UNIVERSIDADE FEDERAL DO CEARÁ

CENTRO DE CIÊNCIAS

DEPARTAMENTO DE COMPUTAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

DAVIDSON ALVES NUNES

USO CLÍNICO DOS CÓDIGOS CID: UM ESTUDO EXPLORATÓRIO EM LARGA ESCALA

FORTALEZA

2021

DAVIDSON ALVES NUNES

USO CLÍNICO DOS CÓDIGOS CID: UM ESTUDO EXPLORATÓRIO EM LARGA ESCALA

Dissertação apresentada ao Curso de do Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Ciência da Computação. Área de Concentração: Engenharia de Software

Orientador: Prof. Dr. João Bosco Ferreira Filho

FORTALEZA

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

N924u Nunes, Davidson Alves.
Uso clínico dos códigos CID : Um estudo exploratório em larga escala / Davidson Alves Nunes. – 2021.
65 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação, Fortaleza, 2021.
Orientação: Prof. Dr. João Bosco Ferreira Filho.

1. Classificação Internacional de Doenças. 2. software. 3. administradora de planos de saúde. 4. distribuições estatísticas. 5. estatísticas de assistência médica. I. Título.

CDD 005

DAVIDSON ALVES NUNES

USO CLÍNICO DOS CÓDIGOS CID: UM ESTUDO EXPLORATÓRIO EM LARGA ESCALA

Dissertação apresentada ao Curso de do Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Ciência da Computação. Área de Concentração: Engenharia de Software

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. João Bosco Ferreira Filho (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Luciano Heitor Gallegos Marin
Universidade Federal do Paraná

Prof. Dr. Flávio Rubens de Carvalho Sousa
Universidade Federal do Ceará (UFC)

Prof. Dr. João Paulo Pordeus Gomes
Universidade Federal do Ceará (UFC)

À minha família, esposa (Renata Adele de Lima Nunes) e filho (Ivan César de Lima Nunes). Aos serviços de saúde e às pessoas da linha de frente, que diante das grandes dificuldades, continuam a acolher a população.

AGRADECIMENTOS

Ao Prof. Dr. João Bosco Ferreira Filho por me orientar em minha dissertação de mestrado.

Ao Prof. Dr. Tadeu Mello e Souza, professor da Universidade Federal do Rio Grande do Sul (UFRGS) do Programa de Pós-Graduação em Neurociência que me atentou às diversas possibilidades de distribuições estatísticas possíveis para os dados coletados.

Aos professores do MDCC que realizaram ótimas cadeiras. O novo conhecimento adquirido ajudou bastante no desenvolvimento de minhas habilidades em computação.

A todos da Empresa INTMED que me receberam tão bem e se dispuseram a ajudar nos processos de obtenção de dados, sugestões e implantação das ferramentas. Agradeço especialmente a Matheus Souza de Carvalho, Bruno Barreto Freitas e Rafael da Rocha Borges.

Ao Doutorando em Ciência da Computação, Thiago Queiroz de Oliveira, companheiro de trabalho na universidade.

Ao Doutorando em Engenharia Elétrica, Ednardo Moreira Rodrigues, e seu assistente, Alan Batista de Oliveira, aluno de graduação em Engenharia Elétrica, pela adequação do *template* utilizado neste trabalho para que o mesmo ficasse de acordo com as normas da biblioteca da Universidade Federal do Ceará (UFC).

Aos médicos Renata Adele de Lima Nunes, José Alberto Alves Oliveira e a todos os médicos que participaram respondendo questionários ou fazendo sugestões para o engrandecimento do trabalho aqui apresentado.

E à Fundação Cearense de Apoio ao Desenvolvimento (FUNCAP), na pessoa do Presidente Tarcísio Haroldo Cavalcante Pequeno pelo financiamento da pesquisa de mestrado via bolsa de estudos.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

CIAP	Classificação Internacional de Atenção Primária
CID	Classificação Internacional de Doenças
CIF	Classificação Internacional de Funcionalidade
DATASUS	Departamento de Informática do Sistema Único de Saúde
EHR	Eletronic Health Records, prontuários eletrônicos em saúde
INTMED	Empresa que mantém o Software de atendimentos clínicos dos pontos de atendimento dos serviços de saúde abordados
OMS	Organização Mundial de Saúde
REGEX	Expressão Regular
SAM	Sistema de Atendimento Médico
SNOMED	Systematized Nomenclature of Human Medicine
SUS	Sistema Único de Saúde
UFRGS	Universidade Federal do Rio Grande do Sul

SUMÁRIO

1 INTRODUÇÃO

A Engenharia de *Software* é a disciplina da engenharia responsável por tratar todos os passos da produção de *softwares*. Estuda desde as primeiras iniciativas, como as especificações da construção de sistemas, até as suas manutenções. Para isso, os engenheiros aplicam teorias, métodos e ferramentas no sentido de planejar e solucionar problemas. Esta abordagem sistemática da produção de *softwares* inclui todos os aspectos dos projetos, incluindo as teorias que dão apoio à produção de *software* (IAN, 2003).

Quase todos os países dependem de complexos sistemas. Tais softwares usam diversos algoritmos, com muitas linhas de código, envolvendo muitas equipes envolvidas no desenvolvimento e manutenção (IAN, 2003). Os sistemas de informação estão em constante evolução para apoiar a sociedade moderna. A criação e evolução rápida dos *softwares* acontecem para satisfazer muitas demandas em muitas áreas (HAMEED, 2003).

Dentre os diversos usos dos *softwares*, está a aplicação destes em: 1) repositórios de conhecimento, sistemas que oferecem dados relevantes aos usuários; 2) instrução e supervisionamento de pessoas, gerando conselhos aos usuários e 3) relato de medições, onde por meio de sensores, os usuários recebem *feedbacks* dos dispositivos (TSENG; FOGG, 1999).

Os repositórios de conhecimento em saúde são muito importantes. Os sistemas que informatizam a ficha clínica substituíram os antigos registros de papel (COSTA; ORLOVSKI, 2013). A maioria das instituições de saúde aderiu às novas técnicas operacionais como repositórios, instrução, supervisionamento e medição. Os registros médicos eletrônicos melhoram a prestação dos cuidados. Contudo vão além disso. Essa aquisição de dados funciona como um catalisador para o desenvolvimento da prestação de cuidados em saúde (WILLIAMS; BOREN, 2008). Os dados históricos em saúde são tão importantes que estão protegidos por leis específicas na maioria dos países do mundo (DEHNAVI; BAGHINI, 2019; GERUM, 2015). O intrincado processo de aquisição e manutenção da informação dos profissionais de saúde depende do formulário preenchido pelo médico, dos mecanismos de salvamentos e recuperação de dados (história clínica do paciente), até a disponibilização destes para os muitos interessados. A instrução para a atividade ajuda os médicos e gestores nos prognósticos. Modelos de Conjuntos Fuzzy, algoritmos evolutivos, redes neurais, modelos baseados na estatística e na teoria da informação ajudam a direcionar a tomada de decisão (BONISSONE, 2006). Iniciativas de supervisionamento registram as atividades profissionais (HERBST; JUVEKAR; BHATTACHARJEE; BANGHA *et al.*, 2015). Tais informatizações permitiram o melhor acompanhamento do histórico médico e

auditorias facilitadas ao serviço, entre outros benefícios.

A demanda por informação aumenta a cada dia, tanto na produção como no consumo dos dados, gerando uma revolução na saúde. O correto e rápido compartilhamento de informações levam a produção de novas diretrizes clínicas baseadas em evidências. O Secretário de Estado do Reino Unido, no momento da implementação do *National Health Service Plan*, em 2001, destacou a difícil tarefa de implantar sistemas de saúde modernos, que possam atender as expectativas do fornecimento de informações. Chamou a atenção para a dificuldade de formalizar processos eficazes de tratar e gerar informações de qualidade. Naquela ocasião, chamou o núcleo de informações de "coração de qualquer organização em saúde" (HAMEED, 2003).

No Brasil, esta iniciativa é realizada pelo Ministério da Saúde com financiamento para o registro e disponibilização de dados no Departamento de Informática do Sistema Único de Saúde (DATASUS). Erros no preenchimento dos códigos pelas unidades públicas de saúde são classificados como dados nulos e impactam diretamente no repasse dos valores pagos pelo Sistema Único de Saúde (SUS) (LOPES; BRASIL, 2003). O mesmo fenômeno do erro no registro das informações acontece em clínicas conveniadas a planos de saúde, que realizam a glosa, negação ou retenção do pagamento por serviços prestados, dos procedimentos realizados (DOS SANTOS; DA ROSA, 2013).

Os dados são de tamanha importância, que existe uma Política Nacional de Disponibilização de Dados (BERTOLINI; FORTUNA; VIDAL; NEVES *et al.,*2020) A CGU, Controladoria Geral da União, órgão de fiscalização técnica e controle federal, recomenda a abertura completa das bases de dados dos serviços de planos de saúde, cuidando para não violar a privacidade do cidadão (ANS, 2019). Essa iniciativa facilita o acesso aos registros do Classificação Internacional de Doenças (CID).

Diante do problema do registro em saúde, a Organização Mundial de Saúde (OMS) cria a referência internacional para a identificação e codificação de doenças e condições de saúde. A funcionalidade CID é a base para o armazenamento, o compartilhamento e a recuperação de dados para análises futuras. Diversos dados epidemiológicos são calculados a partir dos registros gerados nos serviços de saúde (ORGANIZATION, 2021). Os códigos CID descrevem doenças e condições de saúde em todo o mundo (OTERO VARELA; DOKTORCHIK; WIEBE; QUAN *et al.*, 2021; WINKLER; OTT; BECHER, 2010; REILLY; SHULMAN; GILBERT; JOMON *et al.*, 2020).

A padronização da codificação pelo CID ajudou bastante a produção de bancos de

dados mundiais em saúde. Contudo, escolher o código mais apropriado para os contextos gera erros (O'MALLEY; COOK; PRICE; WILDES *et al.*, 2005; HAMEED, 2003).

– códigos que não se aplicam ao achado clínico.

– códigos gerais, supercódigos, não geram quase nenhuma informação para os receptores.

– códigos sem manutenção podem permanecer sem atualização, consequentemente errados.

Os médicos podem produzir códigos inespecíficos ou menos desejáveis, se não forem devidamente treinados (HORSKY; DRUCKER; RAMELSON, 2017).

Os CIDs suportam uma ampla gama de tarefas diferentes: acesso aos dados do paciente para recuperação e suporte à decisão, banco de dados experimental, comparação de casos clínicos, agrupamento de casos clínicos semelhantes, epidemiologia clínica, geração de estatísticas, pesquisa, interoperabilidade semântica, gerenciamento de qualidade, estruturação de repositórios de literatura, faturamento e contabilidade (SCHULZ; KLEIN, 2008). Existem muitos códigos disponíveis no CID. A ferramenta tenta suportar a codificação para diferentes especialidades; no entanto, ainda existem reclamações quanto à insuficiência de códigos (LAURENTI; NUBILA; QUADROS; CONDE *et al.*, 2013). Os códigos estão em constante evolução, à medida que o conhecimento em saúde avança (FUNG; XU; BODENREIDER, 2020).

O CID soluciona parcialmente o problema de codificação. Alguns países, observaram algumas deficiências da própria ferramenta. No sentido de adaptar os códigos a suas realidades locais, mas sem abandonar a primeira iniciativa da OMS, os estados nacionais adequaram-na, a exemplo do: CID-10-AM (Austrália), CID-10-CA (Canadá), CID-10-GM (Alemanha), CID-10-TM (Tailândia) e CID-10-CM (Estados Unidos) (JETTÉ; QUAN; HEMMELGARN; DROSLER *et al.*, 2010).

Outras formas de contornar deficiências do CID justificaram a formação de outras estruturas de codificação. A Classificação Internacional de Atenção Primária (CIAP) foi produzida para codificar melhor contextos sociais e a atenção primária (VAN MENS; ELZINGA; NIELEN; LOKKERBOL *et al.*, 2020). No domínio das deficiências físicas, a Classificação Internacional de Funcionalidade (CIF) abordou os estados de saúde dando foco a elementos incapacitantes (BÖLTE; LAWSON; MARSCHIK; GILDLER, 2021). Por exemplo, abordam as deficiências das funções da visão e suas estruturas correlatas,funções mentais ou psicológicas, desvios de padrões populacionais, etc.

As evoluções na codificação geraram a *Systematized Nomenclature of Human Medicine* (SNOMED) que pode vir até a substituir o CID. Essa nova ferramenta incorpora princípios

de lógica e ontologia. Contudo, ainda está longe de ser tão amplamente adotada quanto o CID (SCHULZ; KLEIN, 2008).

O fato é que todas as classificações de morbidade na atenção primária têm alguma ligação com a CID. Mesmo a SNOMED instituiu um mapeamento semântico correlacionado ao CID (RODRIGUES; ROBISON; DELLA MEA; CAMPBELL *et al.*,2015). Inclusive, os Estados Membros da OMS assumiram a responsabilidade de adotar o CID e o usam como padrão para a publicação de estatísticas mundiais (KARJALAINEN; ORGANIZATION, 1999). Atualmente, o CID-10 ainda é o padrão de codificação (DI NUBILA; BUCHALLA, 2008).

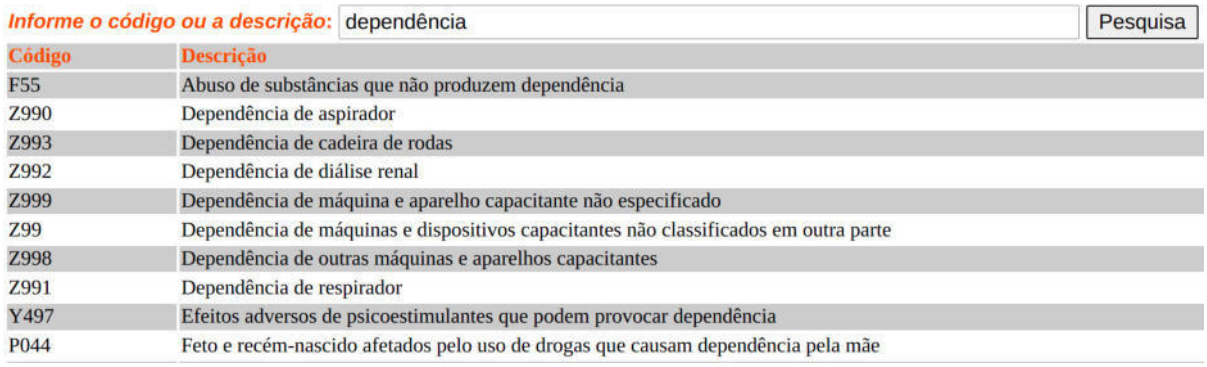
No consultório médico, o preenchimento do campo "código CID", geralmente, é compulsório e bloqueante, impedindo o salvamento do formulário eletrônico. Preencher tal campo é uma tarefa difícil, contudo facilmente negligenciada se o médico usar um código tão geral quanto os seguintes exemplos: T14 - traumatismo de região não especificada do corpo, B99 - Doenças infecciosas, outras e as não especificadas, R52.9 - dor não especificada, etc.

Estas fichas eletrônicas também não podem ser estáticas, sem a possibilidade de modificações ou acréscimo. Isso é um problema. Com a aquisição de novos dados, a partir das ações em saúde (exames, novas consultas, avaliação de históricos, etc.), os CIDs poderão ser atualizados. Códigos CIDs em primeiras consultas e emergências são naturalmente mais genéricos. Na triagem, por exemplo, as instituições médicas admitem códigos gerais, com poucas informações. Conforme ocorre a continuidade do processo analítico, há maior compreensão entre as partes gerando até mesmo a alteração do diagnóstico inicial e consequente código (O'MALLEY; COOK; PRICE; WILDES *et al.*, 2005) Na vida, os indivíduos atravessarão diversas fases, com o tempo, os registros precisam ser obrigatoriamente atualizados. Uma importante informação pode ser perdida se o sistema não tratar do registro cronológico dos diversos códigos aplicados.

Algumas tecnologias já são associadas a procura por códigos CID, como exemplo:

- Expressão Regular (REGEX), conforme Figura 1. O profissional de saúde digita o nome da patologia e é auxiliado por sugestões de *strings* semelhantes (PRUDENTE, 2020).
- ligações automáticas de registros. Exemplo disso, o sistema de previdência social pode gerar um benefício ao portador de certos tipos de câncer capturando dados da ficha (CONTIERO; TITTARELLI; TAGLIABUE; MAGHINI *et al.*, 2005).
- acesso a *Cloud* de históricos clínicos e diagnósticos anteriores e suas codificações (OH; CHA; JI; KANG *et al.*, 2015; FORCHESATTO; SANTIN, 2013).

Figura 1 – Mostra da busca de códigos CID por expressão regular.



Código	Descrição
F55	Abuso de substâncias que não produzem dependência
Z990	Dependência de aspirador
Z993	Dependência de cadeira de rodas
Z992	Dependência de diálise renal
Z999	Dependência de máquina e aparelho capacitante não especificado
Z99	Dependência de máquinas e dispositivos capacitantes não classificados em outra parte
Z998	Dependência de outras máquinas e aparelhos capacitantes
Z991	Dependência de respirador
Y497	Efeitos adversos de psicoestimulantes que podem provocar dependência
P044	Feto e recém-nascido afetados pelo uso de drogas que causam dependência pela mãe

Fonte: Prefeitura Municipal de Presidente Prudente (2021).

No sentido de gerar a produção de funcionalidades melhores, propomos o estudo exploratório dos códigos CIDs históricos de um grande administrador de planos de saúde brasileiro. É importante entender os critérios de codificação empregados nos atendimentos para identificar padrões e ferramentas de simplificação da atividade de codificação.

Desta forma, o tempo dispendido na consulta, na tarefa de eleger o melhor código, seria reduzido. O menor atendimento resultaria, então, em uma economia de recursos para o serviço de saúde.

A pergunta que deu origem a este trabalho foi: É possível facilitar a escolha de códigos de saúde ou partes dessa tarefa a partir do estudo de dados históricos robustos? Ao delimitar ainda mais a pergunta obtemos as seguintes perguntas de pesquisa:

- **1- Quais são os códigos CID mais prevalentes nos serviços analisados?**
- **2- Como tais códigos podem ser agrupados no domínio médico?**
- **3- Os agrupamentos podem gerar um algoritmo para sugestões destes códigos?**

2 FUNDAMENTAÇÃO TEÓRICA

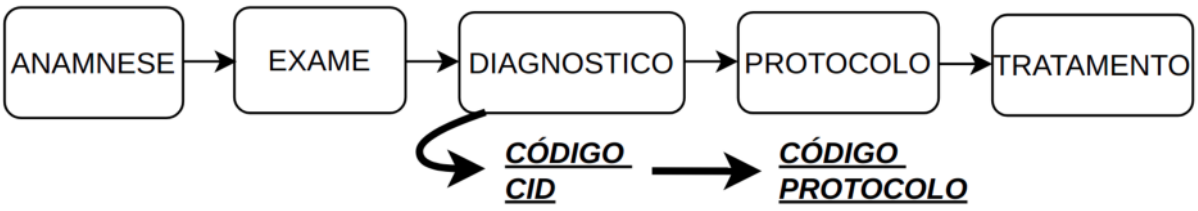
2.1 Codificação Clínica

A nosologia (classificação sistemática de doenças) já era praticada para identificar causas de patologias nas sociedades ocidentais dos séculos XVII e XVIII (O'MALLEY; COOK; PRICE; WILDES *et al.*, 2005). Esta ciência evoluiu bastante ao ponto de permitir a criação de códigos classificatórios e agrupamentos das doenças conforme causas.

No século XX, os seguros médicos, impulsionados pelas entidades pagadoras de serviços públicos e privados, adotaram práticas de identificação e codificação dos achados e procedimentos (O'MALLEY; COOK; PRICE; WILDES *et al.*, 2005).

O processo de trabalho dos profissionais de saúde envolve: a definição de diagnósticos, a escolha por protocolos clínicos e a aplicação de tratamentos. O conjunto de informações recolhidas pelo médico a respeito de um doente e de sua doença, pode ser chamada de anamnese. As associações entre as manifestações das doenças, sinais e sintomas, são os atributos que fundamentam e discriminam as patologias. Isto é, em uma consulta, os profissionais da saúde fazem perguntas, observam exames, avaliam históricos médicos de prontuários para, a partir de seus conhecimentos científicos e de suas experiências, identificar a enfermidade. Os tratamentos são elencados em resposta à enfermidade (CRUZ; PIMENTA, 2005). Com a evolução da computação, os códigos das patologias passaram a alimentar os bancos de dados de saúde. Neste sentido, a atividade "codificação clínica" é realizada paralelamente ao diagnóstico e passo subsequentes. A figura 2 mostra, simplificadamente, um processo comum de atendimento, destacando a geração de códigos de CID e protocolos.

Figura 2 – Processo de atendimento médico e codificação.



Fonte: elaborado pelo autor.

