

## Modelos para previsão das taxas de incidência e mortalidade do câncer do colo do útero

### Models for predicting cervical cancer incidence and mortality rates

### Modelos para predecir las tasas de incidencia y mortalidad del cáncer del cuello uterino

Regina Célia Mendes Anffe<sup>1</sup>, Mariano Martinez Espinosa<sup>2</sup>,  
Paulo Cesar Fernandes de Souza<sup>3</sup>, Noemi Dreyer Galvão<sup>4</sup>

#### RESUMO

**Objetivo:** analisar e prever as taxas de incidência e mortalidade do câncer do colo do útero, utilizando modelos de séries temporais. **Método:** estudo ecológico de uma série histórica, realizado na Grande Cuiabá. Os dados de incidência foram referentes ao registro de câncer de base populacional (2000 a 2012) e os de mortalidade (1981 a 2018), obtidos no Sistema de Informação sobre Mortalidade. Os dados foram analisados pelos modelos Arima e Holt. **Resultados:** na série da taxa de incidência do carcinoma *in situ*, o modelo adequado foi AR (1), cujas projeções estimadas (2013-2014) flutuaram entre 15,54 e 16,45 por 100 mil mulheres. Na série da taxa do colo do útero invasor (modelo Holt), as previsões (2013-2014) foram entre 11,28 a 8,9 por 100 mil mulheres. Na taxa de mortalidade, o modelo satisfatório foi ARMA (1,1), cujas previsões (2019-2020) flutuaram entre 9,46 e 9,62 por 100 mil mulheres. **Conclusão:** nos três casos, as taxas previstas apresentaram níveis próximos dos observados, indicando que os modelos permitiram estimar e prever, de maneira adequada. Também foi possível observar um discreto aumento no câncer *in situ* e um declínio na incidência do câncer invasor. Na mortalidade evidenciou uma estabilidade nos resultados e nas previsões.

**Descritores:** Fatores do Tempo; Neoplasias do Colo do Útero; Incidência; Mortalidade; Epidemiologia.

#### ABSTRACT

<sup>1</sup>Bacharel em Estatística pela Universidade Federal de Mato Grosso (UFMT) e em Sistemas de Informação pela Universidade de Cuiabá (UNIC). Especialista em Segurança da Informação. E-mail: [reginanffe@gmail.com](mailto:reginanffe@gmail.com) ORCID ID: <http://orcid.org/0000-0001-7002-6810> **Autor para Correspondência** - Endereço: Rua: Cinco, 232 Boa Esperança, CEP: 78.068-390, Cuiabá, Mato Grosso, Brasil.

<sup>2</sup>Bacharel em Estatística. Doutor em Ciências e Engenharia de Materiais. Professor Titular do Departamento de Estatística do Instituto de Ciências Exatas e da Terra da Universidade Federal de Mato Grosso (UFMT). Cuiabá, Mato Grosso, Brasil. E-mail: [marianomphd@gmail.com](mailto:marianomphd@gmail.com) ORCID ID: <http://orcid.org/0000-0002-0461-5673>

<sup>3</sup>Licenciatura em Pedagogia. Doutorando do Programa de Pós-Graduação em Saúde Coletiva do Instituto de Saúde Coletiva da Universidade Federal de Mato Grosso (UFMT). Técnico Científico de Telessaúde da Secretaria Estadual de Saúde de Mato Grosso (SES/MT). Cuiabá, Mato Grosso, Brasil. E-mail: [pcfernandesouza@gmail.com](mailto:pcfernandesouza@gmail.com) ORCID ID: <http://orcid.org/0000-0002-0032-1125>

<sup>4</sup>Farmacêutica Bioquímica. Doutora em Enfermagem. Professora Titular da Universidade Federal de Mato Grosso (UFMT). Cuiabá, Mato Grosso, Brasil. E-mail: [noemidreyer@hotmail.com](mailto:noemidreyer@hotmail.com) ORCID ID: <http://orcid.org/0000-0002-8337-0669>



Este artigo está licenciado sob forma de uma licença Creative Commons Atribuição 4.0 Internacional, que permite uso irrestrito, distribuição e reprodução em qualquer meio, desde que a publicação original seja corretamente citada.

**Objective:** to analyze and predict cervical cancer incidence and mortality rates, using time series models. **Method:** ecological study of a historical series, with a secondary base, carried out in Greater Cuiabá. The incidence data referred to the population based cancer registry (2000 to 2012), and mortality data (1981 to 2018), obtained from the Mortality Information System. Data were analyzed by Arima and Holt models. **Results:** in the series of incidence rate of carcinoma in situ, the appropriate model was AR (1), whose estimated projections (2013-2014) fluctuated between 15,54 and 16,45 per 100.000 women. In the invasive cervix rate series (Holt model) the forecasts (2013-2014), with values between 11,28 to 8,9 per 100.000 women. In the mortality rate, a satisfactory model was ARMA (1,1), whose forecasts (2019-2020) fluctuated between 9,46 and 9,62 per 100.000 women. **Conclusion:** in the three cases, the predicted rates showed levels close to those observed, indicating that the models allowed estimating and predicting adequately. It was also possible to observe a slight increase in in situ cancer and a decline in the incidence of invasive cancer, in mortality, showing stability in results and predictions.

