras gramaticais e o significado real do texto, respectivamente. Algoritmos de PLN usam dados não estruturados, como notas clínicas ou textos de livros ou artigos de pesquisa (**Machine Learning for Natural Language Processing**, 2019; LING; KURIAN; CASWELL-JIN *et al.*, 2019; XIAO, CAO; CHOI, EDWARD; SUN, JIMENG, 2018).

2.5. O PROBLEMA DE CLASSIFICAÇÃO DE TEXTO

Estamos diante de uma esmagadora onda de crescimento de dados em todos os setores. Em pesquisa feita pela Ernst Young registrou que 90% dos dados no mundo foram criados nos últimos 2 anos, e que destes 80% eram dados não estruturados (EY, 2019). Constatações como essa evidenciam um paradigma sobre avançar em tecnologia que permita estruturar e entender esses dados, ou se seguir com trabalho manual que permite o alcance de análises menos representativas e escaláveis (XIAO, CAO; CHOI, EDWARD; SUN, JIMENG, 2018).

Infelizmente os avanços de uso de PLN na saúde ainda andam em ritmo mais lento que outros mercados e são pouco explorados na literatura de saúde (XIAO, CAO; CHOI, EDWARD; SUN, JIMENG, 2018). Wang et al. (2018) em sua revisão sistemática aponta que a falta de padronização na linguagem clínica e a dificuldade de acesso aos dados de pacientes

ainda é a principal barreira para o desencontro de estudos que interoperem entre PLN e EHR. Por meio dos esforços de colaboração existentes ou da construção e aproveitamento de uma infraestrutura de computação que preserva a privacidade, será mais comum usar dados de EHR para estruturar narrativas clínicas e apoiar a extração de informações clínicas para apoiar tomada de decisão (WANG, SOPHIA Y.; PERSHING, SUZANN; TRAN, ELAINE et al., 2020).

Além do banco de dados bruto - caracterizado por um documento com vários textos em sequência – é necessário um conjunto estruturado para o propósito de treinamento (KOWSARI; JAFARI; HEIDARYSAFA et al., 2019). Em geral, o sistema de classificação de texto contém quatro níveis diferentes de escopo que podem ser aplicados para construção do vetor, do documento, de parágrafo, de frase, e de sub-sentença (KOWSARI; JAFARI; HEIDARYSAFA et al., 2019), sendo a definição do mais apropriado dependente do objetivo da análise. No nível do documento, o algoritmo obtém as categorias relevantes de um completo documento. No nível de parágrafo, o algoritmo obtém as categorias relevantes de um único parágrafo (uma parte de um documento). No nível de frase, obtém as categorias relevantes de uma única frase (uma parte de um parágrafo). Por fim, no nível de sub-sentença, o algoritmo obtém as categorias relevantes de sub-expressões dentro de uma frase.

A extração e o pré-processamento de atributos são etapas cruciais para aplicações de classificação de texto. Neste sentido, não é incomum que métodos para limpar conjuntos de dados de texto, removendo assim o ruído implícito e permitindo caracterização, sejam a nível de sentença (KOWSARI; JAFARI; HEIDARYSAFA et al., 2019). No entanto, a desvantagem em trabalhar com nível de sentença é que dependendo do método que a quebra dos dados é feita, o contexto formado entre as sentenças pode fornecer informações importantes para a classificação, e pode ser ignorado pelo fato da marcação estar ocorrendo neste nível e não no nível da frase (KOWSARI; JAFARI; HEIDARYSAFA et al., 2019). Detalhes sobre o pré-processamento utilizado no experimento do artigo 2 é apresentado no APÊNDICE A.

2.5.1. Representação de texto

Em geral, textos e documentos são conjuntos de dados não estruturados. Contudo, essas sequências de texto não estruturadas devem ser convertidas em um espaço de atributo estruturado para só assim ser associado a um classificador. Primeiro, os dados precisam ser limpos para omitir caracteres e palavras. Com o texto de saída do pré-processamento, são

utilizados como entrada nos métodos que criam as representações. As técnicas comuns de representação de texto são *Term Frequency-Inverse Document Frequency* (TF-IDF)(ANDRADE; GONÇALVES, 2020), frequência de termo (do inglês, *term-frequency* (TF)) (SALTON; BUCKLEY, 1988), Word2Vec (GOLDBERG; LEVY, 2014), vetores globais para palavras representação (GloVe)(PENNINGTON; SOCHER; MANNING *et al.*, 2014) e *bag-of-words* (BoW).

O TF-IDF é constituído por duas partes, o TF quantifica a frequência dos termos, e o IDF quantifica a frequência inversa de documento, que atribui maior peso as palavras que ocorrem com menos frequência nos documentos (ANDRADE; GONÇALVES, 2020; CUNHA; MANGARAVITE; GOMES *et al.*, 2020). Dessa forma, o escore final de um termo t é obtido pelo $TF * IDF$, que potencializa palavras raras considerando todos os documentos e frequências no documento específico (ANDRADE; GONÇALVES, 2020; CUNHA; CANUTO; VIEGAS *et al.*, 2020; CUNHA; MANGARAVITE; GOMES *et al.*, 2020). Definimos que o TF-IDF seja gerado com *unigrams* e *bigrams*, em que o *bigram* é a combinação do termo atual com o termo posterior. Ainda, podem existir variações do TF-IDF, as mais comuns neste objetivo são Word TF-IDF que corresponde a analisar os termos presentes no documento, e Char TF-IDF que analisa uma sequência de caracteres definida aqui entre o intervalo de 2 a 6 caracteres. Em relação ao Word TF-IDF e Char TF-IDF, o Char consegue capturar melhor termos com erros de digitação porque está analisando parte de um termo definido no intervalo, por exemplo, se estamos analisando o termo ‘tabagista’ e ‘tabaisgta’, ele considera que ambos os termos são similares para um intervalo definido dos quatro primeiros termos. É comum fazer uso da concatenação dos vetores de ambas as representações, Word TF-IDF e Char TF-IDF, em uma única representação e este vetor é utilizado como entrada para um classificador, que serão apresentados na próxima seção. O objetivo desta combinação é potencializar os benefícios de ambas as representações(CUNHA; CANUTO; VIEGAS *et al.*, 2020).

Meta-atributos (do inglês, *metafeatures* ((MFs)) correspondem a representações de documentos derivadas de representações mais básicas, como TFIDF (CANUTO; MARTINS; COUTO *et al.*, 2015; S.; X.; A. *et al.*, 2018). As pontuações dos MFs são obtidas por funções que relacionam dados textuais com categorias (YANG; GOPAL, 2012). Neste trabalho, MFs foram explorados com base nas funções de distância entre exemplos de texto que recentemente obtiveram os resultados mais efetivos em diversos problemas de categorização de texto (CANUTO; MARTINS; COUTO *et al.*, 2015; CUNHA; CANUTO; VIEGAS *et al.*,

2020; CUNHA; MANGARAVITE; GOMES *et al.*, 2020; S.; X.; A. *et al.*, 2018). A racional por trás desses meta-atributos consiste na suposição de que se as distâncias entre um exemplo de texto e os vizinhos mais próximos pertencentes à categoria “c” (e seu centróide correspondente) forem pequenas, então o exemplo provavelmente pertence a “c”. Tais pontuações de distância são evidências adicionais, que quando concatenadas à representação original, têm potencial de auxiliar na construção de modelos de classificação em coleções com uma grande quantidade de exemplos rotulados (CANUTO; SALLES; GONÇALVES *et al.*, 2014).

O word2vec é um método de *word embedding*, esta técnica de aprendizagem de atributos em que cada palavra ou frase do vocabulário é mapeada para um vetor de dimensão N de números reais. É o usual ponto de partida para as tarefas mais profundas de PLN, pois permite que o aprendizado de máquina seja eficaz em conjuntos de dados menores, pois geralmente é a primeira entrada de um modelo e a maneira mais popular de transferir o aprendizado no PLN (KOWSARI; JAFARI; HEIDARYSAFA *et al.*, 2019).

O modelo BoW, é uma representação reduzida e simplificada de um texto documento de partes selecionadas do texto, com base em critérios específicos, como frequência de palavras. A técnica BoW é usada em vários domínios, como visão computacional, PLN, bem como classificação de documentos, e recuperação de informações por aprendizado de máquina (KOWSARI; JAFARI; HEIDARYSAFA *et al.*, 2019).

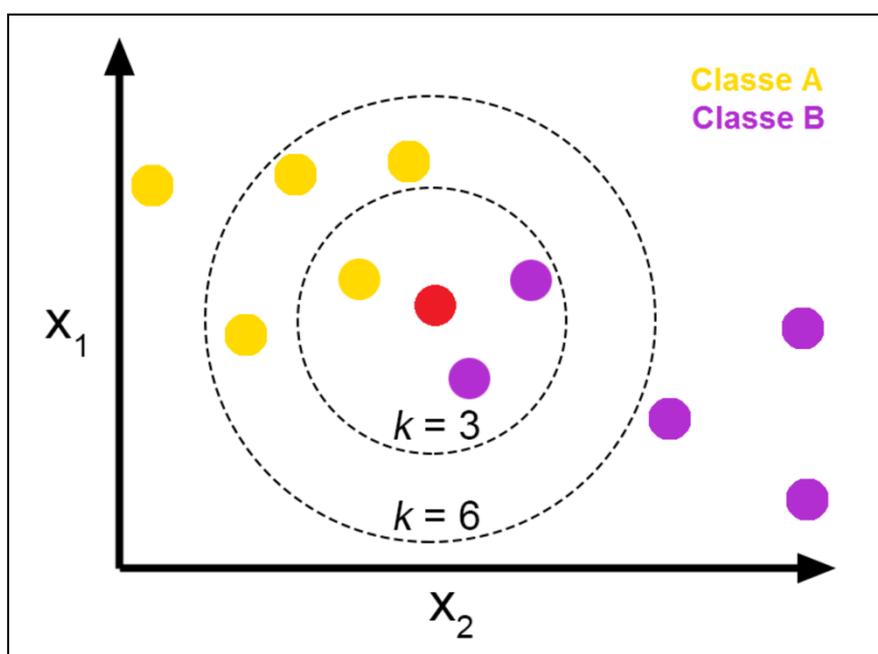
2.5.2. Classificadores do texto

A etapa mais importante de classificação de texto é escolher o melhor classificador. Um dos algoritmos de classificação mais simples é a regressão logística (LR) que foi abordado na maioria dos domínios de mineração de dados (CHEN; XIE; WANG *et al.*, 2017; VELUPILLAI; SUOMINEN; LIAKATA *et al.*, 2018; WANG; WANG; RASTEGAR-MOJARAD *et al.*, 2018). Na história mais antiga da recuperação de informações como uma aplicação viável, o *Naïve Bayes Classifier* (NBC) foi muito popular, e ainda é bastante usado por ser computacionalmente barato e precisar de uma quantidade muito baixa de memória (LARSON, 2009), esta técnica é um modelo generativo, que é o método mais tradicional de categorização de texto (KOWSARI; JAFARI; HEIDARYSAFA *et al.*, 2019).

Técnicas não paramétricas foram estudadas e usadas como tarefas de classificação, como *K-Nearest Neighbor* (KNN) (LI; WEINBERG; DARDEN *et al.*, 2001). Dado um

documento de teste x , o algoritmo KNN encontra os k vizinhos mais próximos de x entre todos os documentos no conjunto de treinamento e pontua os candidatos da categoria com base na classe de k vizinhos. A semelhança de x e o documento de cada vizinho pode ser a pontuação da categoria do vizinho documentos (**Figura 6**). Vários documentos KNN podem pertencer à mesma categoria; neste caso, o somatório de essas pontuações seriam a pontuação de similaridade da classe k em relação ao documento de teste x . Depois de classificar os valores de pontuação, o algoritmo atribui o candidato à classe com a maior pontuação do teste documento x (KOWSARI; JAFARI; HEIDARYSAFA *et al.*, 2019). KNN é um dos os algoritmos de aprendizado de máquina mais simples e tem bom desempenho no básico problemas de reconhecimento (LAYEGHIAN JAVAN; SEPEHRI; AGHAJANI, 2018).

Figura 6- Algoritmo *K-Nearest Neighbor* (KNN)

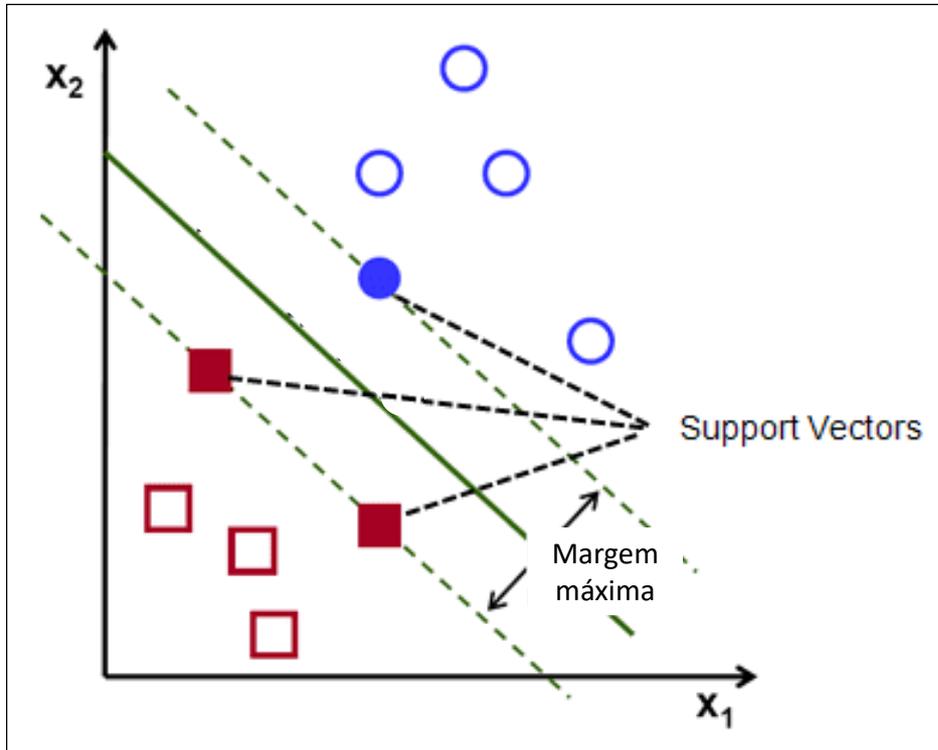


Traduzido e adaptado pela autora, 2020. Fonte: Pedregosa *et al.*, 2011.

Support Vector Machine (SVM) é outra técnica popular e com bons resultados que emprega um classificador discriminativo para categorização de documentos (CORTES; WOLOSZYN; BARONE, 2018; CRISTIANINI; SHAWE-TAYLOR, 2000; CUNHA; MANGARAVITE; GOMES *et al.*, 2020; HAN; KARYPIS, 2000; MANEVITZ; YOUSEF, 2001), e esta entre os métodos mais empregados na área de saúde (LAYEGHIAN JAVAN; SEPEHRI; AGHAJANI, 2018; WANG; KIIK; PEEK *et al.*, 2020). Representa-se o texto através de um vetor de ocorrência de palavras, desconsiderando gramática e ordem das palavras, mas mantendo a multiplicidade das mesmas (**Figura 7**) (YE, 2016). Esta técnica também pode ser usada em todos os domínios da mineração de dados, como bioinformática,

imagem, vídeo, classificação da atividade humana, segurança e saúde. Este modelo também é usado como referência de comparação em muitas pesquisas para destacar novidades e contribuições (KOWSARI; JAFARI; HEIDARYSAFA *et al.*, 2019)./

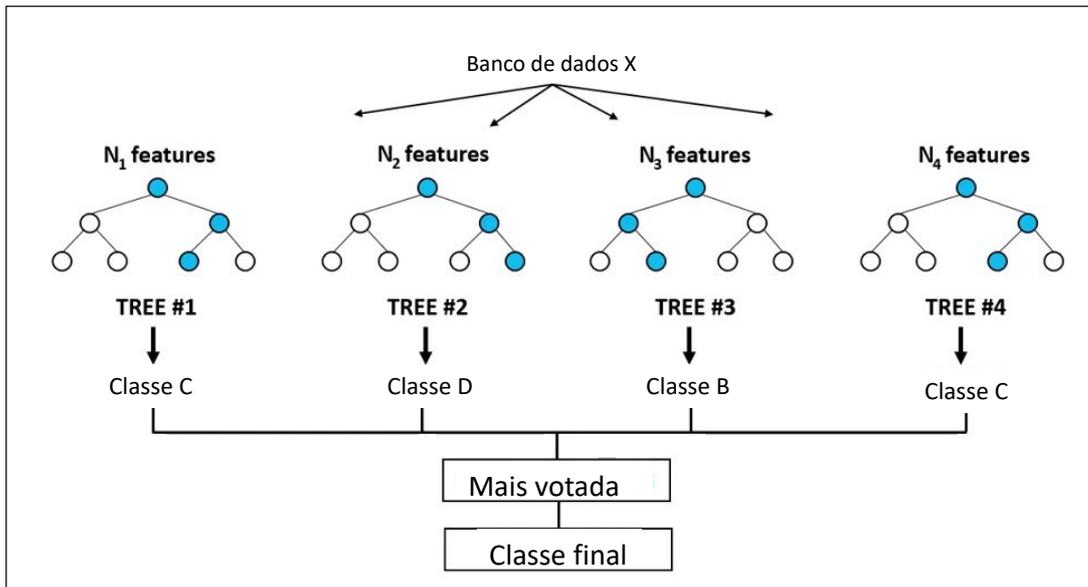
Figura 7- Algoritmo *Support Vector Machine* (SVM)



Traduzido e adaptado pela autora, 2020. Fonte: Pedregosa *et al.*, 2011.

Classificadores “baseados em árvore”, como árvore de decisão e *random forest* (BREIMAN, 2001) também foram estudados com no que diz respeito à categorização do documento (XU; YE; CHENG). O primeiro modelo consiste em um conjunto de decisões hierárquicas que chega a uma decisão final, enquanto a segunda abordagem tem como estratégia o uso de um conjunto de árvores de decisão (WANG; KIIK; PEEK *et al.*, 2020). Nos últimos anos, as classificações “baseadas em árvore” são usadas principalmente para sumarização de documentos (SHEN; SUN; LI *et al.*, 2007) e extração automática de palavras-chave (ZHANG, 2008).

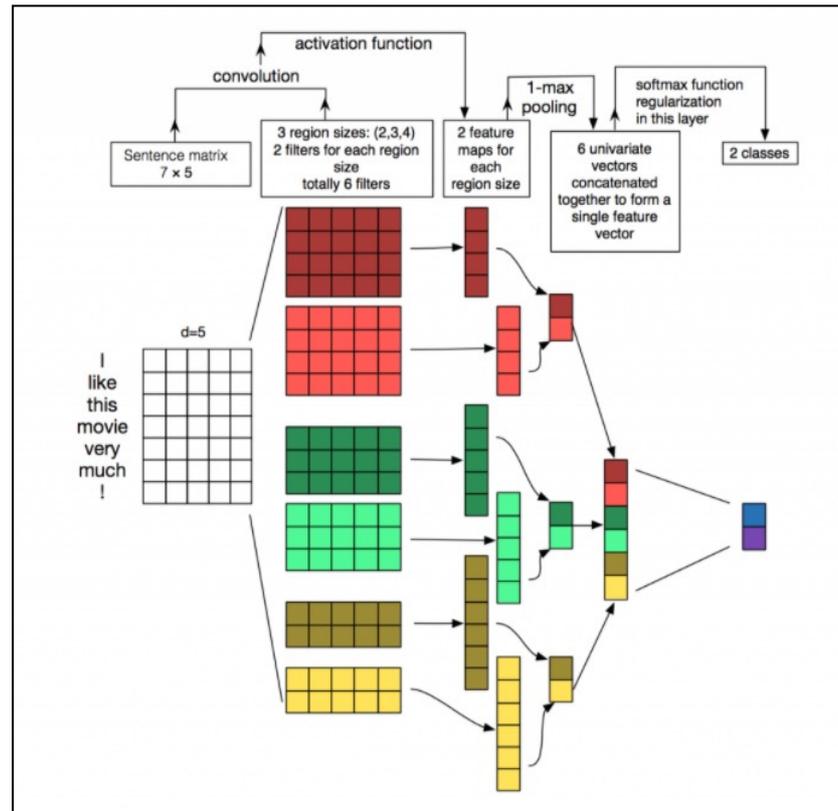
Figura 8- Algoritmo *Random Forest*



Traduzido e adaptado pela autora, 2020. Fonte: Pedregosa et al., 2011.

As abordagens de aprendizagem profunda, *Convolutional Neural Networks* (CNN), alcançaram resultados superiores em comparação com os anteriores algoritmos de aprendizado de máquina (GEORGAKOPOULOS; TASOULIS; VRAHATIS *et al.*, 2018) em tarefas como classificação de imagens, processamento de linguagem natural, reconhecimento facial, entre outros (JOHNSON; ZHANG, 2017).

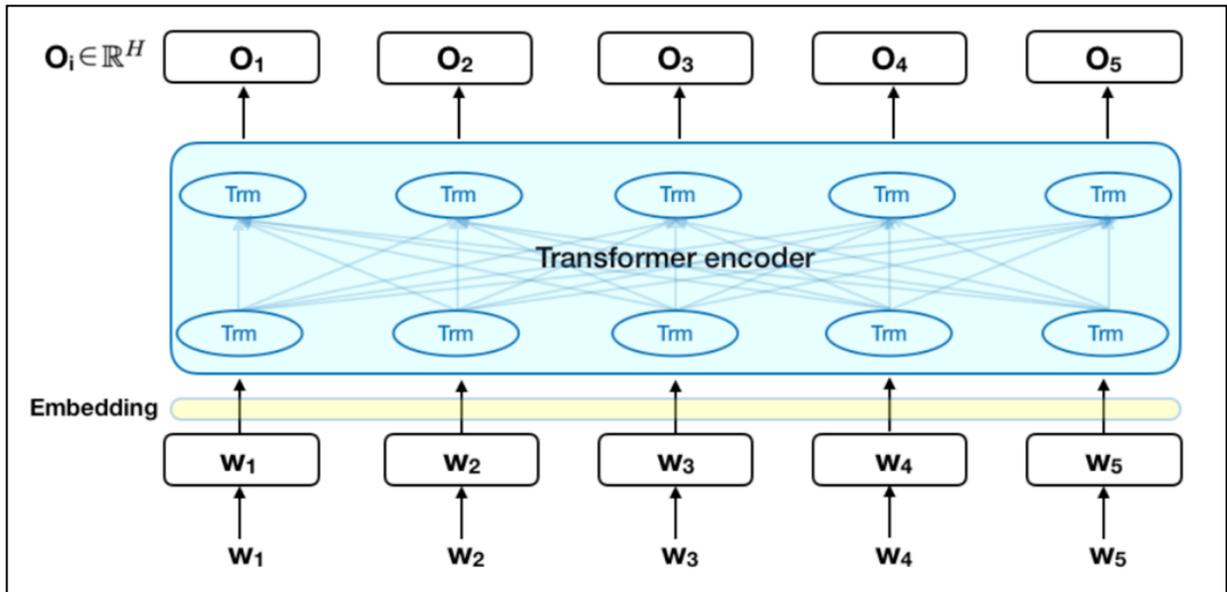
Figura 9- Figura clássica do algoritmo *Convolutional Neural Networks (CNN)*



Traduzido e adaptado pela autora, 2020. Fonte: Pedregosa et al., 2011.

Finalmente, o modelo BERT é método de aprendizado profundo de ponta-a-ponta pré-treinado com um corpus de 3,3 bilhões de palavras, incluindo *BooksCorpus* (800 milhões de palavras) e *Wikipedia* em inglês (2,5 bilhões de palavras). Dessa forma, ele prevê a falta de palavras em uma frase (**Figura 10**). BERT redefiniu o estado da arte para linguagem natural tarefa de processamento (CUNHA; MANGARAVITE; GOMES *et al.*, 2020). O sucesso desses algoritmos de aprendizado profundo depende de sua capacidade de modelar relacionamentos complexos e não lineares dentro dos dados (LECUN; BENGIO; HINTON, 2015).

Figura 10- Algoritmo BERT



Traduzido e adaptado pela autora, 2020. Fonte: Pedregosa et al., 2011.

2.5.3. Ontologias como modelo de classificação

Diferente dos processos descritos anteriormente que requerem a combinação de representadores de textos com algoritmos classificadores treinados, uma segunda opção para a classificação de textos livres é através da utilização de ontologias (ALMEIDA; BAX, 2003). Essa técnica tem mostrado eficácia de resultados equivalente aos modelos de classificação por aprendizado de máquina sem a necessidade de grande quantidade de dados e tempo de treinamento que comumente são requeridos por esses métodos (ALLAHYARI; KOCHUT; JANIK, 2014; CHI; LIN; HSIEH, 2014; GARLA; BRANDT, 2012; WANG; MCKAY; ABBASS *et al.*, 2003).

Ontologias são especificações, explícitas e formais, da conceitualização de um domínio de interesse, onde entende-se por formal uma especificação compreensível por um computador que com esse entendimento consiga realizar deduções sobre os conceitos (ALMEIDA; BAX, 2003; DAVIES; FENSEL; HARMELEN, 2003). Normalmente é incluso um vocabulário de termos e as especificações dos seus significados, e ainda as definições e as relações de como os conceitos estão inter-relacionados impondo uma estrutura sobre um domínio e limitando as possíveis interpretações dos termos (GOMEZ-PEREZ; CORCHO; FERNÁNDEZ-LÓPEZ, 2004).

Na ontologia, são explicitados conjuntos de termos que são relevantes para cada

variável e as regras de como estes se relacionam e se apresentam no texto de forma a categorizar determinada sentença como pertencente a uma determinada variável. Estes termos são elencados a partir do conhecimento de domínio e exemplos identificados nas evoluções. Esses termos são expandidos com uso de *word-embeddings* visando melhorar a abrangência desse conhecimento.

Com um algoritmo de busca textual é identificado os termos nas sentenças e definida uma classe na ontologia para essa sentença que contém todos os termos encontrados e as suas relações. Baseado nos termos e relações o processo de inferência é executado onde este deduz a qual variável a dada sentença pertence (LOPES; VIEIRA; FINATTO *et al.*, 2009).

2.6. AVC

As doenças cerebrovasculares estão entre as principais causas de morte no Brasil e no Mundo, apenas atrás das cardiovasculares e pesquisa sugerem que esse índice se mantenha até 2030 (MINISTÉRIO DA SAÚDE, 2018). O Acidente Vascular Cerebral (AVC) é uma das principais causas de mortalidade e invalidez de longo prazo em todo o mundo (JOHNSON; NGUYEN; ROTH *et al.*, 2019; OURIQUES MARTINS; SACKS; HACKE *et al.*, 2019). No Brasil, é um importante problema de saúde pública correspondendo a 107.658 mortes por ano, e um total de 80% dos pacientes com AVC são atendidos no sistema público de saúde (MINISTÉRIO DA SAÚDE, 2018). Ficando para trás apenas para doenças cardiovasculares, afecções do coração e circulação, que correspondem a mais de 1100 mortes por dia (MINISTÉRIO DA SAÚDE, 2018).

Uma iniciativa do Ministério de Saúde em 2008 em organizar serviços de urgência para melhorar a qualidade do atendimento a todos os pacientes fez com que a Rede Brasil AVC fosse criada (REDE BRASIL AVC, 2008). O projeto conta hoje com aproximadamente 160 hospitais de 19 Estados do Brasil. Através da Rede, profissionais e organizações são instruídos e apresentados ao que há de inovação e tecnologias aptas a serem incorporadas ao atendimento do AVC, permitindo tornar o atendimento nacional mais qualificado. Entre os avanços alcançados tem-se a implementação das Unidades de AVC que são áreas físicas preparadas e qualificadas para melhor atender aos pacientes com AVC. Elas também são caracterizadas pelo forte engajamento na educação continuada de pacientes, familiares e profissionais de saúde no âmbito da morbidade, e incentivo a pesquisa e desenvolvimento de tecnologias que possam auxiliar na prevenção e/ou tratamento de doença (REDE BRASIL

AVC, 2008).

Em 2013, pesquisadores brasileiros publicaram o manual de rotinas para a atenção do AVC com o objetivo de apresentar protocolos, escalas e orientações aos profissionais de saúde no manejo clínico ao paciente acometido por AVC (MINISTÉRIO DA SAÚDE, 2013a; b). Este manual tem sido adotado pelos hospitais da rede, permitindo o alcance de melhores resultados. Complementarmente a Rede AVC Brasil, o Programa *Angels* lançado em 2018, é uma iniciativa internacional da Boehringer Ingelheim que busca qualificar os centros de AVC já existentes e auxiliar na implementação de novos centros. O *Angels Award* classifica os hospitais como Prontos para o atendimento do AVC – quando tem a estrutura mínima e monitora os dados – e aqueles com melhores desempenhos são classificados em ouro, platina e diamante (REDE BRASIL AVC, 2008). Estes hospitais recebem certificados deste status e são reconhecidos nos grandes congressos nacionais e internacionais. Mais de 150 hospitais no Brasil participam e mais de 2800 hospitais no mundo.

Por ser uma doença tempo-dependente, esforços de redesenho da linha de cuidado AVC, que permitam uma avaliação mais precisa e rápida, são foco de pesquisa em várias partes do mundo (MICHAEL E. PORTER; JAMES S MOUNT FORD; KAMALINI RAMDAS *et al.*, 2012; OURIQUES MARTINS; SACKS; HACKE *et al.*, 2019). No âmbito de VBHC, a literatura mostra abordagens de atendimento mais personalizado para as necessidades de pacientes individuais, reduzindo custos e restrições de recursos (W. DAVID FREEMAN ; KEVIN M. BARRETT ; LISA NORDAN *et al.*, 2018). Recentemente, avanços em detectar características do AVC, como severidade pela escala do *National Institute of Health Stroke Scale* (NIHSS), com técnicas de aprendizado de máquina foram exploradas por Kogan *et al.* (2020) (KOGAN; TWYMAN; HEAP *et al.*, 2020), no entanto os mesmos autores reconhecem que a faltam de avaliação desses dados é uma limitação de evidência de mundo real dos pacientes AVC, e incentivam ao tecnologias que respondam o conhecimento estratificado de risco do AVC e outras condições médicas.

A capacidade de monitoramento e controle da adoção das orientações estabelecidas nos manuais e o alinhamento com modelos de reembolso são desafios ainda a serem explorados (VAN VEGHEL; SCHULZ; VAN STRATEN *et al.*, 2018). Uma recente revisão sistemática teve como objetivo identificar e avaliar criticamente o relato e o desenvolvimento de modelos de aprendizado de máquina para prever resultados após AVC (WANG; KIIK; PEEK *et al.*, 2020). Apesar de contemplar apenas dados estruturados, e não de texto, os autores reportaram que a maioria dos estudos abordam desfechos duros, como mortalidade e

rastreio da capacidade funcional pelo Rankin (mRS), e que a presença de dados desbalanceados (isto é, em que a variável de destino tem mais observações em uma classe específica do que as outras) penaliza o treinamento de algoritmos mais complexos. Ainda, os mesmos autores apontam que existe grande variabilidade entre os estudos no quesito tamanho amostral, e que nenhum dos estudos discutiu a implementação do modelo na vida real em prática clínica, embora o objetivo final seja, presumivelmente, ajudar os médicos a fazer decisões de tratamento e estimativa de prognósticos (WANG; KIIK; PEEK *et al.*, 2020). No mesmo estudo, os modelos de aprendizado de máquina mais utilizados, conforme relatado pelos autores Wang et al. (2020) (WANG; KIIK; PEEK *et al.*, 2020) foram *random forest*, *support vector machine* (SVM), árvore de decisão e redes neurais.

Neste sentido, pela relevância do AVC na população brasileira e das evidências sobre condutas adequadas que devem ser adotadas na fase aguda, a linha de cuidado do AVC isquêmico é objeto no estudo piloto deste projeto no que diz respeito a etapa de desenvolvimento de algoritmos para extração de desfechos em dados de texto livre a partir de prontuários eletrônicos.

3. OBJETIVOS

- Identificar na literatura quais medidas de desfechos são considerados em estudos que seguiram a agenda de valor, e analisar a origem dos dados usados para relatar a iniciativa de cuidado em saúde baseada em valor.
- Explorar modelos e regras que possibilitem a automatização da coleta de dados em prontuário eletrônico e seu potencial uso para analisar indicadores de desfechos de saúde.

4. REFERÊNCIAS BIBLIOGRÁFICAS

AGRAWAL, A.; GOLDFARB, A.; GANS, J. **Prediction Machines: The Simple Economics of Artificial Intelligence**. 2018.

ALLAHYARI, M.; KOCHUT, K. J.; JANIK, M. Ontology-based Text Classification into Dynamically Defined Topics. *In: IEEE International Conference on Semantic Computing*, 2014, p. 273-278. DOI: 10.1109/ICSC.2014.51.

ALMEIDA, M. B.; BAX, M. P. **Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção**. *Ciência da Informação*, 32, p. 7-20, 2003.

ANDRADE, C. M. d.; GONÇALVES, M. A. **Combining Representations For Effective Citation Classification**. Association for Computational Linguistics, 2020.

ATEFA SHAH; BLASZCZYNSKI, E. **Value-based healthcare: A global assessment**. 2016.

BALYAN, R.; CROSSLEY, S. A.; BROWN, W., 3rd; KARTER, A. J. *et al.* **Using natural language processing and machine learning to classify health literacy from secure messages: The ECLIPPSE study**. *PLoS One*, 14, n. 2, p. e0212488, 2019.

BARBA, P. Machine Learning for Natural Language Processing. Lexalytics. 2019. Disponível em: <https://www.lexalytics.com/lexablog/machine-learning-natural-language-processing>.

BEAM, A. L.; KOHANE, I. S. **Big Data and Machine Learning in Health Care**. *JAMA*, 319, n. 13, p. 1317-1318, 2018/04/03/ 2018.

BEAM, A. L.; KOHANE, I. S. **Big Data and Machine Learning in Health Care**. *JAMA*, 319, n. 13, p. 1317-1318, Apr 3 2018.

BELL, J.; KILIC, C.; PRABAKARAN, R.; WANG, Y. Y. *et al.* **Use of electronic health records in identifying drug and alcohol misuse among psychiatric in-patients**. *The Psychiatrist*, 37, n. 1, p. 15-20, 2013.

BLUMENTHAL, D.; TAVENNER, M. **The “Meaningful Use” Regulation for Electronic Health Records**. *New England Journal of Medicine*, 363, n. 6, p. 501-504, 2010/08/05/ 2010.

BOAVENTURA, E. M. **Metodologia da pesquisa: monografia, dissertação, tese**. São Paulo: Atlas, 2004. São Paulo: Atlas, 2004.