No sentido de buscar uma padronização, os códigos foram criados para facilitar:

– a comunicação e a transferência de informações médicas nos contextos locais e internacionais.

- criação e manutenção de banco de dados em saúde, gerando, posteriormente, mais eficiência na pequisa e segurança da informação produzida.
- a tradução clara, mesmo em contextos com barreira de idiomas.
- a identificação da patologia para garantir tratamento adequado.
- o gerenciamento de recursos.
- a geração de declarações e de atestados.
- referências estatísticas e epidemiológicas (ORGANIZATION, 2021).

Dentre os benefícios da codificação, o contexto estatístico e epidemiológico dos códigos foi o mais aplicado neste estudo. Para a vigilância epidemiologia, outros benefícios são:

- cálculo de indicadores;
- definição de metas de equipes de saúde;
- controle de surtos: atitudes dos serviços disparados por determinados códigos;
- desvios de comportamento profissionais;
- controle de prevalências: recursos para problemas frequentes, localmente ou não;
- vigilância à saúde do trabalhador: avalia indiretamente condições de trabalho inadequadas;
- controle de pré-natal e cuidados aos recém-nascidos;
- identificação de procedimentos conveniados ou autorizados de serviços de saúde públicos e privados;
- identificação de áreas de atuação de profissionais, como cirurgiões-dentistas, enfermeiros, médicos, fisioterapeutas e demais trabalhadores em saúde. (LOPES; BRASIL, 2003)

Enfim, vale a pena aplicar e estudar os códigos dos serviços, pois geram conhecimentos para o gerenciamento racional dos recursos. Tal modelo é chamado de Modelo de Gestão em Saúde Baseada em Evidências. Evidências científicas, atenção sistemática aos fatos organizacionais, pensamento crítico sobre as informações e considerações éticas devem fundamentar as ações e soluções. Tal atitude promove diminuições de custo, melhora da qualidade e aumento dos resultados positivos (ROUSSEAU, 2012).

2.1.1 Semântica dos Códigos em Saúde

A Semântica dos códigos em saúde refere-se ao estudo dos significados e sentidos das representações da informação em saúde. Um conjunto de símbolos, representação, vão carregar com ele um conceito ligado a saúde. O código e a descrição linguagem natural geram, respectivamente, facilidades ao computador e aos usuários. As ferramentas de codificação

de achados, utilizadas em saúde, procuram mapear símbolos de uso computacional para as descrições menos ambíguas possíveis ao entendimento do profissional de saúde. Quanto mais profundo e fácil entendimento gerado ao usuário, melhor a semântica do código.

2.1.2 Classificação Internacional de Doenças - CID

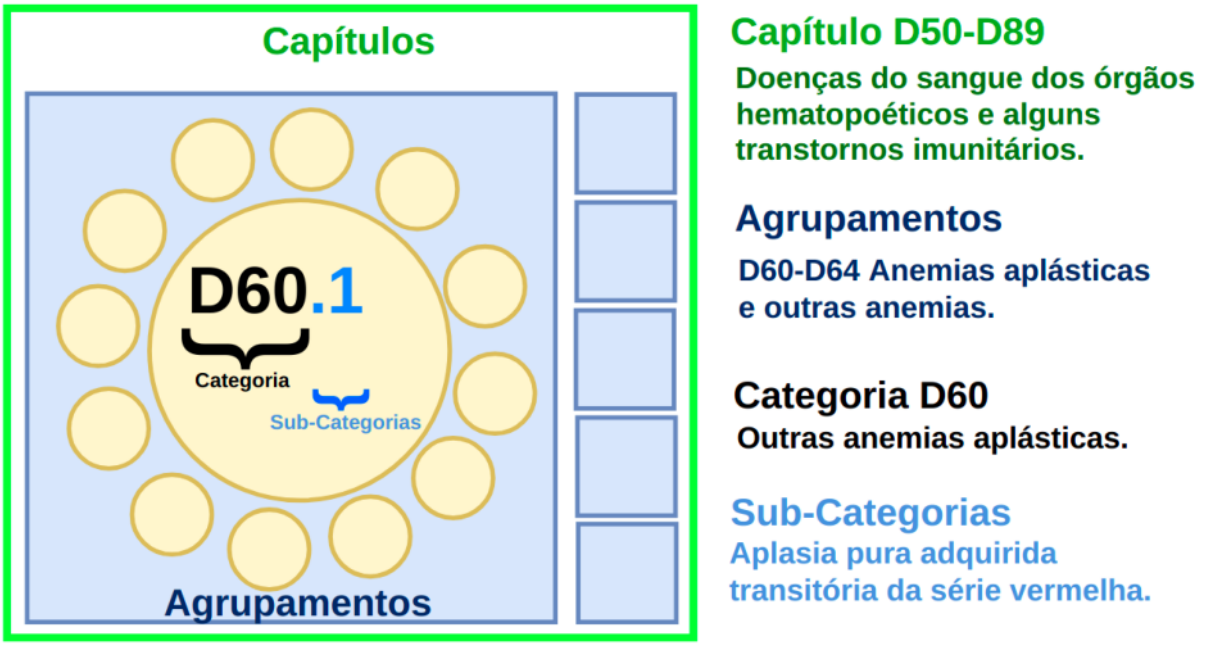
O CID é uma referência internacional, mantida pela Organização Mundial da Saúde, que gera um padrão mundial para a identificação e codificação de doenças e condições de saúde. Permite o armazenamento, o compartilhamento e a recuperação de dados para análises futuras. Diversos dados epidemiológicos são calculados a partir dos registros codificados no CID, gerados nos serviços de saúde (ORGANIZATION, 2021).

Cada CID é composto do nome do elemento (doença, procedimento, ocorrência) e do seu código de identificação. A parte de código usa as letras para denotar funções e estruturas corporais, atividades, e fatores ambientais. Essas letras são seguidas por um numérico indicativo do capítulo (um dígito), pelos agrupamentos, segundo nível, (dois dígitos), categorias, terceiro nível, e subcategorias, quarto nível, (dígitos precedidos do ponto), conforme figura 3. As categorias do CID são "aninhadas"; as categorias mais amplas são definidas para incluir subcategorias mais detalhadas (ORGANIZATION, 2021).

Semanticamente, os capítulos do CID são agrupados em ordem alfabética, conforme os critérios a seguir:

- Doenças infecciosas e parasitárias (A00 – B99).
- Neoplasias [tumores] (C00 – D48).
- Doenças do sangue e dos órgãos hematopoiéticos e alguns transtornos imunitários (D50 – D89).
- Doenças endócrinas (E00 - E90).
- Nutricionais e metabólicas (E00 – E90).
- Transtornos mentais e comportamentais (F00 – F99).
- Doenças do sistema nervoso (G00 – G99).
- Doenças do olho e anexos (H00 – H59).
- Doenças do ouvido e da apófise mastoide (H60 – H95).
- Doenças do aparelho circulatório (I00 – I99).
- Doenças do aparelho respiratório (J00 – J99).
- Doenças do aparelho digestivo (K00 – K93).

Figura 3 – Estrutura do CID em 4 níveis. O capítulo, representado pelo retângulo verde. O capítulo é constituído por agrupamentos, retângulos azuis. O agrupamento é constituído por categorias, círculos amarelos, representadas pelos primeiros 3 dígitos. As categorias também podem ser subdivididas em subcategorias, representadas por dígitos depois do ponto.



Fonte: elaborado pelo autor.

- Doenças da pele e do tecido subcutâneo (L00 – L99).
- Doenças do sistema osteomuscular e do tecido conjuntivo (M00 – M99).
- Doenças do aparelho geniturinário (N00 – N99).
- Gravidez, parto e puerpério (O00 – O99).
- Algumas afecções originadas no período perinatal (P00 – P96).
- Malformações congênitas, deformidades e anomalias cromossômicas (Q00 – Q99).
- Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte (R00 – R99).
- Lesões, envenenamento e algumas outras consequências de causas externas (S00 – T98).
- Causas externas de morbidade e de mortalidade (V01 – Y98).
- Fatores que influenciam o estado de saúde e o contato com os serviços de saúde (Z00 – Z99).
- Códigos para propósitos especiais (U04 – U99).

2.1.3 Protocolos

Cada doença, definida pelo seu respectivo código, gera diversas abordagens. Os padrões em saúde para tratamentos são chamados protocolos e visam definir os tratamentos

vigentes. Uma vez identificada uma patologia, o protocolo determina as ações de prevenção, diagnóstico, cura e reabilitação. As práticas devem ser baseadas em protocolos, pois gera segurança à clínica e aos procedimentos ao validar os passos empregados (DE CASTRO; SHIMAZAKI, 2006). O protocolo é uma linha-guia para uma condição ou doença, porque normatiza e integra o processo dos serviços existentes. Estabelece os fluxos entre as partes da rede de atendimento (WERNECK; DE FARIAS; CAMPOS, 2009).

Da mesma forma, na pesquisa, a gestão da geração dos conhecimentos científicos é realizada por meio de protocolos (GONÇALO, 2007). Desta forma, o pesquisador pode atuar respeitando os direitos e deveres das pessoas e animais envolvidos nos testes em saúde.

Para melhor definir, os protocolos são conjuntos de passos pré-definidos por especialistas, geralmente de Conselhos de Classes, comitês de especialistas, que guiam o andamento da abordagem em saúde. Antes mesmo do diagnóstico de uma patologia, os profissionais, muitas vezes, realizam protocolos de identificação da doença (Ex. protocolo de pedido de exames, protocolo de exame de sangue em emergência, etc.). Os procedimentos de classificação de risco também dependem de protocolos preestabelecidos (PICON; GADELHA; BELTRAME, 2014).

Os protocolos estão em constante evolução. Os profissionais motivados buscam por estratégias metodológicas para adaptação das ações de saúde coletivas conforme as necessidades regionais (ARAUJO; ACIOLI; NETO; DE MELLO *et al.*, 2017).

Os protocolos clínicos geram vários benefícios quando empregados adequadamente em um serviço de saúde. Alguns benefícios estão descritos a seguir:

- Eles orientam o processo de trabalho de toda a equipe de saúde por padronização (WERNECK; DE FARIAS; CAMPOS, 2009). Para o diagnóstico, os protocolos direcionam a entrevista clínica, o exame físico, exames laboratoriais e os procedimentos de avaliação.
- Servem como referência para o ensino de procedimentos e abordagens em saúde. As atividades das faculdades de medicina devem estar em consonância com protocolos de eficácia reconhecida (UNIGRANRIO, 2018). Permitem transmissão dos procedimentos idealizados pelos especialistas da área e de revisões da literatura.
- Permitem a avaliação de condutas e de decisões, já que diversos protocolos podem ser criados para um mesmo objetivo clínico. A comparação de dois protocolos é um rico instrumento de pesquisa médica.
- Para doenças identificadas, o protocolo ajuda o profissional na decisão clínica responsável.
- Consolida as normas, políticas e pesquisas científicas conforme as melhores práticas com

autonomia e responsabilidade (ROSSO; CRUVINEL; SILVA; ALMEIDA *et al.*, 2014). Dão apoio às ações dos profissionais. São a referência ética da relação profissional/paciente nos procedimentos de saúde. Geram segurança aos profissionais que realizam o "melhor esforço" de trabalhar conforme regras amplamente aceitas pelas sociedades de classe. Promovem segurança ao paciente, quando procedimentos destoantes dos protocolos existentes podem indicar erros de abordagem médica.

- Protocolos de comunicação orientam os termos, as abordagens entre médico e paciente (BIASIBETTI; HOFFMANN; RODRIGUES; WEGNER *et al.*, 2019). Também servem de instrumento de comunicação para a equipe de saúde, pois carregam conceitos do domínio de trabalho.
- Servem como referência nas pesquisas médicas, desde que o protocolo seja previamente aprovado para o uso em seres humanos ou animais (MEDICINA, 2018).
- Permitem uma abordagem multiprofissional, pois deve integrar em sua elaboração diferentes profissionais (EBSERH, 2017).

As condutas médicas devem adotar as melhores práticas vigentes. Segundo o Código de Ética Médica brasileiro:

> "A conduta adotada deve ser cientificamente reconhecida, o que por um lado veda a utilização de tratamentos ditos experimentais, e por outro, obriga ao médico a estar permanentemente atualizado, sendo assim capaz de indicar a mais acertada conduta, para o bem-estar e o benefício do seu paciente." (MEDICINA, 2018)

Os defensores dos protocolos destacam a qualidade do atendimento, a redução de variações indesejadas na prática e a ajuda a tornar a prática médica mais científica. Contudo, existem os críticos, sugerindo que os protocolos limitam as respostas possíveis a uma doença, burocratizando e regulamentando excessivamente a prática da saúde (BERG, 1997). Grandes empresas como a deste estudo estão dependentes aos controles de qualidade baseados nos protocolos.

2.1.4 Protocolos Codificados

Os protocolos podem ser também codificados. Nos serviços de saúde deste estudo, ele é codificado de forma similar ao CID, código formado por letras e números e descrição em linguagem natural. Assim, podem ser utilizados como uma ferramenta de simplificação da

codificação clínica. O plano de saúde estudado criou *links* entre alguns CIDs com protocolos clínicos produzidos pelos especialistas da própria empresa.

Desta forma, o processo de trabalho do médico foi modelado nas duas etapas. Na primeira, o diagnóstico clínico gera um código CID. Como a mesma entidade clínica pode ser representada por vários CIDs, os especialistas agruparam os códigos semelhantes. Na segunda, o tratamento deve seguir os protocolos vigentes. A empresa configura automaticamente em seu sistema de fichas clínicas determinados grupos de CIDs para protocolos específicos, orientando os passos dos profissionais.

O protocolo, utilizado desta forma, gera a simplificação da codificação CID. De certa forma, os especialistas médicos escolhem as principais demandas do serviço ou as que precisam de mais orientação para estabelecer um maior controle sobre os atos médicos, no sentido de minimizar os erros oriundos de fatores humanos.

Atualmente, são aproximadamente 100 protocolos modelados com seus respectivos códigos de identificação. Os médicos mais antigos do serviço já os identificam pelo nome e pelo código. Alguns deles sugeriram até a colocação da opção computacional do recebimento do código do protocolo no lugar do código CID em casos clínicos específicos, a Figura 4 mostra a possibilidade de gerar principalmente o código do protocolo. O código CID poderia ser selecionado posteriormente no momento da aplicação do protocolo. Como os tratamentos são mais específicos que a codificação, os CIDs advindos do protocolo aplicado tendem a ser mais acurados.

Figura 4 – Este diagrama é a simplificação do Processo de atendimento médico com a priorização da codificação do protocolo.

Fonte: elaborado pelo autor.

2.1.5 Outras Tentativas de Codificação

Mesmo o CID sendo o mais usado e de iniciativa da OMS, vários outros grupos de códigos coexistem em sistemas de saúde. Isso é justificado devido a outras possibilidades semânticas e incompletudes da codificação básica.

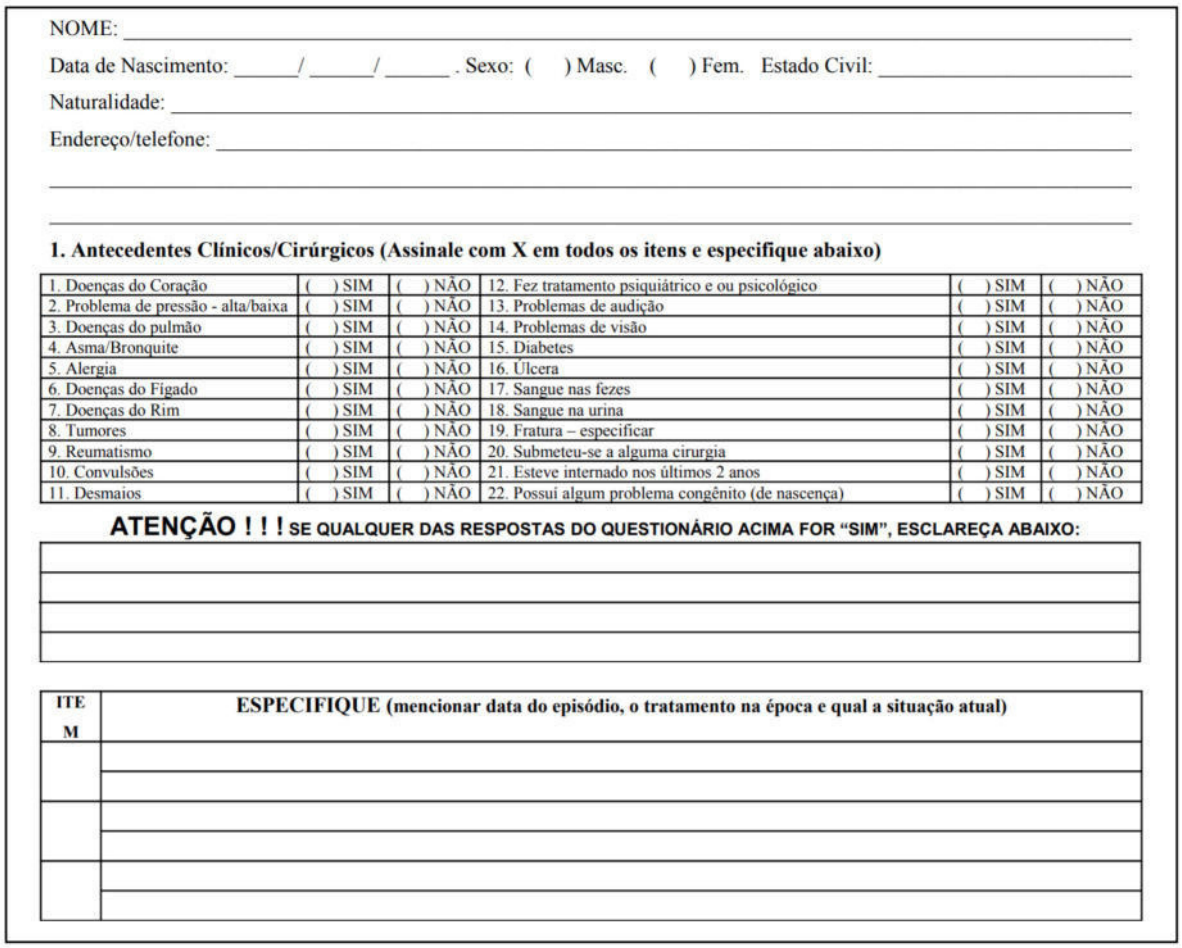
Como exemplo, de codificações auxiliares, mostramos:

- CIAP: muito usada pelo SUS brasileiro, permite a abordagem de problemas de diagnósticos e motivos das consultas e intervenções. Sendo que o CID, no que se refere a morbimortalidade, e o CIAP, contexto social, se complementam (GUSSO, 2009).
- CIF: abordou os estados de saúde dando foco a elementos incapacitantes (BÖLTE; LAWSON; MARSCHIK; GILDLER, 2021).
- A Terminologia Processual Atual da Associação Médica Americana (CPT): analisa os processos envolvidos na saúde;
- Administração de Financiamento de Cuidados de Saúde (HCFA) com o Sistema de Codificação de Procedimentos Comuns de Cuidados de Saúde (HCPCS): é uma codificação obrigatória para a responsabilização e portabilidade dos seguros saúde americanos (SCHAUM, 2011);
- o Manual Diagnóstico e Estatístico de Transtornos Mentais da *American Psychiatric Association* (DSM-IV): específico para transtornos mentais;
- a Classificação Europeia de Operações e Procedimentos Cirúrgicos (OPCS - 4): com enfoque na cirurgia (O'MALLEY; COOK; PRICE; WILDES *et al.*, 2005);
- SNOMED é uma tentativa de gerar terminologias mais compreensíveis que o CID. Possui:
    1. Estrutura em árvore para identificação consistente dos códigos. Códigos gerais na raiz da árvore geram códigos mais específicos a medida que se caminha em direção as folhas.
    2. Define relações que permitem associações baseadas em lógica (iguais, subtipos, local, método utilizado, entre outros elementos que ligam conceitos). Assim, a ferramenta pode agrupar conceitos semelhantes e descrições para o mesmo conceito.
    3. Permite a combinação de códigos para o detalhamento dos achados. Grande abrangência clínica para evitar outros sistemas de codificação (ORGANIZATION, 2020; SCHULZ; KLEIN, 2008).

2.2 Prontuários Eletrônicos e Seus Registros

Os profissionais de saúde, por lei, devem preencher a ficha clínica para cada consulta. Eles escrevem em prontuários e registram as evoluções para gerar históricos do paciente, conforme Figura 5. Tais históricos são muito importantes para a contextualização de casos clínicos e tomadas de decisões com segurança (LOPES; BRASIL, 2003).

Figura 5 – Modelo de prontuário impresso para realização de atendimentos médicos.