**Descriptors:** Time Factors; Uterine Cervical Neoplasms; Incidence; Mortality; Epidemiology.

## RESUMEN

**Objetivo:** analizar y predecir tasas de incidencia y mortalidad por cáncer cervicouterino mediante modelos de series temporales. **Método:** estudio ecológico de una serie histórica, con base secundaria, realizado en la Grande Cuiabá. Los datos de incidencia se refieren a registro de cáncer de base poblacional (2000 a 2012), y los de mortalidad (1981 a 2018), obtenidos del Sistema de Información de Mortalidad. Los datos fueron analizados por los modelos de Arima y Holt. **Resultados:** en la tasa de incidencia de carcinoma in situ, el modelo adecuado fue AR(1), cuyas proyecciones estimadas (2013-2014) oscilaron entre 15,54 y 16,45 por 100.000 mujeres. En la serie de tasa del cuello uterino invasivo (modelo de Holt) las previsiones (2013-2014), con valores entre 11,28 y 8,9 por 100.000 mujeres. En la tasa de mortalidad, el modelo satisfactorio fue ARMA(1,1), cuyas previsiones (2019-2020) oscilaron entre 9,46 y 9,62 por 100.000 mujeres. **Conclusión:** en los tres casos, las tasas pronosticadas mostrarán niveles próximos a los observados, indicando que los modelos permitieron estimar y predecir adecuadamente. También fue posible observar un ligero aumento en el cáncer in situ y una disminución en la incidencia del cáncer invasor. En la mortalidade mostrando estabilidad en los resultados y predicciones.

**Descriptorios:** Factores de Tiempo; Neoplasias del Cuello Uterino; Incidencia; Mortalidad; Epidemiología.

## INTRODUÇÃO

O câncer do colo do útero ou câncer cervical é considerado um problema de saúde pública devido a sua magnitude e ocorrência em países em desenvolvimento, a exemplo do Brasil. Em 2020/2022 esta neoplasia foi uma das

mais incidentes em mulheres brasileiras (7,4%), ficando atrás apenas dos cânceres de mama (29,7%) e cólon e reto (9,2%). Perfil semelhante é observado no Estado de Mato Grosso e, particularmente na capital, onde o câncer do colo do útero, ocupando a segunda posição entre os mais

incidentes, e a terceira causa de morte, excluindo os tumores de pele não melanoma, para a população feminina<sup>1-4</sup>.

A principal causa para o desenvolvimento do câncer do colo do útero é a infecção pelo HPV, na qual os tipos de vírus mais frequentes são o 16 e 18. A maioria dessas infecções torna-se indetectável em um a dois anos. A infecção só se torna crônica em uma pequena porcentagem de mulheres, e só uma parcela dessas infecções persistentes por HPV são consideradas como responsáveis pela progressão das lesões precursoras para o câncer invasivo<sup>5</sup>.

O câncer do colo do útero possui desenvolvimento lento, que pode cursar sem sintomas na fase inicial, e é facilmente detectável<sup>6-8</sup>. A vacinação, em conjunto com o exame preventivo (Papanicolau), complementa como ações de prevenção deste câncer. Alerta-se para a proteção do contato com o HPV, uma vez que, a vacina não protege contra todos os subtipos oncogênicos<sup>9-11</sup>.

Nos últimos anos, tem-se observado uma baixa procura das mulheres, principalmente no âmbito do Sistema Único de Saúde (SUS) para a realização do exame de Papanicolau, e isso coloca em evidência a necessidade de conhecer melhor o território e prever

cenários futuros de neoplasias do colo do útero. Assim, o objetivo do estudo foi analisar e prever as taxas de incidência e mortalidade do câncer do colo do útero, utilizando modelos de séries temporais.

## MÉTODO

Trata-se de um estudo ecológico de uma série histórica, considerando os dados do Registro de Câncer de Base Populacional (RCPB) de Cuiabá, disponíveis para a incidência (2000-2012) e do Sistema de Informação sobre Mortalidade (SIM) referente a mortalidade (1981-2018) nos municípios da Grande Cuiabá (Cuiabá e Várzea Grande). A população de estudo foi formada por mulheres, que se apresentaram nos serviços de atenção secundária, historicamente interpretada como procedimentos de média complexidade, residentes nos municípios da Grande Cuiabá e com diagnóstico confirmado do câncer do colo do útero.

Foram incluídas nas análises os casos novos do carcinoma *in situ* do colo do útero e o câncer do colo do útero invasor, correspondendo, respectivamente, aos códigos D06 e C53 da 10ª edição da Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde. Esses dados

foram extraídos do site do Instituto Nacional do Câncer José Alencar Gomes da Silva (INCA) e selecionado o RCBP de Cuiabá. Enquanto que, os dados do SIM do câncer do colo do útero, correspondendo ao código C53 da CID-10<sup>a</sup> foram extraídos do site Atlas da Mortalidade<sup>12-14</sup>. A população do estudo foi constituída por mulheres diagnosticadas com neoplasia ou que foram a óbito por câncer do colo do útero.