BREIMAN, L. **Random Forests**. *Machine Learning*, 45, n. 1, p. 5-32, 2001/10/01 2001.

BUTTE, A. J.; WEINSTEIN, D. A.; KOHANE, I. S. Enrolling patients into clinical trials faster using real time recruiting. *In: AMIA Annual Symp Proceedings*, 2000, p. 111-115.

CANUTO, S.; MARTINS, W. S.; COUTO, T.; GONÇALVES, M. A. Efficient and Scalable MetaFeature-based Document Classification using Massively Parallel Computing. *In: In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2015, p. 333--342.

CANUTO, S.; SALLES, T.; GONÇALVES, M. A.; ROCHA, L. *et al.* On Efficient Meta-Level Features for Effective Text Classification. *In: In Proceedings of the 23rd ACM International Conference on Information and Knowledge Management -- CIKM*, 2014, p. 1709--1718.

CAROPRESO, M. F.; MATWIN, S. Beyond the bag of words: A text representation for sentence selection. *In: In Conference of the Canadian Society for Computational Studies of Intelligence*, 2006, Germany. p. 324--335.

CAWLEY, G. C.; TALBOT, N. L. C. **On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation.** *Journal of Machine Learning Research*, p. 2079–2107, 2010.

CHEN, W.; XIE, X.; WANG, J.; PRADHAN, B. *et al.* **A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility.** *Catena*, 151, p. 147-160, 2017.

CHI, N.-W.; LIN, K.-Y.; HSIEH, S.-H. **Using ontology-based text classification to assist Job Hazard Analysis.** *Advanced Engineering Informatics*, 28, n. 4, p. 381-394, 2014.

CONSELHO FEDERAL DE MEDICINA. Resolução nº 1638 de 09 de Agosto de 2002. Define prontuário médico e torna obrigatória a criação da Comissão de Revisão de Prontuários nas instituições de saúde. Resolução nº 1638 de 09 de Agosto de 2002. MEDICINA, C. F. D. Brasília, DF: Diário Oficial da União: 184-185 p. 2002.

COOPER, R. **Target costing and value engineering.** Routledge, 2017. 1351411764.

CORTES, E. G.; WOLOSZYN, V.; BARONE, D. A. C., 2018, Cham. **When, Where, Who, What or Why? A Hybrid Model to Question Answering Systems.** Springer International Publishing. 136-146.

CRISTIANINI, N.; SHAWE-TAYLOR, J. **An introduction to Support Vector Machines: and other kernel-based learning methods.** Cambridge University Press, 2000. 0521780195.

CUNHA, W.; CANUTO, S.; VIEGAS, F.; SALLES, T. *et al.* **Extended pre-processing pipeline for text classification: On the role of meta-feature representations, sparsification and selective sampling.** *Information Processing & Management*, 57, n. 4, 2020.

CUNHA, W.; MANGARAVITE, V.; GOMES, C.; CANUTO, S. *e. et al.* **On the Cost-Effectiveness of Neural and Non-Neural Approaches and Representations for Text Classification: A Comprehensive Comparative Study (no prelo).** *Information Processing & Management*, 2020.

DAVIES, J.; FENSEL, D.; HARMELEN, F. **Towards the Semantic Web: Ontology-Driven Knowledge Management.** 03/01 2003.

DONABEDIAN, A. The quality of care. How can it be assessed? *In: Jama.* 1988/09/23 ed., 1988. v. 260, p. 1743-1748.

DONABEDIAN, A. **An Introduction to Quality Assurance in Health Care.** 2002. (Oxford University Press. 9780195158090.

DRIESSEN, S. R.; VAN ZWET, E. W.; HAAZEBROEK, P.; SANDBERG, E. M. *et al.* **A dynamic quality assessment tool for laparoscopic hysterectomy to measure surgical outcomes.** *Am J Obstet Gynecol*, 215, n. 6, p. 754 e751-754 e758, Dec 2016.

ETGES, A. P. B. d. S.; RUSCHEL, K. B.; POLANCZYK, C. A.; URMAN, R. D. **Advances in Value-Based Healthcare by the Application of Time-Driven Activity-Based Costing for Inpatient Management: A Systematic Review.** *Value in Health*, 23, n. 6, p. 812-823, 2020/06// 2020.

EY. *Dados Ernest Young.* 2019.

FALCONER, N.; NAND, S.; LIOW, D.; JACKSON, A. *et al.* **Development of an electronic patient prioritization tool for clinical pharmacist interventions.** *Am J Health Syst Pharm*, 71, n. 4, p. 311-320, Feb 15 2014.

GARLA, V. N.; BRANDT, C. **Ontology-guided feature engineering for clinical text classification.** *J Biomed Inform*, 45, n. 5, p. 992-998, Oct 2012.

GEORGAKOPOULOS, S. V.; TASOULIS, S. K.; VRAHATIS, A. G.; PLAGIANAKOS, V. P. Convolutional Neural Networks for Toxic Comment Classification. *In: Proceedings of the 10th Hellenic Conference on*

Artificial Intelligence, 2018, Patras, Greece. Association for Computing Machinery, p. Article 35. DOI: 10.1145/3200947.3208069. Disponível em: <https://doi.org/10.1145/3200947.3208069>.

GLASER, J. It's Time for a New Kind of Electronic Health Record. Harvard Business Review. ANALYTICS: Harvard Business Print 2020.

GOLAS, S. B.; SHIBAHARA, T.; AGBOOLA, S.; OTAKI, H. *et al.* **A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data.** BMC Medical Informatics and Decision Making, 18, n. 1, p. 44, 2018/06/22 2018.

GOLDBERG, Y.; LEVY, O. **Word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method.** arXiv, 2014.

GOMEZ-PEREZ, A.; CORCHO, O.; FERNÁNDEZ-LÓPEZ, M. **Ontological engineering: with examples from the areas of knowledge management, e-Commerce and the Semantic Web. (advanced information and knowledge processing).** 01/01 2004.

GOPAL, G.; SUTER-CRAZZOLARA, C.; TOLDO, L.; EBERHARDT, W. **Digital transformation in healthcare - Architectures of present and future information technologies.** Clinical Chemistry and Laboratory Medicine (CCLM), 57, 12/11 2018.

GROSSMAN, E.; CARDOSO, M. H. C. d. A. **As narrativas em medicina: contribuições à prática clínica e ao ensino médico.** Rev. bras. educ. med., 30, n. 1, p. 6-14, 2006.

HAN, E. H. S.; KARYPIS, G. Centroid-based document classification: Analysis and experimental results. . *In: In European Conference on Principles of Data Mining and Knowledge Discovery; Springer: Berlin/Heidelberg,, 2000, Berlin/Heidelberg, Germany. p. 424-431.*

HERNANDEZ A.; KAPLAN R.S.; WITKOWSKI M.L.; FORREST FAISON C. *et al.* **Navy Medicine Introduces value-based health care.** Health affairs (Project Hope), 8, 38(8), n. 8, p. 1393-1400, 2019.

HONNIBAL, M.; MONTANI, I. **spaCy.** Explosion AI, 2015. spaCy is an open-source software library for advanced natural language processing, written in the programming languages Python and Cython.

HYUN, S.; JOHNSON, S., B.; BAKKEN, S. **Exploring the ability of natural language processing to extract data from nursing narratives.** . Comput Inform Nurs, 27, n. 4, p. 215-225, 2009.

ICHOM; 2012. **ICHOM 2012 | Healthcare Improvement | Patient-Reported Outcomes.** Disponível em: <https://www.ichom.org/files/28502/www.ichom.org.html>.

IORIO, R.; CLAIR, A. J.; INNEH, I. A.; SLOVER, J. D. *et al.* **Early Results of Medicare's Bundled Payment Initiative for a 90-Day Total Joint Arthroplasty Episode of Care.** The Journal of Arthroplasty, 31, n. 2, p. 343-350, 2016/02// 2016.

JAGANNATHA, A.; LIU, F.; LIU, W.; YU, H. **Overview of the First Natural Language Processing Challenge for Extracting Medication, Indication, and Adverse Drug Events from Electronic Health Record Notes (MADE 1.0).** Drug Saf, 42, n. 1, p. 99-111, Jan 2019.

JAMES, G.; WITTEN, D.; TIBSHIRANI, T. H. R. **An Introduction to Statistical Learning.** New York Heidelberg Dordrecht London: Springer, 2013. 181-184 p.

JOHNSON, C. O.; NGUYEN, M.; ROTH, G. A.; NICHOLS, E. *et al.* **Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016.** The Lancet Neurology, 18, n. 5, p. 439-458, 2019.

JOHNSON, R.; ZHANG, T. Deep Pyramid Convolutional Neural Networks for Text Categorization. *In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017, Vancouver, Canada. p. 567-570.*

KATZ, M.; FRANKEN, M.; MAKDISSE, M. **Value-Based Health Care in Latin America: An Urgent Discussion.** *J Am Coll Cardiol*, 70, n. 7, p. 904-906, Aug 15 2017.

KLIE, J.-C.; BUGERT, M.; BOULLOSA, B.; ECKART DE CASTILHO, R. *et al.* **The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation.** In Proceedings of System Demonstrations of the 27th International Conference on Computational Linguistics (COLING 2018): Santa Fe, New Mexico, USA 2018.

KNOX; RITA; EID, T.; WHITE, A. **“Management Update: Companies should align their structured and unstructured dat.** Gartner Research, 2005.

KOGAN, E.; TWYMAN, K.; HEAP, J.; MILENTIJEVIC, D. *et al.* **Assessing stroke severity using electronic health record data: a machine learning approach.** *BMC Med Inform Decis Mak*, 20, n. 1, p. 8, Jan 8 2020.

KOWSARI; JAFARI, M.; HEIDARYSAFA; MENDU *et al.* **Text Classification Algorithms: A Survey.** *Information*, 10, n. 4, 2019.

KREIMEYER, K.; FOSTER, M.; PANDEY, A.; ARYA, N. *et al.* **Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review.** *J Biomed Inform*, 73, p. 14-29, Sep 2017.

KUHN, M.; JOHNSON, K. **Applied Predictive Modeling.** Springer, 2013. 978-1-4614-6848-6.

LARSON, R. R. **Introduction to Information Retrieval.** *Journal of the American Society for Information Science and Technology*, p. n/a-n/a, 2009.

LAYEGHIAN JAVAN, S.; SEPEHRI, M. M.; AGHAJANI, H. **Toward analyzing and synthesizing previous research in early prediction of cardiac arrest using machine learning based on a multi-layered integrative framework.** *J Biomed Inform*, 88, p. 70-89, Dec 2018.

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning.** *Nature*, 521, n. 7553, p. 436-444, May 28 2015.

LEE, T. H. **Putting the Value Framework to Work.** *New England Journal of Medicine*, 363, n. 26, p. 2481-2483, 2010/12/23/ 2010.

LEUNG, T. I.; VAN MERODE, G. G. Value-Based Health Care Supported by Data Science. *In: KUBBEN, P.; DUMONTIER, M., et al (Ed.). Fundamentals of Clinical Data Science.* Cham (CH): Springer Copyright 2019, The Author(s). 2019. p. 193-212.

LI, L.; WEINBERG, C. R.; DARDEN, T. A.; PEDERSEN, L. G. **Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method.** *Bioinformatics*, 17, n. 12, p. 1131-1142, Dec 2001.

LING, A. Y.; KURIAN, A. W.; CASWELL-JIN, J. L.; SLEDGE, G. W., Jr. *et al.* **Using natural language processing to construct a metastatic breast cancer cohort from linked cancer registry and electronic medical records data.** *JAMIA Open*, 2, n. 4, p. 528-537, Dec 2019.

LOPES, L.; VIEIRA, R.; FINATTO, M. J.; MARTINS, D. *et al.* **Automatic extraction of composite terms for construction of ontologies: an experiment in the health care area.** *Reciis*, 3, n. 1, 2009.

LYNCH, S. M.; MOORE, J. H. **A call for biological data mining approaches in epidemiology.** *BioData Mining*, 9, n. 1, p. 1, 2016/01/04 2016.

Machine Learning for Natural Language Processing. Lexalytics, 2019/11/25/T07:00:53+00:00 2019.

MAKDISSE, M.; KATZ, M.; RAMOS, P.; PEREIRA, A. *et al.* **What Is a Value Management Office? An Implementation Experience in Latin America.** *Value in Health Regional Issues*, 17, n. (Makdisse M., marcia.makdisse@einstein.br; Katz M.; Ramos P.; Pereira A.; Shiramizo S.; Neto M.C.; Klajner S.) Hospital

Israelita Albert Einstein, São Paulo, Brazil, p. 71-73, 2018 2018.

MANEVITZ, L. M.; YOUSEF, M. **One-class SVMs for document classification**. *J. Mach. Learn*, 2, p. 139-154, 2001.

MANNING, C.; RAGHAWAN, P.; SCHUTZE, H. **An introduction to information retrieval**. Cambridge, U.K.: Cambridge University Press, 2009.

MANNING, C.; SCHUTZE, H. **Foundations of statistical natural language processing**. MIT Press: Cambridge, MA., 1999.

MESKÓ, B.; GÖRÖG, M. **A short guide for medical professionals in the era of artificial intelligence**. *npj Digital Medicine*, 3, n. 1, 2020.

MESKÓ, B.; RADÓ, N. **Data annotators: the unsung heroes of artificial intelligence development.**, 2019. Disponível em: <https://medicalfuturist.com/data-annotation/>.

MEULEN, M. v. d. **Artificial Intelligence as a Driver of Value in Value Based Health Care Systems**. 2019.

MEYSTRE, S. M.; KIM, Y.; GOBBEL, G. T.; MATHENY, M. E. *et al.* **Congestive heart failure information extraction framework for automated treatment performance measures assessment**. *J Am Med Inform Assoc*, 24, n. e1, p. e40-e46, Apr 1 2017.

MICHAEL E. PORTER; JAMES S MOUNT FORD; KAMALINI RAMDAS; TAKVORIAN, S. **Reconfiguring Stroke Care in North Central London**. 2012.

MICHAEL E. PORTER; KAPLAN, R. S. How to pay for Health Care. *Havard Bussiness Review*. : HBR Reprint 2016.

MINISTÉRIO DA SAÚDE. Manual de Rotinas Para Atenção do AVC. 2013a.

MINISTÉRIO DA SAÚDE. Protocolo Clínico e Diretrizes Terapêuticas: Trombólise No Acidente Vascular Cerebral Isquêmico Agudo. 2013b.

MINISTÉRIO DA SAÚDE. Uma Análise de Situação de Saúde e Das Doenças e Agravos Crônicos: Desafios e Perspectiva. MINISTÉRIO DA SAÚDE, B. Brasília: MS 2018.

MIOTTO, R.; WENG, C. **Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials**. *J Am Inform Assoc*, 1, 22, p. 141-150, 2015.

NI, Y.; WRIGHT, J.; PERENTESIS, J.; LINGREN, T. *et al.* **Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients**. *BMC Med Inform Decis Mak*, 15, p. 28, Apr 14 2015.

NORIA, S.; NEEDLEMAN, B.; MIKAMI, D.; LAYTON, K. *et al.* **The impact of care coaching on hospital length of stay, readmission rates and overall patient satisfaction post-bariatric surgery**. *Surgery for Obesity and Related Diseases*, 11, n. 6, p. S8-S9, 2015 2015.

OURIQUES MARTINS, S. C.; SACKS, C.; HACKE, W.; BRAININ, M. *et al.* **Priorities to reduce the burden of stroke in Latin American countries**. *The Lancet Neurology*, 18, n. 7, p. 674-683, 2019.

PBGH. **Variation in NTSV C-section rates among California hospitals**. Pacific Business Group on Health. LeapFrog Group. 2014.

PEDREGOSA; ET AL. **Scikit-learn: Machine Learning in Python.**, 2011. pp. 2825-2830.

PENNINGTON, J.; SOCHER, R.; MANNING, C. D.; DOHA, Q. Glove: Global Vectors for Word Representation. *In: In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, **14**. p. 1532–1543.

PESTIAN, J.; NASRALLAH, H.; MATYKIEWICZ, P.; BENNETT, A. *et al.* **Suicide Note Classification Using Natural Language Processing: A Content Analysis.** Biomedical Informatics Insights, 3, 2010.

PORTER, M. E. **What Is Value in Health Care?** New England Journal of Medicine, 363, n. 26, p. 2477-2481, 2010/12/23/ 2010.

PORTER, M. E., Alee, H., Witkowski, M.L. and Feeley, Thomas W. . **“Navy Medicine: Moving to Value-Based Care”** Harvard Business School,. 2019.

PORTER, M. E., and Elizabeth O. Teisberg. **"Cleveland Clinic: Transformation and Growth 2015."** Harvard Business School Case: February 2009, p. 709-473. 2009.

PORTER, M. E., Jens Deerberg-Wittram, and Clifford Marks. **"Martini Klinik: Prostate Cancer Care."** Harvard Business School Case Collection, p. 714-471. 2014.

PORTER, M. E.; LARSSON, S.; LEE, T. H. **Standardizing Patient Outcomes Measurement.** N Engl J Med, 374, n. 6, p. 504-506, Feb 11 2016.

PORTER, M. E.; LEE, T. H. **The Strategy That Will Fix Health Care.** Harvard Business Review, 91, n. 10, p. 50-70, 2013/10// 2013.

PORTER, M. E.; TEISBERG, E. O. **Redefining health care: creating value-based competition on results.** Boston, Mass: Harvard Business School Press, 2006 2006. 506 p. 978-1-59139-778-6.

PORTER, M. E., Thomas H. Lee, and Meredith A. Alger. **"Oak Street Health: A New Model of Primary Care."** Harvard Business School Case : February 2017, p. 717-437. 2017.

RAGHAVAN, P.; CHEN, J. L.; LUSSIER, E. F.; LAI, A. M. How essential are unstructured clinical narratives and information fusion to clinical trial recruitment? . *In: AMIA Jt Summits Transl Sci Proc.*, 2014, p. 218-223. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4333685/pdf/1861388.pdf>>.

REDE BRASIL AVC. **RedeBrasilAVC.** 2008. Disponível em: <http://www.redebrasilavc.org.br>.

ROSENBLOOM, S. T.; DENNY, J. C.; XU, H.; LORENZI, N. *et al.* **Data from clinical notes: a perspective on the tension between structure and flexible documentation.** Journal of the American Medical Informatics Association, 18, n. 2, p. 181-186, 2011.

RUIZ, E. E. S. **Mineração de texto em saúde.** 2016.

RUSSELL, S. J. **Artificial intelligence : a modern approach.** New Jersey: 1995. (Artificial intelligence. 013103805-2.

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach.** Englewood Cliffs, New Jersey: Alan Apt, 2016.

S., C.; X., S. D.; A., G. M.; C., R. T. **A Thorough Evaluation of Distance-Based Meta-Features for Automated Text Classification.** IEEE Transactions on Knowledge and Data Engineering, 30, n. 12, p. 2242-2256, 2018.

SALTON, G.; BUCKLEY, C. **Term-weighting approaches in automatic text retrieval.** . Inf. Process. Manag, 24, p. 513-523, 1988.

SHEIKHALISHAHI, S.; MIOTTO, R.; DUDLEY, J. T.; LAVELLI, A. *et al.* **Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review.** JMIR Med Inform, 7, n. 2, p. e12239, Apr 27 2019.

SHEN, D.; SUN, J. T.; LI, H.; YANG, Q. *et al.* **Document Summarization Using Conditional Random Fields.** 7, p. 2862-2867, 2007.

- SHEN, Y. **Selection incentives in a performance-based contracting system.** Health Serv Res, 38, n. 2, p. 535-552, Apr 2003.
- SHIVADE, C.; HEBERT, C.; LOPETEGUI, M.; MARNEFFE *et al.* **Textual inference for eligibility criteria resolution in clinical trials.** J Biomed Inform, S211-S218, 58, 2015.
- SILVA, A. M. P. d. Dissertação de Mestrado: Processamento de Linguagem Natural na Identificação de Critérios de Elegibilidade para a Pesquisa Clínica SAÚDE, P. d. P. G. d. T. e. : Pontifca Universidade Católica do Paraná: Escola Politécnica 2018.
- SUN, W.; CAI, Z.; LI, Y.; LIU, F. *et al.* **Data Processing and Text Mining Technologies on Electronic Medical Records: A Review.** J Healthc Eng, 2018, p. 4302425, 2018.
- THAKER, N. G.; PUGH, T. J.; MAHMOOD, U.; CHOI, S. *et al.* **Defining the value framework for prostate brachytherapy using patient-centered outcome metrics and time-driven activity-based costing.** Brachytherapy, 15, n. 3, p. 274-282, 2016/05// 2016.
- TRANFIELD, D.; DENYER, D.; SMART, P. **Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review.** British Journal of Management, 14, n. 3, p. 207-222, 2004.
- TSAI, M. H.; PORTER, J. C.; ADAMS, D. C. **The Denominator in Value-Based Health Care: Porter's Hidden Costs.** Anesth Analg, 127, n. 1, p. 317, Jul 2018.
- UZUNER, Ö. e. a. **Challenge on concepts, assertions, and relations in clinical text.** Journal of the American Medical Informatics Association, 18, n. 5, p. 552-556, 2011.
- VALUCK, T.; BLAISDELL, D.; DUGAN, D. P.; WESTRICH, K. *et al.* **Improving Oncology Quality Measurement in Accountable Care: Filling Gaps with Cross-Cutting Measures.** J Manag Care Spec Pharm, 23, n. 2, p. 174-181, Feb 2017.
- VAN CALSTER, B.; VICKERS, A. J. **Calibration of risk prediction models: impact on decision-analytic performance.** Med Decis Making, 35, n. 2, p. 162-169, Feb 2015.
- VAN DEN BERG, K.; DIJKSMAN, L. M.; KEUS, S. H. J.; SCHEELE, F. *et al.* **Value-based health care in obstetrics.** Journal of Evaluation in Clinical Practice, 26, n. 1, p. 101-108, 2020 2020.
- VAN LEIJEN-ZEELLENBERG, J. E.; ELISSEN, A. M. J.; GRUBE, K.; VAN RAAK, A. J. A. *et al.* **The impact of redesigning care processes on quality of care: a systematic review.** BMC Health Serv Res, 16, p. 19, Jan 19 2016.
- VAN VEGHEL, D.; SCHULZ, D. N.; VAN STRATEN, A. H. M.; SIMMERS, T. A. *et al.* **Health insurance outcome-based purchasing: The case of hospital contracting for cardiac interventions in the Netherlands.** International Journal of Healthcare Management, 11, n. 4, p. 371-378, 2018.
- VANWINCKELEN, G. On Estimating Model Accuracy with Repeated Cross-Validation. *In:* Belgian-Dutch Conference on Machine Learning (BeneLearn), 2012, Ghent.
- VELUPILLAI, S.; SUOMINEN, H.; LIAKATA, M.; ROBERTS, A. *et al.* **Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances.** J Biomed Inform, 88, p. 11-19, Dec 2018.
- VIEIRA, R.; LIMA, V. L. S. d. Linguística computacional: princípios e aplicações. IX Escola de Informática da SBC-Sul.: SBC-Sul 2001.
- W. DAVID FREEMAN ; KEVIN M. BARRETT ; LISA NORDAN ; AARON C. SPAULDING *et al.* **Lessons from Mayo Clinic's Redesign of Stroke Care.** Harvard Business Review, n. Operations, 2018.
- WANG, B.; MCKAY, R.; ABBASS, H.; BARLOW, M. **A Comparative Study for Domain Ontology Guided**

Feature Extraction. 2003. 69-78 p.

WANG, S. Y.; PERSHING, S.; TRAN, E.; HERNANDEZ-BOUSSARD, T. **Automated extraction of ophthalmic surgery outcomes from the electronic health record.** International Journal of Medical Informatics, 133, p. 104007, 2020/01// 2020.

WANG, S. Y.; PERSHING, S.; TRAN, E.; HERNANDEZ-BOUSSARD, T. **Automated extraction of ophthalmic surgery outcomes from the electronic health record.** Int J Med Inform, 133, p. 104007, Jan 2020.

WANG, W.; KIIK, M.; PEEK, N.; CURCIN, V. *et al.* **A systematic review of machine learning models for predicting outcomes of stroke with structured data.** PLoS One, 15, n. 6, p. e0234722, 2020.

WANG, Y.; WANG, L.; RASTEGAR-MOJARAD, M.; MOON, S. *et al.* **Clinical information extraction applications: A literature review.** J Biomed Inform, 77, p. 34-49, Jan 2018.

XIAO, C.; CHOI, E.; SUN, J. **Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review.** Journal of the American Medical Informatics Association, 25, n. 10, p. 1419-1428, 2018/10/01/ 2018.

XIAO, C.; CHOI, E.; SUN, J. **Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review.** J Am Med Inform Assoc, 25, n. 10, p. 1419-1428, Oct 1 2018.

XU, B. G., X.; YE, Y.; CHENG, J. **An Improved Random Forest Classifier for Text Categorization.** . JCP 2012, 7, p. 2913-2920.

YANG, Y.; GOPAL, S. **Multilabel classification with meta-level features in a learning-to-rank framework.** JMLR, 88, p. 47-68, 2012.

YE, F., 2016, **Simultaneous Support Vector selection and parameter optimization using Support Vector Machines for sentiment classification.** 59-62.

YIN, R. K. **Case study research: design and methods.** Los Angeles London New Delhi Singapore Washington, DC: SAGE, 2014.

ZHANG, C. **Automatic keyword extraction from documents using conditional random fields.** J. Comput. Inf. Syst, 4, p. 1169-1180, 2008.

5. ARTIGO(S)

Os artigos a seguir estão no formato e normas da revista as quais serão submetidos.

5.1. ARTIGO 1

Value-based healthcare initiatives in the real-world setting: a systematic review**Running title:** Value-based healthcare in the real-world setting

Bruna Stella Zanotto^{1,2}, Ana Paula Beck da Silva Etges^{1,3}, Miriam Allein Zago Marcolino^{1,2}, Carisi Anne Polanczyk^{1,2,4}

1 – National Institute of Health Technology Assessment - INCT/IATS (CNPQ 465518/2014-1), Universidade Federal do Rio Grande do Sul - Porto Alegre (RS), Brasil.

2- Graduate Program in Epidemiology, Universidade Federal do Rio Grande do Sul, Porto Alegre (RS), Brazil

3- Pontificia Universidade Catolica do Rio Grande do Sul, Escola Politécnica, Porto Alegre, BR 90619-900

4- Graduate Program in Cardiology and Cardiovascular Sciences, Universidade Federal do Rio Grande do Sul, Porto Alegre (RS), Brazil

Corresponding author: Bruna Stella Zanotto, National Institute of Science and Technology for Health Technology Assessment (IATS)/ CNPq, 2350 Ramiro Barcelos, room 21507, Porto Alegre, RS 90035-903, Brazil

brunaszanotto@gmail.com

Word count: 3678 / Tables: 1 / Figures: 3 / Appendix Tables: 4

Funding: This study was supported by the National Institute of Science and Technology for Health Technology Assessment (IATS/INCT, CNPq) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Submetido para Journal of Healthcare Management em 30/09/2020

RESUMO

Iniciativas baseadas em valor estão crescendo como modelos estratégicos de gestão de saúde trazendo consigo a necessidade de explorar de forma mais aprofundada as medidas de resultados da equação de valor. Esta revisão sistemática teve como objetivo identificar quais medidas estão sendo utilizadas em estudos que seguiram a Agenda de Valor proposta por Porter.

Múltiplas bases de dados eletrônicas (PubMed / MEDLINE, Embase, Scopus e Cochrane Central) foram pesquisadas até 04 de março de 2020 (CRD42020156365). Os estudos elegíveis relataram implementações de iniciativas de saúde baseadas em valor. Uma abordagem qualitativa foi usada para analisar as medidas de resultados nos estudos. Os resultados foram classificados de acordo com a hierarquia de resultados de nível correspondente. Em um gráfico de radar, comparamos a literatura com a coleção de casos da Harvard Business School. O efeito-valor relatado foi descrito como impactos em cada domínio da equação de valor.

Um total de 7.195 registros foi recuperado e 47 estudos foram incluídos. Quarenta estudos utilizaram prontuários eletrônicos de saúde como origem dos dados. Apenas 16 estudos usaram pesquisas de resultados relatados por pacientes para cobrir as camadas de resultados, e três estudos relataram resultados para todos os seis níveis da Hierarquia de Resultados, sendo diferentes em tais aspectos dos casos avaliados da Harvard Business School. Uma parte considerável dos estudos (36%) alcançou resultados que contribuíram para resultados financeiros baseados em valor com foco na redução de custos. No entanto, essas oportunidades não são medidas por custos precisos e métodos econômicos.

Embora as iniciativas de valor tenham efeitos potencialmente positivos em um cenário do mundo real, ainda há uma lacuna na medição de resultados que são importantes para os pacientes. Transformar a agenda de valores mais inteiramente aplicada pelas organizações de saúde exige avanços na capacidade de tecnologia e gestão de mudança de cultura.

ABSTRACT

Objectives: Value-based initiatives are growing as strategic models of healthcare management, urging the need to explore more in-depth outcome measures of the value equation. This systematic review aimed identifying which measures are being used in studies that followed the Value Agenda in the real-world.

Methods: Multiple electronic databases (PubMed/MEDLINE, Embase, Scopus, and Cochrane Central) were searched up to March 04, 2020 (CRD42020156365). Eligible studies reported implementations of value-based healthcare initiatives. A qualitative approach was used to analyze outcome measurements in the studies. Outcome were classified according to the corresponding tier-level Outcome Hierarchy. In a radar chart, we compared literature with Harvard Business School Case Collection. The value-effect reported was described as impacts on each domain of the value equation.

Results: A total of 7195 records were retrieved, and 47 studies were included. Forty studies used electronic health records as data origin. Only 16 studies used patient-reported outcome surveys to cover outcome tiers, and three studies reported outcomes to all six levels of the Outcome Hierarchy, being different in such aspects from Harvard Business School's cases evaluated. A considerable part of the studies (36%) achieved results that contributed to value-based financial outcomes focused on cost-saving. However, those opportunities are not measured by accurate costs and economical methods.

Conclusions: Even though value initiatives have potentially positive effects on a real-world setting, remain a gap in measuring outcomes that matter to patients. Turning the value agenda more entirely applied by the health organizations require advances on technology capacity and culture change management.

INTRODUCTION

Historically, the quality of health services delivered to patients is not managed in synchrony with observed patient outcomes (Hernandez A. et al.,2019). Health organizations usually do not connect general company management to patient requirements, and the movement to pivot to healthcare management centered on the patient was a premise introduced by value-based healthcare (VBHC). Porter & Teisberg (2006) defined VBHC as an innovative way to rebuild the healthcare system. Measuring outcomes and costs for each patient is part of the strategic agenda for moving to a high-value healthcare delivery system comprised of six interdependent components (Porter & Lee,2013). The applied cases reported by the Harvard Business School (HBS) – where the VBHC concept originated – deserve highlighting within the application and investigation of value-based management models. Many other institutions are currently adopting components of VBHC in clinical practice. Unfortunately, rigorous scientific reports on the outcomes observed with these approaches are presently lacking (Van Deen et al.,2017).

The VBHC suggests that the health system needs to be managed in terms of the outcomes that matters to patients (Porter,2010). However, it is still a recurring practice to measure performance through generalized outcomes such as overall hospital mortality, infection rates, and medication errors. Although those measures represent an important role in institutional sustainability, overall clinical outcomes or hospital measures are not enough to capture dimensions highly crucial to the patient. It stands out that all conceptions of value-based healthcare and implementing such a management format replace the central role of patient-reported outcomes (PROs) in the equation (Porter & Lee,2013;Tseng & Hicks,2016).

In an attempt to translate the VBHC theory into health system operations, Porter idealized the outcome hierarchy that goes against the need to establish a consensus on what constitutes an outcome and presents domains to cover all phases of the cycle of care. The Outcome Measures Hierarchy recognizes that the definition of success for any medical condition may have a set of outcomes invariably broad, but yet it must comprise a three-tiered hierarchy: the degree of health or recovery, the process of recovery and sustainability of health (Porter,2010).

We are currently well appointed in terms of the existence of metrics and scales to measure outcomes (generic or tailored to disease classes); however, the selection of standard and tested measures will improve validity and enable comparisons across providers

(Porter,2010;Tsai et al.,2018;Van Der Wees et al.,2014). The great barrier for the successful implementation of outcome measurement in VBHC initiatives is that it requires the strategic engagement of healthcare managers, data collection and analysis of technological advances (Tsai et al.,2018).

A further question that remains in the VBHC area is about the effect and feasibility of following the value agenda. The need more in-depth exploration of the analysis of value-based initiatives in terms of outcome measurement can begin for a subset of medical conditions and expand over time as infrastructure and experience grow (Porter,2010). Considering the increased interest concerning this topic and the amount of recently published material, this systematic review aimed to identify which outcomes were considered to measure value in studies that followed the value agenda and to analyze the origin of the data used to report a value initiative.

METHODS

This systematic review was described according to the Preferred Reporting Items for Systematic Review and Meta-Analysis (Moher et al.,2009) statement and is consistent with the methods of systematic review proposed by the Cochrane Collaboration (Chalmers et al.,2018). The review protocol was registered prospectively on PROSPERO (CRD42020156365).

Patient and Public Involvement

Due to the qualitative and exploratory nature of this systematic review, patients were not involved in this study.

Literature Search Strategy

The MEDLINE (via PubMed), Embase, Scopus, and Cochrane electronic databases were searched for studies indexed until 03/04/2020. A manual search of specific journals and the reference lists of the retrieved articles was also performed. The search strategy combined indexed words and free terms (variations and synonyms) related to the search question and were substantially terms for VBHC (see table, Supplemental Digital Content 1, which demonstrates the full strategy). The search terms were adjusted to fit the requirements of each electronic database. No restriction for the date of publication or language was adopted. The results of these database searches were cross-checked to locate and eliminate duplicate entries.

Eligibility Criteria and Study Selection

Two reviewers were responsible for the independent screening of all titles and

abstracts identified by the electronic search. Potentially eligible and uncertain studies were retrieved for full-text assessments. If a disagreement arose or a consensus was not reached, a third reviewer made the final decision. The included studies were those with an application of the VBHC initiative as defined by Porter in “The Value Agenda” (Porter & Lee,2013). Only studies available in English, Spanish,or Portuguese were considered. The exclusion criteria applied were specific cost analysis studies, studies of the effectiveness of drugs or diagnostic tests,and studies from an insurance perspective or those that did not evaluate outcomes at the patient level. Editorials, commentaries, or views were only considered if the publication presented results from an application case of VBHC.