NOME: _____

Data de Nascimento: _____/_____/_____. Sexo: ( ) Masc. ( ) Fem. Estado Civil: _____

Naturalidade: _____

Endereço/telefone: _____

_____

_____

1. Antecedentes Clínicos/Cirúrgicos (Assinale com X em todos os itens e especifique abaixo)

1. Doenças do Coração	( ) SIM	( ) NÃO	12. Fez tratamento psiquiátrico e ou psicológico	( ) SIM	( ) NÃO
2. Problema de pressão - alta/baixa	( ) SIM	( ) NÃO	13. Problemas de audição	( ) SIM	( ) NÃO
3. Doenças do pulmão	( ) SIM	( ) NÃO	14. Problemas de visão	( ) SIM	( ) NÃO
4. Asma/Bronquite	( ) SIM	( ) NÃO	15. Diabetes	( ) SIM	( ) NÃO
5. Alergia	( ) SIM	( ) NÃO	16. Úlcera	( ) SIM	( ) NÃO
6. Doenças do Fígado	( ) SIM	( ) NÃO	17. Sangue nas fezes	( ) SIM	( ) NÃO
7. Doenças do Rim	( ) SIM	( ) NÃO	18. Sangue na urina	( ) SIM	( ) NÃO
8. Tumores	( ) SIM	( ) NÃO	19. Fratura – especificar	( ) SIM	( ) NÃO
9. Reumatismo	( ) SIM	( ) NÃO	20. Submeteu-se a alguma cirurgia	( ) SIM	( ) NÃO
10. Convulsões	( ) SIM	( ) NÃO	21. Esteve internado nos últimos 2 anos	( ) SIM	( ) NÃO
11. Desmaios	( ) SIM	( ) NÃO	22. Possui algum problema congênito (de nascença)	( ) SIM	( ) NÃO

ATENÇÃO ! ! ! SE QUALQUER DAS RESPOSTAS DO QUESTIONÁRIO ACIMA FOR "SIM", ESCLAREÇA ABAIXO:

ITEM	ESPECIFIQUE (mencionar data do episódio, o tratamento na época e qual a situação atual)

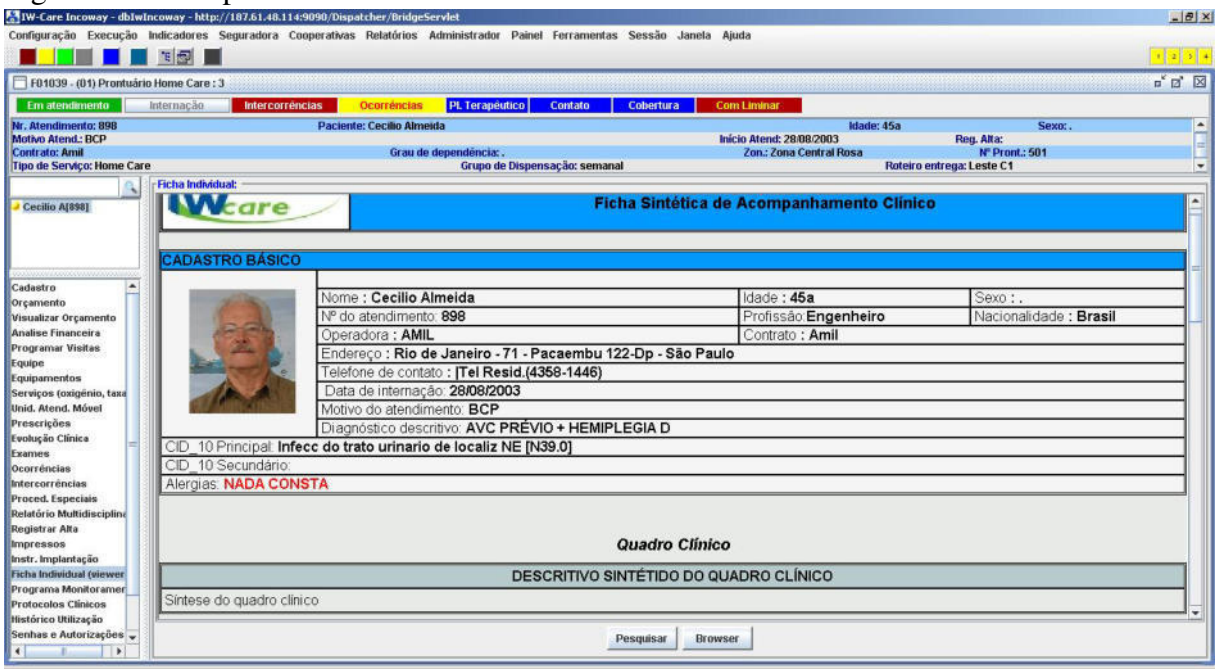
Fonte: elaborado pelo autor.

Nas fichas de papel existem grandes campos de tabelas onde o médico preenche a caneta os achados clínicos, procedimentos, evoluções, observações, exames, entre outras coisas. No exemplo da Figura 5 as tabelas de preenchimento com o título "Atenção" e "Especifique" recebem até as datas da consulta. Comumente, os dados são agrupados sem critérios pré-estabelecidos. As fichas eletrônicas organizam melhor o *input* dos dados e seu armazenamento. Na Figura 6, a empresa IwCare disponibiliza a seus associados o diversos softwares de prateleira que servem para o gerenciamento de ambientes clínicos e hospitalares.

Os *softwares*, sistemas de prontuários eletrônicos, *Eletronic Health Records*, prontuários eletrônicos em saúde, (EHR), a exemplo da Figura 6 para atendimentos médicos melhoraram a organização, gravação e acesso aos dados da ficha médica. Consequentemente, acarretando melhores decisões médicas e administrativas. O custo e a qualidade dos serviços são impactados diretamente (HANNAN, 1996).

Na Figura 7, tais *softwares* reproduzem os grandes campos para o preenchimento do médico. São colocadas grandes caixas de texto para o livre preenchimento do profissional em

Figura 6 – Exemplo de tela de um Sistema de Prontuários Eletrônicos.

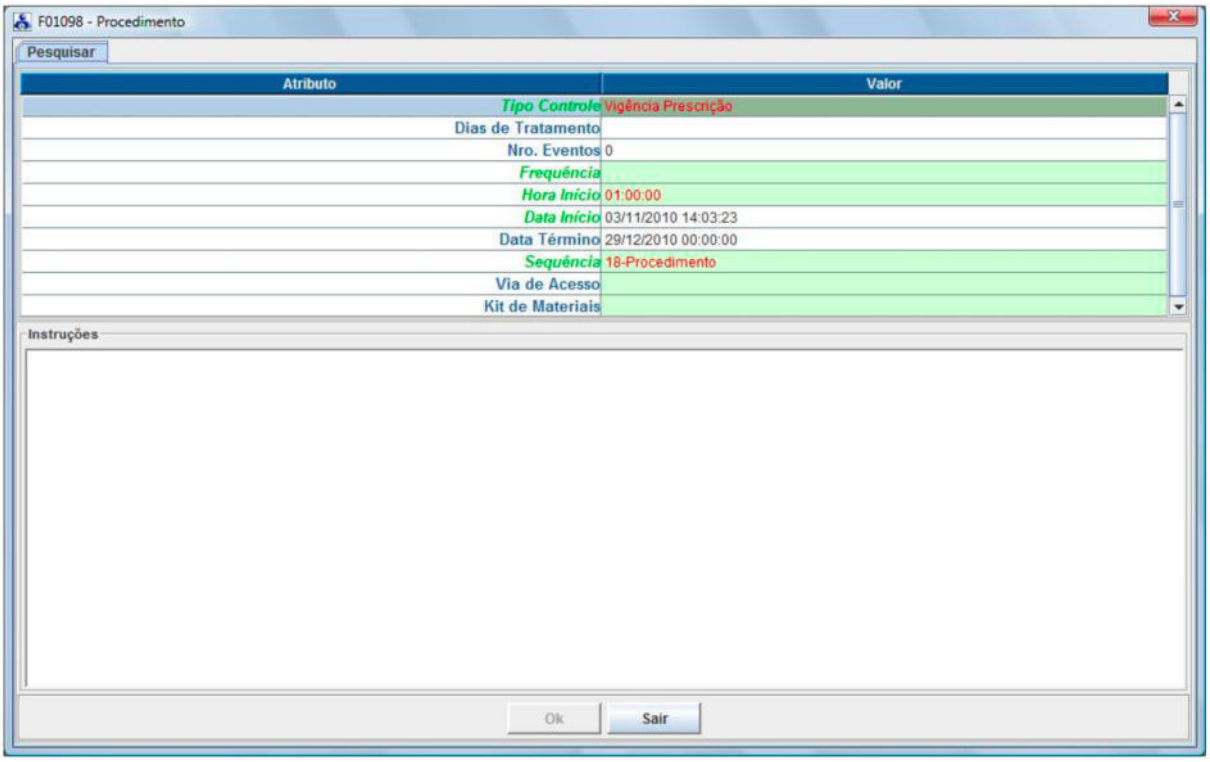


Fonte: (IWSOFTWARE,2009).

linguagem natural nos terminais de atendimento. Nestes, os dados podem ser bem desestruturados e com abreviações de uso corrente da área de atuação do profissional.

Para exemplificar este campo do EHR, um texto real de um médico do serviço em estudo é apresentado a seguir em sua formatação original, Figura 8.

Figura 7 – O campo "Instruções" é uma grande caixa de texto que não valida os dados de entrada.



Fonte: (IWSOFTWARE,2009).

Figura 8 – Exemplo do documento JSON obtido, unidade de dado do estudo.


```
{
"service_code": 12345678,
"pacient_name": "XXXXXXXXXXXX",
"birth_date": "1974-03-20 00:00:00",
"address_name": "Rua XXXXXXXXX",
"service_date": "2019-04-3000: 00: 00",
"hr_service": "62124",
"protocol_code": "H027",
"description_protocol": "Otalgia",
"main_complaint": "O PACIENTE APARECE COM UM RESULTADO DE
MASTOIDE CT QUE EVIDÊNCIA DE MASTOIDE NORMONUCLEAR, UMA GRANDE
QUANTIDADE DE MATERIAL COM DENSIDADE LÍQUIDA ESPESSA QUE VILA
A MAIORIA DAS CÉLULAS CELULARES DE MASTÓIDE ESQUERDA, PODE
REPRESENTAR UMA GRANDE QUANTIDADE DE MATERIAL COM ESPESSA
DENSIDADE LÍQUIDA QUE VILA A MAIORIA DAS CÉLULAS CELULARES DE
MASTÓIDE ESQUERDA, PODE REPRESENTAR UMA PROCURA DE MATERIAL
INFLAMATÓRIO. OTYRRINO. ELA QUER REFERÊNCIA PORQUE O PLANO SÓ
ATENDE A URGÊNCIA ",
"alergias": "DIPIRONA",
"medicine_in_use": "AMOXICILLIN + CLAVULANATE",
"cid10": "H60",
"heart_frequency": 68.0,
"breath_frequency": 14.0,
"blood_systolic_pressure": 110.0,
"blood_diastolic_pressure": 70.0,
"aspecto_geral": "Paciente relata secreção serosa em orelha
esquerda há aproximadamente 8 meses, associada a hipoaquase,
relata que já foi atendida várias vezes em TNE e foi
encaminhado para \" tímpano soprado \ "e solicitada TC
eletiva, referindo a um episódio de febre há 1 semana. O
paciente chega ao pronto-socorro para uma TC urgente. \ n EF:
secreção purulenta cobrindo o ouvido externo, MT não vista \
nCD: transferir caso para a cabeça ",
"temperatura": 36,5,
"saturação de oxigênio": 98,0 }
```


Fonte: elaborado pelo autor.

2.3 Valor Semântico da Codificação

A semântica é o estudo do significado, uma disciplina da linguística, uma análise da formalidade das palavras (MARQUES, 1990). Os CIDs, possuem descrição e código, o que também implica em um significados relacionados a saúde.

O profissional de saúde atribui códigos no seu processo de trabalho, como já relatado. Mas, códigos corretamente atribuídos são necessariamente ótimos códigos? A resposta é não. Um código muito genérico, denominado aqui como "supercódigo", é aquele adequadamente aplicado ao contexto, mas sem especificar bem o achado. O supercódigo engloba muitas doenças, achados, ou outros elementos da clínica e promovem pouca informação específica. Não descrevem o problema ou o fazem de maneira geral. Um exemplo dele é o J00, código da nasofaringite aguda. Pode ser gripe, resfriado, dor de garganta, irritação causada por vírus, bactéria, por contato da mucosa por substâncias irritantes, etc. Todas essas opções tem seus CIDs específicos que deixaram de ser usados. Por isso, o supercódigo tem pouco valor semântico, mesmo correto.

Devido a facilidade de realizar a tarefa obrigatória de definição do CID com o supercódigo, eles podem ser encontrados em grande quantidade nas amostras. Se não houver uma preocupação para a qualidade do registro, eles serão frequentes. Resolvem rapidamente o problema do campo de preenchimento, inadequadamente.

2.4 Tipos de Funcionalidades Existentes para Sugestão de Códigos

Ao estudar os diversos *softwares* existentes para utilizá-los como base para futuras funcionalidades, criou-se uma classificação apenas didática para nível de complexidade computacional. Os *softwares* de ajuda a preenchimento de códigos em saúde podem ser divididos nos seguintes grupos:

- Baseado em Regex,
- Contabilização de dados,
- *Software* especialista,
- Automação de passos,
- Escolhas assistidas.

2.4.1 Baseados em Regex Campo Estruturado

Regex, ou expressões regulares, são padrões de cadeias de caracteres que podem ser identificados (SIDHU; PRASANNA, 2001).

No momento do preenchimento do campo "achado em saúde" a expressão regular procura no banco de dados do código com semelhanças na *string*. Encontrado alguma semelhança, ele dá ao profissional a opção de escolha e autopreenchimento.

Isso facilita a busca por códigos, pois funciona, muitas vezes, incorporando uma ferramenta de busca de códigos no campo de preenchimento. Como exemplo, o Sistema de Atendimento Médico (SAM), em sua parte 1 de formulário, disponibiliza uma pequena lupa que leva o profissional a este tipo de pesquisa em uma base de dados de CID.

Esta abordagem, além de muito simples, evita que os profissionais tenham trabalho em fazer pesquisa em navegadores, aplicativos médicos de celulares e *tablets*, abrir aplicativos, buscar por campos de preenchimento, copiar códigos e colar em espaço adequado.

2.4.2 Avaliação de Grandes Strings de Campos Desestruturados por Regex

Existe também a possibilidade do uso de expressões regulares nos campos de Registros desestruturados do tipo texto. Algumas fichas eletrônicas disponibilizam grandes campos para que os profissionais registrem todo o atendimento incluindo dados de pressão, temperatura, entre outras variáveis e achados. Estes campos reproduzem as fichas de papel e suas muitas linhas onde o médico apenas preenchia o que via desordenadamente.

A dificuldade de gerar estatísticas posteriores a partir dessas grandes *strings* aparecem devido aos diversos contextos possíveis. A linguagem natural aqui empregada é de difícil processamento computacional. Como exemplos, duas frases com a palavra diabetes possuem sentidos antagônicos: "sem histórico de diabetes na família" e "diabete insulinodependente desde 2020".

O uso de termos, sinônimos e abreviados em saúde também dificultam o entendimento para leigos. Estas, muitas vezes, não são bem documentadas. As palavras e abreviações podem estar inseridas no contexto local ou individual dos profissionais. Quem fez sabe o que escreveu, entretanto, mesmo outros médicos, podem não entender.

Como opção para tal, o projeto de prontuários eletrônicos pode criar campos específicos para o carregamento de cada variável por vez. Exemplo, um campo temperatura independente recebe apenas números predefinidos dentro do domínio, com validação. O banco de dados, construído com um valor previamente validado pelo formulário, ajuda a manter a consistência dos registros.

2.4.3 Contabilização de Dados

A contabilização consiste em produzir o somatório dos códigos gerados pelo serviço. Os códigos mais frequentes por si já são importante mecanismo de gerenciamento em saúde. Uma vez estabelecida a moda para a amostragem dos dados, várias conclusões epidemiológicas simples podem ser geradas em uma série temporal. Demandas locais e doenças sazonais, serão identificadas. Desta forma, os gestores podem priorizar atividades preventivas e o melhor emprego de recursos.

Os sistemas poderia sugerir alguns códigos baseados na incidência. Os mais frequentes seriam destacados por listas ou mesmo botões de clique rápido para o preenchimento do campo. Esta simples funcionalidade recebeu grande aceitação dos médicos e desenvolvedores da Empresa que mantém o Software de atendimentos clínicos dos pontos de atendimento dos serviços de saúde abordados (INTMED) entrevistados. A epidemia de *chikungunya* descrita em boletim epidemiológico (CEARÁ, 2018) poderia se beneficiar do destaque do CID e do protocolo referido. O sistema, no mês de março de 2017, identificaria o aumento de casos da doença e passaria a sugerir códigos da doença em questão. A partir do mês de agosto, do mesmo ano, devido a imunidade de rebanho, os casos caem e a sugestão é desabilitada automaticamente.

2.4.4 Softwares Especialistas Baseado em Estatísticas

Um sistema especialista consiste em um *software* que simula o julgamento humano conhecedor do assunto. Comumente, uma base de dados gera regras para situações particulares. Tais regras, quando implementadas, produzem respostas semelhantes a uma contribuição de um perito no assunto tratado (TECHTARGET, 2020).

O uso de um sistema especialista para proposição de códigos se baseia na captura de dados históricos do serviço de saúde. Tais dados, após devido processamento, geram informações para a construção da funcionalidade.

Esta abordagem foi aplicada exemplificada neste trabalho, pois as ricas bases de dados dos atendimentos produziram classificações e conclusões para a implementação de uma funcionalidade.

2.4.5 Algoritmo de Decisão

Algoritmo de Decisão é uma estrutura de dados não-linear onde a predição é realizada por regras de decisão simples. Os dados podem ser divididos por meios estatísticos (SOUZA, 2018). Os algoritmos utilizados aqui foram construídos a partir do conhecimento do domínio médico, *ad hoc*.

O domínio dos códigos médicos já é cheio de classificações de diversas formas (classificação por especialidade, local da incidência, etiologia da patologia, fatores de risco, idade, raça, etc.). Entender e selecionar aquelas que melhor distribuem os dados permitem árvores mais balanceadas e consequentemente algoritmos mais eficientes, com complexidades menores na busca de dados alvo.

Esses algoritmos podem ser compreendidos por médicos, diferentemente dos baseados em redes neurais, pois as regras de decisão são classificadores lógicos dos dados. Os sistemas especialistas devem ser validados por humanos para verificar que os conhecimentos extraídos estão de acordo com os domínios de conhecimento (MONARD; BARANAUSKAS, 2003). Isso foi realizado por médicos que utilizam largamente códigos em seus serviços. A análise estatística dos dados foi usada para a produção de árvore de decisão e deu origem à funcionalidade em estudo.

2.4.6 Software Especialista Baseado em Inteligência Artificial

Com o estabelecimento da inteligência artificial, muitos algoritmos conseguiram realizar diagnósticos de forma similar ao médico (WOYTE; SARR; DE BRADANDERE; RICHTER *et al.*, 2018; YEASMIN, 2019). Dentre os muitos exemplos disponíveis selecionamos:

- A ferramenta *Babylon Triage and Diagnostic System* realizou diagnósticos com uma precisão comparável aos médicos humanos. O sistema baseado em inteligência artificial mostrou-se mais seguro do que os médicos humanos, na média dos casos analisados. A partir da análise dos sinais e dos sintomas, o computador dava o diagnóstico e sugeria os procedimentos clínicos pertinentes. Estes *softwares* vêm ajudando os médicos a realizar diagnósticos complexos, prevenir erros de conduta, classificar mais adequadamente as patologias e melhorar a interpretação dos dados (RAZZAKI; BAKER; PEROV; MIDDLETON *et al.*, 2018).
- os sistemas de diagnósticos em lâminas histopatológicas conseguiram melhores resultados do que humanos realizando mesma tarefa (PELACCIA; FORESTIER; WEMMERT, 2019).
- Fosun, empresa chinesa de tecnologia em saúde, conseguiu estudar os fatores geradores de doenças. Um modelo de tomada de decisão médica resultou nas cabines de diagnóstico e salas de simulação digital. Promete resolver deficiências de diagnósticos médicos, evitar trabalhos repetitivos, melhor a qualidade de serviços e diminuir a necessidade por recursos humanos (FOSUN, 2018).

Mesmo com tantos exemplos, o desafio computacional ainda é grande. Para que aconteça o emprego de tais tecnologias, existe a necessidade de novos estudos. A Geração de Diagnóstico Automatizado está em franco desenvolvimento, contudo, no mundo real, existem ainda problemas de validação dos resultados obtidos e rigor científico na análise dos dados (PARK; KRESSEL, 2018).

2.5 Análise Gráfica de Distribuições Estatísticas

A Inferência Estatística permite obter informações sobre uma população baseada nos resultados de uma amostra (LOPES, 2003). Neste trabalho, tivemos acesso a um grande volume de dados referentes a codificação das patologias. Foram analisados conforme protocolos de pesquisa para a geração de inferências.