Neste trabalho, as variáveis dependentes foram as taxas de incidência e mortalidade e a variável independente o ano calendário. A taxa de incidência, é definida como o número de casos novos de uma doença ou outra afecção de saúde dividido pela população em risco da doença (população exposta) em um espaço

geográfico durante um tempo específico. Pode ser expresso como um número absoluto de casos por ano (o volume de novos pacientes que se apresentam para tratamento) ou a taxa de 100.000 pessoas por ano. Este último fornece uma aproximação ao risco médio de desenvolver câncer e se for necessário comparar o risco de doença entre populações de países, ou grupos étnicos ou diferentes períodos de tempo em um país<sup>15</sup>.

E a mortalidade específica por neoplasias malignas, é a taxa por 100.000 habitantes, da população residente em determinado espaço geográfico, no ano considerado. Estima o risco de morte por neoplasias malignas e dimensiona a sua magnitude como problema de saúde pública<sup>15</sup>.

$$\text{Incidência} = \frac{\text{Número de casos novos ocorridos em um lugar X em determinado tempo}}{\text{Total de indivíduos na população - base (em risco) do lugar X no determinado tempo}} * 10^n$$

$$\text{Mortalidade} = \frac{\text{Número de óbitos de residentes por neoplasia maligna}}{\text{População total residente dor período definido}} * 100.000$$

O procedimento de análise dos dados, iniciou pelas análises dos gráficos temporais, que descreveram o comportamento e identificaram tendências, sazonalidades e variabilidade dos dados. Em seguida, a visualização do correlograma e do correlograma parcial e o teste de KPSS (estacionariedade).

Para as taxas de incidência e mortalidade do câncer do colo do útero

foram aplicados dois modelos estatísticos de séries temporais: modelo de

suavização exponencial de Holt e os modelos de Box-Jenkins.

O modelo de suavização exponencial de Holt é definido por três equações:

$$L_t = \alpha y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$0 \leq \alpha \leq 1, \quad (1)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$0 \leq \beta \leq 1, \quad (2)$$

$$\hat{y}_{t+k} = L_t + kT_t, \quad (3)$$

Nas quais:

$L_t$  é o valor exponencialmente suavizado em período  $t$  para componente nível;

$T_t$  é o valor exponencialmente suavizado em período  $t$  para a componente tendência;

$y_t$  é o dado da série histórica em um período  $t$ ;

$\alpha$  é a constante de suavização, que vai de 0 a 1;

$\beta$  é a constante de suavização para a estimativa de tendência, que vai de 0 a 1;

$y_t$  é o valor real observado no período  $t$  da série;

$k$  número de períodos a serem previstos;

$\hat{y}_{t+k}$  previsão para o período  $t+k$ .

A equação (1) realiza a estimativa do nível, a equação (2) da inclinação e a equação (3) realiza o

cálculo da previsão para os próximos períodos da série temporal<sup>16</sup>.

Se uma série temporal exhibe uma tendência linear (e nenhuma sazonalidade), o modelo de Holt oferece bons resultados, pois possui larga utilização em virtude de sua simplicidade, baixo custo e razoável precisão<sup>17</sup>.

Os modelos de Box-Jenkins utilizam operadores lineares para a representação de uma série temporal ( $Wt$ ), tanto estacionária como não-estacionária, mas que foi diferenciada para se tornar estacionária. Diferentemente de outros modelos, os de Box-Jenkins permitem o estudo das correlações existentes nas séries de tempo, a partir da introdução de modelos auto-regressivos - AR que explicam os processos que dependem dos valores das observações passadas; de modelos de média móvel - MA, que permitem explicar aqueles processos que estão representados pela média ponderada das correlações passadas do resíduo; da combinação destes forma-se o modelo autorregressivo de média móvel - ARMA, e com a introdução do estado de não estacionariedade, formam-se os modelos auto-regressivos integrados e de média móvel - ARIMA<sup>18</sup>.

Como a maioria das séries são não estacionárias, os modelos ARIMA em geral são os mais utilizados<sup>14</sup>. Nestes modelos a componente  $d$  representa o número de vezes que a série original foi diferenciada para se tornar estacionária com a utilização dos operadores de diferenciação  $\nabla^d = (1 - B)^d$  na série não estacionária  $Z_t$ , para, assim, transformá-la em estacionária  $W_t$ ,<sup>16</sup>:

$$W_t = \nabla^d Z_t = (1 - B)^d Z_t$$

A forma geral de um modelo ARIMA  $(p, d, q)$  para uma série não estacionária homogênea, é definida por:

$$\phi_p(B) \cdot \nabla^d Z_t = \theta_q(B) \cdot \alpha_t$$

No qual,  $p$  define a ordem do componente auto-regressivo,  $d$  o número de diferenças satisfeitas para que se torne estacionária e  $q$  a ordem da componente de média móvel.

A construção dos modelos Box-Jenkins é baseada em um ciclo iterativo, na qual a escolha do modelo é feita com base nas autocorrelações e autocorrelações parciais estimadas. São quatro os estágios para a construção do modelo<sup>19</sup>:

- a. Identificação: é a parte mais crítica do método, consiste em tomar diferenças da série, tantas vezes quanto necessárias para se obter uma série estacionária, de modo

que o processo  $\nabla^d Z_t$  seja reduzido a um ARMA  $(p, q)$ . Identificar no processo ARMA  $(p, q)$  resultante por meio da análise das autocorrelações (FAC) e autocorrelações parciais (FACP) estimadas.

