



AVALIAÇÃO DE BASES DE DADOS NA INFORMÁTICA EM SAÚDE: UM ESTUDO DE CASO SOBRE OS REGISTROS HOSPITALARES DE CÂNCER¹

EVALUATION OF DATABASES IN HEALTH INFORMATICS: A CASE OF STUDY ABOUT THE CANCER HOSPITAL REGISTRIES

**Leonardo Minelli², Emilia Damásia de Sousa Xavier³, Eduardo de Paula Alencar³
Maurício De Campos⁴, Paulo Sérgio Sausen⁵**

¹ Trabalho da disciplina de Modelagem Matemática.

² Estudante do Programa de Pós-Graduação em Modelagem Matemática e Computacional. Bolsista Capes.

³ Estudante do Programa de Pós-Graduação em Modelagem Matemática e Computacional.

⁴ Professor da disciplina de Modelagem Matemática.

⁵ Professor orientador.

RESUMO

Devido ao aumento gradual dos casos de câncer, existe a necessidade de se utilizar recursos informáticos para armazenar informações de tais acontecimentos. Isto é justificado pela atenção que deve ser dada ao caso registrado, a fim de facilitar o seguimento do tratamento. Porém, existem diferentes tipos de métodos de registros, sendo a avaliação destes registros hospitalares e dos sistemas que executam este armazenamento um processo primordial para garantir a qualidade dos dados registrados. Os resultados encontrados auxiliam na escolha do sistema de registros hospitalares de câncer a ser utilizado em posteriores pesquisas, o que garantirá a integridade e a confiabilidade das informações geradas posteriormente.

Palavras-chave: Banco de Dados. Avaliação. Informática. Saúde. Câncer.

ABSTRACT

Due to the gradual cancer cases incidence increase, there is need of using computing resources to store such events information. This is justified by the attention that must be given to the case registered in order to facilitate the treatment follow-up. However, there are different types of record methods, with reference to this, the evaluation of hospital records and storage systems that perform this process is paramount to ensure the data record quality. The results help to choose a system of hospital cancer registry to be used in further research, which will ensure the integrity and reliability of the information generated from this.

Keywords: Database. Evaluation. Informatics. Health. Cancer.

INTRODUÇÃO

O câncer há muito tempo foi considerado uma doença que possui grandes números de registros de casos, comprovados em países desenvolvidos e com alto Índice de



Desenvolvimento Humano. Porém, atualmente, estes registros têm ganhado maior dimensão com relação à situação global existente, atingindo também níveis superiores de registros nos países em desenvolvimento (BRASIL, 2019). A Organização Mundial de Saúde (OMS), a partir de estudos que englobam o aumento gradual dos casos de câncer, prevê que a cada ano cerca de 400.000 crianças desenvolvam câncer (WHO, 2022).

Em virtude deste aumento de casos de câncer, foram desenvolvidos sistemas computacionais e bases de dados especializadas para efetuar o armazenamento e a gerência dos casos de câncer. Os registros nestes sistemas, chamados Registros Hospitalares de Câncer (BRASIL, 2000), possuem grande importância com referência à assistência prestada ao paciente, coletando informações pessoais, culturais e sociais, sendo estas utilizadas para a investigação da doença e tratamento com maior probabilidade de sucesso baseado em experiências anteriores armazenadas. Dada esta importância com relação ao seguimento do caso de câncer e da evolução da doença, tais bases de dados devem ser consistentes, atualizadas, padronizadas e devem possuir todas as informações possíveis capazes de auxiliar no tratamento do câncer.

Visando os requisitos funcionais que tais bases de dados possuem, existe a necessidade de avaliação prática, tanto da estrutura destas bases de dados quanto de seus registros, condizente com o armazenamento e a recuperação de informações relacionadas com os casos de câncer. O presente estudo foi fundamentado com o propósito de avaliar bases de dados, objetivando a recuperação dos registros armazenados e a posterior representação do conhecimento gerado a partir destas bases.

REGISTROS HOSPITALARES DE CÂNCER

Os Registros Hospitalares de Câncer (RHC) são definidos como uma coleção de dados registrados a partir de casos de câncer incidentes em um hospital. Os sistemas que possuem tais registros têm por objetivo principal auxiliar administrativamente o hospital no qual está instalado. O estabelecimento de sistemas para efetuar este tipo de acompanhamento é extremamente benéfico, em virtude da crença histórica de que os pacientes são mais bem servidos através da presença de um registro, uma vez que os pacientes retornarão para exames de acompanhamento (SHANMUGARATNAM, 1991).