As distribuições podem definir curvas e linhas em um gráfico. Histogramas, polígo-

nos de frequência, ogivas, gráficos em segmentos de reta vertical são exemplos de simplificação e organização dos dados que permitem conclusões futuras (LOPES, 2003).

Um gráfico pode ser comparado a outros modelos preexistentes, gerando confirmações estatisticamente observáveis. Esses gráficos de distribuição são confrontados com distribuições conhecidas na literatura (distribuições ideais) e podem gerar conhecimento. Usamos o teste Kolmogorov Smirnov (MASSEY JR, 1951) para confirmar as distribuições de cada um dos Bancos de Dados. Os gráficos produzidos a partir do agrupamento de registros CIDs mostram tendências quando transformados pelas funções logit e o logaritmo Neperiano. O *software* de estatística aplicado é usado para esses testes (WESSA, 2021).

2.6 Transformações dos Dados e Justificativa de Uso

Ronald Fisher propôs alguns princípios para os experimentos científicos: justificou o porquê do uso de dados aleatórios, da replicação e do controle de heterogeneidade das condições experimentais. Sugeriu a busca por estimativas ortogonais e eficientes e o ajuste em funções para gerar estimativas ortogonais mais precisas. Essas transformações nos experimentos podem gerar simplificação nos futuros cálculos. Isto é, se uma modificação não gera ortogonalidade, os pesquisadores podem usar variações do experimento na escala original ou na escala logarítmica. Estatisticamente, o conhecimento pregresso sobre os experimentos permite escolher os melhores valores das variáveis experimentais. Valores ótimos das variáveis experimentais podem gerar resultados melhores em relação tempo, custo e precisão. Quanto maior o experimento, mais difícil é o controle das condições experimentais. As ferramentas auxiliares são de grande valor, pois possibilitam a concretização destes procedimentos. Baseadas nos princípios da estimação de parâmetros e fazendo uso da matemática associada à computação, modelos mais rápidos e eficientes avaliam melhor as propriedades de interesse. A metodologia se aplica a qualquer área de pesquisa experimental que procura ajustar modelos estatísticos (FERREIRA; TRINCA; FERREIRA, 2014).

Os pesquisadores costumam realizar transformações arbitrárias de resultados para cumprir a suposição de normalidade (SCHMIDT; FINAN, 2018).

Há sempre uma razão objetiva para uma transformação matemática. A pergunta a fazer é: como ou por que a distribuição amostral está se deformando, fugindo à normalidade ou da linearidade. Contudo, na prática, todas as transformações são facilmente realizadas por *softwares*. O pesquisador pode tentar e avaliar o resultado obtido, já que a transformação mais indicada

geralmente coincide com aquela que apresentar a probabilidade mais elevada de produzir uma distribuição conhecida. Se a transformação não for aplicável tende a dificultar a visualização de padrões (CAMPOS, 2002).

2.7 Transformação Logarítmica

A transformação logarítmica mantém a hierarquia dos dados originais. Se um dado é maior que o outro, a ordenação é mantida. Os dados transformados em logaritmos não terão a mesma média aritmética dos valores de origem.

A transformação logarítmica dos dados é utilizada na estatística paramétrica. Nela, a distribuição dos erros, a homogeneidade das variâncias, e os efeitos dos fatores de variação são observados para gerar previsões (CAMPOS, 2002). Essa transformação permite que sejam feitas comparações entre distribuições muito diferentes. A transformação logarítmica pode reduzir a tendência de *outliers* (YAN; ZHANG; HUANG; SUN *et al.*, 2015).

A transformação logarítmica redimensiona as observações reais do experimento. A suposição geral nas estatísticas é a variabilidade de alguma resposta homogênea "**p**" entre a variável preditora **i** e esta é mantida (CURRAN-EVERETT, 2018).

2.8 Função Logit e Log-Log

O modelo logit procura diminuir as limitações do modelo linear. Esse possui uma função, logística, que gera uma transformação sobre os dados baseada no logaritmo natural. Como resultado, os dados passam a ter imagens dentro dos limites da probabilidade, entre 0 e 1. Intervalos que tendem ao "-∞" (menos infinito) atingem valor 0. Já intervalos que tendem ao "+∞" (mais infinito) atingem valor 1 (WOOLDRIDGE, 2006). A Figura 9 mostra matematicamente a representação das funções de ligação mais comuns. Este estudo utiliza as funções logit e log-log. Estas funções podem ser representadas graficamente pela Figura 10.

2.9 Processo de Apego Preferencial

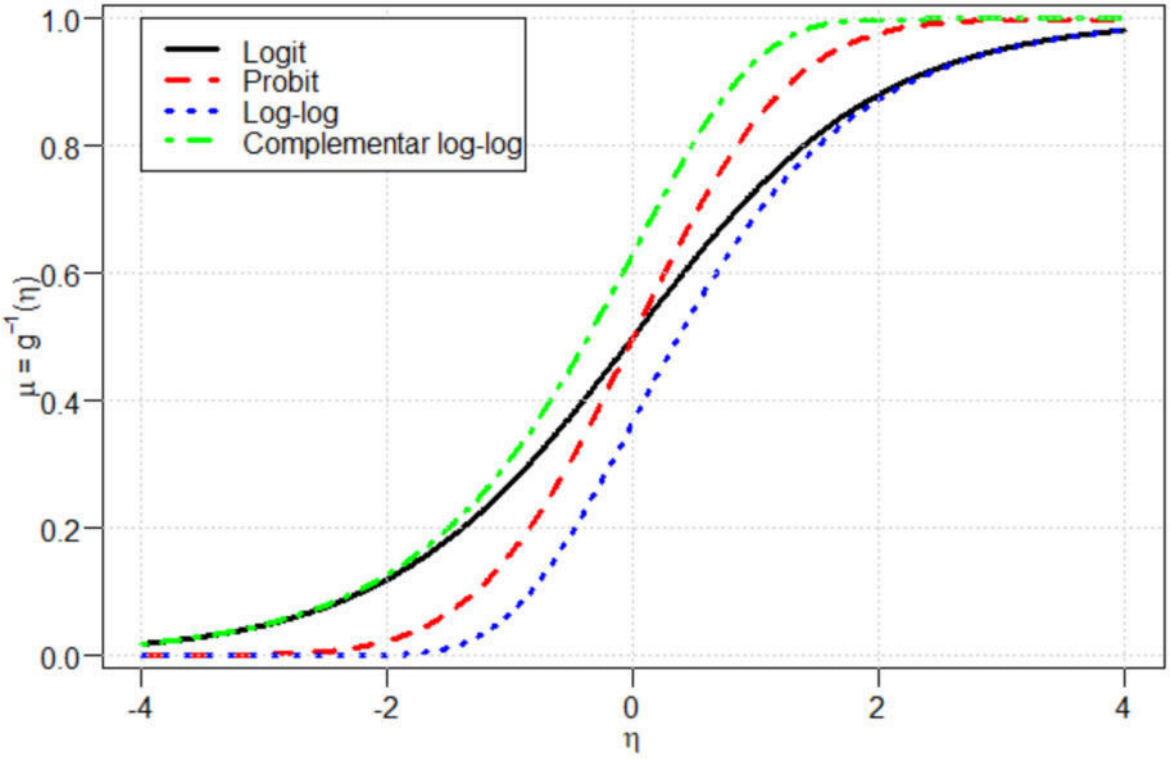
Também chamados de "ricos mais ricos" e "vantagem cumulativa", pois geram distribuições que seguem a lei da potência (JIANG; SUN; FIGUEIREDO; RIBEIRO *et al.*, 2015). Muitas vezes, o processo de apego preferencial é denominado de modelo livre de escala (CANCHO; SOLÉ, 2001). A distribuição de potência modela o apego preferencial. Gera uma

Figura 9 – Representação de funções de ligação. A segunda coluna mostra as fórmulas das funções e a terceira coluna mostra suas inversas.

Ligação	$\eta_i = g(\mu_i)$	$\mu_i = g^{-1}(\eta_i)$
Identidade	μ_i	η_i
Log	$\log_e(\mu_i)$	e^{η_i}
Inversa	μ_i^{-1}	η_i^{-1}
Inversa quadrada	μ_i^{-2}	$\eta_i^{-1/2}$
Raiz quadrada	$\sqrt{\mu_i}$	η_i^2
Logit	$\log_e\left(\frac{\mu_i}{1-\mu_i}\right)$	$\frac{1}{1+\exp(-\eta_i)}$
Progit	$\Phi^{-1}(\mu_i)$	$\Phi(\eta_i)$
Log-log	$-\log_e(-\log_e(\mu_i))$	$\exp(-\exp(-\eta_i))$
Complementar log-log	$\log_e(-\log_e(1-\mu_i))$	$1-\exp(-\exp(\eta_i))$

Fonte: Peres (2021).

Figura 10 – Funções de ligação logit, probit, log-log e complementar log-log.



Fonte: Peres (2021).

distribuição de cauda longa, semelhante a distribuição de Pareto (SIMON, 1955).

Há muitas palavras parecidas em uma língua, pois é comum a formação de umas a partir de outras (CANCHO; SOLÉ, 2001). O mecanismo de criação de códigos CIDs são

parecidos com o da criação de palavras. Existem códigos vinculados ou mesmo derivados. Por isso, possivelmente aconteça a distribuição de potência referentes aos códigos existentes.

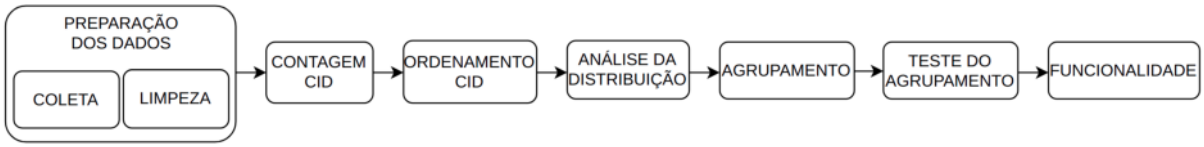
Neste estudo, aplicamos essa teoria para observar o processo de trabalho do médico codificando o CID. Os médicos podem estar beneficiando os códigos mais comuns em sua memória e os estruturalmente mais próximos a estes.

2.10 Justificativa por Modelos Lineares Mais Simples

Modelos que seguem a lei da potência são complexos. Esta distribuição pode ser representada graficamente por uma reta. Por ser extremamente simples, gera uma fácil observação das relações e das tendências entre variáveis (YAN; ZHANG; HUANG; SUN *et al.*, 2015). Os modelos lineares são os mais utilizados na pesquisa (HOPKINS, 2010).

3 METODOLOGIA

Figura 11 – Processo Geral.



Fonte: elaborado pelo autor.

A metodologia deste trabalho é mostrada na Figura 11. As fases do estudo são: **preparação do conjunto de dados** (coleta de dados do plano de saúde e remoção de CIDs nulos ou administrativos), **contagem CID** (identificação do código e contagem igual), **classificação CID** (ordenação dos códigos , daqueles com maior incidência para aqueles com menor incidência), **análise de distribuição** (criação de gráficos e identificação de semelhanças entre distribuições conhecidas), **grupo** (geração de grupos de códigos de acordo com o domínio da saúde, produção de algoritmo), **teste de grupo** (produção de *software* que implementa algoritmo e testes dele por médicos) e **funcionalidade** (proposta de uma nova funcionalidade nas fichas clínicas).

3.1 Preparação Conjunto de Dados

Os EHR são heterogêneos. Vem de 256 locais de atendimentos, incluindo 40 hospitais em 5 estados brasileiros (Alagoas, Amazonas, Bahia, Ceará, Goiás). O contexto da coleta é o do pronto atendimento e emergência, principalmente.

3.2 Coleta

O *"service_code"* é uma chave que identifica o EHR e se refere a um atendimento. Através de um programa *python* e com a lista de chaves do período, realizamos a consulta ao banco de dados. Desta forma, os atributos preenchidos de cada EHR foram reagrupados. O resultado de cada chave consultada gerava um arquivo JSON, como na estrutura anonimizada da Figura 8. Esta é a unidade de dados para este estudo.

Obtemos duas séries temporais que formam os registros:

– 1º de maio a outubro de 2019, 6 meses, 179.842 unidades de dados, doravante identificadas como **Base de dados 1**. Este banco foi obtido diretamente por meio de uma *query SQL*, reconstruindo perfeitamente cada EHR. Possui menos registros, porque o sistema

eletrônico de fichas clínicas estava iniciando sua implantação.

– Dezembro de 2019 até o final de novembro de 2020, com um ano, 2.441.229 formulários, denominado **Base de dados 2**. Os dados não possuem muitos atributos da EHR. Como não tínhamos acesso direto aos dados, a *query* utilizada era preestabelecida por código que intermediava o acesso aos dados.

Portanto, 2.621.071 unidades de dados foram utilizadas neste trabalho. Esta amostra é representativa para planos de saúde, pois além de ser grande, é retirada de ambientes assistenciais heterogêneos em 5 estados brasileiros.

Entre os atributos de cada ficha clínica estão: sexo, idade, queixas principais, códigos CID, códigos de protocolo clínico, temperatura, pressão, condições neurológicas, etc., conforme mostrado na Figura 8. Existem atributos de preenchimento obrigatórios e opcionais. O código CID é um atributo obrigatório, pois, sem ele, não se pode passar para a próxima fase do atendimento, o acesso aos protocolos médicos.

3.3 Limpeza

Para que um EHR seja incluído no estudo, é necessário:

– possua código CID válido, proveniente do banco de dados dos serviços prestados pelo plano de saúde, representando diagnósticos, achados clínicos e procedimentos;
– não represente códigos administrativos e
– não represente solicitações de exames.

Dados inconsistentes, ruídos, foram removidos. Eles estavam fora dos domínios das variáveis estudadas e representavam valores nulos. Atributos do tipo *string* 'sim' ou 's' e 'não' ou 'n' foram homogeneizados, mas substituídos por *False* e *True*, respectivamente.

3.4 Contagem dos CIDs

Os CIDs encontrados, aqueles utilizados nos bancos de dados, foram agrupados. Após cada código contabilizado por meio de um *scanner python* nos dois conjuntos de dados, obtivemos a lista de CIDs utilizados pelos serviços com as suas respectivas quantidades. Chamamos "p" a variável que recebe as quantidades dos CIDs.

Os códigos são ordenados em ordem decrescente em relação à sua incidência. Depois disso, eles são substituídos por números de acordo com suas posições (*rank*). O código com o

maior número de incidência é substituído pelo número 1, o segundo pelo número 2 e assim por diante. Esta abordagem torna mais fácil construir gráficos e visualizar distribuições.

Paralelamente a contagem dos CIDs, foi feita a contagem dos códigos dos protocolos clínicos para comparações posteriores.

3.5 Análise da Distribuição

O banco de dados 2 é priorizado porque tem muito mais registros do que o banco de dados 1. A análise do banco de dados 1 é apresentada por sua fórmula e gráfico de tendências.

As quantidades das variáveis em estudo são discretas e quantitativas. Permite a construção de gráficos de linhas. Primeiro, construímos o gráfico geral da distribuição, sem transformações, o rank i na abcissa e p na ordenada das bases de dados.

No segundo momento, realizamos a primeira transformação dos dados de acordo com a Eq. 1 na base de dados 2, mais representativa.

Por fim, cada probabilidade incondicional das distribuições do Banco de Dados 1 e Banco de Dados 2 foi transformada de acordo com a Eq. 2, assumindo que a probabilidade de CID segue uma função *logit* mista.

A primeira transformação é representada pela fórmula: **yi = -*logit*pi, onde X=i**, denominada equação 1. A segunda transformação possui a fórmula: **yi = ln (-*logit*pi), onde X=i**, identificada como equação 2.

Essa transformação permite que sejam feitas comparações entre probabilidades muito diferentes. Assumindo que a distribuição da quantidade é enviesada com uma extremidade superior e uma cauda longa, a transformação logarítmica reduz a tendência de outliers (YAN; ZHANG; HUANG; SUN *et al.*, 2015).

Esses gráficos de distribuição são comparados com distribuições conhecidas na literatura (distribuições ideais) e podem gerar conhecimento. Usamos o teste Kolmogorov Smirnov (MASSEY JR, 1951) para confirmar as distribuições de cada um dos Bancos de Dados. Os gráficos produzidos a partir do agrupamento de registros CID mostram tendências quando transformados pelas funções logit e o logaritmo neperiano. Um *software* de estatística aplicado é usado para esses testes (WESSA, 2021).

3.6 Agrupamento

O próprio código CID já possui uma semântica pré-estabelecida por meio de letras e números, e indica os capítulos de codificação (ALHARBI; ISOUARD; TOLCHARD, 2021).

Agrupamos os códigos mais comuns nas bases de dados de acordo com o conhecimento do domínio médico (classificações baseadas na especialidade e achados clínicos). O agrupamento produz conjuntos de CIDs. Os grupos formam posteriormente algoritmos *adhoc*. O algoritmo pode organizar buscas de códigos nos serviços, pois reduz o espaço de amostragem para escolha do médico.

Os códigos protocolos também foram agrupados seguindo a mesma metodologia. Da mesma forma geraram algoritmos para comparações.

3.7 Teste do Agrupamento

Novos EHR são obtidos aleatoriamente de amostras diferentes daquelas selecionadas nos Bancos de Dados. Dois médicos, que não participaram da construção do algoritmo, foram convidados a testá-lo.

O processo usado para testar o algoritmo anterior segue estas etapas:

1. Seleção aleatória de EHR que não foram usados nas etapas anteriores.
2. O médico lê os dados históricos e entende a consulta apresentada.
3. O médico escolhe uma das opções dos grupos possíveis (exemplo: o médico escolhe o grupo das vias aéreas).
4. O programa exibe automaticamente os códigos CID mais usados do grupo selecionado.
5. O sistema soma um a variável acerto, quando o CID do EHR está no grupo escolhido pelo médico voluntário. Caso contrário, o programa soma um algarismo a variável erro.
6. O médico voluntário lê as opções de CIDs do grupo escolhido e avalia se existe código válido entre os códigos apresentados.
7. Cada médico repetiu este procedimento 100 vezes, em intervalos de aproximadamente 45 minutos para todo o teste.

3.8 Proposta de Funcionalidade

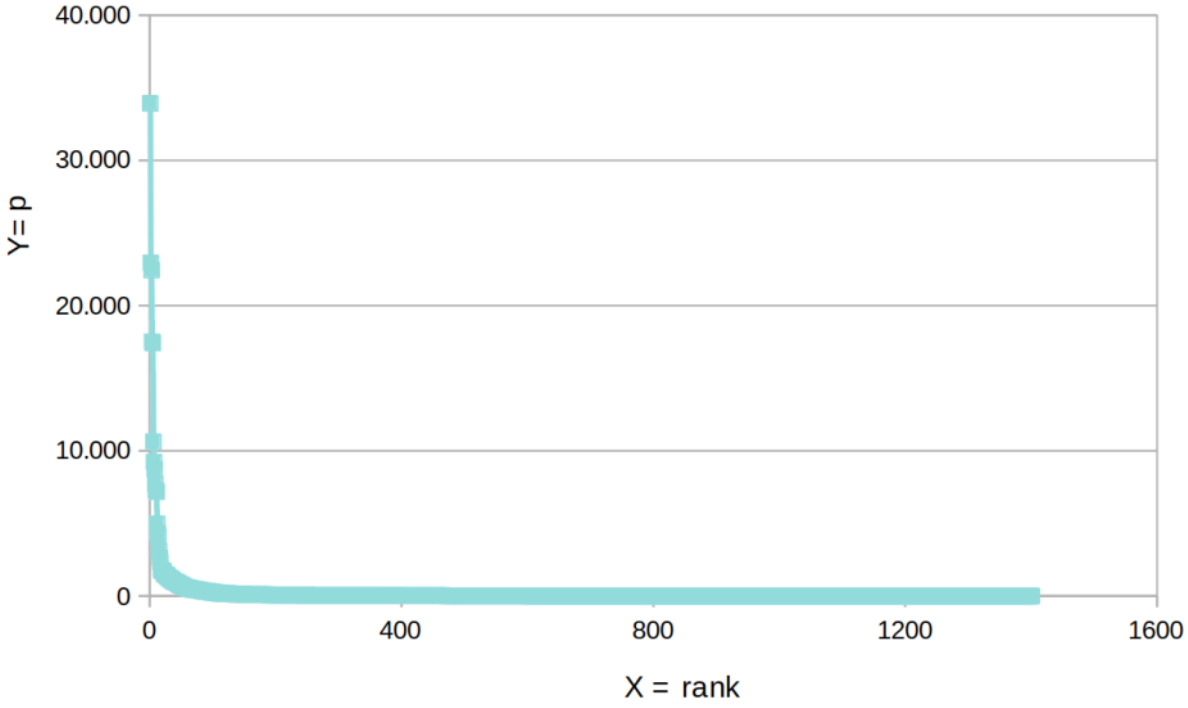
Ao final, produzimos uma modificação na interface do *software* de prontuários médicos da empresa de planos de saúde para futuros testes *on-line* em uma unidade de atendimento.

