



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
CENTRO DE CIÊNCIAS DA SAÚDE  
INSTITUTO DE ESTUDOS EM SAÚDE COLETIVA

**ÉDNEI CÉSAR DE ARRUDA SANTOS JUNIOR**

QUALIDADE DOS DADOS DOS REGISTROS HOSPITALARES DE CÂNCER:  
uma análise dos casos registrados de câncer de mama e de colo do útero no Brasil e grandes  
regiões entre 2008 e 2016

Rio de Janeiro

2023

ÉDNEI CÉSAR DE ARRUDA SANTOS JUNIOR

QUALIDADE DOS DADOS DOS REGISTROS HOSPITALARES DE CÂNCER:  
uma análise dos casos registrados de câncer de mama e de colo do útero no Brasil e grandes  
regiões entre 2008 e 2016

Monografia apresentada ao Instituto de Estudos  
em Saúde Coletiva, da Universidade Federal do  
Rio de Janeiro, como requisito parcial para a  
obtenção do grau de bacharel em Saúde Coletiva.

Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Natália Santana Paiva

Coorientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Gulnar Azevedo e Silva

Rio de Janeiro

2023

## **FOLHA DE APROVAÇÃO**

**ÉDNEI CÉSAR DE ARRUDA SANTOS JUNIOR**

**QUALIDADE DOS DADOS DOS REGISTROS HOSPITALARES DE CÂNCER: Uma  
análise dos casos registrados de câncer de mama e de colo do útero no Brasil e Grandes  
Regiões entre 2008 e 2016**

Monografia apresentada ao Instituto de Estudos em Saúde Coletiva, da Universidade Federal do Rio de Janeiro, como requisito parcial para a obtenção do grau de bacharel em Saúde Coletiva.

Aprovado em: 16 de janeiro de 2023.

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Natália Santana Paiva (Orientador)**  
**Instituto de Estudos em Saúde Coletiva – IESC/UFRJ**

---

**Prof. Dr. Antonio José Leal Costa**  
**Instituto de Estudos em Saúde Coletiva – IESC/UFRJ**

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Rejane Sobrino Pinheiro**  
**Instituto de Estudos em Saúde Coletiva – IESC/UFRJ**

Kaô Kabecilê, Xangô  
Eparrey, bela Oyá

## AGRADECIMENTOS

Gostaria de agradecer a minha família, minha mãe Patricia, meu padrasto Abreu e meu irmão Emanuel, sem o apoio e auxílio de vocês, eu não conseguiria chegar até aqui. Agradeço também a minha tia Manu, meu tio Cristiano e minha prima Adriana, por terem sempre uma palavra de cuidado e força nos momentos complicados da graduação.

Não poderia deixar de agradecer às minhas mães de santo, mãe Iraci, mãe Jacyra e mãe Jojo por todo o cuidado e auxílio durante essa jornada. Agradeço também ao pai Felício, Tete, Dedeia, Ana Rosa e Grazi, e a toda a família do Cantinho de Oxóssi e Iansã, por sempre me incentivarem e apoiarem durante esse processo.

Agradeço a professora Natália Santana Paiva pela dedicação, auxílio e aprendizado que tive durante a graduação. Você é um exemplo de profissional para mim, sem todos os conselhos e dicas de R, eu não teria trilhado metade desse caminho.

Agradeço a professora Gulnar Azevedo e Silva pela oportunidade que me foi dada no início quando não entendia nada sobre câncer e procurava uma oportunidade de aprendizado. Agradeço por toda a dedicação, carinho e compreensão durante a escrita do TCC.

Agradeço aos professores, Antônio José Leal Costa e Roberto de Andrade Medronho pela oportunidade de conseguir trabalhar com vocês ao longo da graduação, pelos debates e discussões sobre epidemiologia e por compartilharem seus conhecimentos comigo.

Agradeço as minhas amigas da graduação, Alice Neris, Deborah Cavalcante e Mariana Lopes por todo incentivo, palavra amiga e apaziguamento dos meus surtos durante a escrita do TCC, esse trabalho não chegaria até a metade se eu não tivesse vocês ao meu lado.

Por fim, agradeço a todo o corpo discente, docente e técnicos administrativos que contribuíram direta e indiretamente para a conclusão desse trabalho. Um agradecimento especial a Roberto Urger, Sheila Ferreira e Luciana Narciso da biblioteca do IESC, muito obrigado por todo o apoio e ajuda durante a escrita e finalização do TCC.

## RESUMO

SANTOS JUNIOR, Édnei César de Arruda. **Qualidade dos dados dos Registros Hospitalares de Câncer**: uma análise dos casos registrados de câncer de mama e de colo do útero no Brasil e grandes regiões entre 2008 e 2016. Monografia (Graduação em Saúde Coletiva) – Instituto de Estudos em Saúde Coletiva, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2023.

O câncer de mama e colo do útero tem um peso importante na população feminina no Brasil. Os Registros Hospitalares de Câncer (RHC) são fontes de informações essenciais para a avaliação da assistência oncológica no SUS possibilitando a condução de pesquisas clínicas-epidemiológicas e a definição de eixos estratégicos para o planejamento das políticas de saúde e ações de vigilância e controle do câncer no país. **Objetivo:** Avaliar a qualidade dos dados de câncer de mama e de colo do útero registrados nos RHC para o Brasil e Grandes Regiões, 2008 a 2016. **Métodos:** Trata-se de um estudo descritivo avaliativo com base nos dados dos RHC para o Brasil e Grandes Regiões entre 2008 a 2016. Os dados foram coletados no Sistema de Informação de Registros Hospitalares de Câncer (SisRHC) disponibilizados de forma pública pelo Instituto Nacional de Câncer. Utilizou-se o escore de Romero & Cunha para a classificação da qualidade dos dados segundo incompletude e inconsistência. **Resultados:** As maiores incompletudes foram observadas para as variáveis: TNM, Estadiamento Clínico do Tumor e Estado da doença ao final do tratamento para o Brasil. Dentre as regiões, as maiores proporções de incompletude e inconsistência foram observadas na região Centro-Oeste. As variáveis que se destacaram com os piores graus de completude foram as variáveis TNM e Estadiamento clínico do tumor na região Norte; TNM e Estado da doença ao final do tratamento na região Nordeste; TNM e Escolaridade na região Centro-Oeste; TNM e Escolaridade na região Sudeste e TNM e Estado da doença ao final do tratamento na região Sul para câncer de mama. Para o câncer do colo do útero, os piores graus de completude foram vistos em TNM e Estado da doença ao final do tratamento para o Brasil, TNM e Raça/Cor da pele na região Norte; TNM e Estado da doença ao final do tratamento na região Nordeste; TNM, Estadiamento clínico do tumor e Escolaridade na região Centro-Oeste; TNM e Estado da doença ao final do tratamento para na região Sudeste e TNM e Estado da doença ao final do tratamento na região Sul. Em relação às inconsistências, a maior proporção ocorreu para “Data da primeira consulta < Data do diagnóstico” para o Brasil e Regiões em ambos os cânceres, com exceção da região Norte, que apresentou maior inconsistência em TNM & Estadiamento para câncer do colo do útero. **Conclusão:** A qualidade dos dados do SisRHC para o Brasil e Grandes Regiões com relação a completude e inconsistência, se mostrou heterogênea entre si, com os piores percentuais localizados na região Centro-Oeste. Contudo, apesar destas limitações, a utilização desta fonte de informação pode oferecer subsídios para a vigilância e controle do câncer no Brasil.

Palavras-chave: Melhoria de qualidade. Neoplasias. Registros hospitalares. Saúde pública. Sistemas de Informação em saúde.

## LISTA DE FIGURAS

Figura 1 - Estrutura anatômica da mama feminina.....	15
Figura 2 - Progressão para o câncer do colo do útero.....	16
Figura 3 - Tipo mais comum de incidência de câncer em 2020 em cada país entre mulheres, excluindo-se câncer de pele não melanoma .....	17
Figura 4 - Fluxo da Informação no IntegradorRHC .....	27
Figura 5 - Distribuição Territorial do Brasil por Grande Região e Unidade de Federação .....	33

## **LISTA DE QUADROS**

Quadro 1 - Roteiro para avaliação de Sistemas de Informação em Saúde.....	20
Quadro 2 - Panorama dos estudos que avaliaram a qualidade de dados para Registros Hospitalares de Câncer no Brasil.....	25
Quadro 3 - Lista de variáveis selecionadas para o estudo.....	34
Quadro 4 - Variáveis selecionadas para análise de inconsistências .....	36



## LISTA DE TABELAS

Tabela 1 - Comparação entre os casos analíticos de câncer de mama e colo do útero registrados na FOSP e SisRHC, São Paulo, 2008 a 2016.....	32
Tabela 2 - Escore de Romero & Cunha, correspondência com a completude e classificação .	35
Tabela 3 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Brasil, 2008-2016.....	40
Tabela 4 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Região Norte, 2008-2016 .....	43
Tabela 5 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Região Nordeste, 2008-2016 .....	45
Tabela 6 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Região Centro-Oeste, 2008-2016.....	48
Tabela 7 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Região Sudeste, 2008-2016 .....	50
Tabela 8 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Região Sul, 2008-2016.....	53
Tabela 9 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer do Colo do Útero, Brasil, 2008-2016 .....	56
Tabela 10 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer do Colo do Útero, Região Norte, 2008-2016.....	59
Tabela 11 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer do Colo do Útero, Região Nordeste, 2008-2016.....	61
Tabela 12 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer do Colo do Útero, Região Centro-Oeste, 2008-2016 .....	64
Tabela 13 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer do Colo do Útero, Região Sudeste, 2008-2016.....	66
Tabela 14 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer do Colo do Útero, Região Sul, 2008-2016.....	68
Tabela 15 - Proporção (%) de inconsistências em variáveis combinadas da ficha de registro do tumor para Câncer de Mama no Brasil e Grandes Regiões, 2008 a 2016.....	71
Tabela 16 - Proporção (%) de inconsistências em variáveis combinadas da ficha de registro do tumor para Câncer do Colo do Útero no Brasil e Grandes Regiões, 2008 a 2016 .....	74

## LISTA DE ABREVIATURAS E SIGLAS

CACON	Centros de Referência de Alta Complexidade em Oncologia
CDC	Centers for Disease Control and Prevention
CID - O	Classificação Internacional de Doenças para Oncologia
CID -10	Classificação Internacional de Doenças
DATASUS	Departamento de Informática do Sistema Único de Saúde
DCNT	Doenças Crônicas Não Transmissíveis
FOSP	Fundação Oncocentro de São Paulo
HPV	Papilomavírus humano
IACR	International Association of Cancer Registries
IARC	International Agency for Research on Cancer
INCA	Instituto Nacional do Câncer
OMS	Organização Mundial da Saúde
OPAS	Organização Pan-Americana da Saúde
RHC	Registros Hospitalares de Câncer
RCBP	Registros de Câncer de Base Populacional
SIS	Sistemas de Informação em Saúde
SISRHC	Sistema de Registro Hospitalar de Câncer
SIM	Sistema de Informações sobre Mortalidade
SINAN	Sistema de Informação de Agravos de Notificação
UF	Unidade Federativa
UICC	União Internacional Contra o Câncer
UNACON	Unidades de Assistência de Alta Complexidade em Oncologia

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>12</b>
1.1 REGISTROS DE CÂNCER NO BRASIL.....	12
1.2 REGISTROS HOSPITALARES DE CÂNCER.....	12
1.3 IMPLANTAÇÃO DOS RHC NO BRASIL.....	13
1.7 AVALIAÇÃO DOS SISTEMAS DE INFORMAÇÃO EM SAÚDE.....	19
1.8 PANORAMA DE ESTUDOS QUE AVALIARAM A QUALIDADE DA INFORMAÇÃO DOS RHC.....	22
1.9 FLUXO DA INFORMAÇÃO.....	27
<b>2 JUSTIFICATIVA.....</b>	<b>29</b>
<b>3 OBJETIVOS.....</b>	<b>30</b>
3.1 OBJETIVO GERAL.....	30
3.2 OBJETIVOS ESPECÍFICOS.....	30
<b>4 METODOLOGIA.....</b>	<b>31</b>
4.1 DELINEAMENTO DO ESTUDO.....	31
4.2 POPULAÇÃO E ÁREA DE ESTUDO.....	31
<b>4.2.1 Brasil e Grandes Regiões.....</b>	<b>32</b>
4.3 FONTE DE DADOS E VARIÁVEIS DE ESTUDO.....	33
4.4 ANÁLISE DE DADOS.....	34
4.5 DISPOSIÇÕES ÉTICAS.....	36
<b>5 RESULTADOS.....</b>	<b>37</b>
5.1 ANÁLISE DE COMPLETUDE.....	37
<b>5.1.1 Câncer de Mama.....</b>	<b>37</b>
<b>5.1.2 Câncer de Colo do Útero.....</b>	<b>54</b>
5.2 INCONSISTÊNCIA.....	69
<b>5.2.1 Câncer de Mama.....</b>	<b>69</b>
<b>5.2.2 Câncer do Colo do Útero.....</b>	<b>72</b>
5.3 SCRIPT PARA ANÁLISE DA QUALIDADE DOS DADOS.....	75
<b>6 DISCUSSÃO.....</b>	<b>76</b>
<b>7 CONSIDERAÇÕES FINAIS.....</b>	<b>82</b>
<b>REFERÊNCIAS.....</b>	<b>84</b>
<b>ANEXOS.....</b>	<b>94</b>
ANEXO A – FICHA DE REGISTRO DO TUMOR.....	95

ANEXO B – SCRIPT NA LINGUAGEM R PARA AVALIAÇÃO DE COMPLETUDE E INCONSISTÊNCIA NA BASE DE DADOS DO SISRHC PARA CÂNCER DE MAMA E COLO DO ÚTERO.....	97
---	----

## 1 INTRODUÇÃO

### 1.1 REGISTROS DE CÂNCER NO BRASIL

No Brasil, a vigilância do câncer se dá por meio do Programa de Epidemiologia e Vigilância do Câncer e seus fatores de risco (PAV), originado pelo Instituto Nacional do Câncer (INCA) em 1999 (OLIVEIRA *et al.*, 2021). O programa é responsável pelas ações relacionadas aos registros de câncer como a implantação, acompanhamento e aprimoramento dessas informações (D'ALESSANDRO; ANTONIAZZI, 2010).

Os registros de câncer se dividem em Registros de Câncer de Base Populacional (RCBP) e Registros Hospitalares de Câncer (RHC). Os RCBP trazem informações sobre a distribuição dos casos de câncer em determinada população geograficamente definida, já os RHC, dispõem de informações sobre a qualidade da assistência prestada nas unidades hospitalares de forma individual ou coletiva (BRASIL, 2007).

### 1.2 REGISTROS HOSPITALARES DE CÂNCER

Os Registros Hospitalares de Câncer (RHC) são alocados em estabelecimentos hospitalares que realizam atendimentos oncológicos, sendo responsáveis pela coleta, armazenamento, tabulação, análise e divulgação das informações sobre os pacientes diagnosticados por neoplasias em suas unidades. Desse modo, a informação obtida pelos RHC é coletada de forma sistêmica e contínua nos indivíduos diagnosticados com neoplasias e tratados nas unidades hospitalares, e auxiliam na realização de estudos que visem entender o cuidado prestado ao usuário, permitindo assim, analisar o desempenho da assistência oncológica oferecida ao paciente pelas unidades de saúde (PINTO *et al.*, 2012; KLIGERMAN, 2001; INSTITUTO NACIONAL DE CÂNCER, 2000, 2010).

A criação de um RHC tem como base a melhoria do cuidado prestado aos pacientes, servindo como uma ferramenta de avaliação do trabalho oferecido pelos estabelecimentos hospitalares, portanto, um RHC vai permitir a realização da vigilância do câncer por parte dos funcionários da unidade que disponibilizam o tratamento aos seus pacientes, e ainda vai oferecer informações ao Ministério da Saúde para melhorias da rede oncológica nos estados e municípios do país (INSTITUTO NACIONAL DE CÂNCER, 2000).

Os RHC utilizam duas fichas para a coleta de dados, sendo elas: A Ficha de Registro de Tumor e a Ficha de Seguimento. Para um preenchimento adequado das fichas é necessária a utilização do Manual de Rotinas e Procedimentos. Diante da necessidade de comparabilidade entre os RHC do Brasil, a ficha de coleta é padronizada e revisada a cada cinco anos. (INSTITUTO NACIONAL DE CÂNCER, 2000).

### 1.3 IMPLANTAÇÃO DOS RHC NO BRASIL

No Brasil, o primeiro Registro Hospitalar de Câncer surgiu em 1983 por iniciativa do Instituto Nacional do Câncer, a criação do RHC baseou-se na padronização das orientações e a utilização de apoio técnico capacitado que visassem a melhoria da qualidade assistencial prestada pelos hospitais, contribuindo para a gestão e criação de pesquisas clínicas oncológicas. Devido ao início da consolidação do RHC no INCA, ocorreu uma distribuição nacional dos RHC em unidades hospitalares que prestassem atendimentos oncológicos. (INSTITUTO NACIONAL DE CÂNCER, 2010).

Relatada a importância dos RHC para a coleta e difusão de informação sobre câncer, surge em 1998, a portaria Nº 3.535/GM, publicada pelo Ministério da Saúde, tornando obrigatório para os centros de atendimento em oncologia dispor e manter em funcionamento o Registro Hospitalar de Câncer em consonância com as normas técnico-operacionais preconizadas pelo Ministério da Saúde. Já em 2005, urge-se a necessidade de organização da rede de atenção oncológica no Brasil, em vista disso, o Ministério da Saúde cria a Política Nacional de Atenção Oncológica, estabelecendo a implantação e credenciamento das Unidades de Assistência de Alta Complexidade em Oncologia (UNACON) e os Centros de Referência de Alta Complexidade em Oncologia (CACON), e a obrigatoriedade de funcionamento dos RHC até doze meses após a publicação da Política Nacional de Atenção Oncológica (INSTITUTO NACIONAL DE CÂNCER, 2010; BRASIL, 2007).

Conforme a implantação e consolidação dos RHC nas unidades hospitalares brasileiras constatou-se a importância desses centros de informações para a vigilância do câncer, dado que possibilitou a criação de pesquisas baseadas na descrição de casos e óbitos, estudos de caso-controle, de sobrevivência e séries temporais para monitorar a situação do câncer nos estados e municípios brasileiros (BRASIL, 2007).

#### 1.4 IMPORTÂNCIA DOS RHC PARA A SAÚDE COLETIVA

É preconizado por órgãos internacionais como a Organização Pan-Americana da Saúde (OPAS), *International Agency for Research on Cancer* (IARC) e a *International Association of Cancer Registries* (IACR) que todos os novos casos diagnosticados de câncer sejam registrados e acompanhados em um RHC para que seja possível existir uma continuidade da assistência ao paciente oncológico (CHAFFIN, 2016). Em decorrência das transformações observadas no perfil demográfico, nutricional e epidemiológico da população brasileira, a carga global de Doenças Crônicas Não Transmissíveis (DCNT), principalmente do câncer, é um problema de saúde pública (D’ALESSANDRO; ANTONIAZZI, 2010). Dessa maneira, torna-se imprescindível uma vigilância epidemiológica do câncer na população, conseqüentemente, os registros hospitalares de câncer surgem como uma forma de não apenas avaliar a assistência prestada na rede oncológica, mas também como uma ferramenta de pesquisa epidemiológica, clínica, de planejamento e avaliação das ações de controle do câncer no Brasil (BRASIL, 2007).

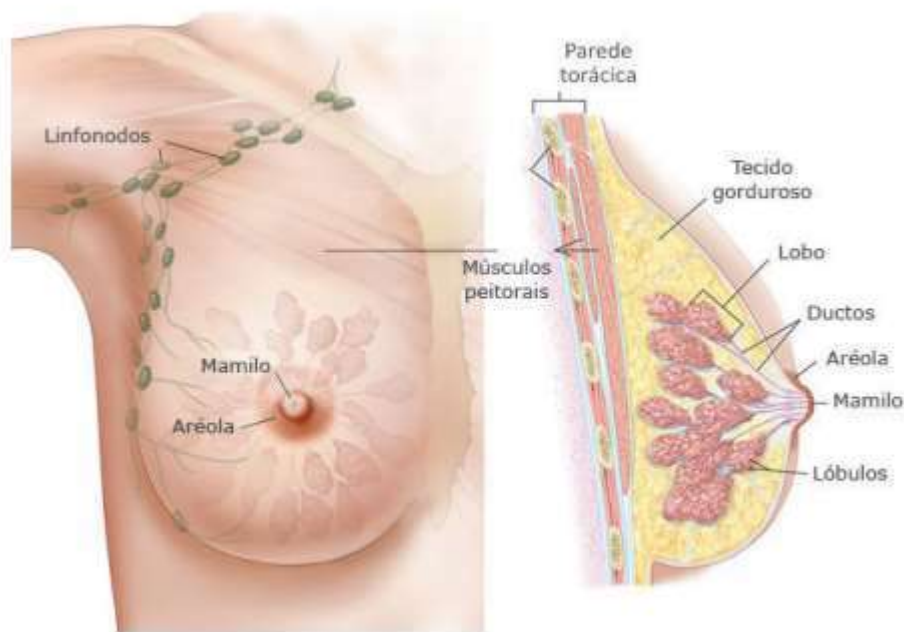
#### 1.5 HISTÓRIA NATURAL DO CÂNCER DE MAMA E COLO DO ÚTERO

O câncer é um resultado de alterações no DNA em uma única célula, ou clones dessa célula, que resultam na perda funcional, crescimento celular descontrolado e metástases. Diferente dos cânceres de origem genética que são adquiridos a partir de uma mutação específica, os cânceres também podem advir de mutações ocasionadas por exposição a agentes externos ou internos, como carcinógenos do tabaco, fatores dietéticos, agentes infecciosos e fatores hormonais (BRENNAN, 2002). Câncer é um termo utilizado para um conjunto de mais de 100 doenças diferentes. Os tipos de câncer correspondem aos tipos de células no corpo humano, se um câncer começa em um tecido epitelial, é denominado de carcinoma. Já se for desenvolvido no tecido conjuntivo, é chamado de sarcoma (INSTITUTO NACIONAL DE CÂNCER, 2022).

O câncer de mama é um tumor que pode se manifestar de forma heterogênea, com variações nas manifestações clínicas e morfológicas. Com relação a evolução do tumor, ele pode progredir de forma agressiva, com rápida evolução para metástase, ou de forma mais branda onde se tem um melhor prognóstico. O câncer de mama é observado mais frequentemente nos ductos mamários (carcinoma ductal) e nos lobos mamários (carcinoma lobular) (Figura 1). Ele pode ser detectado na fase de tumor *in situ*, quando não ultrapassou a

membrana basal do tecido epitelial, ou na fase invasivo, quando ultrapassou a membrana basal. (INSTITUTO NACIONAL DE CÂNCER, 2021).

**Figura 1 - Estrutura anatômica da mama feminina**



Fonte: Instituto Nacional de Câncer José Alencar Gomes da Silva, 2021.

O câncer de colo do útero se desenvolve na parte interior do útero, que é chamado de colo, e fica no fundo da vagina. O desenvolvimento desse câncer ocorre principalmente na zona de transformação do colo uterino, onde o epitélio colunar é substituído pelo novo epitélio escamoso metaplásico. O câncer de colo do útero é mais comumente observado no carcinoma epidermoide que acomete o epitélio escamoso e no adenocarcinoma que atinge o epitélio glandular (INSTITUTO NACIONAL DE CÂNCER, 2021).

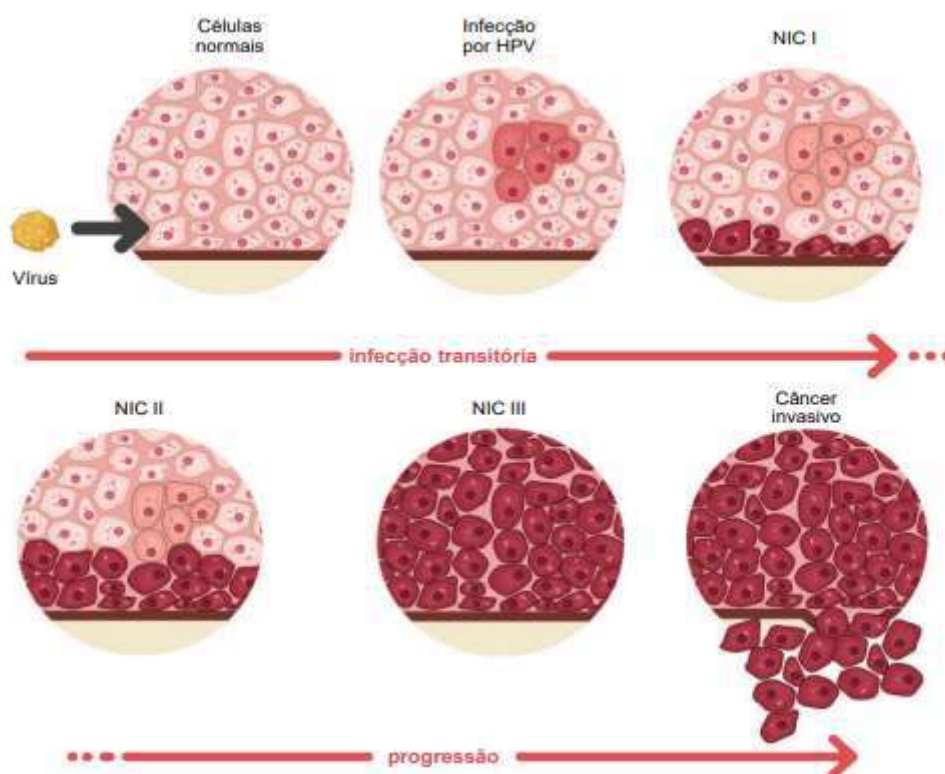
Esse tipo de câncer é causado pela infecção persistente de oncogênicos do Papilomavírus Humano (HPV). Existem mais de 150 tipos de HPV, que estão associados a lesões benignas como verrugas e papilomas laríngeos (HPV não oncogênico ou de baixo risco), lesões precursoras e a vários tipos de câncer (HPV oncogênico ou de alto risco). Os tipos de HPV mais relacionados ao câncer do colo útero são HPV16 (53%), HPV18 (15%), HPV45 (9%), HPV31 (6%) e HPV33 (3%) (INSTITUTO NACIONAL DE CÂNCER, 2021, 2022).

A infecção pelo vírus do HPV não é uma causa suficiente para o desenvolvimento de câncer do colo do útero, a maioria das alterações são eliminadas pela resposta imunológica do organismo. O risco de se desenvolver esse tipo de câncer é de 30% se as lesões precursoras



não forem tratadas. As lesões nas células epiteliais acarretadas pelo HPV progridem de forma lenta para o câncer de colo do útero, passando em torno de 10 a 20 anos como lesões precursoras (Neoplasia Intraepitelial Cervical (NIC) II e III, também chamadas de lesão de alto grau) (Figura 2) (INSTITUTO NACIONAL DE CÂNCER, 2021).

**Figura 2 - Progressão para o câncer do colo do útero**

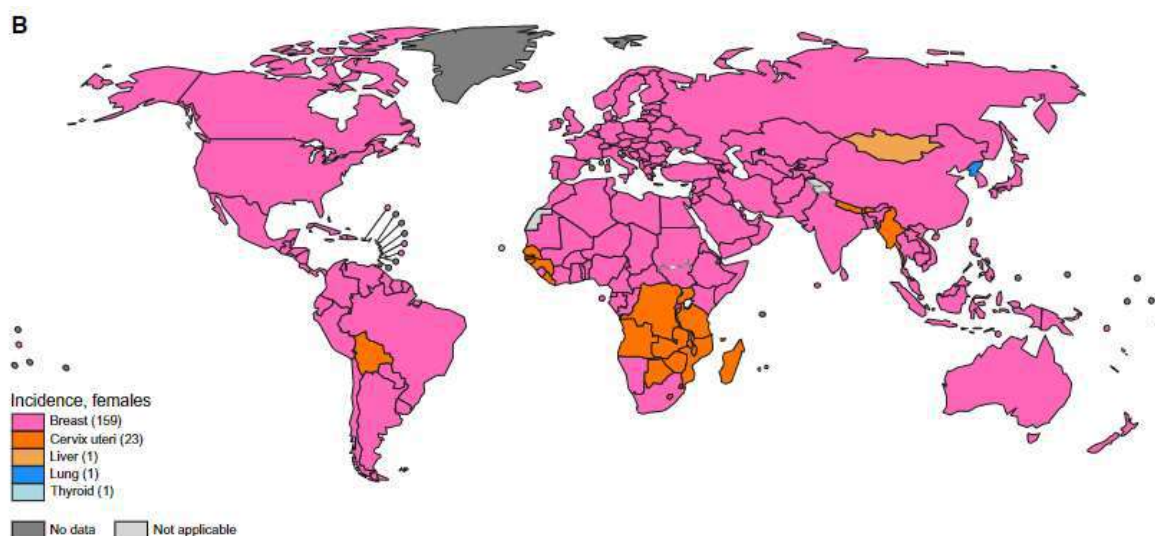


Fonte: Instituto Nacional de Câncer José Alencar Gomes da Silva, 2021.

## 1.6 VIGILÂNCIA DO CÂNCER DE MAMA E COLO DO ÚTERO NO BRASIL

Os cânceres mais incidentes em mulheres no mundo em 2020 foram o câncer de mama e colo do útero, respectivamente, mostrando o impacto que esses cânceres possuem nesse grupo em específico. (Figura 3). No Brasil, o Instituto Nacional do Câncer estimou 66.280 casos novos de câncer de mama e 16.590 casos novos de câncer do colo do útero para o ano de 2020. A distribuição do câncer é heterogênea entre as regiões brasileiras, predominando o câncer de mama na região Sul e Sudeste, e o câncer do colo do útero apresenta um maior impacto na região Norte e Nordeste do país, apesar do câncer de mama ser o mais incidente na população (INSTITUTO NACIONAL DE CÂNCER, 2019).

**Figura 3 - Tipo mais comum de incidência de câncer em 2020 em cada país entre mulheres, excluindo-se câncer de pele não melanoma**



Fonte: GLOBOCAN 2020.

O câncer de mama é o tumor mais incidente em mulheres no Brasil, excetuando-se os tumores de pele não melanoma. Observou-se em países desenvolvidos o aumento da sobrevivência em cinco anos para o câncer de mama, entretanto, circunstâncias relacionadas às barreiras de acesso ao serviço de saúde, diminuem a sobrevivência pela doença devido ao atraso no diagnóstico e tratamento oportuno para esse câncer (INSTITUTO NACIONAL DE CÂNCER, 2019).

A tendência de mortalidade por câncer de mama tem diminuído em países desenvolvidos, em dissonância, os países em desenvolvimento apresentam um aumento na mortalidade por esse tipo de câncer. No Brasil, a distribuição populacional heterogênea entre as cinco macrorregiões brasileiras pode ocasionar em fatores socioeconômicos e culturais que interferem na detecção precoce do câncer de mama, influenciando em um pior prognóstico e possível óbito pela doença (RODRIGUES; BUSTAMANTE-TEIXEIRA, 2011).

O câncer de mama e seus fatores de risco são alvos de pesquisas para se entender a distribuição dessa neoplasia na população, estudos realizados comprovam que hereditariedade, idade avançada da primeira gestação, baixa paridade e amamentar por períodos curtos são considerados fatores de risco, contudo, não são passíveis de modificação. No entanto, a incidência do câncer é influenciada não apenas por fatores biológicos individuais, hábitos comportamentais como o uso de álcool, excesso de peso e a inatividade física após a menopausa se consolidam como fatores de risco para o aparecimento da doença, todavia, é

possível a criação de ações de saúde que tenham como objetivo a diminuição da prevalência desses hábitos comportamentais na população (MIGOWSKI *et al.*, 2018).

Devido a alta incidência da neoplasia de mama na população brasileira, métodos de diagnóstico precoce do câncer se tornam essenciais para a condução de ações de controle da doença no país (TOMAZELLI *et al.*, 2017). O rastreamento para essa neoplasia no país é realizado de forma oportunística através da mamografia de rotina em mulheres de 50 a 69 anos a cada dois anos. Em países que possuem programas de rastreamento efetivo observou-se uma redução na mortalidade, pois possibilita o diagnóstico precoce, e assim, o tratamento adequado e oportuno (BARBOSA *et al.*, 2015; INSTITUTO NACIONAL DE CÂNCER, 2021).

O câncer do colo do útero é o quarto tipo de câncer mais comumente diagnosticado e a quarta principal causa de óbito por câncer em mulheres no mundo, apresentou-se uma estimativa de 604.000 novos casos e 342.000 mortes para o ano de 2020. A doença dispõe de uma carga global com cerca de 85% dos casos novos identificados em países de média e baixa renda (SUNG *et al.*, 2021; BARBOSA *et al.*, 2016).

A neoplasia de colo do útero é um tipo de câncer pré-evitável devido às medidas de prevenção primária com base na aplicação da vacina contra o papilomavírus humano (HPV) e prevenção secundária pautadas no rastreamento (SUNG *et al.*, 2021). Em estudo que buscou avaliar a mortalidade por câncer nas capitais e interior, relatou-se um declínio na tendência da mortalidade por câncer do colo do útero no Brasil em todas as regiões brasileiras, com exceção do interior da Região Norte (SILVA *et al.*, 2020).

O câncer do colo uterino pode ser causado através da infecção genital do papilomavírus humano (HPV), uma doença com alta prevalência na população, contudo, somente alguns indivíduos desenvolvem alterações celulares que podem evoluir para o câncer. A infecção do HPV é considerada uma causa necessária, mas não suficiente para o desenvolvimento desse câncer. Entre os fatores que podem influenciar no aparecimento do câncer cervical se encontram o início precoce da atividade sexual e múltiplos parceiros, tabagismo e uso prolongado de pílulas anticoncepcionais (INSTITUTO NACIONAL DE CÂNCER, 2019; SUNG *et al.*, 2021).

O rastreamento para o câncer do colo do útero é oportunístico através do exame de Papanicolau em mulheres de 25 a 64 anos que já tiveram práticas sexuais (INSTITUTO NACIONAL DE CÂNCER, 2016). O Brasil possui como meta a cobertura de 80% das mulheres rastreadas para câncer do colo do útero (MALTA; SILVA JR, 2013). Países que relatam uma alta cobertura de rastreio para o câncer uterino em sua população observam uma

redução na incidência e mortalidade devido ao diagnóstico precoce e melhor prognóstico da doença (GIRIANELLI; GAMARRA; AZEVEDO E SILVA, 2014).

A Organização Mundial da Saúde (OMS) estabeleceu que 90% das meninas devem ser vacinadas contra o HPV até 15 anos de idade, 70% das mulheres sejam rastreadas com um teste de alto desempenho até 35 anos de idade e novamente aos 45 anos de idade, e 90% das mulheres identificadas com câncer do colo do útero devem receber tratamento. Conforme o cumprimento dessas metas será possível eliminar a neoplasia do colo do útero como um problema de saúde pública (WORLD HEALTH ORGANIZATION, 2020).

## 1.7 AVALIAÇÃO DOS SISTEMAS DE INFORMAÇÃO EM SAÚDE

Os Sistemas de Informação em Saúde (SIS) tem como função a coleta, padronização, codificação e gerenciamento de informações em saúde (WORLD HEALTH ORGANIZATION, 2017). A utilização dos SIS possibilita o monitoramento dos agravos de saúde na população, dessa maneira, os dados disponíveis nestes sistemas servem como base para o planejamento em saúde, avaliação de programas e serviços de saúde e criação de ações de controle para doenças (AGRANONIK; JUNG, 2019). O Brasil conta com a implantação de inúmeros sistemas de informação que abrangem aspectos epidemiológicos, vitais, assistenciais, demográficos e socioeconômicos disponíveis para acesso público no Departamento de Informática do Sistema Único de Saúde (DATASUS) (AGRANONIK; JUNG, 2019; LIMA *et al.*, 2009).