- b. Estimação: consiste em estimar os parâmetros  $\phi$  do componente auto-regressivo, os parâmetros  $\theta$  do componente de médias móveis e a variância de  $\alpha_t$ .
- c. Verificação: avalia se o modelo identificado e estimado é adequado para descrever o comportamento dos dados, através da realização de testes de comprovação de sua validade pela analogia das autocorrelações do modelo e os valores dos parâmetros estimados.

Caso o modelo não seja adequado, o ciclo é repetido, voltando-se à fase de identificação. Se o modelo é satisfatório, passa-se para a etapa principal da metodologia: realizar previsões.

- d. Previsão: processo iterativo que visa estender a valores futuros, características

comportamentais de uma série, usando um modelo que foi construído para descrever os valores passados e o valor presente da variável. As melhores e eficazes previsões, são aquelas que apresentam um erro quadrático médio (EQM) mínimo e um período curto<sup>20</sup>.

Pelo princípio da parcimônia deve buscar um modelo com melhor ajuste e menor número possível de parâmetros. E uma forma de escolher entre eles é por meio dos critérios de informação, que tem por finalidade a comparação entre diferentes modelos que foram ajustados a uma mesma série temporal.

Dentre os critérios, um dos mais utilizados é o do Akaike (AIC), neste caso, quanto menor AIC, melhor o modelo se ajusta aos dados. Outro critério utilizado é o erro percentual médio absoluto (MAPE), no qual valores menores indicam um ajuste melhor. Também outro critério que pode ser utilizado é o erro quadrático médio de previsão (EQMP), neste critério o melhor modelo será o que apresentar o menor valor de EQMP<sup>20-23</sup>.

Os dados foram digitados na planilha do Excel e após transferidos

para o *software* livre R, versão 4.0.2<sup>24</sup>. E para a análise dos dados foram utilizadas as bibliotecas *forecast*, *TSA*, *tseries*, *fpp2*, *summarytools* e *snpar* ([https://cran.r-project.org/web/packages/available\\_packages\\_by\\_name.html](https://cran.r-project.org/web/packages/available_packages_by_name.html)).

Este estudo cumpriu as diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos<sup>25</sup>. Após assinatura da Carta de Anuência pelos Secretários Municipais e Diretor da Regional da Saúde, o protocolo desta pesquisa foi submetido para apreciação do Comitê de Ética em Pesquisa (CEP) Saúde da Universidade Federal de Mato Grosso (UFMT), sob CAAE N°. 48121421.0.0000.8124 e parecer de aprovação n. 4.858.521.

## RESULTADOS

Na primeira série temporal explorada com os dados de incidência do carcinoma *in situ* do colo do útero - D06 (TAI-D06), observa-se que, nas informações no período compreendido entre 2000 a 2012, inicialmente, evidenciou-se um crescimento rápido da tendência da série no ano de 2002, com pico em 2005, ultrapassando 44,6 casos diagnosticados/100.000 em mulheres. Após 2005, a tendência da série se

inverteu, alcançando em 2010 um pico relativamente baixo (Figura 1a).

De acordo com a análise residual do modelo da série do carcinoma *in situ* do colo do útero - D06 (TAI-D06), houve a necessidade de uma transformação de Box-Cox para estabilizar a variância dos dados, que foi confirmado pelo teste de Shapiro-Wilk ( $p = 0,0041$ ) o qual evidenciou pouca homogeneidade nos dados. Na transformação, revelou um valor de lambda de  $- 0,9999242$  ( $\lambda = - 0,9999242$ ) (Figura 1b).

Com a variância estabilizada ( $p=0,3642$ ), foram realizados os testes para verificar a existência de tendência ( $p=0,6875$ ) e sazonalidade ( $p=0,0544$ ). Pelos resultados obtidos, a série mostrou ausência de tendência e sem efeito de sazonalidade. Além disso, foi confirmada a estacionariedade da série aplicando-se o teste de KPSS ( $p=0,1385$ ).