4 RESULTADOS

O CID-10 possui quase 70.000 códigos (HORSKY; DRUCKER; RAMELSON, 2017) No Banco de Dados 1, 1.401 códigos CIDs diferentes foram encontrados e, no Banco de Dados 2, 2.468 códigos.

Em uma análise superficial, os gráficos de distribuição de classificação em relação ao número de bancos de dados são muito semelhantes. A Figura 12, Banco 1, e a Figura 14, Banco 2, mostra algumas tendências quanto ao uso de códigos:

Figura 12 – O gráfico representa a distribuição do Conjunto de Dados 1. X = rank e Y = p.



Fonte: elaborado pelo autor.

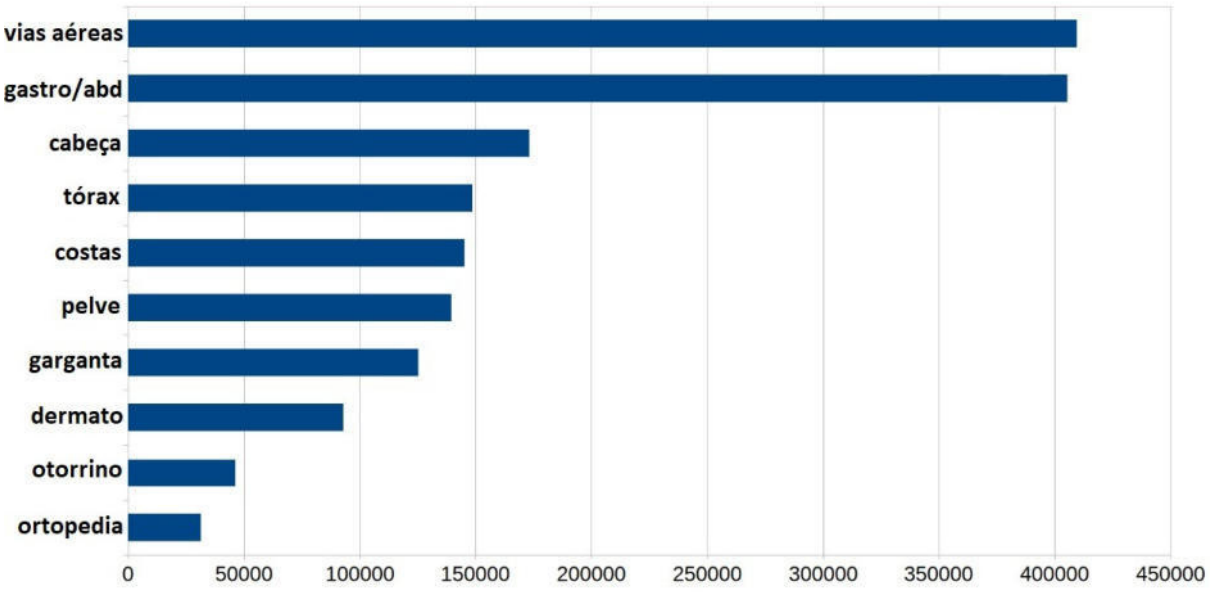
1. os códigos mais usados representam quase a totalidade dos dois bancos.
2. a distribuição geral pode ser dividida em duas.
3. podemos agrupar os códigos mais incidentes em grupos do domínio médico.

Dada a importância dos códigos mais usados, iremos direcionar os esforços em avaliar os 24 códigos mais usados no banco 2, mais representativo. Também, iremos identificar os órgãos e sistemas do corpo humano que mais geram atendimento e impactam os serviços prestados pela empresa.

4.1 Principais Demandas dos Serviços

A avaliação da localização corporal dos códigos CIDs gerou a Figura 13 que mostra as áreas corporais que mais geraram códigos do tipo CID no banco de treino. Este gráfico é importante para o dimensionamento do serviço e identificação da demanda.

Figura 13 – Gráfico das regiões corporais/especialidades, com mais frequência, geram códigos CIDs.



Fonte: elaborado pelo autor.

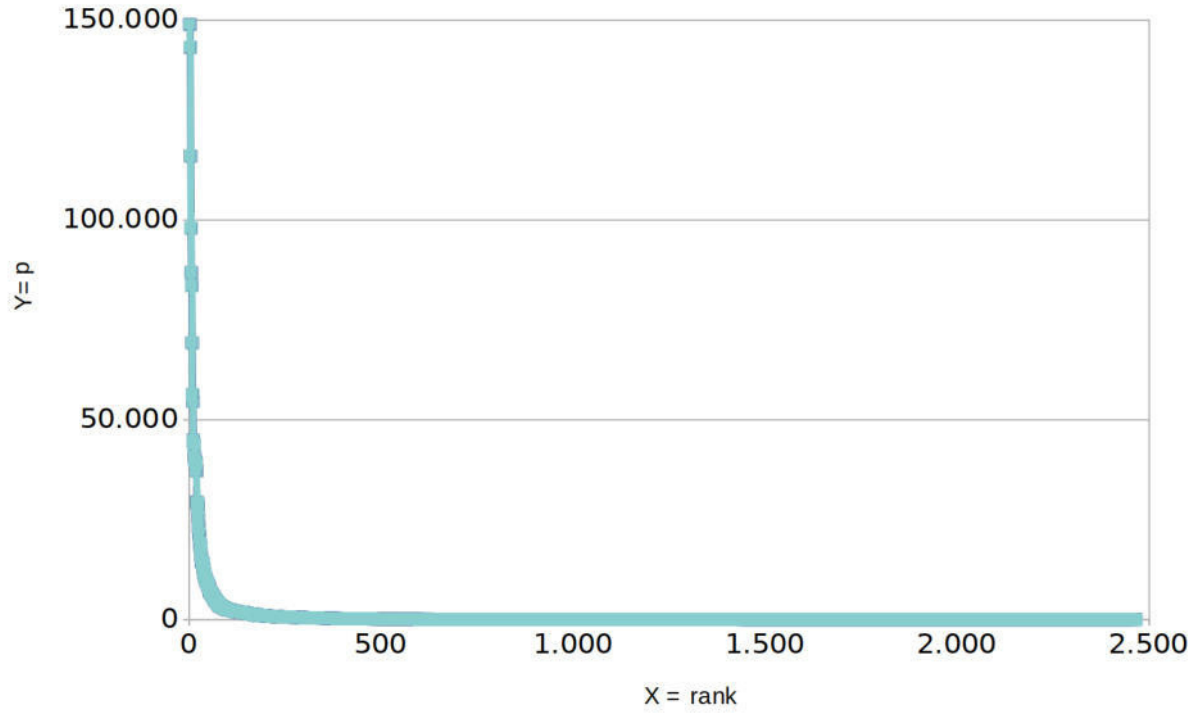
O serviço recebe muitas afecções respiratórias, muito comuns no pronto atendimento adulto e infantil: nasofaringite, rinites, amigdalite, pneumonia comunitária, tosse, asma, etc. As afecções intestinais também geram muitos atendimentos: diarreia, náuseas, vômitos, gastroenterite e colite, infecção intestinal, etc.

4.2 Códigos Mais Usados

4.2.1 Análise dos 24 mais usados

A Figura 14, referente ao Banco de Dados 2 e com gráfico semelhante no Banco de Dados 1, permite uma análise mais precisa e menos sujeita a variações. Possui mais dados e mais tempo de coleta. O código mais usado no Banco de Dados 1 é J00 (Nasofaringite aguda). Ele sozinho representa 13,14 % dos CIDs usados, seguido pelo R50, com 9,09 % (Febre de origem desconhecida) e por A09, com 6,73 % (Diarreia e gastroenterite de origem infecciosa). No Banco de Dados 2, a maioria dos incidentes são A09, com 7,55 %, J00, com 7,25 % e J069,

Figura 14 – O gráfico representa a distribuição do Conjunto de Dados 2. X = rank e Y = p.



Fonte: elaborado pelo autor.

com 6,97 % (infecção aguda não especificada das vias aéreas superiores).

Quando usamos apenas os 24 CIDs mais frequentes, obtemos no Banco de Dados 1 uma representação de 77,6 %, Tabela 1, e no Banco de Dados 2 de 64,48 %, Tabela 2. Isso mostra como poucos códigos representam grande parte do preenchimento dos médicos.

Grupo	**24 CIDs mais frequentes no DB 1**										**Total**
Vias Aéreas	J00 (13,14%)	R05 (8,71%)	J03 (6,74%)	J02 (3,37%)	J06 (2,95%)	J069 (2,76%)	J01 (1,95%)	J039 (1,01%)	J45 (0,90%)	K30 (0,65%)	42,18%
Clínica	R50 (9,09%)	R10 (4,08%)	M545(2,80)	H920 (0,71%)	I10 (0,58%)						17,26%
Gastroenterologia	A09 (6,73%)	R11 (3,56%)	M529 (1,66%)	A08 (0,68%)	K549 (0,65%)						13,28%
Neurologia	R51(1,67)	G43 (1,56%)									3,23%
Urologia	N30 (0,58%)	N300 (0,56%)	N390 (0,51%)								1,65%
										Estes códigos representam 77,60 % de todo banco	77,60%

Tabela 1 – Tabela Banco de Dados 1.

Os principais CIDs encontrados são códigos genéricos que não especificam bem as patologias. Achados de igual valor médico legal também podem ser considerados como: J00 (nasofaringe aguda), J02 (faringite aguda), J069 (infecção aguda não especificada das vias aéreas superiores) e J398 (outras doenças específicas das vias aéreas superiores).

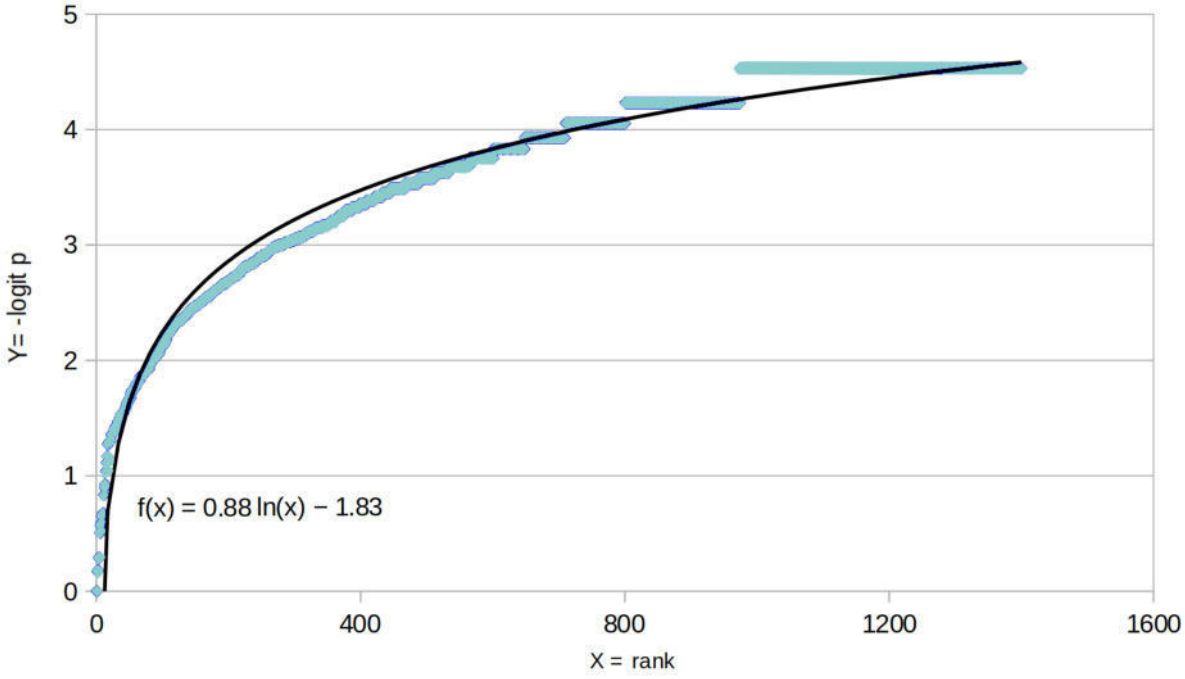
Grupos	**24 mais frequentes no DB 2**									**Total**
Vias Aéreas	J00 (7,25%)	J069 (6,97%)	J03 (2,74%)	R05 (2,18%)	J398 (1,94%)	J02 (1,93%)	J06 (1,43%)	J01 (1,05%)	K30 (0,93%)	26,42%
Dor	R10 (5,65%)	M545 (4,77%)	M549(1,98%)	R074(1,41%)						13,81%
Clínica	R50 (1,81%)	I10 (1,12%)	H920 (0,89%)							3,82%
Gastroenterologia	A09 (7,55%)	R11 (2,18%)	K549 (0,71%)							10,44%
Neurologia	R51 (4,23%)	G43 (1,89%)	R42 (1,01%)							7,13%
Urologia	N390 (1,37%)	N300 (0,78%)	N30 (0,71%)							2,86%
									Estes códigos representam 64,48 % de todo o banco	64,48%

Tabela 2 – Tabela Banco de Dados 2.

4.3 A Distribuição Pode Ser Dividida

A transformação dos dados do Banco 1, Figura 15 não permite observar as duas distribuições. O número maior de registros do Banco 2 gerou um gráfico com quantidades maiores de CIDs distintos pouco incidentes. Assim, houve o melhor preenchimentos do gráfico de tendência, Figura 16.

Figura 15 – Tendência no banco de dados 1. Y = logit p.



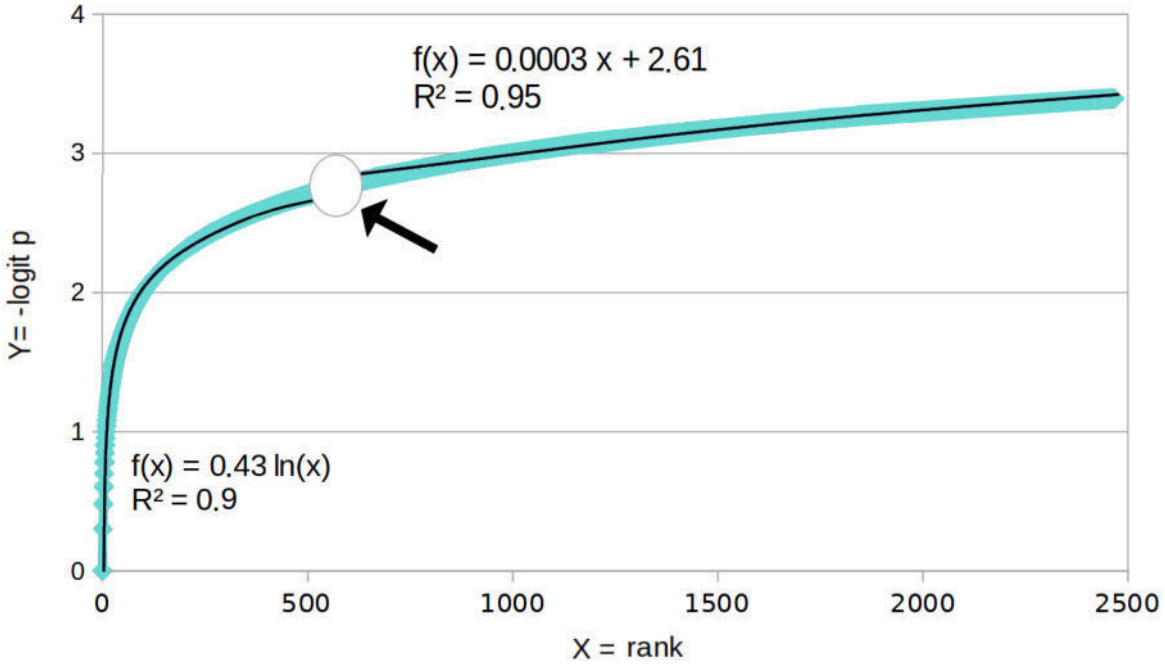
Fonte: elaborado pelo autor.

Com a grande quantidade de dados no Banco 2 foi possível identificar duas distribuições, pois houve o aumento do número dos códigos, bem como a suavização dos valores decrescidos.

O estudo de Savaglio foi crucial para avaliar estas distribuições de dados. No trabalho, é proposto a coexistência de dois fenômenos críticos diferentes representados graficamente por duas distribuições (SAVAGLIO; CARBONE, 2000). Situação semelhante foi encontrada na distribuição estatística dos códigos, a existência de tendências distintas no mesmo gráfico, 16.

Vimos na transformação da Eq. 1 Figura 16 a possibilidade de dividi-lo na posição marcada pelo círculo. Isso permitiu a representação da distribuição por duas fórmulas. Na parte esquerda do círculo, observamos uma distribuição sem escala e, à direita, uma distribuição linear.

Figura 16 – Duas tendências no banco de dados 2. Y = logit p.

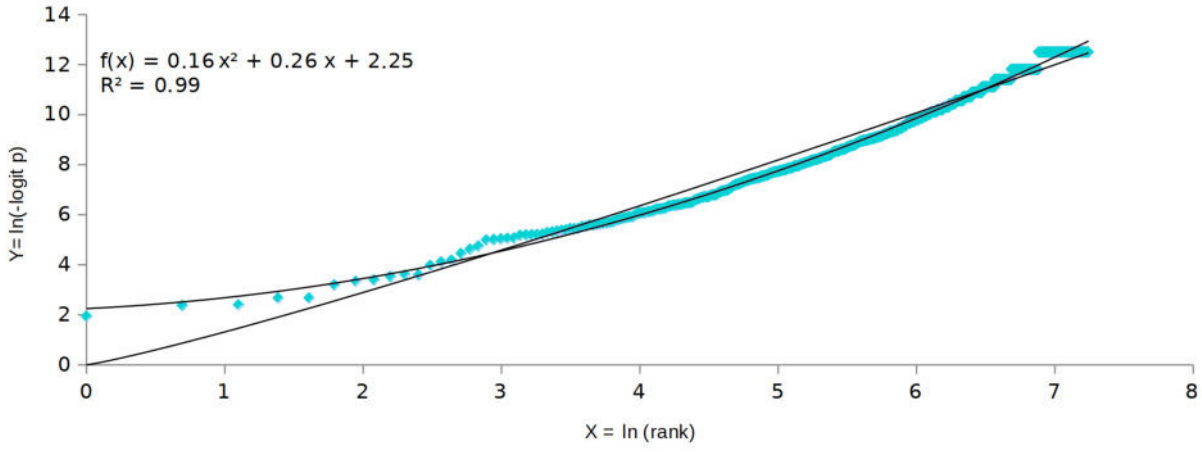


Fonte: elaborado pelo autor.

4.4 Análise das Tendências Globais

A transformação de probabilidade incondicional pela Eq. 2 gera gráficos que mostram as tendências globais dos bancos de dados. No Banco de Dados 1, identificamos um polinômio de distribuição de segunda ordem com tendência à linearidade. Esta equação é válida para quase todos os ICDs, exceto os seis mais prevalentes. A Figura 17 mostra que esta transformação gera um polinômio de segunda ordem usando o logaritmo dos valores de classificação nas abscissas, confirmadas pelo teste Teste de Kolmogorov Smirnov (A = 0,17, B = 0,27, C = 2,25 e $R^2 = 0,99, p < 0,00001$).

Figura 17 – Distribuição do Banco 1 de acordo com Eq. 2.



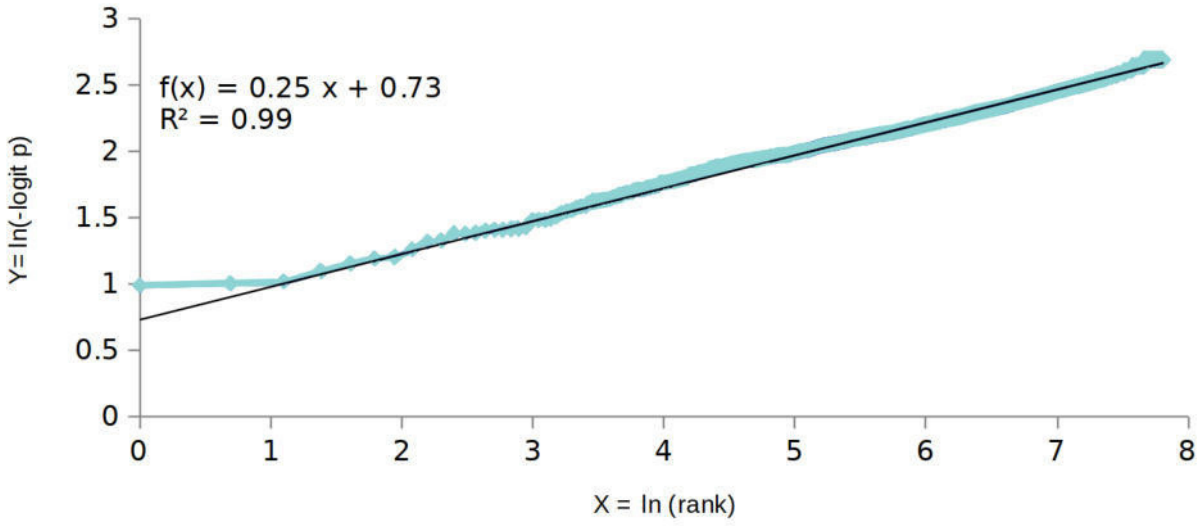
Fonte: elaborado pelo autor.