Data Extraction Process

Data collection was performed independently by the two reviewers; if uncertainty persisted, a third reviewer was contacted to support the decision. The data extraction started with the general characteristics of the studies: year of publication, setting, healthcare field, value initiative, cost measurement methodology (if applied) and to attend our objectives, we extracted information on which outcomes the study collected, the origin of the data to evaluate these outcomes and whether any outcome instrument was used as a collection tool. All data were consolidated with Microsoft Office© Excel 2010 software (Microsoft Corporation, Redmond, WA, USA).

To classify outcomes used by the authors to report a value result, we categorized data into the three-tiered hierarchy defined previously (Porter,2010). Each tier of the hierarchy contains two broad levels, as presented in Figure 1.

Patient’s initial conditions, demographics and disease-related factors are fundamental to evaluate patient outcomes adjusted to their risk (Porter,2010). Therefore, we also assessed whether baseline characteristics were a variable considered in the studies’ methods.

Data Analyses

According to the studies’ initial purpose and the elements of the value agenda, value-added initiatives were distinguished by three classes: (i)performing clinical or surgical pathway redesign, (ii)developing telemedicine or computational intelligence programs, (iii)and measuring clinical, process, and financial outcomes (traditional VBHC program). Clinical or surgical pathway redesign brings up the effort to standardize care, reorganize structure to improve access and efficiency, which is strongly related to the Value Agenda components of integrated practice units (IPUs), and bundle payment for care cycle (Porter & Lee,2013). The second class, telemedicine or computational intelligence supporting value

programs, comprise the information technology element. It proposes a computation solution to advance in value-based implementations, either using artificial intelligence to compose the numerator of the value equation or telemedicine programs to optimize care and access. The last class, the traditional value program, consists of studies practicing the first premise of value, the organization of the care pathway as a function of each patient's clinical condition, and the capability of measuring outcome and cost for each patient.

The country of the study, year of publication, healthcare field, and setting were also assessed. The setting was defined as 'system' when the study covered a multicentric or national perspective and as 'hospital' whether the scenario was at the provider or institution level.

For each article, outcome information was retrieved and classified according to the corresponding tier-level it was related to so that we could map what outcome driver of each tier was more frequent in the studies. We also assessed what the data source of the outcome was, and whether any outcome measurement instruments were used. Moreover, the saturation of tier-level outcome reporting was determined by counting on how many of the six tier-level hierarchies the outcomes of each study could be mapped.

By exploring the value equation and understanding the measure of the value as a difference in outcomes and costs before and after initiating, the value effect reported in the studies was qualitatively described as actions that impact some domain of the value equation. We classified these effects into the following categories: financial outcomes, clinical outcome improvements, PRO improvement, providers' education, and value culture and management.

Furthermore, we used a radar chart depicting the metrics of outcomes, baseline characteristics, and costs to illustrate the balance of outcome measurements within the literature. For this last analysis, to recognize gaps and opportunities in the evolution of VBHC studies and the comprehensive definition of value, we retrieved classics and well-disseminated VBHC cases from the HBS Case Collection (Harvard Business School Publishing,2020). These cases served as a standard reference for the selected studies in the systematic review, using the same eligibility criteria described above. We understand that the HBS cases' primary purpose was not to serve as a definitive research study or academic paper. However, we still believe that these cases could be a source of knowledge and value in describing benchmark initiatives.

RESULTS

Study Selection

The literature search found 7195 registries, of which 105 full-text articles were assessed, and 47 studies ultimately fulfilled the inclusion criteria for the review. Figure 2 shows the flow diagram of the entire review process.

Study Characteristics

The publication years ranged from 2010 to early 2020, with 2019 being the year with most publications (n=18). The majority of selected studies were performed in the United States (n=39), and others were from the United Kingdom (n=3), the Netherlands (n=3), Brazil (n=1), and Italy (n=1) (see table, Supplemental Digital Content 2, which summarizes the characteristics of each of the studies included).

Regarding the healthcare field, 17 studies (Ahn et al.,2019;Berglund et al.,2019;Bolz & Iorio,2016;Colegate-Stone et al.,2016;DiGioia & Greenhouse,2012;Dundon et al.,2016;Featherall et al.,2019;Featherall et al.,2018;Gabriel et al.,2019;Gray et al.,2019;Iorio et al.,2016;Johnson et al.,2019;Lee et al.,2016;McCreary et al.,2019;Navarro et al.,2018;Pelt et al.,2016;Rosner et al.,2018) presented orthopedic surgeries as the primary surgical procedure, and the other surgery types were as follows: gynecological (n=2) (Danilyants et al.,2019;Van Den Berg et al.,2020), cardiac (n=4) (Glotzbach et al.,2018;Kirkpatrick et al.,2015;van Veghel et al.,2016;J. B. Williams et al.,2019), bariatric (n=2) (Goretti et al.,2020;Noria et al.,2015), and spinal (n=1) (Parker et al.,2017). Seven studies focused on oncologic-surgical treatment, which included surgeries for the patient's episode of care for breast (n=2) (Ackerman et al.,2019;Smith et al.,2016), pancreatic (n=1) (Bateni et al.,2019), colorectal (n=1) (Gustafsson et al.,2016), prostate (n=1) (Peard et al.,2019), esophageal (n=1) (Khullar et al.,2015), and thyroid (n=2) (Abdulla et al.,2012;Kulkarni et al.,2011) cancer. Clinical approaches were assessed in 10 studies and were related to general practice (n=5) (Bernstein et al.,2018;Boltz et al.,2019;Hernandez A. et al.,2019;Van Deen et al.,2017;D. V. Williams et al.,2019), cardiovascular care (n=2)(Ebinger et al.,2018;Golas et al.,2018), oncological care (n=2) (Thaker et al.,2016;van Egdom et al.,2019) and pediatric care (n=1) (Weiss et al.,2019). Studies that involved both medical and surgical cohorts of patients (n=2) (Featherall et al.,2018) or value offices (n=1) (Makdisse et al.,2018) were classified as mixed fields.

We found 15 articles exploring the 'system' setting, especially multicenter or national

perspectives studies, and 32 articles that considered the application at a local level (hospital setting). The focus of the value program in the studies was more concentrated on pathway redesign (n=21) and traditional VBHC studies (n=20). For telemedicine or computational intelligence supporting value programs, only 6 studies were identified.

Outcomes Measurement

Financial outcomes were evaluated in 37 studies (79%); among them, 13 applied microcosting methods, with the time-driven activity-based costing (TDABC) method – the method recommended in the literature to be used in VBHC – used by only 6 studies (Chatfield et al.,2019;Danilyants et al.,2019;Hernandez A. et al.,2019;Noria et al.,2015;Peard et al.,2019;Pelt et al.,2016). The remaining 24 studies used reimbursement (n=6) (Berglund et al.,2019;Bolz & Iorio,2016;Dundon et al.,2016;Iorio et al.,2016;Navarro et al.,2018;Weiss et al.,2019), institutional accounting systems (n=6) (Berglund et al.,2019;Gray et al.,2019;Kulkarni et al.,2011;Peard et al.,2019;Pelt et al.,2016;Rosner et al.,2018), external databases (n=6) (Featherall et al.,2018;Golas et al.,2018;Khullar et al.,2015;Kirkpatrick et al.,2015;Smith et al.,2016;Van Deen et al.,2017), hospital charges(n=5) (Abdulla et al.,2012;Ackerman et al.,2019;Bateni et al.,2019;Berglund et al.,2019;Bernstein et al.,2018), and diagnosis-related groups (n=2) (Parker et al.,2017;Weiss et al.,2019), cost of implementation (n=1) (Boltz et al.,2019) as measures (see table, Supplemental Digital Content 3, which is presented the tier levels measured, financial outcomes, instruments used to support data collection, and data origin for all studies).

The main data source used was electronic health records (EHRs), including medical and hospital records (85%); the remaining studies used data from an external database (15%). Only 16 studies (34%) used PRO surveys to cover outcome tiers. Among those, generic metrics of multiple conditions appeared in 8 studies, and metrics tailored to disease classes were reported in 12 studies. Other surveys relating to patient experience (e.g., Hospital Consumer Assessment of Healthcare Providers and Systems [HCAHPS]) were conducted in 9 studies, and scales completed by professionals (e.g.,Activity Measure in Post-Acute Care [AMPAC]) were used in 3 studies.

The saturation of tier-level outcomes reported showed a limited coverage for value assessments in the literature reporting VBHC initiatives. Only 3 studies (Gray et al.,2019; Noria et al.,2015; Thaker et al.,2016) reported outcomes to all 6 levels of the tier hierarchy. Studies most commonly covered 3 levels of the outcome hierarchy (n=24,51%), followed by 4 levels (19%) and 5 levels (15%) levels.

We evaluated the profiles of the studies in regard to outcome measures considered in each tier-level. Among the 19 studies that measured the first level of tier 1, Mortality (or Survival), all but one (Khullar et al.,2015) considered in-hospital death, one-fifth showed specific surgical survival measures (Danilyants et al.,2019;Glotzbach et al.,2018;Makdisse et al.,2018;Noria et al.,2015) or 30-day mortality (Featherall et al.,2019;Golas et al.,2018;Khullar et al.,2015;Parker et al.,2017), and only two assessed one-year mortality (Featherall et al.,2019;Gray et al.,2019).

For Degree of Health/Recovery, the most representative measures applied by the studies were discharge-related, such as discharge disposition to home or care facilities (n=18) (Ahn et al.,2019;Bernstein et al.,2018;Bolz & Iorio,2016;DiGioia & Greenhouse,2012;Dundon et al.,2016;Ebinger et al.,2018;Featherall et al.,2019;Featherall et al.,2018;Glotzbach et al.,2018;Gray et al.,2019;Gustafsson et al.,2016;Hernandez A. et al.,2019;Iorio et al.,2016;Johnson et al.,2019;Parker et al.,2017;van Egdom et al.,2019;Weiss et al.,2019;D. V. Williams et al.,2019), and symptoms solved (n=2) (Hernandez A. et al.,2019;Kirkpatrick et al.,2015), followed by aspects of physical well-being, such as factors related to physical function(n=15) (Ahn et al.,2019;Berglund et al.,2019;Bernstein et al.,2018;DiGioia & Greenhouse,2012;Gabriel et al.,2019;Goretti et al.,2020;Hernandez A. et al.,2019;Johnson et al.,2019;McCreary et al.,2019;Parker et al.,2017;Pelt et al.,2016;Thaker et al.,2016;van Egdom et al.,2019;van Veghel et al.,2016;D. V. Williams et al.,2019) (e.g.,early ambulation) or pain-level achieved (n=3) (Ackerman et al.,2019;Berglund et al.,2019;Glotzbach et al.,2018). Two studies also evaluated diagnosis-related domains such as freedom from disease (Thaker et al.,2016;van Egdom et al.,2019), and 7 considered surgical outcomes (Danilyants et al.,2019;Goretti et al.,2020;McCreary et al.,2019;Peard et al.,2019;Pelt et al.,2016;Van Den Berg et al.,2020;J. B. Williams et al.,2019) directly related to function as organ preservation, method of tissue extraction, and estimated blood loss.

The total of 22 studies considered Time to Recovery, the first level of Tier 2. The outcomes could be divided into time needed to complete different phases of care, expressed as time to return to usual activities (n=8) (Berglund et al.,2019;Gabriel et al.,2019;Goretti et al.,2020;Makdisse et al.,2018;Noria et al.,2015;Parker et al.,2017;Thaker et al.,2016;van Egdom et al.,2019), time to care initiation (n=7) (Colegate-Stone et al.,2016;DiGioia & Greenhouse,2012;Glotzbach et al.,2018;Lee et al.,2016;McCreary et al.,2019;Van Den Berg et al.,2020;D. V. Williams et al.,2019), operative time (n=3) (Ackerman et al.,2019;DiGioia & Greenhouse,2012;McCreary et al.,2019), and time in the recovery phase (Bernstein et

al.,2018;DiGioia & Greenhouse,2012;Glottzbach et al.,2018;Gray et al.,2019;Hernandez A. et al.,2019;Johnson et al.,2019;Pelt et al.,2016).

The most coincident tier level among the studies, Disutility of the care process or Treatment Process, essentially comprised measures that providers directly control or traditionally measured clinical indicators, such as length of inpatient stay (n=33) (Ackerman et al.,2019;Bateni et al.,2019;Bernstein et al.,2018;Boltz et al.,2019;Bolz & Iorio,2016;Chatfield et al.,2019;Colegate-Stone et al.,2016;Danilyants et al.,2019;DiGioia & Greenhouse,2012;Dundon et al.,2016;Ebinger et al.,2018;Featherall et al.,2019;Featherall et al.,2018;Glottzbach et al.,2018;Goretti et al.,2020;Gray et al.,2019;Gustafsson et al.,2016;Iorio et al.,2016;Johnson et al.,2019;Khullar et al.,2015;Kirkpatrick et al.,2015;Kulkarni et al.,2011;McCreary et al.,2019;Navarro et al.,2018;Noria et al.,2015;Parker et al.,2017;Peard et al.,2019;Pelt et al.,2016;Ravikumar et al.,2010;Van Den Berg et al.,2020;Weiss et al.,2019;D. V. Williams et al.,2019;J. B. Williams et al.,2019), short-term complications (n=14) (Bateni et al.,2019;Danilyants et al.,2019;Ebinger et al.,2018;Golas et al.,2018;Khullar et al.,2015;Kirkpatrick et al.,2015;Kulkarni et al.,2011;Makdisse et al.,2018;Noria et al.,2015;Peard et al.,2019;Rosner et al.,2018;Smith et al.,2016;Thaker et al.,2016;van Egdom et al.,2019), intensive care unit (ICU) days (n=7) (Ackerman et al.,2019;Bernstein et al.,2018;Johnson et al.,2019;Khullar et al.,2015;Kirkpatrick et al.,2015;Kulkarni et al.,2011;J. B. Williams et al.,2019), infection rate (n=3) (DiGioia & Greenhouse,2012;Lee et al.,2016;Smith et al.,2016), and target medication usage(n=10) (Ackerman et al.,2019;Berglund et al.,2019;Bernstein et al.,2018;Glottzbach et al.,2018;Golas et al.,2018;Gray et al.,2019;Hernandez A. et al.,2019;Kulkarni et al.,2011;Van Deen et al.,2017;J. B. Williams et al.,2019). Patient satisfaction with care or service was a process measure relatively frequently considered (n=13) (Berglund et al.,2019;Boltz et al.,2019;Chatfield et al.,2019;Colegate-Stone et al.,2016;Danilyants et al.,2019;DiGioia & Greenhouse,2012;Dundon et al.,2016;Featherall et al.,2018;Gray et al.,2019;Hernandez A. et al.,2019;Kulkarni et al.,2011;Rosner et al.,2018;van Veghel et al.,2016) and was included in this level because it refers to satisfaction while undergoing treatment, thus indirectly reflecting the patient's perception of his or her entire cycle care (Porter,2010). Although rare, measures of anxiety, discomfort, and ability to work or function normally while undergoing treatment also appeared in the studies (n=8) (Glottzbach et al.,2018;Gray et al.,2019;Hernandez A. et al.,2019;Iorio et al.,2016;Kulkarni et al.,2011;Noria et al.,2015;Parker et al.,2017;Smith et al.,2016).

Sustainability of Health/Recovery mainly included measures such as 30-day (n=20) (Bateni et al.,2019;Bernstein et al.,2018;Boltz et al.,2019;Chatfield et al.,2019;Dundon et al.,2016;Golas et al.,2018;Goretti et al.,2020;Gustafsson et al.,2016;Hernandez A. et al.,2019;Iorio et al.,2016;Khullar et al.,2015;Kirkpatrick et al.,2015;Lee et al.,2016;Noria et al.,2015;Ravikumar et al.,2010;Van Deen et al.,2017;van Veghel et al.,2016;Weiss et al.,2019;D. V. Williams et al.,2019;J. B. Williams et al.,2019) and 90-day readmissions (n=6) (Ahn et al.,2019;Bolz & Iorio,2016;Dundon et al.,2016;Gray et al.,2019;Khullar et al.,2015;Rosner et al.,2018), also including the need for reoperation (n=6) (Abdulla et al.,2012;Goretti et al.,2020;Makdisse et al.,2018;Smith et al.,2016;Van Deen et al.,2017;J. B. Williams et al.,2019), and after-discharge complications (n=6) (Ahn et al.,2019;Featherall et al.,2019;Featherall et al.,2018;Parker et al.,2017;Rosner et al.,2018;Smith et al.,2016). Proportionally, the second level of tier 3, Long-term Consequences of Therapy, was the least expressed in the studies (n=15), and when reported, the measures focused on patient-reported health status.

Reported Results that Increased Value

The reported results that triggered a value increase in each case studied were consolidated in Table 1, which also describes whether the contribution was observed on financial outcome, clinical outcome improvement, PRO improvement, provider education and value culture, or hospital management.

A considerable proportion of the studies (36%) achieved results that contributed to value-based financial outcomes focused on cost savings. An important common finding was that the calculated savings were derived from reductions in readmissions and inpatient stays, which is accounted for as an indirect financial impact. However, these opportunities for future cost savings are not measured by accurate costs and economical methods. Two studies (Van Deen et al.,2017; Johnson et al.,2019) identified neutral effects or were not able to consistently observe improved results, even though they showed these effects as potential improvements.

Regarding PRO improvement, 25% of the studies reported increased PROs; however, when retrieving studies that used PRO measurement instruments (n=16),11 achieved positive results. Management effects, not surprisingly, were mainly related to hospital capacity (n=15), improved quality through risk adjustments (n=5), and better resource allocation (n=3).

On the Radar: Literature and HBS Cases

Twelve HBS cases were selected regarding value initiatives in the healthcare fields of

prostate cancer (Porter,Deerberg-Wittram,et al.,2014), orthopedic surgeries (Kaplan et al.,2020;Kaplan et al.,2012;Porter,Marks,et al.,2014), pediatric care (Porter,Bachmann,et al.,2014;Porter et al.,2016), and primary general practice (Kaplan et al.,2018;Porter et al.,2007;Porter et al.,2019;Porter,Landman,et al.,2014;Porter et al.,2017;Porter & Teisberg,2009) (see table, Supplemental Digital Content 4, which summarizes the main characteristics of these cases). In all HBS cases, the PRO measures were used. Figure 3 presents the comparison of the outcome tier coverage profiles of the included studies from the literature and the selected HBS cases. Of note, one initiative was reported in both metrics (Hernandez A. et al.,2019;"Porter, Michael E.,Alee,H.,Witkowski,M.L. and Feeley,Thomas W. "Navy Medicine: Moving to Value-Based Care" Harvard Business School,2019,").

The most conflicting information is regarding tier-levels more dependent on a PRO, such as Time to recovery and Long-term consequences, which are less expressed in the literature (47% and 32%,respectively); in contrast, in HBS cases, the same levels were more commonly considered (75% and 92%,respectively. This was not surprising, as predicted by the number of studies using PRO measures earlier in the results. Any method or financial measure of cost was included in a more significant proportion of both the literature and HBS cases (79% and 92%, respectively). Regarding microcosting or TDABC methods, the studies showed similar low prevalence in both the literature (27%) and HBS cases (28%), as revealed by the dotted lines in the figure. Mortality or survival-related metrics were also not widely computed in the selected studies (40% in the literature and 60% in HBS cases). Regarding the literature, only two studies (Abdulla et al.,2012;Ahn et al.,2019) did not measure the Disutility of the care process or Treatment Process level, and only six studies did not measure baseline characteristics for risk-adjustment data, making these two levels the most prevalent in the literature (96% and 87%,respectively). This pattern was also verified for HBS cases because those two levels were reported in all cases.

DISCUSSION

This systematic review intended to map how outcomes are being measured in real-world cases that followed the value agenda. We identified a significant imbalance of outcome measurements in many aspects, such as the configuration of tier-levels chosen in value initiatives, instruments applied to support data, and the rare use of microcosting methods regarding financial outcomes. Tier 3, especially Long-term consequences, was the least explored, whereas traditional clinical and process outcomes, such as length of stay and infections, were still the most frequent measures considered by the literature.

VBHC was first introduced to reduce waste in healthcare and to increase the quality of care delivered to patients (Porter & Lee,2013). As shown in this review, the increase in the quality of care is usually measured by the hospital and clinical outcomes, not necessarily by patient perception, and the financial results are not being reported with highly precise accounting methods. PRO measures play a central role in the value agenda model. Nevertheless, studies evaluating long-term consequences and new conditions are currently rare (Halpern et al.,2020); however, on the contrary, these factors received the most attention when we looked at the applied cases of VBHC in the HBS (Kaplan et al.,2018;Kaplan et al.,2020;Kaplan et al.,2012;Porter,Bachmann,et al.,2014;Porter et al.,2007;Porter,Deerberg-Wittram,et al.,2014;Porter et al.,2016;Porter et al.,2019;Porter,Landman,et al.,2014;Porter,Marks,et al.,2014;Porter & Teisberg,2009)

Our results agreed that widespread and consistent use of PRO measurements has proven to be ambiguous for a range of reasons, including the complexity of the measures tracked and the fluctuating reliability of patient assessments on many measures (Schupbach et al.,2016), which may be the prime factor for the gap seen in the radar chart in regard to the comprehensiveness of the value definition.

In agreement with the factors listed by Martin et al., (2019), our results demonstrate that measuring outcomes in healthcare has been difficult: first, current outcome measurements consist of nonstructured and condition-related data that are difficult to access; second, the adherence to only evidence-based processes rather than clinically and patient-reported driven results is limiting; and third, value initiatives still often do not incorporate a holistic view of the patient. The evolution to registries that inform and provide practicable patient-centered care may have two main routes: first, the education and dissemination of a value culture, which instantly reinforce staff to register important outcomes about the patient, either through validated questionnaires or more effective multidisciplinary meetings; and second, the creation of an EHR-integrated real-time outcome measurement platform. This reflection leads us to understand that VBHC is not feasible without investment in information technology (Boscolo et al.,2020). Once a functionally integrated EHR is implemented, it must be validated to ensure that it provides quality measurements, an essential component of quality improvement (Etges et al.,2020). Addressing suboptimal outcomes and comparing cost data for treatment options will facilitate process improvement and value (Thaker et al.,2016).

In an attempt to address the shortcomings of outcome measurement, the International Consortium for Health Outcomes Measurement was created (ICHOM | Healthcare

Improvement | Patient-Reported Outcomes). The ICHOM working group understands that the role of such parameterization is not to devise new measures of results but to agree on which well-assessed outcome measure indicator everyone should use to cover a much broader spectrum of the outcome hierarchy for a health condition. The use of the ICHOM questionnaires in the literature is still restricted to a few studies concentrated on the fields of orthopedic surgery (Berglund et al.,2019;Glotzbach et al.,2018;Pelt et al.,2016), general practice (Hernandez A. et al.,2019;Kulkarni et al.,2011), bariatric surgery (Noria et al.,2015), obstetrics (Van Den Berg et al.,2020), breast cancer (van Egdom et al.,2019), and prostate cancer (Thaker et al.,2016).

Only the ICHOM, however, was not enough to add all the tiers of value, and some processual measures are still needed (Thaker et al.,2016). It is evident in the HBS cases that measuring outcomes – clinical, processual, financial, and patient-reported outcomes – is a valuable tool that helps the system to be more intentional about quality, efficiency, and especially, patient outcomes (Porter,2010;Schupbach et al.,2016), and this model has been demonstrated to have a higher success rate and permanence (Porter,2010;Thaker et al.,2016).

Regarding the financial outcomes, we attest that cost studies were not part of our main scope. Nevertheless, in the spectrum of value initiatives, we would expect an exploration of both numerators and denominators to compose the value equation. However, by identifying that only 13 studies and 3 cases applied advanced methods to evaluate real costs, we suggest that scaled VBHC adoption requires more methodological rigor when evaluating financial outcomes (Etges et al.,2020; Tsai et al.,2018).

Developments in electronic medical records are already making outcomes far less costly to measure (Porter & Teisberg,2006). The majority of VBHC studies used medical records to collect data to evaluate the value of healthcare. However, as verified from the studies that covered the full range of the outcome hierarchy (Iorio et al.,2016;Noria et al.,2015), the electronic medical records do not uniformly capture the three tiers of outcomes described, requiring additional providers to manually maintain parallel control of the data and update the research databases (Noria et al.,2015).

There are weaknesses and strengths to consider. To the best of our knowledge, this is the first systematic review with an objective to perform a broad literature search of studies with a priori defined methods and well-established methodological guidelines. However, as there is not a valid instrument to assess the methodological quality of VBHC initiative studies, the methodological quality of these studies was not determined. We also did not

identify studies with negative results associated with the outcome measurement in the VBHC initiative; thus, there is potential publication bias toward those only reporting successful results in this field. Additionally, as the searches were conducted in early March 2020, this study does not include or reflect the possible movement in VBHC initiatives driven by the coronavirus disease 2019 pandemic; thus, we consider this a pre-pandemic systematic review on VBHC initiatives.

CONCLUSIONS

Our systematic review suggests that, in the real-world setting, there is still a gap in measuring outcomes that matter to patients and measuring financial outcomes through rigorous methodological methods. Advances in technology capacity and a culture of change in management appear to be the main barriers to making the value agenda more easily reproducible and are still requirements necessary for the items above to occur.

REFERENCES

- Abdulla,A. ,Ituarte,P. G.,Wiggins,R.,Teisberg,E.,Harari,A.,& Yeh,M. (2012,2012). Endocrine surgery as a model for value-based health care delivery. *Surgical Neurology International*,3(1),163. <https://doi.org/10.4103/2152-7806.105102>
- Ackerman,R. S.,Hirschi,M.,Alford,B.,Evans,T.,Kiluk,J. V.,& Patel,S. Y. (2019,2019/03//). Enhanced REVENUE After Surgery? A Cost-Standardized Enhanced Recovery Pathway for Mastectomy Decreases Length of Stay. *World Journal of Surgery*,43(3),839-845. <https://doi.org/10.1007/s00268-018-4850-0>
- Ahn,A.,Ferrer,C.,Park,C.,Snyder,D. J.,Maron,S. Z.,Mikhail,C.,Keswani,A.,Molloy,I. B.,Bronson,M. J.,Moschetti,W. E.,Jevsevar,D. S.,Poeran,J.,Galatz,L. M.,& Moucha,C. S. (2019,2019). Defining and Optimizing Value in Total Joint Arthroplasty From the Patient,Payer,and Provider Perspectives. *The Journal of Arthroplasty*,34(10),2290-2296.e2291. <https://doi.org/10.1016/j.arth.2019.05.024>
- Batani,S. B.,Gingrich,A. A.,Hoch,J. S.,Canter,R. J.,& Bold,R. J. (2019,2019). Defining Value for Pancreatic Surgery in Early-Stage Pancreatic Cancer. *JAMA Surgery*,154(10). <https://doi.org/10.1001/jamasurg.2019.3019>
- Berglund,D. D.,Law,T. Y.,Rosas,S.,Kurowicki,J.,Giveans,M. R.,Mijic,D.,& Levy,J. C. (2019,2019/02//). The procedure value index: a new method for quantifying value in shoulder arthroplasty. *Journal of shoulder and elbow surgery*,28(2(Berglund,2019 #1705)),335-340. <https://doi.org/10.1016/j.jse.2018.07.031>
- Bernstein,J. M.,Graven,P.,Drago,K.,Dobbertin,K.,& Eckstrom,E. (2018,2018). Higher Quality,Lower Cost with an Innovative Geriatrics Consultation Service. *Journal of the American Geriatrics Society*,66(9),1790-1795. <https://doi.org/10.1111/jgs.15473>
- Boltz,M.,Cuellar,N. G.,Cole,C.,& Pistorese,B. (2019,2019/05//). Comparing an on-site nurse practitioner with telemedicine physician support hospitalist programme with a traditional physician hospitalist programme. *Journal of telemedicine and telecare*,25(4),213-220. <https://doi.org/10.1177/1357633X18758744>
- Bolz,N. J.,& Iorio,R. (2016,2016). Bundled Payments: Our Experience at an Academic Medical Center. *Journal of Arthroplasty*,31(5),932-935. <https://doi.org/10.1016/j.arth.2016.01.055>
- Boscolo,P. R.,Callea,G.,Ciani,O.,& Tarricone,R. (2020,Jan). Measuring Value in Health Care: A Comparative Analysis of Value-based Frameworks. *Clin Ther*,42(1),34-43. <https://doi.org/10.1016/j.clinthera.2019.11.017>
- Chalmers,K.,Badgery-Parker,T.,Pearson,S.-A.,Brett,J.,Scott,I. A.,& Elshaug,A. G. (2018,2018/03/05//). Developing indicators for measuring low-value care: mapping Choosing Wisely recommendations to hospital data. *BMC research notes*,11(1),163-163. <https://doi.org/10.1186/s13104-018-3270-4>

Chatfield,S. C.,Volpicelli,F. M.,Adler,N. M.,Kim,K. L.,Jones,S. A.,Francois,F.,Shah,P. C.,Press,R. A.,& Horwitz,L. I. (2019,2019). Bending the cost curve: Time series analysis of a value transformation programme at an academic medical centre. *BMJ Quality and Safety*,28(6),449-458. <https://doi.org/10.1136/bmjqs-2018-009068>

Colegate-Stone,T.,Tavakkolizadeh,A.,Moxham,J.,& Sinha,J. (2016,2016). Increasing value: The King's College Hospital experience. *British Journal of Health Care Management*,22(6),326-334. <https://doi.org/10.12968/bjhc.2016.22.6.326>

Danilyants,N.,MacKoul,P.,van der Does,L.,Haworth,L.,& Baxi,R. (2019,2019). A value-based evaluation of minimally invasive hysterectomy approaches. *Gynecological Surgery*,16(1). <https://doi.org/10.1186/s10397-019-1057-9>

DiGioia,A. M.,3rd,& Greenhouse,P. K. (2012,2012). Care experience-based methodologies: performance improvement roadmap to value-driven health care. *Clinical orthopaedics and related research*,470(4),1038-1045. <https://doi.org/10.1007/s11999-011-2051-3>

Dundon,J. M.,Bosco,J.,Slover,J.,Yu,S.,Sayeed,Y.,& Iorio,R. (2016,2016/12/07/). Improvement in Total Joint Replacement Quality Metrics: Year One Versus Year Three of the Bundled Payments for Care Improvement Initiative. *The Journal of bone and joint surgery. American volume*,98(23),1949-1953. <https://doi.org/10.2106/JBJS.16.00523>

Ebinger,J. E.,Strauss,C. E.,Garberich,R. R.,Bradley,S. M.,Rush,P.,Chavez,I. J.,Poulose,A. K.,Porten,B. R.,& Henry,T. D. (2018,2018). Value-based ST-segment-elevation myocardial infarction care using risk-guided triage and early discharge. *Circulation: Cardiovascular Quality and Outcomes*,11(4). <https://doi.org/10.1161/CIRCOUTCOMES.118.004553>

Etges,A. P. B. d. S.,Ruschel,K. B.,Polanczyk,C. A.,& Urman,R. D. (2020,2020/06/). Advances in Value-Based Healthcare by the Application of Time-Driven Activity-Based Costing for Inpatient Management: A Systematic Review. *Value in Health*,23(6),812-823. <https://doi.org/10.1016/j.jval.2020.02.004>

Featherall,J.,Brigati,D. P.,Arney,A. N.,Faour,M.,Bokar,D. V.,Murray,T. G.,Molloy,R. M.,& Higuera Rueda,C. A. (2019,2019). Effects of a Total Knee Arthroplasty Care Pathway on Cost,Quality,and Patient Experience: Toward Measuring the Triple Aim. *Journal of Arthroplasty*,34(11),2561-2568. <https://doi.org/10.1016/j.arth.2019.06.011>

Featherall,J.,Brigati,D. P.,Faour,M.,Messner,W.,& Higuera,C. A. (2018,2018/06/). Implementation of a Total Hip Arthroplasty Care Pathway at a High-Volume Health System: Effect on Length of Stay,Discharge Disposition,and 90-Day Complications. *The Journal of Arthroplasty*,33(6),1675-1680. <https://doi.org/10.1016/j.arth.2018.01.038>

Gabriel,L.,Casey,J.,Gee,M.,Palmer,C.,Sinha,J.,Moxham,J.,& Colegate-Stone,T. J. (2019,2019/06/09/). Value-based healthcare analysis of joint replacement surgery for patients with primary hip osteoarthritis. *BMJ open quality*,8(2),e000549-e000549. <https://doi.org/10.1136/bmjoq-2018-000549>

Glotzbach, J. P., Sharma, V., Tonna, J. E., Pettit, J. C., McKellar, S. H., Eckhauser, A. W., Varghese, T. K., & Selzman, C. H. (2018, 2018). Value-driven cardiac surgery: Achieving “perfect care” after coronary artery bypass grafting. *Journal of Thoracic and Cardiovascular Surgery*, 156(4), 1436-1448. e1432. <https://doi.org/10.1016/j.jtcvs.2018.03.177>

Golas, S.

B., Shibahara, T., Agboola, S., Otaki, H., Sato, J., Nakae, T., Hisamitsu, T., Kojima, G., Felsted, J., Kakarmath, S., Kvedar, J., & Jethwani, K. (2018, 2018/06/22/). A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC medical informatics and decision making*, 18(1), 44-44. <https://doi.org/10.1186/s12911-018-0620-z>

Goretti, G., Marinari, G. M., Vanni, E., & Ferrari, C. (2020, 2020). Value-Based Healthcare and Enhanced Recovery After Surgery Implementation in a High-Volume Bariatric Center in Italy. *Obesity Surgery* (Goretti G., giulia.goretti@humanitas.it) Quality Department, Humanitas Clinical and Research Center –IRCCS, via Manzoni, Rozzano, MI, Italy). <https://doi.org/10.1007/s11695-020-04464-w>

Gray, C. F., Prieto, H. A., Deen, J. T., & Parvataneni, H. K. (2019, 2019/02//). Bundled Payment “Creep”: Institutional Redesign for Primary Arthroplasty Positively Affects Revision Arthroplasty. *The Journal of Arthroplasty*, 34(2), 206-210. <https://doi.org/10.1016/j.arth.2018.10.025>

Gustafsson, U. O., Ooppelstrup, H., Thorell, A., Nygren, J., & Ljungqvist, O. (2016, 2016/07//). Adherence to the ERAS protocol is Associated with 5-Year Survival After Colorectal Cancer Surgery: A Retrospective Cohort Study. *World Journal of Surgery*, 40(7), 1741-1747. <https://doi.org/10.1007/s00268-016-3460-y>

Halpern, M. T., Gasoyan, H., & Aaronson, W. E. (2020, May-Jun). Metrics to Evaluate the Performance of Cancer Center Leadership: A Systematic Review. *J Healthc Manag*, 65(3), 217-235. <https://doi.org/10.1097/JHM-D-19-00064>

Harvard Business School Publishing. (2020). <https://store.hbr.org/case-studies/>

Hernandez A., Kaplan R.S., Witkowski M.L., Forrest Faison C., & Porter M.E. (2019). Navy Medicine Introduces value-based health care. *Health affairs (Project Hope)*, 38(8)(8), 1393-1400. <https://doi.org/10.1377/hlthaff.2019.00280>

ICHOM | Healthcare Improvement | Patient-Reported Outcomes. <https://www.ichom.org/files/28502/www.ichom.org.html>

Iorio, R., Clair, A. J., Inneh, I. A., Slover, J. D., Bosco, J. A., & Zuckerman, J. D. (2016, 2016/02//). Early Results of Medicare's Bundled Payment Initiative for a 90-Day Total Joint Arthroplasty Episode of Care. *The Journal of Arthroplasty*, 31(2), 343-350. <https://doi.org/10.1016/j.arth.2015.09.004>

Johnson, J. K., Erickson, J. A., Miller, C. J., Fritz, J. M., Marcus, R. L., & Pelt, C. E. (2019, 2019). Short-term functional recovery after total joint arthroplasty is unaffected by bundled payment participation. *Arthroplasty Today*, 5(1), 119-125. <https://doi.org/10.1016/j.artd.2018.12.003>

Kaplan, R. S., Porter, M. E., Feeley, T. W., & Hernandez, A. (2018). "Medtronic: Navigating a Shifting Healthcare Landscape."