Dada a importância das informações presentes nos SIS para a prevenção, promoção, planejamento e monitoramento das condições de saúde da população, os dados alocados nesses bancos de informação devem dispor de alta proporção de cobertura populacional, de modo que ocorra um fluxo de informações fidedignas à população para a gestão municipal, estadual e federal, que propiciem melhorias no planejamento e ações de saúde (CORREIA; PADILHA; VASCONCELOS, 2014; AGRANONIK; JUNG, 2019; BRAZ *et al.*, 2013). A informação é um recurso estratégico para o setor saúde devido a necessidade da produção de indicadores que retratem as condições sanitárias presentes na população (CORREIA; PADILHA; VASCONCELOS, 2014; LIMA *et al.*, 2009; SILVA *et al.*, 2014).

A qualidade da informação é essencial para o fortalecimento do sistema de saúde. Entretanto, alguns SIS apresentam dados incompletos ou não suficientes para o monitoramento adequado da população, ou seja, a criação de indicadores de saúde que façam uso de dados encontrados em bases de dados não representativas da população subestima a carga de uma

doença nos indivíduos (WORLD HEALTH ORGANIZATION, 2017; FRIAS; SZWARCOWALD; LIRA, 2014). O Brasil não possui uma estratégia definida pelo Ministério da Saúde para a avaliação da qualidade dos dados encontrados nos SIS. Utilizam-se iniciativas não sistemáticas e isoladas para o monitoramento dos dados (LIMA *et al.*, 2009; AGRANONIK; JUNG, 2019).

A avaliação de um Sistema de Informação em Saúde é algo complexo e sem nenhuma diretriz técnica publicada a nível nacional. Dessa maneira, uma forma de avaliar os SIS é através do *Updated Guidelines for Evaluating Public Health Surveillance Systems* publicado em 2001 pelo *Centers for Disease Control and Prevention* (CDC), que propõe um roteiro de avaliação dos sistemas de vigilância em saúde pública (Quadro 1) (CINTHO; MACHADO; MORO, 2016; MONTALVÃO, 2017)

**Quadro 1 - Roteiro para avaliação de Sistemas de Informação em Saúde**

<b>Atributos do SIS</b>	<b>Definição</b>
<b>Simplicidade</b>	A simplicidade de um sistema de vigilância em saúde pública refere-se tanto à sua estrutura quanto à facilidade de operação.
<b>Flexibilidade</b>	Um sistema flexível de vigilância em saúde pública pode se adaptar às mudanças nas necessidades de informação ou condições operacionais com pouco tempo adicional, pessoal ou fundos alocados.
<b>Qualidade dos dados</b>	A qualidade dos dados reflete a completude e validade dos dados registrados no sistema de vigilância em saúde pública.
<b>Aceitabilidade</b>	A aceitabilidade reflete a disposição das pessoas e organizações em participar do sistema de vigilância
<b>Sensibilidade</b>	A sensibilidade de um sistema de vigilância pode ser considerada em dois níveis. Em primeiro lugar, no nível de notificação de casos, a sensibilidade refere-se à proporção de casos de uma doença (ou outro evento relacionado à saúde) detectado pelo sistema de vigilância. Em segundo lugar, a sensibilidade pode se referir à capacidade de detectar surtos, incluindo a capacidade de monitorar mudanças no número de casos ao longo do tempo.
<b>Valor Preditivo Positivo</b>	O valor preditivo positivo (PVP) é a proporção de casos notificados que realmente têm o evento relacionado à saúde sob vigilância
<b>Representatividade</b>	Um sistema de vigilância em saúde pública representativo descreve com precisão a ocorrência de um evento relacionado à saúde ao longo do tempo e sua distribuição na população por local e pessoa.
<b>Pontualidade</b>	A pontualidade reflete a velocidade entre as etapas em um sistema de vigilância em saúde pública
<b>Estabilidade</b>	A estabilidade refere-se à confiabilidade e disponibilidade do sistema de vigilância em saúde pública.

Fonte: Updated Guidelines for Evaluating Public Health Surveillance Systems (CDC, 2001).

Conforme o avanço na consolidação dos Sistemas de Informação em Saúde no Brasil, as informações contidas nesses grandes bancos de dados subsidiam as decisões da gestão na saúde do país. As informações obtidas através dos SIS compõem o maior eixo das informações em saúde produzidas através de fontes secundárias de dados. Desse modo, os SIS constituem a

principal fonte de informação em saúde no país, e, portanto, é necessário que haja dados de qualidade sobre as doenças e agravos. (LIMA *et al.*, 2009; CORREIA; PADILHA; VASCONCELOS, 2014; SILVA *et al.*, 2014).

Devido à abrangência, baixo custo e rapidez dos SIS no território nacional, eles representam uma valiosa fonte de informação sobre as condições de saúde na população (GIRIANELLI *et al.*, 2008). Dessa forma, a avaliação da qualidade da informação dos SIS auxilia na observação das fragilidades e potencialidades das informações produzidas a partir dos dados obtidos pelos Sistemas de Informação em Saúde (CORREIA; PADILHA; VASCONCELOS, 2014). Entretanto, o termo qualidade da informação é considerado uma categoria multidimensional, e, assim, não existe um conceito específico determinando a sua abordagem metodológica (SILVA *et al.*, 2014, LIMA *et al.*, 2009).

No campo da saúde necessita-se de informações de qualidade para a criação de indicadores para o planejamento, gestão, avaliação de políticas e ações de vigilância que contribuem na tomada de decisões visando a melhoria do contexto sanitário da população (CORREIA; PADILHA; VASCONCELOS, 2014; OLIVEIRA *et al.*, 2015). Os dados presentes nos Sistemas de Informação em Saúde se tornam um eixo norteador para a identificação de padrões no adoecimento populacional, corroborando a possibilidade de intervenção. Ademais, a avaliação de um sistema de informação permite a averiguação das informações produzidas no âmbito do direcionamento das políticas públicas de saúde, identificação de população vulnerável, riscos coletivos e impactos de doenças na população e seus territórios (OLIVEIRA *et al.*, 2015).

Para que ocorra uma vigilância em saúde eficiente, é importante a utilização de dados representativos, de boa qualidade e fácil acesso. No entanto, os SIS sofrem com o mau preenchimento das fichas de notificação de doenças e agravos, que influencia diretamente na qualidade da informação e pode inviabilizar a construção de determinados indicadores para medir o impacto de uma condição de saúde na população (MARQUES; SIQUEIRA; PORTUGAL, 2020). É relatado em diversos estudos a baixa confiabilidade dos SIS devido à má qualidade dos dados, gerados pelo mau preenchimento das fichas de notificação ou por inconsistência da informação (SILVA *et al.*, 2014). Em países que apresentam condições de desigualdades socioeconômicas na população, a informação completa, precisa e oportuna é de extrema relevância para o planejamento, monitoramento, execução e avaliação das ações de saúde (ROMERO; CUNHA, 2007).

Em 2000, na 11ª Conferência Nacional de Saúde, enfatizou-se que a informação é um eixo importante dos serviços de saúde do Sistema Único de Saúde, visto que com o acesso a

informação de qualidade se torna possível avaliar e monitorar as desigualdades de acesso ao sistema de saúde, fatores de risco e condições socioeconômicas, culturais e ambientais que influenciam na saúde da população (ROMERO; CUNHA, 2006, 2007).

Definida a importância de uma informação completa, precisa e oportuna para o planejamento, monitoramento e controle de doenças. A avaliação das informações contidas nos Sistema de Informação em Saúde se torna uma atividade de relevância para o campo da saúde, dado que possibilitará a identificação de fragilidades e melhorias a serem feitas nestes sistemas. Conforme publicado pelo *Centers for Disease Control and Prevention*, a avaliação de um Sistema de Vigilância abrange atributos qualitativos e quantitativos, e, entre eles, a qualidade dos dados é um dos atributos mais recomendados para análise, principalmente quanto a sua confiabilidade, validade, cobertura e completude. (CORREIA; PADILHA; VASCONCELOS, 2014).

Devido à importância da dados de qualidade, diversos estudos avaliam a completude da informação dos SIS para a identificação de padrões que estão influenciando na má qualidade dos dados. Em análise realizada por Groseclose & Buckeridge, observou-se elevada frequência de dados incompletos em estudos que realizaram avaliação de qualidade dos dados (GROSECLOSE; BUCKERIDGE, 2017). No geral, a classificação de incompletude é divergente entre si, com propostas e pontos de cortes específicos de acordo com a metodologia da avaliação de dados proposta. O parâmetro mais utilizado em estudos que visam a avaliação de completude é o proposto por Romero e Cunha em 2006 (LIMA *et al.*, 2009).

## 1.8 PANORAMA DE ESTUDOS QUE AVALIARAM A QUALIDADE DA INFORMAÇÃO DOS RHC

A informação em saúde é um fator essencial na tomada de decisões. Dado que compõem o processo de *informação-decisão-ação*, dessa maneira, é de extrema necessidade entender a abrangência, potencialidades e fragilidades dos Sistemas de Informação em Saúde (SILVA *et al.*, 2014). Para tanto, é necessária uma informação de qualidade dentro dos SIS.

Os Registros Hospitalares de Câncer podem ser utilizados para pesquisas epidemiológicas, clínicas, de planejamento e avaliação das ações de controle do câncer. Todavia, poucos são os estudos que avaliam a qualidade dos dados advindos dos RHC.

Em estudo realizado por Pinto *et al.* (2012) onde avaliou-se a qualidade dos dados dos RHC no Brasil entre 2000 e 2006, observou-se boa completude e consistência dos dados, exceto para TNM, estadiamento, estado da doença ao final do primeiro tratamento e grau de

instrução (escolaridade). O estudo de Oliveira *et al.* (2021), que avaliou a completude dos dados e inconsistência nos RHC do Mato Grosso entre 2008 e 2014, encontrou resultados semelhantes ao de Pinto *et al.* (2012), com maior incompletude em TNM, escolaridade, estado da doença ao final do primeiro tratamento e ocupação.

No estudo de Grabois *et al.* (2014) analisou-se a completude da variável *ocupação* entre 2000 e 2008 nos RHC para o Brasil, observou-se que o percentual médio de ausência de informação sobre ocupação foi superior a 45%. Ocorreu melhoras durante o período analisado, contudo, não muito significativo.

Entre os resultados apresentados por D'Alessandro e Antoniazzi (2010) com relação à consistência das informações dos RHC de Minas Gerais, encontrou-se que 69% dos dados estavam sem informação do grau de instrução, 26% sem definição do estadiamento clínico do tumor, 50% sem informação de alcoolismo e 46% sem informação de tabagismo. As variáveis alcoolismo e tabagismo são itens opcionais de preenchimento, ou seja, o alto percentual de informações faltantes dessas variáveis pode ter sido acarretado por este motivo.

Os achados no estudo de Cardoso (2018) que avaliou a qualidade dos dados do RHC de Santa Maria, Rio Grande do Sul, nos anos de 2010 a 2015, corroboram os resultados encontrados por Pinto *et al.* (2012), Oliveira *et al.* (2021) e Grabois *et al.* (2014). Observando baixa completude para as variáveis alcoolismo, instrução, histórico familiar, ocupação, tabagismo, pTNM, estadiamento e TNM.

No estudo de Chaffin (2016) que analisou os RHC para o monitoramento do câncer infanto-juvenil foi observado baixa completude nas variáveis diagnóstico e tratamento anteriores, estado da doença no final do primeiro tratamento, origem do encaminhamento e razão para não tratar. O estudo de Lopes-Júnior *et al.* (2022) constatou completude ruim nas variáveis escolaridade, tabagismo, consumo de álcool, ocupação, histórico familiar de câncer e estadiamento clínico do tumor, em relação às inconsistências, as variáveis “estado civil”, “tabagismo”, “consumo de álcool” e “histórico familiar de câncer” apresentaram nas 4 variáveis o mesmo grau de 4,49% de inconsistência.

O estudo realizado por Oliveira *et al.* (2017) analisou 5 variáveis da ficha de registro do tumor, sendo elas: Razão para não tratar, Estado da doença ao final do 1º tratamento, Ocupação, Escolaridade e Estadiamento TNM. As maiores proporções de dados incompletos foram observadas nas variáveis Escolaridade (35,5%) e Estadiamento TNM (36,6%). No estudo de COSTA (2016) observou-se completude ruim nas variáveis raça/cor, escolaridade e ocupação principal, estadiamento, TNM e estado da doença ao final do primeiro tratamento. Notou-se também maiores inconsistências nas variáveis TNM e estadiamento.



No estudo de Keske (2021) notou-se completude muito ruim e ruim em TNM e ruim para ocupação, estadiamento, histórico familiar, consumo de bebida alcoólica e consumo de tabaco. A pior completude foi em 2017 para TNM (74,97%).

No que diz respeito às inconsistências nota-se uma frequência da verificação nas variáveis data da primeira consulta e data do diagnóstico, data da primeira consulta e data do início do tratamento, data do diagnóstico e data do início do tratamento, sexo e topografia, topografia e morfologia, topografia e lateralidade, TNM e grupamento de estadiamento e localização primária desconhecida e localização provável do tumor primário.

O estudo de Pinto *et al.* (2012) constatou a maior proporção de inconsistência em topografia e lateralidade (18,7%), seguido de data do diagnóstico e data do início do tratamento (7,8%). As menores inconsistências estavam em topografia e sexo (0,001%) e topografia e morfologia (0,02%). No estudo de Oliveira *et al.* (2021), as inconsistências entre topografia e lateralidade foram de 6,69%. Já no estudo de Cardoso (2018), as inconsistências entre as datas do diagnóstico e tratamento foram de 11,05% e entre lateralidade e topografia foi menor que 5% para as topografias mama (1,99%), de próstata (4,2%) e entre 5% a 10% para pulmão (6,3%).

No estudo de Chaffin (2016) notou-se que as inconsistências para razão para não tratar e 1º tratamento recebido no hospital foi de 73,3%. O estudo de Lopes-Júnior *et al.* (2022) observou inconstância de 4,49% entre lateralidade e topografia. O estudo de Costa (2016) foi observado inconsistência de 4,8% para TNM e estadiamento, já em topografia e sexo e topografia e a morfologia não foram encontradas inconsistências.

Em suma, os estudos que avaliaram os aspectos de qualidade dos dados de RHC observaram resultados semelhantes entre si, com boa completude e baixa inconsistência nas variáveis, com exceção de TNM, estadiamento, estado da doença ao final do primeiro tratamento e grau de instrução, ocupação, alcoolismo, tabagismo e histórico familiar para a completude, já na inconsistência foi observado que elas se concentravam principalmente entre topografia e lateralidade e data do diagnóstico e data do início do tratamento. A organização dos principais achados desses estudos está no Quadro 2.

**Quadro 2 - Panorama dos estudos que avaliaram a qualidade de dados para Registros Hospitalares de Câncer no Brasil**

Nome do estudo	Autor	Período do estudo	Local	Atributos avaliados	Parâmetro utilizado para avaliação da qualidade dos dados	Resultados
Completude e consistência dos dados dos registros hospitalares de câncer no Brasil	PINTO <i>et al.</i> , 2012	2000 a 2006	Brasil	Completude e consistência	Escore proposto por Romero e Cunha	Grande parte das variáveis obrigatórias apresentaram grau de completude regular e excelente, com exceção de grau de instrução, TNM, grupamento de estadiamento e estado da doença ao final do primeiro tratamento. observou-se boa consistência dos dados.
Registros Hospitalares de Câncer de Mato Grosso: análise da completude e da consistência	OLIVEIRA <i>et al.</i> , 2021	2008 a 2014	Mato Grosso	Completude e validade	Escore proposto por Romero e Cunha	33,33% das variáveis dos RHC/Cuiabá e 38,89% das variáveis do banco do interior apresentaram grau de completude “excelente” ou “bom”, as que tiveram grau de completude “ruim” ou “muito ruim” foram superiores, com 44,44% e 55,56%
Completude da Informação “Ocupação” nos Registros Hospitalares de Câncer do Brasil: Bases para a Vigilância do Câncer Relacionado ao Trabalho	GRABOIS <i>et al.</i> , 2014	2000 a 2008	Brasil <sup>1</sup>	Completude	-	Elevado percentual de ausência de informação sobre ocupação
Registros hospitalares de câncer de Minas Gerais: análise de consistência das bases de dados	D’ALESSANDRO; ANTONIAZZI, 2010	1998 a 2006	Minas Gerais	Consistência	Proporção de casos registrados em duplicidade e a proporção do preenchimento de informações em variáveis selecionadas	As variáveis apresentaram altos percentuais de preenchimento em procedência, local de nascimento, data de diagnóstico, diagnóstico e tratamentos anteriores, topografia (local do tumor primário), morfologia (tipo histológico). Desses registros, 69% estavam sem informação do grau de instrução, 26% sem definição do estadiamento clínico do tumor e 50 e 46% sem informação de alcoolismo e tabagismo. 2,7% dos casos apresentaram duplicidade ou registro duvidoso
Avaliação da qualidade das informações do Registro Hospitalar de Câncer do município de Santa Maria, Rio Grande do Sul, Brasil	CARDOSO, 2018	2010 a 2015	Santa Maria (Rio Grande do Sul)	Duplicidade, completude e consistência	Escore proposto por Romero e Cunha // Escore proposto por Abath	Apresentaram grau de completude “ruim” para as variáveis alcoolismo, instrução, histórico familiar, ocupação e tabagismo e “excelentes” para raça/cor, e cidade. As variáveis pTNM, estadiamento e TNM alcançaram graus de completude e consistência “muito ruim”. Em relação à consistência, observaram-se baixos percentuais de códigos inexistentes e de inconsistência entre as variáveis. Quanto à duplicidade, 8,2% do total de casos foi cadastrado mais de uma vez durante os anos analisados
Análise dos Registros Hospitalares de câncer para o monitoramento do câncer infantojuvenil	CHAFFIN, 2016	2000 a 2015	Brasil	abrangência, estabilidade temporal, completude <sup>2</sup> e adequação	Escore proposto por Romero e Cunha	A completude não se mostrou consistente, seis variáveis que descreviam aspectos clínicos e de identificação do paciente apresentaram excelente grau de completude (falha menor que 5%) exceto para a variável estado da doença no final do primeiro tratamento. As variáveis que dependiam da história do paciente antes de ingressar no hospital como origem do encaminhamento e razão para não tratar apresentaram pior completude

Completeness and Consistency of Epidemiological Variables from Hospital-Based Cancer Registries in a Brazilian State	LOPES-JÚNIOR <i>et al.</i> , 2022	2010 a 2017	Espírito Santo	Completeness e inconsistência	Escore proposto por Romero e Cunha	A completeness foi excelente nas variáveis idade, estado de residência, data do primeiro diagnóstico, diagnóstico e tratamento anteriores, localização primária do tumor, primeiro tratamento recebido no hospital, data do óbito e ruim nas variáveis escolaridade, tabagismo, consumo de álcool, ocupação, histórico familiar de câncer e estadiamento clínico do tumor. As variáveis “estado civil”, “tabagismo”, “consumo de álcool” e “histórico familiar de câncer” apresentaram 4,49% de inconsistência.
Registros Hospitalares de Câncer em Pernambuco: da Gestão ao Registro	OLIVEIRA <i>et al.</i> , 2017	2012 a 2015	Pernambuco	Documentos da gestão estadual e completeness <sup>2</sup>	-	As variáveis analisadas com relação a completeness foram razão para não tratar, estado da doença ao final do 1º tratamento, ocupação, escolaridade e estadiamento TNM. As maiores proporções de dados incompletos estavam nas variáveis Escolaridade (35,5%) e Estadiamento TNM (36,6%). A menor proporção de incompleteness foi na variável razão para não tratar (3,9%).
Hospital Federal de Bonsucesso: ações estratégicas de gestão da informação para o registro de câncer.	COSTA, 2016	2012	Rio de Janeiro	Processos adotados no RHC; informar os problemas de gestão da informação; avaliar a qualidade do banco de dados dos tumores <sup>2</sup>	Escore proposto por Romero e Cunha	A completeness foi ruim nas variáveis raça/cor, escolaridade e ocupação principal, estadiamento, TNM e estado da doença ao final do primeiro tratamento. Ocorreram inconsistências nas variáveis clínica do início do 1º tratamento, primeiro tratamento recebido no hospital, estado da doença ao final do 1º tratamento, data do início do tratamento, TNM e estadiamento. É importante frisar que as maiores inconsistências estavam em TNM (4,8%) e estadiamento (4,8%).
Análise e proposta de gestão para a qualidade de informação do registro hospitalar de câncer do Hospital Universitário de Santa Maria – RS	KESKE, 2021	2017 a 2019	Rio Grande do Sul	Completeness	Escore proposto por Romero e Cunha	A completeness muito ruim e ruim em TNM e ruim para ocupação, estadiamento, histórico familiar, consumo de bebida álcool e de consumo de tabaco. A pior completeness foi em 2017 para TNM (74,97%).

<sup>1</sup>Os estados do Amapá, Goiás, Maranhão, Rondônia, Roraima e São Paulo não tinham casos registrados no integrador na data da extração dos dados

<sup>2</sup> Levou-se em conta apenas os resultados para completeness ou inconsistência

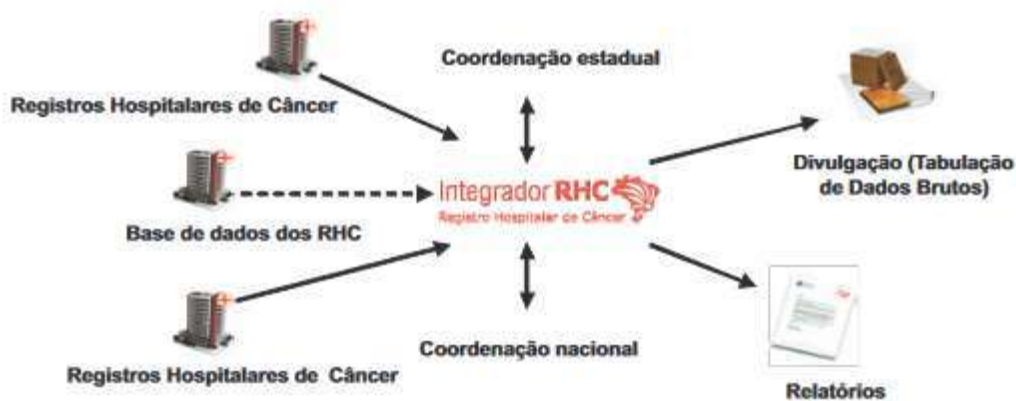
Fonte: Elaborado pelo autor.

## 1.9 FLUXO DA INFORMAÇÃO

As informações produzidas pelos Registros Hospitalares de Câncer são processadas dentro do Sistema de Informação de Registro Hospitalar de Câncer (SisRHC) e alimentam o IntegradorRHC que consolida e disponibiliza os dados hospitalares oncológicos provenientes dos RHC do Brasil (INSTITUTO NACIONAL DE CÂNCER, 2011). O SisRHC é um programa gratuito criado pelo Instituto Nacional do Câncer, e distribuído nacionalmente aos RHC desde 2002, que tem como função o processamento eletrônico de dados. O IntegradorRHC é um sistema *web* que foi criado e implantado pelo Instituto Nacional do Câncer, em 2007, para a consolidação dos dados advindos do SisRHC. (INSTITUTO NACIONAL DE CÂNCER, 2011; BRASIL, 2007).

A chegada da informação ao IntegradorRHC é feita em duas fases. Primeiro o Coordenador Estadual consolida todas as bases provenientes do seu estado. Logo após, a Coordenação Nacional faz a consolidação de todas as bases dos estados do Brasil (Figura 2). O envio da base de dados consolidada é feito pelos RHC no SisRHC respeitando os critérios, normas, prazos e temporalidade estabelecidos na portaria nº 741 estabelecida em dezembro de 2005 (INSTITUTO NACIONAL DE CÂNCER, 2011).

**Figura 4 - Fluxo da Informação no IntegradorRHC**



Fonte: Instituto Nacional de Câncer (2011).

A informação contida dentro do IntegradorRHC permite aos gestores estaduais e nacionais o entendimento do tratamento ofertado por suas unidades, além de que possibilita a compreensão da realidade enfrentada em cada estado e município que possui um RHC em funcionamento. Ademais, a facilidade de acesso a informação sobre a assistência ao câncer no

Brasil contribui para o conhecimento da distribuição da doença na população, guiando a criação de políticas de saúde e ações de controle e vigilância do câncer (INSTITUTO NACIONAL DE CÂNCER, 2011).

## 2 JUSTIFICATIVA

As transformações ocorridas no perfil demográfico, nutricional e epidemiológico da população influenciaram na carga global de Doenças Crônicas Não Transmissíveis no Brasil. Desse modo, doenças cardiovasculares, câncer, diabetes e doenças respiratórias crônicas se tornam relevantes fontes de estudo para melhoria das condições de saúde presentes na população (D’ALESSANDRO; ANTONIAZZI, 2010). Entre os cânceres que afetam a população, os de mama e os do colo do útero apresentam incidência e mortalidade significativas na população feminina do Brasil, estimando-se para o ano de 2020, 66.280 casos novos de câncer de mama e 16.590 casos novos de câncer do colo do útero (INSTITUTO NACIONAL DE CÂNCER, 2019).

Em decorrência da magnitude do câncer de mama e colo do útero na população brasileira, a informação de qualidade é de extrema necessidade para o planejamento de políticas de saúde e ações de controle e vigilância do câncer, que visem minimizar o impacto desses cânceres na população (CORREIA; PADILHA; VASCONCELOS, 2014; LIMA *et al.*, 2009; SILVA *et al.*, 2014). É preconizado por órgãos internacionais que todos os casos de câncer sejam devidamente registrados e acompanhados por um Registro Hospitalar de Câncer (CHAFFIN, 2016). Os registros disponíveis nos Registros Hospitalares de Câncer possibilitam a criação de estudos que observem a assistência prestada nas unidades, fatores de risco e tempo de sobrevivência. Entretanto, os Sistemas de Informação demonstram questões relacionadas à má qualidade da informação devido ao mau preenchimento das fichas de notificação, falta de tempo ou treinamento, que influenciam na criação de indicadores subestimados para o monitoramento das situações de saúde que afetam a população (MARQUES; SIQUEIRA; PORTUGAL, 2020).

Considerando o exposto, a avaliação da qualidade dos dados disponíveis nos Registros Hospitalares no Brasil é de grande valia para observar as fragilidades e potencialidades dos dados presentes no IntegradorRHC, visto que as informações obtidas através deste *website* são gratuitas e de fácil acesso para a população. Além disso, as informações encontradas podem servir para guiar a melhoria da qualidade dos dados desses registros. Vale ressaltar que estudos que avaliaram a qualidade de dados dos RHC são mais frequentes nas regiões Sul e Sudeste do país (CORREIA; PADILHA; VASCONCELOS, 2014). Dessa maneira, a avaliação da qualidade dos dados dos Registros Hospitalares de Câncer para o Brasil e por Grande Região é essencial para aprimorar a vigilância do câncer no cenário regional possibilitando a criação de ações que levem em conta a heterogeneidade entre as regiões do Brasil.

### **3 OBJETIVOS**

#### **3.1 OBJETIVO GERAL**

Avaliar a qualidade dos dados sobre câncer de mama e colo do útero na base de Registros Hospitalares de Câncer no Brasil e Grandes Regiões, 2008 a 2016.

#### **3.2 OBJETIVOS ESPECÍFICOS**

- Avaliar a completude dos registros de câncer de mama e colo do útero de variáveis selecionadas no Brasil e Grandes Regiões, 2008 a 2016;
- Avaliar a inconsistência dos registros de câncer de mama e colo do útero de variáveis selecionadas no Brasil e Grandes Regiões, 2008 a 2016;
- Disponibilizar uma rotina de análise para avaliação de completude e inconsistência dos dados de Registros Hospitalares de Câncer no Brasil e Grandes Regiões.

## 4 METODOLOGIA

### 4.1 DELINEAMENTO DO ESTUDO

Trata-se de um estudo descritivo de avaliação dos dados de Registros Hospitalares de Câncer no Brasil e Grandes Regiões entre 2008 e 2016. Os dados utilizados foram extraídos do IntegradorRHC que é uma base de dados não nominais de acesso público que consolida todos os casos de pacientes com diagnóstico confirmado de câncer que realizaram atendimento em uma unidade hospitalar habilitada pelo SUS nos Registros Hospitalares de Câncer (RHC) de todo Brasil.

### 4.2 POPULAÇÃO E ÁREA DE ESTUDO

A população de estudo é constituída por todos os casos analíticos e sem recidivas de câncer de mama (Classificação Internacional de Doenças – CID, Código C50) e colo do útero (Classificação Internacional de Doenças – CID, Código C53) notificados, sendo a 1ª Consulta realizada entre 1º de janeiro de 2008 e 31 de dezembro de 2016, nos Registros Hospitalares de Câncer do Brasil, excetuando-se o estado de São Paulo. Esse período foi escolhido considerando que os registros desses casos estão mais consolidados do que os notificados em anos mais recentes.

O estado de São Paulo foi excluído do estudo por não enviar todos os dados produzidos para o IntegradorRHC. A única forma de acesso aos microdados do estado de São Paulo é através do Tabnet da Fundação Oncocentro de São Paulo (FOSP), entretanto, a sua utilização inviabilizaria o contexto do estudo que é a avaliação dos dados de Registros Hospitalares de Câncer disponíveis no IntegradorRHC. A comparação entre os dados da FOSP e do SisRHC para casos analíticos de câncer pode ser visto na tabela 1.

É importante frisar que os casos analíticos são aqueles que chegam a uma instituição sem tratamento, com ou sem diagnóstico de câncer, e o tratamento passa a ser executado no próprio hospital. Dessa forma, é possível reduzir as duplicidades de registro devido ao fato de os indivíduos nunca terem recebido tratamento em nenhuma unidade hospitalar (FUNDAÇÃO ONCOCENTRO DE SÃO PAULO, c2022).



**Tabela 1 - Comparação entre os casos analíticos de câncer de mama e colo do útero registrados na FOSP e SisRHC, São Paulo, 2008 a 2016**

Ano <sup>1</sup>	Câncer de Mama				Câncer de Colo do Útero			
	FOSP <sup>2</sup>	% <sup>2</sup>	SisRHC	%	FOSP <sup>2</sup>	% <sup>2</sup>	SisRHC	%
2008	6.226	8,48	5.811	8,11	2.225	9,59	2.114	9,29
2009	6.664	9,08	6.636	9,26	2.386	10,28	2.378	10,45
2010	7.098	9,67	7.083	9,88	2.288	9,86	2.285	10,04
2011	7.958	10,84	7.951	11,09	2.507	10,8	2.504	11,00
2012	8.437	11,5	8.399	11,72	2.662	11,47	2.657	11,67
2013	9.069	12,36	9.016	12,58	2.787	12,01	2.778	12,2
2014	9.702	13,22	9.638	13,45	2.777	11,97	2.772	12,18
2015	9.202	12,54	9.099	12,69	2.866	12,35	2.856	12,55
2016	9.034	12,31	8.043	11,22	2.705	11,66	2.421	10,63
Total	73.390	100,00	71.676	100,00	23.203	100,00	22.765	100,00

<sup>1</sup> Ano de diagnóstico

<sup>2</sup>Fundação Oncocentro de São Paulo

Fonte: Elaborado pelo autor com dados da Fundação Oncocentro de São Paulo e IntegradorRHC.

#### 4.2.1 Brasil e Grandes Regiões

O local de estudo foi o Brasil e as Grandes Regiões (Norte, Nordeste, Sudeste, Sul e Centro-Oeste), compostas por 26 Unidades Federativas (UF) e o Distrito Federal (Figura 5). A região Norte abrange 7 estados, que são: Acre, Amapá, Amazonas, Pará, Rondônia, Roraima e Tocantins. Segundo as projeções populacionais do Instituto Brasileiro de Geografia e Estatística (IBGE) para o ano de 2020, a Região Norte possuía uma população de 18.672.591, destes 9.289.845 eram mulheres. Já a Região Nordeste, continha uma população de 57.374.243, onde 29.564.960 eram mulheres. Na região Sudeste, a população total é 89.012.240, e 45.649.897 são mulheres. A região Sul contém uma população de 30.192.315, sendo 15.394.432 do sexo feminino. A região Centro-Oeste apresenta a menor população, sendo 16.504.303 ao todo, e 8.328.869 de mulheres. O Brasil possuía uma população de 211.755.692, onde 103.527.689 (49%) eram do sexo masculino e 108.228.003 (51%) do sexo feminino (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2018).

**Figura 5 - Distribuição Territorial do Brasil por Grande Região e Unidade de Federação**



Fonte: Instituto Brasileiro de Geografia e Estatística, 2018.

#### 4.3 FONTE DE DADOS E VARIÁVEIS DE ESTUDO

Os dados referentes aos Registros Hospitalares de Câncer foram obtidos através do *website* IntegradorRHC<sup>1</sup> do Instituto Nacional do Câncer no dia 4 de setembro de 2022.

Para a seleção das variáveis de estudo optou-se por aquelas que possuíssem o preenchimento obrigatório (exceto Ocupação e pTNM) e disponibilizadas dentro da base de dados, visto que é uma conduta padronizada em todos os Registros Hospitalares de Câncer no Brasil. As variáveis foram selecionadas através da ficha de registro do tumor (disponível em Anexo) e estão dispostas no quadro 3.

<sup>1</sup> Disponível em: <https://irhc.inca.gov.br/RHCNet/visualizaTabNetExterno.action>

**Quadro 3 - Lista de variáveis selecionadas para o estudo**

<b>Blocos da ficha de registro de câncer</b>	<b>Nome da Variável</b>
<b>Identificação do Paciente</b>	Sexo
	Data de nascimento
	Idade
	Raça/Cor da pele
	Escolaridade
	UF de procedência
<b>Caracterização do diagnóstico</b>	Data da primeira consulta
	Data do primeiro diagnóstico
	Diagnóstico e tratamento anteriores
	Base mais importante para o diagnóstico do tumor
<b>Caracterização do tumor</b>	Tipo histológico
	Estadiamento clínico do tumor (TNM)
	TNM
<b>Caracterização do primeiro tratamento</b>	Data do início do primeiro tratamento
	Principal razão para a não realização do tratamento
	Primeiro tratamento recebido
	Estado da doença ao final do primeiro tratamento
	Data do óbito

Fonte: Elaborado pelo autor.

#### 4.4 ANÁLISE DE DADOS

Foi realizada a manipulação e análise das bases de dados no *software* livre R versão 4.0.5 e utilizaram-se os seguintes pacotes: *rio*, *dplyr*, *kableExtra*, *reshape2* e *tidyr*. Para identificar as Grandes Regiões do Brasil (Norte, Nordeste, Centro-Oeste, Sudeste e Sul) foram agrupados os estados de residência correspondentes a essas regiões para a criação de uma nova variável que tivesse como resultado uma das cinco macrorregiões do Brasil. Ademais, foram calculados indicadores de completude e inconsistência para a realização da avaliação da qualidade dos dados presentes no IntegradorRHC baseando-se no guia do *Center for Disease and Control* para a avaliação de Sistemas de Vigilância em Saúde Pública.

Para os resultados da análise, será realizado a criação de 12 tabelas referentes à proporção de completude por ano, Brasil e Grandes Regiões, sendo uma tabela sobre o Brasil e cinco tabelas sobre cada uma das Grandes Regiões, tanto para câncer de mama quanto para câncer do colo do útero. Já para a análise de inconsistência, foram criadas duas tabelas

contendo a proporção de inconsistências por ano, Brasil e Grande Regiões para câncer de mama e câncer do colo do útero.

A seguir serão detalhadas as análises de completude e inconsistência abordadas no estudo. Todas as análises foram realizadas para o local (Brasil e Grandes Regiões) e período do estudo, 2008 a 2016.