	Resultado 1	Resultado 2
O código do PES está presente no grupo de CIDs escolhido em:	57,31 %	58,00
O médico achou a sugestão dos CIDs válida em:	61,10 %	72,00

Tabela 3 – Resultado dos testes realizados com dois voluntários médicos especialistas.

Da mesma forma, as tendências globais da transformação de probabilidade incondicional do Banco de Dados 2 pela Eq. 2 são válidas para quase todos os CIDs, exceto os dois mais prevalentes e alguns outros casos. A Figura 18, validada com o Teste de Kolmogorov Smirnov, mostra que esta transformação gera uma linha reta usando o logaritmo dos valores de classificação nas abscissas (A = 0,248 e B = 0,7321, $R^2 = 0,9911, p < 0,00001$).

Figura 18 – Distribuição do Banco 2 de acordo com a Eq. 2.



Fonte: elaborado pelo autor.

4.5 Análise do Especialista

Avaliamos os dados e identificamos a possibilidade de endereçar a maioria dos códigos nos bancos de dados, conforme mostrado na Tabela 1 e 2. As classificações de código usaram agrupamentos de conhecimento geral do domínio médico. São eles: dor (achado clínico), clínica geral (cotidiano de médicos não especialistas), vias aéreas (problemas respiratórios), gastroenterologia (especialidade médica), neurologia (especialidade médica) e urologia (especialidade médica). Os resultados do teste do algoritmo mostram a viabilidade de agrupamento dos códigos, conforme Tabela 3.

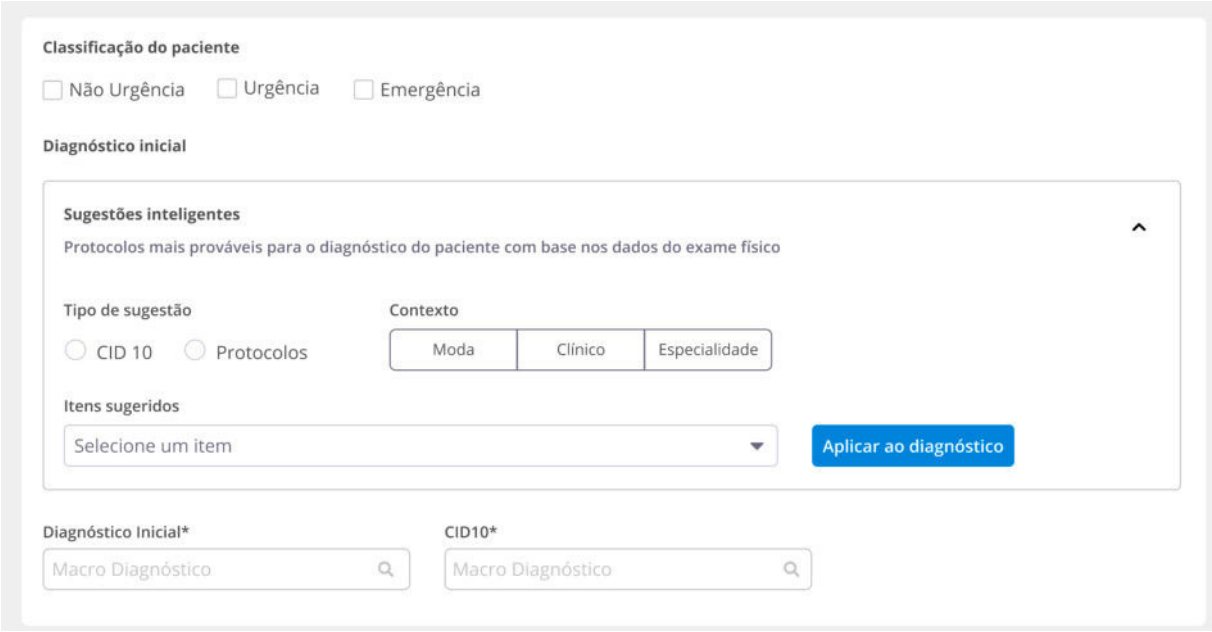
4.6 Campos de Linguagem Natural Concentradores

Foram encontrados muitos relatos médicos nas variáveis *"main_complain"*, queixa principal, e *"genera_aspect"*, aspecto geral, queixa principal e aspecto geral. Como exemplo, selecionamos um exemplo de preenchimento, com na Figura 8. Os dados deste estudo ainda são incompletos e desestruturados para a realização desse tipo de mapeamento. É preciso criar uma biblioteca semântica especializada para a realidade da empresa geradora dos códigos em análise. Os formulários, se completos, poderiam gerar estudos similares dos obtidos pela empresa Fosun. Algumas variáveis dos atendimentos clínicos costumam estar dispersas no campo "queixa principal" que é muitas vezes usado como ficha clínica.

4.7 Construção de Nova Funcionalidade

A partir dos resultados obtidos, propomos uma nova funcionalidade para o *software* de gerenciamento de fichas. Conforme Figura 19, esta poderia ser inserida como um mecanismo de ajuda a codificação médica.

Figura 19 – Funcionalidade sugestão inteligente.



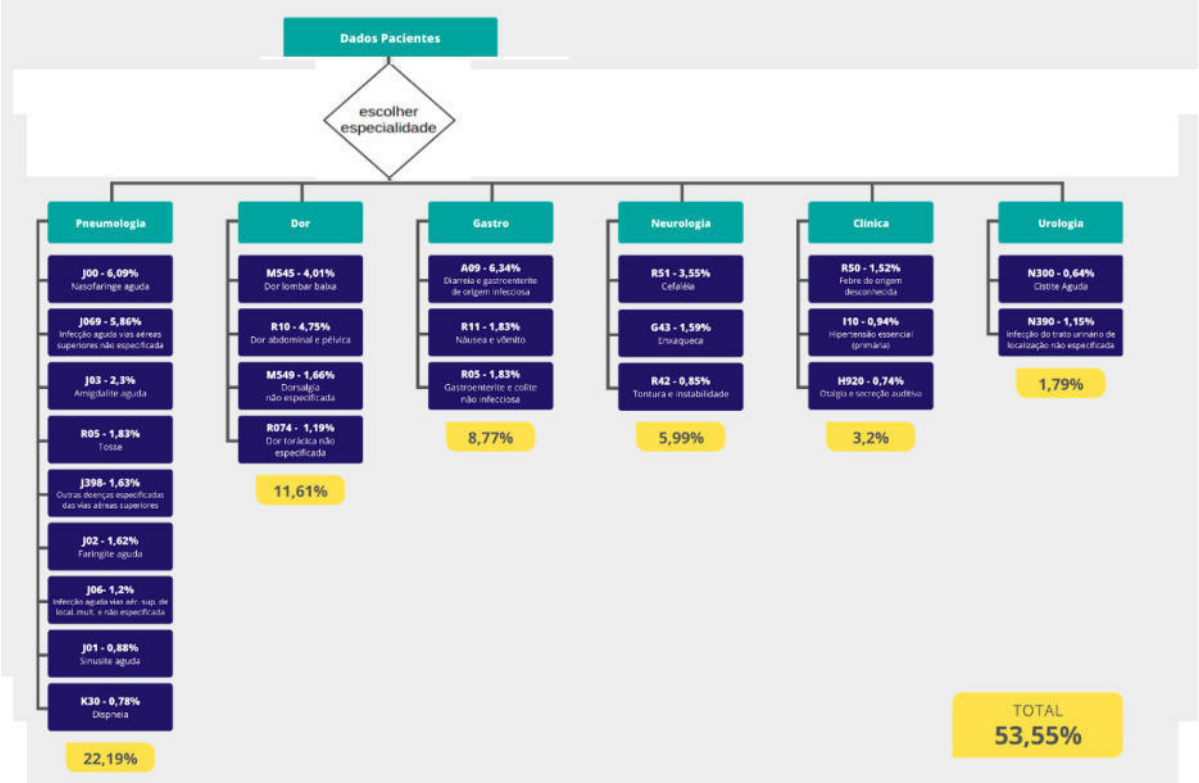
Fonte: elaborado pelo autor.

Algoritmos foram modeladas como ilustrado na Figura 20. Estas mostram o possível acesso rápido a folhas a partir de pequenas decisões médicas. Elas ratificam a importância da escolha e classificação de bons códigos para a construção de novas funcionalidades computacionais.

Na Figura 21, referente ao algoritmo para os protocolos, destacamos a importância da simplificação dos códigos para a melhoria do acesso. Cerca de 100 códigos permitem o funcionamento e a orientação do processo de trabalho médico da grande maioria dos atendimentos do serviço.

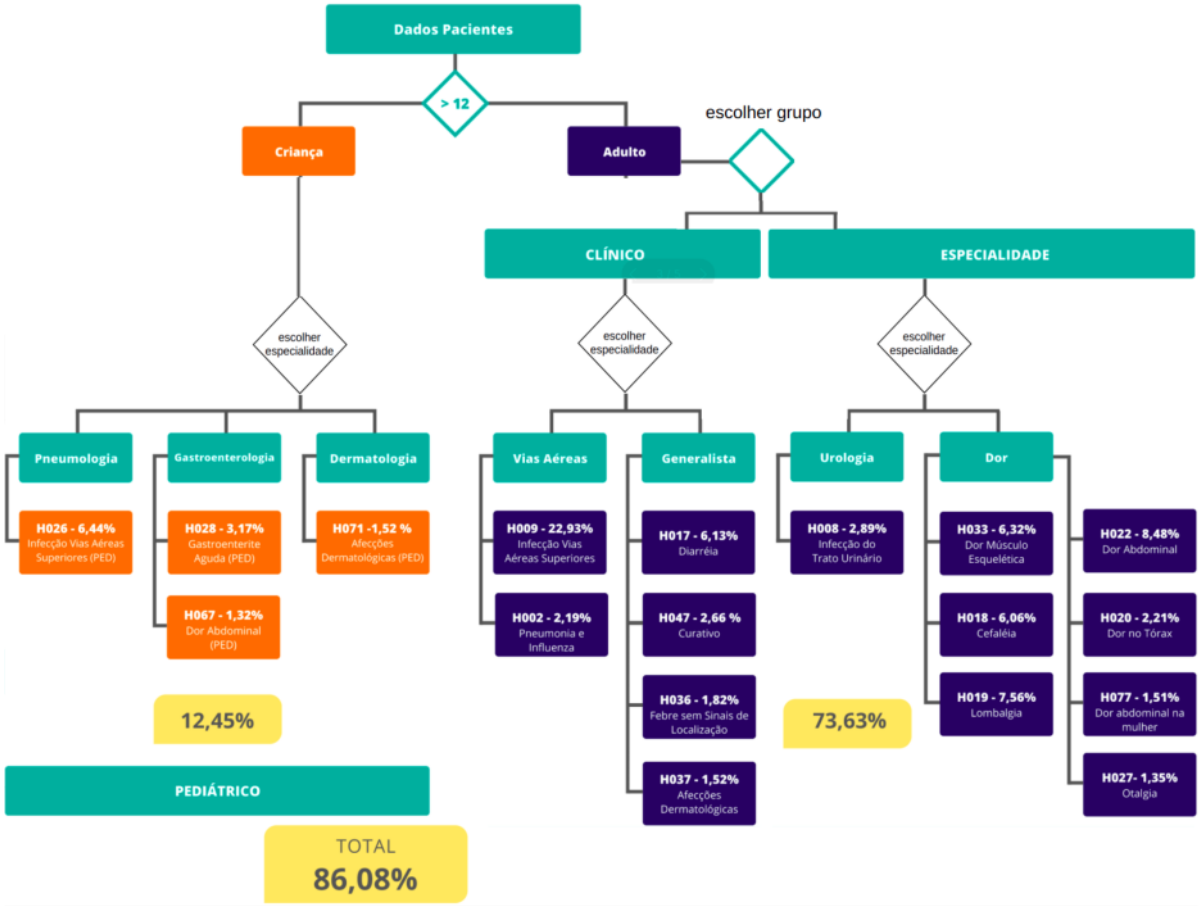
A participação dos códigos mais usados é bastante grande e representam significantemente o banco de dados avaliado, 24 códigos com 53,55%, para o CID, e 17 códigos, 90,37%, para os códigos dos protocolos.

Figura 20 – Algoritmo CID com um passo, escolha por especialidade, permite o acesso dos 24 códigos mais usados. Este modelo representa 53,55% de todo o banco.



Fonte: elaborado pelo autor.

Figura 21 – Algoritmo Protocolo com um passo automático, se idade>12, e um passo do médico, escolha por especialidade, permite o acesso dos 17 códigos mais usados. Este modelo representa 90,37% de todo o banco.



Fonte: elaborado pelo autor.

5 CONCLUSÕES E TRABALHOS FUTUROS

A captura de informações de saúde por este registro é exaustiva e complexa. Prontuários de papel comumente possuem a data do atendimento e, logo depois, todas as observações e dados da consulta. Seguindo o mesmo mapa mental, muitos médicos preenchem os campos para *strings* grandes como a ficha de papel. Lá diversas variáveis de grande valor na geração do banco são anexadas inadequadamente. A pressa e o costume do preenchimento em campo único podem levar a concentração informações e na dificuldade de acesso à variáveis como temperatura, pressão, número de batimentos, etc. Os formulários eletrônicos possuem campos projetados para o recebimento e validação dos valores, contudo nem sempre recebem a devida atenção.

Muito estudo ainda precisa ser realizado para melhorar os sistemas de codificação. O uso de super códigos (códigos muito gerais para muitas doenças), prática que foi confirmada por este estudo, deve ser evitado. Só um protocolo, o H026, infecções de vias aéreas respondeu por 31,05% do banco. Essas duas codificações envolvem todas as doenças infecciosas da garganta, do seio maxilar e do nariz como um todo, a exemplo da covid-19 e podem ter o mesmo valor semântico. Isso foi identificado nos códigos CIDs, A09 - diarreia e gastroenterite de origem infecciosa, representava 6,34% de todo o banco. Na verdade, o CID está sendo subutilizado com os supercódigos. Dos 55.000 usa-se aproximadamente 2500 em um ano. E poucos CIDs respondem por quase todo o banco.

No quadro atual do serviço, o sistema aceleraria a escolha do CID e diminuiria a cognição na realização da tarefa pelo médico. No entanto, da forma como estão sendo escolhidos os códigos poucos dados podem ser realmente aferidos.

A ferramenta se mostrou promissora, mas qual o nível de automatização o *software* deve implementar? O médico pode se valer negativamente da ferramenta? Quais os questionamentos éticos para esse tipo de ferramenta?

A funcionalidade pode evoluir para não aceitar códigos muito gerais. Pode propor novas opções mais ricas no registro das informações. Os resultados obtidos neste trabalho sugerem a construção de uma ferramenta que não só identifique códigos, mas os coloquem dentro de contextos para a melhora semântica. Ouvir os usuários finais e coletar requisitos também ajudaria a definir grupos de códigos mais precisos e mais importantes. Em cada serviço, existem os códigos mais específicos que podem classificar melhor os achados mais recorrentes. Para tal, é preciso a capacitação e disposição da equipe que iria definir os códigos de maior valor

semântico.

Hoje, já é consenso a importância de avaliar o histórico clínico do paciente (ASIF; MOHIUDDIN; HASAN; PAULY, 2017). Existem apresentações semelhantes para doenças diferentes, contudo, o histórico de visitas anteriores, dará muito mais pistas para o fechamento do diagnóstico correto. Com a evolução da Inteligência Artificial aplicada aos códigos clínicos, a codificação com assistência poderá ajudar os profissionais a identificar códigos mais específicos.

As codificações evoluem com o tempo. O CID, neste momento, está na versão 11. Estudos como este demonstram que de 55 mil códigos definidos, a maioria deles não são usados. Tecnologias já incorporadas a computação, como a usabilidade poderiam vir a contribuir e propor um agrupamento com "uma quantidade menor de códigos".

As propostas de simplificação do CID, como feita pelos códigos de protocolos é valida. No entanto, um protocolo clínico não é aplicável a todos os casos. A palavra final sempre será do médico. Por isso, é importante que o CID disponibilize o conjunto de opções o mais abrangente possível. Assim, o profissional terá a liberdade de buscar a máxima exatidão e individualização do achado no registro.

Dados como impressão diagnóstica podem servir muito mais do que dados de achados clínicos. Entretanto, quando um médico coloca seu carimbo em um prontuário, as informações devem estar corretas sob pena de processos judiciais por erro médico. Como, na maioria das vezes os médicos não têm acesso a exames rápidos e assertivos, preferem não se comprometer com diagnósticos específicos.

As estatísticas dos hospitais forneceram importante registros que impactam no gerenciamento futuro do serviço. Contudo, os dados não podem ser esquecidos, precisam de constante análise por pessoal qualificado para a aquisição de novas informações.

Como contribuições, duas distribuições peculiares dos dados foram encontradas e estudadas durante o processo de coleta, limpeza e análise (uma distribuição livre de escala nunca antes detectada na saúde). Tal descoberta pode levar a novas decisões de projeto para futuras codificações, já que padrões de busca por códigos foram estabelecidas. A descobertos novos padrões de codificação médica é provavel.

Os resultados estatísticos também permitiram a construção de algoritmos que se mostraram promissores na sugestão de códigos clínicos e hospitalares. Estes podem ou ajudar ou ser usados como bases para novas funcionalidades que visem o aumento da qualidade semântica dos códigos gerados por serviços de saúde.

5.1 Respondendo perguntas de pesquisa

As perguntas de pesquisas, localizadas ao final da introdução, podem ser abordadas.

A primeira pergunta era: quais os códigos mais prevalentes? A resposta é: acometimento das vias aéreas e gastrointestinais. Os códigos deste gruo respondem por grande parte da demanda do serviço. São seguidos pelos CIDs ortopédicos, neurológicos e urogenitais. Essa informação mostra-se de grande relevância para a sociedade onde o serviço se encontra, pois confirma perfis epidemiológicos registrados também no serviço público.

A segunda pergunta: como agrupar os códigos para melhorar o acesso e a seleção? As melhores formas de agrupamento para acesso aos CIDs são aqueles que dividem as opções de forma equilibrada. Dentro do contextos dos serviços, são os relacionados as especialidades médicas básicas (clínica, otorrinolaringologia, gastroenterologia, ortopedia, urologia, neurologia, etc.) a idade (infantil, jovem, adulto e idoso) e a dor (diversos CIDs relatam dor). Os médicos conseguem agrupar os códigos com facilidade, sem necessitar novas aquisições de aprendizados. O contexto é muito importante para o agrupamento eficiente.

A terceira pergunta: uma nova funcionalidade pode ser proposta para ajudar na tarefa de escolha dos códigos? A resposta é sim. Os algoritmos foram válidos e simplificaram a tarefa. Obtiveram a aprovação dos médicos que relataram querer algo semelhante nos seus serviços.

REFERÊNCIAS

ALHARBI, M. A.; ISOUARD, G.; TOLCHARD, B. Historical development of the statistical classification of causes of death and diseases. **Cogent Medicine**, 8, n. 1, p. 1893422, 2021.

ANS. **Plano de Dados Abertos.** Rio de Janeiro-RJ, 2019. Disponível em: https://www.gov.br/ans/pt-br/arquivos/acesso-a-informacao/perfil-do-setor/dados-abertos/pda-2019-2021-pdf. Acesso em: 30 março 2021.

ARAUJO, M. C. C.; ACIOLI, S.; NETO, M.; DE MELLO, A. S. *et al.* Nursing protocols: motivation and methodology in the shared construction process/Protocolos de enfermagem: motivacao e metodologia no processo de construcao compartilhada/Protocolos de enfermería: la motivación y la metodología en el proceso de construcción compartida. **Enfermagem Uerj**, 25, n. 1, p. NA-NA, 2017.

ASIF, T.; MOHIUDDIN, A.; HASAN, B.; PAULY, R. R. Importance of thorough physical examination: a lost art. **Cureus**, 9, n. 5, 2017.

BARROS JR, E. D. A. Código de ética médica: comentado e interpretado [Internet]. **São Paulo: Cia do eBook**, 2019.

BERG, M. Problems and promises of the protocol. **Social science & medicine**, 44, n. 8, p. 1081-1088, 1997.

BERTOLINI, G.; FORTUNA, G. G.; VIDAL, M. D. B.; NEVES, O. M. D. C. *et al.* **Manual de Elaboração de Planos de Dados Abertos**. Brasília, DF, 2020. Disponível em: https://www.gov.br/cgu/pt-br/centrais-de-conteudo/publicacoes/transparencia-publica/arquivos/manual-pda.pdf. Acesso em: 30 março 2021.