Kaplan, R. S., Witkowski, M. L., & Emanuel, T. (2020). "New England Baptist Hospital: Paying for Value"

Kaplan, R. S., Witkowski, M. L., & Hohman, J. A. (2012). "Schön Klinik: Measuring Cost and Value."

Khullar, O. V., Jiang, R., Force, S. D., Pickens, A., Sancheti, M. S., Ward, K., Gillespie, T., & Fernandez, F. G. (2015, 2015). Transthoracic versus transhiatal resection for esophageal adenocarcinoma of the lower esophagus: A value-based comparison. *Journal of Surgical Oncology*, 112(5), 517-523. <https://doi.org/10.1002/jso.24024>

Kirkpatrick, J. R., Marks, S., Slane, M., Kim, D., Cohen, L., Cortelli, M., Plate, J., Perryman, R., & Zapas, J. (2015, 2015). Using value-based analysis to influence outcomes in complex surgical systems. *Journal of the American College of Surgeons*, 220(4), 461-468. <https://doi.org/10.1016/j.jamcollsurg.2014.12.034>

Kulkarni, R. P., Ituarte, P. H. G., Gunderson, D., & Yeh, M. W. (2011, 2011/01//). Clinical Pathways Improve Hospital Resource Use in Endocrine Surgery. *Journal of the American College of Surgeons*, 212(1), 35-41. <https://doi.org/10.1016/j.jamcollsurg.2010.09.016>

Lee, V. S., Kawamoto, K., Hess, R., Park, C., Young, J., Hunter, C., Johnson, S., Gulbransen, S., Pelt, C. E., Horton, D. J., Graves, K. K., Greene, T. H., Anzai, Y., & Pendleton, R. C. (2016, 2016/09/13/). Implementation of a Value-Driven Outcomes Program to Identify High Variability in Clinical Costs and Outcomes and Association With Reduced Cost and Improved Quality. *JAMA*, 316(10), 1061-1072. <https://doi.org/10.1001/jama.2016.12226>

Makdisse, M., Katz, M., Ramos, P., Pereira, A., Shiramizo, S., Neto, M. C., & Klajner, S. (2018, 2018). What Is a Value Management Office? An Implementation Experience in Latin America. *Value in Health Regional Issues*, 17((Makdisse M., marcia.makdisse@einstein.br; Katz M.; Ramos P.; Pereira A.; Shiramizo S.; Neto M.C.; Klajner S.) Hospital Israelita Albert Einstein, São Paulo, Brazil), 71-73. <https://doi.org/10.1016/j.vhri.2018.02.002>

Martin, J., BL., P., TF., F., & TW., F. (2019). Value-based Health Care: Lessons for the Anesthesiologist. *Int. Anesthesiol Clin*, 57(1), 63-80.

McCreary,D. L.,Dugarte,A. J.,Vang,S.,Plowman,B.,Williams,B. R.,Parikh,H. R.,& Cunningham,B. P. (2019,2019). Patient-Level Value Analysis: An Innovative Approach to Optimize Care Delivery. *Journal of Orthopaedic Trauma*, S49-S52. <https://doi.org/10.1097/BOT.0000000000001624>

Moher,D.,Liberati,A.,Tetzlaff,J.,Altman,D. G.,& for the,P. G. (2009,2009/07/21/). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*,339(jul21 1),b2535-b2535. <https://doi.org/10.1136/bmj.b2535>

Navarro,S. M.,Wang,E. Y.,Haeberle,H. S.,Mont,M. A.,Krebs,V. E.,Patterson,B. M.,& Ramkumar,P. N. (2018,2018). Machine Learning and Primary Total Knee Arthroplasty: Patient Forecasting for a Patient-Specific Payment Model. *Journal of Arthroplasty*,33(12),3617-3623. <https://doi.org/10.1016/j.arth.2018.08.028>

Noria,S.,Needleman,B.,Mikami,D.,Layton,K.,Jalilvand,A.,Hornor,M.,Suzo,A.,Abdel-Rasoul,M.,& Macadam,L. (2015,2015). The impact of care coaching on hospital length of stay,readmission rates and overall patient satisfaction post-bariatric surgery. *Surgery for Obesity and Related Diseases*,11(6),S8-S9. <http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L72280035>

Parker,S. L.,Chotai,S.,Devin,C. J.,Tetreault,L.,Mroz,T. E.,Brodke,D. S.,Fehlings,M. G.,& McGirt,M. J. (2017,2017). Bending the cost curve - Establishing value in spine surgery. *Clinical Neurosurgery*,80(3),S61-S69. <https://doi.org/10.1093/neuros/nyw081>

Peard,L.,Goodwin,J.,Hensley,P.,Dugan,A.,Bylund,J.,& Harris,A. M. (2019,2019). Examining and Understanding Value: The Impact of Preoperative Characteristics,Intraoperative Variables,and Postoperative Complications on Cost of Robot-Assisted Laparoscopic Radical Prostatectomy. *Journal of endourology*,33(7),541-548. <https://doi.org/10.1089/end.2019.0066>

Pelt,C. E.,Anderson,M. B.,Pendleton,R.,Foulks,M.,Peters,C. L.,& Gililland,J. M. (2016,2016/04/08/). Improving value in primary total joint arthroplasty care pathways: changes in inpatient physical therapy staffing. *Arthroplasty Today*,3(1),45-49. <https://doi.org/10.1016/j.artd.2016.02.003>

Porter,M. E. (2010,2010/12/23/). What Is Value in Health Care? *New England Journal of Medicine*,363(26),2477-2481. <https://doi.org/10.1056/NEJMp1011024>

Porter,M. E.,Bachmann,J. M.,& C.,L. Z. (2014). "Texas Children's Hospital: Congenital Heart Disease Care."

Porter,M. E.,Clemens Guth,& Dannemiller,E. M. (2007). "The West German Headache Center: Integrated Migraine Care."

Porter,M. E.,Deerberg-Wittram,J.,& Marks,C. (2014). "Martini Klinik: Prostate Cancer Care."

Porter, M. E., Feeley, T. W., & Okanlawon, T. J. (2016). "The Children's Hospital of Philadelphia: Network Strategy 2016."

Porter, M. E., Hernandez, A., Witkowski, M. L., & Feeley, T. W. (2019). "Navy Medicine: Moving to Value-Based Care"

Porter, M. E., Landman, Z. C., & Haas, D. A. (2014). "Vanderbilt: Transforming a Health Care Delivery System."

Porter, M. E., & Lee, T. H. (2013, 2013/10//). The Strategy That Will Fix Health Care. *Harvard Business Review*, 91(10), 50-70.

Porter, M. E., Lee, T. H., & Alger, M. A. (2017). "Oak Street Health: A New Model of Primary Care."

Porter, M. E., Marks, C. M., & Landman, Z. C. (2014). "OrthoChoice: Bundled Payments in the County of Stockholm (A)."

Porter, M. E., & Teisberg, E. O. (2006). *Redefining health care: creating value-based competition on results*. Harvard Business School Press.

Porter, M. E., & Teisberg, E. O. (2009). "Cleveland Clinic: Transformation and Growth 2015."

Porter, Michael E., Alee, H., Witkowski, M.L. and Feeley, Thomas W. "Navy Medicine: Moving to Value-Based Care" Harvard Business School, 2019.

Ravikumar, T. S., Sharma, C., Marini, C., Steele, G. D., Jr., Ritter, G., Barrera, R., Kim, M., Safyer, S. M., Vandervoort, K., De Geronimo, M., Baker, L., Levi, P., Pierdon, S., Horgan, M., Maynor, K., Maloney, G., Wojtowicz, M., & Nelson, K. (2010, 2010). A validated value-based model to improve hospital-wide perioperative outcomes: adaptability to combined medical/surgical inpatient cohorts. *Annals of Surgery*, 252(3), 486-498. <https://doi.org/10.1097/SLA.0b013e3181f1c412>

Rosner, B. I., Gottlieb, M., & Anderson, W. N. (2018, 2018). Effectiveness of an Automated Digital Remote Guidance and Telemonitoring Platform on Costs, Readmissions, and Complications After Hip and Knee Arthroplasties. *The Journal of Arthroplasty*, 33(4), 988-996.e984. <https://doi.org/10.1016/j.arth.2017.11.036>

Schupbach, J., Chandra, A., & Huckman, R. S. (2016, 2016/12/08/T20:50:15Z). A Simple Way to Measure Health Care Outcomes. *Harvard Business Review*. <https://hbr.org/2016/12/a-simple-way-to-measure-health-care-outcomes/files/28494/a-simple-way-to-measure-health-care-outcomes.html>

Smith, B. D., Jiang, J., Shih, Y.-C., Giordano, S. H., Huo, J., Jagsi, R., Momoh, A. O., Caudle, A. S., Hunt, K. K., Shaitelman, S. F., Buchholz, T. A., & Shirvani, S. M. (2016, 2016/09/27/). Cost and Complications of Local Therapies for Early-Stage Breast Cancer. *Journal of the National Cancer Institute*, 109(1), djw178.

<https://doi.org/10.1093/jnci/djw178>

Thaker, N. G., Pugh, T. J., Mahmood, U., Choi, S., Spinks, T. E., Martin, N. E., Sio, T. T., Kudchadker, R. J., Kaplan, R. S., Kuban, D. A., Swanson, D. A., Orio, P. F., Zelefsky, M. J., Cox, B. W., Potters, L., Buchholz, T. A., Feeley, T. W., & Frank, S. J. (2016, 2016/05//). Defining the value framework for prostate brachytherapy using patient-centered outcome metrics and time-driven activity-based costing. *Brachytherapy*, 15(3), 274-282. <https://doi.org/10.1016/j.brachy.2016.01.003>

Tsai, M. H., Porter, J. C., & Adams, D. C. (2018, 2018/07//). The Denominator in Value-Based Health Care: Porter's Hidden Costs. *Anesthesia & Analgesia*, 127(1), 317. <https://doi.org/10.1213/ANE.0000000000003401>

Tseng, E. K., & Hicks, L. K. (2016, 2016/08//). Value Based Care and Patient-Centered Care: Divergent or Complementary? *Current hematologic malignancy reports*, 11(4), 303-310. <https://doi.org/10.1007/s11899-016-0333-2>

Van Deen, W. K., Spiro, A., Burak Ozbay, A., Skup, M., Centeno, A., Duran, N. E., Lacey, P. N., Jatulis, D., Esrailian, E., Van Oijen, M. G. H., & Hommes, D. W. (2017, 2017). The impact of value-based healthcare for inflammatory bowel diseases on healthcare utilization: A pilot study. *European Journal of Gastroenterology and Hepatology*, 29(3), 331-337. <https://doi.org/10.1097/MEG.0000000000000782>

Van Den Berg, K., Dijkman, L. M., Keus, S. H. J., Scheele, F., & Van Pampus, M. G. (2020, 2020). Value-based health care in obstetrics. *Journal of Evaluation in Clinical Practice*, 26(1), 101-108. <https://doi.org/10.1111/jep.13208>

Van Der Wees, P. J., Nijhuis-Van Der Sanden, M. W. G., Ayanian, J. Z., Black, N., Westert, G. P., & Schneider, E. C. (2014, 2014/12//). Integrating the Use of Patient-Reported Outcomes for Both Clinical Practice and Performance Measurement: Views of Experts from 3 Countries: Patient-Reported Outcomes, Clinical Practice, Performance Measurement. *Milbank Quarterly*, 92(4), 754-775. <https://doi.org/10.1111/1468-0009.12091>

van Egdom, L. S. E., Lagendijk, M., van der Kemp, M. H., van Dam, J. H., Mureau, M. A. M., Hazelzet, J. A., & Koppert, L. B. (2019, 2019). Implementation of Value Based Breast Cancer Care. *European Journal of Surgical Oncology*, 45(7), 1163-1170. <https://doi.org/10.1016/j.ejso.2019.01.007>

van Veghel, D., Martijn, M., & de Mol, B. (2016, 2016/06//). First results of a national initiative to enable quality improvement of cardiovascular care by transparently reporting on patient-relevant outcomes. *European Journal of Cardio-Thoracic Surgery*, 49(6), 1660-1669. <https://doi.org/10.1093/ejcts/ezw034>

Weiss, M. A., Marchese, S., & Zhang, L. (2019, 2019). Effective Care Management for Children With Special Health Care Needs in the Era of Value-Based Payment. *Clinical pediatrics*, 58(9), 949-956. <https://doi.org/10.1177/0009922819839231>

Williams, D. V., Liu, T. C., Zywielski, M. G., Hoff, M. K., Ward, L., Bozic, K. J., & Koenig, K. M. (2019, 2019). Impact of an integrated practice unit on the value of musculoskeletal care for

uninsured and underinsured patients. *Healthcare (Amsterdam, Netherlands)*,7(2),16-20.
<https://doi.org/10.1016/j.hjdsi.2018.10.001>

Williams, J. B., McConnell, G., Allender, J. E., Woltz, P., Kane, K., Smith, P. K., Engelman, D. T., & Bradford, W. T. (2019, 2019/05//). One-year results from the first US-based enhanced recovery after cardiac surgery (ERAS Cardiac) program. *The Journal of thoracic and cardiovascular surgery*,157(5),1881-1888.
<https://doi.org/10.1016/j.jtcvs.2018.10.164>

Figure 1 – Outcome Hierarchy and measures, adapted from Porter (2010).

| | |
|---|---|
| Tier 1 - Health Status Achieved - | |
| Mortality (survival) <i>- Common mortality/survival metrics</i> | Degree of health or recovery <i>- Functional status achieved;</i> <i>- Freedom of disease</i> |
| Outcomes may be related to structures that allow safe and effective service | |
| Tier 2 - Process of Recovery - | |
| Time <i>-Cycle time to complete a care phase</i> <i>- Time-related burden of recovery</i> | Disutility of the care and treatment process <i>- Inconvenience, harm and distress while undergoing treatment</i> |
| Outcomes may include short-term complications, medications usage, length-of-stay during cycle of care | |
| Tier 3 - Sustainability of Health- | |
| Sustainability of Health <i>-Maintainance of health status</i> <i>-Need for revision/readmission</i> | Long-term consequences <i>- New health problems</i> <i>- Care-inducend condition</i> |
| Outcomes may include recurrence of events, readmission rates, incidence of secondary health conditions | |

Figure 2- Flowchart of study selection according to the PRISMA statement.

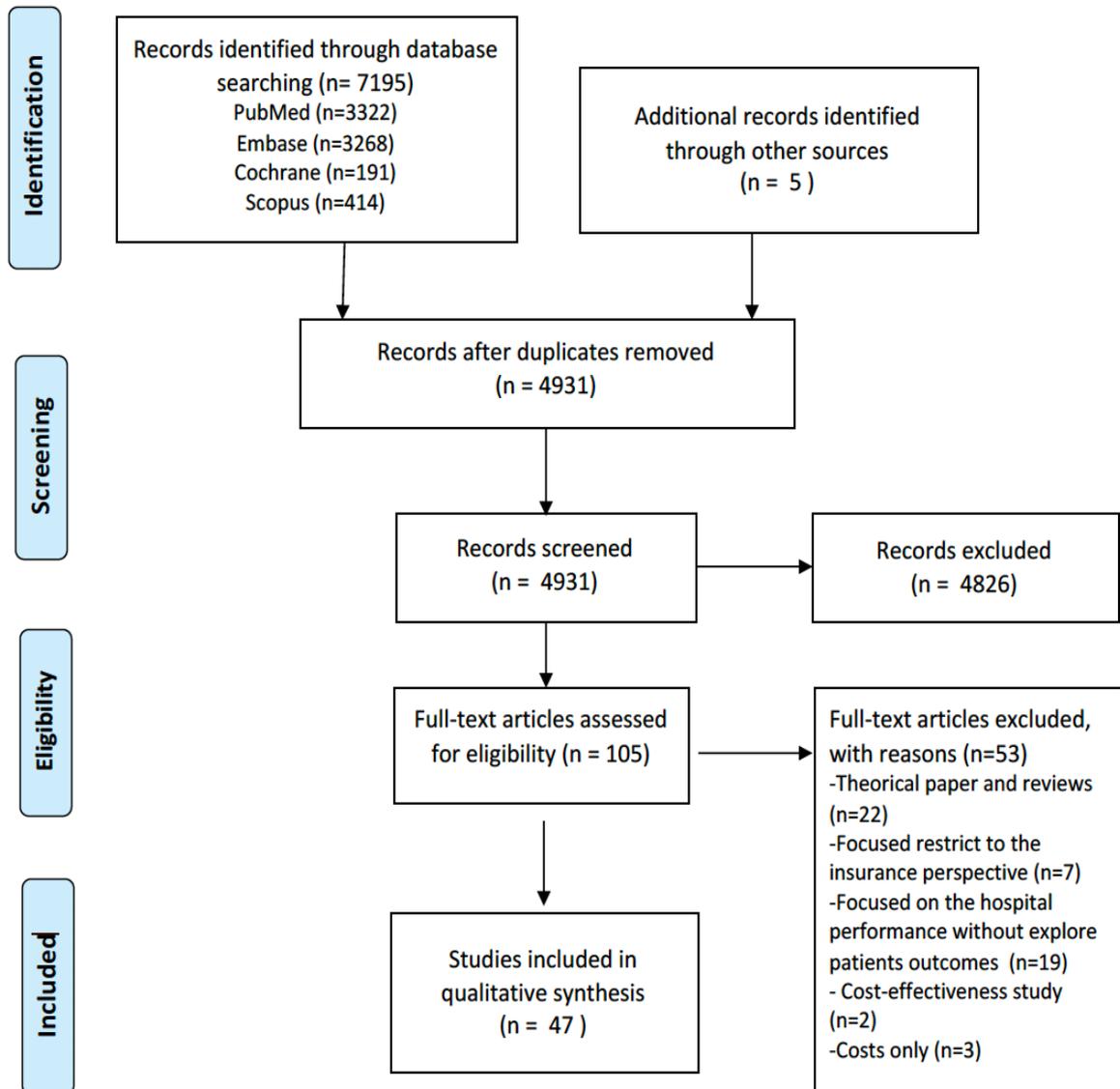
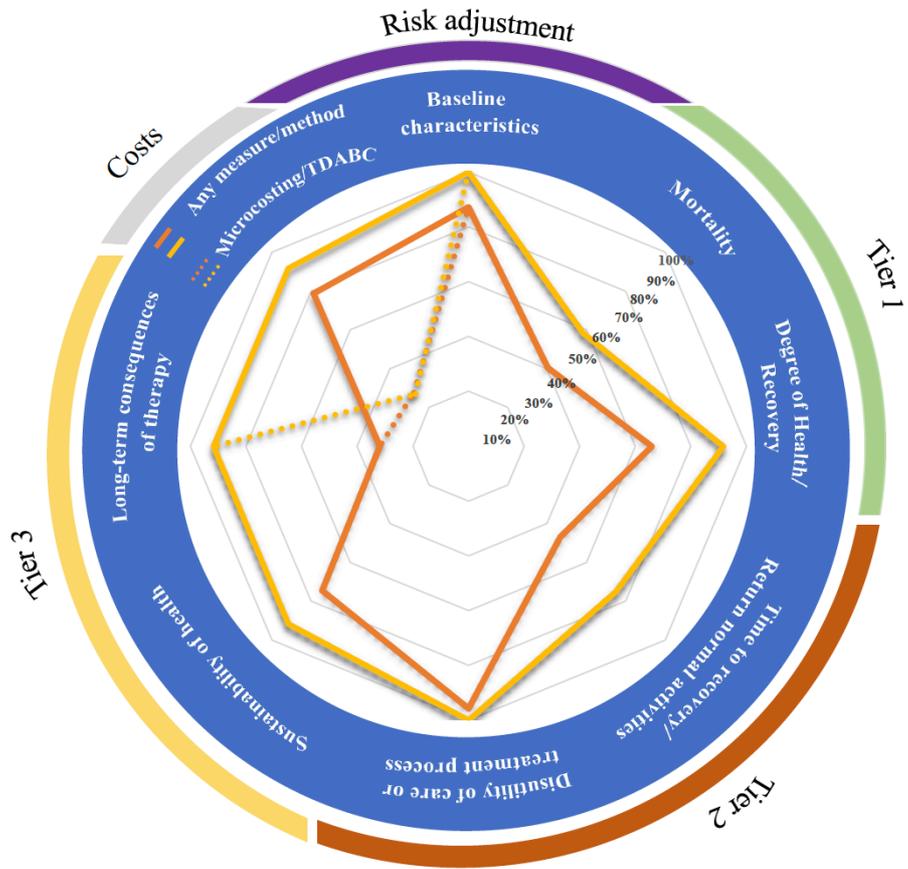


Table 1- A summary of value-effect and domains reported in real-world settings.

| Value-effect | Domain | Reported |
|---|---|---|
| <p>Value = $\frac{\text{Outcomes (t=1)} - \text{Outcomes (t=0)}}{\text{Costs (t=1)} - \text{Costs (t=0)}}$</p> <p>Where: Outcomes include measures stratified in the Tiers 1,2 and 3. Costs may consider costs over the complete pathway. t= time</p> | <p>Financial Outcome</p> | <p>Direct cost-savings^{(Ackerman et al.,2019;(Bernstein et al.,2018;Boltz et al.,2019;Bolz & Iorio,2016;Chatfield et al.,2019;Dundon et al.,2016;Ebinger et al.,2018;Featherall et al.,2019;Glotzbach et al.,2018;Goretti et al.,2020;Gray et al.,2019;Iorio et al.,2016;Lee et al.,2016;Pelt et al.,2016)}</p> <p>Indirect cost-savings^(DiGioia & Greenhouse,2012;Weiss et al.,2019)</p> <p>Reduce variance in cost^(Ackerman et al.,2019)</p> <p>Sustainable^(Goretti et al.,2020)</p> |
| | <p>Clinical Outcome Improvement</p> | <p>Reduce complications^{(Danilyants et al.,2019;(Goretti et al.,2020;Rosner et al.,2018)}</p> <p>Reduce mortality^(Colegate-Stone et al.,2016;DiGioia & Greenhouse,2012;Iorio et al.,2016)</p> <p>Improved laboratories and recovered from comorbidities^(Abdulla et al.,2012;Goretti et al.,2020;Iorio et al.,2016;D. V. Williams et al.,2019)</p> <p>Perioperative Outcomes^(J. B. Williams et al.,2019)</p> <p>Reduced pharmacological treatment time^(Hernandez A. et al.,2019;Kirkpatrick et al.,2015;Lee et al.,2016)</p> |
| | <p>Patient-reported Outcome Improvement</p> | <p>Patient satisfaction with service^(Colegate-Stone et al.,2016;DiGioia & Greenhouse,2012;Noria et al.,2015;van Egdome et al.,2019)</p> <p>Improved work and function relationships^(Hernandez A. et al.,2019;Ahn et al.,2019;Goretti et al.,2020;Parker et al.,2017;Weiss et al.,2019)</p> <p>Improved/Favorable quality of life scores^(Ahn et al.,2019;Iorio et al.,2016;Parker et al.,2017;J. B. Williams et al.,2019)</p> <p>Improved well-being^(Hernandez A. et al.,2019;Bateni et al.,2019;Goretti et al.,2020)</p> |
| | <p>Providers Education and Value Culture</p> | <p>Support for innovative implementations^(Boltz et al.,2019)</p> <p>Value consciousness and engagement^{(Ackerman et al.,2019;Chatfield et al.,2019;Gustafsson et al.,2016;Navarro et}</p> |

| | | |
|--|----------------------------|--|
| | | al.,2018;Noria et al.,2015;Ravikumar et al.,2010) Replicable (Goretti et al.,2020) |
| | Hospital Management | <p>Increase in hospital capacity (Abdulla et al.,2012;Ackerman et al.,2019;Bolz & Iorio,2016;Chatfield et al.,2019;Dundon et al.,2016;Featherall et al.,2019;Gabriel et al.,2019;Gray et al.,2019;Johnson et al.,2019;Kirkpatrick et al.,2015;Kulkarni et al.,2011;Noria et al.,2015;Pelt et al.,2016;Weiss et al.,2019;D. V. Williams et al.,2019)</p> <p>Improve discharge efficiency(Bolz & Iorio,2016;(DiGioia & Greenhouse,2012;Dundon et al.,2016;Ebinger et al.,2018;Featherall et al.,2019;Featherall et al.,2018)</p> <p>Better resource and capacity allocating (Colegate-Stone et al.,2016;Gustafsson et al.,2016;Van Den Berg et al.,2020)</p> <p>Value-office (Hernandez A. et al.,2019;(Makdisse et al.,2018)</p> <p>Improved quality through risk adjustment (Bernstein et al.,2018;Golas et al.,2018;Khullar et al.,2015;Smith et al.,2016;D. V. Williams et al.,2019)</p> <p>Benchmarking (Van Den Berg et al.,2020;van Veghel et al.,2016)</p> |

Figure 3- Radar chart of literature and cases profile in outcome information coverage.



Legend Figure 3: The yellow line represents the proportion of HBS cases in each tier, and the orange line demonstrates the same profile for the systematic review studies. The dotted line represents the studies, in its respective colors, within the financial result category that used micro-costing methods or TDABC.

SDC List:

- Supplemental Digital Content 1: Full search strategies used in each database for the systematic review.
- Supplemental Digital Content 2: Characteristics of the selected studies.
- Supplemental Digital Content 3: Expanded Outcomes Measurement and Drive Assessment in included studies
- Supplemental Digital Content 4: Expanded Outcomes Measurement and characteristics of Harvard Business School (HBS) cases.

Supplemental Digital Content 1. Full search strategies used in each database for the systematic review.

| Base | Query | Hits |
|------------------|--|------|
| Pubmed | ((("value-based health insurance"[MeSH Terms] OR "VBHC"[Title/Abstract]) OR "value health care"[Title/Abstract] OR "value health care"[Title/Abstract]) OR (((("value based"[Title/Abstract] OR "value based"[Title/Abstract]) OR "value driven"[Title/Abstract] OR "value driven"[Title/Abstract]) AND (((("Health"[Title/Abstract] OR "Care"[Title/Abstract]) OR "health care"[Title/Abstract] OR "health care"[Title/Abstract]) OR "Health Insurance"[Title/Abstract] OR "Insurance Design"[Title/Abstract])))) | 3322 |
| Embase | ('value-based insurance design'/exp OR 'vbhc':ab,ti OR 'value health care':ab,ti OR 'value health-care':ab,ti OR (('value based':ab,ti OR 'value-based':ab,ti OR 'value-driven':ab,ti OR 'value driven':ab,ti) AND ('health':ab,ti OR 'care':ab,ti OR 'health care':ab,ti OR 'health-care':ab,ti OR 'health insurance':ab,ti OR 'insurance design':ab,ti))) AND [embase]/lim | 3268 |
| Cochrane CENTRAL | ((MeSH descriptor: [Value-Based Health Insurance] explode all trees) OR ("VBHC" OR "Value Health Care" OR "Value Health-Care")) OR ((("Value Based" OR "Value-Based" OR "Value-driven" OR "Value driven") AND ("Health" OR "Care" OR "Health Care" OR "Health-Care" OR "Health Insurance" OR "Insurance Design"))) | 191 |
| Scopus | TITLE-ABS-KEY ({Value-Based Health Insurance} OR {Value Based Health Insurance} OR {VBHC} OR {Value-Based Health Care} OR {Value Based Health Care} OR {Value driven health care} OR {Value-driven health care}) | 414 |

Supplemental Digital Content 2. Characteristics of the selected studies.

| Refs. | Title | Setting | Origin Country | Publication year | Value initiative | Healthcare Field |
|-----------------------------|---|----------|----------------|------------------|--|------------------|
| (Hernandez A. et al., 2019) | Navy medicine introduces value-based health care | System | USA | 2019 | Traditional value program: measure of clinical, process and financial outcomes | Clinical |
| (Van Deen et al., 2017) | The impact of value-based healthcare for inflammatory bowel diseases on healthcare utilization: A pilot study | Hospital | USA | 2017 | Traditional value program: measure of clinical, process and financial outcomes | Clinical |
| (Abdulla et al., 2012) | Endocrine surgery as a model for value-based health care delivery | System | USA | 2012 | Clinical or surgical pathway redesign | Surgery |
| (Ackerman et al., 2019) | Enhanced REVENUE After Surgery? A Cost-Standardized Enhanced Recovery Pathway for Mastectomy Decreases Length of Stay | Hospital | USA | 2019 | Clinical or surgical pathway redesign | Surgery |
| (Ahn et al., 2019) | Defining and Optimizing Value in Total Joint Arthroplasty From the Patient, Payer, and Provider Perspectives | Hospital | USA | 2019 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Bateni et al., 2019) | Defining Value for Pancreatic Surgery in Early-Stage Pancreatic Cancer | System | USA | 2019 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Berglund et al., 2019) | The procedure value index: a new method for quantifying value in shoulder arthroplasty | Hospital | USA | 2019 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Bernstein et al., 2018) | Higher Quality, Lower Cost with an Innovative Geriatrics Consultation Service | Hospital | USA | 2018 | Traditional value program: measure of clinical, process and financial outcomes | Clinical |
| (Boltz et al., 2019) | Comparing an on-site nurse practitioner with telemedicine physician support hospitalist | Hospital | USA | 2019 | Telemedicine or Computational Intelligence supporting value programs | Clinical |

| | | | | | | |
|-------------------------------|---|----------|-----|------|--|----------|
| | programme with a traditional physician hospitalist programme | | | | | |
| (Bolz & Iorio, 2016) | Bundled Payments: Our Experience at an Academic Medical Center | Hospital | USA | 2016 | Clinical or surgical pathway redesign | Surgery |
| (Chatfield et al., 2019) | Bending the cost curve: Time series analysis of a value transformation programme at an academic medical centre | Hospital | USA | 2019 | Traditional value program: measure of clinical, process and financial outcomes | Mixed |
| (Colegate-Stone et al., 2016) | Increasing value: The King's College Hospital experience | Hospital | UK | 2016 | Clinical or surgical pathway redesign | Surgery |
| (Danilyants et al., 2019) | A value-based evaluation of minimally invasive hysterectomy approaches | Hospital | USA | 2019 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (DiGioia & Greenhouse, 2012) | Care experience-based methodologies: performance improvement roadmap to value-driven health care | Hospital | USA | 2012 | Clinical or surgical pathway redesign | Surgery |
| (Dundon et al., 2016) | Improvement in Total Joint Replacement Quality Metrics: Year One Versus Year Three of the Bundled Payments for Care Improvement Initiative | Hospital | USA | 2016 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Ebinger et al., 2018) | Value-based ST-segment-elevation myocardial infarction care using risk-guided triage and early discharge | System | USA | 2018 | Telemedicine or Computational Intelligence supporting value programs | Clinical |
| (Featherall et al., 2019) | Effects of a Total Knee Arthroplasty Care Pathway on Cost, Quality, and Patient Experience: Toward Measuring the Triple Aim | System | USA | 2019 | Clinical or surgical pathway redesign | Surgery |
| (Featherall et al., 2018) | Implementation of a Total Hip Arthroplasty Care Pathway at a High-Volume Health System: Effect on Length of Stay, Discharge Disposition, and 90-Day Complications | System | USA | 2018 | Clinical or surgical pathway redesign | Surgery |
| (Gabriel et al., 2019) | Value-based healthcare analysis of joint replacement surgery for patients with primary hip osteoarthritis | Hospital | UK | 2019 | Clinical or surgical pathway redesign | Surgery |

| | | | | | | |
|----------------------------|---|----------|-------|------|--|----------|
| (Glotzbach et al., 2018) | Value-driven cardiac surgery: Achieving “perfect care” after coronary artery bypass grafting | Hospital | USA | 2018 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Golas et al., 2018) | A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data | System | USA | 2018 | Telemedicine or Computational Intelligence supporting value programs | Clinical |
| (Goretti et al., 2020) | Value-Based Healthcare and Enhanced Recovery After Surgery Implementation in a High-Volume Bariatric Center in Italy | Hospital | Italy | 2020 | Clinical or surgical pathway redesign | Surgery |
| (Gray et al., 2019) | Bundled Payment "Creep": Institutional Redesign for Primary Arthroplasty Positively Affects Revision Arthroplasty | Hospital | USA | 2019 | Clinical or surgical pathway redesign | Surgery |
| (Gustafsson et al., 2016) | Adherence to the ERAS protocol is associated with 5-year survival after colorectal cancer surgery: a retrospective cohort study. | Hospital | USA | 2016 | Clinical or surgical pathway redesign | Surgery |
| (Iorio et al., 2016) | Early results of medicare’s bundled payment initiative for a 90-day total joint arthroplasty episode of care. | Hospital | USA | 2016 | Clinical or surgical pathway redesign | Surgery |
| (Johnson et al., 2019) | Short-term functional recovery after total joint arthroplasty is unaffected by bundled payment participation | Hospital | USA | 2019 | Clinical or surgical pathway redesign | Surgery |
| (Khullar et al., 2015) | Transthoracic versus transhiatal resection for esophageal adenocarcinoma of the lower esophagus: A value-based comparison | System | USA | 2015 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Kirkpatrick et al., 2015) | Using value-based analysis to influence outcomes in complex surgical systems | System | USA | 2015 | Clinical or surgical pathway redesign | Surgery |
| (Kulkarni et al., 2011) | Clinical pathways improve hospital resource use in endocrine surgery | hospital | USA | 2011 | Clinical or surgical pathway redesign | Surgery |

| | | | | | | |
|--------------------------|---|----------|--------|------|--|---------|
| (Lee et al., 2016) | Implementation of a Value-Driven Outcomes Program to Identify High Variability in Clinical Costs and Outcomes and Association With Reduced Cost and Improved Quality | System | USA | 2016 | Telemedicine or Computational Intelligence supporting value programs | Surgery |
| (Makdisse et al., 2018) | What Is a Value Management Office? An Implementation Experience in Latin America | Hospital | Brazil | 2018 | Clinical or surgical pathway redesign | Mixed |
| (McCreary et al., 2019) | Patient-Level Value Analysis: An Innovative Approach to Optimize Care Delivery | Hospital | USA | 2019 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Navarro et al., 2018) | Machine Learning and Primary Total Knee Arthroplasty: Patient Forecasting for a Patient-Specific Payment Model | System | UK | 2018 | Telemedicine or Computational Intelligence supporting value programs | Surgery |
| (Noria et al., 2015) | The impact of care coaching on hospital length of stay, readmission rates and overall patient satisfaction post-bariatric surgery | Hospital | USA | 2015 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Parker et al., 2017) | Bending the cost curve - Establishing value in spine surgery | Hospital | USA | 2017 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Peard et al., 2019) | Examining and Understanding Value: The Impact of Preoperative Characteristics, Intraoperative Variables, and Postoperative Complications on Cost of Robot-Assisted Laparoscopic Radical Prostatectomy | Hospital | USA | 2019 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Pelt et al., 2016) | Improving value in primary total joint arthroplasty care pathways: changes in inpatient physical therapy staffing | Hospital | USA | 2016 | Clinical or surgical pathway redesign | Surgery |
| (Ravikumar et al., 2010) | A validated value-based model to improve hospital-wide perioperative outcomes: Adaptability to combined medical/surgical inpatient cohorts | System | USA | 2010 | Clinical or surgical pathway redesign | Mixed |
| (Rosner et al., 2018) | Effectiveness of an Automated Digital Remote Guidance and Telemonitoring Platform on Costs, Readmissions, and | System | USA | 2018 | Telemedicine or Computational Intelligence supporting value programs | Surgery |

| | | | | | | |
|-------------------------------|--|----------|-----------------|------|--|----------|
| | Complications After Hip and Knee Arthroplasties | | | | | |
| (Smith et al., 2016) | Cost and Complications of Local Therapies for Early-Stage Breast Cancer | System | USA | 2016 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Thaker et al., 2016) | Defining the Value Framework for Prostate Brachytherapy using Patient-Centered Outcome Metrics and Time-Driven Activity- Based Costing | Hospital | USA | 2016 | Traditional value program: measure of clinical, process and financial outcomes | Clinical |
| (Van Den Berg et al., 2020) | Value-based health care in obstetrics | Hospital | The Netherlands | 2020 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (van Egdom et al., 2019) | Implementation of Value Based Breast Cancer Care | Hospital | The Netherlands | 2019 | Traditional value program: measure of clinical, process and financial outcomes | Clinical |
| (van Veghel et al., 2016) | First results of a national initiative to enable quality improvement of cardiovascular care by transparently reporting on patient-relevant outcomes. | System | The Netherlands | 2016 | Traditional value program: measure of clinical, process and financial outcomes | Surgery |
| (Weiss et al., 2019) | Effective Care Management for Children With Special Health Care Needs in the Era of Value-Based Payment | Hospital | USA | 2019 | Clinical or surgical pathway redesign | Clinical |
| (D. V. Williams et al., 2019) | Impact of an integrated practice unit on the value of musculoskeletal care for uninsured and underinsured patients | Hospital | USA | 2019 | Clinical or surgical pathway redesign | Clinical |
| (J. B. Williams et al., 2019) | One-year results from the first US-based enhanced recovery after cardiac surgery (ERAS Cardiac) program | Hospital | USA | 2019 | Clinical or surgical pathway redesign | Surgery |