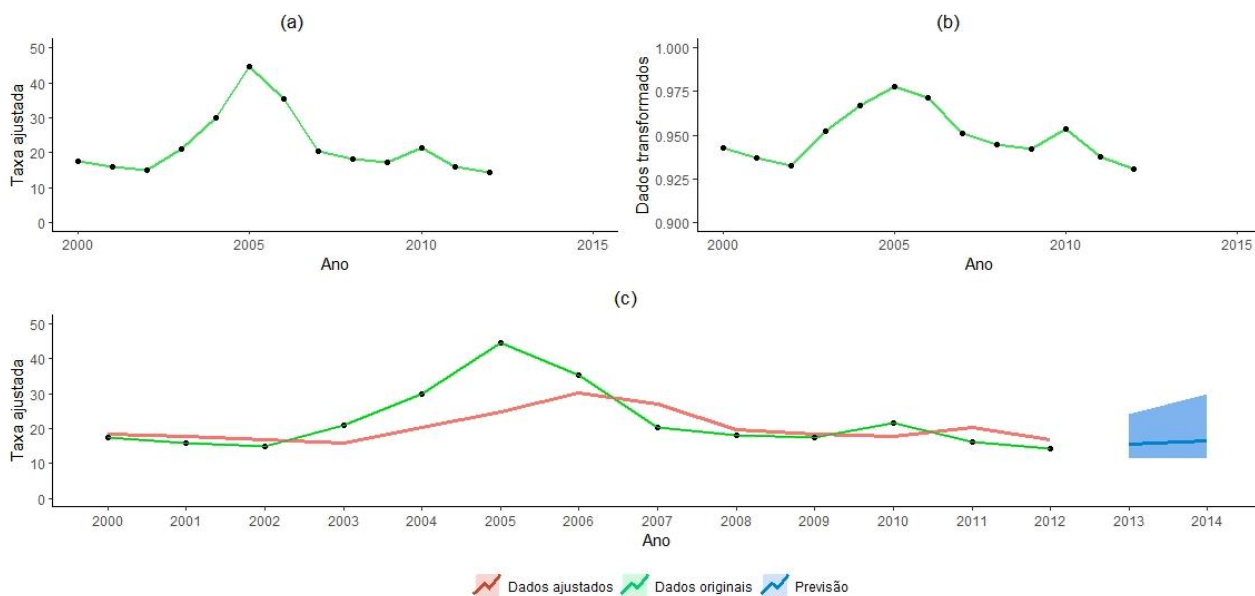


Figura 1 - Comparação gráfica da série temporal: (a) série TAI-D06 com valores originais, (b) série com dados transformados e (c) série original TAI-D06, modelo ajustado AR (1) e a previsão entre os anos 2013 e 2014 fora da amostra.

Os gráficos da FAC e FACP da série estacionária, demonstram *lags* correlacionados com picos significativos um pouco acima dos limites de significância, indicando correlação

diferente de zero. O pico significativo no *lag* 1 na FAC, sugeriu uma componente MA de ordem 1 não sazonal,  $q=1$ , enquanto que o pico significativo no *lag* 1 na FACP, sugeriu uma componente AR



de ordem 1, não sazonal,  $p=1$ . O modelo de ajuste identificado aos dados foi ARIMA (1,0,1) e na sua estimação chegou-se apenas ao modelo AR (1) com seus coeficientes estatisticamente significativos, e foi o mais bem adequado aos dados da TAI-D06, da Grande Cuiabá (Tabela 1).

Dessa forma, o modelo é escrito como:

$$TAI.D06 = 0,94 + 0,66 TAI.D06_{t-1} + \alpha_t$$

No diagnóstico dos resíduos não se observou nenhum *lag* que exceda os limites de significância, logo nenhum deles apresentaram autocorrelação estatisticamente significativa. Aplicando o teste Ljung-Box obteve-se valor-p = 0,1126, o que levou à não existência de autocorrelação entre os resíduos, e de acordo com o observado no gráfico da autocorrelação. A partir do teste de Shapiro-Wilk, encontrou valor-p = 0,0981, cuja hipótese de normalidade dos resíduos não foi rejeitada ao nível de 5% de significância, demonstrando que o modelo de previsão utilizado se apresentou adequadamente a estes pressupostos e poderá ser utilizado para fazer previsões dos valores futuros fora da amostra (2013 a 2014).

Portanto, foi previsto, no período de 2013 a 2014, que a taxa

ajustada de incidência do carcinoma *in situ* do colo do útero - D06 (TAI-D06) cresceu discretamente, atingindo nesses anos valores de 15,54 (IC95%:11,47 a 24,10) e 16,45 (IC95%:11,34 a 29,98) casos diagnosticados/100 mil mulheres (Figura 1c). Neste caso, verificou-se que o modelo escolhido apresentou um ajuste adequado e compatível aos dados, e com a previsão obteve uma boa precisão, dentro dos intervalos de previsão de 95%.

A precisão da performance preditiva do modelo da série do carcinoma *in situ* do colo do útero - D06 (TAI-D06), mostrou um MAPE de 19,54%, EQMP de 48,57 e o AIC de -74,44. Estes valores evidenciaram adequação do modelo ajustado aos dados, revelando um bom desempenho.

A segunda série temporal estudada com dados da incidência, foi a taxa ajustada de incidência com diagnóstico confirmado do câncer do colo do útero invasor - C53 (TAI-C53), que dispunha de informações com períodos compreendidos entre 2000 a 2012.

Na Figura 2, são apresentados os dados originais ou sem transformação do TAI-C53, os quais revelaram uma tendência de queda ao longo do tempo (13 anos estudados) e nota-se, a não

estacionariedade da série temporal e sem sazonalidade. Nesta figura, a primeira observação é o valor mais alto da série com 45,33 casos

diagnosticados/100 mil mulheres no ano de 2000 e o menor valor em 2012 foi de 14,69 casos diagnosticados/100 mil mulheres.

Tabela 1 - Resultado da estimação do modelo AR (1).