## Completude

O cálculo de completude foi realizado a partir da proporção de dados completos preenchidos subtraindo-se a proporção de dados com campos em branco, ignorado ou preenchimento divergente do solicitado no dicionário de dados, ou seja, 100% - a proporção de incompletude. Para avaliação de completude, utilizou-se o escore proposto por Romero & Cunha (Tabela 2) para classificar as variáveis selecionadas no estudo (Quadro 3).

**Tabela 2 - Escore de Romero & Cunha, correspondência com a completude e classificação**

Score de Romero & Cunha		
Proporção de incompletude	Correspondência com a completude	Classificação
<5%	100% a 96%	Excelente
5% a 10%	95% a 90%	Bom
10% a 20%	90% a 80%	Regular
20% a 50%	80% a 50%	Ruim
50%+	50% a 0	Muito Ruim

Fonte: Romero e Cunha (2006).

## Inconsistência

A proporção de registros com inconsistências foi calculada para 9 das 18 variáveis selecionadas. A seleção das variáveis para análise de inconsistência baseou-se em uma busca bibliográfica de estudos cujos temas eram correlatos ao presente estudo, uma vez que não há nos dicionários e manuais do RHC uma forma específica de análise de inconsistências (PINTO *et al.*, 2012; OLIVEIRA *et al.*, 2021; CARDOSO, 2018; CHAFFIN, 2016; LOPES-JÚNIOR *et al.*, 2022; COSTA, 2016). Detalha-se mais profundamente os estudos utilizados no Quadro 1.

Para a classificação da inconsistência entre Localização primária & Tipo histológico utilizou-se a Classificação Internacional de Doenças para Oncologia (CID-O), 3ª edição. No caso de Estadiamento clínico do tumor (TNM) e Estadiamento, buscou-se a Classificação

TNM da União Internacional Contra o Câncer (UICC), 8ª edição. A avaliação das inconsistências baseou-se no escore proposto por Romero e Cunha em 2006.

#### Quadro 4 - Variáveis selecionadas para análise de inconsistências

Análise de Inconsistência
Data da primeira consulta < Data do diagnóstico
Data da primeira consulta > Data do início do tratamento
Data do diagnóstico > Data do início do tratamento
Sexo & Localização primária*
Localização primária & Tipo histológico*
TNM & Estadiamento*

- A inconsistência entre Sexo & Localização primária é referente ao sexo biológico do indivíduo e a localização do tumor, por exemplo, um homem cisgênero com câncer do colo do útero.
- A inconsistência entre Localização primária e Tipo histológico refere-se a um tumor na mama ou colo útero com tipo histológico não correspondente a esses tumores.
- A inconsistência entre TNM & Estadiamento é devido ao preenchimento das categorias de TNM para a descrição da extensão anatômica da doença (T – Extensão do tumor primário, N – Ausência ou presença e a extensão de metástase em linfonodos regionais e M – Ausência ou presença de metástase a distância), e a partir do preenchimento do TNM, é definido um estadiamento clínico do tumor, então, TNM não existentes ou não correspondentes ao estadiamento são considerados inconsistências.
- Fonte: Romero e Cunha (2006).

A rotina de autoria própria, na linguagem R, que auxilia na manipulação dos dados e a análise de completude e inconsistência dos casos de câncer notificados nos Registros Hospitalares de Câncer no Brasil (Objetivo 3) encontra-se no Apêndice e a rotina completa por Brasil e Grandes regiões está disponível no GitHub<sup>2</sup>.

#### 4.5 DISPOSIÇÕES ÉTICAS

Os dados secundários utilizados neste trabalho são públicos, de acesso irrestrito e foram analisados de forma agregada, sem a identificação de indivíduos. Desta forma, não foi necessária a apreciação do sistema CEP/CONEP para o uso dos mesmos.

---

<sup>2</sup> Disponível em: [https://github.com/natpaiva/qualidadedados\\_rhc](https://github.com/natpaiva/qualidadedados_rhc)

## 5 RESULTADOS

A organização dos resultados foi iniciada pela análise de completude, por tipo de câncer (mama e colo do útero) e local de residência (Brasil e regiões Norte, Nordeste, Centro-Oeste, Sudeste e Sul), e posteriormente a análise de inconsistências, por tipo de câncer e local de residência. Isto é,

- Completude
  - Câncer de mama
    - ✓ Brasil
    - ✓ Grandes regiões
  - Câncer de colo do útero
    - ✓ Brasil
    - ✓ Grandes regiões
- Inconsistência
  - Câncer de mama
    - ✓ Brasil e Grandes regiões
  - Câncer de colo do útero
    - ✓ Brasil e Grandes regiões
- Script para análise da qualidade dos dados

### 5.1 ANÁLISE DE COMPLETUDE

A apresentação da análise de completude baseou-se nas divisões da ficha de registro do tumor. Dessa maneira, os resultados obtidos no estudo serão apresentados seguindo a ordem dos blocos de Identificação do paciente, Caracterização do diagnóstico, Caracterização do tumor e Caracterização do 1º tratamento.

#### 5.1.1 Câncer de Mama

No Brasil, com exceção de São Paulo, entre 2008 e 2016 foram notificados no IntegradorRHC 195.899 casos novos de câncer de mama.

Com relação à completude por Grande Região, entre os anos de 2008 e 2016, foram notificados pelo IntegradorRHC, 9.810 (5,00%) casos novos de câncer de mama na Região

Norte, 66.343 (33,87%) na Região Nordeste, 8.197 (4,18%) na Região Centro-Oeste, 52.943 (27,03%) na Região Sudeste, sem levar em conta São Paulo, e 57.040 (29,11%) na Região Sul.

A contabilização dos casos de câncer de mama por Grande Região não é o mesmo valor que o observado no Brasil, dado que por existirem casos que foram notificados sem o preenchimento do Estado de residência não foi possível alocá-los em nenhuma região (n = 1.566; 0,8%).

#### • Câncer de Mama – Brasil:

Com relação às variáveis do bloco de Identificação do paciente, a variável Sexo apresentou completude Excelente em todo o período de análise, notou-se também uma melhora no preenchimento dessa variável, visto que nos anos de 2008 e 2009 apresentou completude em torno de 99,9%, já de 2010 a 2016, o preenchimento alcançou 100%. A variável Raça/Cor da pele oscilou entre completude Regular (2008 e 2009) e Bom (2010 a 2016), apresentando pequenos aumentos no preenchimento, variando de 91,54% a 92,63%. Na variável Escolaridade entre 2008 e 2013 a completude foi classificada como Ruim, variando de 68,12% a 79,61%, já em 2014 se encontrava com preenchimento Regular (80,52%), decrescendo em 2015 (79,05%) e 2016 (78,23%) atingindo completude Ruim. A variável Estado de residência apresentou completude Excelente em todo o período observado, oscilando entre 98% e 99%. Na variável idade observou-se uma completude Excelente entre 2008 e 2016 com 100% do preenchimento na maioria dos anos, com exceção de 2011 (99,98%), 2013 (99,97%) e 2014 (99,99%) (Tabela 3)

No bloco de Caracterização do diagnóstico, a variável Data da primeira consulta obteve completude Excelente com 100% dos registros preenchidos em todos os anos. A variável Data de diagnóstico mostrou completude Excelente em todos os anos variando de 96% a 98%. Na variável diagnóstico e tratamento anteriores a completude foi Excelente em todos os anos, intercalando-se entre 97% a 99%. A variável Base mais importante para o diagnóstico do tumor demonstrou completude Excelente de 97% a 99% (Tabela 3)

No bloco de caracterização do tumor, a variável tipo histológico possui completude excelente de 99% dos registros preenchidos. A variável Estadiamento clínico do tumor (TNM) demonstrou completude Ruim em todo o período analisado, com destaque ao ano de 2015 onde atingiu 74,48% sendo o menor percentual dessa variável no período. Na variável TNM observou-se completude Ruim em todos os anos, e no ano de 2012 (64,24%) atingiu o menor percentual de preenchimento dessa variável (Tabela 3).

No bloco de Caracterização do 1º tratamento, a completude da variável Data de início do tratamento variou entre Bom e Excelente, nos anos de 2008 (95,45%) a 2010 (95,45%) atingiu grau de preenchimento Bom, já de 2011 (96,13%) a 2016 (96,79%) a completude foi Excelente. Em relação a variável Principal razão para a não realização do tratamento antineoplásico a completude foi Regular entre os anos de 2008 (89,41%) a 2010 (80,30%), Bom entre 2011 (91,98%) e 2012 (95,65%) e Excelente entre 2013 (97,01%) a 2016 (96,95%). A variável Primeiro tratamento recebido foi Excelente em todos os anos observados com percentual de preenchimento dos dados de 98%. Notou-se completude Ruim na variável Estado da doença ao final do tratamento em todo o período com destaque para 2008 (69,10%) com o menor percentual observado na variável. A variável Data do óbito demonstrou completude Excelente em todos os anos, atingindo 100% nos anos de 2008 a 2015, e caindo para 99,94% em 2016 (Tabela 3).



**Tabela 3 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Brasil, 2008-2016**

Bloco de variáveis	Câncer de Mama								
	2008	2009	2010	2011	2012	2013	2014	2015	2016
<b>Identificação do paciente</b>									
Sexo	99,96	99,98	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Raça/Cor da pele	88,34	89,18	91,54	90,27	90,76	90,93	92,22	92,43	92,63
Escolaridade	68,12	68,61	70,65	72,7	78,25	79,61	80,52	79,05	78,23
Estado de residência	99,66	99,77	99,53	99,01	98,77	98,71	99,05	98,91	99,58
Idade	100,00	100,00	100,00	99,98	100	99,97	99,99	100,00	100,00
<b>Caracterização do diagnóstico</b>									
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Data de diagnóstico	98,27	97,56	96,89	97,13	97,71	97,77	98,19	98,02	98,07
Diagnóstico e tratamento anteriores	97,83	98,31	98,28	97,44	97,9	98,03	98,57	98,93	99,3
Base mais importante para o diagnóstico do tumor	97,94	99,09	99,2	99,2	99,15	98,86	99,2	99,4	99,53
<b>Caracterização do tumor</b>									
Tipo histológico	99,58	99,78	99,93	99,95	99,95	99,92	99,97	99,99	99,89
Estadiamento clínico do tumor (TNM)	78,85	79,08	77,37	77,25	75,6	76,15	76	74,48	75,3
TNM	66,87	69,41	67,29	65,63	64,24	65,14	66,12	65,67	66,83
<b>Caracterização do 1º tratamento</b>									
Data de início do tratamento	95,45	95,66	95,45	96,13	96,2	96,04	96,41	96,56	96,79
Principal razão para a não realização do tratamento antineoplásico no hospital	89,41	85,93	80,83	91,98	95,65	97,01	96,73	96,47	96,95
Primeiro tratamento recebido no hospital	98,32	98,28	98,39	98,07	98,29	98,1	98,32	98,78	98,49
Estado da doença ao final do tratamento	69,1	69,47	72,55	74,4	73,1	71,18	72,3	72,84	73,74
Data de óbito	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	99,94

Excelente
Bom
Regular
Ruim
Muito Ruim

Fonte: Elaborado pelo autor com dados do IntegradorRHC.

### • Câncer de Mama - Região Norte

No bloco de Identificação do paciente, a variável Sexo apresentou completude Excelente de 100% em todos os anos analisados. A completude na variável Raça/Cor da pele oscilou entre Ruim e Regular (2012 e 2016), os piores percentuais de completude nessa variável foram observados nos anos de 2008 (60,92%) e 2009 (63,56%). Com relação a variável Escolaridade, notou-se uma variação entre completude de grau Bom, Regular e Ruim, destacam-se os anos de 2008 (91,81%) e 2015 (91,24%) com completude de grau Bom e 2009 (79,42%), 2010 (79,39%) e 2013 (78,51%) com completude Ruim. A variável Idade apresentou completude Excelente de 99% a 100% em todos os anos (Tabela 4).

Sobre o bloco de Caracterização do diagnóstico, a variável Data da primeira consulta demonstrou completude Excelente de 100% em todo o período. A variável Data de diagnóstico seguiu com Completude Excelente em todos os anos oscilando entre 99% e 100%. Na variável Diagnóstico e tratamento anteriores existe uma variação nos anos entre a completude de grau Bom e Excelente, nos anos de 2011 (91,55%), 2012 (94,68%) e 2013 (94,85%) a completude foi de grau Bom, o restante dos anos seguiu com uma completude Excelente entre 96% e 99%. Já a variável Base mais importante para o diagnóstico do tumor dispôs de completude Excelente (99% a 100%) em todos os anos (Tabela 4).

O preenchimento do bloco de Caracterização do tumor apresentou variação entre Excelente, Bom, Regular e Ruim nas variáveis. A completude da variável Tipo histológico foi Excelente de 100% em todos os anos. A variável Estadiamento clínico do tumor (TNM) demonstrou grau de completude Ruim para todo o período, estando os piores percentuais nos anos de 2012 (59,70%) e 2016 (62,58%). Já na variável TNM, a completude foi classificada como Ruim em todos os anos, destacando-se 2011 (53,94%) e 2012 (60,17%) com as piores porcentagens (Tabela 4).

Entre as variáveis do bloco de Caracterização do 1º tratamento, a variável Data de início do tratamento apresentou completude de grau Bom em 2008 (93,80%) e no resto do período a completude foi Excelente de 96% a 99%. Na variável Principal razão para a não realização do tratamento antineoplásico no hospital o grau de completude foi crescente entre 2008 e 2010, passando de Regular em 2008 (88,34%) para Bom em 2009 (94,22%), e a partir de 2011 a completude foi Excelente de 97% a 99% no restante dos anos. Na variável Primeiro tratamento recebido no hospital nota-se completude foi Excelente de 99% a 100% em todos os anos. A variável Estado da doença ao final do tratamento mostrou estabilidade no seu

preenchimento com grau de completude Bom (90% a 94%) em todo o período. Já a variável Data de óbito apresentou completude Excelente de 100% em todos os anos (Tabela 4).

**Tabela 4 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Região Norte, 2008-2016**

Bloco de variáveis	Câncer de Mama									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Identificação do paciente</b>										
Sexo	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Raça/Cor da pele	60,92	63,56	75,88	79,28	80,23	79,57	70,28	78,00	86,37	
Escolaridade	91,81	79,43	79,39	89,00	80,42	78,51	89,33	91,24	89,40	
Estado de residência	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Idade	100,00	100,00	100,00	100,00	100,00	99,82	99,92	100,00	100,00	
<b>Caracterização do diagnóstico</b>										
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Data de diagnóstico	99,50	99,77	99,89	99,31	100,00	100,00	99,84	99,71	99,88	
Diagnóstico e tratamento anteriores	96,53	98,05	96,15	91,55	94,68	94,85	99,43	98,31	99,20	
Base mais importante para o diagnóstico do tumor	99,88	100,00	100,00	99,88	99,81	99,64	99,67	99,56	99,94	
<b>Caracterização do tumor</b>										
Tipo histológico	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Estadiamento clínico do tumor (TNM)	76,92	78,39	76,67	64,93	59,70	69,98	65,64	64,68	62,58	
TNM	70,35	71,38	60,36	53,94	60,17	69,89	66,69	66,23	61,34	
<b>Caracterização do 1º tratamento</b>										
Data de início do tratamento	93,80	96,32	98,87	98,50	97,91	99,56	99,67	99,34	98,58	
Principal razão para a não realização do tratamento antineoplásico no hospital	88,34	90,69	94,22	99,42	99,24	99,20	99,67	97,79	98,64	
Primeiro tratamento recebido no hospital	100,00	100,00	100,00	100,00	99,52	99,47	99,67	99,71	99,88	
Estado da doença ao final do tratamento	90,20	91,15	94,56	94,79	90,02	91,39	93,57	91,17	94,94	
Data de óbito	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	

	Excelente
	Bom
	Regular
	Ruim
	Muito Ruim

Fonte: Elaborado pelo autor com dados do IntegradorRHC.

### • Câncer de Mama - Região Nordeste

No preenchimento do bloco de Identificação do paciente, a variável sexo apresentou completude Excelente de 99% a 100% em todo o período. Já na variável Raça/Cor da pele, o grau de completude é Regular entre 85% a 88% nos anos de 2008 a 2013, em 2014 até 2016 o grau de completude é Bom de 90 a 92%. A variável escolaridade teve completude Ruim em todos os anos, destacando os anos de 2008 (54,26%) e 2009 (60,71%) com os piores percentuais. Na variável Idade a completude foi Excelente de 99% a 100% em todos os anos (Tabela 5).

No bloco de Caracterização do diagnóstico, a variável Data da primeira consulta apresentou completude Excelente de 100% em todos os anos. Na Data de diagnóstico a completude foi Excelente entre 97% a 98% em todo o período. A variável Diagnóstico e tratamento anteriores demonstrou completude Excelente de 99% em todos os anos. A variável Base mais importante para o diagnóstico do tumor teve completude Excelente de 99% em todo o estudo (Tabela 5).

No bloco de Caracterização do Tumor, a variável Tipo histológico obteve completude Excelente de 100% em todos os anos. No Estadiamento clínico do tumor (TNM) observou-se completude Ruim em todos os anos, com os piores percentuais em 2015 (72,62%) e 2016 (73,37%). O TNM seguiu o padrão observado no Estadiamento clínico com completude Ruim em todo o período, entretanto os piores percentuais se localizavam nos anos de 2013 (61,08%) e 2015 (62,48%) (Tabela 5).

No bloco de Caracterização do 1º tratamento, a variável Data de início do tratamento teve completude de grau Bom de 94% a 95% nos anos de 2008 a 2010, já a partir de 2011 até 2016 a completude foi Excelente de 96% a 97% em todos os anos. A variável Principal razão para a não realização do tratamento antineoplásico no hospital variou entre completude de grau Regular, Bom e Excelente, em 2008 (91,62%) a completude foi de grau Bom, já em 2009 (88,76%) a completude foi Regular, em 2010 e 2011 a completude foi de grau Bom de 91% a 94%, de 2012 a 2016 a completude foi Excelente de 98%. Na variável Primeiro tratamento recebido no hospital a completude foi Excelente de 99% em todos os anos. A variável Estado da doença ao final do tratamento teve grau de preenchimento Ruim em todos os anos, destacando o ano de 2015 (65,95%) com o pior percentual. A data de óbito apresentou preenchimento Excelente de 99% a 100% em todos os anos (Tabela 5).

**Tabela 5 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Região Nordeste, 2008-2016**

Bloco de variáveis	Câncer de Mama									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Identificação do paciente</b>										
Sexo	99,94	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Raça/Cor da pele	85,72	86,36	88,85	88,45	86,89	87,99	91,30	92,30	90,86	
Escolaridade	54,26	60,71	71,63	74,95	77,96	78,82	79,19	79,20	74,08	
Estado de residência	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Idade	100,00	100,00	100,00	99,99	100,00	99,98	99,99	100,00	100,00	
<b>Caracterização do diagnóstico</b>										
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Data de diagnóstico	98,14	98,06	97,91	97,32	97,91	98,52	98,66	98,11	98,80	
Diagnóstico e tratamento anteriores	99,84	99,78	99,69	99,30	99,40	99,56	99,51	99,43	99,51	
Base mais importante para o diagnóstico do tumor	99,70	99,60	99,71	99,62	99,52	99,02	98,96	99,22	99,74	
<b>Caracterização do tumor</b>										
Tipo histológico	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Estadiamento clínico do tumor (TNM)	77,14	76,80	75,84	79,96	76,26	76,46	74,61	72,62	73,37	
TNM	66,81	65,77	65,74	67,17	64,13	61,08	63,15	62,48	62,72	
<b>Caracterização do 1º tratamento</b>										
Data de início do tratamento	95,13	95,26	94,66	97,01	97,11	97,13	97,45	96,64	97,08	
Principal razão para a não realização do tratamento antineoplásico no hospital	91,62	88,76	91,25	94,35	98,72	98,80	98,32	98,67	98,53	
Primeiro tratamento recebido no hospital	99,70	99,52	99,89	99,83	99,59	99,77	99,50	99,87	99,79	
Estado da doença ao final do tratamento	68,80	69,58	72,27	77,49	74,46	68,98	67,22	65,95	68,63	
Data de óbito	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	99,84	



Fonte: Elaborado pelo autor com dados do IntegradorRHC.

### • Câncer de Mama - Região Centro-Oeste

No bloco de Identificação do paciente, a variável sexo demonstrou completude Excelente de 99% a 100% em todos os anos. A variável Raça/Cor da pele apresentou completude Regular de 84% a 89% entre 2008 a 2011, já em 2012 (90,01%) o grau de completude foi Bom, entre 2013 a 2015 a completude foi Regular de 80% a 89% e em 2016 (73,58%) a completude foi Ruim. Na variável Escolaridade a completude oscilou em alguns momentos entre Ruim e Muito ruim, 2008 a 2010 a completude estava Ruim e caindo de 65% para 50%, em 2011 (45,14%) a completude foi Muito ruim, em 2012 observou-se uma subida para 59,48%, de 2013 a 2015 a completude Ruim caiu de 58% para 50% e em 2016 (46,72%) a completude estava Muito ruim. Na variável Idade a completude foi Excelente de 99% a 100% em todos os anos (Tabela 6).

No bloco de Caracterização do diagnóstico, a variável Data da primeira consulta apresentou completude Excelente de 100% em todos os anos. A variável Data de diagnóstico demonstrou completude Excelente de 97% a 100% em todo o período. Na variável Diagnóstico e tratamento anteriores a completude seguiu Excelente de 97% a 100% entre 2008 a 2013, em 2014 (95,30%) o grau de completude foi Bom, de 2015 a 2016 a completude foi Excelente de 98%. A variável Base mais importante para o diagnóstico do tumor apresentou completude Excelente de 95% a 99% em todos os anos (Tabela 6).

Com relação ao bloco de Caracterização do tumor, a completude foi Excelente de 100% na variável Tipo histológico em todos os anos. A completude na variável Estadiamento clínico do tumor (TNM) foi ruim em todo o período, com o pior percentual no ano de 2011 (50,55%). Já a variável TNM demonstrou completude Ruim entre 2008 (53,16%) e 2009 (50,24%), entre 2010 a 2013 a completude foi Muito Ruim variando de 43% a 35%, de 2014 a 2016 a completude foi Ruim entre 69% e 54% (Tabela 6).

No bloco de Caracterização do 1º tratamento, a variável Data de início do tratamento apresentou completude Excelente de 97% a 99% em todo o período. A variável Principal razão para a não realização do tratamento antineoplásico no hospital oscilou entre completude de grau Bom e Excelente, nos anos de 2008 a 2009 a completude foi Excelente de 97% e 96%, em 2010 e 2011 a completude apresentou grau Bom de 93% e 95%, em 2012 (96,96%) a completude foi Excelente, de 2013 a 2015 o grau de completude foi Bom de 95% a 92% e em 2016 (96,46%) a completude foi Excelente. A completude na variável Primeiro tratamento recebido no hospital foi Excelente de 99% e 100% nos anos de 2008 a 2012, em 2013 (95,35%) e 2014 (94,98%) o grau de completude foi Bom, de 2015 a 2016 a completude foi

Excelente de 98% e 99%. Na variável Estado da doença ao final do tratamento a completude oscilou entre Regular e Ruim, de 2008 a 2011 a completude foi Ruim de 58% a 67%, em 2012 (80,61%) a completude foi Regular, entre 2013 e 2014 a completude foi Ruim de 75% a 71%, já em 2015 (83,31%) a completude foi Regular e em 2016 (72,19%) a completude foi Ruim. A variável Data de óbito demonstra completude Excelente de 100% em todo o período (Tabela 6).



**Tabela 6 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Região Centro-Oeste, 2008-2016**

Bloco de variáveis	Câncer de Mama								
	2008	2009	2010	2011	2012	2013	2014	2015	2016
<b>Identificação do paciente</b>									
Sexo	99,87	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Raça/Cor da pele	89,14	88,53	83,20	84,13	90,01	87,53	89,10	80,10	73,58
Escolaridade	65,66	58,51	50,50	45,14	59,48	58,28	55,02	50,76	46,72
Estado de residência	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Idade	100,00	100,00	100,00	100,00	100,00	99,89	100,00	100,00	100,00
<b>Caracterização do diagnóstico</b>									
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Data de diagnóstico	99,49	99,29	99,89	99,38	99,71	100,00	97,86	98,82	99,31
Diagnóstico e tratamento anteriores	99,87	100,00	98,88	98,03	97,11	95,69	95,30	98,06	98,70
Base mais importante para o diagnóstico do tumor	99,75	99,88	99,33	98,65	97,54	95,69	96,58	99,33	99,65
<b>Caracterização do tumor</b>									
Tipo histológico	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Estadiamento clínico do tumor (TNM)	63,64	62,29	58,68	50,55	53,84	70,18	77,99	74,87	65,20
TNM	53,16	50,24	43,90	42,19	35,75	54,54	55,24	69,14	63,39
<b>Caracterização do 1º tratamento</b>									
Data de início do tratamento	99,12	99,05	98,66	98,15	98,70	98,98	97,22	97,72	97,75
Principal razão para a não realização do tratamento antineoplásico no hospital	97,10	96,81	93,95	95,57	96,96	95,12	94,02	92,50	96,46
Primeiro tratamento recebido no hospital	100,00	100,00	99,89	100,00	99,57	95,35	94,98	99,24	98,45
Estado da doença ao final do tratamento	58,96	63,95	67,75	67,90	80,61	75,06	71,79	83,31	72,19
Data de óbito	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

	Excelente
	Bom
	Regular
	Ruim
	Muito Ruim

Fonte: Elaborado pelo autor com dados do IntegradorRHC.

- **Câncer de Mama - Região Sudeste**

No bloco de Identificação do paciente, a variável sexo apresentou completude Excelente de 99% a 100% durante todo o período. Na variável Raça/Cor da pele, o grau de completude foi Bom de 90% a 94% em todos os anos analisados. A variável Escolaridade teve completude Ruim de 67% a 77% em todo o período. A variável Idade apresentou completude Excelente de 99% a 100% em todos os anos observados (Tabela 7).

No bloco de Caracterização do diagnóstico, a variável Data da primeira consulta obteve completude Excelente de 100% em todos os anos. A variável Data de diagnóstico demonstrou completude Excelente de 97% a 99% no período analisado. Na variável Diagnóstico e tratamento anteriores, a completude foi Excelente de 99% em todos os anos. A variável Base mais importante para o diagnóstico do tumor apresentou completude Excelente de 99% em todos os anos (Tabela 7).

No bloco de Caracterização do tumor, a variável Tipo histológico apresentou completude Excelente de 99% a 100% em todos os anos. Na variável Estadiamento clínico do tumor (TNM), a completude foi Regular de 86% a 83% em 2008 a 2009, Ruim em 2010 (79,58%), Regular em 2011 (81,99%) e Ruim de 2012 (78,59%) a 2016 (79,22%). Já a variável TNM apresentou completude Ruim de 68% a 76% em todo o período (Tabela 7).

No bloco de Caracterização do 1º tratamento, a variável Data de início do tratamento demonstrou completude Excelente de 97% a 98% entre 2008 a 2016. A variável Principal razão para a não realização do tratamento antineoplásico no hospital apresentou completude Regular em 2008 (89,60%) e 2009 (88,38%), já em 2010 (79,00%) a completude foi Ruim e de 2011 a 2016 a completude foi Excelente de 97% a 99%. A completude foi Excelente de 99% na variável Primeiro tratamento recebido no hospital em todos os anos. Na variável Estado da doença ao final do tratamento, a completude foi Ruim de 72% a 79% em todo o período. A variável Data de óbito apresentou completude Excelente de 100% (Tabela 7).

**Tabela 7 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Região Sudeste, 2008-2016**

Bloco de variáveis	Câncer de Mama									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Identificação do paciente</b>										
Sexo	99,98	99,98	100,00	100,00	100,00	100,00	99,99	100,00	100,00	
Raça/Cor da pele	90,32	93,52	94,34	92,23	91,68	90,46	93,12	94,32	93,89	
Escolaridade	74,89	74,07	67,14	67,72	77,39	76,49	77,41	76,28	76,82	
Estado de residência	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Idade	100,00	100,00	100,00	100,00	100,00	99,98	100,00	100,00	100,00	
<b>Caracterização do diagnóstico</b>										
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Data de diagnóstico	99,60	99,43	97,87	99,40	99,33	99,77	99,65	99,48	99,52	
Diagnóstico e tratamento anteriores	99,92	99,39	99,53	99,76	99,18	99,80	99,66	99,18	99,90	
Base mais importante para o diagnóstico do tumor	99,74	99,75	99,30	99,76	99,64	99,73	99,63	99,76	99,73	
<b>Caracterização do tumor</b>										
Tipo histológico	99,88	99,94	99,98	100,00	99,94	99,95	99,99	100,00	99,99	
Estadiamento clínico do tumor (TNM)	86,68	83,44	79,58	81,99	78,59	77,13	77,21	78,13	79,22	
TNM	76,40	78,42	73,28	72,07	70,08	69,82	69,21	68,16	70,41	
<b>Caracterização do 1º tratamento</b>										
Data de início do tratamento	98,45	97,11	97,30	97,69	98,76	98,25	98,10	98,25	98,54	
Principal razão para a não realização do tratamento antineoplásico no hospital	89,60	88,38	79,00	99,42	99,14	99,22	97,74	98,15	98,42	
Primeiro tratamento recebido no hospital	99,98	99,94	99,79	99,93	99,89	99,80	99,87	99,84	99,96	
Estado da doença ao final do tratamento	79,86	76,14	77,58	78,89	72,86	76,01	75,70	78,88	79,69	
Data de óbito	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	



Fonte: Elaborado pelo autor com dados do IntegradorRHC.

### • Câncer de Mama - Região Sul

No bloco de Identificação do paciente, a variável Sexo apresentou completude Excelente de 99% a 100% em todo os anos. A variável Raça/Cor da pele teve grau de completude Bom entre 2008 (93,15%) e 2009 (92,45%), Excelente em 2010 (95,92%), Bom em 2011 (93,37%) e a partir de 2012 até 2016 a completude foi Excelente de 96% a 98%. Na variável Escolaridade, a completude foi Ruim de 71% a 77% entre 2008 e 2011, e de 2012 a 2016 a completude foi Regular de 81% a 88%. A variável Idade apresentou completude Excelente de 99% a 100% em todos os anos (Tabela 8).

No bloco de Caracterização do diagnóstico, a variável Data da primeira consulta demonstrou completude Excelente de 100% em todo o período. A variável Data de diagnóstico oscilou entre completude de grau Bom e Excelente, em 2008 (96,81%) a completude foi Excelente, de 2009 até 2013 o grau de completude foi Bom de 94% a 95%, em 2014 (95,77%) e 2015 (95,98%) a completude foi Excelente e em 2016 (95,15%) o grau de completude foi Bom. Na variável Diagnóstico e tratamento anteriores, o grau de completude em 2008 (93,94%) e 2009 (95,46%) foi Bom, em 2010 (95,82%) foi Excelente, em 2011 (94,53%) foi Bom e a partir de 2012 até 2016 a completude foi Excelente de 95% até 99%. A variável Base mais importante para o diagnóstico do tumor apresentou grau de completude Bom em 2008 (94,13%) e de 2009 a 2016 a completude foi Excelente de 97% a 99% (Tabela 8).

No bloco de Caracterização do Tumor, a variável Tipo histológico demonstrou completude Excelente de 98% a 99% em todos os anos. Na variável Estadiamento clínico do tumor (TNM) a completude foi Ruim em 2008 (75,80%), Regular em 2009 (80,26%) e 2010 (80,14%), e entre 2011 a 2016 a completude foi Ruim de 74% a 79%. Já na variável TNM, a completude foi Ruim de 59% a 70% entre 2008 a 2016 (Tabela 8).

No bloco de Caracterização do 1º tratamento, a variável Data de início do tratamento apresentou grau de completude Bom de 92% a 94% entre 2008 a 2016. A variável Principal razão para a não realização do tratamento antineoplásico no hospital demonstrou completude Regular em 2008 (86,32%), Ruim em 2009 (78,14%) e 2010 (67,12%), Regular em 2011 (81,17%) e 2012 (88,76%) e a partir de 2013 até 2016 o grau de completude foi Bom de 93%. A completude na variável Primeiro tratamento recebido no hospital foi de grau Bom de 94% a 95% entre 2008 a 2014, em 2015 (96,18%) foi Excelente e em 2016 (95,16%) foi Bom. Na variável Estado da doença ao final do tratamento, a completude foi Ruim de 58% a 72% em todo o período, ressalta-se o ano de 2008 (58,20%) e 2009 (60,82%) com os piores

percentuais. A completude da variável Data de óbito foi Excelente de 100% em todos os anos (Tabela 8).

**Tabela 8 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer de Mama, Região Sul, 2008-2016.**

Bloco de variáveis	Câncer de Mama									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Identificação do paciente</b>										
Sexo	99,96	99,95	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Raça/Cor da pele	93,15	92,45	95,92	93,37	96,76	97,83	97,12	97,61	98,21	
Escolaridade	71,95	72,54	75,05	77,25	81,91	87,39	88,00	86,00	88,19	
Estado de residência	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Idade	100,00	100,00	100,00	99,95	100,00	100,00	99,98	100,00	100,00	
<b>Caracterização do diagnóstico</b>										
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Data de diagnóstico	96,81	94,68	94,05	94,43	95,53	94,40	95,77	95,98	95,15	
Diagnóstico e tratamento anteriores	93,94	95,46	95,82	94,53	95,76	95,77	96,62	98,33	99,13	
Base mais importante para o diagnóstico do tumor	94,13	97,67	98,46	98,37	98,48	98,74	99,39	99,28	99,17	
<b>Caracterização do tumor</b>										
Tipo histológico	98,97	99,53	99,90	99,98	99,97	99,87	99,92	99,97	99,66	
Estadiamento clínico do tumor (TNM)	75,80	80,26	80,14	74,93	77,23	76,59	77,66	75,05	79,03	
TNM	59,71	67,80	68,25	62,62	62,99	65,83	67,37	66,75	70,85	
<b>Caracterização do 1º tratamento</b>										
Data de início do tratamento	92,74	94,21	93,61	93,92	93,78	94,19	94,43	94,19	94,39	
Principal razão para a não realização do tratamento antineoplásico no hospital	86,32	78,14	67,12	81,17	88,76	93,21	93,59	93,37	93,41	
Primeiro tratamento recebido no hospital	95,03	94,81	94,99	93,84	95,18	95,19	95,55	96,18	95,16	
Estado da doença ao final do tratamento	58,20	60,82	65,78	65,88	69,57	67,83	72,58	69,35	69,51	
Data de óbito	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	



Fonte: Elaborado pelo autor com dados do IntegradorRHC.

### 5.1.2 Câncer de Colo do Útero

No Brasil, com exceção de São Paulo, entre 2008 e 2016, foram notificados no IntegradorRHC 97.022 casos novos de câncer do colo do útero.

Com relação ao número de casos por Grande Região, entre os anos de 2008 a 2016, foram notificados pelo IntegradorRHC, 11.979 (12,35%) casos novos de câncer de colo do útero na Região Norte, 39.933 (41,16%) na Região Nordeste, 4.899 (5,05%) na Região Centro-Oeste, 19.387 (19,98%) na Região Sudeste, não considerando São Paulo, e 20.284 (20,91%) na Região Sul.

Ressalta-se novamente que a contabilização dos casos de câncer de colo do útero por Grande Região não é o mesmo valor que o observado no Brasil, uma vez que casos que foram notificados sem o preenchimento do Estado de residência (n = 540; 0,55%) não são alocados a nenhuma das cinco regiões do Brasil.