BIASIBETTI, C.; HOFFMANN, L. M.; RODRIGUES, F. A.; WEGNER, W. *et al.* Comunicação para a segurança do paciente em internações pediátricas. **Revista Gaúcha de Enfermagem**, 40, 2019.

BONISSONE, P. P. **Knowledge and time: a framework for soft computing applications in prognostics and health management (PHM).** NY, USA, 2006. Disponível em: https://www.researchgate.net/profile/P-Bonissone/publication/268666213_Knowledge_and_Time_A_Framework_for_Soft_Computing_Applications_in_PHM/links/550ef5f20cf21287416afc4a/Knowledge-and-Time-A-Framework-for-Soft-Computing-Applications-in-PHM.pdf. Acesso em: 18 julho 2021.

BRÄMER, G. R. International statistical classification of diseases and related health problems. Tenth revision. **World health statistics quarterly. Rapport trimestriel de statistiques sanitaires mondiales**, 41, n. 1, p. 32-36, 1988.

BÖLTE, S.; LAWSON, W. B.; MARSCHIK, P. B.; GIRDLER, S. Reconciling the seemingly irreconcilable: The WHO's ICF system integrates biological and psychosocial environmental determinants of autism and ADHD: The International Classification of Functioning (ICF) allows to model opposed biomedical and neurodiverse views of autism and ADHD within one framework. **Bioessays**, 43, n. 9, p. 2000254, 2021.

CAMPOS, G. M. Estatística prática para docentes e pós-graduandos. Faculdade de Odontologia de Ribeirão Preto da Universidade de São Paulo 2002.

CANCHO, R. F. I.; SOLÉ, R. V. The small world of human language. **Proceedings of the Royal Society of London. Series B: Biological Sciences**, 268, n. 1482, p. 2261-2265, 2001.

CEARÁ, G. D. E. **Monitoramento dos Casos de Dengue, Chikungunya e Doença Aguda pelo Vírus Zika até a Semana Epidemiológica 45 DE 2018**. 2018. Disponível em: https://www.saude.ce.gov.br/wp-content/uploads/sites/9/2018/06/Boletim-Arboviroses-SE-45_2018.pdf. Acesso em: 16 janeiro 2021.

CESAR, C. L. G.; LAURENTI, R.; BUCHALA, C. M.; FIGUEIREDO, G. M. *et al.* Uso da Classificação Internacional de Doenças em inquéritos de saúde. **Revista Brasileira de Epidemiologia**, 4, n. 2, p. 120-129, 2001.

CONTIERO, P.; TITTARELLI, A.; TAGLIABUE, G.; MAGHINI, A. *et al.* The EpiLink record linkage software. **Methods of Information in Medicine**, 44, n. 01, p. 66-71, 2005.

COSTA, K.; ORLOVSKI, R. A Importância da Utilização do Software na Área da Saúde. **Revista Científica Semana Acadêmica**, 50, 2014/03/06 2014.

CRUZ, D. D. A. L. M. D.; PIMENTA, C. A. D. M. Prática baseada em evidências, aplicada ao raciocínio diagnóstico. **Revista latino-americana de enfermagem**, 13, p. 415-422, 2005.

CURRAN-EVERETT, D. Explorations in statistics: the log transformation. **Advances in physiology education**, 42, n. 2, p. 343-347, 2018.

DE CASTRO, A. J. R.; SHIMAZAKI, M. E. **Protocolos clínicos para unidades básicas de saúde**. Belo Horizonte-MG: Escola de Saúde Pública, 2006. 85-7526-203-3.

DEHNAVI, M.; BAGHINI, M. S. National Medical Record Retention Laws. **Spec J Med Res Heal Sci**, 4, n. 4, p. 35-48, 2019.

DI NUBILA, H. B. V.; BUCHALLA, C. M. O papel das Classificações da OMS-CID e CIF nas definições de deficiência e incapacidade. **Revista Brasileira de Epidemiologia**, 11, p. 324-335, 2008.

DOS SANTOS, M. P.; DA ROSA, C. D. P. Auditoria de contas hospitalares: análise dos principais motivos de glosas em uma instituição privada. **Revista da Faculdade de Ciências Médicas de Sorocaba**, 15, n. 4, p. 125-132, 2013.

EBSERH. **Nucleo visa a padronização de protocolos multiprofissionais**. 2017. Disponível em: https://www.gov.br/ebserh/pt-br/hospitais-universitarios/regiao-sudeste/hc-uftm/comunicacao/noticias/nucleo-visa-a-padronizacao-de-protocolos-multiprofissionais. Acesso em: 21 janeiro 2021.

FERREIRA, I. E.; TRINCA, L. A.; FERREIRA, C. P. Delineamentos experimentais eficientes para estudos de cinética química. **Química Nova**, 37, n. 4, p. 589-596, 2014.

FERREIRA, L.; HOCHMAN, B. Padronização da ficha clínica em cirurgia plástica. **Revista Brasileira de Cirurgia Plástica**, 18, n. 2, p. 56-60, 2001.

FORCHESATTO, A. L.; SANTIN, F. A. ESTUDO DE CASO COM OS FRAMEWORK PLAY E PHONEGAP: GERENCIAMENTO E BUSCA DE INFORMAÇÕES PARA CLÍNICAS PEDIÁTRICAS. **Seminário de Iniciação Científica e Seminário Integrado de Ensino, Pesquisa e Extensão**, 2013.

FOSUN. **Fosun Showcased a Diagnostic Lab for the Future at WAIC 2018**. 2018. Disponível em: https://www.fosun.com/language/en/p/29239.html. Acesso em: 03 junho 2019.

FUNG, K. W.; XU, JULIA; BODENREIDER, O. The new International Classification of Diseases 11th edition: a comparative analysis with ICD-10 and ICD-10-CM. **Journal of the American Medical Informatics Association**, 27, n. 5, p. 738-746, 2020.

GERUM, A. C. A. Comparação de modelos formais de segurança da informação: estudo de caso do sistema de controle de registros de saúde em Unidade de Saúde da Família (USF). 2015-09-30 2015.

GONÇALO, C. R. **Gestão Estratégica da Criação do Conhecimento nas Organizações Hospitalares: um estudo baseado na construção de protocolos médico-assistenciais**. 2007. - Programa de Pós-graduação em Administração, UNISINOS, Porto Alegre-RS.

GUSSO, G. Classificação Internacional de Atenção Primária: capturando e ordenando a informação clínica. **Ciência & Saúde Coletiva**, 25, p. 1241-1250, 2020.

HAMEED, K. The application of mobile computing and technology to health care services. **Telematics and Informatics**, 20, n. 2, p. 99-106, 2003.

HANNAN, T. J. Electronic medical records. **Health informatics: An overview**, 133, 1996.

HERBST, K.; JUVEKAR, S.; BHATTACHARJEE, T.; BANGHA, M. *et al.* The INDEPTH Data Repository: an international resource for longitudinal population and health data from Health and Demographic Surveillance Systems. **Journal of Empirical Research on Human Research Ethics**, 10, n. 3, p. 324-333, 2015.

HIRSCH, J.; NICOLA, G.; MCGINTY, G.; LIU, R. *et al.* ICD-10: history and context. **American Journal of Neuroradiology**, 37, n. 4, p. 596-599, 2016.

HOPKINS, W. G. Linear models and effect magnitudes for research, clinical and practical applications. **Sportscience**, 14, p. 49-59, 2010.

HORSKY, J.; DRUCKER, E. A.; RAMELSON, H. Z., 2017, **Accuracy and completeness of clinical coding using ICD-10 for ambulatory visits**. American Medical Informatics Association. 912.

IAN, S. Engenharia de software. **6a. edição, Addison-Wesley/Pearson**, 2003.

IWSOFTWARE. **Ficha Individual dos Pacientes**. 2009. Disponível em: http://iw1.iwsoftware.com.br:9090/IwHelp/Ficha_Individual.html. Acesso em: 18 julho 2021.

JETTÉ, N.; QUAN, H.; HEMMELGARN, B.; DROSLER, S. *et al.* The development, evolution, and modifications of ICD-10: challenges to the international comparability of morbidity data. **Medical care**, p. 1105-1110, 2010.

JIANG, B.; SUN, L.; FIGUEIREDO, D. R.; RIBEIRO, B. *et al.* On the duration and intensity of cumulative advantage competitions. **Journal of Statistical Mechanics: Theory and Experiment**, 2015, n. 11, p. P11022, 2015.

KARJALAINEN, A.; ORGANIZATION, W. H. **International statistical classification of diseases and related health problems (ICD-10) in occupational health**. World Health Organization. 1999.

LAURENTI, R.; NUBILA, H. B. V. D.; QUADROS, A. A. J.; CONDE, M. T. R. P. *et al.* The International Classification of Diseases, the Family of International Classifications, the ICD-11, and post-polio syndrome. **Arquivos de Neuro-Psiquiatria**, 71, p. 3-10, 2013.

LOPES, J. M. C.; BRASIL. Manual de assistência domiciliar na atenção primária à saúde. **Porto Alegre: Serviço de Saúde Comunitária do Grupo Hospitalar Conceição**, p. 48, 2003.

LOPES, L. F. D. **Apostila de Estatística**. 2003. Disponível em: https://www.inf.ufsc.br/~vera.carmo/LIVROS/LIVROS/Luis%20Felipe%20Dias%20Lopes.pdf. Acesso em: 12 janeiro 2019.

MARQUES, M. H. D. **Iniciação à semântica**. J. Zahar, 1990. 8571100861.

MASSEY JR, F. J. The Kolmogorov-Smirnov test for goodness of fit. **Journal of the American statistical Association**, 46, n. 253, p. 68-78, 1951.

MEDICINA, C. F. D. **Código de Ética Médica**. Brasília, DF, 2018. Disponível em: https://portal.cfm.org.br/images/PDF/cem2019.pdf. Acesso em: 10 junho 2021.

MONARD, M. C.; BARANAUSKAS, J. A. Indução de regras e árvores de decisão. **Sistemas Inteligentes-Fundamentos e Aplicações**, 1, p. 115-139, 2003.

O'MALLEY, K. J.; COOK, K. F.; PRICE, M. D.; WILDES, K. R. *et al.* Measuring diagnoses: ICD code accuracy. **Health services research**, 40, n. 5p2, p. 1620-1639, 2005.

OH, S.; CHA, J.; JI, M.; KANG, H. *et al.* Architecture design of healthcare software-as-a-service platform for cloud-based clinical decision support service. **Healthcare informatics research**, 21, n. 2, p. 102-110, 2015.

ORGANIZATION, W. H. **International Statistical Classification of Diseases and Related Health Problems**. 2020. Disponível em: htt://www.who.int/standards/classification-of-diseases. Acesso em: 18 julho 2021.

OTERO VARELA, L.; DOKTORCHIK, C.; WIEBE, N.; QUAN, H. *et al.* Exploring the differences in ICD and hospital morbidity data collection features across countries: an international survey. **BMC health services research**, 21, n. 1, p. 1-9, 2021.

PARK, S. H.; KRESSEL, H. Y. Connecting technological innovation in artificial intelligence to real-world medical practice through rigorous clinical validation: what peer-reviewed medical journals could do. **Journal of Korean Medical Science**, 33, n. 22, 2018.

PAULO, G. D. E. D. S. **Pesquisa CID 10**. 2021. Disponível em: http://www.saudepp.sp.gov.br/farmacia/cons_CID10.asp. Acesso em: 16 janeiro 2021.

PELACCIA, T.; FORESTIER, G.; WEMMERT, C. Deconstructing the diagnostic reasoning of human versus artificial intelligence. **CMAJ**, 191, n. 48, p. E1332-E1335, 2019.

PEREZ, F. L. **Modelos Lineares Generalizados**. 2020. Disponível em: http://leg.ufpr.br/~lucambio/CE225/20211S/CE225.html. Acesso em: 10 junho 2021.

PICON, P. D.; GADELHA, M. I. P.; BELTRAME, A. Protocolos clínicos e diretrizes terapêuticas. Ministério da Saúde, Secretaria de Atenção à Saúde 2014.

PRUDENTE, P. M. D. P. **Pesquisa CID 10**. 2020. Disponível em: http://www.saudepp.sp.gov.br/farmacia/cons_CID10.asp. Acesso em: 18 julho 2021.

RAZZAKI, S.; BAKER, A.; PEROV, Y.; MIDDLETON, K. *et al.* A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. **arXiv preprint arXiv:1806.10698**, 2018.

REILLY, J. R.; SHULMAN, M. A.; GILBERT, A. M.; JOMON, B. *et al.* Towards a national perioperative clinical quality registry: The diagnostic accuracy of administrative data in identifying major postoperative complications. **Anaesthesia and Intensive Care**, 48, n. 3, p. 203-212, 2020.

RODRIGUES, J.-M.; ROBINSON, D.; DELLA MEA, V.; CAMPBELL, J. *et al.* Semantic alignment between ICD-11 and SNOMED CT. *In*: **MEDINFO 2015: eHealth-enabled Health**: IOS Press, 2015. p. 790-794.

ROSSO, C. F. W.; CRUVINEL, K. P. D. S.; SILVA, M. A. D. S.; ALMEIDA, N. A. M. *et al.* Protocolo de enfermagem na atenção primária à saúde no Estado de Goiás. *In*: **Protocolo de enfermagem na atenção primária à saúde no estado de Goiás**, 2014.

ROUSSEAU, D. M. Envisioning evidence-based management. *In*: **The Oxford handbook of evidence-based management**, 2012.

SAVAGLIO, S.; CARBONE, V. Scaling in athletic world records. **Nature**, 404, n. 6775, p. 244-244, 2000.

SCHAUM, K. D. Did you implement the 2011 HCPCS code changes? **Advances in Skin & Wound Care**, 24, n. 2, p. 60-62, 2011.

SCHMIDT, A. F.; FINAN, C. Linear regression and the normality assumption. **Journal of clinical epidemiology**, 98, p. 146-151, 2018.

SCHULZ, S.; KLEIN, G. O. SNOMED CT–advances in concept mapping, retrieval, and ontological foundations. Selected contributions to the Semantic Mining Conference on SNOMED CT (SMCS 2006). Springer. 8**:** 1-3 p. 2008.

SIDHU, R.; PRASANNA, V. K., 2001, **Fast regular expression matching using FPGAs**. IEEE. 227-238.

SIMON, H. A. On a class of skew distribution functions. **Biometrika**, 42, n. 3/4, p. 425-440, 1955.

SNOMED. **SNOMED Home page**. 2021. Disponível em: https://www.snomed.org/. Acesso em: 11 janeiro 2021.

SOUZA, L. Aplicação de aprendizado de máquina para predição de prioridade em gestão de incidentes. 2017. **Acessado em**, 9, p. 12, 2018.

TECHTARGET. **Most Popular White Papers and Webcasts Expert Systems Research**. 2020. Disponível em: https://www.bitpipe.com/tlist/Expert-Systems.html?asrc=RSS_BP. Acesso em: 29 março 2020.

TSENG, S.; FOGG, B. Credibility and computing technology. **Communications of the ACM**, 42, n. 5, p. 39-44, 1999.

UNIGRANRIO. **Curso de Medicina Projeto Pedagógico**. 2018. Disponível em: http://www.unigranrio.com.br/_docs/cursos/PPC_MEDICINA_2018_atualizado.pdf. Acesso em: 18 julho 2021.

VAN MENS, K.; ELZINGA, E.; NIELEN, M.; LOKKERBOL, J. *et al.* Applying machine learning on health record data from general practitioners to predict suicidality. **Internet Interventions**, 21, p. 100337, 2020.

WERNECK, M. A. F.; DE FARIA, H. P.; CAMPOS, K. F. C. **Protocolos de cuidados à saúde e de organização do serviço**. Belo Horizonte-MG: Editora Coopmed, 2009. 84 p. 978-85-7825-021-8.

WESSA, P. **Free Statistics Software, Office for Research Development and Education, version 1.2.1**. 2021. Disponível em: http://wessa.net./. Acesso em: 18 julho 2021.

WILLIAMS, F.; BOREN, S. A. The role of electronic medical record in care delivery in developing countries. **International journal of information management**, 28, n. 6, p. 503-507, 2008.

WINKLER, V.; OTT, J. J.; BECHER, H. Reliability of coding causes of death with ICD-10 in Germany. **International journal of public health**, 55, n. 1, p. 43-48, 2010.

WOOLDRIDGE, J. M. **Introdução à econometria: uma abordagem moderna**. Pioneira Thomson Learning, 2006. 8522104212.

WOYTE, A.; SARR, B.; DE BRABANDERE, K.; RICHTER, M. *et al.* Better fault detection and diagnosis with artificial intelligence: methods, examples and business cases. **35th EU-PVSEC**, p. 1545-1547, 2018.

YAN, Y.; ZHANG, J.; HUANG, B.; SUN, X. *et al.*, 2015, **Distributed outlier detection using compressive sensing**. 3-16.

YEASMIN, S., 2019, Riyadh, Saudi Arabia. **Benefits of artificial intelligence in medicine**. IEEE. 1-6.

APÊNDICE A – SOFTWARES ACESSÓRIOS AO TRABALHO

1. gerador de unidade de dado: acessa o banco de dados, coleta e salva as unidades de dados.
2. contabilizaram os códigos existentes: agrupa e quantifica cada código.
3. gerador das *strings* dos códigos: buscar em páginas web a descrição semântica do código (nome do elemento). A entrada lista de códigos gera a saída lista de códigos ligadas a descrição conforme a OMS
4. filtros: para retirar das consultas atributos irrelevantes.

Relacionados a Bancos de Dados

Fases do acesso ao banco:

1. criar arquivo json.
2. criação dicionário para acesso a dados.
3. seleção de dados de retorno.
4. estabelecimento da conexão com banco.
5. realização da query do dia selecionado.
6. agrupamento em função do atendimento.
7. cria lista de registros do atendimento.
8. alimenta o json.
9. vai para o próximo registro de atendimento.

Figura 22 – Código utilizado para reagrupar consultas.

```
#Criar o json a partir da string
import json

#intertools ferramentas de internet e groupby, agrupamento
from itertools import groupby

#pprint imprimir organizadamente
from pprint import pprint

#cria o objeto pacientes, eh uma lista
pacientes = []

# str_json = json.dumps(obj, indent=4)

#dicionario mapeamento do banco. Dois elementos forma um num array de retorno
FL_MAP = {
    'ATRIBUTOS': "chave",
    ...
    }

#funcao que cria a string a partir dos atributos e retorna o num mapeado
def fl_map_func(fl_tipo_atributo, fl_tipo_valor):
    return FL_MAP[f'{fl_tipo_atributo}_{fl_tipo_valor}']

#seleciona o dado para retorno na tabela correspondente
def key_paciente_func(obj):
    return obj[0]

#retorna o protocolo que está na coluna 1, 2a coluna
def key_protocolo_func(obj):
    return obj[1]

#estabelece a conexao e faz a query
with P.connect("argumentos") as connection:
    cursor = connection.cursor()
    result = cursor.execute("""
        SELECT
            "Query SQL"
#groupby aceita funcao para agrupamento recebe lista = result e funcao key [0] codigo atendimento

for codigo_atendimento, registros in groupby(result, key_paciente_func):
    paciente = {
        'codigo_atendimento': codigo_atendimento,
        (...)
    }
    #groupby agrupa outra vez tudo com funcao key[1] protocolo
    for cd_protocolo, registros_clinicos in groupby(registros, key_protocolo_func):
        protocolo = {
            (...)
        }
        #cria uma lista de registros-clinicos e captura todos os registros
        for registro_clinico in registros_clinicos:
            protocolo['registros_clinicos'].append({
                (...)
            })

        paciente['protocolos'].append(protocolo)
    pacientes.append(paciente)

#gera um json do tipo string
str_json = json.dumps({'results': pacientes}, indent=4)

#abre o arquivo e escreve
with open('08_10_2019.json', 'w') as f:
    f.write(str_json)
```

Fonte: elaborado pelo autor (2019).

Figura 23 – Exemplo de código para formação de um banco MongoDB.