Parâmetros	Coefficientes	Erro padrão	Estatística t	Valor-p
$\Phi 1$	0,6630	0,2020	3,2812	0,0010*
Intercepto	0,9463	0,0080	118,1282	<0,0001

$\Phi 1$ : Auto-regressivo de ordem 1; \*: Significativo para valores de  $p < 0,05$  %.



Figura 2 - Representação gráfica da série temporal: série original TAI-C53, modelo ajustado de Holt e a previsão para os anos 2013 e 2014 - para frente da amostra.

Durante a análise da série temporal, não houve necessidade de se aplicar uma transformação na escala dos dados que foi confirmada pelo teste de correlação entre a média e amplitude ( $p=0,8613$ ). Observou-se pelo teste de Cox-Stuart a existência de tendência ( $p=0,0312$ ). Dessa forma, foi necessário realizar uma diferenciação, e novamente realizado teste de Cox-Stuart e

constatada ausência de tendência ( $p=0,6875$ ). O teste de Fisher demonstrou ausência de sazonalidade ( $p=0,1260$ ) e periodograma indicando mudança a cada três períodos. As análises das FAC e FACP, demonstraram lags não correlacionados. A estacionariedade da série foi confirmada pelo teste de KPSS ( $p=0,1714$ ).

A identificação do modelo que melhor representou os dados da TAI-C53, pela metodologia de Box-Jenkins, foi o modelo ARIMA (0,1,0), e por não possuir correlação não havia como usar valores passados para prever o futuro. Por esse motivo, foi necessária a utilização de outra técnica de modelagem, a suavização exponencial de Holt, pois a série apresentou apenas a existência de tendência, sendo assim o mais adequado para a análise.

Os parâmetros de suavização do nível ( $\alpha=0,0001$ ) e da tendência ( $\beta=0,0001$ ) deste modelo, foram otimizados pelo próprio modelo que minimizaram a amplitude dos erros de previsão. Também se verificou os pressupostos de variância constante e de distribuição normal dos resíduos, os quais foram cumpridos. A série temporal ajustada em análise foi adequadamente modelada por este modelo, o qual foi utilizado para prever dois anos para frente (2013-2014), demonstrando que a série seguiu com um comportamento de declínio da taxa de incidência até 2014, e valores de 11,29 (IC95%: 3,35 a 19,22) e 8,94 (IC95%: 1,01 a 16,88) casos/100 mil mulheres, para os respectivos anos considerados, após a aplicação do modelo de Holt.

A precisão da qualidade das previsões do modelo da série do carcinoma invasor do colo do útero - C53 (TAI-C53), revelou um MAPE de 11,28%, EQMP de 11,35 e o AIC de 74,92. Considerando estes valores, o modelo previsto foi o mais adequado aos dados, apresentando um bom desempenho.

A terceira e última série do câncer do colo do útero - C53 (TAM-C53), evidenciou um decréscimo gradual nas taxas ajustadas de mortalidade, na qual a observação mais alta foi de 18,07 óbitos/100 mil mulheres no ano de 1983 e a mais baixa 4,55 óbitos/100 mil mulheres no ano de 2002. Neste ano os dados mostraram uma mudança da tendência, e os óbitos voltaram a crescer, atingindo em 2008, 13,85 óbitos entre 100 mil mulheres (Figura 3). Na série analisada, não houve necessidade de transformação dos dados, pois a variância foi homogênea ( $p=0,388$ ), com ausência de tendência (teste de Cox-Stuart:  $p=0,0635$ ) e sazonalidade (teste de Fisher:  $p=0,1299$ , periodograma a cada 13 períodos). Foi confirmada a estacionariedade da série (KPSS,  $p=0,5167$ ). Os gráficos da FAC e FACP da série estacionária, demonstraram *lags* correlacionados com picos significativos um pouco acima dos limites de significância, indicando correlação

diferente de zero. Isso identificou que o modelo de Box-Jenkins ARIMA (1,0,2) e na sua estimação, chegou-se apenas ao ARMA (1,1) com coeficientes estatisticamente diferentes de zero mais a constante (Tabela 2).

O modelo ARMA (1,1) descreveu o comportamento da TAM-C53 e foi dado por:

$$TAM.C53 = 10,42 + 0,82 TAM.C53_{t-1} - 0,55a_{t-1} + a_t$$

No diagnóstico do modelo ARMA (1,1), observou-se que os resíduos variavam em torno de zero, não havendo

nenhuma informação discrepante. A FAC dos resíduos, evidenciou que todos os lags são não correlacionados, pois não ultrapassaram os limites de confiança e a observação visual do valor-p do teste de Ljung-Box igual a 0,6271, indicando que as observações são conjuntamente não correlacionadas ao longo do tempo. A normalidade dos resíduos também foi verificada com o teste de Shapiro Wilk (p=0,7852).