#### • Câncer no Colo do Útero - Brasil

No que diz respeito as variáveis do bloco de Identificação do paciente, a variável Sexo apresentou completude Excelente de 99% e 100% em todo os anos de estudo. A variável Raça/Cor da pele oscilou entre completude Regular nos anos de 2008 (87,23%) e 2009 (89,27%) e grau de completude Bom entre 90% a 92% nos anos de 2010 a 2016. Já na variável Escolaridade observou-se completude Ruim 72% a 77% entre 2008 a 2011, e a partir de 2012 (80,92%) a completude foi Regular até 2014 (82,73%) e de 2015 (70,39%) a 2016 (78,42%) a completude voltou a ficar Ruim. A variável Estado de residência obteve completude Excelente de 99% em todo o período. A variável idade apresentou completude Excelente de 100% em todos os anos (Tabela 9).

Nas variáveis do Bloco de Caracterização do diagnóstico, a variável Data da primeira consulta obteve completude Excelente com 100% de preenchimento. Na variável Data do diagnóstico, a completude foi Excelente entre 97% a 98% em todo o seguimento. A variável Diagnóstico e tratamento anteriores apresentou completude Excelente de 98% a 99% em todos os anos. A variável Base mais importante para o diagnóstico do tumor dispôs de completude Excelente de 99% em todo o estudo (Tabela 9).

No bloco de Caracterização do tumor, a completude da variável Tipo histológico foi Excelente de 99% em todo o estudo. Na variável Estadiamento clínico do tumor (TNM) a completude foi Regular nos anos de 2008 (86,81%) e 2009 (83,93%) e a partir de 2010 a 2016

a completude foi Ruim entre 79% e 75%. A variável TNM apresentou completude Ruim em todos os anos variando de 69% a 52%, o pior percentual de completude nessa variável foi observado no ano de 2011 (52,40%) (Tabela 9).

No bloco de Caracterização do 1º tratamento, a variável Data de início do tratamento obteve grau de completude Bom entre 2008 (92,11%) e 2011 (95,21%), em 2012 (95,84%) e 2013 (95,99%) a completude foi Excelente e de 2014 a 2016 a completude foi de grau Bom de 95%. A variável Principal razão para a não realização do tratamento antineoplásico no hospital obteve completude Regular nos anos de 2008 (89,68%) e 2010 (85,58), grau de completude Bom em 2009 (90,14%) e 2011 (94,27%) e completude Excelente de 96% a 97% entre 2012 a 2016. Na variável Primeiro tratamento recebido no hospital a completude foi Excelente de 99% em todos os anos. A variável Estado da doença ao final do tratamento apresentou completude Ruim de 72% a 78% em todo o período. A variável Data de óbito demonstrou completude Excelente de 100% em quase todos os anos com exceção de 2008 (99,92%) (Tabela 9).



**Tabela 9 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer do Colo do Útero, Brasil, 2008-2016**

Bloco de variáveis	Câncer do Colo do Útero									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Identificação do paciente</b>										
Sexo	99,99	99,99	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Raça/Cor da pele	87,23	89,27	91,79	90,74	91,42	92,04	91,76	90,43	90,97	
Escolaridade	72,28	72,23	74,91	77,78	80,92	83,36	82,73	79,39	78,42	
Estado de residência	99,76	99,83	99,57	99,45	99,25	99,47	99,29	98,96	99,46	
Idade	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
<b>Caracterização do diagnóstico</b>										
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Data de diagnóstico	98,52	98,18	97,58	97,70	97,58	97,92	98,45	97,93	98,35	
Diagnóstico e tratamento anteriores	99,28	99,32	99,27	98,72	98,86	99,05	99,22	98,68	98,95	
Base mais importante para o diagnóstico do tumor	99,16	99,45	99,67	99,65	99,65	99,44	99,52	99,64	99,61	
<b>Caracterização do tumor</b>										
Tipo histológico	99,87	99,82	99,90	99,93	99,90	99,90	99,97	99,96	99,89	
Estadiamento clínico do tumor (TNM)	86,81	83,93	79,85	79,51	78,55	78,14	78,24	75,78	76,50	
TNM	69,06	63,84	55,43	52,40	54,30	56,18	55,50	53,87	53,61	
<b>Caracterização do 1º tratamento</b>										
Data de início do tratamento	92,11	93,27	94,63	95,21	95,84	95,99	95,62	95,59	95,37	
Principal razão para a não realização do tratamento antineoplásico no hospital	89,68	90,14	85,58	94,27	97,24	97,88	97,66	96,82	97,24	
Primeiro tratamento recebido no hospital	99,28	99,24	99,33	99,08	99,16	99,28	99,08	99,28	99,22	
Estado da doença ao final do tratamento	72,68	75,28	78,01	78,47	78,63	75,74	77,31	76,64	77,53	
Data de óbito	99,92	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	



Fonte: Elaborado pelo autor com dados do IntegradorRHC.

### • Câncer de Colo do Útero - Região Norte

Dentre o bloco de Identificação do paciente, a variável Sexo apresentou completude Excelente de 100% de preenchimento em todos os anos. A variável Raça/Cor da pele demonstrou completude Ruim em 2008 (68,73%) e 2009 (71,31%), Regular em 2010 (81,29%), 2011 (82,55%) e 2012 (82,85%), Ruim em 2013 (79,84%), 2014 (76,85%) e 2015 (78,36%), e em 2016 (84,55%) a completude foi Regular. Na variável Escolaridade observou-se uma variação entre completude de grau Bom, Regular e Ruim, em 2008 (92,57%) a completude foi de grau Bom, em 2009 (79,66%) e 2010 (79,87%) a completude foi Ruim, o grau de completude foi Bom em 2011 (90,41%), Regular entre 2012 (83,72%) e 2013 (82,90%) e entre 2014 a 2016 a completude foi de grau Bom oscilando entre 90% e 94%. A variável Idade dispunha de completude Excelente de 100% em todos os anos (Tabela 10).

No bloco de caracterização do diagnóstico, a variável Data da primeira consulta possui completude Excelente de 100% em todo o período. A variável Data de diagnóstico apresentou entre 2008 a 2016 completude Excelente de 99% a 100%. Na variável Diagnóstico e tratamento anteriores a completude foi Excelente de 97% a 99% em todos os anos. A completude da variável Base mais importante para o diagnóstico do tumor foi Excelente de 99% em todo o período (Tabela 10).

Com relação ao bloco de Caracterização do tumor, a variável Tipo histológico apresentou completude Excelente de 100% em todos os anos. Na variável Estadiamento clínico do tumor (TNM) constatou-se uma oscilação entre completude de grau Bom, Regular e Ruim, sendo Bom em 2008 (90,30%), Regular entre 2009 (88,19%) a 2011 (82,94%), Ruim em 2012 (77,13%), Regular de 2013 (82,11%) a 2014 (80,24%), Ruim em 2015 (76,64%) e Regular em 2016 (80,16%). A variável TNM demonstrou completude Regular no ano de 2008 (84,12%), Ruim em 2009 (69,27%), Muito ruim em 2010 (46,54%) e 2011 (42,77%) e entre 2012 a 2016 a completude foi Ruim entre 53% e 65% (Tabela 10).

No bloco de Caracterização do 1º tratamento, a variável Data de início do tratamento apresentou grau de completude Regular em 2008 (88,82%), Bom em 2009 (91,47%) e a partir de 2010 até 2016 a completude foi Excelente de 98% a 99%. A variável Principal razão para a não realização do tratamento antineoplásico no hospital variou entre completude Regular, Bom e Excelente, em 2008 (85,07%) e 2009 (88,54%) foi Regular, em 2010 (92,61%) foi Bom e entre 2011 a 2016 foi Excelente de 98% a 99%. A completude na variável Primeiro tratamento recebido no hospital foi Excelente de 99% a 100% em todos os anos. Já a variável Estado da doença ao final do tratamento apresentou grau de completude Regular em 2008 (89,99%) e

Bom entre 2009 (92,63%) a 2016 (93,79%). Na Data de óbito entre 2008 a 2016 a completude foi Excelente de 100% de preenchimento (Tabela 10).

**Tabela 10 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer do Colo do Útero, Região Norte, 2008-2016**

Bloco de variáveis	Câncer do Colo do Útero									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Identificação do paciente</b>										
Sexo	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Raça/Cor da pele	68,73	71,31	81,29	82,55	82,85	79,84	76,85	78,36	84,55	
Escolaridade	92,57	79,66	79,87	90,41	83,72	82,90	92,45	94,48	90,35	
Estado de residência	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Idade	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
<b>Caracterização do diagnóstico</b>										
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Data de diagnóstico	99,77	99,64	100,00	99,69	99,86	99,93	99,79	99,78	99,93	
Diagnóstico e tratamento anteriores	99,30	98,40	97,48	97,25	97,76	98,30	99,79	99,10	99,46	
Base mais importante para o diagnóstico do tumor	99,92	99,73	99,84	99,69	99,86	99,93	99,86	99,70	99,73	
<b>Caracterização do tumor</b>										
Tipo histológico	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Estadiamento clínico do tumor (TNM)	90,30	88,19	86,48	82,94	77,13	82,11	80,24	76,64	80,16	
TNM	84,21	69,27	46,54	42,77	59,12	65,65	61,61	63,81	53,78	
<b>Caracterização do 1º tratamento</b>										
Data de início do tratamento	88,82	91,47	98,19	98,11	98,91	99,43	98,87	98,58	98,52	
Principal razão para a não realização do tratamento antineoplásico no hospital	85,07	88,54	92,61	99,45	99,78	99,65	99,72	98,66	98,11	
Primeiro tratamento recebido no hospital	99,92	99,82	100,00	99,84	100,00	100,00	100,00	99,70	99,93	
Estado da doença ao final do tratamento	89,99	92,63	95,05	94,50	91,90	93,47	94,99	92,39	93,79	
Data de óbito	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	

Fonte: Elaborado pelo autor com dados do IntegradorRHC.

	Excelente
	Bom
	Regular
	Ruim
	Muito Ruim

### • Câncer de Colo do Útero - Região Nordeste

Com relação ao bloco de Identificação do paciente, a variável Sexo apresentou completude Excelente de 100% em todos os anos. A variável Raça/Cor da pele demonstrou completude Regular em 2008 (86,27%) e 2009 (89,24%) e a partir de 2010 até 2016 a completude foi de grau Bom variando de 91% a 93%. Já a variável Escolaridade dispôs de completude Ruim de 61% a 78% entre os anos de 2008 até 2011, de 2012 a 2014 a completude é Regular e em 2015 (79,79%) e 2016 (78,59%) a completude foi Ruim. Na variável Idade a completude foi Excelente com 100% de preenchimento em todos os anos (Tabela 11).

No bloco de Caracterização do diagnóstico, a variável Data da primeira consulta apresentou completude Excelente de 100% em todo o período. Na variável Data de diagnóstico entre 2008 a 2016 a completude foi Excelente de 96% a 98%. A variável Diagnóstico e tratamento anteriores demonstrou completude Excelente de 98% a 99% entre 2008 e 2016. A variável Base mais importante para o diagnóstico do tumor dispôs de completude Excelente de 99% em todos os anos (Tabela 11).

No bloco de Caracterização do tumor, a variável Tipo histológico mostrou completude Excelente de 100% em todos os anos. Na variável Estadiamento clínico do tumor (TNM) ocorreu oscilações entre completude Regular e Ruim, onde em 2008 (87,42%) e 2009 (82,59%) foi Regular, 2010 (79,75%) foi Ruim, 2011 (80,65%) foi Regular e entre 2012 a 2016 foi Ruim decrescendo de 78% a 76%. A variável TNM demonstra completude Ruim em todo o período, destacando-se o ano de 2016 (57,06%) com o pior percentual nessa variável. (Tabela 11).

No bloco de caracterização do 1º diagnóstico, a variável Data de início do tratamento demonstrou completude Regular em 2008 (89,14%) e entre 2009 a 2016 a completude foi de grau Bom de 91% a 95%. A variável Principal razão para a não realização do tratamento antineoplásico no hospital apresentou completude de grau Bom de 90% a 95% entre 2008 e 2011, a partir de 2012 até 2016 a completude foi Excelente de 97% a 98%. Já a variável Primeiro tratamento recebido no hospital dispôs de completude Excelente de 99% em todos os anos. A variável Estado da doença ao final do tratamento apresentou de completude Ruim de 65% a 78% na grande maioria dos anos, excetuando-se o ano de 2011 (80,32%) que a completude foi Regular. Na variável Data de óbito a completude foi Excelente de 100% em todos os anos (Tabela 11).

**Tabela 11 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer do Colo do Útero, Região Nordeste, 2008-2016**

Bloco de variáveis	Câncer do Colo do Útero									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Identificação do paciente</b>										
Sexo	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Raça/Cor da pele	86,27	89,24	92,65	92,41	92,50	93,42	93,02	92,74	91,97	
Escolaridade	61,34	67,01	73,48	78,05	81,43	83,78	81,05	79,79	78,59	
Estado de residência	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Idade	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
<b>Caracterização do diagnóstico</b>										
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Data de diagnóstico	98,37	97,86	97,29	97,38	96,47	97,85	98,05	97,25	97,42	
Diagnóstico e tratamento anteriores	99,74	99,64	99,87	99,49	99,50	99,55	99,48	98,83	98,75	
Base mais importante para o diagnóstico do tumor	99,74	99,73	99,78	99,91	99,81	99,39	99,24	99,44	99,67	
<b>Caracterização do tumor</b>										
Tipo histológico	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Estadiamento clínico do tumor (TNM)	87,42	82,59	79,75	80,65	78,75	78,41	76,38	76,65	76,71	
TNM	76,83	70,11	60,34	58,39	59,65	58,77	57,18	57,09	57,06	
<b>Caracterização do 1º tratamento</b>										
Data de início do tratamento	89,14	91,36	93,00	94,76	94,97	94,94	95,27	94,88	94,29	
Principal razão para a não realização do tratamento antineoplásico no hospital	90,88	93,28	93,65	95,01	97,91	98,25	98,05	98,29	98,16	
Primeiro tratamento recebido no hospital	99,67	99,24	99,84	99,81	99,50	99,73	99,71	99,84	99,65	
Estado da doença ao final do tratamento	65,69	69,60	73,83	80,32	78,44	73,71	74,32	74,78	74,45	
Data de óbito	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	



Fonte: Elaborado pelo autor com dados do IntegradorRHC.

### • Câncer de Colo do Útero - Região Centro-Oeste

Referindo-se ao bloco de Identificação do paciente, a variável Sexo demonstrou completude Excelente de 100% de preenchimento em todos os anos. Na variável Raça/Cor da pele, a completude variou entre grau Bom, Regular e Ruim, sendo em 2008 (94,68%) e 2009 (92,02%) de grau Bom, em 2010 (85,88%) foi Regular, em 2011 (74,83%) e 2012 (78,23%) foi Ruim, em 2013 (92,89%) foi Bom, em 2014 (89,31%) foi Regular e entre 2015 a 2016 foi Ruim de 68% a 66%. Na variável Escolaridade entre os anos de 2008 e 2016, a completude oscilou entre Ruim e Muito ruim, de 2008 (70,64%) a 2010 (56,86%) foi Ruim, em 2011 (44%) Muito Ruim, de 2012 (56,01%) a 2014 (50,94%) foi Ruim e em 2015 (37,55%) e 2016 (30,53%) foi Muito ruim. A variável Idade apresentou completude Excelente de 100% em todos os anos (Tabela 12).

No bloco de Caracterização do diagnóstico, a completude da variável Data da primeira consulta foi Excelente de 100% em todo o período. A variável Data de diagnóstico demonstrou completude Excelente de 97% a 99% em todos os anos. Na variável Diagnóstico e tratamento anteriores, a completude foi Excelente de 99% a 100% em grande parte dos anos, com exceção do ano de 2015 (95,47%) que apresentou completude de grau Bom. Já a variável Base mais importante para o diagnóstico do tumor dispõe de completude Excelente de 99% a 100% em todos os anos (Tabela 12).

No bloco de Caracterização do tumor, a variável Tipo histológico demonstrou completude Excelente de 100% entre 2008 e 2016. Na variável Estadiamento clínico do tumor (TNM), a completude foi Ruim de 51% a 73% na maioria dos anos, com exceção de 2010 (45,29%), 2011 (43,33%) e 2016 (41,18%) que a completude atingiu o grau Muito ruim. A variável TNM apresentou completude Muito ruim em todos os anos, tendo o ano de 2010 (23,33%) o pior percentual de preenchimento (Tabela 12).

No bloco de Caracterização do 1º tratamento, a variável Data de início do tratamento variou entre completude de grau Excelente, Bom e Regular, em 2008 (98,51%) e 2009 (96,11%) a completude foi Excelente, em 2010 (92,55%) foi Bom, em 2011 (88,33%) foi Regular, em 2012 (97,96%) foi Excelente, de 2013 a 2015 a completude foi de grau Bom de 90% a 95% e em 2016 (96,80%) foi Excelente. A variável Principal razão para a não realização do tratamento antineoplásico no hospital dispõe de completude oscilante entre grau Excelente, Bom e Regular, em 2008 (97,45%) foi Excelente, de 2009 a 2011 foi Bom de 90% a 94%, em 2012 e 2013 a completude foi Excelente de 96% a 98%, em 2014 (92,45%) foi Bom, em 2015 (86,95%) foi Regular e em 2016 (92,30%) foi Bom. Já a variável Primeiro

tratamento recebido no hospital apresentou em sua grande maioria completude Excelente de 96% a 100%, excetuando-se o ano de 2014 (93,50%) que a completude foi de grau Bom. Na variável Estado da doença ao final do tratamento a completude foi Ruim de 56% a 79% em todo o período, o ano de 2008 (56,60%) apresentou o pior percentual de completude da variável. A variável Data de óbito demonstrou completude Excelente de 100% em todos os anos (Tabela 12).



**Tabela 12 - Proporção de campos preenchidos e completez das variáveis do IntegradorRHC para Câncer do Colo do Útero, Região Centro-Oeste, 2008-2016**

Bloco de variáveis	Câncer do Colo do Útero									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Identificação do paciente</b>										
Sexo	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Raça/Cor da pele	94,68	92,02	85,88	74,83	78,23	92,89	89,31	68,04	66,39	
Escolaridade	70,64	64,98	56,86	44,00	56,01	63,74	50,94	37,55	30,53	
Estado de residência	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Idade	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
<b>Caracterização do diagnóstico</b>										
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Data de diagnóstico	99,79	98,83	99,22	97,83	99,09	99,76	99,58	99,47	99,58	
Diagnóstico e tratamento anteriores	100,00	99,61	100,00	100,00	99,77	99,29	99,16	95,47	96,36	
Base mais importante para o diagnóstico do tumor	100,00	99,81	99,80	99,83	100,00	99,29	99,58	99,87	99,30	
<b>Caracterização do tumor</b>										
Tipo histológico	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Estadiamento clínico do tumor (TNM)	58,94	51,56	45,29	43,33	59,86	65,88	73,17	50,07	41,18	
TNM	45,53	28,02	23,33	28,83	33,79	43,84	43,82	37,42	32,07	
<b>Caracterização do 1º tratamento</b>										
Data de início do tratamento	98,51	96,11	92,55	88,33	97,96	95,02	90,99	95,34	95,80	
Principal razão para a não realização do tratamento antineoplásico no hospital	97,45	94,55	90,59	91,83	98,87	96,68	92,45	86,95	92,30	
Primeiro tratamento recebido no hospital	99,57	99,42	100,00	99,50	99,09	99,29	93,50	96,80	97,20	
Estado da doença ao final do tratamento	56,60	72,37	72,35	71,83	79,37	76,54	64,15	76,30	68,07	
Data de óbito	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	



Fonte: Elaborado pelo autor com dados do IntegradorRHC.

### • Câncer de Colo do Útero - Região Sudeste

Com relação ao bloco de Identificação do paciente, a variável Sexo demonstrou completude Excelente de 100% em todos os anos. A variável Raça/Cor da pele em sua grande maioria demonstrou completude de grau Bom de 91% a 95%, com exceção do ano de 2015 (95,90%) que apresentou completude Excelente. Na variável Escolaridade observou-se uma variação entre completude Regular e Ruim, em 2008 (81,63%) foi Regular, de 2009 a 2011 foi Ruim de 72% a 79%, de 2012 a 2015 foi Regular de 81% a 82% e em 2016 (78,83%) foi Ruim. A variável Idade apresentou completude Excelente de 100% em todos os anos (Tabela 13).

No bloco de Caracterização do diagnóstico, a variável Data da primeira consulta demonstrou completude Excelente de 100% em todo o período. Na variável Data de diagnóstico a completude foi Excelente de 98% a 99% entre 2008 a 2016. A variável Diagnóstico e tratamento anteriores dispunha de completude Excelente de 99% em todos os anos. A completude da variável Base mais importante para o diagnóstico do tumor foi Excelente de 99% para 2008 a 2016 (Tabela 13).

No preenchimento do bloco de Caracterização do tumor, a variável Tipo histológico apresentou completude Excelente de 100% em todo o período. A variável Estadiamento clínico do tumor (TNM) apresentou completude oscilante entre Regular e Ruim, entre 2008 a 2012 a completude foi Regular de 89% a 80%, em 2013 e 2014 a completude foi Ruim entre 76% e 78%, em 2015 (80,23%) a completude foi Regular e em 2016 (79,15%) a completude foi Ruim. Na variável TNM, a completude foi Ruim de 2008 (61,51%) a 2010 (52,57%), e a partir de 2011 a completude foi Muito ruim de 33% a 44%, destaca-se o ano de 2012 (33,44%) com o menor percentual de completude (Tabela 13).

Já no bloco de Caracterização do 1º tratamento, a variável Data de início do tratamento apresentou completude Excelente de 96% a 98% em quase todos os anos, com exceção de 2009 (95,19%) que o grau de completude foi Bom. Na variável Principal razão para a não realização do tratamento antineoplásico no hospital a completude em 2008 (88,69%) e 2009 (87,10%) foi Regular, em 2010 (71,27%) foi Ruim e entre 2011 a 2016 a completude foi Excelente de 97% a 99%. A variável Primeiro tratamento recebido no hospital obteve completude Excelente de 99% a 100% em todos os anos. Na variável Estado da doença ao final do tratamento a completude foi Regular em 2008 (86,34%) e 2009 (82,14%) e de 2010 a 2016 a completude foi Ruim de 70% a 78%. A variável Data de óbito demonstrou completude Excelente de 100% em quase todo o período, com exceção de 2008 (99,73%) (Tabela 13).

**Tabela 13 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer do Colo do Útero, Região Sudeste, 2008-2016**

Bloco de variáveis	Câncer do Colo do Útero									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Identificação do paciente</b>										
Sexo	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Raça/Cor da pele	93,30	93,77	94,08	94,51	93,68	91,79	95,57	95,90	95,63	
Escolaridade	81,63	79,59	72,52	77,32	82,69	81,81	82,87	82,24	78,83	
Estado de residência	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Idade	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
<b>Caracterização do diagnóstico</b>										
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Data de diagnóstico	99,74	99,12	99,22	98,68	99,59	99,53	99,79	99,56	99,84	
Diagnóstico e tratamento anteriores	99,95	99,95	99,64	99,85	99,44	99,86	99,83	99,44	99,96	
Base mais importante para o diagnóstico do tumor	99,90	99,46	99,64	99,51	99,85	99,76	99,71	99,92	99,84	
<b>Caracterização do tumor</b>										
Tipo histológico	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Estadiamento clínico do tumor (TNM)	89,92	88,71	82,81	84,03	80,89	76,64	78,11	80,23	79,15	
TNM	61,51	57,36	52,57	44,68	33,44	35,66	39,18	41,86	42,55	
<b>Caracterização do 1º tratamento</b>										
Data de início do tratamento	97,59	95,19	96,47	97,94	97,89	97,25	96,69	98,03	97,17	
Principal razão para a não realização do tratamento antineoplásico no hospital	88,69	87,10	71,27	99,71	99,49	99,19	97,72	98,83	97,77	
Primeiro tratamento recebido no hospital	99,95	100,00	99,84	100,00	99,90	99,91	99,75	99,96	99,84	
Estado da doença ao final do tratamento	86,34	82,14	78,34	73,84	72,11	70,37	76,00	76,42	76,11	
Data de óbito	99,73	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	



Fonte: Elaborado pelo autor com dados do IntegradorRHC.

### • Câncer de Colo do Útero - Região Sul

Entre as variáveis do bloco de Identificação do paciente, a variável Sexo apresenta completude Excelente de 99% a 100% em todo o período. A variável Raça/Cor da pele demonstrou completude de grau Bom de 93% a 95% entre 2008 e 2011, em 2012 (95,91%) e 2013 (97,06%) foi Excelente, em 2014 (95,49%) foi Bom e entre 2015 (96,24%) e 2016 (96,42%) foi Excelente. Na variável Escolaridade, a completude foi Ruim nos anos de 2008 (76,21%) e 2009 (74,30%) e entre 2010 a 2016 a completude foi Regular de 80% a 88%. Na variável Idade, a completude foi Excelente de 100% em todo o período (Tabela 14).

No bloco de Caracterização do diagnóstico, a variável Data da primeira consulta foi Excelente de 100% em todos os anos. A variável Data de diagnóstico obteve completude Excelente de 95 a 97% em quase todo o período, com exceção de 2010 (95,43%) e 2013 (95,09%) que o grau de completude foi Bom. Na variável Diagnóstico e tratamento anteriores observou-se completude Excelente de 99% a 100% em todos os anos. A completude na variável Base mais importante para o diagnóstico do tumor foi Excelente de 96% a 99% em todo o período (Tabela 14).

No bloco de Caracterização do tumor, a Tipo histológico apresentou completude Excelente de 99% a 100% entre 2008 a 2016. A variável Estadiamento clínico do tumor (TNM) mostrou completude Regular de 80% a 87% em grande parte do tempo, excetuando-se os anos de 2013 (78,91%) e 2015 (78,05%) que a completude foi Ruim. Já a variável TNM apresentou completude Ruim de 54% a 67% entre 2008 e 2016 (Tabela 14).

Nas variáveis do bloco de Caracterização do 1º diagnóstico, a Data de início do tratamento apresentou completude oscilante entre Excelente e Bom, sendo em grande parte Bom com percentual entre 92% a 95% e Excelente nos anos de 2009 (95,83%) e 2014 (95,67%). A variável Principal razão para a não realização do tratamento antineoplásico no hospital demonstrou completude Regular em 2008 (88,98%) e 2009 (86,33%), Ruim em 2010 (77,19%), Regular em 2011 (86,34%), Bom entre 2012 (92,34%) e 2013 (95,26%), Excelente em 2014 (96,69%), Bom em 2015 (95,22%) e Excelente em 2016 (96,06%). A completude na variável Primeiro tratamento recebido no hospital foi Excelente de 96% a 98% entre 2008 e 2016. Na variável Estado da doença ao final do tratamento, a completude foi Ruim entre 2008 e 2016, com destaque para o ano de 2008 (68,74%) com o pior percentual. A variável Data de óbito apresentou completude Excelente de 100% em todo o período (Tabela 14)

**Tabela 14 - Proporção de campos preenchidos e completude das variáveis do IntegradorRHC para Câncer do Colo do Útero, Região Sul, 2008-2016**

Bloco de variáveis	Câncer do Colo do Útero									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Identificação do paciente</b>										
Sexo	99,95	99,95	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Raça/Cor da pele	93,79	94,23	95,47	93,25	95,91	97,06	95,49	96,24	96,42	
Escolaridade	76,21	74,30	81,93	80,25	82,74	88,55	87,32	82,52	86,42	
Estado de residência	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Idade	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
<b>Caracterização do diagnóstico</b>										
Data da primeira consulta	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	
Data de diagnóstico	96,60	97,07	95,43	96,42	96,43	95,09	96,73	95,84	97,18	
Diagnóstico e tratamento anteriores	97,48	98,41	98,63	97,41	97,68	97,86	97,70	98,44	99,15	
Base mais importante para o diagnóstico do tumor	96,50	98,61	99,36	99,26	99,10	99,12	99,73	99,60	99,55	
<b>Caracterização do tumor</b>										
Tipo histológico	99,66	99,65	99,96	100,00	99,96	99,87	99,96	99,91	99,96	
Estadiamento clínico do tumor (TNM)	86,75	87,97	81,55	80,58	80,89	78,91	81,54	78,05	82,44	
TNM	54,81	62,48	59,97	58,56	63,41	65,87	67,54	61,60	65,41	
<b>Caracterização do 1º tratamento</b>										
Data de início do tratamento	94,13	95,83	95,26	94,73	94,27	95,51	95,67	92,85	93,68	
Principal razão para a não realização do tratamento antineoplásico no hospital	88,98	86,33	77,19	86,34	92,34	95,26	96,69	95,22	96,06	
Primeiro tratamento recebido no hospital	97,28	98,16	97,44	96,46	97,42	97,74	97,79	98,12	97,80	
Estado da doença ao final do tratamento	68,74	72,02	77,92	73,09	77,18	74,30	77,92	71,61	77,87	
Data de óbito	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	



Fonte: Elaborado pelo autor com dados do IntegradorRHC.

## 5.2 INCONSISTÊNCIA

### 5.2.1 Câncer de Mama

As inconsistências escolhidas na análise de câncer de mama para o Brasil e Grandes Regiões foram: Data da primeira consulta < Data do diagnóstico; Data da primeira consulta > Data de início do tratamento; Data do diagnóstico > Data do início do tratamento; Localização primária do tumor & Tipo histológico; e TNM & Estadiamento

Notificaram-se 195.899 casos de câncer de mama no período de 2008 a 2016 no Brasil, destes casos, a inconsistência Data da primeira consulta < Data do diagnóstico mostrou um pequeno aumento ao longo dos anos, em 2008 o percentual foi de 3,12% e em 2016 foi de 4,82%. A Data da primeira consulta > Data de início do tratamento não teve nenhum aumento no decorrer dos anos, sempre oscilando entre 0,01% e 0,04%. Já na Data do diagnóstico > Data do início do tratamento, a inconsistência se mostrou crescente atingindo em 2008 (0,31%) e em 2016 (0,59%). Na Localização primária do tumor & Tipo histológico e TNM & Estadiamento observou-se comportamento semelhante das outras inconsistências, aumentando de 0,87% e 0,43% em 2008 para 1,83% e 1,67% em 2016, respectivamente (Tabela 15).

Adentrando-se na análise por Grande Região, entre 2008 a 2016, notou-se atuação similar ao observado no Brasil, a Data da primeira consulta < Data do diagnóstico mostrou aumento da inconsistência ao longo dos em todas as regiões, sendo o maior percentual na região Centro-Oeste em 2016 (5,77%). Vale ressaltar que por mais que tenha ocorrido o aumento na região Centro-Oeste e Sudeste, elas apresentaram variações ao longo do período analisado, não sofrendo crescimento constante como as demais regiões.

Na Data da primeira consulta > Data de início do tratamento, os percentuais de inconsistências foram relativamente baixos em todas as regiões, sendo o maior na região Norte em 2010 (0,10%), nessa inconsistência observou-se uma redução e estabilidade em quase todas as regiões no decorrer dos anos, com exceção da região Centro-Oeste que em 2016 subiu para 0,05%.

A Data do diagnóstico > Data do início do tratamento sofreu um aumento constante nas inconsistências entre 2008 a 2016 em grande parte das regiões, com exceção da região Centro-Oeste e Sul. O maior grau de inconsistência na combinação dessas duas variáveis foi no Sudeste em 2016 (1,01%), além disso, o Sudeste apresentou as maiores inconsistências ao longo dos anos em comparação com as outras regiões. As inconsistências na Localização primária do tumor & Tipo histológico sugeriu crescimento em quase todos os anos e regiões,

sendo o maior percentual na região Norte 2016 (3,04%). Nessas inconsistências a região Centro-oeste aumentou em 2016 (2,92%) com relação a 2015 (2,44%), entretanto, oscilou entre aumento e queda, e a região Sudeste variou entre aumento e queda durante 2008 e 2011, porém as inconsistências passaram a crescer a partir de 2012.

A inconsistência no TNM & Estadiamento foi oscilante entre as regiões ao longo do período, a região Norte oscilou entre 0,89% e 2,34%, a região Nordeste mostrou crescimento ao longo dos anos, a região Centro-Oeste oscilou entre 2008 (0,83%) e 2012 (0,66%), mas aumentou entre 2013 (1,11%) e 2015 (2,55%), a região Sudeste oscilou durante todo o período entre 0,36% a 1,50% e a região Sul sofreu um aumento crescente entre 2008 a 2016. No TNM & Estadiamento, a inconsistência aumentou em quase todas as regiões ao compará-las com 2015, com exceção da região Centro-Oeste que diminuiu para 2,32%. É importante frisar que a região Centro-Oeste apresentou o percentual de inconsistência mais alto de todas as regiões, que foi de 2,55% em 2015 para TNM & Estadiamento (Tabela 15).

Tabela 15 - Proporção (%) de inconsistências em variáveis combinadas da ficha de registro do tumor para Câncer de Mama no Brasil e Grandes Regiões, 2008 a 2016

Inconsistências	Câncer de Mama (%)									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	
<b>Brasil</b>	Data da primeira consulta < Data do diagnóstico	3,12	3,46	3,59	3,74	3,80	4,23	4,54	4,59	4,82
	Data da primeira consulta > Data de início do tratamento	0,01	0,02	0,03	0,04	0,04	0,04	0,03	0,03	0,04
	Data do diagnóstico > Data do início do tratamento	0,31	0,32	0,37	0,28	0,39	0,47	0,54	0,57	0,59
	Localização primária do tumor & Tipo histológico	0,87	1,02	0,97	1,05	1,10	1,28	1,47	1,66	1,83
	TNM & Estadiamento	0,43	0,51	0,85	1,09	1,06	1,25	1,41	1,43	1,67
<b>Norte</b>	Data da primeira consulta < Data do diagnóstico	2,38	2,71	2,73	3,07	3,31	2,66	3,56	3,28	3,75
	Data da primeira consulta > Data de início do tratamento	0,02	0,07	0,10	0,07	0,01	0,02	0,00	0,00	0,01
	Data do diagnóstico > Data do início do tratamento	0,30	0,32	0,42	0,43	0,33	0,48	0,24	0,30	0,53
	Localização primária do tumor & Tipo histológico	0,61	0,87	1,01	0,73	1,16	1,27	1,34	1,74	3,04
	TNM & Estadiamento	0,89	1,20	0,90	0,66	1,31	2,23	1,77	1,85	2,34
<b>Nordeste</b>	Data da primeira consulta < Data do diagnóstico	3,52	3,93	3,92	4,42	4,60	5,01	5,19	5,07	5,42
	Data da primeira consulta > Data de início do tratamento	0,01	0,03	0,03	0,04	0,04	0,06	0,08	0,06	0,06
	Data do diagnóstico > Data do início do tratamento	0,37	0,26	0,20	0,20	0,28	0,33	0,40	0,44	0,42
	Localização primária do tumor & Tipo histológico	0,70	0,87	0,75	0,91	0,91	1,10	1,22	1,49	1,61
	TNM & Estadiamento	0,35	0,47	0,80	1,24	1,28	1,31	1,45	1,53	1,71
<b>Centro-Oeste</b>	Data da primeira consulta < Data do diagnóstico	4,22	3,88	4,17	3,37	2,12	3,43	3,22	5,23	5,77
	Data da primeira consulta > Data de início do tratamento	0,01	0,01	0,00	0,00	0,02	0,01	0,00	0,00	0,05
	Data do diagnóstico > Data do início do tratamento	0,12	0,06	0,05	0,04	0,09	0,29	0,26	0,24	0,44
	Localização primária do tumor & Tipo histológico	1,07	0,94	0,84	0,96	0,85	1,61	2,99	2,44	2,92
	TNM & Estadiamento	0,83	1,05	0,99	0,71	0,66	1,11	1,96	2,55	2,32
<b>Sudeste</b>	Data da primeira consulta < Data do diagnóstico	3,07	3,48	3,46	3,19	3,00	3,65	4,32	4,32	4,43
	Data da primeira consulta > Data de início do tratamento	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,02
	Data do diagnóstico > Data do início do tratamento	0,46	0,53	0,59	0,46	0,51	0,64	0,87	0,99	1,01
	Localização primária do tumor & Tipo histológico	0,83	0,80	0,73	0,72	0,85	0,92	1,19	1,25	1,25
	TNM & Estadiamento	0,36	0,25	0,73	0,87	0,80	1,05	1,28	1,10	1,50
<b>Sul</b>	Data da primeira consulta < Data do diagnóstico	2,73	3,03	3,41	3,63	3,99	4,30	4,41	4,39	4,57
	Data da primeira consulta > Data de início do tratamento	0,01	0,01	0,04	0,08	0,09	0,06	0,02	0,02	0,03
	Data do diagnóstico > Data do início do tratamento	0,13	0,26	0,40	0,21	0,46	0,53	0,49	0,40	0,44
	Localização primária do tumor & Tipo histológico	1,15	1,46	1,48	1,59	1,54	1,74	1,81	2,01	2,27
	TNM & Estadiamento	0,44	0,61	1,00	1,23	1,00	1,15	1,28	1,36	1,61

Fonte: Elaborado pelo autor com dados do IntegradorRHC.