A iniciativa de efetuar o acompanhamento dos casos de câncer consolidou-se no ano de 1929 na Alemanha. Seguido deste, vários mapeamentos de registros de casos de câncer de base populacional foram elaborados, conforme listado na Tabela 1.

Tabela 1. Registros de câncer de base populacional

País (região)	Ano de estabelecimento	Participação
Alemanha (Hamburgo)	1929	Voluntária
EUA (Nova York)	1940	Obrigatória
EUA (Connecticut)	1941	Obrigatória
Dinamarca	1942	Obrigatória
Canadá (Saskatchewan)	1944	Obrigatória
Inglaterra e País de Gales (Sudoeste)	1945	Voluntária

Fonte: WAGNER (1991).

No ano de 1946, várias mudanças foram sugeridas e recomendadas pela Comissão Interina da OMS para que fosse possível o estabelecimento de registros mundiais dos casos de câncer. São elas, segundo Clemmensen (1977): Os dados registrados devem possuir um padrão, para que seja possível executar comparações; Cada país deve possuir um centro de registros para o armazenamento e coleta de tais dados; Deve haver uma organização internacional cujo dever é correlacionar os dados obtidos de cada país.

A partir destas recomendações, vários centros de registros de casos de câncer foram implantados mundialmente, alguns atendendo mais ou menos estes requisitos. As informações básicas que são inseridas nestes tipos de registros condizem com o paciente e com o câncer (CLEMMENSEN, 1977). Segundo Gherstorff (2011), as informações armazenadas no Registro de Câncer da Dinamarca se referem ao paciente (e.g. número único de identificação pessoal nome, sexo, idade, município, região, país, data de nascimento e data de falecimento ou emigração) e ao câncer (e.g. diagnóstico CID-10, histologia, localização primária, comportamento, lateralidade, estágio, grau, data de diagnóstico, base de diagnóstico, número do câncer).

METODOLOGIA



A avaliação realizada neste trabalho tem o propósito de verificar quais os dados presentes nos registros armazenados, como é dada a padronização destes registros, qual a sua capacidade de atualização e qual o nível de qualidade destas informações no RHC.

Com a finalidade de garantir um padrão de avaliação, uma metodologia de avaliação foi proposta para que os RHC fossem submetidos. Antes que fosse realizada tal avaliação, fez-se um processo seletivo dos RHC disponíveis na literatura e da utilização dos mesmos.

DEFINIÇÃO DA AVALIAÇÃO

A definição da metodologia de avaliação dos RHC é dada a partir dos seguintes pontos gerais:

a) Informações dos registros coletados: havendo a necessidade de possuir a maior quantidade de dados possível do registro de caso de câncer, este ponto é justificado pelas posteriores decisões que o especialista tomará perante o tratamento e acompanhamento do caso. Hábitos do paciente, localização primária e detalhada do tumor, entre outros dados são relevantes neste ponto.

b) Padronização: os dados registrados devem seguir os padrões mais utilizados pela *International Agency for Research on Cancer (IARC)*.

c) Atualização: os registros devem ser atualizados periodicamente, garantindo que as informações ilustrem o cenário atual e proporcionem feedbacks precisos aos especialistas com relação aos resultados de tratamentos e prevalências do câncer.

d) Qualidade: a qualidade dos registros é algo importante. Antes de serem consolidados nas bases dos RHC devem passar por filtros e técnicas de pré-processamento que garantem a integridade dos dados publicados.

Os quatro pontos devem estar presentes em um RHC para que seja considerado ideal e para que seja possível a sua utilização em projetos de pesquisa do câncer que envolva cálculos probabilísticos, estatísticas, índices e níveis de prevalência desta doença.

SELEÇÃO DOS RHC

Os RHC utilizados para elaborar o estudo avaliativo proposto foram pesquisados com as seguintes premissas: o órgão competente e detentor dos registros deve ser reconhecido pelo



governo do país no qual exerce sua função; os registros ou base de dados deve ser de utilidade pública, ou seja, estar disponível para uso e desenvolvimento de pesquisa.