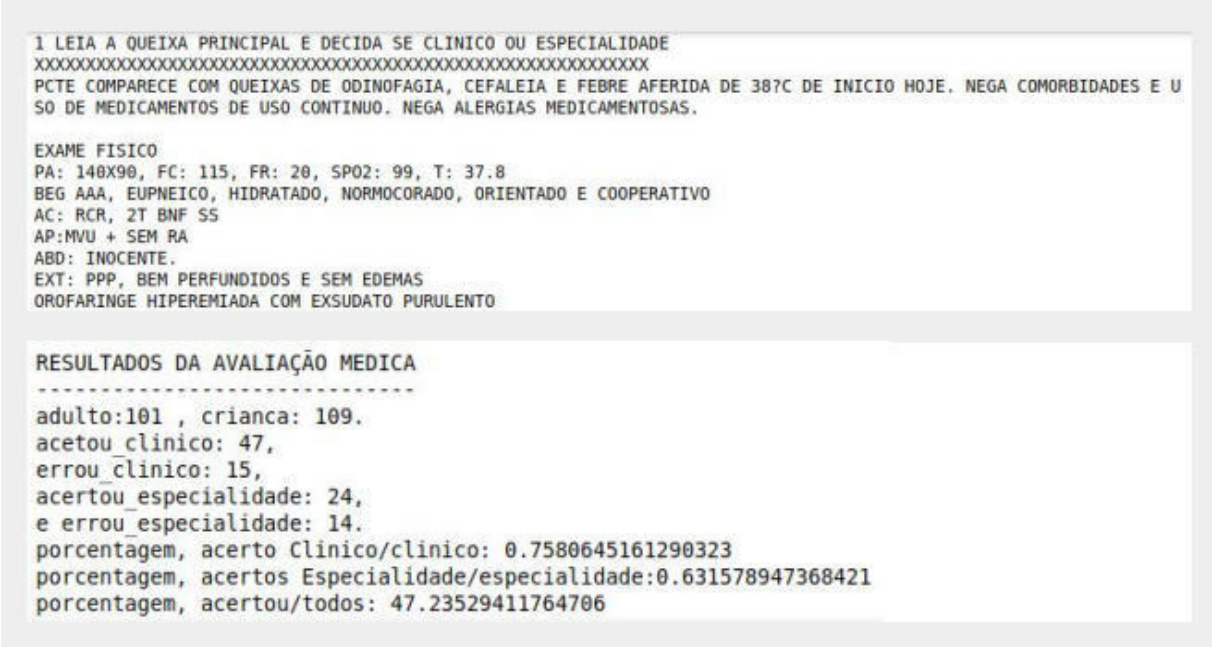
```
#acesso ao banco mongo
from pymongo import MongoClient

#funcoes do mongo
import json

#criando conexao com o mongo
conn = MongoClient('localhost', 27017)

#criando banco com nome bdHap
db = conn.bdHap

#criando collection nome collHap
collection = db.collHap

#cria lista inicio 1 ate 31
for item in range(1,32):
    # se o item for maior ou = 10 retorne o item (1o) else 0 concatenado com item
    dia = item if item >= 10 else f'0{item}'
    strDia = f'./{dia}_10_2019.json'
    print(strDia)

    # consumir json
    # abre o dia  e cria um apelido
    with open(strDia) as json_file:
        # usa funcao load que transforma para um modelo python de intepretacao
        data = json.load(json_file)
        #encontra o dicionario no array resultados
        #todos os atendimentos
        results = data['results']

        for item in results:
            #datatime converte string funcao strptime do item dt_atendimento no formato definido
            item['dt_atendimento'] = datetime.strptime(item['dt_atendimento'], '%Y-%m-%d')
            item['hr_atendimento'] = int(item['hr_atendimento'])

        # insere todo a array no mongo
        collection.insert_many(results)
```

Fonte: elaborado pelo autor (2019).

Programa de Teste Médico

Programa no Jupyter Notebook para avaliação dos agrupamentos obtidos por médicos voluntários.

Figura 24 – Tela do Jupyter Notebook que rodava o teste para os médicos.



Fonte: elaborado pelo autor (2020).

APÊNDICE B – PROJETO PILOTO PARA PESQUISA *SURVEY*

Como explicar tal diferença entre as distribuições? Quais os eventos justificariam? Essas perguntas foram parcialmente respondidas a partir do questionário aplicado aos médicos do serviço. Os médicos têm duas formas diferentes de preencher códigos como o CID. Isso só pode ser respondido a partir de uma pesquisa futura.

Hipótese

Os CIDs dos grupos gerais, que podem ser usados em muitos contextos, são os de alta frequência. As unidades de atendimento generalistas representam a maioria dos serviços de saúde, por isso, no montante, os seus códigos relacionados terão um peso maior.

As distribuições de línguas seguem a lei de escala. Isso também acontece porque a criação de palavras novas é influenciada pelas palavras preexistentes (cite). Sendo o CID também um código com significado semântico, era de se esperar mesmo comportamento.

Primeira Distribuição

A primeira distribuição corresponde aos CIDs que, na maioria das vezes, já são memorizados ou pré-selecionados. Tabelas de fácil acesso pelos profissionais de saúde são comuns em ambientes clínicos e hospitalares.

Segunda distribuição

A segunda distribuição, códigos desconhecidos pelos médicos, que, possivelmente, foram buscados em uma lista ou mesmo em pesquisas da internet. O profissional de saúde precisa, diante de uma situação atípica, registrar bem o achado clínico no sentido de destacar sua conclusão diagnóstica ou achado. Este "destaque semântico" acontece por necessidade: 1) de encaminhamentos a outros serviços de saúde de maior nível de complexidade, 2) de medicamentos diferenciados, 3) de maior segurança documental médico-legal, 4) de notificações compulsórias, entre outros motivos. O profissional de saúde, neste momento, vai fazer uma busca de um código mais específico na internet ou em mídia disponível por códigos específicos. Essa tarefa, não dá vantagem a escolhas semânticas, justificando a distribuição linear. Pesquisa piloto em 25 médicos do sistema de saúde em estudo.

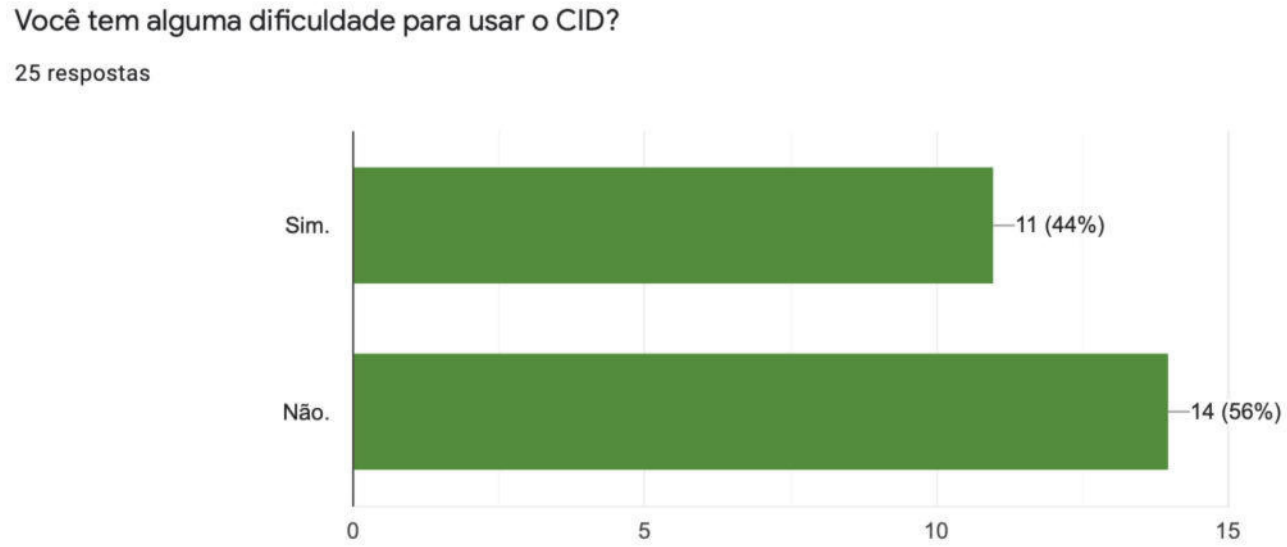


Figura 25 – O que o médico acha de usar o CID.

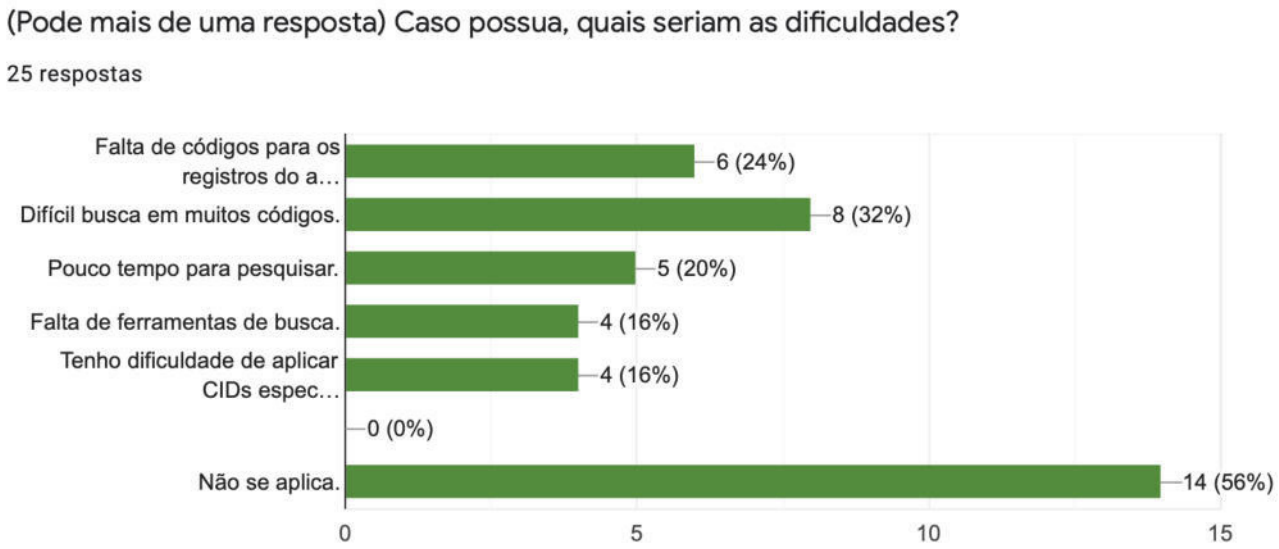


Figura 26 – Identifica dificuldades para o uso do CID.

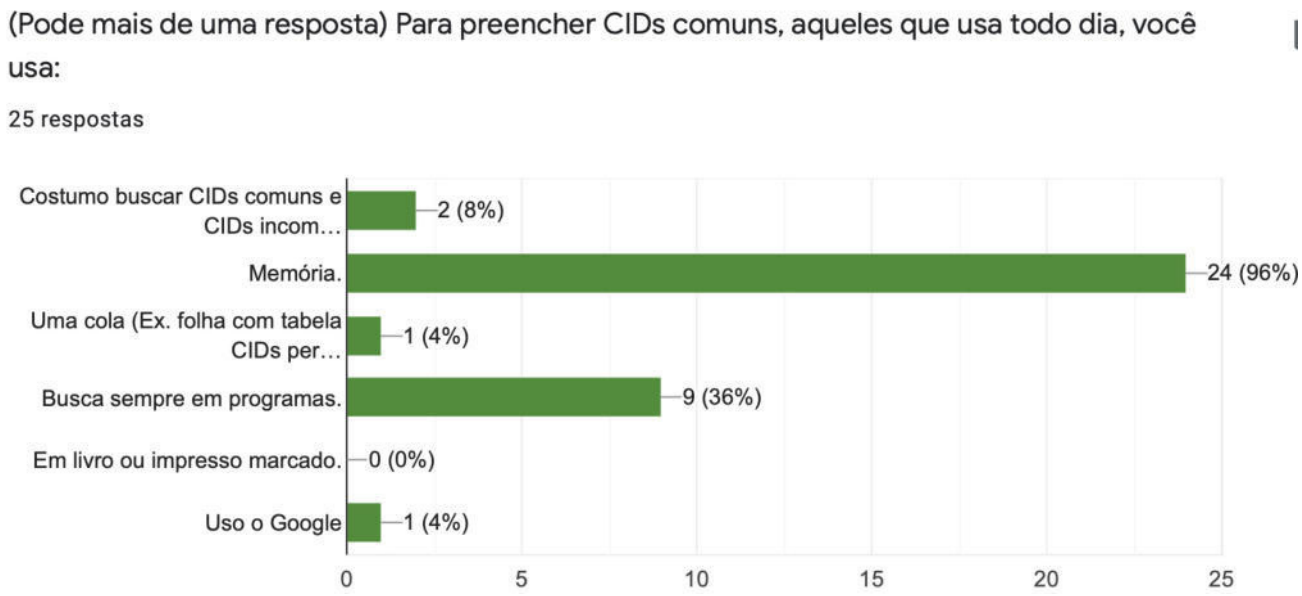


Figura 27 – O uso da memória dos códigos justifica a primeira distribuição.

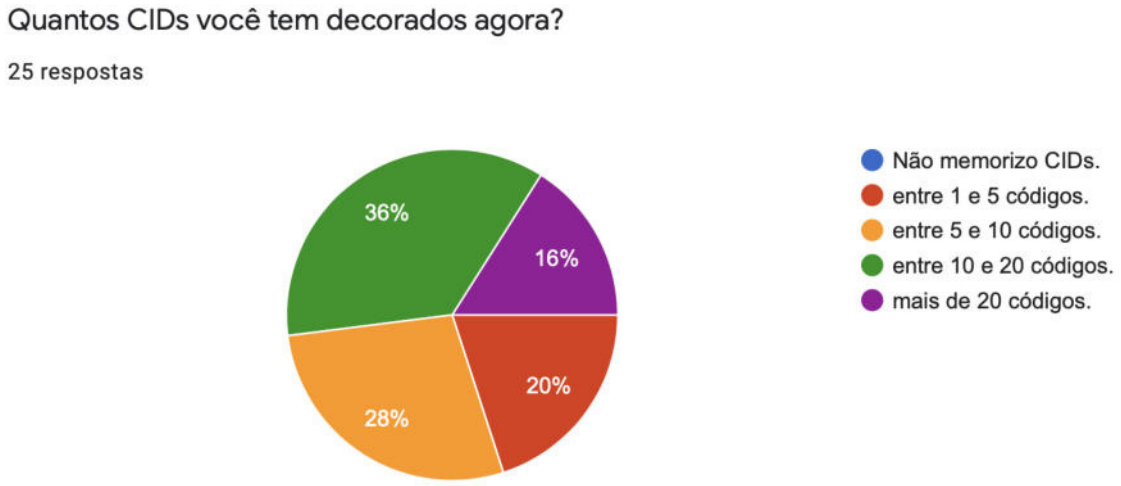


Figura 28 – O gráfico identifica a dimensão do uso dos CIDs memorizados.

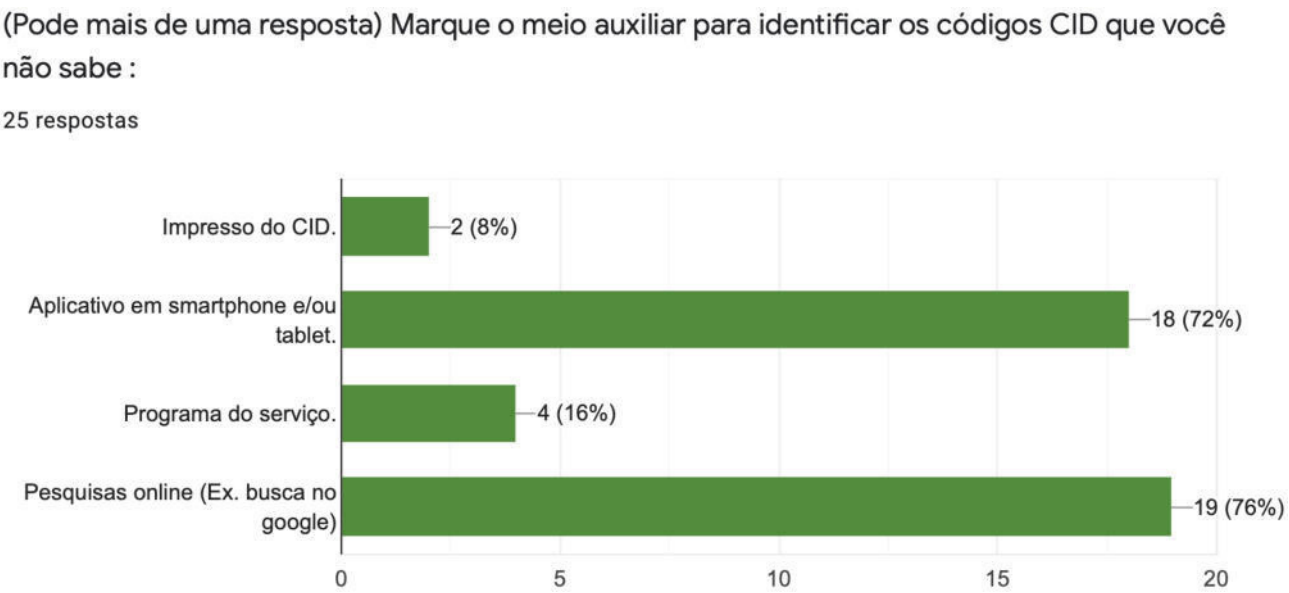


Figura 29 – A pergunta aborda as ferramentas de ajuda que podem justificar a segunda distribuição.

Figura 30 – Avalia o uso dos códigos mais gerais em detrimento aos mais específicos, com melhores valores semânticos.

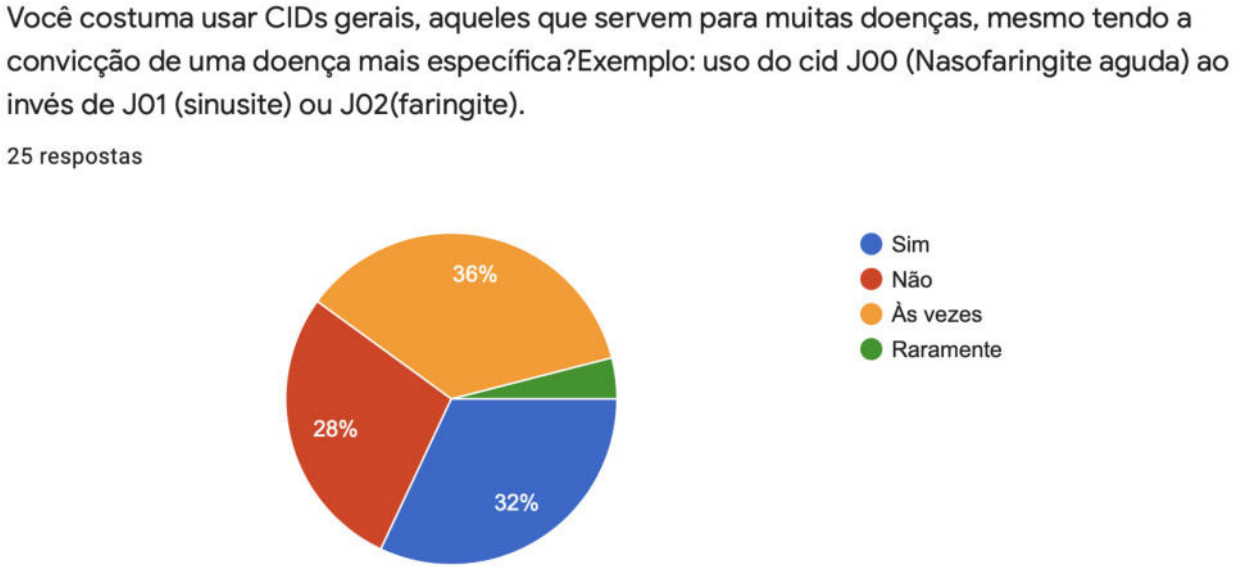


Figura 31 – Também, avalia o uso de códigos mais gerais em detrimento aos mais específicos.

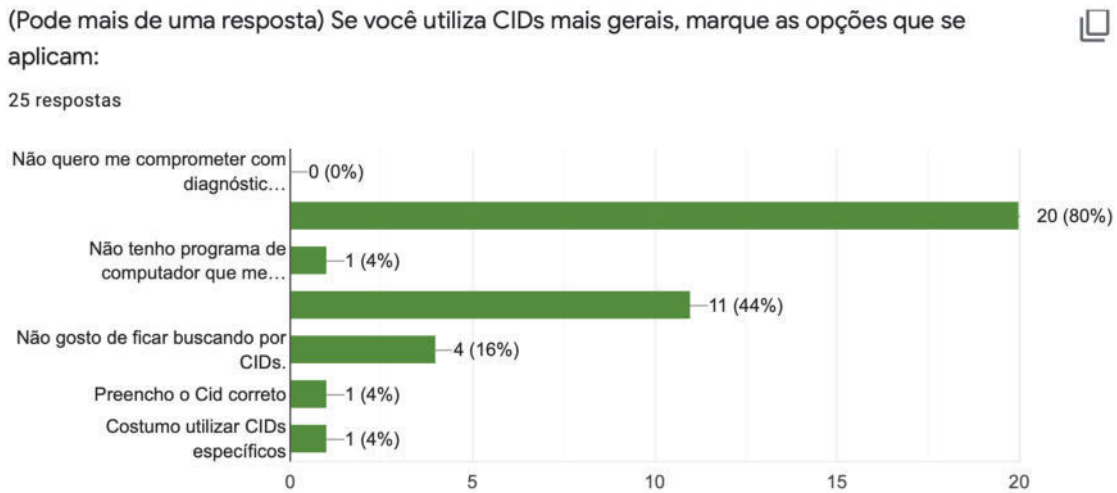


Figura 32 – Identifica os principais motivos para o uso de códigos gerais.