### 5.2.2 Câncer do Colo do Útero

Na análise de inconsistências para câncer de colo do útero no Brasil e Grandes Regiões, serão abordadas as seguintes combinações: Data da primeira consulta < Data do diagnóstico; Data da primeira consulta > Data de início do tratamento; Data do diagnóstico > Data do início do tratamento; Sexo & Localização primária do tumor; Localização primária do tumor & Tipo histológico; e TNM & Estadiamento.

No Brasil foram notificados 97.022 casos de câncer do colo do útero entre 2008 e 2016. Em relação às inconsistências, a Data da primeira consulta < Data do diagnóstico se mostrou crescente no decorrer do período, atingindo em 2008 (3,13%) e 2016 (3,50%) percentuais de inconsistências.

Já a Data da primeira consulta > Data de início do tratamento se manteve estável em todo o período variando entre 0,01% e 0,03% percentuais de inconsistência. A inconsistência na Data do diagnóstico > Data do início do tratamento apresentou crescimento a partir de 2014 (0,40%) até 2016 (0,76%), e entre 2008 a 2013 notou-se inconstância nos percentuais, uma vez que eles aumentavam e diminuem a cada ano.

As inconsistências em Sexo & Localização primária do tumor foram baixas ( $n = 6$ ), devido a esse motivo o resultado fica como 0,00%, entretanto, ocorreram 0 inconsistências em 2008, 1 em 2009, 0 em 2010, 0 em 2011, 3 em 2012, 1 em 2013, 0 em 2014, 0 em 2015 e 1 em 2016.

Na Localização primária do tumor & Tipo histológico e no TNM & Estadiamento observou-se um crescimento das inconsistências ao longo dos anos, atingindo em 2008 (0,97%) e (0,89%) e em 2016 (1,21%) e (1,50%), respectivamente (Tabela 16).

O padrão de inconsistências observados no Brasil se mostrou similar para as regiões Centro-Oeste, Sudeste e Sul. As inconsistências na Data da primeira consulta < Data do diagnóstico foram variantes entre 2008 (1,97%) a 2016 (1,83%) sendo os percentuais mais altos em 2011 (2,64%) e 2012 (2,36%) no Norte, no Nordeste têm-se uma queda de inconsistências entre 2008 (4,38%) até 2012 (2,97%) e a partir de 2013 (3,36%) nota-se um crescimento até 2016 (3,61%), já o Centro-Oeste apresentou crescimento contínuo entre 2008 a 2016, iniciando a análise com 3,02% e terminando com 7,08% percentuais de inconsistência, o Sudeste demonstrou variações no crescimento ao longo dos anos, terminando o período com 2,65% de percentuais de inconsistência, e por fim, no Sul, observou-se inconstância nas inconsistências, diminuindo entre 2008 (2,63%) a 2009 (2,40%) e aumentando entre 2010 (3,18%) até 2016 (4,20%).

Na Data da primeira consulta > Data de início do tratamento, as inconsistências foram baixas variando de 0,01% a 0,04% em quase todas as regiões e anos, excetuando-se a região Centro-oeste em 2012 que apresentou percentual de inconsistência de 0,33%. Entre a inconsistência de Data do diagnóstico > Data do início do tratamento notou-se uma variação nos percentuais conforme a região de análise, os maiores percentuais de inconsistências ao longo dos anos foram vistos na região Centro-Oeste, com destaque para o ano de 2016 (2,27%).

A inconsistência de Sexo & Localização primária do tumor foi igual 0,00% em quase todas as regiões, com exceção da região Norte em 2009 (0,01%) e Sudeste em 2016 (0,01%) (Tabela 16).

Conforme a combinação de Localização primária do tumor & Tipo histológico observou-se aumento na região Norte e Nordeste, variações na região Centro-Oeste e Sudeste, e diminuição na região Sul. No Norte e Nordeste, as inconsistências apresentaram percentuais de 0,40% e 0,65% em 2008 e 0,99% e 1,05% em 2016. Já a região Centro-Oeste variou entre diminuir e aumentar de 2008 (0,84%) a 2013 (0,86%) e a partir de 2014 (1,20%) teve um crescimento constante até 2016 (2,82%). Na região Sudeste há uma variação entre 1,22% em 2008 a 0,72% em 2012, já a partir de 2013 (0,84%) até 2016 (1,40%) ocorreu aumento das inconsistências. Na região Sul, nota-se um aumento entre 2008 (1,76%) a 2010 (2,47%), entretanto, de 2011(2,34%) a 2016 (1,06%) ocorre uma diminuição das inconsistências nessa combinação. A maior inconsistência dessa combinação foi vista em 2016 (2,82%) no Centro-Oeste (Tabela 16).

Na combinação de TNM & Estadiamento, a região Norte mostrou diminuição das inconsistências entre 2008 (1,49%) a 2010 (1,31%) e um crescimento de 2011 (1,42%) a 2016 (3,31%), no Nordeste as inconsistências foram crescendo e diminuindo ao longo dos anos, o percentual de inconsistência em 2008 foi de 0,86%, em 2012 foi de 1,79% e em 2016 de 1,45%. Na região Centro-Oeste ocorre um aumento entre 2008 (0,39%) e 2011 (1,51%) e uma diminuição entre 2014 (1,53%) e 2016 (1,18%). Já na região Sudeste ocorre uma subida das inconsistências entre 2012 (0,47%) até 2016 (0,91%). A região Sul dispõe de um aumento constante nos percentuais de inconsistência ao longo dos anos, atingindo em 2008 (0,54%) e em 2016 (1,20%). O maior percentual de inconsistência nessa combinação foi na região Norte em 2016 de 3,31% (Tabela 16).

Tabela 16 - Proporção (%) de inconsistências em variáveis combinadas da ficha de registro do tumor para Câncer do Colo do Útero no Brasil e Grandes Regiões, 2008 a 2016

Inconsistências		Câncer do colo do útero (%)								
		2008	2009	2010	2011	2012	2013	2014	2015	2016
Brasil	Data da primeira consulta < Data do diagnóstico	3,13	2,75	2,91	2,92	2,84	3,10	3,29	3,33	3,50
	Data da primeira consulta > Data de início do tratamento	0,01	0,02	0,01	0,01	0,03	0,01	0,01	0,02	0,02
	Data do diagnóstico > Data do início do tratamento	0,22	0,20	0,28	0,25	0,37	0,30	0,40	0,58	0,76
	Sexo & Localização primária do tumor	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Localização primária do tumor & Tipo histológico	0,97	1,03	1,13	1,08	1,11	1,11	1,22	1,16	1,21
	TNM & Estadiamento	0,89	0,93	1,09	1,14	1,36	1,38	1,43	1,48	1,50
Norte	Data da primeira consulta < Data do diagnóstico	1,97	1,89	2,35	2,64	2,36	2,16	1,99	1,44	1,83
	Data da primeira consulta > Data de início do tratamento	0,02	0,04	0,06	0,02	0,01	0,00	0,00	0,01	0,00
	Data do diagnóstico > Data do início do tratamento	0,35	0,41	0,45	0,51	0,36	0,35	0,48	0,36	0,73
	Sexo & Localização primária do tumor	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Localização primária do tumor & Tipo histológico	0,40	0,37	0,66	0,68	0,78	0,63	0,81	0,80	0,99
	TNM & Estadiamento	1,49	1,37	1,31	1,42	2,00	2,90	3,27	3,21	3,31
Nordeste	Data da primeira consulta < Data do diagnóstico	4,38	3,45	3,29	3,18	2,97	3,36	3,61	3,53	3,61
	Data da primeira consulta > Data de início do tratamento	0,01	0,02	0,01	0,01	0,01	0,00	0,01	0,01	0,02
	Data do diagnóstico > Data do início do tratamento	0,24	0,16	0,28	0,26	0,38	0,25	0,41	0,59	0,65
	Sexo & Localização primária do tumor	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Localização primária do tumor & Tipo histológico	0,65	0,55	0,72	0,70	0,76	0,88	0,97	1,02	1,05
	TNM & Estadiamento	0,86	1,07	1,57	1,59	1,79	1,60	1,46	1,58	1,45
Centro-Oeste	Data da primeira consulta < Data do diagnóstico	3,02	3,16	3,49	3,12	1,94	2,18	2,20	6,72	7,08
	Data da primeira consulta > Data de início do tratamento	0,02	0,00	0,00	0,06	0,33	0,00	0,00	0,08	0,06
	Data do diagnóstico > Data do início do tratamento	0,02	0,00	0,18	0,16	0,22	0,10	0,16	1,31	2,27
	Sexo & Localização primária do tumor	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Localização primária do tumor & Tipo histológico	0,84	0,88	0,78	0,76	0,98	0,86	1,20	1,94	2,82
	TNM & Estadiamento	0,39	0,20	0,96	1,51	1,49	0,59	1,53	1,47	1,18
Sudeste	Data da primeira consulta < Data do diagnóstico	1,87	2,15	1,98	1,87	1,69	2,22	2,60	2,67	2,65
	Data da primeira consulta > Data de início do tratamento	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,01	0,04
	Data do diagnóstico > Data do início do tratamento	0,35	0,32	0,19	0,20	0,26	0,27	0,38	0,66	0,82
	Sexo & Localização primária do tumor	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01
	Localização primária do tumor & Tipo histológico	1,22	1,29	0,94	0,84	0,72	0,84	0,94	1,23	1,40
	TNM & Estadiamento	1,10	1,21	0,57	0,39	0,47	0,46	0,62	0,60	0,91
Sul	Data da primeira consulta < Data do diagnóstico	2,63	2,40	3,18	3,49	4,21	4,24	4,38	3,82	4,20
	Data da primeira consulta > Data de início do tratamento	0,01	0,01	0,00	0,02	0,04	0,01	0,02	0,03	0,02
	Data do diagnóstico > Data do início do tratamento	0,02	0,11	0,32	0,16	0,49	0,46	0,40	0,45	0,56
	Sexo & Localização primária do tumor	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	Localização primária do tumor & Tipo histológico	1,76	2,18	2,47	2,34	2,35	2,12	2,15	1,27	1,06
	TNM & Estadiamento	0,54	0,31	0,60	0,71	0,91	1,08	1,00	1,12	1,20

Fonte: Elaborado pelo autor com dados do IntegradorRHC.

### 5.3 SCRIPT PARA ANÁLISE DA QUALIDADE DOS DADOS

Foi criado um script visando a disponibilização gratuita e de fácil acesso à análises de completude e inconsistência para dados dos SisRHC, com ênfase nos dados de câncer de mama e de colo do útero.

Os passos para a criação do script se deram a partir de:

1. Ativação dos pacotes necessários para análise;
2. Junção das bases de dados referente a cada ano analisado;
3. Manipulação da base de dados (mudança do nome nas variáveis, classificação das variáveis e datas, transferência das inequivalências da variável para o registro de ignorado, criação das variáveis completude e incompletude, redistribuição dos casos entre os estados em uma região correspondente, que são: Norte, Nordeste, Centro-Oeste, Sudeste e Sul);
4. Tabulação dos dados.

O script baseou-se na análise dos dados de câncer de mama e de colo do útero, entretanto, é possível a sua utilização na avaliação de qualidade de dados de outros tipos de câncer desde que ocorra a correção nas variáveis clínicas específicas. Além disso, o script pode ser utilizado para auxiliar na avaliação de completude e inconsistência de diferentes Sistemas de Informação em Saúde, bem como o Sistema de Informações sobre Mortalidade (SIM), Sistema de Informação de Agravos de Notificação (SINAN), Sistema de Informações sobre Nascidos Vivos (SINASC), Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP) e entre outros.

## 6 DISCUSSÃO

As informações em saúde demonstram uma importante ferramenta na gestão da atenção e controle de doenças e vetores que influenciam na saúde humana, entretanto, é de grande valia uma informação de qualidade, precisa e representativa da população. Dessa forma, o câncer por se tratar de uma das principais causas de morbimortalidade no Brasil e no mundo deve passar por avaliações periódicas que identifiquem as fragilidades no Sistema de Informação de Registros Hospitalares de Câncer (SisRHC).

É importante frisar que os Registros Hospitalares de Câncer dispõem de informações de casos de câncer tratados em hospitais referenciados pelo SUS, portanto, permitem o estudo da assistência oncológica pública (BORGES *et al.*, 2020). Dessa maneira, o estudo da qualidade de dados do SisRHC possibilitará identificar além das fragilidades nos RHC e no cuidado ao paciente oncológico, os principais pontos de necessidade de atuação para a melhoria da qualificação do preenchimento.

Na completude do câncer de mama para o Brasil, notou-se que o Bloco de Identificação do paciente apresenta completude Ruim e Regular para Escolaridade. O bloco de Caracterização do tumor apresentou preenchimento Ruim das variáveis Estadiamento Clínico do tumor e TNM. Essas variáveis são de extrema importância para o seguimento do tratamento, visto que o Estadiamento clínico do tumor e o TNM influenciam no desenvolvimento da doença (PINTO *et al.*, 2012). Na Caracterização do 1º tratamento, completude Ruim foi observada no preenchimento da variável Estado da doença ao final do primeiro tratamento.

Com relação às diferenças regionais no câncer de mama, observou-se o preenchimento Ruim e Regular na variável Raça/Cor da pele e Escolaridade (Bom, Ruim e Regular) na região Norte. O bloco de Caracterização do tumor seguiu o mesmo padrão do Brasil com completude Ruim.

No Nordeste, constatou-se o preenchimento Ruim da variável Escolaridade no bloco de Identificação do paciente. O bloco de Caracterização do tumor demonstrou completude Ruim no Estadiamento Clínico do tumor e TNM. Na Caracterização do 1º tratamento, a variável Estado da doença ao final do primeiro tratamento obteve completude Ruim.

Na região Centro-Oeste a completude do bloco de Identificação do paciente foi de preenchimento Regular, Bom e Ruim na variável Raça/cor da pele, já em Escolaridade, o preenchimento foi Ruim e Muito ruim. Entre as variáveis do bloco de Caracterização do tumor, a completude foi Ruim em Estadiamento Clínico do tumor e o TNM apresentou

completude Ruim e Muito ruim. No bloco de Caracterização do 1º tratamento, o preenchimento foi Ruim e Regular na variável Estado da doença ao final do primeiro tratamento.

A região Sudeste dispôs de preenchimento Ruim da variável Escolaridade no bloco de Identificação do paciente. A completude foi Regular e Ruim na variável Estadiamento Clínico do tumor e Ruim para o TNM no bloco de Caracterização do diagnóstico. Já no bloco de Caracterização do 1º tratamento, a variável Principal razão para a não realização do tratamento antineoplásico no hospital obteve completude Regular, Ruim e Excelente, e na variável Estado da doença ao final do primeiro tratamento, a completude foi Ruim.

Na região Sul a completude foi Ruim e Regular na variável Escolaridade do bloco de Identificação do paciente. Entre as variáveis de Caracterização do primeiro diagnóstico, a variável Estadiamento Clínico do tumor apresentou completude Ruim em quase todo o período e Regular em 2009 e 2010, e na variável TNM, a completude do preenchimento foi Ruim. Para o bloco de Caracterização do 1º tratamento, a Principal razão para a não realização do tratamento antineoplásico no hospital seguiu-se entre completude Regular, Ruim e Bom e o Estado da doença ao final do tratamento teve preenchimento Ruim.

No que tange à completude para câncer do colo do útero, o Brasil deteve preenchimento Ruim e Regular para Escolaridade no bloco de Identificação do paciente. O bloco de Caracterização do tumor contou com completude Regular e Ruim para Estadiamento Clínico do tumor e Ruim para o TNM. Entre as variáveis de Caracterização do 1º tratamento, o Estado final da doença ao final do primeiro tratamento obteve completude Ruim.

Para a região Norte, notou-se completude Ruim e Regular em Raça/cor da pele e em Escolaridade preenchimento Bom, Ruim e Regular. No bloco de Caracterização do tumor, o Estadiamento Clínico do tumor atingiu preenchimento Bom, Regular e Ruim e o TNM alcançou preenchimento Regular, Ruim e Muito ruim.

Na região Nordeste observou-se completude Ruim e Regular para a variável Escolaridade do bloco de Identificação do paciente. Nas variáveis do bloco de Caracterização do diagnóstico, a variável Estadiamento Clínico do tumor dispôs de completude Regular e Ruim e o TNM de completude Ruim. No bloco de caracterização do 1º tratamento, notou-se completude Ruim e Regular para o Estado da doença ao final do tratamento.

A região Centro-Oeste obteve completude de grau Bom, Regular e Ruim em Raça/cor da pele e completude Ruim e Muito ruim em escolaridade. O bloco de Caracterização do diagnóstico contou com preenchimento Ruim e Muito ruim do Estadiamento Clínico do tumor

e preenchimento Muito ruim do TNM. No bloco de caracterização do 1º tratamento, percebeu-se completude Ruim em estado da doença ao final do primeiro tratamento.

Na região Sudeste, a completude da variável Escolaridade do bloco de Identificação do paciente foi Regular e Ruim. O preenchimento das variáveis do bloco de Caracterização do diagnóstico foi Regular e Ruim para o Estadiamento Clínico do tumor e preenchimento Ruim e Muito ruim do TNM. No bloco de caracterização do 1º tratamento constatou-se preenchimento Regular, Ruim e Excelente na variável Principal razão para a não realização do tratamento antineoplásico no hospital e a variável Estado da doença ao final do primeiro tratamento contaram com preenchimento Regular e Ruim.

Na região Sul, a completude foi Ruim e Regular para a variável escolaridade do bloco de Identificação do paciente. Constatou-se no bloco de Caracterização do diagnóstico completude de preenchimento Regular e Ruim em Estadiamento Clínico do tumor e Ruim em TNM. Entre as variáveis do bloco de caracterização do 1º tratamento, a Razão para não tratar obteve preenchimento Regular, Ruim, Bom e Excelente e o preenchimento da completude do Estado da doença ao final do primeiro tratamento foi Ruim.

Com relação à completude, constatou-se que ocorrem diferenças regionais no preenchimento das variáveis quando comparadas com o Brasil e as demais regiões, todavia, as variáveis com mau preenchimento entre as regiões seguem sendo similares às observadas no Brasil para câncer de mama e colo do útero. Destaca-se que dentre as regiões, a região Centro-Oeste apresentou as piores completudes entre ambos os cânceres estudados.

Os achados de nosso estudo corroboram estudos anteriores que realizaram qualidade dos dados voltados para a completude de Registros Hospitalares de Câncer no Brasil (PINTO *et al.*, 2012, OLIVEIRA *et al.*, 2021, D’ALESSANDRO; ANTONIAZZI, 2010, CARDOSO, 2018, CHAFFIN, 2016, LOPES-JÚNIOR *et al.*, 2022, OLIVEIRA *et al.*, 2017, COSTA, 2016, KESKE, 2021).

No estudo de Pinto *et al.* (2012) evidenciou-se completude Ruim para as variáveis Grau de instrução (31,2%), Estadiamento (29,7%), TNM (49,8%) e Estado da doença ao final do primeiro tratamento (21,3%), sendo o pior grau de completude observado no TNM. O resultado encontrado é similar em nosso estudo para TNM com os piores percentuais de completude tanto para câncer de mama quanto para câncer do colo do útero. É importante frisar que a pior completude de TNM localizou-se na análise de câncer do colo do útero.

O estudo de Oliveira *et al.* (2021) encontrou completude Muito ruim para Escolaridade (57,53%) e TNM (77,22%) e Ruim para Estadiamento (42,17%) e doença ao final do primeiro tratamento (46,95%). A pior completude foi encontrada em TNM (77,22%), desse modo,

corroborando com o encontrado em nosso estudo. Esse estudo subdividiu a análise entre Cuiabá e interior do Mato Grosso observado que as piores completudes de dados estavam no interior. Já no estudo de D'Alessandro e Antoniazzi (2010), os piores percentuais de casos sem informação entre as variáveis obrigatórias estavam em Escolaridade (69%), TNM e pTNM conjuntamente (35%) e Estadiamento (26%).

Em Cardoso (2018) observou-se que nas variáveis obrigatórias, as piores completude estavam em TNM (70%), Ocupação (66%), pTNM (58%) e Estadiamento (52%). Nesse estudo a variável Instrução (21%) foi classificada como Ruim, o resultado para Instrução foi de encontro ao nosso estudo com Escolaridade classificada como Ruim e algumas vezes Regular em ambos os cânceres.

Na variável Estado da doença ao final do tratamento, nosso estudo observou completude Ruim em ambos os cânceres para o Brasil. Esse resultado vai de encontro ao observado em estudos anteriores (CHAFFIN, 2016; PINTO *et al.*, 2012; OLIVEIRA *et al.*, 2017, 2021).

O estudo de Lopes-Júnior *et al.* (2022) encontrou nas variáveis obrigatórias completude Ruim para as variáveis escolaridade (37,89%), ocupação (20,00%) e estadiamento clínico do tumor (45,65%) e Muito ruim para TNM (67,84%). No estudo de Costa (2016), observou-se completude Ruim nas variáveis Raça/cor (47,9%) e Estado da doença ao final do primeiro tratamento (30,4%) e Muito ruim em Escolaridade (72,8%) e Ocupação principal (63,3%), Estadiamento (78,6%), TNM (77%).

No estudo de Oliveira *et al.* (2017) as maiores proporções de dados incompletos estavam nas variáveis Escolaridade (35,5%) e Estadiamento TNM (36,6%). Em Keske (2021), as variáveis Estadiamento clínico do tumor (Ruim) e Estado da doença ao final do primeiro tratamento no hospital apresentaram completude Ruim e nas variáveis Ocupação e TNM foi Muito ruim.

Os resultados encontrados em nosso estudo são similares aos encontrados em estudos anteriores, com exceção da variável Escolaridade, visto que em comparação com alguns estudos, a Escolaridade foi classificada como Muito ruim e em nosso estudo foi Ruim.

O preenchimento da variável Escolaridade é importante devido a necessidade de aferição das condições socioeconômicas na falta de um dado que representativo socioeconomicamente, desse modo, a escolaridade pode indicar a desigualdade social e atuar como um proxy da situação socioeconômica dos indivíduos (KESKE, 2021; LOPES-JÚNIOR *et al.*, 2022). A Escolaridade obteve completude Ruim no câncer de mama e oscilou entre Ruim e Regular para câncer do colo do útero no Brasil, e entre as regiões, variou entre Bom, Regular, Ruim e Muito



ruim, destaca-se que os piores percentuais de completude dessa variável foram na Região Centro-Oeste para ambos os cânceres.

Nosso estudo relatou ausência de informações em diversos blocos da ficha de registro do tumor, sendo as maiores proporções encontradas nas variáveis TNM e Estado da doença ao final do tratamento para o Brasil e entre as regiões notou-se que os piores graus de completude estavam nas variáveis TNM e Estadiamento clínico do tumor para o Norte, TNM e Estado da doença ao final do tratamento para o Nordeste, TNM e Escolaridade para o Centro-Oeste, TNM e Escolaridade para o Sudeste e TNM e Estado da doença ao final do tratamento para o Sul para câncer de mama. Para o câncer do colo do útero, as piores completudes estavam nas variáveis TNM e Estado da doença ao final do tratamento para o Brasil, já entre as regiões, as maiores ausências de informações estavam em TNM e Raça/Cor da pele no Norte, TNM e Estado da doença ao final do tratamento para o Nordeste, TNM, Estadiamento clínico do tumor e Escolaridade no Centro-Oeste, TNM e Estado da doença ao final do tratamento para o Sudeste e TNM e Estado da doença ao final do tratamento no Sul.

Em sua maioria as piores completudes alocaram-se nas variáveis TNM e Estadiamento clínico do tumor para os dois tipos de câncer, todavia, a informação contida nessas variáveis permite conhecer a extensão da doença ao diagnosticar, e, assim, realizar a avaliação do tratamento que será administrado ao paciente oncológico e padronizar os procedimentos e a troca de experiências entre os UNACONs e CACONs. Além disso, o estadiamento clínico do tumor possibilita avaliar a assistência prestada na unidade hospitalar e atua como base na criação de políticas públicas de saúde voltadas para o diagnóstico precoce do câncer (OLIVEIRA *et al.*, 2021; LOPES-JÚNIOR *et al.*, 2022). O registro do estadiamento clínico do tumor é uma informação que não pode ser resgatada posteriormente, desse modo, é preciso o preenchimento correto e imediato no registro do paciente (OLIVEIRA *et al.*, 2017).

Vale salientar que o preenchimento Ruim da variável Escolaridade não está restrito apenas ao SisRHC, diversos outros estudos que trabalharam com qualidade de dados no Sistema de Informação sobre Mortalidade (SIM) e Sistema Nacional de Agravos de Notificação (SINAN) constataram preenchimento Ruim da variável Escolaridade (FELIX *et al.*, 2012; BARBUSCIA; RODRIGUES-JÚNIOR, 2011; MARQUES; SIQUEIRA; PORTUGAL, 2020; MUZY; CASTANHEIRA; ROMERO, 2021; SOUSA *et al.*, 2020).

O mau preenchimento dos registros hospitalares de câncer para a variável Raça/Cor da pele na região Norte para câncer do colo do útero chama atenção por ser uma variável não ligada somente ao contexto biológico, mas também um marcador social de desigualdade. Inúmeros estudos realizados evidenciam as diferenças dos óbitos por câncer entre mulheres

negras quando comparado a outras etnias (OLIVEIRA *et al.*, 2021). Ademais, as desigualdades sociais provocadas pela Raça/Cor da pele provocam dificuldades de acesso ao diagnóstico e tratamento para o câncer, portanto, essa é uma variável que a completude deve abranger um grau “Bom” ou “Excelente”.

A maior inconsistência para câncer de mama foi na Data da primeira consulta < Data do diagnóstico atingindo o ápice em 2016, mas não ultrapassando o valor de 5%, tanto para o Brasil quanto para Regiões. O câncer do colo do útero seguiu o mesmo padrão que o observado em câncer de mama, com a maior inconsistência sendo localizada em Data da primeira consulta < Data do diagnóstico e atingindo o ápice em 2016 e não ultrapassando o valor de 7%. A região Norte diferenciou-se das outras devido a sua maior inconsistência ser em TNM & Estadiamento (3,31%) no ano de 2016. Salienta-se que dentre a análise, a região Centro-Oeste obteve as maiores inconsistências na Data da primeira consulta < Data do diagnóstico em ambos os cânceres. O resultado encontrado demonstra que as inconsistências no SisRHC são baixas variando de acordo o critério de Romero & Cunha entre grau “Excelente” e “Bom”.

Ressalta-se que é de responsabilidade do registrador garantir a ficha de registro do tumor e os procedimentos realizados sejam documentados de forma precisa e oportuna para garantir um dado de qualidade (D’ALESSANDRO; ANTONIAZZI, 2010).

## 7 CONSIDERAÇÕES FINAIS

A qualidade dos dados do SisRHC para o Brasil e Grandes Regiões em que tange a completude e inconsistência, se mostrou heterogênea entre si, com os piores percentuais localizados na região Centro-Oeste. Todavia, constatou-se que a base de dados do SisRHC apesar de apresentar incompletude e inconsistência em alguns campos, pode servir como subsídio para a vigilância e controle do câncer no Brasil.

A ausência de informações em alguns campos pode estar relacionada a organização do serviço nos RHC, digitação incorreta, dificuldade no entendimento do sistema e inexistência de um fluxo de comunicação entre o corpo clínico e os registradores. Dessa maneira, para que haja uma melhoria na qualidade da informação é necessária à criação de rotinas periódicas de avaliação da base de dados por parte dos registradores, uma maior interação entre os registradores e os profissionais de saúde responsáveis pelo cuidado do paciente oncológico, criação de cursos de capacitação para os registradores dos RHC que demonstrem o preenchimento de forma correta e análise crítica dos dados produzidos pelo RHC.

A garantia de informações de qualidade sobre o câncer nos Sistemas de Informação em Saúde deve ser uma meta permanente para a manutenção da atenção ao câncer no Brasil. Para tanto, buscamos produzir um *script* no *Software R*, que é um software gratuito e de fácil acesso, que aborde os aspectos relacionados a completude e inconsistência das informações produzidas em RHC para os profissionais de saúde e gestores hospitalares com objetivo de contribuir com a melhoria da qualidade da vigilância do câncer nos estados e municípios do Brasil, bem como, direcionar com base nas informações encontradas estratégias mais eficientes para o controle da qualidade da informação sobre câncer no país.

Este estudo apresentou limitações devido ao uso de base de dados secundárias de domínio público, sem informações pessoais como o número de cartão do SUS, número do cadastro de pessoa física, nome da mãe e data de nascimento, portanto, não foi possível a utilização de técnicas de linkage com outras bases de dados para constatar o quanto a informação melhoraria após a vinculação das bases de dados. O SisRHC não identifica duplicidade no registro caso o indivíduo possua mais de um número de prontuário ou o nome tenha sido digitado errado, e, nesses casos, seria necessária uma abordagem de linkage de dados. Por fim, a base de dados de SisRHC do Brasil não está totalmente integrada entre si, desse modo, a informação está disponível de forma fragmentada, como é o caso do estado de São Paulo.

Contudo, mesmo com as limitações apontadas, os resultados encontrados na avaliação dos dados do SisRHC demonstram bom preenchimento e baixas inconsistências das variáveis obrigatórias da ficha de registro do tumor para o Brasil e Grandes Regiões, com exceção da escolaridade, TNM, estadiamento e estado final da doença. A utilização destes dados deve ser estimulada e o aprimoramento das informações dos RHC é essencial para a qualidade do sistema de informação em câncer, e, em especial para assegurar a melhoria da assistência oncológica na rede de serviços do SUS.

Além disso, recomenda-se a criação de análises futuras que contemplem o estudo de completude e inconsistências para os RHC do Brasil e Grandes Regiões levando em conta o nome das unidades hospitalares, dado que possibilitará a identificação de estabelecimentos hospitalares com fragilidades no preenchimento correto da informação, portanto, será possível elaborar estratégias para a melhoria da informação nestes locais.

## REFERÊNCIAS

- AGRANONIK, Marilyn; JUNG, Renata Oliveira. Qualidade dos sistemas de informações sobre nascidos vivos e sobre mortalidade no Rio Grande do Sul, Brasil, 2000 a 2014. **Ciência & Saúde Coletiva**, v. 24, p. 1945–1958, 2019. DOI: <https://doi.org/10.1590/1413-81232018245.19632017>.
- BALLANTINE, Kirsten R. *et al.* Childhood cancer registration in New Zealand: A registry collaboration to assess and improve data quality. **Cancer Epidemiology**, v. 55, p. 104–109, 2018.
- BARBOSA, Isabelle Ribeiro *et al.* Mortalidade por câncer de mama nos estados do nordeste do Brasil: tendências atuais e projeções até 2030. **Revista Ciência Plural**, v. 1, n. 1, p. 04–14, 2015. Disponível em: <https://periodicos.ufrn.br/rcp/article/view/7318>. Acesso em: 16 jul. 2022.
- BARBOSA, Isabelle Ribeiro *et al.* Desigualdades regionais na mortalidade por câncer de colo de útero no Brasil: tendências e projeções até o ano 2030. **Ciência & Saúde Coletiva**, v. 21, p. 253–262, 2016. DOI: <https://doi.org/10.1590/1413-81232015211.03662015>.
- BARBOSA, Jakeline Ribeiro *et al.* Avaliação da qualidade dos dados, valor preditivo positivo, oportunidade e representatividade do sistema de vigilância epidemiológica da dengue no Brasil, 2005 a 2009. **Epidemiologia e Serviços de Saúde**, v. 24, n. 1, p. 49–58, 2015. DOI: <https://doi.org/10.5123/S1679-49742015000100006>.
- BARBUSCIA, Denise Machado; RODRIGUES-JÚNIOR, Antonio Luiz. Completude da informação nas Declarações de Nascido Vivo e nas Declarações de Óbito, neonatal precoce e fetal, da região de Ribeirão Preto, São Paulo, Brasil, 2000-2007. **Cadernos de Saúde Pública**, v. 27, p. 1192–1200, 2011. DOI: <https://doi.org/10.1590/S0102-311X2011000600016>.
- BARROS, Natalia Gonçalves Meneses. **Registros de câncer: uma necessidade de vigilância em saúde para área de influência ambiental da Refinaria Abreu e Lima**. 2010. 62 f. Trabalho de Conclusão de Curso (Especialização em Saúde Pública) – Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2010.
- BATISTA, Francinne Laureth. **Avaliação da qualidade dos dados, oportunidade e representatividade do Sistema de Vigilância da COVID-19 na Macrorregião de Saúde Jequitinhonha, Minas Gerais, 2020**. 2021. 150 f. Dissertação (Mestrado em Ensino em Saúde) – Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, 2021. Disponível em: <http://acervo.ufvjm.edu.br/jspui/handle/1/2729>. Acesso em: 13 jul. 2022.
- BORGES, Anne Karin da Mota *et al.* Câncer de tireoide no Brasil: estudo descritivo dos casos informados pelos registros hospitalares de câncer, 2000-2016\*. **Epidemiologia e Serviços de Saúde**, v. 29, n. 4, 2020. DOI: <https://doi.org/10.5123/S1679-49742020000400012>.
- BRASIL. Ministério da Saúde. Departamento de Ciência e Tecnologia, Secretaria de Ciência e Tecnologia e Insumos Estratégicos do Ministério da Saúde. Integração de informações dos registros de câncer brasileiros. **Revista de Saúde Pública**, v. 41, p. 865–868, 2007. Disponível em: <https://www.scielosp.org/article/rsp/2007.v41n5/865-868/pt/>. Acesso em: 13 jul. 2022.

BRAZ, Rui Moreira *et al.* Avaliação da completude da variável raça/cor nos sistemas nacionais de informação em saúde para aferição da equidade étnico-racial em indicadores usados pelo Índice de Desempenho do Sistema Único de Saúde. **Saúde em Debate**, v. 37, p. 554–562, 2013. Disponível em: <http://www.scielo.br/j/sdeb/a/ZqDr6yqgFryL5zXqCyrLVLc/abstract/?lang=pt>. Acesso em: 13 jul. 2022.

BRAY, Freddie; PARKIN, D. Max. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. **European Journal of Cancer**, v. 45, n. 5, p. 747–755, 2009. DOI: <https://doi.org/10.1016/j.ejca.2008.11.032>.

BRENNAN, Paul. Gene–environment interaction and aetiology of cancer: what does it mean and how can we measure it? **Carcinogenesis**, v. 23, n. 3, p. 381–387, 2002. DOI: <https://doi.org/10.1093/carcin/23.3.381>.

CARDOSO, Andreza Lima. **Avaliação da qualidade das informações do registro hospitalar de câncer do município de Santa Maria, Rio Grande do Sul, Brasil**. 2018. 60 f. Dissertação (Mestrado Profissional em Ciências da Saúde) – Universidade Federal de Santa Maria, Santa Maria, 2018. Disponível em: <http://repositorio.ufsm.br/handle/1/20850>. Acesso em: 13 jul. 2022.