Supplemental Digital Content 3. Expanded Outcomes Measurement and Drive Assessment in included studies.

| Reference | Baseline characteristics / risk-adjusted | Tier 1 | | Tier 2 | | Tier 3 | | Financial Outcome | Instruments used (if applied) | Data Origin |
|-----------------------------|--|-----------|---------------------------|--|---|--------------------------|-----------------------------------|--|--|-------------------|
| | | Mortality | Degree of health/recovery | Time to recovery and return to normal activities | Disutility of the care or treatment process | Sustainability of health | Long-term consequences of therapy | | | |
| (Hernandez A. et al., 2019) | yes | no | yes | yes | yes | no | yes | TDABC | ICHOM, HEDIS | EHR |
| (Van Deen et al., 2017) | no | no | no | no | yes | yes | yes | External Database | NA | External Database |
| (Abdulla et al., 2012) | yes | no | no | no | no | yes | no | Hospital Charges | NA | EHR |
| (Ackerman et al., 2019) | yes | no | yes | yes | yes | no | no | Hospital Charges | 0–10 pain intensity numeric rating scale. | EHR |
| (Ahn et al., 2019) | yes | no | yes | no | no | yes | yes | NA | HOOS, KOOS | EHR |
| (Bateni et al., 2019) | yes | yes | no | no | yes | yes | no | Hospital Charges | NA | EHR |
| (Berglund et al., 2019) | yes | no | yes | yes | yes | no | yes | Hospital charges / reimbursement / institutional accounting system | SST, ASES, VAS, SANE, patient satisfaction | EHR |
| (Bernstein et al., 2018) | yes | yes | yes | yes | yes | yes | no | Hospital Charges | AMPAC | EHR |
| (Boltz et al., 2019) | no | yes | no | no | yes | yes | no | Cost of Implementation | HCAHPS | EHR |
| (Bolz & Iorio, 2016) | yes | no | yes | no | yes | yes | no | Reimbursement | NA | EHR |
| (Chatfield et al., 2019) | yes | yes | no | no | yes | yes | no | Microcosting | Experience survey | EHR |
| (Colegate-Stone et al., | yes | yes | no | yes | yes | no | no | TDABC | NA | EHR |

| | | | | | | | | | | |
|------------------------------|-----|-----|-----|-----|-----|-----|-----|---------------------------------|-------------------|-------------------|
| 2016) | | | | | | | | | | |
| (Danilyants et al., 2019) | yes | no | yes | no | yes | yes | yes | TDABC | Experience survey | EHR |
| (DiGioia & Greenhouse, 2012) | no | yes | yes | yes | yes | no | no | NA | HCAHPS | EHR |
| (Dundon et al., 2016) | yes | no | yes | no | yes | yes | no | Reimbursement | NA | EHR |
| (Ebinger et al., 2018) | yes | yes | yes | no | yes | no | no | Microcosting | NA | EHR |
| (Featherall et al., 2019) | yes | no | yes | no | yes | yes | no | Microcosting | HCAHPS | EHR |
| (Featherall et al., 2018) | yes | no | yes | no | yes | yes | no | External Database | NA | EHR |
| (Gabriel et al., 2019) | yes | yes | yes | yes | yes | no | yes | Microcosting | OHS, EQ-5D, VAS | EHR |
| (Glotzbach et al., 2018) | yes | yes | yes | yes | yes | no | no | Microcosting | NA | EHR |
| (Golas et al., 2018) | yes | no | no | no | yes | yes | no | External Database | NA | EHR |
| (Goretti et al., 2020) | yes | Microcosting | BAROS | EHR |
| (Gray et al., 2019) | yes | no | yes | yes | yes | yes | no | Institutional accounting system | NA | EHR |
| (Gustafsson et al., 2016) | yes | yes | yes | no | yes | yes | no | NA | NA | EHR |
| (Iorio et al., 2016) | no | no | yes | no | yes | yes | no | Reimbursement | NA | EHR |
| (Johnson et al., 2019) | yes | no | yes | yes | yes | no | yes | NA | AMPAC, PF-CAT | EHR |
| (Khullar et al., 2015) | yes | yes | no | no | yes | yes | no | External Database | NA | External Database |
| (Kirkpatrick et al., 2015) | no | yes | no | no | yes | yes | no | External Database | NA | EHR |
| (Kulkarni et | yes | no | yes | no | yes | yes | no | Institutional | NA | EHR |

| | | | | | | | | | | |
|-----------------------------|-----|-----|-----|-----|-----|-----|-----|---------------------------------|------------|-------------------|
| al., 2011) | | | | | | | | accounting system | | |
| (Lee et al., 2016) | yes | yes | no | yes | yes | yes | yes | DRG | PROMIS | External Database |
| (Makdisse et al., 2018) | yes | no | no | yes | yes | yes | yes | TDABC | ICHOM | EHR |
| (McCreary et al., 2019) | yes | no | yes | yes | yes | no | no | TDABC | SMFA | EHR |
| (Navarro et al., 2018) | yes | no | no | no | yes | no | no | Reimbursement | NA | External Database |
| (Noria et al., 2015) | yes | yes | no | yes | yes | yes | no | NA | HCAHPS | EHR |
| (Parker et al., 2017) | yes | no | yes | yes | yes | yes | yes | Microcosting | ODI, EQ-5D | EHR |
| (Peard et al., 2019) | yes | no | yes | no | yes | yes | no | Institutional accounting system | NA | EHR |
| (Pelt et al., 2016) | yes | no | yes | yes | yes | no | no | Institutional accounting system | ASA | EHR |
| (Ravikumar et al., 2010) | yes | yes | no | no | yes | yes | no | DRG | NA | EHR |
| (Rosner et al., 2018) | yes | no | no | no | yes | yes | no | Institutional accounting system | NA | External Database |
| (Smith et al., 2016) | yes | no | no | yes | yes | yes | no | External Database | NA | External Database |
| (Thaker et al., 2016) | yes | TDABC | ICHOM | EHR |
| (Van Den Berg et al., 2020) | yes | yes | yes | yes | yes | no | no | NA | ICHOM | EHR |

| | | | | | | | | | | |
|-------------------------------|-----|-----|-----|-----|-----|-----|-----|---------------|---|-------------------|
| (van Egdom et al., 2019) | yes | NA | EORTC-QLQ-C30, EORTC-QLQ-B23, BREAST-Q , EQ-5D-5L, RCS-NL, CarerQoL-7D, Distress Thermometer, and Experience survey | EHR |
| (van Veghel et al., 2016) | yes | yes | yes | no | yes | yes | yes | NA | SF-36 | External database |
| (Weiss et al., 2019) | yes | no | yes | no | yes | yes | no | Reimbursement | NA | EHR |
| (D. V. Williams et al., 2019) | no | no | yes | yes | yes | yes | yes | NA | HOOS, KOOS, PROMIS | EHR |
| (J. B. Williams et al., 2019) | yes | no | yes | no | yes | yes | no | NA | NA | EHR |

Supplemental Digital Content 4. Expanded Outcomes Measurement and characteristics of Harvard Business School (HBS) cases.

| Ref | Setting | Origin Country | Publication year | Value initiative | Healthcare Field | Baseline characteristics | Mortality | Degree of health/recovery | Time to recovery | Disutility of the care or treatment process | Sustainability of health | Long-term consequence of therapy | Financial Outcome | Instruments used (if applied) | Data Origin |
|--|----------|----------------|------------------|--|------------------|--------------------------|-----------|---------------------------|------------------|---|--------------------------|----------------------------------|-------------------------------|------------------------------------|-------------|
| (Porter, Deerberg - Wittram, et al., 2014) | hospital | Germany | 2014 | Traditional value program: measure of clinical, process and financial outcomes | surgery | yes | yes | yes | yes | yes | yes | yes | NA | QLQ0C30, IIEF-5, ICHOM; EPIC-26 | EHR |
| (Kaplan et al., 2012) | hospital | Germany | 2012 | Traditional value program: measure of clinical, process and financial outcomes | surgery | yes | yes | yes | yes | yes | yes | yes | TDABC | EQ5D, WOMAC | EHR |
| (Kaplan et al., 2020) | hospital | USA | 2020 | Clinical or surgical pathway redesign | surgery | yes | no | yes | no | yes | yes | yes | Reimbursement | VAS, KOOS, HOOS, PROMIS PF. | EHR |
| (Porter, Marks, et al., 2014) | system | Sweden | 2014 | Clinical or surgical pathway redesign | surgery | yes | no | yes | yes | yes | yes | yes | Reimbursement | EQ-5D, PROMIS Global 10, PROMIS PF | EHR |
| (Porter, Bachman, et al., 2014) | hospital | USA | 2014 | Clinical or surgical pathway redesign | surgery | yes | yes | yes | yes | yes | yes | yes | Institution Accounting System | BSID-II, PDI, MDI | EHR |

| | | | | | | | | | | | | | | | |
|---------------------------------|--------|-----------------|------|--|----------|-----|-----|-----|-----|-----|-----|-----|---------------|-------------------|-----|
| (Porter et al., 2016) | system | USA | 2016 | Clinical or surgical pathway redesign | clinical | yes | yes | yes | no | yes | yes | yes | Reimbursement | HEDIS, and PROs | EHR |
| (Kaplan et al., 2018) | system | The Netherlands | 2018 | Clinical or surgical pathway redesign | surgery | yes | Reimbursement | PROMIS | EHR |
| (Porter et al., 2007) | system | Germany | 2007 | Clinical or surgical pathway redesign | clinical | yes | no | yes | yes | yes | yes | yes | Reimbursement | PROMIS | EHR |
| (Porter & Teisberg, 2009) | system | USA | 2009 | Traditional value program: measure of clinical, process and financial outcomes | clinical | yes | yes | no | no | yes | yes | yes | Reimbursement | PROMIS | EHR |
| (Porter, Landman, et al., 2014) | system | USA | 2014 | Clinical or surgical pathway redesign | clinical | yes | TDABC | NOSES | EHR |
| (Porter et al., 2017) | system | USA | 2017 | Clinical or surgical pathway redesign | clinical | yes | no | yes | yes | yes | yes | no | Reimbursement | HEDIS | EHR |
| (Porter et al., 2019) | system | USA | 2019 | Traditional value program: measure of clinical, process and financial outcomes | clinical | yes | no | yes | yes | yes | no | yes | TDABC | ICHOM, HEDIS, ODI | EHR |

5.2. ARTIGO 2

An artificial intelligence solution for outcomes classification to scale a value-based healthcare program

Bruna Stella Zanoto^{1,2}, Ana Paula Beck da Silva Etges^{1,3}, Avner dal Bosco³, Eduardo Cortes⁴, Renata Ruschell¹, Sheila Ouriques Martins⁵, Ana Claudia Souza⁵, Claudio Valiense⁶, Felipe Viegas⁶, Sergio Canuto⁶, Washington Luiz⁶, Renata Vieira³, Marcos Gonçalves⁶, Carisi Anne Polanczyk^{1,2}

1 – National Institute of Health Technology Assessment - INCT/IATS (CNPQ 465518/2014-1), Universidade Federal do Rio Grande do Sul - Porto Alegre (RS), Brasil.

2- Graduate Program in Epidemiology, Universidade Federal do Rio Grande do Sul, Porto Alegre (RS), Brazil

3- Pontificia Universidade Católica do Rio Grande do Sul, Porto Alegre, BR 90619-900

4- Graduate Program of Computer Science, Universidade Federal do Rio Grande do Sul, Porto Alegre (RS), Brazil

5- Neurology Service, Rede AVC Brasil, Hospital Moinhos de Vento, Porto Alegre (RS), Brazil.

6- Universidade Federal de Minas Gerais, Belo Horizonte (MG), Brazil.

Journal cotado: Lancet Digital Health

RESUMO

Introdução: Com a rápida adoção de registros eletrônicos de saúde (RES), é desejável que se consiga coletar informações e conhecimento que possam dar suporte à gestão centrada no paciente e baseada em valor. Esta pesquisa tem como objetivo apresentar um framework com os melhores métodos para automatizar a medição de resultados de saúde e a classificação das características dos pacientes pelos dados de texto livre de RES para apoiar um programa baseado em valor para a linha de cuidado do AVC.

Métodos: O problema computacional é de extração de informação e classificação de textos. Variáveis e recursos que podem ser considerados importantes para um programa de gestão baseado em valor foram selecionados. Foram usadas 44206 sentenças de prontuários médicos em português para treinar e desenvolver algoritmos. Foram usados 11 modelos computacionais diferentes baseados em aprendizado de máquina e regras criadas por especialistas e avaliamos seu desempenho com base na pontuação do macro score F1.

Resultados: Os três modelos que apresentaram melhor desempenho na extração de dados em texto livre em português do RES foram Word + Character TDIDF + SVM, Linear SVM + BoW e metafeatures, cobrindo um total de 14 tarefas com Macro F1 > 80. A ontologia apresentou desempenho superior a outros algoritmos para múltiplas tarefas.

Conclusão: Avanços em capacidade tecnológica informacional são essenciais para a escalabilidade e agilidade da capacidade de medição de resultados em saúde. Este estudo permitiu sugerir as melhores soluções de IA para automatizar a classificação de desfechos de tarefas específicas da condição clínica do AVC.

ABSTRACT

Background: With the rapid adoption of electronic health records (EHRs), it is desirable to harvest information and knowledge from EHRs to support patient-centered and value-based management. This research aims to present a framework with the best methods to automate the measurement of health outcomes and the classification of patient characteristics by the free text data of EHR to support a value-based program for the stroke care line.

Methods: The study is a computational problem of information extraction and text classification. Variables and features that may be considered for a value-based program on stroke were selected. Electronic medical record data from 188 patients admitted for acute stroke were identified, and 44,206 sentences were used from free-text records in Portuguese to train and developed algorithms. We used 11 different computational models based on machine learning and rules created by specialists and evaluated their performance based on macro F1-score.

Findings: The three models that showed the best performance in extracting free text data in Portuguese from the RES were Word + Character TDIDF + SVM, Linear SVM + BoW and metafeatures, covering a total of 14 tasks with Macro F1 > 80. The ontology presented superior performance to other algorithms for multiple tasks.

Interpretation: Advances in information technology capacity are essential for the scalability and agility of the ability to measure health outcomes. This study allowed to suggest the best AI solutions to automate the classification of outcomes of specific tasks of the clinical condition of the stroke.

INTRODUCTION

The last decade is marked by the increase of incentive programs to hospitals and ambulatory care providers to the adherence of electronic health records (EHRs) and the purported benefits of its implementation, including immediate access to patient information ¹. However, in agreement with Blumenthal et al. (2019), the implementation is not enough to achieve the meaningful use of EHRs. While its use has simple, pure administrative purposes of informing and register conduct, we will still not improve caregivers' decisions and patient outcomes. By the time we propose to demystify all medical conditions and its consequences, urging healthcare organizations to develop manners of capturing data will allow them to assess the quality of service delivery to patients and the real costs of providing these results ².

Outcome measurement is perhaps the most potent tool for revamping the healthcare system ³. Useful data on outcomes and costs is one of the most urgent to designing and optimizing value-based care⁴, unfortunately, the absence of adequate tools for patient-level outcome follow-up turns value-focused work generally gone undetected thus unrewarded ^{5,6}.

Stroke is the second leading cause of mortality and disability-adjusted life years in the world ^{7,8}. The complexity of a condition such as a stroke potentially lends itself well to the use of machine learning methods, which can incorporate a large variety of variables and observations into one predictive framework without the need for preprogrammed rules. Therefore, there has been an increasing interest in the use of ML to track stroke outcomes, with the hope that such methods could make use of large, routinely collected datasets and deliver accurate, personalized prognoses. However, studies applying machine learning methods to stroke, while published regularly, the main focus of these has been on stroke imaging application ⁹⁻¹¹, and structured data retrieval¹².

The information technology gap between automating data collection from electronic medical records has been described in the literature as a decelerator of value initiatives ^{4,13,14}. A promising approach is the incorporation of Natural Language Processing (NLP) in clinical outcomes evaluation studies¹⁴⁻¹⁶. The availability of a robust information technology platform to process information from the free-text medical records can accelerate the adoption of value-based health management. Efforts from different expertise must transpose and reinvent the EHR not as a document but as a system that supports the generation and tracking of multiple events, processes, and

patient outcomes⁴.

Having value for the patient as the central guide of a requires advances on strategies to automatize the capturing data that will allow them to assess the quality-of-service delivery to patients and the real costs of providing these results^{1,2}. This research aims to present a framework suggesting the best methods to automatize outcomes and patients' characteristics classification by the free-text data from the EHRs to support a value-based program. Our specific aim is to identify the outcomes and patients' clinical characteristics that may be considered in a value-care program for ischemic stroke.

METHODS

The study is a computational problem of information extraction and text classification. It is divided into five stages, and study architecture is demonstrated in **Figure 1**: (i) Variables and features a selection of clinical characteristics and outcomes; (ii) Extraction and processing of medical records data; (iii) Development of computational models based on machine learning and rules created by specialists; (iv) Statistical Analysis; and (v) Framework development

Data Sources

We retrospectively built a medical records database covering two years of patients hospitalized for ischemic stroke from a digital hospital system. The hospital in is a private institution of excellence in the Southern Brazil. The EHR system used is MV (Porto Alegre, Brazil). Since 2017, the hospital has introduced the ICHOM standard sets' data collection in its routine for different clinical pathways and created the institutional value office. For the stroke pathway, data collection started on October 15, 2015. In 2019, the hospital incorporated the Angel Awards Program¹⁷, being certified with the Platinum category at the end of the first year.

Variables and Features

In an attempt to translate the VBHC theory into health system operations, literature review and multidisciplinary expert interviews (n=8) were used to define specific outcomes dimensions, and measures that may be considered in a value-based healthcare program for ischemic stroke patients. The outcome identification step was based on following the value agenda elements dimensions to cover the Outcome Hierarchy's tiers such as functionality dimensions there process of recovery, and outcomes that matter to patients¹⁷. Although overall clinical outcomes or hospital quality measures do not capture dimensions highly crucial to the patient, they represent an important role in institutional sustainability regarding care quality. Therefore, they

were also considered to be extracted from the electronic medical record. These dimensions may include risk events, healthcare status achieved, scales discussing stroke outcomes (National Health Stroke Scale [NIHSS], Rankin, strength scales)^{18,19}.

Extraction and processing of medical records data

Patients' medical records went through a pre-processing that included the Spacy python library²⁰ to stratified texts into sentences. We produce a total of 44206 sentences derived from 188 patients. The approach was used as a hypothesis of managing unbalanced data like electronic health records²¹⁻²³. During the text stratification process, the Spacy²⁰ employs rules-based algorithms that set the sentence edges according to characters' patterns delimiting its beginning and end. Also, the names of patients and the medical staff were de-identification, representing the removal of confidential information from the database. They were done through a regular expression algorithm that allows searching for writing patterns to replace specific passages or characters in the texts.

Model development

We employed different supervised models selected from literature due to their performance in text classification tasks. To address that, two independent annotators registered data categorically classified relevant information for each variable in each sentence. In case of disagreement, a commission compound of researchers was retrieved to consensus. **Figure 1** summarizes the methods used.

In terms of text representations, we compare traditional term-weighting alternatives (TFIDF), weighted by word and character and recent representations based on MetaFeatures that have obtained state-of-the-art effectiveness in some of the experimented datasets²⁴⁻²⁶. As classification algorithms, we exploit SVM, which is still most robustest non-neural network text classification algorithm, using all the representations mentioned above²⁷. In terms of neural networks, we confirmstrong representatives of the most important neural network architectures applied to ATC, including convolutional neural network (CNN)²⁸, and the state-of-art transformer architectures: Bidirectional Encoder Representations from Transformers (BERT)²⁹.

As presented in the study architecture, such methods are divided into two categories: 2-phases (2P) methods and end-to-end (E2E) methods. The first category of methods consists of approaches whose document representation is intrinsically independent of the class prediction phase's classification method. In other words, the classifier used to predict the class of the documents is not used in the phase of

construction of the document representation.

In contrast, E2E methods use a discriminative classifier function to transform the document representation space into a new and more informed (usually more reduced and compact) space, using this classifier to predict the document class. In general, these approaches use an iterative process of representation, classification, evaluation, and parameter adaptation (for instance: transform, predict, evaluate loss function, and backpropagate, respectively).

The rules-based models were developed through logical conditions implemented in an Ontology Web Language (OWL) format, built with specialists who, through the inference processes, identified clinical variables in medical record texts^{30,31}.

Each task results in an individual model. For the model's validation, we use a 5-fold cross-validation and stratification. Information from the same patient was always assigned to the same fold to test the model's ability to predict new data that was not used in the estimate, signaling problems such as overfitting or selection bias³². This validation allows us to train and test all the annotated dataset, reducing bias and bring more reliable results.

Model performance and Statistical Analysis

To evaluate the ability to classify the relevant Portuguese medical records free-text correctly, we used F₁-score (**Equation 1**). This metric is based on a "confusion matrix" and is defined as:

$$F_1 = \frac{PRECISION * RECALL}{PRECISION + RECALL} = \frac{2TP}{2TP + FP + FN} \quad (1)$$

, where:

TP = true positive

TN = true negative

FP = false positive

FN = false negative

The macro-F₁ measures each class's classification effectiveness by averaging them, being one of the most popular aggregated evaluation metrics for classifier evaluation is skewed datasets^{33,34}.

To compare the average results on our cross-validation experiments, we assess the statistical significance employing the paired t-test with 95% confidence^{35,36}. To account for multiple tests, we adopt the Friedman-Nemenyi-Test³⁷ for multiple

comparisons of mean rank sums. The Friedman test is used to compare multiple methods. It is a non-parametric statistical test similar to the parametric repeated measures ANOVA. Friedman's test is used to detect differences in treatments across multiple test attempts^{37,38}, and Bonferroni correction was adopted. We understand ontology methods could not be ranked with other models described above that require the combination of text representations with trained classifying algorithms, so we compare ontology methods with the three best models.

Framework development

Finally, the experiment aimed to provide relations among the supervised models, text representation strategies, and tasks. The framework potential is to boost and contribute to the adequacy of parameters that agree with what is sought in value-based management, and the adoption of indicators that allow their control. It may allow to observe which combinations of models' representation strategies can be recommended for each task.

Framework was developed as a heatmap, which adopted color red for $F1 < 20$, a gradual color scale from orange to yellow to $21 < F1 < 79$, and green for $F1 > 80$. Tasks (represented by the lines) were ordered by the average of the performed models, while the columns ordering shows the rank position on each method accordingly to the statistical analysis.

RESULTS

Variables and Features

For any medical condition, multiple outcomes collectively define success. The set of outcomes is invariably broad, ranging from immediate procedural outcomes to longer-term functional status, recovery time, complications, and recurrences³⁹. We defined 30 tasks presented in **Table 1** that were considered feasible to be extracted from EHR, including two levels suggested by Porter et al.⁴⁰: healthcare status achieved (Tier 1); and recovery (Tier 2). For the first one, the standard sets usually used to evaluate clinical stroke outcomes were used, including the Rankin¹⁸ and the National Health Stroke Scale (NIHSS) scales¹⁹, in addition to traditional outcomes such as mortality and pain level. For Tier 2, the ICHOM standard set already developed for ischemic stroke was used^{17,41}, considered measures of mobility, ability to communicate, ability to feed orally, the ability to understand, and measures and scales for strength level. Indicators of the hospitalization process of care used in the institution were also included^{42,43}, such

as rating scales and risk events tracked by fall risk, pressure ulcer risk, fall event during inpatient, infection indicative, intracranial hemorrhage, therapy care (thrombolytic, thrombectomy or both), and location of the patient during the inpatient path⁴⁴. Finally, baseline characteristics important for tracking population and further risk-adjusted analysis were included ^{17,42,45}, such as high blood pressure, smoking status, coronary artery disease, atrial fibrillation, diabetes, prior stroke, active cancer, alcoholism, obesity, dyslipidemia.

Each task could have two or more classes of an answer, depending on the sentence meaning. For example, the task "smoking status" has four classes. The sentence could indicate the patient was an active smoker, a former smoker, a non-smoker, or even the sentence could a non-informed (NI) sentence related to the patient's smoking status. The classes are also reported in **Table 1**.

Data Source and Dataset Characteristics

The patients' descriptive characteristics, including previous comorbidities, NIHSS, and clinical care, are presented in **Table 2**.

The details for each task in terms of class and sentence distribution demonstrate the dataset's unbalanced nature, which means it has many more sentences corresponding to the NI class. For example, according to **Supplementary Material 1**, "thrombolytic therapy" has three classes, the minor represented class appeared in 107 sentences, and the primary represented class in 43708 sentences. The same table informs the sentence distribution in words for each task, for the same example, the median words per annotate sentence are 13, and the mean is 14.4.

Model performance and framework development

Results for macro F1 for each of the 30 tasks and 10 compared models are shown in **Table 3**. Colors represent the F1 performance of each model for each task gradually, being red the worse (F1 score < 20), and green the better (F1 score > 80), and results in bold correspond to the best statistical in each given task for the specific performance metric with 95% confidence. Word+Character TFIDF and BoW were the overall bests text representations methods used in this dataset. They both generate used term weighting representations to be used alongside SVM classifier. Metafeatures was on average, the third best placed model to cover more tasks with good performance. Although the methods mentioned were not the winner in any task, it often got a

consistent high performance in most of them.

Thirteen tasks have at least three models that performed better than 80% in macro F1. For other tasks, it is noteworthy that some models are much superior to the others tested, for example, "Pressure Ulcer Risk" and "ability to ambulate" has two models that outstands and could be invested for further increase performance. The tasks with best results and with more coverage in all models were "ability to feed orally," "location", "fall risk", "diabetes," "paresis," and "thrombolytic therapy."

Table 4 provides the information of performance compared with ontology methods. It highlights that most of the best performed tasks in ontology do not coincide with the best used by the models in **Table 3**. This is noticing in patient's comorbidities category, which has a very good resolution with ontology classification models. One hypothesis is due to characteristics of the dataset, while patients' comorbidities can be more easily identified by rules than tasks that are identify by description and numbers scales, for example, NIHSS. **Figure 5** shows a subsampling of the best tasks performed by ontology basis ("rankin", "alcoholism", "atrial fibrillation", "dyslipidemia") compared with their results in the 3 top ranked models, to reinforce the potential of ontology to tasks neglected by other models.

Nowadays, outcome evaluation is based on tailor electronic health records review and questionnaires applied to patients, providers, and families, requiring exhaustive and time-consuming work. These results indicate an advance in automation and in informational capacity to make possible outcomes classification in VBHC initiatives.

DISCUSSION

For any system to be successful, it urges the need to track performance through metrics. In the health area, this has always been a challenge due to the amount of data available, with a very heterogeneous, non-standardized base register and culturally done in printed format.

Our experiment intended to recognize the path we may advance in terms technological capacity to support an VBHC program. The success of VBHC implementation starts on the establishment of the correct outcomes to be measured and the intelligence to analyze them in a way to support decisions in healthcare⁴⁶. We identified which models better performed in extracting Portuguese free-text data from the patient's EHR. The three best models were Word+Character TDIDF+SVM, Linear SVM+BoW and Metafeatures, covering a total of 14 tasks with Macro F1 > 80.

Ontology surprisingly went very well in terms of performance, being even superior to other algorithms. There was an interesting unbalance in tasks performance models, which may indicate the usage of different models' usage to capture tasks.

To truly improve the quality of care and ultimately deliver value to patients, outcomes must be routinely measured and reported at the level of the medical condition⁴⁷⁻⁵⁰. Increasingly, this entails use of an electronic health record. There are limited reports in the literature about how to integrate patient-centered outcomes measurement into the unstructured data from EHR⁵¹. This reflection leads us to understand that VBHC is not feasible without investment in information technology⁴⁷.

Once there are metrics monitoring what is happening with each patient since his starting point of the treatment until the discharge day, the decision-making process is fed by rich information to guide assertive decisions for quality monitoring and improvement⁵², and might facilitate process improvement and value⁵³. Efforts to adapt an automation of outcome data collection creates the possibility for benchmarking and comparative effectiveness research⁵⁴.

Due to its rule-based nature, ontology models have better performance on comorbidities than other models^{30,55}. This suggests ontology methods to be used for this kind of task and the potential to be used in VBHC initiatives of other clinical conditions such as cardiovascular and endocrine⁵⁶⁻⁵⁸. A mixed-method use of top 3 ranked models plus ontology should cover a considerable part of tasks considered in such value-based program.

Compared to structured data, free text is a more natural and expressive method to document clinical events and facilitate communication among the care team in the health care environment. One critical component, however, is find ways to facilitate the use of EHR data for clinical decision support and quality improvement, which automatically extracts and encodes clinical information from text. Healthcare operators and hospitals can make use of artificial intelligence use of mixed methods for implementing technology in the models of tracking outcomes and reimbursement⁵⁹. Advances of this nature are important when considering use on a large scale such as in hospitals or health care providers.^{15,16,27} The results above show that the use of mixed methods can turn feasible tracking patients' outcomes and feeding hospital indicators.

The establishment of automation that support the control of the patient's outcomes during stroke episode, according to patients' risks, outcomes and conditions to each therapy, make value-based healthcare ensure its purpose⁴⁰. Proposing new

reimbursement models or pathway redesign have the opportunity be sustained once we provide advances to facilitate the outcomes extraction for the ischemic stroke.

The development of clinical artificial intelligence solutions requires NLP experts to work closely with clinicians who can provide the necessary domain knowledge²⁷. Technological capacity exists to support the adoption of programs to monitor outcomes in real time, but integration between expertise is essential to make it applied.

LIMITATIONS

The study has limitations. For clinical NLP method development to advance further globally and, for example, become an integral part of clinical outcomes research or have a natural place in clinical practice, there are still challenges ahead. Our work is based on the EHR of a single center, with necessarily limited numbers of patients annotated. Future work is needed to test this approach in EHRs from different centers with different patients, who may use different language for clinical documentation. We have no access to exams or hospital indicators, that is the reason infection identification (for example) could be any report of antibiotics use, typical symptoms of infection, and tests described.

We should acknowledge that although we have used cross validation methods, the characteristic of an unstructured real-life database is very unbalanced, this weighs a lot for some tasks that have many classes such as NIHSS or Rankin, or tasks that happened less frequently in our sample, such as patients that have a falling event or indicative of active cancer. It even should be considered that few tasks are not reported in the EHR on the basis of a value-based program.

CONCLUSION

Efforts in machine learning models built on EHR-based data can be used to determine proxies for value-based initiatives. This potentializes for real-time outcomes measurement to be incorporated in studies of stroke patient outcomes using Portuguese unstructured free-text EHR databases. Future directions should explore alternative annotation approaches for increase tasks performance and how it reflects on its external validation.