Com os requisitos supracitados para a seleção, foram escolhidas as seguintes organizações: Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA), OMS, *American College of Surgeons* (ACS) e o *National Cancer Institute* (NCI). Tais organizações e seus respectivos RHC são listados na Tabela 2.

Tabela 2. Relação de organizações e RHC selecionados

Organização	RHC
INCA	Integrador de Registros Hospitalares do Câncer (Integrador RHC)
OMS	<i>Cancer Incidence in Five Continents</i> (CI5)
ACS	<i>National Cancer Database</i> (NCDB)
NCI	<i>Surveillance, Epidemiology, and End Results</i> (SEER)

Fonte: Autor.

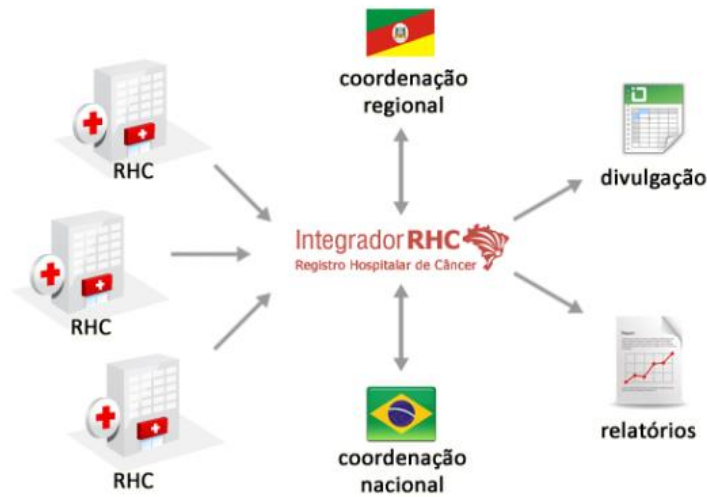
Todos os registros encontrados e selecionados possuem uma base de dados já consolidada dos RHC, ou possuem publicações numéricas de valores de interesse para a elaboração de prevalências e pesquisas condizentes com a oncologia.

PROCESSO DE USO DOS RHC

Para que fosse possível obter resultados de avaliação dos RHC, foi necessário elaborar um processo de população de base de dados distinto para cada RHC.

O Integrador de Registros Hospitalares do Câncer (Integrador RHC) é uma base de dados em nível nacional, desenvolvido pelo INCA que possibilita o envio, a consolidação, o acompanhamento e a análise de dados hospitalares sobre o câncer. O desenvolvimento do Integrador RHC é justificado pela facilidade de acesso a estas informações, rapidez na transmissão dos dados, utilização de dados padronizados, a criação de um banco de dados nacional, amplo acesso na divulgação e armazenamento central em local seguro (BRASIL, 2010). O fluxo de informações do Integrador RHC pode ser visualizado na Figura 1.

Figura 1. Fluxo das informações do Integrador RHC.



Fonte: BRASIL (2011)

O fluxo de informações é dado a partir da seguinte sequência: o paciente com caso de câncer inicialmente é registrado no hospital que possui uma unidade especializada para tratamento de câncer (e.g. CACON – Centro de Alta Complexidade em Oncologia). Com isto, o hospital armazena os registros para posteriormente enviar aos cuidados do INCA. Quando são submetidos estes registros, a coordenação regional do Integrador RHC, presente em todos os estados, avalia tais dados e repassa para a coordenação nacional que executa ajustes e adequa os mesmos para serem posteriormente disponibilizados e utilizados em relatórios que auxiliem na vigilância dos casos de câncer.

Os pontos positivos que podem ser citados do Integrador RHC são: quantidade de informação, confiabilidade dos dados, alta frequência de atualização e dados condizentes com os padrões internacionais adotados.

O *Cancer Incidence in Five Continents* (CI5) tem por objetivo agrupar dados comparáveis sobre a incidência do câncer a partir de uma variedade de localizações geográficas, sendo atualização feita de cinco em cinco anos. Diferente do Integrador RHC, este possui suas informações disponibilizadas em volumes literários contendo tabulações de dados agrupados por gênero, faixa etária e localização do câncer. Em cada volume são disponibilizadas atualizações ou informações de uma região ou tipo de câncer, além de percentuais de casos comprovados microscopicamente, casos de óbito, base de diagnóstico, casos de óbito por tipo e por localização do câncer.