CINTHO, Lilian Mie; MACHADO, Roni Rodrigues; MORO, Claudia Maria Cabral. Métodos para Avaliação de Sistema de Informação em Saúde. **Journal of Health Informatics**, v. 8, n. 2, 2016. Disponível em: <http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/346>. Acesso em: 19 jul. 2022.

CORREIA, Lourani Oliveira dos Santos; PADILHA, Bruna Merten; VASCONCELOS, Sandra Mary Lima. Métodos para avaliar a completude dos dados dos sistemas de informação em saúde do Brasil: uma revisão sistemática. **Ciência & Saúde Coletiva**, v. 19, p. 4467–4478, 2014. DOI: <https://doi.org/10.1590/1413-812320141911.02822013>.

COSTA, Ligia Maria Cantarino da. **Avaliação do sistema de vigilância epidemiológica da influenza no Brasil, 2010-2013**. 2015. 233 f. Tese (Doutorado em Ciências da Saúde) – Universidade de Brasília, Brasília, 2015. Disponível em: <https://repositorio.unb.br/handle/10482/19643>. Acesso em: 13 jul. 2022.

COSTA, Luciano Mesentier da. **Hospital Federal de Bonsucesso: ações estratégicas de gestão da informação para o registro de câncer**. 2016. 18 f. Monografia (Graduação em Administração Pública) – Universidade Federal Fluminense, Volta Redonda, 2016. Disponível em: <http://app.uff.br/riuff/handle/1/3365>. Acesso em: 2 dez. 2022.

COSTA, Juliana Martins Barbosa da Silva; FRIAS, Paulo Germano de. Avaliação da completude das variáveis da Declaração de Nascido Vivo de residentes em Pernambuco, Brasil, 1996 a 2005. **Cadernos de Saúde Pública**, v. 25, p. 613–624, 2009. Disponível em: <https://www.scielosp.org/article/csp/2009.v25n3/613-624/>. Acesso em: 13 jul. 2022.

D’ALESSANDRO, Thays Aparecida Leão; ANTONIAZZI, Berenice Navarro; ABREU, Daisy Maria Xavier de. Registros hospitalares de câncer de Minas Gerais: análise de consistência das bases de dados. **Cadernos Saúde Coletiva**, v. 18, n. 3, 2010. Disponível em: [http://www.cadernos.iesc.ufrj.br/cadernos/images/csc/2010\\_3/artigos/CSCv18n3\\_pag410-7.pdf](http://www.cadernos.iesc.ufrj.br/cadernos/images/csc/2010_3/artigos/CSCv18n3_pag410-7.pdf). Acesso em: 2 jan. 2023.

BRASIL. Ministério da Saúde. **Asis - Análise de Situação de Saúde**. Brasília: Ministério da Saúde, 2015. 282 p. v. 1. Disponível em: [https://bvsms.saude.gov.br/bvs/publicacoes/asis\\_analise\\_situacao\\_saude\\_volume\\_1.pdf](https://bvsms.saude.gov.br/bvs/publicacoes/asis_analise_situacao_saude_volume_1.pdf). Acesso em: 2 jan. 2023.

DIMITROVA, Nadya; PARKIN, Donald Maxwell. Data quality at the Bulgarian National Cancer Registry: An overview of comparability, completeness, validity and timeliness. **Cancer Epidemiology**, v. 39, n. 3, p. 405–413, 2015.

DUARTE, Daniela de Almeida Pereira; BUSTAMANTE-TEIXEIRA, Maria Teresa. Iniquidade social e mortalidade por câncer de mama e colo do útero: uma revisão integrativa social. **Revista de Pesquisa Cuidado é Fundamental Online**, p. 877–888, 2018. Disponível em: [http://www.seer.unirio.br/index.php/cuidadofundamental/article/viewFile/6154/pdf\\_1](http://www.seer.unirio.br/index.php/cuidadofundamental/article/viewFile/6154/pdf_1). Acesso em: 16 jul. 2022.

FELIX, Janaina Daumas *et al.* Avaliação da completude das variáveis epidemiológicas do Sistema de Informação sobre Mortalidade em mulheres com óbitos por câncer de mama na Região Sudeste: Brasil (1998 a 2007). **Ciência & Saúde Coletiva**, v. 17, p. 945–953, 2012. DOI: <https://doi.org/10.1590/S1413-81232012000400016>.

FRIAS, Paulo Germano de; SZWARCOWALD, Célia Landmann; LIRA, Pedro Israel Cabral de. Avaliação dos sistemas de informações sobre nascidos vivos e óbitos no Brasil na década de 2000. **Cadernos de Saúde Pública**, v. 30, n. 10, p. 2068–2280, 2014. DOI: <https://doi.org/10.1590/0102-311X00196113>.

FUNDAÇÃO ONCOCENTRO DE SÃO PAULO. RHC - Registro Hospitalar de Câncer **FOSP**, c2021. Disponível em: <http://www.fosp.saude.sp.gov.br/fosp/diretoria-adjunta-de-informacao-e-epidemiologia/rhc-registro-hospitalar-de-cancer/>. Acesso em: 19 jun. 2022.

GABRIEL, Guilherme Paiva *et al.* Avaliação das informações das Declarações de Nascidos Vivos do Sistema de Informação sobre Nascidos Vivos (Sinasc) em Campinas, São Paulo, 2009. **Revista Paulista de Pediatria**, v. 32, p. 183–188, 2014. DOI: <https://doi.org/10.1590/0103-0582201432306>.

GERMAN, R. R. *et al.* Updated guidelines for evaluating public health surveillance systems. **Morbidity and Mortality Weekly Report**, v. 50, n. RR-13, p. 1–35 July 2001. Disponível em: <https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm>. Acesso em: 13 jul. 2022.

GIRIANELLI, Vania Reis; FERREIRA, Aldo Pacheco; VIANNA, Marcos Besserman; *et al.* Qualidade das notificações de violências interpessoal e autoprovocada no Estado do Rio de Janeiro, Brasil, 2009-2016. **Cadernos Saúde Coletiva**, v. 26, p. 318–326, 2018. DOI: <https://doi.org/10.1590/1414-462X201800030075>.

GIRIANELLI, Vania Reis; GAMARRA, Carmen Justina; AZEVEDO E SILVA, Gulnar. Disparities in cervical and breast cancer mortality in Brazil. **Revista de Saúde Pública**, v. 48, p. 459–467, 2014. DOI: <https://doi.org/10.1590/S0034-8910.2014048005214>.

GIRIANELLI, Vania Reis; THULER, Luiz Claudio Santos; SILVA, Gulnar Azevedo e. Qualidade do sistema de informação do câncer do colo do útero no estado do Rio de Janeiro. **Revista de Saúde Pública**, v. 43, p. 580–588, 2009. DOI: <https://doi.org/10.1590/S0034-89102009005000043>.

GRABOIS, Marília Fornaciari *et al.* Completude da Informação “Ocupação” nos Registros Hospitalares de Câncer do Brasil: Bases para a Vigilância do Câncer Relacionado ao Trabalho. **Revista Brasileira de Cancerologia**, v. 60, n. 3, p. 207–214, 2014. Disponível em: <https://rbc.inca.gov.br/index.php/revista/article/view/465>. Acesso em: 22 jun. 2022.

GROSECLOSE, Samuel L.; BUCKERIDGE, David L. Public health surveillance systems: recent advances in their use and evaluation. **Annual Review of Public Health**, v. 38, n. 1, p. 57–79, 2017. DOI: <https://doi.org/10.1146/annurev-publhealth-031816-044348>.

GUERRA, Maximiliano Ribeiro *et al.* Magnitude e variação da carga da mortalidade por câncer no Brasil e Unidades da Federação, 1990 e 2015. **Revista Brasileira de Epidemiologia**, v. 20, p. 102–115, 2017. DOI: <https://doi.org/10.1590/1980-5497201700050009>.

GUERRA, Maximiliano Ribeiro *et al.* Sobrevida por câncer de mama e iniquidade em saúde. **Cadernos de Saúde Pública**, v. 31, p. 1673–1684, 2015. DOI: <https://doi.org/10.1590/0102-311X00145214>.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Sistema IBGE de Recuperação Automática – SIDRA. Tabela 7358: População, por sexo e idade. Disponível em: <https://sidra.ibge.gov.br/tabela/7358#resultado>. Acesso em: 2 jan. 2023.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Mapas 2010**. [S. l.]: IBGE, 2018. Disponível em: [https://censos.ibge.gov.br/images/agenciadenoticias/ibge/2018\\_01/mapas-2010.png](https://censos.ibge.gov.br/images/agenciadenoticias/ibge/2018_01/mapas-2010.png). Acesso em: 2 jan. 2023.

INSTITUTO NACIONAL DE CÂNCER (Brasil). Detecção precoce. **INCA**, 16 set. 2022. Disponível em: <https://www.inca.gov.br/controlado-cancer-de-mama/acoes-de-controlado/deteccao-precoce>. Acesso em: 18 jul. 2022.

INSTITUTO NACIONAL DE CÂNCER (Brasil). O que é câncer? **INCA**, 31 maio 2022. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/o-que-e-cancer/o-que-e-cancer>. Acesso em: 9 out. 2022.

INSTITUTO NACIONAL DE CÂNCER (Brasil) (org.). **Informação dos Registros Hospitalares de Câncer como estratégia de transformação**: perfil do Instituto Nacional de Câncer José Alencar Gomes da Silva em 25 anos. Rio de Janeiro: INCA, 2012.

INSTITUTO NACIONAL DE CÂNCER (Brasil). **Registros hospitalares de câncer**: planejamento e gestão. [S. l.: s. n.], 2010. Disponível em: <https://www.inca.gov.br/publicacoes/manuais/registros-hospitalares-de-cancer>. Acesso em: 2 jan. 2023.



INSTITUTO NACIONAL DE CÂNCER (Brasil). **Registros hospitalares de câncer**: rotinas e procedimentos. [S. l.: s. n.], 2000. Disponível em: <https://www.inca.gov.br/publicacoes/manuais/registros-hospitalares-de-cancer>. Acesso em: 2 jan. 2023.

INSTITUTO NACIONAL DE CÂNCER (Brasil). **Deteção Precoce do Câncer**. [S. l.: s. n.], 2021. Disponível em: <https://www.inca.gov.br/publicacoes/livros/deteccao-precoce-do-cancer>. Acesso em: 2 jan. 2023.

INSTITUTO NACIONAL DE CÂNCER (Brasil). **Estimativa 2020**: incidência de câncer no Brasil. [S. l.: s. n.], 2019. Disponível em: <https://www.inca.gov.br/publicacoes/livros/estimativa-2020-incidencia-de-cancer-no-brasil>. Acesso em: 2 jan. 2023.

INSTITUTO NACIONAL DE CÂNCER (Brasil). **IntegradorRHC**: ferramenta para a vigilância hospitalar de câncer no Brasil. [S. l.: s. n.], 2011. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/folder-integrador-rhc-2011.pdf>. Acesso em: 2 jan. 2023.

KESKE, Greice de Lima. **Análise e proposta de gestão para a qualidade de informação do registro hospitalar de câncer do Hospital Universitário de Santa Maria – RS**. 2021. 120 f. Dissertação (Mestrado Profissional em Ciências da Saúde) – Universidade Federal de Santa Maria, Santa Maria, 2021. Disponível em: <http://repositorio.ufsm.br/handle/1/23575>. Acesso em: 2 dez. 2022.

KLIGERMAN, Jacob. Registro Hospitalar de Câncer no Brasil. **Revista Brasileira de Cancerologia**, v. 47, n. 4, p. 357–359, 2001. Disponível em: <https://rbc.inca.gov.br/index.php/revista/article/view/2302>. Acesso em: 13 jul. 2022.

LIMA, Claudia Risso de Araujo *et al.* Revisão das dimensões de qualidade dos dados e métodos aplicados na avaliação dos sistemas de informação em saúde. **Cadernos de Saúde Pública**, v. 25, n. 10, p. 2095–2109, 2009. DOI: <https://doi.org/10.1590/S0102-311X2009001000002>.

LINO, Renata Rodrigues Garcia *et al.* Tendência da incompletude das estatísticas vitais no período neonatal, estado do Rio de Janeiro, 1999-2014. **Epidemiologia e Serviços de Saúde**, v. 28, n. 2, 2019. DOI: <http://dx.doi.org/10.5123/s1679-49742019000200014>.

LOPES-JÚNIOR, Luís Carlos *et al.* Completeness and consistency of epidemiological variables from hospital-based cancer registries in a Brazilian State. **International Journal of Environmental Research and Public Health**, v. 19, n. 19, p. 12003, 2022. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9566550/>. Acesso em: 19 nov. 2022.

MADEIRO, Alberto; RUFINO, Andréa Cronemberger. Cobertura e fatores associados à não realização do exame citopatológico do colo do útero entre mulheres brasileiras de 18 a 39 anos. **Journal of Health & Biological Sciences**, v. 10, n. 1, p. 1–9, 2022. Disponível em: <https://periodicos.unichristus.edu.br/jhbs/article/view/3521>. Acesso em: 16 jul. 2022.

- MALTA, Deborah Carvalho; SILVA JR, Jarbas Barbosa da. O Plano de Ações Estratégicas para o Enfrentamento das Doenças Crônicas Não Transmissíveis no Brasil e a definição das metas globais para o enfrentamento dessas doenças até 2025: uma revisão. **Epidemiologia e Serviços de Saúde**, v. 22, n. 1, p. 151–164, 2013. DOI: <http://dx.doi.org/10.5123/S1679-49742013000100016>.
- MANICA, Silvia Troyahn; DRACHLER, Maria de Lourdes; TEIXEIRA, Luciana Barcellos; *et al.* Desigualdades socioeconômicas e regionais na cobertura de exames citopatológicos do colo do útero. **Revista Gaúcha de Enfermagem**, v. 37, 2016. DOI: <https://doi.org/10.1590/1983-1447.2016.01.52287>.
- MANSUR, Antonio de Padua; FAVARATO, Desiderio. Taxas de Mortalidade por Doenças Cardiovasculares e Câncer na População Brasileira com Idade entre 35 e 74 Anos, 1996-2017. **Arquivos Brasileiros de Cardiologia**, v. 117, p. 329–340, 2021. DOI: <https://doi.org/10.36660/abc.20200233>.
- MARQUES, Carla Adriana; SIQUEIRA, Marluce Mechelli de; PORTUGAL, Flávia Batista. Avaliação da não completude das notificações compulsórias de dengue registradas por município de pequeno porte no Brasil. **Ciência & Saúde Coletiva**, v. 25, p. 891–900, 2020. DOI: <https://doi.org/10.1590/1413-81232020253.16162018>.
- MELO, Enirtes Caetano Prates; TRAVASSOS, Claudia; CARVALHO, Marília Sá. Qualidade dos dados sobre óbitos por infarto agudo do miocárdio, Rio de Janeiro. **Revista de Saúde Pública**, v. 38, p. 385–391, 2004. Disponível em: <https://www.scielosp.org/article/rsp/2004.v38n3/385-391/>. Acesso em: 13 jul. 2022.
- MIGOWSKI, Arn *et al.* Diretrizes para detecção precoce do câncer de mama no Brasil. III - Desafios à implementação. **Cadernos de Saúde Pública**, v. 34, 2018. DOI: <https://doi.org/10.1590/0102-311X00046317>.
- MIGOWSKI, Arn *et al.* Diretrizes para detecção precoce do câncer de mama no Brasil. II - Novas recomendações nacionais, principais evidências e controvérsias. **Cadernos de Saúde Pública**, v. 34, p. e00074817, 2018. DOI: <https://doi.org/10.1590/0102-311X00074817>.
- MOHAMMADZADEH, Zeinab *et al.* Systematic Review of Hospital Based Cancer Registries (HBCRs): necessary tool to improve quality of care in cancer patients. **Asian Pacific Journal of Cancer Prevention : APJCP**, v. 18, n. 8, p. 2027–2033, 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5697455/>. Acesso em: 19 jul. 2022.
- MONTALVÃO, Elisa Alves. **Avaliação de atributos do Sistema de Vigilância Sentinela da Síndrome Gripal no Município do Rio de Janeiro, Brasil, 2013-2014**. 2017. 52 f. Dissertação (Mestrado em Saúde Pública) – Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz, Rio de Janeiro, 2017.
- MUZY, Jéssica; CASTANHEIRA, Débora; ROMERO, Dália. Análise da qualidade da informação da mortalidade prematura por doenças crônicas não transmissíveis e sua utilização nos Objetivos de Desenvolvimento Sustentável. **Cadernos Saúde Coletiva**, v. 29, p. 152–164, 2021. DOI: <https://doi.org/10.1590/1414-462X202199010456>.

- OLIVEIRA, Andrea Santos de *et al.* Registros Hospitalares de Câncer em Pernambuco: da Gestão ao Registro. **Revista Brasileira de Cancerologia**, v. 63, n. 1, p. 21–28, 2017. Disponível em: <https://rbc.inca.gov.br/index.php/revista/article/view/152>. Acesso em: 2 dez. 2022.
- OLIVEIRA, Jânia Cristiane de Souza *et al.* Registros Hospitalares de Câncer de Mato Grosso: análise da completitude e da consistência. **Cadernos Saúde Coletiva**, v. 29, n. 3, p. 330–343, 2021. DOI: <https://doi.org/10.1590/1414-462X202129030230>.
- OLIVEIRA, Max Moura de *et al.* Avaliação do Sistema de Informações sobre Nascidos Vivos. Brasil, 2006 a 2010. **Epidemiologia e Serviços de Saúde**, v. 24, p. 629–640, 2015. Disponível em: <https://www.scielo.org/article/ress/2015.v24n4/629-640/>. Acesso em: 13 jul. 2022.
- OLIVEIRA, Max Moura de *et al.* Cobertura de exame Papanicolaou em mulheres de 25 a 64 anos, segundo a Pesquisa Nacional de Saúde e o Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico, 2013. **Revista Brasileira de Epidemiologia**, v. 21, 2018. DOI: <https://doi.org/10.1590/1980-549720180014>.
- OLIVEIRA, Nayara Priscila Dantas *et al.* Completeness of cervical cancer staging information in Brazil: A national hospital-based study. **Cancer Epidemiology**, v. 79, p. 102191, 2022. DOI: <https://doi.org/10.1016/j.canep.2022.102191>.
- PARKIN, D. Max.; MUIR, C. S. Cancer incidence in five continents. Comparability and quality of data. **IARC Scientific Publications**, n. 120, p. 45–173, 1992.
- PARKIN, D Max; BRAY, Freddie. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. **European Journal of Cancer**, v. 45, n. 5, p. 756–764, 2009. DOI: <https://doi.org/10.1016/j.ejca.2008.11.033>.
- PEDRAZA, Dixis Figueroa. Qualidade do Sistema de Informações sobre Nascidos Vivos (SINASC): análise crítica da literatura. **Ciência & Saúde Coletiva**, v. 17, n. 10, p. 2729–2737, 2012. DOI: <https://doi.org/10.1590/S1413-81232012001000021>.
- PEDRAZA, Dixis Figueroa. Sistema de informações sobre nascidos vivos: uma análise da qualidade com base na literatura. **Cadernos Saúde Coletiva**, v. 29, n. 1, p. 143–152, 2021. DOI: <https://doi.org/10.1590/1414-462X202129010106>.
- PIÑEROS, Marion *et al.* The role and utility of population-based cancer registries in cervical cancer surveillance and control. **Preventive Medicine**, v. 144, Mar. 2021. DOI: <https://doi.org/10.1016/j.ypmed.2020.106237>.
- PIÑEROS, Marion *et al.* Essential TNM: a registry tool to reduce gaps in cancer staging information. **The Lancet Oncology**, v. 20, n. 2, p. e103–e111, 2019. DOI: [https://doi.org/10.1016/S1470-2045\(18\)30897-0](https://doi.org/10.1016/S1470-2045(18)30897-0).
- PIÑEROS, Marion *et al.* A global cancer surveillance framework within noncommunicable disease surveillance: making the case for population-based cancer registries. **Epidemiologic Reviews**, v. 39, n. 1, p. 161–169, 2017. DOI: <https://doi.org/10.1093/epirev/mxx003>.

PINTO, Isabella Vitral *et al.* Completude e consistência dos dados dos registros hospitalares de câncer no Brasil. **Cadernos Saúde Coletiva**, v. 20, n. 1, p. 113-120, 2012.

RODRIGUES, Anselmo Duarte; BUSTAMANTE-TEIXEIRA, Maria Teresa. Mortalidade por câncer de mama e câncer de colo do útero em município de porte médio da Região Sudeste do Brasil, 1980-2006. **Cadernos de Saúde Pública**, v. 27, p. 241–248, 2011. DOI: <https://doi.org/10.1590/S0102-311X2011000200005>.

ROMERO, Dalia E.; CUNHA, Cynthia Braga da. Avaliação da qualidade das variáveis epidemiológicas e demográficas do Sistema de Informações sobre Nascidos Vivos, 2002. **Cadernos de Saúde Pública**, v. 23, n. 3, p. 701–714, 2007. DOI: <https://doi.org/10.1590/S0102-311X2007000300028>.

ROMERO, Dalia E.; CUNHA, Cynthia Braga da. Avaliação da qualidade das variáveis sócio-econômicas e demográficas dos óbitos de crianças menores de um ano registrados no Sistema de Informações sobre Mortalidade do Brasil (1996/2001). **Cadernos de Saúde Pública**, v. 22, n. 3, p. 673–681, 2006. DOI: <https://doi.org/10.1590/S0102-311X2006000300022>.

ROMERO, Dalia Elena; MAIA, Leo; MUZY, Jessica. Tendência e desigualdade na completude da informação sobre raça/cor dos óbitos de idosos no Sistema de Informações sobre Mortalidade no Brasil, entre 2000 e 2015. **Cadernos de Saúde Pública**, v. 35, 2019. DOI: <https://doi.org/10.1590/0102-311X00223218>.

SÁ, Marcia Martiniano de Sousa e. **Registros Hospitalares de Câncer de São Paulo: produção e utilização de informações**. 2017. Dissertação (Mestrado em Medicina Preventiva) – Universidade de São Paulo, São Paulo, 2017. Disponível em: <http://www.teses.usp.br/teses/disponiveis/5/5137/tde-11052017-144425/>. Acesso em: 13 jul. 2022.

SANCHES, Ana Maria. **Acurácia dos dados de topografia, morfologia e estadiamento clínico dos tumores**. São Paulo: Fundação Oncocentro, 2009. [Slide de capacitação]. Disponível em: [http://fosp.saude.sp.gov.br:443/docs/rhc/V\\_Encontro/Apresentacao\\_05.pdf](http://fosp.saude.sp.gov.br:443/docs/rhc/V_Encontro/Apresentacao_05.pdf). Acesso em: 2 jan. 2023.

SCHMIDTMANN, Irene. Estimating completeness in cancer registries: comparing capture-recapture methods in a simulation study. **Biometrical Journal. Biometrische Zeitschrift**, v. 50, n. 6, p. 1077–1092, 2008.

SIGURDARDOTTIR, Lara Gudrun *et al.* Data quality at the Icelandic Cancer Registry: comparability, validity, timeliness and completeness. **Acta Oncologica (Stockholm, Sweden)**, v. 51, n. 7, p. 880–889, 2012.

SILVA, Aline Almeida da *et al.* Avaliação do Sistema de Vigilância do Programa Nacional de Imunizações - Módulo Registro do Vacinado, Brasil, 2017. **Epidemiologia e Serviços de Saúde**, v. 30, n. 1, p. e2019596, 2021. DOI: <https://doi.org/10.1590/S1679-49742021000100028>.

SILVA, Gabriela Drummond Marques da *et al.* Avaliação da qualidade dos dados, oportunidade e aceitabilidade da vigilância da tuberculose nas microrregiões do Brasil. **Ciência & Saúde Coletiva**, v. 22, n. 10, p. 3307–3319, 2017. DOI: <https://doi.org/10.1590/1413-812320172210.18032017>.

SILVA, Laura Pedroza da *et al.* Avaliação da qualidade dos dados do Sistema de Informações sobre Nascidos Vivos e do Sistema de Informações sobre Mortalidade no período neonatal, Espírito Santo, Brasil, de 2007 a 2009. **Ciência & Saúde Coletiva**, v. 19, n. 7, p. 2011–2020, 2014. DOI: <https://doi.org/10.1590/1413-81232014197.08922013>.

SILVA, Gulnar Azevedo e *et al.* Tendência da mortalidade por câncer nas capitais e interior do Brasil entre 1980 e 2006. **Revista de Saúde Pública**, v. 45, p. 1009–1018, 2011. DOI: <https://doi.org/10.1590/S0034-89102011005000076>.

SILVA, Gulnar Azevedo e *et al.* Mortalidade por câncer nas capitais e no interior do Brasil: uma análise de quatro décadas. **Revista de Saúde Pública**, v. 54, p. 126, 2020. Disponível em: <https://www.revistas.usp.br/rsp/article/view/180449>. Acesso em: 18 jul. 2022.

SOUSA, Cyntia Meneses de Sá *et al.* Incompletude do preenchimento das notificações compulsórias de violência - Brasil, 2011-2014. **Cadernos Saúde Coletiva**, v. 28, p. 477–487, 2020. DOI: <https://doi.org/10.1590/1414-462X202028040139>.

STEVANATO, Jaqueline Monte; GAÍVA, Maria Aparecida Munhoz; MATHIAS, Thais Aidar de Freitas. <b>Análise da qualidade do sistema de informações sobre nascidos vivos. **Ciência, Cuidado e Saúde**, v. 16, n. 2, 2017. DOI: <https://doi.org/10.4025/ciencuidsaude.v16i2.36509>.

SUNG, Hyuna *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA: A Cancer Journal for Clinicians**, v. 71, n. 3, p. 209–249, 2021. Disponível em: <https://acsjournals.onlinelibrary.wiley.com/doi/full/10.3322/caac.21660>. Acesso em: 16 jul. 2022.

TIENSOLI, Sabrina Daros. **Fatores individuais e contextuais associados ao rastreamento do câncer de mama e colo do útero**. 2021. 119 f. Dissertação (Doutorado em Enfermagem) – Universidade Federal de Minas Gerais, Belo Horizonte, 2021. Disponível em: <https://repositorio.ufmg.br/handle/1843/39775>. Acesso em: 2 jan. 2023.

TOMAZELLI, Jeane Glaucia *et al.* Rastreamento do câncer de mama no Brasil: uma avaliação da oferta e utilização da rede assistencial do Sistema Único de Saúde no período 2010-2012\*. **Epidemiologia e Serviços de Saúde**, v. 26, n. 4, p. 713–724, 2017. DOI: <http://dx.doi.org/10.5123/s1679-49742017000400004>.

TOMAZELLI, Jeane; RIBEIRO, Caroline Madalena; DIAS, Maria Beatriz Kneipp. Cobertura dos Sistemas de Informação dos Cânceres do Colo do Útero e de Mama no Brasil, 2008-2019. **Revista Brasileira de Cancerologia**, v. 68, n. 1, 2022. Disponível em: <https://rbc.inca.gov.br/revista/index.php/revista/article/view/1544>. Acesso em: 16 jul. 2022.

WANNER, Miriam *et al.* Indicators of data quality at the cancer registry Zurich and Zug in Switzerland. **BioMed Research International**, v. 2018, p. 7656197, 2018. DOI: 10.1155/2018/7656197.

WORLD HEALTH ORGANIZATION. **Global strategy to accelerate the elimination of cervical cancer as a public health problem**. Geneva: WHO, 2020.

WORLD HEALTH ORGANIZATION. **Health information systems and rehabilitation.** [S. l.: s. n.], 2017. Disponível em: <https://apps.who.int/iris/bitstream/handle/10665/354362/WHO-NMH-NVI-17.2-eng.pdf>. Acesso em: 2 jan. 2023.

ZANETTI, R. *et al.* Completeness and timeliness: Cancer registries could/should improve their performance. **European Journal of Cancer**, v. 51, n. 9, p. 1091–1098, 2015. DOI: 10.1016/.ejca.2013.11.040

**ANEXOS**







## ANEXO B – SCRIPT NA LINGUAGEM R PARA AVALIAÇÃO DE COMPLETUDE E INCONSISTÊNCIA NA BASE DE DADOS DO SISRHC PARA CÂNCER DE MAMA E COLO DO ÚTERO

```
### TCC ÉDNEI###
#ANÁLISE DE COMPLETUDE E INCOSISTENCIA DOS RHC PARA CANCER DE MAMA E COLO DO UTERO
# UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
# INSTITUTO DE ESTUDOS EM SAÚDE COLETIVA
# ELABORAÇÃO: ÉDNEI CÉSAR DE A. S. JUNIOR (IESC/UFRJ)
# BASE DE 2008-2016 BAIXADA NO DIA 04/09/2022

library(dplyr)
library(ggplot2)
library(rio)
library(kableExtra)
library(readxl)
library(stringr)
library(reshape2)
library(readr)
library(patchwork)
library(openxlsx)
library(tidyr)
library(descr)
library(glue)
library(tableone)
library(openxlsx)

# Aumentar o limite de memoria

memory.limit(9999999999)

# Diretorio

setwd("C:/Users/junio/Downloads/TCC - ÉDNEI/bases separadas RHC - 2008.2016")

# base de dados

RHC2008 <- import("rhc08.dbf")
RHC2009 <- import("rhc09.dbf")
RHC2010 <- import("rhc10.dbf")
```

```
RHC2011 <- import("rhc11.dbf")
RHC2012 <- import("rhc12.dbf")
RHC2013 <- import("rhc13.dbf")
RHC2014 <- import("rhc14.dbf")
RHC2015 <- import("rhc15.dbf")
RHC2016 <- import("rhc16.dbf")

# Juntando todas as bases

RHC.CCU <- rbind(RHC2008,RHC2009,RHC2010, RHC2011, RHC2012,
                RHC2013,RHC2014, RHC2015, RHC2016)

# FILTRANDO CASOS ANALITICOS

table(RHC.CCU$TPCASO) #1 CASOS ANALITICOS

RHC.CCU <- filter(RHC.CCU, TPCASO == "1")

# Filtrando para Cancer de Mama (C50) e Colo do Utero (C53)

table(RHC.CCU$LOCTUDET)

RHC.CCU <- filter(RHC.CCU, LOCTUDET == "C50" | LOCTUDET == "C53")

# Salvando em XLSX

write.xlsx(RHC.CCU, "base_filtrada.xlsx")

# Excluindo as bases anteriores

rm(RHC2008)
rm(RHC2009)
rm(RHC2010)
rm(RHC2011)
rm(RHC2012)
rm(RHC2013)
rm(RHC2014)
rm(RHC2015)
rm(RHC2016)
```

```
# Selecionando apenas as variáveis que iremos trabalhar
```

```
RHC.CCU <- read.xlsx("base_filtrada.xlsx")
```

```
##### Câncer do colo do útero#####
```

```
# Selecionando as variáveis que irei trabalhar
```

```
RHC.CCU <- RHC.CCU%>%
```

```
  select(SEXO, IDADE, RACACOR, INSTRUC, ESTADRES, DATAPRICON, DTDIAGNO, DIAGANT,  
  BASMAIMP,
```

```
    LOCTUDET, TIPOHIST, ESTADIAM, LATERALI, DATAINITRT, RZNTR, PRITRATH, ESTDFIMT,  
    DATAOBITO, TNM)
```

```
# Selecionando apenas cancer de colo do utero
```

```
RHC.CCU <- RHC.CCU%>%
```

```
  filter(LOCTUDET == "C53")
```

```
# Analisando a completude das variáveis
```

```
names(RHC.CCU)[1] <- "Sexo"
```

```
names(RHC.CCU)[2] <- "Idade"
```

```
names(RHC.CCU)[3] <- "Raça/Cor da pele"
```

```
names(RHC.CCU)[4] <- "Escolaridade"
```

```
names(RHC.CCU)[5] <- "Estado de residência"
```

```
names(RHC.CCU)[6] <- "Data da primeira consulta"
```

```
names(RHC.CCU)[7] <- "Data de diagnóstico"
```

```
names(RHC.CCU)[8] <- "Diagnóstico e tratamento anteriores"
```

```
names(RHC.CCU)[9] <- "Base mais importante para o diagnóstico do tumor"
```

```
names(RHC.CCU)[10] <- "Localização primária do tumor"
```

```
names(RHC.CCU)[11] <- "Tipo histológico"
```

```
names(RHC.CCU)[12] <- "Estadiamento clínico do tumor (TNM)"
```

```
names(RHC.CCU)[13] <- "Lateralidade do tumor"
```

```
names(RHC.CCU)[14] <- "Data de início do tratamento"
```

```
names(RHC.CCU)[15] <- "Principal razão para a não realização do tratamento antineoplásico no hospital"
```

```
names(RHC.CCU)[16] <- "Primeiro tratamento recebido no hospital"
```

```
names(RHC.CCU)[17] <- "Estado da doença ao final do tratamento"
```

```
names(RHC.CCU)[18] <- "Data de óbito"
```

```
names(RHC.CCU)[19] <- "TNM"
```

```
# Corrigindo a classificação das variáveis
```

```
RHC.CCU <- RHC.CCU%>%
  mutate(Sexo = as.factor(Sexo),
         #Idade = as.numeric(Idade),
         `Raça/Cor da pele` = as.factor(`Raça/Cor da pele`),
         `Estado de residência` = as.factor(`Estado de residência`),
         `Diagnóstico e tratamento anteriores` = as.factor(`Diagnóstico e tratamento anteriores`),
         `Base mais importante para o diagnóstico do tumor` = as.factor(`Base mais importante para o diagnóstico do
tumor`),
         `Localização primária do tumor` = as.factor(`Localização primária do tumor`),
         `Tipo histológico` = as.factor(`Tipo histológico`),
         `Estadiamento clínico do tumor (TNM)` = as.factor(`Estadiamento clínico do tumor (TNM)`),
         `Lateralidade do tumor` = as.factor(`Lateralidade do tumor`),
         `Principal razão para a não realização do tratamento antineoplásico no hospital` = as.factor(`Principal razão
para a não realização do tratamento antineoplásico no hospital`),
         `Primeiro tratamento recebido no hospital` = as.factor(`Primeiro tratamento recebido no hospital`),
         TNM = as.factor(TNM),
         `Estado da doença ao final do tratamento` = as.factor(`Estado da doença ao final do tratamento`))
```

```
# Transferindo todas as inequivalências (divergências no preenchimento de acordo com o dicionário) nas variáveis
factor para o 9
```

```
# Bloco de identificação do paciente
```

```
# Sexo, Raça/cor da pele, escolaridade e UF de procedencia
```

```
# Sexo - 1 = Masculino e 2 = Feminino | necessário criar o 9
```

```
summary(RHC.CCU$Sexo)
RHC.CCU$Sexo <- factor(RHC.CCU$Sexo, levels = c("0", "1", "2", "3", "9"))
table(RHC.CCU$Sexo)
RHC.CCU$Sexo[RHC.CCU$Sexo %in% c("0", "3")] <- "9"
RHC.CCU$Sexo <- factor(RHC.CCU$Sexo)
table(RHC.CCU$Sexo)
```

```
# Raça/cor da pele - 1 = Branca, 2 = Preta, 3 = Amarela, 4 = Parda, 5 = Indígena, 9 =Sem informação
```

```
summary(RHC.CCU`Raça/Cor da pele`)
RHC.CCU`Raça/Cor da pele`[RHC.CCU`Raça/Cor da pele` == "99"] <- "9"
```

```

RHC.CCU$`Raça/Cor da pele` <- factor(RHC.CCU$`Raça/Cor da pele`)
table(RHC.CCU$`Raça/Cor da pele`)

# Escolaridade - 1.Nenhuma; 2.Fundamental incompleto; 3.Fundamental completo; 4.Nível médio; 5.Nível
superior incompleto; 6.Nível superior completo; 9.Sem informação

summary(RHC.CCU$Escolaridade)
RHC.CCU$Escolaridade <- factor(RHC.CCU$Escolaridade)
RHC.CCU$Escolaridade[RHC.CCU$Escolaridade == "0"] <- "9"
RHC.CCU$Escolaridade <- factor(RHC.CCU$Escolaridade)
table(RHC.CCU$Escolaridade)

# Estado de Residência
# 77 e 99 inconsistencias - criar 9 (sem informação)
summary(RHC.CCU$`Estado de residência`)
RHC.CCU$`Estado de residência` <- factor(RHC.CCU$`Estado de residência`, levels = c("77", "99", "AC", "AL",
"AM", "AP", "BA", "CE", "DF", "ES", "GO", "MA",
"MG", "MS", "MT", "PA", "PB", "PE", "PI", "PR", "RJ",
"RN", "RO",
"RR", "RS", "SC", "SE", "SP", "TO", "99"))
table(RHC.CCU$`Estado de residência`)
RHC.CCU$`Estado de residência` [RHC.CCU$`Estado de residência` %in% c("77", "99")] <- "9"
RHC.CCU$`Estado de residência` <- factor(RHC.CCU$`Estado de residência`)
table(RHC.CCU$`Estado de residência`)

# Idade

RHC.CCU$Idade <- as.numeric(RHC.CCU$Idade)
table(RHC.CCU$Idade)
levels(RHC.CCU$Idade)

RHC.CCU$Idade <- factor(RHC.CCU$Idade, levels = c("-59", "0", "1", "4", "7", "9", "10", "12", "13", "14",
"15", "16", "17", "18", "19", "20", "21", "22", "23", "24",
"25", "26", "27", "28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38",
"39", "40", "41", "42", "43", "44",
"45", "46", "47", "48", "49", "50", "51", "52", "53", "54", "55", "56", "57", "58",
"59", "60", "61", "62", "63", "64",
"65", "66", "67", "68", "69", "70", "71", "72", "73", "74", "75", "76", "77", "78",
"79", "80", "81", "82", "83", "84",