REFERENCES

1. Blumenthal D, Tavenner M. The “Meaningful Use” Regulation for Electronic Health Records. *N Engl J Med*. 2010;363(6):501-4.
2. Lee TH. Putting the Value Framework to Work. *N Engl J Med*. 2010;363(26):2481-3.
3. Porter ME, Teisberg EO. Redefining health care: creating value-based competition on results. Boston, Mass: Harvard Business School Press; 2006 2006. 506 p.
4. Glaser J. It’s Time for a New Kind of Electronic Health Record. *Harvard Business Review*. 2020.
5. L D, T L. Health Care Needs Real Competition. . *Harvard Business Review*. 2016: 76–87.
6. D K, D L, O S. The Big Idea: Before You Make That Big Decision. *Harvard Business Review*. 2011:50-60.
7. Johnson CO, Nguyen M, Roth GA, Nichols E, Alam T, Abate D, et al. Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*. 2019; 18: 439–458. [https://doi.org/10.1016/S1474-4422\(19\)30034-1](https://doi.org/10.1016/S1474-4422(19)30034-1) PMID: 30871944.
8. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017. Institute for Health Metrics and Evaluation (IHME); 2018. Available: <http://ghdx.healthdata.org/gbd-results-tool>.
9. Kamal H, Lopez V, Sheth SA. Machine Learning in Acute Ischemic Stroke Neuroimaging. *Front Neurol*. 2018; 9: 945. <https://doi.org/10.3389/fneur.2018.00945> PMID: 30467491.
10. Feng R, Badgeley M, Mocco J, Oermann EK. Deep learning guided stroke management: a review of clinical applications. *J NeuroIntervent Surg*. 2018; 10: 358–362. <https://doi.org/10.1136/neurintsurg-2017-013355> PMID: 28954825.
11. Lee E-J, Kim Y-H, Kim N, Kang D-W. Deep into the Brain: Artificial Intelligence in Stroke Imaging. *J Stroke*. 2017; 19: 277–285. <https://doi.org/10.5853/jos.2017.02054> PMID: 29037014.
12. Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, et al. (2020) A systematic review of machine learning models for predicting outcomes of stroke with structured data. *PLoS ONE* 15(6): e0234722. <https://doi.org/10.1371/journal.pone.0234722>.
13. Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Medical Informatics and Decision Making*. 2018;18(1):44.
14. Porter ME, Lee TH. The Strategy That Will Fix Health Care. *Harvard Business Review*. 2013;91(10):50-70.
15. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *npj Digital Medicine*. 2020;3(1).

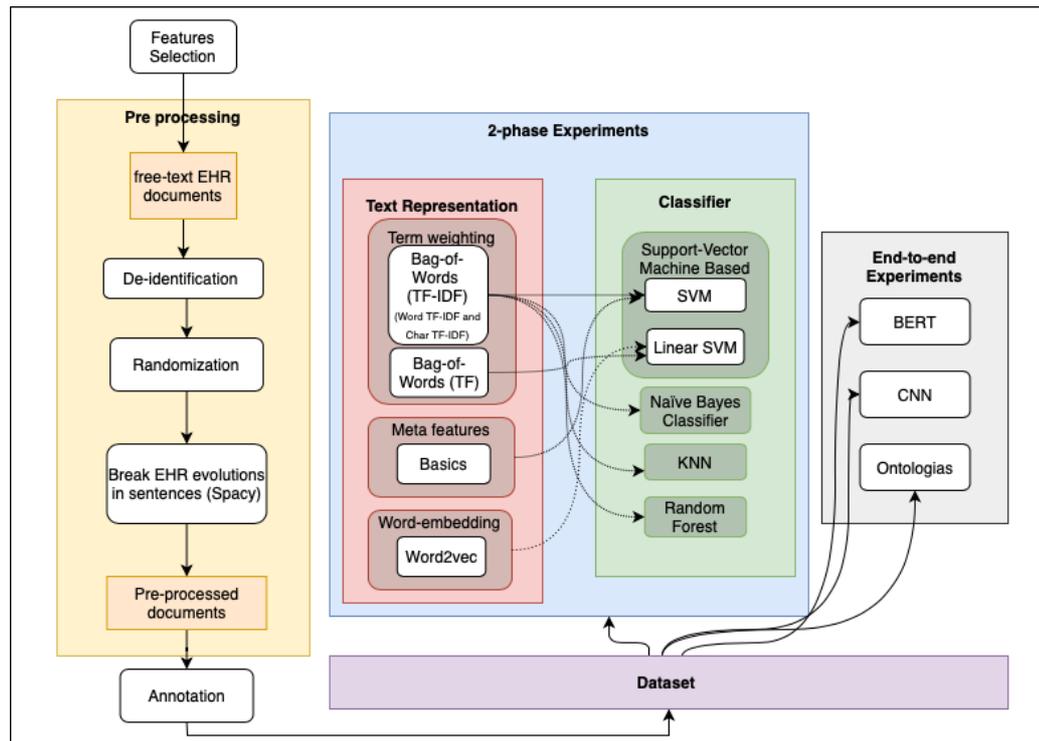
16. Meulen Mvd. Artificial Intelligence as a Driver of Value in Value Based Health Care Systems. 2019.
17. Salinas J, Sprinkhuizen SM, Ackerson T, Bernhardt J, Davie C, George MG, Gething S, Kelly AG, Lindsay P, Liu L. An international standard set of patient-centered outcome measures after stroke. *Stroke*. 2016;47(1):180–6. .
18. Wilson JL, Hareendran A, Grant M, Baird T, Schulz UG, Muir KW, Bone I. Improving the assessment of outcomes in stroke: use of a structured interview to assign grades on the modified Rankin Scale. *Stroke*. 2002;33(9):2243–6. .
19. Lyden PD, Lu M, Levine S, Brott TG, Broderick J. A modified National Institutes of Health Stroke Scale for use in stroke clinical trials. *Stroke*. 2001;32(6):1310–7. .
20. Honnibal M, Montani I. spaCy. Explosion AI; 2015. p. spaCy is an open-source software library for advanced natural language processing, written in the programming languages Python and Cython.
21. Manning C, Schütze H. Foundations of statistical natural language processing. MIT Press: Cambridge, MA.; 1999.
22. Manning C, Raghawan P, Schütze H. An introduction to information retrieval. Cambridge, U.K.: Cambridge University Press; 2009.
23. Klie J-C, Bugert M, Boullosa B, Eckart de Castilho R, Gurevych I. The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In Proceedings of System Demonstrations of the 27th International Conference on Computational Linguistics (COLING 2018): Santa Fe, New Mexico, USA 2018.
24. Canuto S, Salles T, Gonçalves MA, Rocha L, Ramos G, Gonçalves G, et al. On Efficient Meta-Level Features for Effective Text Classification. In Proceedings of the 23rd ACM International Conference on Information and Knowledge Management -- CIKM2014. p. p. 1709--18.
25. S. Canuto, T. Salles, T. C. Rosa, M. A. Goncalves, Similarity-based synthetic document representations for meta-feature generation in text classification, in: Proceedings of the 42nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2019, pp. 355{364.
26. S. C, X. SD, A. GM, C. RT. A Thorough Evaluation of Distance-Based Meta-Features for Automated Text Classification. *IEEE Transactions on Knowledge and Data Engineering*. 2018;30(12):2242-56.
27. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. *J Biomed Inform*. 2018;77:34-49.
28. X. Zhang, J. Zhao, Y. LeCun, Character-level convolutional networks for text classification, in: Advances in Neural Information Processing Systems, Vol. 28, 2016, pp. 649{657.
29. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL), 2019, pp. 4171{4186.
30. Wang B, McKay R, Abbass H, Barlow M. A Comparative Study for Domain Ontology Guided Feature Extraction 2003. 69-78 p.
31. Liyanage H, Krause P, De Lusignan S. Using ontologies to improve semantic interoperability in health data. *J Innov Health Inform*. 2015;22(2):309-15.
32. Cawley GC, Talbot NLC. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research*.

2010:2079–107.

33. Kowsari, Jafari M, Heidarysafa, Mendu, Barnes, Brown. Text Classification Algorithms: A Survey. *Information*. 2019;10(4).
34. Velupillai S, Suominen H, Liakata M, Roberts A, Shah AD, Morley K, et al. Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances. *J Biomed Inform*. 2018;88:11-9.
35. D. Hull, Using statistical testing in the evaluation of retrieval experiments, in: *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 1993, pp. 329{338.
36. Y. Hochberg, A sharper bonferroni procedure for multiple tests of significance, *Biometrika* 75 (4) (1988).
37. J. H. Zar, *Biostatistical Analysis* (5th Edition), Prentice-Hall, Inc., 2007.
38. Cunha W, Mangaravite V, Gomes C, Canuto Se, Elaine Resendea CN, Viegas F, et al. On the Cost-Eectiveness of Neural and Non-Neural Approaches and Representations for Text Classification: A Comprehensive Comparative Study (no prelo). *Information Processing & Management*. 2020.
39. Porter ME. What Is Value in Health Care? *N Engl J Med*. 2010;363(26):2477-81.
40. Porter ME, Larsson S, Lee TH. Standardizing Patient Outcomes Measurement. *N Engl J Med*. 2016;374(6):504-6.
41. Bonnie L. Callen, Jane E. Mahoney, Thelam J. Wells, Myra Enloe, Hughes S. Admission and Discharge Mobility of Frail Hospitalized Older Adults. *MEDSURG Nursing*. 2004;13:164.
42. Feigin VL, Krishnamurthi R. Stroke is largely preventable across the globe: where to next? *The Lancet*. 2016;388(10046):733-4.
43. Martinez MC, Iwamoto VE, Latorre Mdo R, Noronha AM, Oliveira AP, Cardoso CE, et al. Transcultural adaptation of the Johns Hopkins Fall Risk Assessment Tool. *Rev Lat Am Enfermagem*. 2016;24:e2783.
44. W. David Freeman , Kevin M. Barrett , Lisa Nordan , Aaron C. Spaulding , and RSK, Karney M. Lessons from Mayo Clinic’s Redesign of Stroke Care. *Harvard Business Review*. 2018(Operations).
45. Ouriques Martins SC, Sacks C, Hacke W, Brainin M, de Assis Figueiredo F, Marques Pontes-Neto O, et al. Priorities to reduce the burden of stroke in Latin American countries. *The Lancet Neurology*. 2019;18(7):674-83.
46. van Egdome LSE, Lagendijk M, van der Kemp MH, van Dam JH, Mureau MAM, Hazelzet JA, et al. Implementation of Value Based Breast Cancer Care. *Eur J Surg Oncol*. 2019;45(7):1163-70.
47. Boscolo PR, Callea G, Ciani O, Tarricone R. Measuring Value in Health Care: A Comparative Analysis of Value-based Frameworks. *Clin Ther*. 2020;42(1):34-43.
48. Martin J, BL.; P, TF.; F, TW. F. Value-based Health Care: Lessons for the Anesthesiologist. *Int Anesthesiol Clin*. 2019;57(1):63-80.
49. Blumenthal D, Tavenner M. The “Meaningful Use” Regulation for Electronic Health Records. *N Engl J Med*. 2010;363(6):501-4.
50. Etges APBdS, Ruschel KB, Polanczyk CA, Urman RD. Advances in Value-Based Healthcare by the Application of Time-Driven Activity-Based Costing for Inpatient Management: A Systematic Review. *Value in Health*. 2020;23(6):812-23.
51. Carberry K, Landman Z, Xie M, Feeley T, Henderson J, Fraser C, Jr. Incorporating longitudinal pediatric patient-centered outcome measurement into the clinical workflow using a commercial electronic health record: a step toward increasing value for the patient. *J Am Med Inform Assoc*. 2016;23(1):88-93.

52. Lakdawalla DN, Doshi JA, Garrison LP, Jr., Phelps CE, Basu A, Danzon PM. Defining Elements of Value in Health Care-A Health Economics Approach: An ISPOR Special Task Force Report [3]. *Value Health*. 2018;21(2):131-9.
53. Thaker NG, Pugh TJ, Mahmood U, Choi S, Spinks TE, Martin NE, et al. Defining the value framework for prostate brachytherapy using patient-centered outcome metrics and time-driven activity-based costing. *Brachytherapy*. 2016;15(3):274-82.
54. Devine EB, Capurro D, van Eaton E, Alfonso-Cristancho R, Devlin A, Yanez ND, et al. Preparing Electronic Clinical Data for Quality Improvement and Comparative Effectiveness Research: The SCOAP CERTAIN Automation and Validation Project. *EGEMS (Wash DC)*. 2013;1(1):1025.
55. Peng Zhou SMA, Nora El-Gohary AMA. Ontology-Based Multilabel Text Classification of Construction Regulatory Documents. *J Comput Civ Eng*. 2016;30(4).
56. Abdulla A, Ituarte PG, Wiggins R, Teisberg E, Harari A, Yeh M. Endocrine surgery as a model for value-based health care delivery. *Surg Neurol Int*. 2012;3(1):163.
57. Kulkarni RP, Ituarte PHG, Gunderson D, Yeh MW. Clinical Pathways Improve Hospital Resource Use in Endocrine Surgery. *Journal of the American College of Surgeons*. 2011;212(1):35-41.
58. Kirkpatrick JR, Marks S, Slane M, Kim D, Cohen L, Cortelli M, et al. Using value-based analysis to influence outcomes in complex surgical systems. *Journal of the American College of Surgeons*. 2015;220(4):461-8.
59. Meskó B, Radó N. Data annotators: the unsung heroes of artificial intelligence development.: *The Medical Futurist*; 2019 [Available from: <https://medicalfuturist.com/data-annotation/>].

Figure 1- Study Architecture



Dotted lines represent the union of the text representative technique that was used with which classifier in the 2-phase experiments.

Table 1. Eligible Tasks for analysis and classification rules.

| Tasks | Number of classes | Supporting Information for Classes |
|--|-------------------|---|
| Healthcare Status Achieved (Tier 1) | | |
| Rankin | 8 | 0 to 6/ NI |
| National Health Stroke Scale (NIHSS) | 42 | 1 to 41 / NI |
| Death | 3 | Absent of vital signs / Vital signs present / NI |
| Process of recovery (Tier 2) | | |
| Mobility level | 16 | 1 to 15 / NI |
| Self- care | 3 | Able/Unable/ NI |
| Pain | 4 | No pain / low-intermediate pain / intense pain / NI |
| Strenght | 7 | 0 to 5 / NI |
| Paresis | 3 | Yes/No/ NI |
| Ability to feed orally | 3 | Yes/No/ NI |
| Ability to communicate | 4 | Yes/No/Poorly or syntomathic/ NI |
| Ability of understanding | 4 | Yes/No/Poorly or syntomathic/ NI |
| Ability to ambulate | 4 | Yes/No/Poorly or syntomathic/ NI |
| Treatment / Care Related | | |
| Thrombolytic therapy | 3 | No delta/ Yes/ NI |
| Thrombectomy | 3 | No delta/ Yes/ NI |
| Location | 4 | Emergency room/ ICU/ Inpatient Unit/ NI |
| Infection Indication | 3 | Yes/No/ NI |
| Intracranial Hemorrhage | 3 | Yes/No/ NI |
| Fall Risk | 4 | Low risk / Moderate risk / High risk / NI |
| Pressure Ulcer Risk | 4 | Low risk / Moderate risk / High risk / NI |
| Fall Event during Inpatient | 3 | Yes/No/ NI |
| Baseline Carachteristics | | |
| High blood pressure | 3 | Yes/No / NI |
| Smoking Status | 4 | Yes/No/ Former / NI |
| Coronary artery disease | 3 | Yes/No/ NI |
| Atrial fibrillation | 3 | Yes/No/ NI |
| Diabetes | 3 | Yes/No/ NI |

| | | |
|--------------|---|-------------------|
| Prior Stroke | 3 | Yes/No/ NI |
| Cancer | 3 | Yes/No/ NI |
| Alcoholism | 4 | Yes/No/Former/ NI |
| Obesity | 3 | Yes/No/ NI |
| Dyslipidemia | 3 | Yes/No/ NI |

NI, not informed.

Table 2. Descriptive characteristics of the patients.

| Ischemic Stroke Patients evaluated (n=188) | | |
|---|---------------------------|--------------|
| | Median | Range |
| Age, years | 79 | (68-87) |
| LOS, days | 6 | (4-12) |
| | Number of patients | % |
| Sex | | |
| Female | 100 | 53% |
| Male | 88 | 47% |
| Comorbidities | | |
| Previous stroke | 38 | 20% |
| Previous coronary artery disease | 12 | 6% |
| Atrial fibrillation | 33 | 18% |
| Diabetes | 53 | 28% |
| Hipertension | 125 | 66% |
| Tabagism | 15 | 8% |
| Alcohol consumption | 4 | 2% |
| Clinical Care | | |
| Antithrombotic therapy | 131 | 70% |
| Thrombolysis with rtPA | 38 | 20% |
| Thrombectomy | 12 | 6% |
| Trombolysis and Thrombectomy | 7 | 4% |
| NIHSS | | |
| <8 | 147 | 78% |
| >8 and <15 | 24 | 13% |
| >15 | 17 | 9% |

LOS, length of stay

Age and LOS are shown in median (IQR)

Table 3 – Results of macro F1 for each task and comparative models.

| | W+C + SVM | Linear SVM+BoW | Metafeatures | Word_TFID F + SVM | Char_TFID F + SVM | CNN | BERT | KNN + SVM | Random Forest + SVM | Naive Bayes + SVM |
|------------------------|--------------|----------------|--------------|-------------------|-------------------|--------------|--------------|-----------|---------------------|-------------------|
| Ability to feed orally | 89,48 | 89,46 | 89,37 | 88,93 | 87,06 | 85,46 | 88,40 | 77,13 | 87,57 | 82,65 |
| Location | 88,95 | 89,40 | 89,08 | 86,48 | 88,74 | 83,28 | 89,21 | 78,10 | 83,42 | 68,73 |
| Fall Risk | 89,61 | 91,06 | 88,63 | 86,11 | 86,30 | 88,57 | 83,70 | 74,42 | 74,81 | 67,01 |
| Diabetes | 89,01 | 87,90 | 90,10 | 87,83 | 83,19 | 84,44 | 87,41 | 70,40 | 77,72 | 70,28 |
| Paresis | 88,71 | 87,90 | 88,08 | 87,78 | 86,84 | 83,69 | 89,36 | 69,21 | 74,24 | 68,91 |
| Thrombolytic therapy | 85,76 | 84,77 | 88,58 | 85,03 | 82,48 | 79,63 | 79,46 | 62,26 | 68,96 | 58,68 |
| Death | 89,49 | 66,94 | 89,24 | 88,97 | 85,40 | 68,19 | 62,89 | 76,92 | 72,60 | 74,94 |
| High blood pressure | 85,97 | 79,99 | 84,12 | 81,67 | 77,65 | 79,47 | 66,11 | 65,60 | 69,46 | 56,94 |
| Obesity | 81,66 | 85,45 | 75,37 | 76,49 | 86,03 | 75,46 | 75,89 | 64,03 | 52,82 | 73,08 |
| Ability to ambulate | 75,72 | 76,39 | 72,16 | 76,03 | 75,27 | 80,68 | 69,30 | 65,69 | 66,30 | 59,35 |
| Infection Indication | 79,87 | 74,81 | 77,38 | 73,73 | 79,84 | 69,93 | 76,60 | 60,41 | 63,34 | 58,91 |
| Pressure Ulcer Risk | 66,40 | 92,51 | 65,04 | 66,31 | 65,73 | 86,76 | 64,48 | 63,47 | 60,42 | 56,18 |
| Ability to Communicate | 74,38 | 71,91 | 71,59 | 72,77 | 72,57 | 69,95 | 72,10 | 55,85 | 67,87 | 58,10 |
| Dyslipidemia | 83,21 | 68,58 | 80,65 | 72,46 | 75,25 | 71,77 | 67,45 | 62,55 | 53,26 | 47,20 |
| Smoking Status | 82,15 | 82,39 | 74,17 | 82,98 | 71,86 | 76,08 | 73,85 | 46,00 | 42,28 | 40,32 |
| Thrombectomy | 72,59 | 73,29 | 73,75 | 74,06 | 60,79 | 68,11 | 72,78 | 52,25 | 48,28 | 49,55 |

| | | | | | | | | | | |
|---|--------------|--------------|--------------|--------------|--------------|--------------|-------|-------|-------|-------|
| Ability of understanding | 68,71 | 69,71 | 67,97 | 68,16 | 65,58 | 67,22 | 67,35 | 52,54 | 52,23 | 47,83 |
| Intracranial Hemorrhage | 70,31 | 63,23 | 62,32 | 64,81 | 71,28 | 50,54 | 67,30 | 54,78 | 40,53 | 41,09 |
| Atrial Fibrillation | 71,31 | 65,55 | 51,31 | 68,00 | 65,94 | 64,20 | 62,19 | 48,70 | 38,13 | 47,08 |
| Prior stroke | 67,14 | 57,67 | 70,66 | 58,11 | 61,20 | 49,82 | 56,19 | 59,31 | 35,11 | 51,64 |
| Coronary artery disease | 61,19 | 66,72 | 55,42 | 55,76 | 55,83 | 59,87 | 56,85 | 55,90 | 45,71 | 48,72 |
| Degree of disability / dependence after a stroke | 57,58 | 60,81 | 56,45 | 55,16 | 61,29 | 57,28 | 54,83 | 46,97 | 49,53 | 45,59 |
| Fall Event during Inpatient | 49,99 | 54,43 | 49,99 | 49,99 | 49,99 | 51,88 | 49,99 | 50,65 | 49,99 | 52,39 |
| Pain | 52,01 | 51,25 | 47,81 | 52,05 | 49,89 | 47,09 | 45,72 | 47,11 | 43,55 | 44,69 |
| Strenght | 56,15 | 55,42 | 49,42 | 56,26 | 52,03 | 50,51 | 40,72 | 41,83 | 37,83 | 40,49 |
| Active cancer | 48,87 | 52,52 | 50,50 | 50,16 | 47,84 | 45,36 | 52,44 | 39,06 | 35,73 | 36,11 |
| Alcoholism | 38,58 | 56,08 | 49,52 | 35,70 | 46,53 | 46,86 | 46,16 | 34,18 | 28,33 | 34,03 |
| Mobility level | 40,50 | 32,38 | 27,93 | 39,00 | 38,35 | 55,69 | 16,96 | 30,73 | 28,58 | 27,97 |
| Rankin (mRS) | 26,88 | 28,63 | 22,98 | 26,77 | 28,45 | 68,85 | 24,77 | 25,19 | 25,88 | 21,05 |
| NIHSS | 12,41 | 12,86 | 13,48 | 12,48 | 13,40 | 29,44 | 11,39 | 10,73 | 9,39 | 8,77 |

The colors represent the F1 performance of each model for each task gradually, being red the worse (F1 score < 20), and green the better (F1>80). Results in **bold** are the best (with ties) according to the statistical tests.

Table 4. Performance results for ontology-basis models.

| | Ontology |
|---|-----------------|
| Dyslipidemia | 96,3 |
| Diabetes | 93,8 |
| Pressure Ulcer Risk | 92,2 |
| Obesity | 91,3 |
| Location | 88,4 |
| Rankin (mRS) | 87,1 |
| Thrombolytic therapy | 87,0 |
| Atrial Fibrillation | 84,6 |
| Alcoholism | 83,4 |
| High blood pressure | 81,7 |
| Coronary artery disease | 81,7 |
| Thrombectomy | 68,7 |
| Ability to ambulate | 68,4 |
| Ability to Communicate | 67,2 |
| Paresis | 64,1 |
| NIHSS | 61,2 |
| Ability to feed orally | 60,9 |
| Smoking Status | 60,4 |
| Fall Risk | 52,9 |
| Degree of disability / dependence after a stroke | 40,5 |
| Strenght | 39,3 |
| Fall Event during Inpatient | 38,1 |
| Mobility level | 38,1 |
| Infection Indication | 34,9 |
| Ability of understanding | 34,9 |
| Prior stroke | 16,9 |
| Active cancer | 14,6 |
| Pain | 13,2 |
| Intracranial Hemorrhage | 11,2 |
| Death | 0,6 |

The colors represent the F1 performance of each model for each task gradually, being red the worse (F1 score < 20), and green the better (F1>80)

Supplemental Material 1 – Dataset Characteristics

| Task | Size (clinical narrative) | Size (sentences) | # Classes | Class distribution per task (sentences) | | Sentence Distribution (words) | | | | Distinct Patients (n) | Skewness |
|--------------------------|---------------------------------|---------------------|--------------|--|----------------|-------------------------------|------|-----|-----|-----------------------------|----------------------|
| | | | | Minor Class | Major Class | Median | Mean | Min | Max | | |
| Ability of understanding | 1558 | 44206 | 4 | 124 | 43450 | 9.5 | 10.8 | 1 | 405 | 188 | Extremely Imbalanced |
| Ability to communicate | 1558 | 44206 | 4 | 173 | 43076 | 10 | 10.5 | 1 | 405 | 188 | Extremely Imbalanced |
| Intracranial Hemorrhage | 1558 | 44206 | 3 | 87 | 43990 | 5 | 8 | 1 | 405 | 188 | Extremely Imbalanced |
| Strenght | 1558 | 44206 | 7 | 17 | 43541 | 10 | 13.1 | 1 | 405 | 188 | Extremely Imbalanced |
| Location | 1558 | 44206 | 4 | 109 | 42694 | 13 | 13.9 | 1 | 405 | 188 | Extremely Imbalanced |
| Death | 1558 | 44206 | 3 | 9 | 41871 | 6 | 8.12 | 1 | 405 | 188 | Extremely Imbalanced |
| Thrombolytic therapy | 1558 | 44206 | 3 | 107 | 43708 | 13 | 14.4 | 1 | 405 | 188 | Extremely Imbalanced |
| Thrombectomy | 1558 | 44206 | 3 | 25 | 43971 | 16.9 | 11 | 1 | 405 | 188 | Extremely Imbalanced |
| Diabetes | 1558 | 44206 | 3 | 39 | 43859 | 7.8 | 6.7 | 1 | 405 | 188 | Extremely Imbalanced |
| High blood pressure | 1558 | 44206 | 3 | 25 | 43624 | 7.8 | 7 | 1 | 405 | 188 | Extremely Imbalanced |
| Atrial Fibrillation | 1558 | 44206 | 3 | 14 | 43914 | 10.4 | 9.8 | 1 | 405 | 188 | Extremely Imbalanced |

| | | | | | | | | | | | |
|-----------------------------|------|-------|----|-----|-------|------|------|---|-----|-----|----------------------|
| Smoking Status | 1558 | 44206 | 4 | 32 | 43923 | 13.2 | 16.2 | 1 | 405 | 188 | Extremely Imbalanced |
| Alcoholism | 1558 | 44206 | 4 | 22 | 44097 | 20.4 | 16.1 | 1 | 405 | 188 | Extremely Imbalanced |
| Dyslipidemia | 1558 | 44206 | 3 | 8 | 44063 | 7.9 | 6 | 1 | 405 | 188 | Extremely Imbalanced |
| Prior stroke | 1558 | 44206 | 3 | 29 | 43969 | 10.5 | 9.7 | 1 | 405 | 188 | Extremely Imbalanced |
| Fall Event during Inpatient | 1558 | 44206 | 2 | 87 | 43990 | 11.8 | 10.5 | 1 | 405 | 188 | Extremely Imbalanced |
| Fall Risk | 1558 | 44206 | 4 | 115 | 43763 | 8 | 12.5 | 1 | 405 | 188 | Extremely Imbalanced |
| Ability to ambulate | 1558 | 44206 | 4 | 106 | 43479 | 15.2 | 8.4 | 1 | 405 | 188 | Extremely Imbalanced |
| Mobility level | 1558 | 44206 | 16 | 3 | 43560 | 17.9 | 11.6 | 1 | 405 | 188 | Extremely Imbalanced |
| Pressure Ulcer Risk | 1558 | 44206 | 4 | 59 | 43949 | 11.1 | 9.2 | 1 | 405 | 188 | Extremely Imbalanced |
| NIHSS | 1558 | 44206 | 42 | 2 | 43903 | 5 | 8 | 1 | 405 | 188 | Extremely Imbalanced |
| Rankin (mRS) | 1558 | 44206 | 7 | 11 | 44019 | 10.2 | 8.5 | 1 | 405 | 188 | Extremely Imbalanced |
| Pain | 1558 | 44206 | 4 | 23 | 43581 | 9.9 | 7.7 | 1 | 405 | 188 | Extremely Imbalanced |
| Ability to feed orally | 1558 | 44206 | 3 | 689 | 42657 | 9.2 | 6,7 | 1 | 405 | 188 | Extremely Imbalanced |
| Obesity | 1558 | 44206 | 3 | 27 | 44121 | 11.4 | 8,5 | 1 | 405 | 188 | Extremely Imbalanced |
| Paresis | 1558 | 44206 | 2 | 505 | 43701 | 11.1 | 8.5 | 1 | 405 | 188 | Extremely Imbalanced |

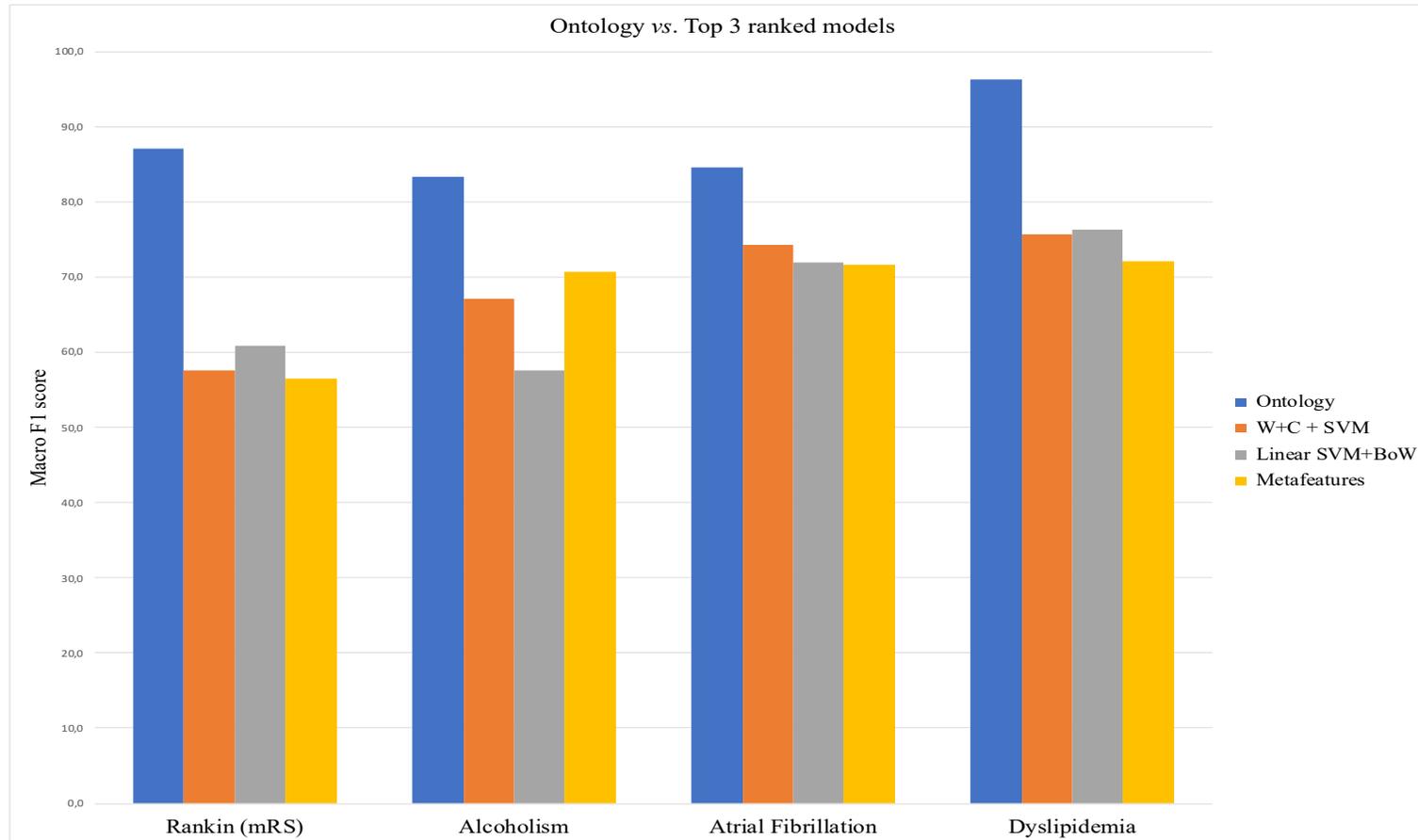
| | | | | | | | | | | | |
|--|------|-------|---|-----|-------|-------|------|---|-----|-----|----------------------|
| Coronary artery disease | 1558 | 44206 | 3 | 24 | 43890 | 13 | 8.3 | 1 | 405 | 188 | Extremely Imbalanced |
| Degree of disability / dependence after a stroke | 1558 | 44206 | 3 | 70 | 43778 | 15.35 | 12.3 | 1 | 405 | 188 | Extremely Imbalanced |
| Infection Indication | 1558 | 44206 | 4 | 362 | 43234 | 11.5 | 8.3 | 1 | 405 | 188 | Extremely Imbalanced |
| Active cancer | 1558 | 44206 | 3 | 55 | 44021 | 13.5 | 8.5 | 1 | 405 | 188 | Extremely Imbalanced |

Minor class, less expressed class in number of sentences

Major class, most present class in number of sentences

Mean, average words per sentence per class

Median, median words per sentence per class

Figure 5 - Macro F1 score por ontology and best 3 models

6. CONCLUSÕES E CONSIDERAÇÕES FINAIS

A partir da contextualização sobre a importância do avanço em capacidade tecnológica informacional para suportar modelos de avaliação de valor em saúde, esta pesquisa se dispôs a avaliar a temática de modo exploratório.

A revisão da literatura realizada deste trabalho atendeu ao objetivo específico de compreender como os resultados em saúde estão sendo medidos em casos do mundo real, e analisar a origem dos dados usados para relatar a iniciativa de cuidado em saúde baseada em valor. Os resultados do primeiro artigo sugerem que ainda há uma lacuna na medição dos resultados que são importantes para os pacientes e na medição dos resultados financeiros por meio de métodos metodológicos rigorosos. Avanços na capacidade tecnológica e uma cultura de mudança na gestão parecem ser as principais barreiras para tornar a agenda de valor mais facilmente reproduzível, e ainda são requisitos necessários para que os itens acima ocorram.

O segundo artigo apresenta os melhores métodos a serem investidos na classificação de texto a partir de dados não estruturados de prontuários eletrônicos para medir de forma automatizada alguns desfechos de saúde de pacientes que sofreram AVC. O framework demonstra visualmente os modelos que apresentam melhor desempenho para avaliar cada tarefa. Com isso é possível orientar e impulsionar a adequação de parâmetros que vão ao encontro do que se busca na gestão por valor e a adoção de indicadores que permitam o seu controle.