O método de utilização proposto para este RHC foi de transcrição dos dados captados nos volumes publicados, além do download de arquivos no formato CSV que também possuem estes registros com dados mais específicos. Tais registros foram utilizados para popular uma base de dados e esta posteriormente utilizada como representação de dados deste RHC.

Um dos benefícios do CI5 é sua abrangência mundial de dados. Isto possibilita a visão global dos casos de câncer, prevalência mundial dos tipos de câncer, diferença de comportamento da doença perante grupos sociais e grupos culturais diferentes, se a localização geográfica é um fator decisivo para o desenvolvimento de células cancerígenas, entre outros pontos interessantes.

O *National Cancer Database* (NCDB) é uma base de dados composta pela Comissão do Câncer da ACS e a *American Cancer Society* para mais de 1.500 comissões credenciadas nos Estados Unidos e Porto Rico. Este é responsável por mais de 70% dos novos registros efetuados de casos de câncer nos Estados Unidos da América (ACS, 2022a).

Os dados do NCDB podem ser acessados remotamente, uma vez que são mantidas várias aplicações web para disponibilizar as informações em domínio público ao alcance de clínicos, médicos, especialistas e pesquisadores. Entre os dados publicados, é possível saber os tipos mais comuns de câncer que afetam a população norte-americana, a quantia de casos registrados em seus estágios específicos, entre outras informações que auxiliam na pesquisa do câncer. Estas informações podem ser visualizadas na Figura 2.

Figura 2. Informações do NCDB.



<i>Site by Stage of Top 10 (out of 14) Sites Cancers Diagnosed in 2019</i>										
<i>All Diagnosis Types - All Types Hospitals in All States - Data from 1330 Hospitals</i>										
Site	Stage								Totals	
	0	I	II	III	IV	OC	NA	UNK	N	%
1 Breast	51842 15.9%	184243 56.6%	37615 11.6%	18430 5.7%	13519 4.2%	0	8430 2.6%	11457 3.5%	325536	26.26%
2 Prostate	1 0	36378 21.2%	68278 39.9%	36957 21.6%	19474 11.4%	0	193 0.1%	9973 5.8%	171254	13.81%
3 Lung, Bronchus - Non-Small Cell Carcinoma	837 0.5%	50228 30.8%	14623 9%	28941 17.7%	57211 35%	73	6827 4.2%	4506 2.8%	163246	13.17%
4 Colon	3316 3.6%	16490 17.9%	22261 24.2%	21925 23.8%	20036 21.8%	0	574 0.6%	7517 8.2%	92119	7.43%
5 Melanoma of the Skin	22122 28.9%	32461 42.4%	8818 11.5%	6683 8.7%	3537 4.6%	0	80 0.1%	2933 3.8%	76634	6.18%
6 Kidney and Renal Pelvis	1518 2.3%	35105 53.5%	4178 6.4%	8716 13.3%	8633 13.2%	0	1066 1.6%	6430 9.8%	65646	5.3%
7 Urinary Bladder	30828 47.5%	13919 21.4%	9131 14.1%	4259 6.6%	3290 5.1%	0	748 1.2%	2742 4.2%	64917	5.24%
8 Corpus Uteri	0 0	39709 66%	2549 4.2%	6957 11.6%	4926 8.2%	0	405 0.7%	5581 9.3%	60127	4.85%
9 Pancreas	402 0.8%	10138 20.1%	7981 15.8%	6722 13.3%	22547 44.6%	0	311 0.6%	2425 4.8%	50526	4.08%
10 Thyroid	0 0	33341 75.4%	5525 12.5%	644 1.5%	1380 3.1%	0	1231 2.8%	2121 4.8%	44242	3.57%
Top 10 Sites Cancers Total 110866 452012 180959 140234 154553 73 19865 55685 1114247 89.89%										

Fonte: ACS, 2022b.

Como método de utilização elaborou-se o mesmo que para o CI5, com o agrupamento dos registros e com a população da base de dados. A diferença para os dois RHC (i.e. CI5 e NCDB) é que no NCDB existem os dados já tabulados, ou seja, estruturados em tabelas das quais os registros são aninhados por ano de diagnóstico.