```

```

"85", "86", "87", "88", "89", "90", "91", "92", "93", "94", "95", "96", "97", "98",
"99", "100", "101", "102", "103", "104",
"105", "106", "107", "108", "116", "118", "999"))

```

```

table(RHC.CCU$Idade)
RHC.CCU$Idade[RHC.CCU$Idade == "-59"] <- "999"
RHC.CCU$Idade <- as.numeric(RHC.CCU$Idade)
table(RHC.CCU$Idade)

```

```

# Bloco de caracterização do diagnóstico-----

```

```

#Data da primeira consulta, Data do primeiro diagnóstico, Diagnóstico e tratamento anteriores e
#Base mais importante para o diagnóstico do tumor

```

```

# Data da primeira consulta

```

```

RHC.CCU$`Data da primeira consulta` <- as.Date(RHC.CCU$`Data da primeira consulta`, format =
"%d/%m/%Y")
summary(RHC.CCU$`Data da primeira consulta`)
table(RHC.CCU$`Data da primeira consulta`)
levels(RHC.CCU$`Data da primeira consulta`)

```

```

# Data do primeiro diagnóstico

```

```

RHC.CCU$`Data de diagnóstico` <- as.Date(RHC.CCU$`Data de diagnóstico`, format = "%d/%m/%Y")
summary(RHC.CCU$`Data de diagnóstico`)
table(RHC.CCU$`Data de diagnóstico`) #NA
levels(RHC.CCU$`Data da primeira consulta`)

```

```

#Diagnóstico e tratamento anteriores

```

```

# 1.Sem diag./Sem trat.; 2.Com diag./Sem trat.; 3.Com diag./Com trat.; 4.Outros; 9. Sem informação

```

```

summary(RHC.CCU$`Diagnóstico e tratamento anteriores`)
table(RHC.CCU$`Diagnóstico e tratamento anteriores`)
RHC.CCU$`Diagnóstico e tratamento anteriores`[RHC.CCU$`Diagnóstico e tratamento anteriores` == "0"] <-
"9"
RHC.CCU$`Diagnóstico e tratamento anteriores` <- factor(RHC.CCU$`Diagnóstico e tratamento anteriores`)
table(RHC.CCU$`Diagnóstico e tratamento anteriores`)

```

```
#Base mais importante para o diagnóstico do tumor
```

```
#1.Clínica; 2.Pesquisa clínica; 3.Exame por imagem; 4.Marcadores tumorais; 5.Citologia; 6.Histologia da metástase; 7.Histologia do tumor primário; 9.Sem informação
```

```
summary(RHC.CCU$`Base mais importante para o diagnóstico do tumor`)
```

```
table(RHC.CCU$`Base mais importante para o diagnóstico do tumor`)
```

```
RHC.CCU$`Base mais importante para o diagnóstico do tumor` [RHC.CCU$`Base mais importante para o diagnóstico do tumor` == "0"] <- "9"
```

```
RHC.CCU$`Base mais importante para o diagnóstico do tumor` <- factor(RHC.CCU$`Base mais importante para o diagnóstico do tumor`)
```

```
table(RHC.CCU$`Base mais importante para o diagnóstico do tumor`)
```

```
#-----
```

```
# Bloco de Caracterização do tumor-----
```

```
#Localização primária, Tipo histológico, Estadiamento clínico do tumor (TNM) e
```

```
#Lateralidade do tumor
```

```
# Tipo histologico
```

```
summary(RHC.CCU$`Tipo histológico`)
```

```
table(RHC.CCU$`Tipo histológico`)
```

```
RHC.CCU$`Tipo histológico` [RHC.CCU$`Tipo histológico` %in% c("8000/2", "8077/3", "8090/3", "9990/3")] <- "9999/9"
```

```
RHC.CCU$`Tipo histológico` <- factor(RHC.CCU$`Tipo histológico`)
```

```
table(RHC.CCU$`Tipo histológico`)
```

```
# Estadiamento
```

```
summary(RHC.CCU$`Estadiamento clínico do tumor (TNM)`)
```

```
RHC.CCU$`Estadiamento clínico do tumor (TNM)` [RHC.CCU$`Estadiamento clínico do tumor (TNM)` %in% c("00", "98")] <- "99"
```

```
RHC.CCU$`Estadiamento clínico do tumor (TNM)` <- factor(RHC.CCU$`Estadiamento clínico do tumor (TNM)`)
```

```
table(RHC.CCU$`Estadiamento clínico do tumor (TNM)`)
```

```
# TNM
```



```

RHC.CCU$TNM <- as.factor(RHC.CCU$TNM)
summary(RHC.CCU$TNM)
RHC.CCU$TNM[RHC.CCU$TNM %in% c("9", "00", "01", "1", "10", "21", "23", "3", "30", "31",
                                "32", "33", "4", "40", "41", "43", "99")] <- "999"
table(RHC.CCU$TNM)

# 999 Sem informação

# Lateralidade

# 1.Direita; 2. Esquerda; 3.Bilateral; 8.Não se aplica; 9.Sem informação

summary(RHC.CCU$Lateralidade do tumor`)
table(RHC.CCU$Lateralidade do tumor`)
RHC.CCU$Lateralidade do tumor`[RHC.CCU$Lateralidade do tumor` == "0"] <- "9"
RHC.CCU$Lateralidade do tumor` <- factor(RHC.CCU$Lateralidade do tumor`)
table(RHC.CCU$Lateralidade do tumor`)

#-----

# Bloco de Caracterização do primeiro tratamento-----

#Data do início do primeiro tratamento, Principal razão para a não realização do tratamento,
#Primeiro tratamento recebido, Estado da doença ao final do primeiro tratamento e Data do óbito

#Data do início do primeiro tratamento

RHC.CCU$`Data de início do tratamento` <- as.Date(RHC.CCU$`Data de início do tratamento`, format =
"%d/%m/%Y")
summary(RHC.CCU$`Data de início do tratamento`)
table(RHC.CCU$`Data de início do tratamento`) #NA

# Principal razão para a não realização do tratamento

#1.Recusa do tratamento; 2.Tratamento realizado fora; 3.Doença avançada, falta de condições clínicas ou outras
doenças associadas; 4.Abandono do tratamento; 5.Complicações de tratamento; 6.Óbito; 7.Outras razões; 8.Não
se aplica; 9. Sem informação

summary(RHC.CCU$`Principal razão para a não realização do tratamento antineoplásico no hospital`)

```

```

# Tudo certo nesse

#Primeiro tratamento recebido

summary(RHC.CCU$`Primeiro tratamento recebido no hospital`) #Tudo certo

#Estado da doença ao final do primeiro tratamento

#1.Sem evidência da doença (remissão completa); 2.Remissão parcial; 3.Doença estável; 4.Doença em
progressão; 5.Suporte terapêutico oncológico; 6. Óbito; 8. Não se aplica; 9. Sem informação

summary(RHC.CCU$`Estado da doença ao final do tratamento`)
table(RHC.CCU$`Estado da doença ao final do tratamento`) #Certo

# Data do óbito

RHC.CCU$`Data de óbito` <- as.Date(RHC.CCU$`Data de óbito`, format = "%d/%m/%Y")
summary(RHC.CCU$`Data de óbito`)
table(RHC.CCU$`Data de óbito`) #NA

# Conferindo se tem a informação 99/99/9999 (sem informação)

RHC.CCU$anoDIAG <- str_sub(RHC.CCU$`Data de diagnóstico`, start = 1, end = 4)
RHC.CCU$anoPRIMEIRACONS <- str_sub(RHC.CCU$`Data da primeira consulta`, start = 1, end = 4)
RHC.CCU$anoINISTRATA <- str_sub(RHC.CCU$`Data de início do tratamento`, start = 1, end = 4)
RHC.CCU$anoOBITO <- str_sub(RHC.CCU$`Data de óbito`, start = 1, end = 4)

table(RHC.CCU$anoDIAG)
table(RHC.CCU$anoPRIMEIRACONS)
table(RHC.CCU$anoINISTRATA)
table(RHC.CCU$anoOBITO)

#-----

# Criando completude e incompletude

##### TROCANDO NA POR 9 #####

replace_factor_na <- function(x){
  x <- as.character(x)

```

```

x <- if_else(is.na(x), "9", x)
x <- as.factor(x)
}

RHC.CCU <- RHC.CCU%>%
  mutate_if(is.factor, replace_factor_na)

RHC.CCU$`Data da primeira consulta` <- lubridate::as_date(RHC.CCU$`Data da primeira consulta`, format =
"%d/%m/%Y")
library(lubridate)
RHC.CCU$Ano <- year(RHC.CCU$`Data da primeira consulta`)
RHC.CCU$Ano
table(RHC.CCU$Ano)
RHC.CCU1 <- RHC.CCU[RHC.CCU$Ano >= 2008 & RHC.CCU$Ano <= 2016,]

# Definindo as categorias

RHC.CCU$Sexo <- ifelse(RHC.CCU$Sexo == "9", "Incompletude", "Completude")
RHC.CCU$`Raça/Cor da pele` <- ifelse(RHC.CCU$`Raça/Cor da pele` == "9", "Incompletude", "Completude")
RHC.CCU$Escolaridade <- ifelse(RHC.CCU$Escolaridade == "9", "Incompletude", "Completude")
RHC.CCU$`Estado de residência` <- ifelse(RHC.CCU$`Estado de residência` == "9", "Incompletude",
"Completude")
RHC.CCU$Idade <- ifelse(RHC.CCU$Idade == "999", "Incompletude", "Completude")
RHC.CCU$`Diagnóstico e tratamento anteriores` <- ifelse(RHC.CCU$`Diagnóstico e tratamento anteriores` ==
"9", "Incompletude", "Completude")
RHC.CCU$`Base mais importante para o diagnóstico do tumor` <- ifelse(RHC.CCU$`Base mais importante para
o diagnóstico do tumor` == "9", "Incompletude", "Completude")
RHC.CCU$`Tipo histológico` <- ifelse(RHC.CCU$`Tipo histológico` == "9999/9", "Incompletude",
"Completude")
RHC.CCU$`Estadiamento clínico do tumor (TNM)` <- ifelse(RHC.CCU$`Estadiamento clínico do tumor
(TNM)` == "99", "Incompletude", "Completude")
RHC.CCU$TNM <- ifelse(RHC.CCU$TNM == "999", "Incompletude", "Completude")
RHC.CCU$`Lateralidade do tumor` <- ifelse(RHC.CCU$`Lateralidade do tumor` == "9", "Incompletude",
"Completude")
RHC.CCU$`Principal razão para a não realização do tratamento antineoplásico no hospital` <-
ifelse(RHC.CCU$`Principal razão para a não realização do tratamento antineoplásico no hospital` == "9",
"Incompletude", "Completude")
RHC.CCU$`Estado da doença ao final do tratamento` <- ifelse(RHC.CCU$`Estado da doença ao final do
tratamento` == "9", "Incompletude", "Completude")

```

```

RHC.CCU$`Primeiro tratamento recebido no hospital` <- ifelse(RHC.CCU$`Primeiro tratamento recebido no
hospital` == "9", "Incompletude", "Completude")
RHC.CCU$`Data da primeira consulta` <- ifelse(is.na(RHC.CCU$`Data da primeira consulta`), "Incompletude",
"Completude")
RHC.CCU$`Data de diagnóstico` <- ifelse(is.na(RHC.CCU$`Data de diagnóstico`), "Incompletude",
"Completude")
RHC.CCU$`Data de início do tratamento` <- ifelse(is.na(RHC.CCU$`Data de início do
tratamento`), "Incompletude", "Completude")
RHC.CCU$`Data de óbito` <- ifelse(RHC.CCU$`Data de óbito` == "0010-02-03", "Incompletude",
"Completude")

rhc.completude <- RHC.CCU%>%
  group_by(Sexo, `Raça/Cor da pele`, Escolaridade, `Estado de residência`, Idade, `Data da primeira consulta`,
`Data de diagnóstico`, `Diagnóstico e tratamento anteriores`,
  `Base mais importante para o diagnóstico do tumor`, `Tipo histológico`, `Estadiamento clínico do tumor
(TNM)`, TNM,
  `Lateralidade do tumor`, `Data de início do tratamento`, `Principal razão para a não realização do tratamento
antineoplásico no hospital`,
  `Primeiro tratamento recebido no hospital`, `Estado da doença ao final do tratamento`, `Data de óbito`, .drop
= F)%>%
  select(Sexo, `Raça/Cor da pele`, Escolaridade, `Estado de residência`, Idade, `Data da primeira consulta`, `Data
de diagnóstico`, `Diagnóstico e tratamento anteriores`,
  `Base mais importante para o diagnóstico do tumor`, `Tipo histológico`, `Estadiamento clínico do tumor
(TNM)`, TNM,
  `Lateralidade do tumor`, `Data de início do tratamento`, `Principal razão para a não realização do tratamento
antineoplásico no hospital`,
  `Primeiro tratamento recebido no hospital`, `Estado da doença ao final do tratamento`, `Data de óbito`, Ano)

# Invertendo linha por coluna

rhc.completude = melt (rhc.completude, id = c("Ano"), variable.name = "variable.name", value.name = "Total")

rhc.completude$completude[rhc.completude$Total == "Completude"] <- "Completude"
rhc.completude$Incompletude[rhc.completude$Total == "Incompletude"] <- "Incompletude"

# Criando uma tabela apenas com as variaves selecionadas

tabela.completude <- rhc.completude%>%
  group_by(Ano, variable.name, Total, .drop = F)%>%
  arrange(Ano)%>%

```

```

tally()

# Filtrando para completude

tabela.completude1 <- tabela.completude%>%
  filter(Total == "Completude")%>%
  select(Ano,variable.name, n)

tabela.completude1$Completude <- rep("Completude")

# Filtrando para incompletude

tabela.incompletude <- tabela.completude%>%
  filter(Total == "Incompletude")%>%
  select(Ano,variable.name, n)

tabela.incompletude$Incompletude <- rep("Incompletude")

# Juntando as bases de completude e incompletude

tabela.geral <- left_join(tabela.completude1, tabela.incompletude, by = c("Ano","variable.name"))

tabela.geral<- tabela.geral%>%
  mutate_all(replace_na, 0)

# Trabalhando as bases para o calculo de completude

tabela.geral$total <- round(tabela.geral$n.x + tabela.geral$n.y)

tabela.completude <- tabela.geral%>%
  group_by(Ano, variable.name, n.y, total)%>%
  arrange(Ano)%>%
  tally()

tabela.completude <- tabela.completude[,c(-5)]

colnames(tabela.completude)[1:4] <- c("Ano", "Variáveis", "N incompletude", "Total")

tabela.completude$`% Incompletude` <- (tabela.completude$`N incompletude` / (tabela.completude$total)*100)
tabela.completude$`% Incompletude` <- round(tabela.completude$`% Incompletude`, 2)

```

```

tabela.completude$totaldados` <- rep("100")

tabela.completude$totaldados <- as.numeric(tabela.completude$totaldados)

tabela.completude$completude <- tabela.completude$totaldados - tabela.completude$`% Incompletude`

# Salvando a tabela

write.csv(tabela.completude, "tabela.completude.csv")

tabela.completa <- tabela.completude%>%
  select(Ano, Variáveis, completude)

tabela.completa <- tabela.completa[,c(-1)]

# Organizando a formatação da tabela

tabela.final <- tabela.completa%>%
  pivot_wider(names_from = Ano,
              values_from = completude)%>%
  unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))

# Gerando a tabela

tab.completude <- knitr::kable(tabela.final, align = rep(c('l','c','c')), format.args = list(decimal.mark = ',', big.mark
= "."),
                             caption = "") %>%
  kable_classic(full_width = F, html_font = "Calibri") %>%
  column_spec(1, bold = T)%>%
  row_spec(0, bold=T) %>%
  #row_spec(6, bold=T)%>%
  add_header_above(c(" " = 1, "Câncer do colo do útero" = 9))%>%
  footnote(general = "IntegradorRHC", general_title = "Fonte de dados:")

tab.completude
#####Câncer de mama #####

RHC.MAMA <- read.xlsx("base_filtrada.xlsx")

RHC.MAMA <- RHC.MAMA%>%

```

```
select(SEXO, IDADE, RACACOR, INSTRUC, ESTADRES, DATAPRICON, DTDIAGNO, DIAGANT,
BASMAIMP,
      LOCTUDET, TIPOHIST, ESTADIAM, LATERALI, DATAINITRT, RZNTR, PRITRATH, ESTDFIMT,
      DATAOBITO, TNM)
```

```
RHC.MAMA <- RHC.MAMA%>%
  filter(LOCTUDET == "C50")
```

```
# Analisando a completude das variáveis
```

```
names(RHC.MAMA)[1] <- "Sexo"
names(RHC.MAMA)[2] <- "Idade"
names(RHC.MAMA)[3] <- "Raça/Cor da pele"
names(RHC.MAMA)[4] <- "Escolaridade"
names(RHC.MAMA)[5] <- "Estado de residência"
names(RHC.MAMA)[6] <- "Data da primeira consulta"
names(RHC.MAMA)[7] <- "Data de diagnóstico"
names(RHC.MAMA)[8] <- "Diagnóstico e tratamento anteriores"
names(RHC.MAMA)[9] <- "Base mais importante para o diagnóstico do tumor"
names(RHC.MAMA)[10] <- "Localização primária do tumor"
names(RHC.MAMA)[11] <- "Tipo histológico"
names(RHC.MAMA)[12] <- "Estadiamento clínico do tumor (TNM)"
names(RHC.MAMA)[13] <- "Lateralidade do tumor"
names(RHC.MAMA)[14] <- "Data de início do tratamento"
names(RHC.MAMA)[15] <- "Principal razão para a não realização do tratamento antineoplásico no hospital"
names(RHC.MAMA)[16] <- "Primeiro tratamento recebido no hospital"
names(RHC.MAMA)[17] <- "Estado da doença ao final do tratamento"
names(RHC.MAMA)[18] <- "Data de óbito"
names(RHC.MAMA)[19] <- "TNM"
```

```
# Corrigindo a classificação das variáveis
```

```
RHC.MAMA <- RHC.MAMA%>%
  mutate(Sexo = as.factor(Sexo),
         #Idade = as.numeric(Idade),
         `Raça/Cor da pele` = as.factor(`Raça/Cor da pele`),
         `Estado de residência` = as.factor(`Estado de residência`),
         `Diagnóstico e tratamento anteriores` = as.factor(`Diagnóstico e tratamento anteriores`),
         `Base mais importante para o diagnóstico do tumor` = as.factor(`Base mais importante para o diagnóstico do
tumor`),
```

```

`Localização primária do tumor` = as.factor(`Localização primária do tumor`),
`Tipo histológico` = as.factor(`Tipo histológico`),
`Estadiamento clínico do tumor (TNM)` = as.factor(`Estadiamento clínico do tumor (TNM)`),
`Lateralidade do tumor` = as.factor(`Lateralidade do tumor`),
`Principal razão para a não realização do tratamento antineoplásico no hospital` = as.factor(`Principal razão
para a não realização do tratamento antineoplásico no hospital`),
`Primeiro tratamento recebido no hospital` = as.factor(`Primeiro tratamento recebido no hospital`),
TNM = as.factor(TNM),
`Estado da doença ao final do tratamento` = as.factor(`Estado da doença ao final do tratamento`))

# Transferindo todas as inequivalencias (divergências no preenchimento de acordo com o dicionário) nas variáveis
factor para o 9

# Bloco de identificação do paciente

# Sexo, Raça/cor da pele, escolaridade e UF de procedencia

# Sexo - 1 = Masculino e 2 = Feminino | necessário criar o 9

summary(RHC.MAMA$Sexo)
RHC.MAMA$Sexo <- factor(RHC.MAMA$Sexo, levels = c("0", "1", "2", "3", "9"))
table(RHC.MAMA$Sexo)
RHC.MAMA$Sexo[RHC.MAMA$Sexo %in% c("0", "3")] <- "9"
RHC.MAMA$Sexo <- factor(RHC.MAMA$Sexo)
table(RHC.MAMA$Sexo)

# Raça/cor da pele - 1 = Branca, 2 = Preta, 3 = Amarela, 4 = Parda, 5 = Indígena, 9 =Sem informação

summary(RHC.MAMA$`Raça/Cor da pele`)
RHC.MAMA$`Raça/Cor da pele`[RHC.MAMA$`Raça/Cor da pele` == "99"] <- "9"
RHC.MAMA$`Raça/Cor da pele` <- factor(RHC.MAMA$`Raça/Cor da pele`)
table(RHC.MAMA$`Raça/Cor da pele`)

# Escolaridade - 1.Nenhuma; 2.Fundamental incompleto; 3.Fundamental completo; 4.Nível médio; 5.Nível
superior incompleto; 6.Nível superior completo; 9.Sem informação

summary(RHC.MAMA$Escolaridade)
RHC.MAMA$Escolaridade <- factor(RHC.MAMA$Escolaridade)
RHC.MAMA$Escolaridade[RHC.MAMA$Escolaridade == "0"] <- "9"
RHC.MAMA$Escolaridade <- factor(RHC.MAMA$Escolaridade)

```



```

table(RHC.MAMA$Escolaridade)

# Estado de Residência
# 77 e 99 inconsistências - criar 9 (sem informação)
summary(RHC.MAMA$`Estado de residência`)
RHC.MAMA$`Estado de residência` <- factor(RHC.MAMA$`Estado de residência`, levels = c("77","99","AC",
"AL", "AM", "AP", "BA", "CE", "DF", "ES", "GO", "MA",
"MG", "MS", "MT", "PA", "PB", "PE", "PI", "PR", "RJ",
"RN", "RO",
"RR", "RS", "SC", "SE", "SP", "TO", "9"))
table(RHC.MAMA$`Estado de residência`)
RHC.MAMA$`Estado de residência`[RHC.MAMA$`Estado de residência` %in% c("77", "99")] <- "9"
RHC.MAMA$`Estado de residência` <- factor(RHC.MAMA$`Estado de residência`)
table(RHC.MAMA$`Estado de residência`)

# Idade

RHC.MAMA$Idade <- as.numeric(RHC.MAMA$Idade)
table(RHC.MAMA$Idade)
summary(RHC.MAMA$Idade)
levels(RHC.MAMA$Idade)
table(RHC.MAMA$Idade)
RHC.MAMA$Idade[RHC.MAMA$Idade %in% c("000", NA)] <- "999"
RHC.MAMA$Idade <- as.numeric(RHC.MAMA$Idade)
table(RHC.MAMA$Idade)

# Bloco de caracterização do diagnóstico-----

#Data da primeira consulta, Data do primeiro diagnóstico, Diagnóstico e tratamento anteriores e
#Base mais importante para o diagnóstico do tumor

# Data da primeira consulta

RHC.MAMA$`Data da primeira consulta` <- as.Date(RHC.MAMA$`Data da primeira consulta`, format =
"%d/%m/%Y")
summary(RHC.MAMA$`Data da primeira consulta`)
table(RHC.MAMA$`Data da primeira consulta`)
levels(RHC.MAMA$`Data da primeira consulta`)

# Data do primeiro diagnóstico

```

```

RHC.MAMA$`Data de diagnóstico` <- as.Date(RHC.MAMA$`Data de diagnóstico`, format = "%d/%m/%Y")
summary(RHC.MAMA$`Data de diagnóstico`)
table(RHC.MAMA$`Data de diagnóstico`) #NA
levels(RHC.MAMA$`Data da primeira consulta`)

#Diagnóstico e tratamento anteriores

# 1.Sem diag./Sem trat.; 2.Com diag./Sem trat.; 3.Com diag./Com trat.; 4.Outros; 9. Sem informação

summary(RHC.MAMA$`Diagnóstico e tratamento anteriores`)
table(RHC.MAMA$`Diagnóstico e tratamento anteriores`)
RHC.MAMA$`Diagnóstico e tratamento anteriores`[RHC.MAMA$`Diagnóstico e tratamento anteriores` == "0"]
<- "9"
RHC.MAMA$`Diagnóstico e tratamento anteriores` <- factor(RHC.MAMA$`Diagnóstico e tratamento
anteriores`)
table(RHC.MAMA$`Diagnóstico e tratamento anteriores`)

#Base mais importante para o diagnóstico do tumor

#1.Clínica; 2.Pesquisa clínica; 3.Exame por imagem; 4.Marcadores tumorais; 5.Citologia; 6.Histologia da
metástase; 7.Histologia do tumor primário; 9.Sem informação

summary(RHC.MAMA$`Base mais importante para o diagnóstico do tumor`)
table(RHC.MAMA$`Base mais importante para o diagnóstico do tumor`)
RHC.MAMA$`Base mais importante para o diagnóstico do tumor`[RHC.MAMA$`Base mais importante para o
diagnóstico do tumor` == "0"] <- "9"
RHC.MAMA$`Base mais importante para o diagnóstico do tumor` <- factor(RHC.MAMA$`Base mais
importante para o diagnóstico do tumor`)
table(RHC.MAMA$`Base mais importante para o diagnóstico do tumor`)

#-----

# Bloco de Caracterização do tumor-----

#Localização primária, Tipo histológico, Estadiamento clínico do tumor (TNM) e
#Lateralidade do tumor

# Tipo histologico

summary(RHC.MAMA$`Tipo histológico`)

```

```

table(RHC.MAMA$`Tipo histológico`)
RHC.MAMA$`Tipo histológico` [RHC.MAMA$`Tipo histológico` %in%
c("/", "8000/2", "8010/", "8100/3", "8500/9", "8503/", "9713/3", "9990/3", "9991/6", "9991/7",
      "9992/0", "9992/2", "9992/6", NA)] <- "9999/9"
RHC.MAMA$`Tipo histológico` <- factor(RHC.MAMA$`Tipo histológico`)
table(RHC.MAMA$`Tipo histológico`)

# Estadiamento

summary(RHC.MAMA$`Estadiamento clínico do tumor (TNM)`)
RHC.MAMA$`Estadiamento clínico do tumor (TNM)` [RHC.MAMA$`Estadiamento clínico do tumor (TNM)`
%in% c("00", "98", "33", "9", NA)] <- "99"
RHC.MAMA$`Estadiamento clínico do tumor (TNM)` <- factor(RHC.MAMA$`Estadiamento clínico do tumor
(TNM)`)
table(RHC.MAMA$`Estadiamento clínico do tumor (TNM)`)

# tnm

RHC.MAMA$TNM <- as.factor(RHC.MAMA$TNM)
summary(RHC.MAMA$TNM)
RHC.MAMA$TNM[RHC.MAMA$TNM %in% c("9", "00", "01", "1", "10", "2", "20", "21", "22", "23", "3", "30",
"31",
      "32", "33", "4", "40", "41", "43", "99", "9")] <- "999"
table(RHC.MAMA$TNM)

# 999 Sem informação

# Lateralidade

# 1.Direita; 2. Esquerda; 3.Bilateral; 8.Não se aplica; 9.Sem informação

summary(RHC.MAMA$`Lateralidade do tumor`)
table(RHC.MAMA$`Lateralidade do tumor`)
RHC.MAMA$`Lateralidade do tumor` [RHC.MAMA$`Lateralidade do tumor` == "0"] <- "9"
RHC.MAMA$`Lateralidade do tumor` <- factor(RHC.MAMA$`Lateralidade do tumor`)
table(RHC.MAMA$`Lateralidade do tumor`)

#-----

# Bloco de Caracterização do primeiro tratamento-----

```

#Data do início do primeiro tratamento, Principal razão para a não realização do tratamento,  
#Primeiro tratamento recebido, Estado da doença ao final do primeiro tratamento e Data do óbito

#Data do início do primeiro tratamento

```
RHC.MAMA$`Data de início do tratamento` <- as.Date(RHC.MAMA$`Data de início do tratamento`, format =
"%d/%m/%Y")
```

```
summary(RHC.MAMA$`Data de início do tratamento`)
```

```
table(RHC.MAMA$`Data de início do tratamento`) #NA
```

# Principal razão para a não realização do tratamento

#1.Recusa do tratamento; 2.Tratamento realizado fora; 3.Doença avançada, falta de condições clínicas ou outras doenças associadas; 4.Abandono do tratamento; 5.Complicações de tratamento; 6.Óbito; 7.Outras razões; 8.Não se aplica; 9. Sem informação

```
summary(RHC.MAMA$`Principal razão para a não realização do tratamento antineoplásico no hospital`)
```

# Tudo certo nesse

#Primeiro tratamento recebido

```
summary(RHC.MAMA$`Primeiro tratamento recebido no hospital`) #Conferir
```

#Estado da doença ao final do primeiro tratamento

#1.Sem evidência da doença (remissão completa); 2.Remissão parcial; 3.Doença estável; 4.Doença em progressão; 5.Suporte terapêutico oncológico; 6. Óbito; 8. Não se aplica; 9. Sem informação

```
summary(RHC.MAMA$`Estado da doença ao final do tratamento`)
```

```
table(RHC.MAMA$`Estado da doença ao final do tratamento`) #Certo
```

# Data do óbito

```
RHC.MAMA$`Data de óbito` <- as.Date(RHC.MAMA$`Data de óbito`, format = "%d/%m/%Y")
```

```
summary(RHC.MAMA$`Data de óbito`)
```

```
table(RHC.MAMA$`Data de óbito`) #NA
```

# Conferindo se tem a informação 99/99/9999 (sem informação)

```

RHC.MAMA$anoDIAG <- str_sub(RHC.MAMA$`Data de diagnóstico`, start = 1, end = 4)
RHC.MAMA$anoPRIMEIRACONS <- str_sub(RHC.MAMA$`Data da primeira consulta`, start = 1, end = 4)
RHC.MAMA$anoINISTRATA <- str_sub(RHC.MAMA$`Data de início do tratamento`, start = 1, end = 4)
RHC.MAMA$anoOBITO <- str_sub(RHC.MAMA$`Data de óbito`, start = 1, end = 4)

table(RHC.MAMA$anoDIAG)
table(RHC.MAMA$anoPRIMEIRACONS)
table(RHC.MAMA$anoINISTRATA)
table(RHC.MAMA$anoOBITO)

#-----

# Criando completude e incompletude

##### TROCANDO NA POR 9 #####

replace_factor_na <- function(x){
  x <- as.character(x)
  x <- if_else(is.na(x), "9", x)
  x <- as.factor(x)
}

RHC.MAMA <- RHC.MAMA%>%
  mutate_if(is.factor, replace_factor_na)

RHC.MAMA$`Data da primeira consulta` <- lubridate::as_date(RHC.MAMA$`Data da primeira consulta`,
format = "%d/%m/%Y")
library(lubridate)
RHC.MAMA$Ano <- year(RHC.MAMA$`Data da primeira consulta`)
RHC.MAMA$Ano
table(RHC.MAMA$Ano)
RHC.MAMA1 <- RHC.MAMA[RHC.MAMA$Ano >= 2008 & RHC.MAMA$Ano <= 2016,]

# Definindo as categorias

RHC.MAMA$Sexo <- ifelse(RHC.MAMA$Sexo == "9", "Incompletude", "Completude")
RHC.MAMA$`Raça/Cor da pele` <- ifelse(RHC.MAMA$`Raça/Cor da pele` == "9", "Incompletude",
"Completude")
RHC.MAMA$Escolaridade <- ifelse(RHC.MAMA$Escolaridade == "9", "Incompletude", "Completude")

```

```

RHC.MAMA$`Estado de residência` <- ifelse(RHC.MAMA$`Estado de residência` == "9", "Incompletude",
"Completude")
RHC.MAMA$Idade <- ifelse(RHC.MAMA$Idade == "999", "Incompletude", "Completude")
RHC.MAMA$`Diagnóstico e tratamento anteriores` <- ifelse(RHC.MAMA$`Diagnóstico e tratamento
anteriores` == "9", "Incompletude", "Completude")
RHC.MAMA$`Base mais importante para o diagnóstico do tumor` <- ifelse(RHC.MAMA$`Base mais importante
para o diagnóstico do tumor` == "9", "Incompletude", "Completude")
RHC.MAMA$`Tipo histológico` <- ifelse(RHC.MAMA$`Tipo histológico` == "9999/9", "Incompletude",
"Completude")
RHC.MAMA$`Estadiamento clínico do tumor (TNM)` <- ifelse(RHC.MAMA$`Estadiamento clínico do tumor
(TNM)` == "99", "Incompletude", "Completude")
RHC.MAMA$TNM <- ifelse(RHC.MAMA$TNM == "999", "Incompletude", "Completude")
RHC.MAMA$`Lateralidade do tumor` <- ifelse(RHC.MAMA$`Lateralidade do tumor` == "9", "Incompletude",
"Completude")
RHC.MAMA$`Principal razão para a não realização do tratamento antineoplásico no hospital` <-
ifelse(RHC.MAMA$`Principal razão para a não realização do tratamento antineoplásico no hospital` == "9",
"Incompletude", "Completude")
RHC.MAMA$`Estado da doença ao final do tratamento` <- ifelse(RHC.MAMA$`Estado da doença ao final do
tratamento` == "9", "Incompletude", "Completude")
RHC.MAMA$`Primeiro tratamento recebido no hospital` <- ifelse(RHC.MAMA$`Primeiro tratamento recebido
no hospital` == "9", "Incompletude", "Completude")
RHC.MAMA$`Data da primeira consulta` <- ifelse(is.na(RHC.MAMA$`Data da primeira
consulta`), "Incompletude", "Completude")
RHC.MAMA$`Data de diagnóstico` <- ifelse(is.na(RHC.MAMA$`Data de diagnóstico`), "Incompletude",
"Completude")
RHC.MAMA$`Data de início do tratamento` <- ifelse(is.na(RHC.MAMA$`Data de início do
tratamento`), "Incompletude", "Completude")
RHC.MAMA$`Data de óbito` <- ifelse(RHC.MAMA$`Data de óbito` == "2022-03-10", "Incompletude",
"Completude")

rhc.completude <- RHC.MAMA%>%
  group_by(Sexo, `Raça/Cor da pele`, Escolaridade, `Estado de residência`, Idade, `Data da primeira consulta`,
`Data de diagnóstico`, `Diagnóstico e tratamento anteriores`,
  `Base mais importante para o diagnóstico do tumor`, `Tipo histológico`, `Estadiamento clínico do tumor
(TNM)`, TNM,
  `Lateralidade do tumor`, `Data de início do tratamento`, `Principal razão para a não realização do tratamento
antineoplásico no hospital`,
  `Primeiro tratamento recebido no hospital`, `Estado da doença ao final do tratamento`, `Data de óbito`, .drop
= F)%>%