A escrita deste trabalho e o interesse pela temática só foi possível devido a oportunidade de participar e auxiliar na coleta de dados de projeto “Proposição de um modelo de gestão de saúde baseada em valor para saúde suplementar e pública do Brasil”. As discussões e experimentos em avanços em capacidade tecnológica para estudos de microcusteio despertaram motivação para explorar o outro fator da equação de valor trabalhado nesta dissertação. Em parte, a construção desta parte da pesquisa foi desencadeada pelo interesse em inovação em saúde e tentativa da união de áreas de ensino e pesquisa que juntas podem atender a um interesse comum no campo da saúde ponderada pelo valor.

Em uma das primeiras reuniões sobre o projeto nos foi dito que “(..) *os modelos computacionais não podem trazer mais trabalho para o prestador, os algoritmos precisam aprender a interpretar os dados tal como eles estão sendo colocados no dia a dia, por isso*

que se chama 'inteligência artificial' (...)". A partir daí pessoas com formações e bagagens diferentes desenvolveram relação de equipe e propósito comum. Por isso finalizo as considerações finais sobre este trabalho com a reflexão do primeiro passo dado como grupo de pesquisa em prol de avanços de capacidade tecnológica para programas de valor em saúde, e que isso inove e reflita positivamente no nosso papel como pesquisador no Brasil.

7. LIMITAÇÕES DO ESTUDO

Existem limitações a serem consideradas. A primeira delas foi o número de pacientes que compreendeu a nossa amostra, embora seja uma característica do banco de dados ser um texto muito desbalanceado em termos de informações extraídas, um número maior de pacientes poderia gerar classes mais homogêneas. Em segundo lugar, embora em termos de sentença seja um grande volume, não testamos a performance dos modelos a nível de evolução e reconhecemos que algumas variáveis poderiam ter uma performance mais adequada se fossem avaliadas a nível de evolução e não de sentença. Uma terceira limitação, é que até o momento foram usados dados de apenas um hospital, sendo difícil fazer alguma inferência sobre a generalizabilidade dos resultados encontrados.

Técnicas de aprendizado de máquina prometem resultados satisfatórios na extração de informações clínicas, mas para isso faz-se necessário um corpus anotado para treinamento de algoritmos. *Corpus* para textos em português são raros, especialmente focados na área da saúde (BUTTE; WEINSTEIN; KOHANE, 2000; JAGANNATHA; LIU; LIU *et al.*, 2019; MIOTTO; WENG, 2015; RAGHAVAN; CHEN; LUSSIER *et al.*, 2014) A grande maioria das ferramentas de PLN são treinadas com corpus jornalístico o que dificulta a identificação de termos pela falta de similaridade entre eles.

A baixa disponibilidade de dados de registros clínicos também pode estar associada a falta de padrão na escrita dos atendimentos realizados pelos profissionais. Os textos utilizados para aplicação do método não seguiam nenhum padrão de escrita e de estrutura, apresentavam erros gramaticais e ortográficos, diferentemente dos trabalhos desenvolvidos por Friedlin e McDonald (2006) e Long (2007) que extraíram informações a partir de um cenário ideal de estrutura e escrita.

8. TRABALHOS FUTUROS

Como perspectivas de trabalhos futuros e maturidade da pesquisa, entende-se a necessidade de expandir o uso dos algoritmos com melhor performance para outras linhas de cuidado e hospitais. Melhorias contínuas de capacidade tecnológica informacional para garantir a disponibilidade de dados de resultados em todo o ciclo de atendimento, e a instituição com cultura baseada em valor entre os provedores são essenciais para conduzir a implementação do VBHC.

Os modelos gerados por esta pesquisa utilizaram dados de um hospital privado do sul do país, o uso de novos dados da mesma instituição pode ampliar a aplicação dos modelos para outras linhas de cuidado. Abre-se espaço, por exemplo, para pesquisa no âmbito do mapeamento epidemiológico da instituição com base nas morbidades da população, controle de escalas de risco, suporte para vigilância de eventos de risco, integração com dados de performance e desfechos reportados pelo paciente. E, dessa forma, contribuir para o desenvolvimento de novas pesquisas de VBHC em países de língua portuguesa.

Em paralelo, entende-se a aplicação dos algoritmos em outros hospitais irá contribuir com a generalizabilidade dos modelos de monitoramento de resultados para linha de cuidado do AVC, facilitando o uso de indicadores que corroborem com uma maior transparência do desempenho das instituições para a linha de cuidado específica. Para isso são necessários mais dados manualmente anotados para conferência em diferentes instituições. Para o aperfeiçoamento do algoritmo de expressão regular, o grupo de pesquisa em recuperação de informação prossegue os estudos na construção de um *corpus* anotado em português para auxiliar no processo de análise semântica de textos clínicos. Isso também abre espaço para uso de plataformas *open-source* de anotação de dados em larga escala, como *Inception* ou *Webanno*.

E finalmente, vê-se a oportunidade da criação de *dashboards* ou plataformas *Business Intelligence* que permitam a visualização desses dados a nível do gestor hospitalar. Visualizar os dados em tempo real da jornada de cuidado e a nível de paciente é um dos grandes objetivos de implementação de VBHC em uma instituição (PORTER; LEE, 2013), que permite maior transparência e objetividade nas comparações entre instituições. Por isso, esforços que permitam tornar a saída desses algoritmos passíveis de serem implementadas na prática é outra tarefa a seguir.

APÊNDICE A - DETALHES METODOLÓGICOS ADICIONAIS

Esta seção inicia-se com as considerações éticas da pesquisa, tipo de estudo, local onde a pesquisa foi desenvolvida e cenário da pesquisa. Posteriormente são apresentadas as etapas e design do experimento realizado.

1. Considerações Éticas da Pesquisa

A disponibilização da base de dados para o grupo de pesquisa de recuperação de informação em saúde ocorreu após aprovação do Comitê de Ética em Pesquisa (CEP) do Hospital Moinhos de Vento de Porto Alegre (HMV), sob parecer nº 29694720.0.0000.5330 obtido em 06 de maio de 2020 (ANEXO A)

Este projeto também atende aos preceitos éticos, diretrizes e normas regulamentadoras em pesquisa envolvendo seres humanos estabelecidos na Lei Geral de Proteção de Dados Pessoais 13.853 de 8 de julho de 2019 que prevê, entre outros preceitos, a de-identificação das informações confidenciais de pacientes, familiares e profissionais da base de dados utilizada. Pacientes não serão identificados individualmente em nenhuma análise que será realizada e os dados das evoluções serão extraídos dos sistemas de informação seguindo os processos internos estabelecidos.

O acesso à base completa de dados para análise dos resultados será apenas dos pesquisadores do HMV que assinaram os termos de compromisso com o uso de dados. Os dados coletados serão mantidos em sigilo a fim de assegurar a privacidade e o anonimato dos participantes. As informações obtidas foram utilizadas apenas para fins científicos vinculados a este projeto.

2. Tipo de estudo

Esta pesquisa, por estar orientada a propor soluções de um problema específico em ambiente real, apresenta natureza exploratória, a qual proporciona um aprofundamento na temática, ainda relativamente pouco explorada, de Gestão de Saúde Baseada em Valor, ao estudá-la no contexto de mensuração de desfechos (BOAVENTURA, 2004; TRANFIELD; DENYER; SMART, 2004; YIN, 2014). No que diz respeito à abordagem, tem-se uma discussão qualitativa e quantitativa pois, ao mesmo tempo em que se analisa de maneira qualitativa quais parâmetros e indicadores são relevantes para uso na linha de cuidado, propõe-se um modelo que permite automatização de análises em registros médicos eletrônicos e a eficácia deste modelo.

3. Local de estudo

Esta é uma pesquisa que faz parte do projeto intitulado “Proposição de modelo de gestão de saúde baseada em valor para os sistemas de saúde pública e suplementar do Brasil” proposto pelo Instituto de Avaliação de Tecnologias em Saúde (IATS/INCT) e Universidade Federal do Rio Grande do Sul (UFRGS), em parceria com alunos de grupos de pesquisa da Universidade Federal de Minas Gerais (UFMG) e da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS).

4. Cenário da Pesquisa

Este estudo está consolidando dados do hospital proponente da pesquisa. O estudo foi desenvolvido em um hospital privado de alta qualidade do Brasil, credenciado pela *Joint Comission International*, sendo classificado como o 5º melhor hospital do país. Desde 2017, o hospital introduziu a coleta de dados dos conjuntos de padrões ICHOM em sua rotina para diferentes caminhos clínicos e criou o escritório de valor institucional. Para o caminho do AVC, a coleta de dados teve início em 15 de outubro de 2015 e, desde agosto de 2018, está estruturada no Redcap e integrado ao Escritório de Valor do hospital. Em 2019, o hospital passou a incorporar o *Angel Awards Program*, sendo certificado com a categoria Platinum ao final do primeiro ano.

Foram levantados os números de prontuário para recuperação dos dados de evolução texto de pacientes com AVC isquêmico internados no hospital de agosto de 2018 a março de 2020. Foram anotadas 44206 sentenças, que corresponde 1558 evoluções de 188 pacientes. A plataforma de prontuários eletrônicos da instituição usa o sistema de prontuários MV.

Antes do processamento dos dados de evolução para anotação das variáveis, foi necessário uma limpeza e preparação da base de dados, que consistiu na remoção dos registros que não fossem da internação correspondente ao AVCi do paciente – que eventualmente poderia ter internado com suspeita do diagnóstico mas não ter sido confirmado durante a internação, ou ainda paciente que internou mais de uma vez nesta instituição no mesmo período.

5. Pré-processamento

O pré-processamento ocorreu em três momentos: i) de-identificação das narrativas clínicas, ii) aleatorização das evoluções, e iii) quebra de sentença.

A de-identificação representa a retirada de informações confidenciais e sigilosas da base de dados e foi feita por meio de um algoritmo de expressão regular que permite a busca por padrões de escrita para substituição de trechos ou caracteres específicos nos textos. A equipe de pesquisadores responsável pela programação dos algoritmos obteve a relação de todos os pacientes e nome de profissionais para que fosse possível essa substituição. Entretanto, alguns problemas surgiram durante a interpretação do algoritmo em virtude da ambiguidade dos nomes. Pelo fato de alguns nomes da listagem coincidirem com pronomes e substantivos que não deveriam ser de-identificados como, por exemplo, Tala, Clara, Bem e Pelo. Desta forma, algumas informações como pronomes de tratamento, apelidos, abreviações de registros de classes profissionais não foram de-identificadas.

As expressões regulares utilizadas para a de-identificação dessas informações são substituídas no texto da seguinte maneira: “NOME_PACIENTE” e “NOME_COMPLETO_PACIENTE”. Os dados de evolução foram aleatorizados antes de serem entregues para anotação aos pesquisadores para abranger um maior número de pacientes durante as anotações, e para não viciar a marcação do anotador por já estar familiarizado com a descrição de um mesmo paciente.

Foi utilizada a estratégia de quebra das evoluções em sentenças baseados nos pressupostos de que as informações das variáveis estão presentes em um espaço curto do texto, o uso de grandes quantidades de dados pode facilitar aprendizado preciso da semântica (CAROPRESO; MATWIN, 2006), redução de ruídos (textos em branco, não informativos)(MANNING; RAGHAWAN; SCHUTZE, 2009; MANNING; SCHUTZE, 1999), e é uma estratégia comum em plataformas de anotação semântica (KLIE; BUGERT; BOULLOSA *et al.*, 2018). A quebra de sentenças pode ser feita de diversas formas, por pontuação, número de palavras, uso de *stop words*, entre outros, a nossa estratégia incluiu o uso da biblioteca spaCy, que é uma biblioteca de código aberto gratuita para PLN avançado em Python. Foi escolhida por apresentar a melhor performance em tarefas de quebra de sentença em textos longos e mal formatados (HONNIBAL; MONTANI, 2015), como é o caso

de evoluções clínicas, e por fazer uma quebra baseada em semântica e não pontuação, como seria uma humana.

Uma limitação importante da quebra de sentenças é a quebra errônea, por exemplo, quebras de sentença em locais que não fazem sentido gerando um dado de má qualidade. Dado de má qualidade é um dado incoerente e sem sentido, e que por vezes quando em formato de sentenças perdem seu significado, isso pode implicar em falhas nas anotações não pela ausência da informação e sim pela quebra que não foi informativa o suficiente.

5.1 Variação das Terminologias

A verificação da variação de terminologias em meio as narrativas clínicas são relevantes para este estudo, pois os termos relacionados as variáveis e suas classes podem ser redigidos de maneiras diferentes, porém com o mesmo significado. Os textos utilizados não seguem nenhum padrão formal de escrita, com alta variação de escrita, além de erros ortográficos e alguns textos inteiramente em letras maiúsculas. Exemplos de tais variações são as sinonímias e as variações da palavra como abreviaturas, acrônimos e estilos de escrita. Um exemplo é o termo “trombólise” que também pode ser encontrado como “terapia de reperfusão” e como acrônimo “rtPA”. Da mesma forma, o termo “hipertensão” e “diabetes” pode ser encontrado respectivamente como “HAS” e “DM”, entre outros.

O intuito de verificar o padrão de escrita resultou na criação de um manual (ANEXO B) para auxiliar os anotadores a manterem a mesma qualidade de anotação. Por vezes, quando se trata de variáveis subjetivas, o fato de ter o manual de classificação auxilia o entendimento das sentenças por parte dos anotadores. Além disso, este manual construído foi utilizado também para os algoritmos de ontologia que foram detalhados na seção 2.5.

6. Rotulação dos dados e validação do experimento

Anotação dos dois foi feita aos pares nos documentos em *Google Drive*. Para definição de cada variável e instrução as anotadoras tinham o manual de anotação para treinamento dos algoritmos (ANEXO B). A planilha de anotação continha as sentenças, e cada coluna representava uma variável. As anotadoras marcavam qual classe de cada variável aquela sentença representava e para as variáveis que não forneciam informação sobre a sentença era marcado -1. Após a anotação, os documentos foram cruzados, e as divergências eram resolvidas por terceiros a depender da especialidade (fisioterapia, neurologia, cardiologia, etc) (MESKÓ; RADÓ, 2019).

A validação cruzada é um método estatístico usado para estimar a habilidade dos

modelos de AP (JAMES; WITTEN; TIBSHIRANI, 2013; RUSSELL; NORVIG, 2016). O procedimento tem um único parâmetro chamado k que se refere ao número de grupos em que uma determinada amostra de dados deve ser dividida. Como tal, o procedimento é frequentemente denominado validação cruzada k -fold (RUSSELL; NORVIG, 2016). Quando um valor específico para k é escolhido, ele pode ser usado no lugar de k na referência ao modelo, por exemplo $k = 5$ tornando-se uma validação cruzada de 5 vezes (MEYSTRE; KIM; GOBBEL *et al.*, 2017). É comumente usado no aprendizado de máquina aplicado para comparar e selecionar um modelo para um determinado problema de modelagem preditiva porque é fácil de implementar e resulta em estimativas de habilidade que geralmente têm um viés menor do que outros métodos (BALYAN; CROSSLEY; BROWN *et al.*, 2019; CUNHA; CANUTO; VIEGAS *et al.*, 2020; GOLAS; SHIBAHARA; AGBOOLA *et al.*, 2018).

Neste estudo foi utilizado uma validação cruzada de 5 *folds*. Informações de um mesmo paciente foram atribuídas sempre ao mesmo *fold* com objetivo de testar a capacidade do modelo de prever novos dados que não foram usados na estimativa, sinalizando problemas como *overfitting* ou viés de seleção (CAWLEY; TALBOT, 2010). Apesar de conservadora, essa estratégia busca ser mais condizente com a aplicação na realidade, uma vez não se espera que a instituição tenha informações prévias de um novo paciente. Além disso, foi feita a estratificação de cada *fold*, ou seja, foi preservada porcentagem de amostras para cada classe de cada tarefa. No caso da classificação binária, isso significa que cada partição contém aproximadamente as mesmas proporções dos dois tipos de rótulos de classe (VANWINCKELEN, 2012). A implementação tem como objetivo gerar conjuntos de teste de forma que todos contenham a mesma distribuição de classes ou o mais próximo possível preserva as dependências de ordem na ordenação do conjunto de dados, gera conjuntos de teste onde o menor e o maior diferem em no máximo uma amostra (PEDREGOSA; ET AL, 2011).

7. Métricas de performance

Estudos de ferramentas de usando PLN são semelhantes a avaliação dos estudos de diagnóstico - testando se um histórico item, descoberta de exame ou resultado do teste está associado a um subsequente diagnóstico. A construção é uma tabela de contingência 2×2 onde o número de valores atribuídos correta e incorretamente para um dado resultado binário e da criação da curva ROC para mensuração da área sobre a curva. Essa tabela permite avaliação de métricas de desempenho, como precisão (Valor preditivo positivo), *recall*

(sensibilidade), precisão, *F-score*. *F-score* é a média ponderada de precisão e sensibilidade para cada classe (KOWSARI; JAFARI; HEIDARYSAFA *et al.*, 2019; KUHN; JOHNSON, 2013; VELUPILLAI; SUOMINEN; LIAKATA *et al.*, 2018). Portanto, essa pontuação leva em consideração tanto os falsos positivos quanto os falsos negativos, e é calculado pela seguinte fórmula:

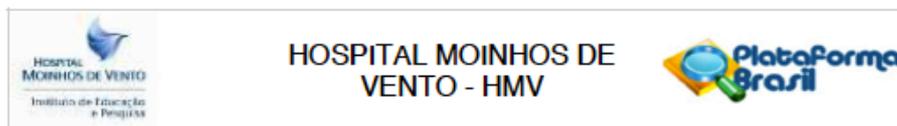
$$(1) \quad F1 \text{ Score} = 2 * (\text{Revocação} * \text{Precisão}) / (\text{Revocação} + \text{Precisão})$$

Fórmula do cálculo de F-score

Como estamos tratando de tarefas com multiníveis de classe, podemos fazer uso da Macro-F1 que retorna a média da métrica F1 para todas as classes de cada tarefa. Assim pra se ter um valor alto de Macro F1 é preciso ter um valor alto de F1 pra cada uma das classes, tanto as maiores (com muitos eventos) quanto as menores (com poucos eventos).

Embora as métricas de avaliação intrínseca sejam importantes e valorizem especialmente quando comparamos diferentes métodos de PLN para a mesma tarefa, para estabelecer um valor de referência é importante ter bem claro o que o modelo propõe responder. Não são necessariamente este valor é informativo quando a tarefa é aplicada em um problema de nível superior (por exemplo, nível do paciente) ou em novos dados. Por exemplo, se o estado da arte atual que é alcançado e a classificação de desfecho é > 80% F-score (UZUNER, 2011; VAN CALSTER; VICKERS, 2015), o que é próximo de acordo humano sobre a mesma tarefa, pode ser satisfatório para os objetivos do modelo. No entanto, se esse sistema tem por objetivo ser implantado na prática clínica e for essencial para tomada de decisão médica, qualquer taxa de erro > 0%, como indicação de medicamento ou história de alergia grave, pode ser vista como inaceitável.

ANEXO A. CÓPIA DO PARECER DE APROVAÇÃO DO PROJETO NO COMITÊ DE ÉTICA



PARECER CONSUBSTANCIADO DO CEP

DADOS DA EMENDA

Título da Pesquisa: Proposição de modelo de gestão de saúde baseada em valor para os sistemas de saúde pública e suplementar do Brasil

Pesquisador: Carisi Anne Polanczyk

Área Temática:

Versão: 4

CAAE: 29894720.0.0000.5330

Instituição Proponente: Hospital Moinhos de Vento - HMV

Patrocinador Principal: Capes Coordenação Aperf Pessoal Nível Superior

DADOS DO PARECER

Número do Parecer: 4.199.370

Apresentação do Projeto:

Trata-se de uma Emenda referente a pesquisa: Proposição de modelo de gestão de saúde baseada em valor para os sistemas de saúde pública e suplementar do Brasil.

Pesquisador Responsável: Carisi Anne Polanczyk

Objetivo da Pesquisa:

Objetivo Primário:

Propor um modelo de gestão de saúde baseada em valor para os sistemas de saúde suplementar e pública que permita orientar a gestão e alternativas de reembolso de saúde no Brasil.

Objetivos Secundários:

- I. Definir medidas de avaliação do desempenho clínico, operacional, de qualidade e segurança ao longo da internação e de percepção do paciente que permitam avaliar os desfechos do serviço de saúde prestado.
- II. Propor cost tool kit baseadas em métodos acurados de microcusteio para aferir custos reais e compor as equações de valor.
- III. Propor framework de gestão de valor para a saúde suplementar.
- IV. Definir método de avaliação de valor para o sistema público de saúde através da proposição da equação de valor incremental.

Endereço: Rua Tiradentes, 198 - Subsolo
Bairro: Floresta **CEP:** 90.560-030
UF: RS **Município:** PORTO ALEGRE
Telefone: (51)3314-3537 **E-mail:** cep.iep@hmv.org.br



Continuação do Parecer: 4.199.370

Avaliação dos Riscos e Benefícios:

A avaliação risco/benefício é favorável e justifica a pesquisa. Estão bem descritos e explicados no projeto de pesquisa.

Comentários e Considerações sobre a Pesquisa:

Nesta emenda, foram incluídas as seguintes informações:

- Cartas de autorização de desenvolvimento do estudo dentro dos serviços de neurologia, e documentações que foram solicitadas pelos centros co-participantes.
- Documentos na plataforma por solicitação dos CEPs dos demais centros.

Cabe destacar que nenhuma alteração foi feita no detalhamento do projeto.

Neste contexto, incluem-se os seguintes documentos:

- carta_justificativa_emendar_HMV.docx
- TCLEpara_pacientes_HospitalSaoJose.pdf
- TCLEpara_HospitalSaoJose.pdf
- anuencia_octavio_USP.pdf
- AUTORIZACAO_SAOJOSE_JOINVILLE.pdf

Considerações sobre os Termos de apresentação obrigatória:

Estão adequados.

Recomendações:

Vide conclusões.

Conclusões ou Pendências e Lista de Inadequações:

Emenda pode ser aprovadas, sem pendências.

Considerações Finais a critério do CEP:

Diante do exposto, o Comitê de Ética em Pesquisa em Seres Humanos do Hospital Moínhos de Vento, de acordo com as atribuições definidas na Resolução 466/2012 do CNS e complementares, e pela Norma Operacional N° 001/2013 do CNS, manifesta-se pela aprovação da Emenda.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

| Tipo Documento | Arquivo | Postagem | Autor | Situação |
|----------------|---------|----------|-------|----------|
|----------------|---------|----------|-------|----------|

Endereço: Rua Tiradentes, 198 - Subsolo
 Bairro: Floresta CEP: 90.560-030
 UF: RS Município: PORTO ALEGRE
 Telefone: (51)3314-3537 E-mail: cep.iep@hmv.org.br



Continuação do Parecer: 4.199.370

| | | | | |
|---|--|------------------------|----------------------------------|--------|
| Informações Básicas do Projeto | PB_INFORMAÇÕES_BÁSICAS_1599698_E2.pdf | 23/07/2020 08:17:10 | | Acelto |
| Outros | carta_justificativa_emendar_HMV.docx | 23/07/2020 08:16:27 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| Outros | TCLEpara_pacientes_HospitalSaoJose.pdf | 22/07/2020 09:01:56 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| Outros | TCLEpara_HospitalSaoJose.pdf | 22/07/2020 09:00:15 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| Outros | anuencia_octavio_USP.pdf | 22/07/2020 08:57:32 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| Outros | AUTORIZACAO_SAOJOSE_JOINVILLE.pdf | 22/07/2020 08:56:36 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| Outros | EMENDA_JUSTIFICATIVA.docx | 02/06/2020 09:03:00 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| Folha de Rosto | FolhaRostoAbril2020.pdf | 17/04/2020 16:03:49 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| Recurso Anexado pelo Pesquisador | cartaResposta_parecr_HMV.docx | 29/03/2020 09:38:49 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| TCLE / Termos de Assentimento / Justificativa de Ausência | AnexoIIconviteprofissionais.docx | 29/03/2020 09:28:05 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| TCLE / Termos de Assentimento / Justificativa de Ausência | AnexoITCLEprofissionais.docx | 29/03/2020 08:51:56 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| TCLE / Termos de Assentimento / Justificativa de Ausência | AnexoIIITCLEIhom.docx | 29/03/2020 08:51:19 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| Projeto Detalhado / Brochura Investigador | projetoVbhchMV.docx | 14/01/2020 10:09:30 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| Outros | ParecerdaComissaocientifica.pdf | 14/01/2020 10:07:18 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| TCLE / Termos de Assentimento / Justificativa de Ausência | ANEXOIVdispensaTCLE.docx | 14/01/2020 10:04:14 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| TCLE / Termos de Assentimento / Justificativa de Ausência | AnexoIIconvitelCHOM.docx | 14/01/2020 10:03:59 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| Declaração de concordância | CarlisiPolanczyk130120pdf.pdf | 14/01/2020 09:55:56 | ANA PAULA BECK DA SILVA ETGES | Acelto |
| Declaração de Pesquisadores | termostodosassinados.pdf | 14/01/2020 09:54:19 | ANA PAULA BECK DA SILVA ETGES | Acelto |

Endereço: Rua Tiradentes, 198 - Subsolo

Bairro: Floresta

CEP: 90.560-030

UF: RS

Município: PORTO ALEGRE

Telefone: (51)3314-3537

E-mail: cep.lep@hmv.org.br



Continuação do Parecer: 4.199.370

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

PORTO ALEGRE, 07 de Agosto de 2020

Assinado por:

Guilherme Alcides Fiôres Soares Rollin
(Coordenador(a))

Endereço: Rua Tiradentes, 198 - Subsolo

Bairro: Floresta

CEP: 90.560-030

UF: RS

Município: PORTO ALEGRE

Telefone: (51)3314-3537

E-mail: cep.lep@hmv.org.br

ANEXO B. MANUAL DE TREINAMENTO PARA OS ALGORITMOS

A construção de algoritmos de classificação de textos baseados em processamento de linguagem natural requer que os dados de texto sejam anotados por especialistas que devem buscar um padrão de anotação. As anotações foram feitas por duas especialistas da área da saúde e conhecedoras do método aplicado, levando em consideração o manual de definição para cada variável. Este manual foi produzido e validado pela equipe do projeto, nas suas diferentes expertises.

Universidade Federal do Rio Grande do Sul
Instituto de Avaliação de Tecnologias em Saúde

Proposta de automatização da análise de desfecho para dimensionar programas de
gestão de saúde baseada em valor.

Manual de Orientações para Classificação de texto das Evoluções

Coordenação do projeto:

Profa Dra. Carisi Anne Polanczyk

Confecção do manual: Bruna Stella Zanotto, Renata Garcia Ruschel

Porto Alegre, 2020.

LISTA DE ABREVIATURAS

| | |
|---------|--|
| AVC | Acidente vascular cerebral |
| AVCi | Acidente vascular cerebral isquêmico |
| AVC ISQ | Acidente vascular cerebral isquêmico |
| AVCh | Acidente vascular cerebral hemorrágico |
| ATB | Antibiótico |
| CI | Cardiopatía isquêmica |
| IAM | Infarto agudo do miocárdio |
| DAC | Doença arterial coronariana |
| DCC | Doença cardíaca coronariana |
| SCA | Síndrome coronariana aguda |
| IC | Insuficiência cardíaca |
| CRM | Cirurgia de revascularização do miocárdio |
| MP | Marca passo |
| FA | Fibrilação atrial |
| ACFA | Fibrilação atrial |
| CA | Câncer |
| DM | Diabetes mellitus |
| DM2 | Diabetes mellitus tipo II |
| HAS | Hipertensão |
| DLP | Dislipidemia |
| TC | Tomografia computadorizada |
| TCC* | Tomografia computadorizada cardíaca / controle |
| RM | Ressonância magnética |
| SV | Sinais vitais |
| AA | Ar ambiente |
| AO | Abertura ocular |
| JH | John Hopkins (escala de risco de queda) |
| GF | Grau de força (escala) |
| FG | Força grau (escala) |
| VO | Via oral |
| GTT | Gastrostomia |

| | |
|--------|---|
| SNE | Sonda nasoenteral |
| BI | Bomba de infusão |
| ITU | Infecção do trato urinário |
| HMC | Hemocultura |
| EVA | Escala visual analógica |
| PAINAD | Pain Assessment in Advanced Dementia |
| BPS | Behavioral Pain Scale |
| MI | Membro inferior |
| MsIs | Membros inferiores |
| MIE | Membro inferior esquerdo |
| MID | Membro inferior direito |
| MS | Membro superior |
| MsSs | Membros superiores |
| MSE | Membro superior esquerdo |
| MSD | Membro superior direito |
| LOC | Lucido, orientado e coerente |
| BEG | Bom estado geral |
| REG | Regular estado geral |
| MEG | Mal estado geral |
| Tmax | Temperatura máxima |
| NIHSS | National Institute of Health Stroke Scale |
| CTI | Centro de tratamento intensivo |
| UTI | Unidade de tratamento intensivo |
| EMG | Emergência |
| UI | Unidade de internação |

1. Apresentação

A construção de algoritmos de classificação de textos baseados em processamento de linguagem natural requer que os dados de texto sejam anotados por especialistas que devem buscar um padrão de anotação.

Este manual foi criado para orientar o(a) anotador (a) no contexto do projeto de pesquisa intitulado “Modelo de Gestão de Saúde Baseado em Valor para o Sistema Único de Saúde e Saúde Suplementar do Brasil”, aprovado pelo Comitê de Ética em Pesquisa (CAAE 29694720000005330).

2. População-Alvo

Esse material é destinado a pesquisadores das instituições hospitalares que venham a replicar em suas instituições as soluções computacionais desenvolvidas no projeto de pesquisa “Modelo de Gestão de Saúde Baseado em Valor para o Sistema Único de Saúde e Saúde Suplementar do Brasil”.

3. Orientações gerais

Ao anotar dados de sua instituição, algumas premissas são importantes:

- Solicite os códigos fonte aos pesquisadores escrevendo um e-mail para anabsetges@gmail.com.

- Ao receber os códigos, não altere a estrutura ou nome das variáveis, caso seja necessário, entre em contato com os pesquisadores do projeto nos e-mails: anabsetges@gmail.com ou brunaszanotto@gmail.com.

- A informação negativa de uma sentença sempre deve ser anotada, por exemplo, “nega hipertensão”, precisa ser classificada como uma sentença que refere a ausência de hipertensão. Diferentemente do caso de uma sentença não ter relação nenhuma com a variável avaliada, quando isso acontecer, deixe-a em branco

- Ao longo do processo de anotação, o objetivo é extrair dados de evoluções de texto com base na melhor evidência descrita. Por isso, desconsidera-se o dado que informa apenas resultados de exames ou hipóteses diagnósticas por mais que esses sejam passíveis de fazer

associações fortes com morbidades do nosso interesse. Sob hipótese alguma os modelos irão indicar ou prever algum diagnóstico, apenas buscam estruturar informações para que os profissionais de saúde possam agir e diagnosticar.

- Considerar erro gramatical desde que fique claro a informação a ser anotada, por exemplo, “diabete melitos”, “isnfuciencia respiratoria”

- Certifique-se de que a base de dados texto que será anotada está anonimizada. Caso nome de pacientes e profissionais sejam muito frequentes nas evoluções, entre em contato com nossa equipe para juntos buscarmos uma forma de ocultá-los.

4. Apresentação do manual para as Tasks (Variáveis)

O formulário é composto por 30 variáveis, divididas em quatro grupos (características clínicas; manejo clínico e processo de cuidado; escalas de avaliação e eventos de risco; indicadores clínicos, desfechos e status do paciente).

Para cada variável é apresentada a definição acrescida de exemplos de sentenças que remetem a classificação do texto em cada variável.