A atualização desta base é dada de forma anual, sendo a publicação dos registros efetuada dois anos após o registro elaborado. Apesar desta margem, bons resultados são proporcionados quando estes dados são comparados com a realidade.

O *Surveillance, Epidemiology, and End Results* (SEER) é um programa da NCI que busca prover informações estatísticas do câncer para a população dos Estados Unidos, com a finalidade de reduzir a quantidade de casos de câncer registrados. Seus esforços são direcionados ao controle e prevenção dos casos de câncer.

Com início em 1973 o SEER abrange todo o território dos Estados Unidos da América e armazena dados demográficos, localização primária do tumor, morfologia do tumor, estágio no diagnóstico, primeiro tratamento recebido, entre outras informações de seguimento em seus registros. Com tais dados, o SEER é a única fonte de informação nos Estados Unidos que possui informações condizentes ao estágio do câncer, diagnósticos e dados da sobrevivência do paciente registrado (NCI, 2022).



Os dados de mortalidade e os dados populacionais utilizados para calcular as taxas de incidência dos casos de câncer presentes no SEER são fornecidos pelo *National Center for Health Statistics* e pelo *Census Bureau*, respectivamente, atualizados anualmente. Porém, somente dados numéricos registrados e informações limitadas são possíveis de se acessar em domínio público, o que resulta em pouco recurso informativo para ser utilizado em pesquisas.

RESULTADOS

Dentre os RHC selecionados, submetidos à metodologia de avaliação proposta e analisados, alguns possuem mais recursos que outros, resultando em diferentes níveis da qualidade dos RHC como um todo. A qualidade contextualizada nesta questão condiz com as informações armazenadas em termos quantitativos (i.e. variedade de dados), o que pode gerar dados estatísticos importantes na vigilância dos casos de câncer, acompanhamento demográfico de incidência das doenças cancerígenas e seguimento do tratamento efetuado no caso de câncer comprovado registrado.

Para mensurar a capacidade de uso dos RHC, especificou-se que será atribuída a nota máxima de 1,0 (um) para o RHC que possuir cada um dos seguintes pontos:

- Possuir maior número de informações, condizente ao item (a) da seção 3.1. Para cada informação indisponível na base de dados em análise, foi decrescido o valor proporcional à quantidade máxima de informações disponíveis em uma das bases de dados. O valor definido é dado a partir de:

$$P_I = I_{bd} / I_{max}$$

Onde P_I é o valor da pontuação a ser atribuído, I_{bd} a quantidade de informações da base de dados em análise e I_{max} a quantidade de informações presentes na base de dados com o maior número de atributos;

- Seguir especificamente as orientações da IARC, padronizações internacionais que possibilitem a compreensão dos registros armazenados de forma única. Os valores possíveis para este ponto são 1,0 (um) ou 0,0 (zero);

- Possuir maior frequência de atualização. Para os demais RHC, a nota foi decrescida de forma gradual para cada mês ou ano de retardo na atualização em relação ao RHC com maior pontuação. O valor para este ponto é definido por:

$$P_A = A_{max} / A_{db}$$



Onde P_A é o valor da pontuação a ser atribuído, A_{max} como sendo os meses de atraso de atualização da base de dados com a maior frequência de atualizações e A_{db} como sendo os meses de atraso na atualização da base de dados em análise;

- Manter a qualidade das informações, com a aplicação de técnicas de pré-processamento, supervisão anteriores à publicação dos registros armazenados na base de dados e a integridade dos dados (i.e. sem a presença de valores desconhecidos). O valor para este ponto é definido por:

$$P_Q = Q_{db} \times 0.5$$

Onde P_Q é o valor da pontuação a ser atribuído, Q_{db} é o somatório dos itens que englobam este ponto (i.e. pré-processamento, supervisão e integridade dos dados) e 0,5 é a constante referente à definição do valor para cada item que englobe este ponto;

- Disponibilizar manuais de uso e material bibliográfico para que seja possível executar pesquisas e demais estudos de vigilância do câncer, facilitando seu uso. São acrescidos 0,33 para a existência de cada item disponível, entre os seguintes: manuais oficiais, trabalhos de pesquisa já realizados ou publicados, além da facilidade no acesso e utilização dos dados. O valor final, portanto, é definido por:

$$P_F = F_{db} \times 0.33$$

Onde P_F é o valor da pontuação a ser atribuído, F_{db} é o somatório dos itens que englobam este ponto (i.e. manuais, publicações, e acesso e utilização dos dados) e 0,33 é a constante referente à definição do valor para cada item que englobe este ponto.