```

```

select(Sexo, `Raça/Cor da pele`, Escolaridade, `Estado de residência`, Idade, `Data da primeira consulta`, `Data
de diagnóstico`, `Diagnóstico e tratamento anteriores`,
  `Base mais importante para o diagnóstico do tumor`, `Tipo histológico`, `Estadiamento clínico do tumor
(TNM)`, TNM,
  `Lateralidade do tumor`, `Data de início do tratamento`, `Principal razão para a não realização do tratamento
antineoplásico no hospital`,
  `Primeiro tratamento recebido no hospital`, `Estado da doença ao final do tratamento`, `Data de óbito`, Ano)

```

```
rhc.completude = melt (rhc.completude, id = c("Ano"), variable.name = "variable.name", value.name = "Total")
```

```

rhc.completude$completude[rhc.completude$Total == "Completude"] <- "Completude"
rhc.completude$Incompletude[rhc.completude$Total == "Incompletude"] <- "Incompletude"

```

```

tabela.completude <- rhc.completude%>%
  group_by(Ano, variable.name, Total,.drop = F)%>%
  arrange(Ano)%>%
  tally()

```

```

tabela.completude1 <- tabela.completude%>%
  filter(Total == "Completude")%>%
  select(Ano,variable.name, n)

```

```
tabela.completude1$Completude <- rep("Completude")
```

```

tabela.incompletude <- tabela.completude%>%
  filter(Total == "Incompletude")%>%
  select(Ano,variable.name, n)

```

```
tabela.incompletude$Incompletude <- rep("Incompletude")
```

```
tabela.geral <- left_join(tabela.completude1, tabela.incompletude, by = c("Ano", "variable.name"))
```

```

tabela.geral<- tabela.geral%>%
  mutate_all(replace_na, 0)

```

```
tabela.geral$total <- round(tabela.geral$n.x + tabela.geral$n.y)
```

```

tabela.completude <- tabela.geral%>%
  group_by(Ano, variable.name, n.y, total)%>%
  arrange(Ano)%>%

```

```

tally()

tabela.completude <- tabela.completude[,c(-5)]

colnames(tabela.completude)[1:4] <- c("Ano", "Variáveis", "N incompletude", "Total")

tabela.completude$`% Incompletude` <- (tabela.completude$`N incompletude` / (tabela.completude$Total)*100)
tabela.completude$`% Incompletude` <- round(tabela.completude$`% Incompletude`, 2)

tabela.completude$totaldados <- rep("100")

tabela.completude$totaldados <- as.numeric(tabela.completude$totaldados)

tabela.completude$completude <- tabela.completude$totaldados - tabela.completude$`% Incompletude`

write.csv(tabela.completude, "tabela.completude.mama.csv")

tabela.completa <- tabela.completude%>%
  select(Ano, Variáveis, completude)

tabela.completa <- tabela.completa[,c(-1)]

tabela.final <- tabela.completa%>%
  pivot_wider(names_from = Ano,
              values_from = completude)%>%
  unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))

tab.completude.mama <- knitr::kable(tabela.final, align = rep(c('l','c','c')), format.args = list(decimal.mark = ',',
big.mark = "."),
                                caption = "") %>%
  kable_classic(full_width = F, html_font = "Calibri") %>%
  column_spec(1, bold = T)%>%
  row_spec(0, bold=T) %>%
  #row_spec(6, bold=T)%>%
  add_header_above(c(" " = 1, "Câncer de mama" = 9))%>%
  footnote(general = "IntegradorRHC", general_title = "Fonte de dados:")

tab.completude.mama

```



```
# ANALISE DE INCONSISTENCIA CANCER DO COLO DO UTERO #####
```

```
# Chamando a base
```

```
RHC.CCU <- read.xlsx("base_filtrada.xlsx")
```

```
# Selecionando as variaveis
```

```
RHC.CCU <- RHC.CCU%>%
```

```
  select(SEXO, IDADE, RACACOR, INSTRUC, ESTADRES, DATAPRICON, DTDIAGNO, DIAGANT,
  BASMAIMP,
  LOCTUDET, TIPOHIST, ESTADIAM, LATERALI, DATAINITRT, RZNTR, PRITRATH, ESTDFIMT,
  DATAOBITO, TNM)
```

```
# Filtrando para colo do utero
```

```
RHC.CCU <- RHC.CCU%>%
```

```
  filter(LOCTUDET == "C53")
```

```
# Mudando o nome das variaveis no banco
```

```
names(RHC.CCU)[1] <- "Sexo"
```

```
names(RHC.CCU)[2] <- "Idade"
```

```
names(RHC.CCU)[3] <- "Raça/Cor da pele"
```

```
names(RHC.CCU)[4] <- "Escolaridade"
```

```
names(RHC.CCU)[5] <- "Estado de residência"
```

```
names(RHC.CCU)[6] <- "Data da primeira consulta"
```

```
names(RHC.CCU)[7] <- "Data de diagnóstico"
```

```
names(RHC.CCU)[8] <- "Diagnóstico e tratamento anteriores"
```

```
names(RHC.CCU)[9] <- "Base mais importante para o diagnóstico do tumor"
```

```
names(RHC.CCU)[10] <- "Localização primária do tumor"
```

```
names(RHC.CCU)[11] <- "Tipo histológico"
```

```
names(RHC.CCU)[12] <- "Estadiamento clínico do tumor (TNM)"
```

```
names(RHC.CCU)[13] <- "Lateralidade do tumor"
```

```
names(RHC.CCU)[14] <- "Data de início do tratamento"
```

```
names(RHC.CCU)[15] <- "Principal razão para a não realização do tratamento antineoplásico no hospital"
```

```
names(RHC.CCU)[16] <- "Primeiro tratamento recebido no hospital"
```

```
names(RHC.CCU)[17] <- "Estado da doença ao final do tratamento"
```

```
names(RHC.CCU)[18] <- "Data de óbito"
```

```
names(RHC.CCU)[19] <- "TNM"
```

```
# CRIANDO AS REGIOES
```

```
RHC.CCU$Regioes = ""
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "AC")] = "Norte"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "AM")] = "Norte"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "RR")] = "Norte"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "AP")] = "Norte"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "PA")] = "Norte"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "TO")] = "Norte"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "RO")] = "Norte"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "MA")] = "Nordeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "PI")] = "Nordeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "CE")] = "Nordeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "RN")] = "Nordeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "PE")] = "Nordeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "PB")] = "Nordeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "SE")] = "Nordeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "AL")] = "Nordeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "BA")] = "Nordeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "MT")] = "Centro-oeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "MS")] = "Centro-oeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "GO")] = "Centro-oeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "SP")] = "Sudeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "RJ")] = "Sudeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "ES")] = "Sudeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "MG")] = "Sudeste"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "PR")] = "Sul"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "RS")] = "Sul"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "SC")] = "Sul"
RHC.CCU$Regioes[str_detect(RHC.CCU$`Estado de residência`, "DF")] = "Centro-oeste"
```

```
library(lubridate)
```

```
RHC.CCU$`Data da primeira consulta` <- lubridate::as_date(RHC.CCU$`Data da primeira consulta`, format =
"%d/%m/%Y")
RHC.CCU$Ano <- year(RHC.CCU$`Data da primeira consulta`)
RHC.CCU$`Data de início do tratamento` <- as.Date(RHC.CCU$`Data de início do tratamento`, format =
"%d/%m/%Y")
RHC.CCU$`Data de diagnóstico` <- as.Date(RHC.CCU$`Data de diagnóstico`, format = "%d/%m/%Y")
```

```
# BASE COM APENAS AS VARIÁVEIS DESEJADAS
```

```
AUX.CCU <- RHC.CCU%>%
```

```
  select(`Data da primeira consulta`, `Data de diagnóstico`, `Data de início do tratamento`, `Sexo`, Regioes,  
         `Localização primária do tumor`, `Tipo histológico`, `TNM`, `Estadiamento clínico do tumor (TNM)`, Ano)
```

```
table(AUX.CCU$`Estadiamento clínico do tumor (TNM)`)
```

```
# DATA DA PRIMEIRA CONSULTA E DATA DE DIAGNOSTICO
```

```
DTPC_DTDIAG <- AUX.CCU%>%
```

```
  filter(`Data da primeira consulta` < `Data de diagnóstico`)%>%
```

```
  group_by(Ano) %>%
```

```
  arrange(Ano)%>%
```

```
  tally()
```

```
DTPC_DTDIAG$Total <- rep("97022")
```

```
DTPC_DTDIAG$Total <- as.numeric(DTPC_DTDIAG$Total)
```

```
DTPC_DTDIAG$`% Inconsistência` <- (DTPC_DTDIAG$n / (DTPC_DTDIAG$Total)*100)
```

```
DTPC_DTDIAG$`% Inconsistência` <- round(DTPC_DTDIAG$`% Inconsistência`, 2)
```

```
DTPC_DTDIAG <- DTPC_DTDIAG[,c(-2,-3)]
```

```
DTPC_DTDIAG <- DTPC_DTDIAG%>%
```

```
  pivot_wider(names_from = Ano,
```

```
              values_from = `% Inconsistência`)%>%
```

```
  unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))
```

```
# DATA DA PRIMEIRA CONSULTA E DATA DE INICIO DO TRATAMENTO
```

```
DTPC_DTINIT <- AUX.CCU%>%
```

```
  filter(`Data da primeira consulta` > `Data de início do tratamento`)%>%
```

```
  group_by(Ano) %>%
```

```
  arrange(Ano)%>%
```

```
  tally()
```

```
DTPC_DTINIT$Total <- rep("97022")
```

```
DTPC_DTINIT$Total <- as.numeric(DTPC_DTINIT$Total)
```

```
DTPC_DTINIT$`% Inconsistência` <- (DTPC_DTINIT$n / (DTPC_DTINIT$Total)*100)
```

```
DTPC_DTINIT$`% Inconsistência` <- round(DTPC_DTINIT$`% Inconsistência`, 2)
```

```

DTPC_DTINIT <- DTPC_DTINIT[,c(-2,-3)]

DTPC_DTINIT <- DTPC_DTINIT%>%
  pivot_wider(names_from = Ano,
              values_from = `% Inconsistência`)%>%
  unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))

# DATA DE DIAGNOSTICO E DATA DE INICIO DO TRATAMENTO

DTDIAG_DTINIT <- AUX.CCU%>%
  filter(`Data de diagnóstico` > `Data de início do tratamento`)%>%
  group_by(Ano) %>%
  arrange(Ano)%>%
  tally()

DTDIAG_DTINIT$Total <- rep("97022")
DTDIAG_DTINIT$Total <- as.numeric(DTDIAG_DTINIT$Total)
DTDIAG_DTINIT$`% Inconsistência` <- (DTDIAG_DTINIT$n / (DTDIAG_DTINIT$Total)*100)
DTDIAG_DTINIT$`% Inconsistência` <- round(DTDIAG_DTINIT$`% Inconsistência`, 2)

DTDIAG_DTINIT <- DTDIAG_DTINIT[,c(-2,-3)]

DTDIAG_DTINIT<- DTDIAG_DTINIT%>%
  pivot_wider(names_from = Ano,
              values_from = `% Inconsistência`)%>%
  unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))

# SEXO E LOCALIZACAO

SEXO_LOCALIZACAO <- AUX.CCU%>%
  filter(Sexo == "1" & `Localização primária do tumor` == "C53")%>%
  group_by(Ano) %>%
  arrange(Ano)%>%
  tally()

# 2009 - 1, 2012 - 3, 2013 - 1 e 2016 -1

SEXO_LOCALIZACAO$Total <- rep("97022")
SEXO_LOCALIZACAO$Total <- as.numeric(SEXO_LOCALIZACAO$Total)

```

```
SEXO_LOCALIZACAO$`% Inconsistência` <- (SEXO_LOCALIZACAO$n /
(SEXO_LOCALIZACAO$Total)*100)
SEXO_LOCALIZACAO$`% Inconsistência` <- round(SEXO_LOCALIZACAO$`% Inconsistência`, 3)
```

```
SEXO_LOCALIZACAO <- SEXO_LOCALIZACAO[,c(-2,-3)]
```

```
SEXO_LOCALIZACAO <- SEXO_LOCALIZACAO %>%
  pivot_wider(names_from = Ano,
              values_from = `% Inconsistência`)%>%
  unnest(cols = c("2009", "2012", "2013", "2016"))
```

```
# LOCALIZACAO E TIPO HISTOLOGICO
```

```
LOCALIZACAO_HISTOLOGIA <- AUX.CCU %>%
  filter(!`Tipo histológico` %in% c("8070/3",
    "8071/3",
    "8072/3",
    "8083/3",
    "8051/3",
    "8051/3",
    "8052/3",
    "8082/3",
    "8120/3",
    "8076/3",
    "8077/2",
    "8070/2",
    "8052/0",
    "8140/3",
    "8480/3",
    "8482/3",
    "8144/3",
    "8490/3",
    "8480/3",
    "8262/3",
    "8380/3",
    "8310/3",
    "8441/3",
    "9110/3",
    "8140/3",
    "8140/2",
```

```

"8560/3",
"8015/3",
"8200/3",
"8098/3",
"8240/3",
"8249/3",
"8041/3",
"8013/3",
"8020/3",
"8890/3",
"8931/3",
"8805/3",
"8910/3",
"9581/3",
"9120/3",
"9540/3",
"8890/0",
"8905/0",
"8980/3",
"8933/3",
"8960/3",
"9013/0",
"8932/0",
"8720/3",
"8780/0",
"9071/3",
"9084/0",
"9080/0"))%>%
group_by(Ano) %>%
arrange(Ano)%>%
tally()

LOCALIZACAO_HISTOLOGIA$Total <- rep("97022")
LOCALIZACAO_HISTOLOGIA$Total <- as.numeric(LOCALIZACAO_HISTOLOGIA$Total)
LOCALIZACAO_HISTOLOGIA$`% Inconsistência` <- (LOCALIZACAO_HISTOLOGIA$n /
(LOCALIZACAO_HISTOLOGIA$Total)*100)
LOCALIZACAO_HISTOLOGIA$`% Inconsistência` <- round(LOCALIZACAO_HISTOLOGIA$`%
Inconsistência`, 2)

LOCALIZACAO_HISTOLOGIA<- LOCALIZACAO_HISTOLOGIA[,c(-2,-3)]

```

```
LOCALIZACAO_HISTOLOGIA<- LOCALIZACAO_HISTOLOGIA%>%
```

```
  pivot_wider(names_from = Ano,
```

```
    values_from = `% Inconsistência`)%>%
```

```
  unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))
```

```
#LOCALIZACAO E LATERALIDADE DO TUMOR
```

```
TNM_ESTADIAMENTO <- AUX.CCU%>%
```

```
  filter(TNM %in% c("000","001","011","100", "200", "300", "400", "110", "210", "310", "410",
```

```
    "120", "220", "320", "420", "130", "230", "330", "430", "101", "201",
```

```
    "301", "401", "111", "211", "311", "411", "121", "221", "321", "421",
```

```
    "131", "231", "331", "431", "I00"))
```

```
TNM_INSITU <- TNM_ESTADIAMENTO%>%
```

```
  filter(TNM %in% c("I00", "000")))%>%
```

```
  filter(!Estadiamento clínico do tumor (TNM)` %in% c("0")))%>%
```

```
  group_by(Ano) %>%
```

```
  arrange(Ano)%>%
```

```
  tally()
```

```
TNM_1A <- TNM_ESTADIAMENTO%>%
```

```
  filter(TNM %in% c("100")))%>%
```

```
  filter(!Estadiamento clínico do tumor (TNM)` %in% c("1", "1A", "1B")))%>%
```

```
  group_by(Ano) %>%
```

```
  arrange(Ano)%>%
```

```
  tally()
```

```
TNM_2A <- TNM_ESTADIAMENTO%>%
```

```
  filter(TNM == "200")))%>%
```

```
  filter(!Estadiamento clínico do tumor (TNM)` %in% c("2", "2A", "2B")))%>%
```

```
  group_by(Ano) %>%
```

```
  arrange(Ano)%>%
```

```
  tally()
```

```
TNM_3A <- TNM_ESTADIAMENTO%>%
```

```
  filter(TNM == "300")))%>%
```

```
  filter(!Estadiamento clínico do tumor (TNM)` %in% c("3", "3A")))%>%
```

```
  group_by(Ano) %>%
```

```
  arrange(Ano)%>%
```

```
  tally()
```

```
TNM_3b <- TNM_ESTADIAMENTO%>%
  filter(TNM %in% c("110", "210", "310"))%>%
  filter(!Estadiamento clínico do tumor (TNM) ` %in% c("3", "3B"))%>%
  group_by(Ano) %>%
  arrange(Ano)%>%
  tally()
```

```
TNM_4A <- TNM_ESTADIAMENTO%>%
  filter(TNM %in% c("400", "410", "420", "430"))%>%
  filter(!Estadiamento clínico do tumor (TNM) ` %in% c("4", "4A"))%>%
  group_by(Ano) %>%
  arrange(Ano)%>%
  tally()
```

```
TNM_4B <- TNM_ESTADIAMENTO%>%
  filter(TNM %in% c("001","011","111", "211", "311", "411", "121", "221", "321", "421",
    "131", "231", "331", "431"))%>%
  filter(!Estadiamento clínico do tumor (TNM) ` %in% c("4", "4B"))%>%
  group_by(Ano) %>%
  arrange(Ano)%>%
  tally()
```

```
TNM_ESTADIAMENTO_FINAL <- bind_rows(TNM_1A, TNM_2A, TNM_3A, TNM_3b, TNM_4A,
  TNM_4B, TNM_INSITU)
```

```
TNM_ESTADIAMENTO_FINAL <- TNM_ESTADIAMENTO_FINAL %>%
  group_by(Ano) %>%
  summarise(Total = sum(n),
    .groups = "drop_last")
```

```
TNM_ESTADIAMENTO_FINAL$Totalgeral <- rep("97022")
TNM_ESTADIAMENTO_FINAL$Totalgeral <- as.numeric(TNM_ESTADIAMENTO_FINAL$Totalgeral)
TNM_ESTADIAMENTO_FINAL$`% Inconsistência` <- (TNM_ESTADIAMENTO_FINAL$Total /
  (TNM_ESTADIAMENTO_FINAL$Totalgeral)*100)
TNM_ESTADIAMENTO_FINAL$`% Inconsistência` <- round(TNM_ESTADIAMENTO_FINAL$`%
  Inconsistência`, 2)
```

```
TNM_ESTADIAMENTO_FINAL<- TNM_ESTADIAMENTO_FINAL[,c(-2,-3)]
```



```

TNM_ESTADIAMENTO_FINAL<- TNM_ESTADIAMENTO_FINAL%>%
  pivot_wider(names_from = Ano,
              values_from = `% Inconsistência`)%>%
  unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))

# CRIACAO DA TABELA DE CCU

tabela.total <- bind_rows(DTPC_DTDIAG, DTPC_DTINIT, DTDIAG_DTINIT,
SEXO_LOCALIZACAO,LOCALIZACAO_HISTOLOGIA,
                        TNM_ESTADIAMENTO_FINAL)

tabela.total<- tabela.total%>%
  mutate_all(replace_na, 0)

tabela.total <- cbind(Inconsistências = "", tabela.total)

TABELA.INCONSISTENCIA.CCU <- knitr::kable(tabela.total, align = rep(c('l','c','c')), format.args =
list(decimal.mark = ',', big.mark = "."))%>%
  kable_classic(full_width = F, html_font = "Calibri") %>%
  pack_rows("Data da primeira consulta < Data do diagnóstico", 1, 1) %>%
  pack_rows("Data da primeira consulta > Data de início do tratamento", 2, 2) %>%
  pack_rows("Data do diagnóstico > Data do início do tratamento", 3, 3) %>%
  pack_rows("Sexo & Localização primária do tumor", 4, 4) %>%
  pack_rows("Localização primária do tumor & Tipo histológico", 5, 5) %>%
  pack_rows("TNM & Estadiamento", 6, 6) %>%
  add_header_above(c(" " = 1, "Câncer do colo do útero" = 9))%>%
  footnote(general = "IntegradorRHC", general_title = "Fonte de dados:")

TABELA.INCONSISTENCIA.CCU

# Cancer de mama

RHC.MAMA <- read.xlsx("base_filtrada.xlsx")

RHC.MAMA <- RHC.MAMA%>%
  select(SEXO, IDADE, RACACOR, INSTRUC, ESTADRES, DATAPRICON, DTDIAGNO, DIAGANT,
BASMAIMP,
        LOCTUDET, TIPOHIST, ESTADIAM, LATERALI, DATAINITRT, RZNTR, PRITRATH, ESTDFIMT,
        DATAOBITO, TNM)

```

```

RHC.MAMA <- RHC.MAMA%>%
  filter(LOCTUDET == "C50")

# Analisando a completude das variáveis

names(RHC.MAMA)[1] <- "Sexo"
names(RHC.MAMA)[2] <- "Idade"
names(RHC.MAMA)[3] <- "Raça/Cor da pele"
names(RHC.MAMA)[4] <- "Escolaridade"
names(RHC.MAMA)[5] <- "Estado de residência"
names(RHC.MAMA)[6] <- "Data da primeira consulta"
names(RHC.MAMA)[7] <- "Data de diagnóstico"
names(RHC.MAMA)[8] <- "Diagnóstico e tratamento anteriores"
names(RHC.MAMA)[9] <- "Base mais importante para o diagnóstico do tumor"
names(RHC.MAMA)[10] <- "Localização primária do tumor"
names(RHC.MAMA)[11] <- "Tipo histológico"
names(RHC.MAMA)[12] <- "Estadiamento clínico do tumor (TNM)"
names(RHC.MAMA)[13] <- "Lateralidade do tumor"
names(RHC.MAMA)[14] <- "Data de início do tratamento"
names(RHC.MAMA)[15] <- "Principal razão para a não realização do tratamento antineoplásico no hospital"
names(RHC.MAMA)[16] <- "Primeiro tratamento recebido no hospital"
names(RHC.MAMA)[17] <- "Estado da doença ao final do tratamento"
names(RHC.MAMA)[18] <- "Data de óbito"
names(RHC.MAMA)[19] <- "TNM"

# CRIANDO AS REGIOES
RHC.MAMA$Regioes = ""
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "AC")] = "Norte"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "AM")] = "Norte"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "RR")] = "Norte"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "AP")] = "Norte"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "PA")] = "Norte"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "TO")] = "Norte"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "RO")] = "Norte"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "MA")] = "Nordeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "PI")] = "Nordeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "CE")] = "Nordeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "RN")] = "Nordeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "PE")] = "Nordeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "PB")] = "Nordeste"

```

```

RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "SE")] = "Nordeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "AL")] = "Nordeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "BA")] = "Nordeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "MT")] = "Centro-oeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "MS")] = "Centro-oeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "GO")] = "Centro-oeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "SP")] = "Sudeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "RJ")] = "Sudeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "ES")] = "Sudeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "MG")] = "Sudeste"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "PR")] = "Sul"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "RS")] = "Sul"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "SC")] = "Sul"
RHC.MAMA$Regioes[str_detect(RHC.MAMA$`Estado de residência`, "DF")] = "Centro-oeste"

```

```
table(RHC.MAMA$Regioes)
```

```
library(lubridate)
```

```
RHC.MAMA$`Data da primeira consulta` <- lubridate::as_date(RHC.MAMA$`Data da primeira consulta`,
format = "%d/%m/%Y")
```

```
RHC.MAMA$Ano <- year(RHC.MAMA$`Data da primeira consulta`)
```

```
RHC.MAMA$`Data de início do tratamento` <- as.Date(RHC.MAMA$`Data de início do tratamento`, format =
"%d/%m/%Y")
```

```
RHC.MAMA$`Data de diagnóstico` <- as.Date(RHC.MAMA$`Data de diagnóstico`, format = "%d/%m/%Y")
```

```
table(RHC.MAMA$Ano)
```

```
tab1(RHC.MAMA$Ano)
```

```
epiDisplay::tab1(RHC.MAMA$Ano)
```

```
epiDisplay::n.for.2means(RHC.MAMA$Ano)
```

```
# BASE COM APENAS AS VARIÁVEIS DESEJADAS
```

```
AUX.MAMA <- RHC.MAMA%>%
```

```
  select(`Data da primeira consulta`, `Data de diagnóstico`, `Data de início do tratamento`, `Sexo`, Regioes,
    `Localização primária do tumor`, `Tipo histológico`, `TNM`, `Estadiamento clínico do tumor (TNM)`, Ano)
```

```
table(AUX.MAMA$`Estadiamento clínico do tumor (TNM)`)
```

```
# DATA DA PRIMEIRA CONSULTA E DATA DE DIAGNOSTICO
```

```
DTPC_DTDIAG <- AUX.MAMA%>%
```

```

filter(`Data da primeira consulta` < `Data de diagnóstico`)%>%
group_by(Ano) %>%
arrange(Ano)%>%
tally()

DTPC_DTDIAG$Total <- rep("195899")
DTPC_DTDIAG$Total <- as.numeric(DTPC_DTDIAG$Total)
DTPC_DTDIAG$`% Inconsistência` <- (DTPC_DTDIAG$n / (DTPC_DTDIAG$Total)*100)
DTPC_DTDIAG$`% Inconsistência` <- round(DTPC_DTDIAG$`% Inconsistência`, 2)

DTPC_DTDIAG <- DTPC_DTDIAG[,c(-2,-3)]

DTPC_DTDIAG <- DTPC_DTDIAG%>%
pivot_wider(names_from = Ano,
             values_from = `% Inconsistência`)%>%
unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))

# DATA DA PRIMEIRA CONSULTA E DATA DE INICIO DO TRATAMENTO

DTPC_DTINIT <- AUX.MAMA%>%
filter(`Data da primeira consulta` > `Data de início do tratamento`)%>%
group_by(Ano) %>%
arrange(Ano)%>%
tally()

DTPC_DTINIT$Total <- rep("195899")
DTPC_DTINIT$Total <- as.numeric(DTPC_DTINIT$Total)
DTPC_DTINIT$`% Inconsistência` <- (DTPC_DTINIT$n / (DTPC_DTINIT$Total)*100)
DTPC_DTINIT$`% Inconsistência` <- round(DTPC_DTINIT$`% Inconsistência`, 2)

DTPC_DTINIT <- DTPC_DTINIT[,c(-2,-3)]

DTPC_DTINIT <- DTPC_DTINIT%>%
pivot_wider(names_from = Ano,
             values_from = `% Inconsistência`)%>%
unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))

# DATA DE DIAGNOSTICO E DATA DE INICIO DO TRATAMENTO

DTDIAG_DTINIT <- AUX.MAMA%>%

```

```

filter(`Data de diagnóstico` > `Data de início do tratamento`)%>%
group_by(Ano) %>%
arrange(Ano)%>%
tally()

```

```

DTDIAG_DTINIT$Total <- rep("195899")
DTDIAG_DTINIT$Total <- as.numeric(DTDIAG_DTINIT$Total)
DTDIAG_DTINIT$`% Inconsistência` <- (DTDIAG_DTINIT$n / (DTDIAG_DTINIT$Total)*100)
DTDIAG_DTINIT$`% Inconsistência` <- round(DTDIAG_DTINIT$`% Inconsistência`, 2)

```

```
DTDIAG_DTINIT <- DTDIAG_DTINIT[,c(-2,-3)]
```

```

DTDIAG_DTINIT<- DTDIAG_DTINIT%>%
pivot_wider(names_from = Ano,
             values_from = `% Inconsistência`)%>%
unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))

```

```
# LOCALIZACAO E TIPO HISTOLOGICO
```

```
LOCALIZACAO_HISTOLOGIA <- AUX.MAMA%>%
```

```

filter(!`Tipo histológico` %in% c("8201/2",
                                "8201/3",
                                "8204/0",
                                "8230/2",
                                "8314/3",
                                "8315/3",
                                "8500/2",
                                "8500/3",
                                "8501/2",
                                "8501/3",
                                "8502/3",
                                "8503/2",
                                "8503/3",
                                "8506/0",
                                "8507/2",
                                "8508/3",
                                "8513/3",
                                "8520/2",
                                "8520/3",
                                "8521/3",

```

```

"8522/2",
"8522/3",
"8523/3",
"8524/3",
"8530/3",
"8540/3",
"8541/3",
"8543/3",
"8983/0",
"9010/0",
"9011/0",
"9016/0",
"9020/0",
"9020/1",
"9020/3",
"9030/0"))%>%
group_by(Ano) %>%
arrange(Ano)%>%
tally()

LOCALIZACAO_HISTOLOGIA$Total <- rep("195899")
LOCALIZACAO_HISTOLOGIA$Total <- as.numeric(LOCALIZACAO_HISTOLOGIA$Total)
LOCALIZACAO_HISTOLOGIA$`% Inconsistência` <- (LOCALIZACAO_HISTOLOGIA$n /
(LOCALIZACAO_HISTOLOGIA$Total)*100)
LOCALIZACAO_HISTOLOGIA$`% Inconsistência` <- round(LOCALIZACAO_HISTOLOGIA$`%
Inconsistência`, 2)

LOCALIZACAO_HISTOLOGIA<- LOCALIZACAO_HISTOLOGIA[,c(-2,-3)]

LOCALIZACAO_HISTOLOGIA<- LOCALIZACAO_HISTOLOGIA%>%
pivot_wider(names_from = Ano,
values_from = `% Inconsistência`)%>%
unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))

#LOCALIZACAO E LATERALIDADE DO TUMOR

table(AUX.MAMA$TNM)

TNM_ESTADIAMENTO <- AUX.MAMA%>%
filter(TNM %in% c("000", "010", "011", "020", "021", "030", "031", "100", "101", "110",

```

```
"111", "120", "121", "130", "131", "200", "201", "210", "211", "220",
"221", "230", "231", "300", "301", "310", "311", "320", "321", "330",
"331", "400", "401", "410", "411", "420", "421", "430", "431", "441"))
```

```
TNM_INSITU <- TNM_ESTADIAMENTO%>%
```

```
  filter(TNM %in% c("I00", "000"))%>%
```

```
  filter(!Estadiamento clínico do tumor (TNM)` %in% c("0"))%>%
```

```
  group_by(Ano) %>%
```

```
  arrange(Ano)%>%
```

```
  tally()
```

```
table(AUX.MAMAS`Estadiamento clínico do tumor (TNM)`)
```

```
TNM_1A <- TNM_ESTADIAMENTO%>%
```

```
  filter(TNM %in% c("100"))%>%
```

```
  filter(!Estadiamento clínico do tumor (TNM)` %in% c("01", "1", "1A", "1B", "1C", "I"))%>%
```

```
  group_by(Ano) %>%
```

```
  arrange(Ano)%>%
```

```
  tally()
```

```
TNM_2A <- TNM_ESTADIAMENTO%>%
```

```
  filter(TNM %in% c("010", "110", "200"))%>%
```

```
  filter(!Estadiamento clínico do tumor (TNM)` %in% c("02", "2", "2A", "II"))%>%
```

```
  group_by(Ano) %>%
```

```
  arrange(Ano)%>%
```

```
  tally()
```

```
TNM_2B <- TNM_ESTADIAMENTO%>%
```

```
  filter(TNM %in% c("300", "210"))%>%
```

```
  filter(!Estadiamento clínico do tumor (TNM)` %in% c("02", "2", "2B", "II"))%>%
```

```
  group_by(Ano) %>%
```

```
  arrange(Ano)%>%
```

```
  tally()
```

```
TNM_3A <- TNM_ESTADIAMENTO%>%
```

```
  filter(TNM %in% c("020", "120", "220", "310", "320"))%>%
```

```
  filter(!Estadiamento clínico do tumor (TNM)` %in% c("03", "3", "3A", "III"))%>%
```

```
  group_by(Ano) %>%
```

```
  arrange(Ano)%>%
```

```
  tally()
```

```
TNM_3B <- TNM_ESTADIAMENTO%>%
  filter(TNM %in% c("400", "410", "420", "430"))%>%
  filter(!Estadiamento clínico do tumor (TNM) ` %in% c("03", "3", "3B", "III"))%>%
  group_by(Ano) %>%
  arrange(Ano)%>%
  tally()
```

```
TNM_3C <- TNM_ESTADIAMENTO%>%
  filter(TNM %in% c("030", "130", "230", "330", "430"))%>%
  filter(!Estadiamento clínico do tumor (TNM) ` %in% c("03", "3", "3C", "III"))%>%
  group_by(Ano) %>%
  arrange(Ano)%>%
  tally()
```

```
TNM_4A <- TNM_ESTADIAMENTO%>%
  filter(TNM %in% c("001", "011", "021", "031", "101", "121", "131", "201",
    "211", "221", "231", "301", "311", "321", "331", "401",
    "411", "421", "431"))%>%
  filter(!Estadiamento clínico do tumor (TNM) ` %in% c("04", "4", "IV"))%>%
  group_by(Ano) %>%
  arrange(Ano)%>%
  tally()
```

```
TNM_ESTADIAMENTO_FINAL <- bind_rows(TNM_1A, TNM_2A, TNM_2B, TNM_3A, TNM_3B,
  TNM_3C, TNM_4A, TNM_INSITU)
```

```
TNM_ESTADIAMENTO_FINAL <- TNM_ESTADIAMENTO_FINAL %>%
  group_by(Ano) %>%
  summarise(Total = sum(n),
    .groups = "drop_last")
```

```
TNM_ESTADIAMENTO_FINAL$Totalgeral <- rep("195899")
TNM_ESTADIAMENTO_FINAL$Totalgeral <- as.numeric(TNM_ESTADIAMENTO_FINAL$Totalgeral)
TNM_ESTADIAMENTO_FINAL$`% Inconsistência` <- (TNM_ESTADIAMENTO_FINAL$Total /
  (TNM_ESTADIAMENTO_FINAL$Totalgeral)* 100)
TNM_ESTADIAMENTO_FINAL$`% Inconsistência` <- round(TNM_ESTADIAMENTO_FINAL$`%
  Inconsistência`, 2)
```

```
TNM_ESTADIAMENTO_FINAL <- TNM_ESTADIAMENTO_FINAL[,c(-2,-3)]
```



```

TNM_ESTADIAMENTO_FINAL<- TNM_ESTADIAMENTO_FINAL%>%
  pivot_wider(names_from = Ano,
              values_from = `% Inconsistência`)%>%
  unnest(cols = c("2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016"))

# CRIACAO DA TABELA DE MAMA

tabela.total.mama <- bind_rows(DTPC_DTDIAG, DTPC_DTINIT,
                              DTDIAG_DTINIT,LOCALIZACAO_HISTOLOGIA,
                              TNM_ESTADIAMENTO_FINAL)

tabela.total.mama <- tabela.total.mama%>%
  mutate_all(replace_na, 0)

tabela.total.mama <- cbind(Inconsistências = "", tabela.total.mama)

TABELA.INCONSISTENCIA.MAMA <- knitr::kable(tabela.total.mama, align = rep(c('l','c','c')), format.args =
list(decimal.mark = ',', big.mark = "."))%>%
  kable_classic(full_width = F, html_font = "Calibri") %>%
  pack_rows("Data da primeira consulta < Data do diagnóstico", 1, 1) %>%
  pack_rows("Data da primeira consulta > Data de início do tratamento", 2, 2) %>%
  pack_rows("Data do diagnóstico > Data do início do tratamento", 3, 3) %>%
  pack_rows("Localização primária do tumor & Tipo histológico", 4, 4) %>%
  pack_rows("TNM & Estadiamento", 5, 5) %>%
  add_header_above(c(" " = 1, "Câncer de mama" = 9))%>%
  footnote(general = "IntegradorRHC", general_title = "Fonte de dados:")

TABELA.INCONSISTENCIA.MAMA

```