4.1. Características Clínicas

| Variável | Definição | Opções de Resposta | Orientações para resposta / Sentenças |
|--------------------|---|--------------------|---|
| Doença coronariana | CID10 I20 – Angina Pectoris | 0- Doença ausente | # Sem histórico de CI |
| | CID10 I21 – Infarto Agudo do Miocárdio | | # Não possui história progressiva de IAM |
| | CID10 I22 – Infarto do Miocárdio Recorrente | | # Paciente sem histórico de DAC |
| | | | #Miocardiopatia hipertrófica |
| | | | # Angina estável # Angina instável |
| | CID10 I24 – Outras Doenças Isquêmicas Agudas do Coração | 1- Doença presente | # ACTP |
| | CID10 I25 – Doença Isquêmica Crônica do Coração | | # IAM # IAMSSST # IAMCSST # CRM # Cardiopatia isquêmica # CI # DAC # SCA # DCC #Aterosclerose coronariana/coronária #Estenose coronariana/coronária |

| Variável | Definição | Opções de Resposta | Orientações para resposta / Sentenças |
|--|--|--------------------|--|
| Fibrilação Atrial | CID10 I48 - Flutter e fibrilação atrial | 0- Doença ausente | # Sem história progressa de FA # Não apresenta fibrilação atrial |
| | | 1- Doença presente | #FA #Fibrilação atrial # Fibrilação atrial paroxística # ACFA |
| Diabetes | CID10-E10 – Diabetes Mellitus Insulino-dependente | 0- Doença ausente | # Nega DM # Não possui diabetes # Sem história progressa de DM |
| | CID10-E11 – Diabetes Mellitus Não-insulino-dependente | | |
| | CID10-E12 – Diabetes Mellitus Relacionado Com a Desnutrição | 1- Doença presente | # DM # Diabetes mellitus |
| | CID10-E13 – Outros Tipos Especificados de Diabetes Mellitus | | # Diabete mellitus # Diabetes |
| CID10-E14 – Diabetes Mellitus Não Especificado | | # DM2 | |
| AVC prévio | CID10-I64 Acidente vascular cerebral, não especificado como hemorrágico ou isquêmico | 0- Não | # Sem história prévia de AVC |
| | CID10-I63 – Infarto Cerebral | 1- Sim | # AVC isquêmico prévio # AVC ISQ em 2018 # História prévia: AVCi AVCh # Acidente vascular cerebral # Acidente vascular encefálico |
| Hipertensão arterial sistêmica | CID10-I10 – Hipertensão Essencial (primária) | 0- Doença ausente | # Nega HAS. # Não possui história progressa de hipertensão # Sem histórico de HAS |

| | | |
|---|-----------------------|--|
| CID10- I11 – Doença Cardíaca Hipertensiva | 1- Doença presente | # HAS # Hipertensão arterial Hipertensa # História prévia de Has. |
|---|-----------------------|--|

| Variável | Definição | Opções de Resposta | Orientações para resposta / Sentenças |
|--------------|--|--------------------|--|
| Obesidade | CID10- E66 – Obesidade | 0- Doença ausente | - |
| | | 1- Doença presente | # Obesidade # Obeso # IMC > 30 Índice de massa corporal > 30 |
| Dislipidemia | CID10-E78 - Distúrbios do metabolismo de lipoproteínas e outras lipidemias | 0- Doença ausente | # Nega DLP # Sem história progressiva de dislipidemia |
| | | 1- Doença presente | # Dislipidemia # Dislipidêmico # DLP |
| Câncer | Registro de câncer em atividade, indicativo de realização de tratamentos para a doença (radioterapia, quimioterapia) ou indicativo de câncer recente. Quando indicar que o paciente já teve câncer curado ou no passado não deverá ser considerado. | 0- Doença ausente | #2005 câncer de pele #Ca no passado |
| | | 1- Doença presente | # Ca em atividade # Câncer em tratamento #em tto quimioterápico # Neoplasia de próstata ativa |
| Tabagismo | Registro do consumo condicionado a dependência de cigarros ou outros produtos que contenham tabaco (possível relato do paciente). | 0- Não tabagista | # Nega tabagismo # Não tabagista |
| | | 1- Tabagista | # Tabagismo # Tabagista # Fumante |
| | | 2- Ex-tabagista | # Ex-tabagista # Tabagista no passado |

| Variável | Definição | Opções de Resposta | Orientações para resposta / Sentenças |
|----------|--|--------------------|--|
| Etilista | Registro do consumo condicionado a dependência de álcool (possível relato do paciente). O relato de negação ou ex-alcoolista também deve ser considerado | 0- Não etilista | # Nega etilismo # Não etilista |
| | | 1- Etilista | # Etilismo # Alcoolismo # Abuso de álcool # Alcoolista |
| | | 2- Ex etilista | # História prévia de alcoolismo # Ex alcoólatra #Ex elitista |

4.2. Manejo Clínico e Processo de cuidado

| Variável | Definição | Opções de Resposta | Orientações para resposta / Sentenças |
|--------------|---|--|---|
| Localização | Indica o local de presente momento do paciente. ***Sentenças que indicam deslocamento do paciente sem informativo do destino são desconsideradas | 1- Emergência | # Evolução diária- emergência- enfermagem |
| | | 3- CTI | ## CTIA ## # Paciente proveniente da emergência por volta das 15:50h pós trombólise por AVCi. # Interna na UTI após procedimento. # Fisioterapia terapia intensiva |
| | | 4-Unidade Internação | # Enfermagem C1. # Paciente deu entrada na UI |
| Trombectomia | Refere ao possível manejo clínico que o paciente AVCi recebeu. | 0- Não realizou ou sem janela para terapia | # Sem indicação de trombólise ou trombectomia. # Sem janela para trombectomia |
| | *A contraindicação, sem delta ou sem janela também deve ser considerada como não realização do procedimento | 1- Trombectomia | # Trombectomia mecânica – # Trombectomia mecânica as 12 horas |
| Trombólise | Refere ao possível manejo clínico que o paciente AVCi recebeu. | 0- Não realizou ou sem janela para terapia | # Não trombolizada - delta # Contraindicação a trombólise (hipodensidade >1/3). # Sem janela trombolítica |
| | *A contraindicação, sem delta ou sem janela também deve ser considerada como não realização do procedimento. | 1- Trombólise | # Trombolise endovenosa # ALTEPLASE # AVCi trombolizado |

4.3. Escalas de Avaliação e Eventos de Risco

| Variável | Definição | Opções de Resposta | Orientações para resposta/ Sentenças |
|--------------------------|--|------------------------|--|
| Hemorragia intracraniana | CID10 - I61 Hemorragia intracerebral. | 0- Hemorragia ausente | # TC controle ontem sem sangramento # TCC sem alterações # Sem indícios de sangramento |
| | CID10 - I62 Outras hemorragias intracranianas não-traumáticas. | 1- Hemorragia presente | # Hemorragia cerebelar # Hemorragia cerebral com transformação hemorrágica # Hemorragia em fossa posterior próximo ao IV ventrículo. |
| Queda | Registro de queda do paciente durante período de internação. | 1- Sim, queda | # Paciente relata insegurança de usar andador após queda # Paciente perdeu o equilíbrio e apresentou queda da própria altura. # Caiu no quarto a noite |
| Escala Braden | Escala de Braden é um recurso utilizado nas Unidades de Terapia Intensiva para medir o risco dos pacientes críticos de desenvolverem lesões por pressão. | 1- Risco baixo | Baixo risco: escore =>17 |
| | | 2-Risco moderado | Risco moderado: escore 16-13 |
| | | 3-Risco alto | Risco alto: escore =<12 |
| Risco de queda | Variável com objetivo de identificar o risco de queda do paciente de acordo com a escala John Hopkins. Atenta-se também a descritivos da escala que podem ser ajustados de acordo com o risco definido por consenso dos profissionais. | 1- Baixo risco | Baixo risco: escore de 0-5 pontos ou atentar para medidas preventivas de queda. Ex.: # Risco para Quedas (Escala John Hopkins 02); # Manter medidas de prevenção de queda; |
| | | 2- Moderado risco | Risco moderado: escore de 6-13 pontos. Ex.: #Reforço orientações com a familiar sobre o risco de quedas. # risco de queda. *Sentenças que apresentam apenas "risco de queda", sem score ou contexto, podem ser consideradas como risco moderado. |

| | | |
|--|--------------|--|
| | 3-Alto risco | Alto risco: escore > 13 pontos. Implementação do protocolo de risco de queda. Ex.: i) Alto risco de queda (14). ii) Protocolo de alto risco para queda; |
|--|--------------|--|

| Variável | Definição | Opções de Resposta | Orientações para resposta/ Sentenças |
|------------------------|---|---|---|
| Indicativo de Infecção | <p>Sentenças que indiquem possível infecção viral, bacteriana ou fúngica no paciente.</p> <p>Observar: sinais flogísticos (Calor (febre), rubor ou hiperemia (vermelhidão), edema (inchaço), hiperestesia (dor ao toque), perda de função), uso de antibióticos, exames de investigação como hemocultura, início de secreção por infecção</p> <p>**desconsiderar resultados de exames sem indicativo explícito de infecção</p> <p>*** em anexo alguns antimicrobianos comuns na prática clínica</p> | 0- Indicativo negativo de infecção | <p>#Sem sinais flogísticos</p> <p>#Afebril, sinais vitais estáveis</p> <p>#Temp máx 37,2°C</p> <p># Culturais HMCs 09/08: negativas ATQ 09/08: negativo HMCs 12/08: negativas ATQ</p> |
| | 1- Indicativo alerta de infecção | <p># Pneumonia base direita-< BACTRIN</p> <p># EQU: urocultura em andamento.</p> <p>#NOVA ITU - MEROPENEN -> E.COLI</p> <p>#sepse</p> | |

4.4. Indicadores Clínicos Desfechos e Status do Paciente

| Variável | Definição | Opções de Resposta | Orientações para resposta/ Sentenças |
|-------------|--|--|--|
| Óbito | Sentenças que indiquem nota de óbito/falecimento. Indicativos de não-óbito podem ser marcados e são considerados: sinais vitais estáveis, registros de paciente presente no leito, bem, respirando espontaneamente, previsão de alta do paciente. | 0- Indicativos de presença e sinais vitais | # No leito, tranquilo, acompanhado. # Condição Ventilatória: ar ambiente, eupneica # Eupneica, ventilando espontâneo em ar ambiente, sem esforço ventilatório; # Hemodinamicamente estável # SV estáveis |
| | | 1- Nota de óbito | # Nota de falecimento # Paciente faleceu hoje pela manhã |
| Dor | Indica se o paciente apresenta sinais clínicos de dor. **observar: escala EVA de dor, escala PAINAD, escala BPS e composições de palavras que justifiquem marcação. | 0- Sem dor | # Paciente sem queixas # Paciente bem, sem dor #Paciente não refere dor # Sem queixas álgicas # Nega dor. |
| | | 1- Dor leve/moderada | Escores (EVA/PAINAD/BPS) 1 a 6 Ex.: #Dor indefinida #Dor leve #Queixa de dor. |
| | | 2- Dor intensa | Escores (EVA/PAINAD/BPS) 7 a 10 Ex.: #Dor difusa #Dor grave #Dor intensa |
| Alimentação | Indica se o paciente precisa de sonda ou gastrostomia para se alimentar. | 1- Alimentação por via oral | #Paciente alimentando-se por via oral. #Dieta para semi pastosa #Boa aceitação VO. |
| | | 2- Uso de sondas ou gastrostomia | #Paciente alimentando-se por via alternativa # Cuidados com SNE; |

| | | |
|--|--|--|
| | | #Dieta em BI; #Cuidados com gastrostomia #GTT #Dieta polimérica |
|--|--|--|

| Variável | Definição | Opções de Resposta | Orientações para resposta/ Sentenças |
|----------|--|--------------------|--|
| Força | <p>Indica a capacidade de superar ou opor-se a uma resistência por meio da atividade muscular, conforme escala de grau de força, composições descritivas ou nível de mobilidade associado.</p> <p>Levar em conta apenas membros superiores e inferiores (facial não), no caso de uma sentença apresentar vários graus de força, priorizar força de Membros Inferiores e proximais e menor grau apresentado.</p> <p>*Verificar relação com escala NIH</p> | Escore de 0 a 5 | <p>Apresentação da escala:</p> <p>0 – Não percebe contração. Plegia/paciente plégico é marcado com força 0.</p> <p>1 – Traço de contração sem produção de movimento. *Passível de correlação com escala NIH domínio Membros Superiores (5) e Membros Inferiores (6) pontuação 4.</p> <p>2 – Contração fraca, elimina gravidade. *Passível de correlação com escala NIH domínio Membros Superiores (5) e Membros Inferiores (6) pontuação 3.</p> <p>3 – Realiza movimento contra a gravidade, porém sem resistência adicional. *Passível de correlação com escala NIH domínio Membros Superiores (5) e Membros Inferiores (6) pontuação 2 ou 1.</p> <p>4 – Realiza movimento contra a gravidade e resistência externa. *Passível de correlação com escala NIH domínio Membros Superiores (5) e Membros Inferiores (6) pontuação 0.</p> <p>5 - Supera maior quantidade de resistência. Sem alteração em força.</p> |
| | | | <p>Movimento limitado ou fraco, motilidade num padrão abaixo do normal. No que se refere à força muscular, precisão do movimento, amplitude do movimento e a resistência muscular localizada, ou seja, refere-se a um comprometimento parcial, uma perda de força.</p> <p>*Levar em conta apenas membros superiores e inferiores (paresia facial não)</p> |

| Variável | Definição | Opções de Resposta | Orientações para resposta/ Sentenças |
|------------|--|---|--|
| Mobilidade | <p>Variável indica se paciente é capaz de andar com autonomia.</p> <p>Associada com escala de nível de mobilidade utilizada pela instituição (Callen, BL., et al. <i>Medsurg nursing, 2004</i>) ou a descrição dos seus níveis, podendo ser adaptada conforme necessário.</p> <p>***Usualmente associada com grau de força e escala de nível de mobilidade da instituição (se aplicável) conforme orientações de resposta.</p> | <p>0 - Incapaz de andar</p> | <p>Incapacidade de andar podem ser relacionados aos níveis de 1 a 8 na escala de mobilidade, ou descrições como:</p> <ul style="list-style-type: none"> i) paciente dependente: equipe promove trocas de decúbito, posicionamento e ADM; ii) paciente participa com a equipe das trocas de decúbito, posicionamento e ADM; iii) paciente é independente no leito; iv) transferido para a cadeira; v) auxílio mecânico ou 3 pessoas para a cadeira /cadeira de rodas; vi) transferência para a cadeira/de rodas com auxílio de duas pessoas; vii) auxílio de duas pessoas , ortostase e pivô para a cadeira; viii) auxílio de uma pessoa, ortostase e pivô para a cadeira; |
| | | <p>1 - Capaz de andar sem ajuda de outra pessoa ou dispositivo</p> | <p>Capaz de andar sem ajuda de outra pessoa ou dispositivo pode ser relacionados aos níveis de mobilidade de 12 a 15 na escala de mobilidade, ou descrições como:</p> <ul style="list-style-type: none"> i) deambula com um auxiliar de prontidão; ii) deambula independente somente no quarto; iii) deambula fora do quarto, distância menor que um corredor; iv) deambula fora do quarto, distância maior que um corredor. <p>*** Quando mobilidade 1, marcar força 5.</p> |
| | | <p>2 - Capaz de andar com ajuda de outra pessoa ou dispositivo</p> | <p>Capaz de andar com a ajuda de outra pessoa ou dispositivo podem ser relacionados aos níveis de 9 a 11 na escala de mobilidade, ou descrições como:</p> <ul style="list-style-type: none"> i) uma pessoa ao lado de prontidão para transferir para a cadeira; ii) deambula com dois auxiliares; iii) deambula com um auxiliar. <p>*** Quando mobilidade 2, marcar força 4.</p> |

| Variável | Definição | Opções de Resposta | Orientações para resposta/ Sentenças |
|---------------------|---|--------------------|---|
| Nível de mobilidade | <p>Facultativo: variável dependente do uso, pela instituição, de escalas de rastreamento para níveis de mobilidade do paciente.</p> <p>Caso não haja uma escala padrão proposta, podem ser feitas de acordo com as descrições indicadas.</p> <p>Escala de nível de mobilidade utilizada pela instituição (<i>Callen, BL., et al. Medsurg nursing, 2004</i>)</p> | Escore de 1 a 15 | <p>1 - Paciente dependente</p> <p>2- Paciente participa com a equipe das trocas de decúbito, posicionamento e ADM</p> <p>3- Paciente é independente no leito</p> <p>4- Transferido para a cadeira</p> <p>5-auxílio mecânico ou 3 pessoas para a cadeira /cadeira de rodas</p> <p>6- Transferência para a cadeira/de rodas com auxílio de duas pessoas</p> <p>7-Auxílio de duas pessoas, ortostase e pivô para a cadeira</p> <p>8-Auxílio de uma pessoa, ortostase e pivô para a cadeira</p> <p>9-Uma pessoa ao lado de prontidão para transferir para a cadeira</p> <p>10-Deambula com dois auxiliares</p> <p>11-Deambula com um auxiliar</p> <p>12-Deambula com um auxiliar de prontidão</p> <p>13-Deambula independente somente no quarto</p> <p>14-Deambula fora do quarto, distância menor que um corredor</p> <p>15-Deambula fora do quarto, distância maior que um corredor</p> |

| Variável | Definição | Opções de Resposta | Orientações para resposta / Sentenças |
|-------------|---|---------------------------|---|
| Comunicação | <p>Variável indica se paciente tem problemas de comunicação verbal.</p> <p>*Verificar relação com escala NIH.</p> | 0- Sem comunicação verbal | <p>Impossibilidade de efetuar comunicação verbal.</p> <p>Afasia grave, toda a comunicação é feita através de expressões fragmentadas; necessidade de interferência, a quantidade de informação que pode ser trocada é limitada.</p> <p>Ex.:</p> <p># Paciente comunica por gestos.</p> <p>*Passível de correlação com escala NIH domínio Linguagem (9) e Disartria (10) pontuação 2 ou 3.</p> |
| | | 1- Comunica bem | <p>Comunicação sem dificuldades/assintomáticas.</p> <p>Ex.:</p> <p># Paciente dá bom dia;</p> <p># Fala ok;</p> <p># Recuperado dos sintomas de fala;</p> <p># Sem afasia, normal;</p> <p># Comunicativo</p> <p>*Passível de correlação com escala NIH domínio Linguagem (9) e Disartria (10) pontuação 0.</p> |
| | | 2- Comunica pouco ou mal | <p>Comunicação verbal com dificuldade/ problemas.</p> <p>Afasia leve a moderada; perda óbvia de alguma fluência, sem limitação significativa das ideias expressas ou formas de expressão.</p> <p>Ex.:</p> <p># Paciente com afasia de expressão e agora disartria.</p> <p># Fala enrolada</p> <p>*Passível de correlação com escala NIH domínio Linguagem (9) e Disartria (10) pontuação 1.</p> |

| Variável | Definição | Opções de Resposta | Orientações para resposta/ Sentenças |
|----------------------|---|--------------------------------|--|
| Capacidade cognitiva | Variável sobre o entendimento, situação de lucidez e orientação do paciente. | 0- Não estado de clareza | <p>Paciente sem clareza ou em estado de sedação.</p> <p>Ex.:</p> <p># Paciente sedado, não responsivo,</p> <p># Paciente no leito, abertura ocular espontânea, não interage, sem atender comandos.</p> <p>*Passível de correlação com escala NIH domínio Nível de Consciência (1,1a,1b) pontuação 2.</p> |
| | Deve-se sempre avaliar o contexto da sentença, pois palavras como "colaborativo", "bem", "interagindo", "atendem comandos" quando sozinhas não indicam entendimento pleno, porém quando agrupadas em contexto podem traduzir algum discernimento do paciente. | 1- Lúcido, orientado, coerente | <p>Situação de lucidez e orientação do paciente.</p> <p>Ex.:</p> <p># LOC</p> <p># Escala glasgow 15</p> <p># Paciente lúcido, responsivo, tranquilo, obedece a comandos</p> <p>*Passível de correlação com escala NIH domínio Nível de Consciência (1,1a,1b) pontuação 0.</p> |
| | *Verificar relação com escala NIH. | 2- Confuso, compreende mal | <p>Paciente confuso, compreendendo mal, delirium.</p> <p>Ex.:</p> <p># Paciente relata confusão em pensamentos, desorientado em espaço e tempo.</p> <p>*Passível de correlação com escala NIH domínio Nível de Consciência (1,1a,1b) pontuação 1.</p> |

| Variável | Definição | Opções de Resposta | Orientações para resposta / Sentenças |
|---------------|--|--|--|
| Escala Rankin | <p>Escala de avaliação funcional pós-AVC. Preenche a variável conforme indicação do score na evolução do paciente (de 0 a 6).</p> <p>*** Quando rankin 4 ou 5, marcar auto-cuidado = 0</p> | Escore de 0 a 6 | <p>Apresentação da escala:</p> <p>0 - Sem sintomas</p> <p>1 - Nenhuma deficiência significativa, a despeito de sintomas</p> <p>2 - Leve deficiência - Incapaz conduzir todas as atividades de antes, mas é capaz de cuidar dos próprios interesses sem assistência</p> <p>3 - Deficiência moderada -Requer alguma ajuda, mas é capaz de caminhar sem assistência (pode usar bengala ou andador)</p> <p>*4 - Deficiência moderadamente grave - Incapaz de caminhar sem assistência e incapaz de atender às próprias necessidades fisiológicas sem assistência</p> <p>*5 - Deficiência grave - Confinado à cama, incontinente, requerendo cuidados e atenção constante de enfermagem</p> |
| | | | 6 – Óbito |
| Auto cuidado | <p>Capacidade de autonomia em atividades básicas como vestir-se, ir sozinho ao banheiro.</p> <p>*A não autonomia também é indicada, inclusive pela relação com rankin 4 e 5.</p> | 0- Dependente para atividades básicas/de higiene | <p># Eliminações espontâneas, em fraldas</p> <p># Rankin 4</p> <p># Restrito ao leito no momento</p> <p># Precisa de auxílio para as atividades da vida diária.</p> |
| | | 1- Independente para atividades básicas/de higiene | <p># Tem vida de relação, conversa, caminha, alimenta-se.</p> <p># Paciente faz uso banheiro/WC</p> |
| NIH | <p>A National Institute of Health Stroke Scale (NIHSS) é uma escala padrão, validada, segura, quantitativa da severidade e magnitude do déficit neurológico após o AVC.</p> <p>Seu escore varia de 1 a 42, porém não é incomum esta escala</p> | Escore de 1 a 42 | <p># NIH 14 (sonolento 1, errou 1 pergunta, paresia face 2, desvio do olhar 1, paresia MSE 3, MIE3, disartria 1, Heminegligência 2)</p> <p># NIH: face: 1, 2, negligência 2, MSE 3, MIE 3 - Total: 11.</p> <p># NIHSS 5</p> |

| | | |
|---|--|--|
| <p>aparecer de forma a indicar o nível de cada domínio considerado, quando isso acontece os valores podem ser traduzidas a outras variáveis que temos presente no manual, conforme instruções acima.</p> <p>*** em anexo a escala NIH</p> | | |
|---|--|--|

A variável “status_chegada” é uma variável-suporte nos casos em que a sentença diz respeito a uma informação claramente no passado do paciente (por exemplo, situação do paciente na chegada ao hospital), e não refere seu quadro atual; isso não desonera o fato da marcação da variável correspondente.

4.4.1. ANEXO A1 – Informação de apoio para antimicrobianos comuns na prática clínica

| Antimicrobianos comuns na prática clínica | Exemplos |
|---|--|
| Penicilinas | Amoxicilina, Ampicilina, Azlocilina, Carbenicilina, Cloxacilina |
| Cefalosporinas | Cefaclor, Cefadroxilo, Cefazolina, Cefixime., Cefoperazona, Cefotaxima, Cefotetan, Cefoxitina, Ceftazidima, Ceftriaxona, Cefuroxima, Cefalexina, Cefalotina, Loracarbef. |
| Quinolonas | Ciprofloxacina, Enoxacina, Levofloxacina, Lomefloxacina, Moxifloxacina, Norfloxacina, Ofloxacina. |
| Aminoglicosídeos | Amicacina, Gentamicina, Canamicina, Neomicina, Estreptomicina, Tobramicina. |
| Macrolídios | Azitromicina, Claritromicina, Eritromicina. |
| Tetraciclínas | Doxiciclina, Minociclina, Tetraciclínas. |
| Outros exemplos | Aztreonam, Clindamicina, Etambutol, Fosfomicina, Imipenem, Isoniazida, Meropenem, Metronidazol, Nitrofurantoina, Pirazinamida, Rifampicina, Trimetoprima, Fluconazol |

4.4.2. ANEXO A2 – Escala NIHSS e definição de cada domínio

| Instrução | Definição da escala |
|---|--|
| <p>1a. Nível de Consciência</p> <p>O investigador deve escolher uma resposta mesmo se uma avaliação completa é prejudicada por obstáculos como um tubo orotraqueal, barreiras de linguagem, trauma ou curativo orotraqueal. Um 3 é dado apenas se o paciente não faz nenhum movimento (outro além de postura reflexa) em resposta à estimulação dolorosa.</p> | <p>0 = Alerta; reponde com entusiasmo.</p> <p>1= Não alerta, mas ao ser acordado por mínima estimulação obedece, responde ou reage.</p> <p>2= Não alerta, requer repetida estimulação ou estimulação dolorosa para realizar movimentos (não estereotipados).</p> <p>3= Responde somente com reflexo motor ou reações autonômicas, ou totalmente irresponsivo, flácido e irreflexo.</p> |
| <p>1b. Perguntas de Nível de Consciência</p> <p>O paciente é questionado sobre o mês e sua idade. A resposta deve ser correta – não há nota parcial por chegar perto. Pacientes com afasia ou esturpor que não compreendem as perguntas irão receber 2. Pacientes incapacitados de falar devido a intubação orotraqueal, trauma orotraqueal, disartria grave de qualquer causa, barreiras de linguagem ou qualquer outro problema não secundário a afasia receberá um 1. É importante que somente a resposta inicial seja considerada e que o examinador não “ajude” o paciente com dicas verbais ou não verbais.</p> | <p>0= Responde ambas as questões corretamente.</p> <p>1= Responde uma questão corretamente.</p> <p>2= Não responde nenhuma questão corretamente.</p> |
| <p>1c. Comandos de Nível de Consciência</p> <p>O paciente é solicitado a abrir e fechar os olhos e então abrir e fechar a mão não parética. Substitua por outro comando de um único passo se as mãos não podem ser utilizadas. É dado crédito se uma tentativa inequívoca é feita, mas não completada devido à fraqueza. Se o paciente não responde ao comando, a tarefa deve ser demonstrada a ele (pantomima) e o resultado registrado (i.e., segue um, nenhum ou ambos os comandos). Aos pacientes com trauma, amputação ou outro impedimento físico devem ser dados comandos únicos compatíveis. Somente a primeira tentativa é registrada.</p> | <p>0= Realiza ambas as tarefas corretamente.</p> <p>1= Realiza uma tarefa corretamente.</p> <p>2= Não realiza nenhuma tarefa corretamente.</p> |

| | |
|---|---|
| <p>2. Melhor olhar conjugado</p> <p>Somente os movimentos oculares horizontais são testados. Movimentos oculares voluntários ou reflexos (óculo-cefálico) recebem nota, mas a prova calórica não é usada.</p> <p>Se o paciente tem um desvio conjugado do olhar, que pode ser sobreposto por atividade voluntária ou reflexa, o escore será 1. Se o paciente tem uma paresia de nervo periférica isolada (NC III, IV ou VI), marque 1. O olhar é testado em todos os pacientes afásicos. Os pacientes com trauma ocular, curativos, cegueira preexistente ou outro distúrbio de acuidade ou campo visual devem ser testados com movimentos reflexos e a escolha feita pelo investigador. Estabelecer contato visual e, então, mover-se perto do paciente de um lado para outro, pode esclarecer a presença de paralisia do olhar.</p> | <p>0= Normal.</p> <p>1= Paralisia parcial do olhar. Este escore é dado quando o olhar é anormal em um ou ambos os olhos, mas não há desvio forçado ou paresia total do olhar.</p> <p>2= Desvio forçado ou paralisia total do olhar que não podem ser vencidos pela manobra óculo-cefálica.</p> |
| <p>3. Visual</p> <p>OS campos visuais (quadrantes superiores e inferiores) são testados por confrontação, utilizando contagem de dedos ou ameaça visual, conforme apropriado. O paciente deve ser encorajado, mas se olha para o lado do movimento dos dedos, deve ser considerado como normal. Se houver cegueira unilateral ou enucleação, os campos visuais no olho restante são avaliados. Marque 1 somente se uma clara assimetria, incluindo quadrantanopsia, for encontrada. Se o paciente é cego por qualquer causa, marque 3. Estimulação dupla simultânea é realizada neste momento. Se houver uma extinção, o paciente recebe 1 e os resultados são usados para responder a questão 11.</p> | <p>0= Sem perda visual.</p> <p>1= Hemianopsia parcial.</p> <p>2= Hemianopsia completa.</p> <p>3= Hemianopsia bilateral (cego, incluindo cegueira cortical).</p> |
| <p>4. Paralisia Facial</p> <p>Pergunte ou use pantomima para encorajar o paciente a mostrar os dentes ou sorrir e fechar os olhos. Considere a simetria de contração facial em resposta a estímulo doloroso em paciente pouco responsivo ou incapaz de compreender. Na presença de trauma /curativo facial, tubo orotraqueal, esparadrapo ou outra barreira física que obscureça a face, estes devem ser removidos, tanto quanto possível.</p> | <p>0= Movimentos normais simétricos.</p> <p>1= Paralisia facial leve (apagamento de prega nasolabial, assimetria no sorriso).</p> <p>2= Paralisia facial central evidente (paralisia facial total ou quase total da região inferior da face).</p> <p>3= Paralisia facial completa (ausência de movimentos faciais das regiões superior e inferior da face).</p> |
| <p>5. Motor para braços</p> <p>O braço é colocado na posição apropriada: extensão dos braços (palmas para baixo) a 90° (se sentado) ou a 45° (se deitado). É valorizada queda do braço se esta ocorre antes de 10 segundos. O paciente afásico é encorajado através de firmeza na voz e de pantomima, mas não com estimulação dolorosa. Cada membro é testado isoladamente, iniciando pelo braço não-parético. Somente em</p> | <p>0= Sem queda; mantém o braço 90° (ou 45°) por 10 segundos completos.</p> <p>1= Queda; mantém o braço a 90° (ou 45°), porém este apresenta queda antes dos 10 segundos completos; não toca a cama ou outro suporte.</p> <p>2= Algum esforço contra a gravidade; o braço não atinge ou não mantém 90° (ou 45°), cai na cama, mas tem alguma força contra a gravidade.</p> <p>3= Nenhum esforço contra a gravidade; braço</p> |

| | |
|--|--|
| <p>caso de amputação ou de fusão de articulação no ombro, o item deve ser considerado não-testável (NT), e uma explicação deve ser escrita para esta escolha.</p> | <p>despenca. 4= Nenhum movimento. NT = Amputação ou fusão articular, explique: _____</p> <p>5a. Braço esquerdo 5b. Braço direito</p> |
| <p>6. Motor para pernas A perna é colocada na posição apropriada: extensão a 30° (sempre na posição supina). É valorizada queda do braço se esta ocorre antes de 5 segundos. O paciente afásico é encorajado através de firmeza na voz e de pantomima, mas não com estimulação dolorosa. Cada membro é testado isoladamente, iniciando pela perna não-parética. Somente em caso de amputação ou de fusão de articulação no quadril, o item deve ser considerado não-testável (NT), e uma explicação deve ser escrita para esta escolha.</p> | <p>0= Sem queda; mantém a perna a 30° por 5 segundos completos. 1= Queda; mantém a perna a 30°, porém esta apresenta queda antes dos 5 segundos completos; não toca a cama ou outro suporte. 2= Algum esforço contra a gravidade; a perna não atinge ou não mantém 30°, cai na cama, mas tem alguma força contra a gravidade. 3= Nenhum esforço contra a gravidade; perna despenca. 4= Nenhum movimento. NT = Amputação ou fusão articular, explique: _____</p> <p>6a. Perna esquerda 6b. Perna direita</p> |
| <p>7. Ataxia de membros Este item é avaliado se existe evidência de uma lesão cerebelar unilateral. Teste com os olhos abertos. Em caso de defeito visual, assegure-se que o teste é feito no campo visual intacto. Os testes índex-nariz e calcanhar-joelho são realizados em ambos os lados e a ataxia é valorizada, somente, se for desproporcional à fraqueza. A ataxia é considerada ausente no paciente que não pode entender ou está hemiplégico. Somente em caso de amputação ou de fusão de articulações, o item deve ser considerado não-testável (NT), e uma explicação deve ser escrita para esta escolha. Em caso de cegueira, teste tocando o nariz, a partir de uma posição com os braços estendidos.</p> | <p>0= Ausente. 1= Presente em 1 membro. 2= Presente em dois membros. NT = Amputação ou fusão articular, explique: _____</p> |

| | |
|--|---|
| <p>8. Sensibilidade</p> <p>Avalie sensibilidade ou mímica facial ao beliscar ou retirada do estímulo doloroso em paciente torporoso ou afásico. Somente a perda de sensibilidade atribuída ao AVC é registrada como anormal e o examinador deve testar tantas áreas do corpo (braços [exceto mãos], pernas, tronco e face) quantas forem necessárias para checar acuradamente uma perda hemisensitiva. Um escore de 2, “grave ou total” devem ser dados somente quando uma perda grave ou total da sensibilidade pode ser claramente demonstrada. Portanto, pacientes em esturpor e afásicos irão receber provavelmente 1 ou 0. O paciente com AVC de tronco que tem perda de sensibilidade bilateral recebe 2. Se o paciente não responde e está quadriplégico, marque 2. Pacientes em coma (item 1a=3) recebem arbitrariamente 2 neste item.</p> | <p>0= Normal; nenhuma perda.</p> <p>1= Perda sensitiva leve a moderada; a sensibilidade ao beliscar é menos aguda ou diminuída do lado afetado, ou há uma perda da dor superficial ao beliscar, mas o paciente está ciente de que está sendo tocado.</p> <p>2= Perda da sensibilidade grave ou total; o paciente não sente que está sendo tocado.</p> |
| <p>9. Melhor linguagem</p> <p>Uma grande quantidade de informações acerca da compreensão pode obtida durante a aplicação dos itens precedentes do exame. O paciente é solicitado a descrever o que está acontecendo no quadro em anexo, a nomear os itens na lista de identificação anexa e a ler da lista de sentença anexa. A compreensão é julgada a partir destas respostas assim como das de todos os comandos no exame neurológico geral precedente. Se a perda visual interfere com os testes, peça ao paciente que identifique objetos colocados em sua mão, repita e produza falas. O paciente intubado deve ser incentivado a escrever. O paciente em coma (Item 1A=3) receberá automaticamente 3 neste item. O examinador deve escolher um escore para pacientes em estupor ou pouco cooperativos, mas a pontuação 3 deve ser reservada ao paciente que está mudo e que não segue nenhum comando simples.</p> | <p>0= Sem afasia; normal.</p> <p>1= Afasia leve a moderada; alguma perda óbvia da fluência ou dificuldade de compreensão, sem limitação significativa das ideias expressão ou forma de expressão. A redução do discurso e/ou compreensão, entretanto, dificultam ou impossibilitam a conversação sobre o material fornecido. Por exemplo, na conversa sobre o material fornecido, o examinador pode identificar figuras ou item da lista de nomeação a partir da resposta do paciente.</p> <p>2= Afasia grave; toda a comunicação é feita através de expressões fragmentadas; grande necessidade de interferência, questionamento e adivinhação por parte do ouvinte. A quantidade de informação que pode ser trocada é limitada; o ouvinte carrega o fardo da comunicação. O examinador não consegue identificar itens do material fornecido a partir da resposta do paciente.</p> <p>3= Mudo, afasia global; nenhuma fala útil ou compreensão auditiva.</p> |
| <p>10. Disartria</p> <p>Se acredita que o paciente é normal, uma avaliação mais adequada é obtida, pedindo-se ao paciente que leia ou repita palavras da lista anexa. Se o paciente tem afasia grave, a clareza da articulação da fala espontânea pode ser graduada. Somente se o paciente estiver intubado ou tiver outras barreiras físicas a produção da fala, este item deverá ser considerado não testável (NT). Não diga ao paciente por que ele está sendo testado.</p> | <p>0= Normal.</p> <p>1= Disartria leve a moderada; paciente arrasta pelo menos algumas palavras, e na pior das hipóteses, pode ser entendido, com alguma dificuldade.</p> <p>2= Disartria grave; fala do paciente é tão empastada que chega a ser ininteligível, na ausência de disfasia ou com disfasia desproporcional, ou é mudo/anártrico.</p> <p>NT = Intubado ou outra barreira física; explique _____</p> |

| | |
|---|---|
| <p>11. Extinção ou Desatenção (antiga negligência) Informação suficiente para a identificação de negligência pode ter sido obtida durante os testes anteriores. Se o paciente tem perda visual grave, que impede o teste da estimulação visual dupla simultânea, e os estímulos cutâneos são normais, o escore é normal. Se o paciente tem afasia, mas parece atentar para ambos os lados, o escore é normal. A presença de negligência espacial visual ou anosagnosia pode também ser considerada como evidência de negligência. Como a anormalidade só é pontuada se presente, o item nunca é considerado não testável.</p> | <p>0= Nenhuma anormalidade. 1= Desatenção visual, tátil, auditiva, espacial ou pessoal, ou extinção à estimulação simultânea em uma das modalidades sensoriais. 2= Profunda hemi-desatenção ou hemidesatenção para mais de uma modalidade; não reconhece a própria mão e se orienta somente para um lado do espaço.</p> |
|---|---|