Todos os pontos possuem a mesma importância, uma vez que o somatório final das pontuações de cada item definirá qual o RHC que melhor obedece aos requisitos e ideal a ser utilizado para demais trabalhos.

Partindo destes pontos, os quatro RHC encontrados foram analisados e suas condições avaliadas com o auxílio destas métricas. A relação dos RHC e as notas atribuídas aos mesmos são exibidas na Tabela 3.

Tabela 3. Relação dos RHC e avaliações pontuais

RHC	Informações	Padrão	Atualização	Qualidade	Facilidade	Total
Integrador RHC	1,0 (38)	1,0	1,0 (24)	1,0 (2)	0,99 (3)	4,99
CI5	0,76 (29)	1,0	0,4 (60)	1,0 (2)	0,99 (3)	4,15
NCDB	0,32 (12)	1,0	1,0 (24)	1,0 (2)	0,66 (2)	3,98



SEER	0,50 (19)	1,0	0,67 (36)	1,0 (2)	0,66 (2)	3,83
------	-----------	-----	-----------	---------	----------	------

Fonte: Autor.

Para o Integrador RHC foi atribuída a nota máxima nos quesitos de registros e de atualização, justificado pela gama de dados armazenados nas bases de dados e pela frequência constante de atualização, respectivamente. Ainda com relação à atualização, existe um intervalo de dois anos entre o ano corrente e a publicação dos registros efetuados, consequência do período de pré-processamento e consolidação dos dados dos Centros de Oncologia das unidades hospitalares. Este cenário também se faz presente nos demais RHC, portanto não é considerado qualquer decréscimo neste quesito.

O Integrador RHC do INCA obteve o melhor resultado total, compreendido da soma da pontuação dos quesitos (4,99) sendo, portanto, superior aos demais RHC na avaliação. No quesito “Registros”, este foi utilizado como referência em virtude da quantidade de dados presentes nos registros armazenados. Com referência ao quesito “Atualização”, sua pontuação máxima é justificada por, anualmente, serem elaboradas diversas atualizações e, por consequência, publicações de novos casos de câncer registrados. Neste quesito o Integrador RHC e os demais RHC possuem pontos negativos, pois existe um intervalo de tempo até estes dados serem consolidados e posteriormente publicados, sendo este intervalo de, em média, dois anos. Por ser uma característica geral, adotou-se por desconsiderar este problema, em específico, na avaliação.

CONSIDERAÇÕES FINAIS

A partir do cenário de constante modificação das doenças cancerígenas com relação ao seu alastro e disseminação, é visto que a sua incidência se torna mais comum na população mundial. Isto é justificado pela variedade de regiões do corpo que esta pode atingir e pelos agentes internos e externos ao organismo que podem potencializar a evolução de células normais para células cancerígenas em um tecido anteriormente saudável. Por ser uma doença que evolui gradualmente e que possui diferentes métodos de tratamento, o seguimento de cada caso de câncer é necessário ser efetuado e, partindo desta necessidade, os RHC auxiliam neste acompanhamento pelo armazenamento de dados que devem ser considerados durante o tratamento do caso de câncer.



Com a constatação de como os RHC auxiliam no acompanhamento do caso de câncer, a IARC especificou padrões para serem adotados no processo de registro dos dados. Além destes padrões, existe também a especificação de quais informações são consideradas imprescindíveis constar no registro do caso de câncer. Têm-se com estas exigências, bases de dados íntegras e confiáveis.

O presente trabalho resultou na elaboração de uma pesquisa, tanto técnica quanto no estado da arte, que disponibilizasse resultados relacionados ao nível da qualidade dos RHC encontrados a partir do método de seleção proposto. Esta pesquisa proporcionou a capacidade de concluir sobre cada doutrina de armazenamento e gerência dos dados adotada pelos RHC. As sugestões e imposições da IARC foram fielmente seguidas pelas organizações em seus respectivos RHC no quesito de padronização de dados armazenados e de informações básicas necessárias para o registro de caso de câncer. Porém, existiram diferenças entre os RHC avaliados com relação às informações disponíveis, frequência de atualização, técnicas de pré-processamento utilizadas, material de apoio disponível e facilidade de uso.