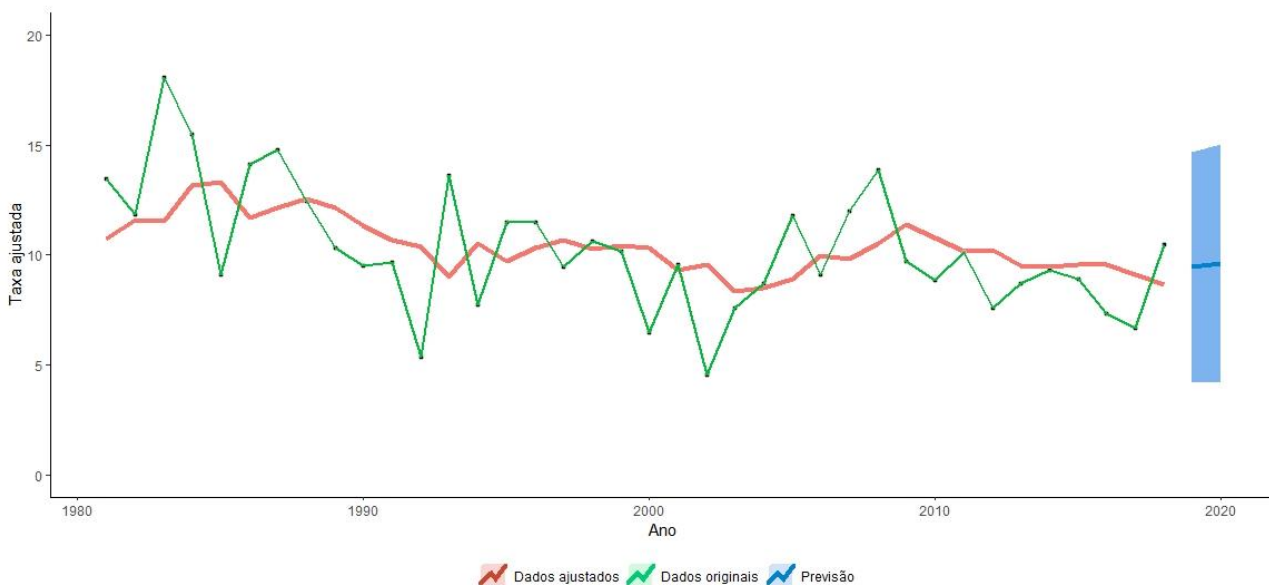


Figura 3 - Representação gráfica da série temporal: série original TAM-C53, modelo ajustado ARMA (1,1) e a previsão entre os anos 2019 e 2020 fora da amostra.

Tabela 2 - Resultado da estimação do modelo ARMA (1, 1).

Parâmetros	Coefficientes	Erro padrão	Estatística t	Valor-p
$\Phi_1$	0,8278	0,1865	4,4386	<0,0001*
$\theta_1$	-0,5550	0,2435	-2,2792	0,0289*
Intercepto	10,4238	1,0004	10,4188	<0,0001

$\Phi_1$ : Auto-regressivo de ordem 1;  $\theta_1$ : Médias móveis de ordem 1; \*: Significativo para valores de  $p < 0,05\%$ .

Uma vez que o modelo ARMA (1,1) satisfaz todas as suposições, pode-se dizer que o modelo se adequa aos dados. Assim, foi previsto para os anos de 2019 e 2020, que a série TAI-C53 teve um comportamento ligeiramente ascendente até 2020, com taxas de 9,46 (IC95%: 4,24 a 14,69) e de 9,62 (IC95%: 4,21 a 15,04) óbitos/100 mil mulheres, para os anos considerados (Figura 3). Neste caso, verificou-se que o modelo escolhido apresentou um ajuste adequado aos dados, e com esta previsão obteve-se também uma precisão adequada, e que os dados previstos e os reais foram compatíveis e ficaram dentro do intervalo de confiança previsto.

Para medir o desempenho preditivo do modelo, utilizou-se os critérios: MAPE (22,32%), EQMP (6,55) e o AIC (187,53) e seus resultados revelaram que o modelo ARMA (1,1) foi adequado para modelar este conjunto de dados, uma vez que apresentou menores valores destas estatísticas.

## DISCUSSÃO

No presente estudo, foi analisado o aumento do comportamento da incidência do carcinoma *in situ* do colo do útero - D06 (TAI-D06), dos dados da base RCPB, no período de 2000 a 2005. Não obstante, a análise da tendência considerando esta base, deve ser utilizada com cuidado, pois, em geral, os dados não apresentam o mesmo comportamento em cada ano. Muito embora, o aumento observado na incidência da neoplasia em análise, pode ter ocorrido pela detecção de mulheres que até então nunca haviam realizado o exame citopatológico.

A partir de 2005 é possível perceber um declínio na tendência da série, pois nesse período, conjuntamente, foi instituído a Política Nacional de Atenção Oncológica e o Pacto pela Saúde, no controle do câncer do colo do útero para compor os planos de saúde estaduais e municipais. Comportamento semelhante, é observado globalmente para um conjunto de países analisados pela Agência Internacional de Pesquisa do Câncer (IARC), os quais evidenciaram decréscimo, mesmo em países em

desenvolvimento, porém, com diferentes magnitudes devido às altas taxas de incidência. O que corrobora com o levantamento de registros e bancos de dados do município de Ladainha (MG), onde houve baixa quantidade de exames citopatológicos realizados dentro da faixa etária de 25 a 64 anos. A dificuldade em alcançar a meta de exames citopatológicos é causada por diversos motivos que atrapalham as estratégias de prevenção e rastreamento precoce para esse tipo de câncer, como medo, falta de tempo, ausência de sintomas, dificuldade de acesso, dentre outros<sup>26-29</sup>.