Ao utilizar a metodologia de avaliação proposta nos RHC selecionados, pode-se considerar o Integrador RHC como a melhor opção a ser utilizada para demais projetos de pesquisa que englobem a necessidade de recuperação de informações condizentes a informática em saúde e a vigilância dos casos de câncer. Além de fácil utilização, o Integrador RHC possui uma metodologia simples e objetiva no processo de registros de casos de câncer e de recuperação destes registros, sendo possível adotá-lo em qualquer artifício técnico como softwares e manipulá-lo com o auxílio de Sistemas Gerenciadores de Bancos de Dados (SGBD).

Com esta avaliação das bases de dados na informática em saúde, pode-se garantir a qualidade dos posteriores trabalhos desenvolvidos neste âmbito que utilizem o Integrador RHC. Alguns exemplos são: a elaboração de prognósticos para os pacientes e a definição de padrões e regras para a ocorrência de casos de câncer. Para tal, o uso de tais bases no Processo de Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases – KDD*) proporciona a capacidade de descobrir e posteriormente representar novos conhecimentos a partir da utilização desta base de dados, detentora de todas as necessidades para esta finalidade.

Como trabalhos futuros, espera-se projetar o prognóstico (i.e. sobrevida) de um paciente de uma unidade hospitalar de acordo com suas características (estágio do câncer,



tratamento efetuado, hábitos socioprofissionais, entre outros). Isto é alcançável a partir da utilização de técnicas de Descoberta de Conhecimento em Bancos de Dados e da representação do conhecimento gerado por intermédio da construção de árvores de decisão, regras de associação ou por outro tipo de método que se enquadre como a melhor forma para esta tarefa.

REFERÊNCIAS BIBLIOGRÁFICAS

- ACS. National Cancer Database (NCDB) | ACS. American College of Surgeons. Disponível em: <https://www.facs.org/quality-programs/cancer-programs/national-cancer-database/>. Acesso em: 14 jul. 2022a.
- ACS. Public Access to Cancer Data | ACS. American College of Surgeons. Disponível em: <https://www.facs.org/quality-programs/cancer-programs/national-cancer-database/publicaccess/>. Acesso em: 14 jul. 2022b.
- BRASIL. Integrador RHC: Ferramenta para a Vigilância Hospitalar de Câncer no Brasil. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA), Rio de Janeiro, 2011.
- BRASIL. Registros Hospitalares de Câncer: Rotinas e Procedimentos. 1ª Edição. Ministério da Saúde, Instituto Nacional de Câncer (INCA), 2000.
- BRASIL. Registros Hospitalares de Câncer: Planejamento e Gestão. 2ª Edição. Ministério da Saúde. Instituto Nacional de Câncer (INCA), Rio de Janeiro, 2010.
- BRASIL. Estimativa 2020: Incidência de Câncer no Brasil. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Rio de Janeiro, 2019.
- CLEMMENSEN, J. Statistical studies in the aetiology of malignant neoplasms. V. Trends and risks. *Acta Pathol Microbiol Scand*, v. 261, p. 1-286, 1977.
- GJERSTORFF, M.L. The Danish cancer registry. *Scandinavian journal of public health*, v. 39, n. 7_suppl, p. 42-45, 2011.
- NCI. Overview of the SEER Program. National Cancer Institute. Disponível em: <http://seer.cancer.gov/about/overview.html>. Acesso em 14 jul. 2022.
- SHANMUGARATNAM, K. Introduction. *In: JENSEN, O.M.; PARKIN, D.M.; MACLENNAN, R.; MUIR, C.S.; SKEET, R.G. (org.). Cancer Registration: Principles and Methods. IARC Scientific Publications n. 95. Lyon, France, 1991. p. 1-2.*
- WAGNER, G. History of cancer registration. *Cancer Registration: Principles and Methods. Volume, v. 95, 1991.*
- WHO. Cancer. World Health Organization. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/cancer>. Acesso em: 14 jul. 2022.