Baseado nos dados da TAI-D06 da base RCPB, também foi analisada a previsão da incidência do carcinoma *in situ*, e observou-se um pequeno aumento, podendo ser pelo reflexo do número de casos de lesões precursoras diagnosticadas precocemente, já que o exame citológico é trabalhado intensamente nas unidades da APS<sup>29</sup>. Por outro lado, esse crescimento pode indicar também, resultados pouco eficientes de ações de promoção a saúde, as ações de rastreamento para o câncer do colo do útero na região da Grande Cuiabá, onde o fenômeno foi observado, sendo assim, considerados não satisfatórios.

A literatura sobre essa neoplasia ainda é escassa, o que dificulta comparações com outras realidades. No entanto, o comportamento do modelo obtido é semelhante ao comportamento dos modelos ARIMA em estudos de incidência de outras doenças, a exemplo da China (1975 a 2008), quando estudaram a incidência da febre hemorrágica com síndrome renal (HFRS) e obtiveram um modelo ARIMA (0,3,1), com indicativo que a incidência da HFRS teria um ligeiro crescimento nos três anos seguintes (2009-2011), com MAPE de 12,20% e AIC no valor de 116,70<sup>30</sup>.

Por outro lado, o modelo de Holt ajustado para o câncer do colo do útero - TAI-C53, evidenciou a acurácia na previsão. Todo modelo ajustado por suavização, quando há previsão com vários passos à frente dos valores reais, tem uma maior variação e conseqüentemente uma maior amplitude nos intervalos de confiança, mas neste caso, como são dois anos apenas, nota-se que o modelo foi eficiente, pois apresentou resultados próximos da realidade<sup>16</sup>.

Houve declínio significativo da incidência do câncer do colo do útero invasor. O que corrobora com os dados analisados em todos os países analisados pela Agência Internacional de Pesquisa

sobre Câncer - IARC evidenciaram decréscimo, mesmo em países em desenvolvimento, porém, com diferentes magnitudes devido às altas taxas de incidência<sup>27</sup>.

O declínio do resultado evidenciado no presente estudo, pode ter sido consequência das ações do “Programa Viva Mulher”, que visou o controle do câncer do colo do útero com a 1ª. Campanha Nacional de Rastreamento e a 2ª Mobilização Nacional para Detecção Precoce da Doença nos anos 2000 e 2002, respectivamente<sup>31</sup>.

De fato, o declínio observado na incidência do câncer invasor (TAI-C53) e o pequeno aumento da incidência do carcinoma *in situ* (TAI-D06) nos períodos descritos, parecem consistentes. Este comportamento corrobora com os resultados obtidos em outros trabalhos, que analisaram a tendência das taxas de incidência do câncer do colo do útero com base nos dados de RCBP de Fortaleza, Porto Alegre, Recife e São Paulo, para obtenção do número de casos para neoplasia maligna do colo do útero e carcinoma *in situ* do colo do útero por idade. Neste estudo, a tendência foi analisada por regressão polinomial nos períodos de 1990 a 2004 e variaram de 7 a 13 anos. A tendência de

incidência para o câncer do colo do útero foi decrescente em Fortaleza e São Paulo. A análise da tendência do carcinoma *in situ* do colo do útero foi crescente em todas as cidades e grupos etários estudados, com exceção de Recife<sup>27</sup>.

Em relação a mortalidade, o pequeno crescimento pode indicar problemas quanto ao rastreamento, busca ativa, descrédito de alguns serviços de saúde junto a população, morosidade quanto a liberação de resultados de exames, dificuldade em acessar centros especializados em oncologia e ter acompanhamento com especialistas, e em consequência descobrir o diagnóstico ou iniciar o tratamento tardiamente. Já nas capitais brasileiras o cenário é diferente, com decréscimo das taxas padronizadas pela população mundial de 8,4 em 1980, para 4,6/100 mil mulheres em 2020<sup>32,33</sup>. Por outro lado, na Grande Cuiabá, as taxas mostraram-se com valores altos (9,62/100 mil óbitos em mulheres) em comparação às taxas de diferentes regiões do mundo, com menos de 2/100 mil óbitos em mulheres na Ásia Ocidental<sup>2,29</sup>.

No presente estudo, também observou que a série da taxa de mortalidade do câncer do colo do útero - TAM-C53, possui várias flutuações que

vão diminuindo ao longo do tempo, porém, sem muita evolução na queda da taxa de óbitos ao longo dos 38 anos estudados. Pode-se dizer que, pela média das flutuações (10,27), a série TAM-C53 se manteve estável.

Cabe destacar que os modelos estatísticos utilizados no presente trabalho, ajustaram-se adequadamente aos dados, embora tenham sido pouco utilizados para prever a incidência e a mortalidade do câncer do colo do útero. No entanto, dentro da epidemiologia, os modelos de séries temporais estão sendo muito utilizados e bem-sucedidos na aplicação para analisar e prever a incidência de doenças infecciosas e não infecciosas<sup>30,34,35</sup>.

Em Gana, o modelo ARIMA (1,1,1) foi o mais adequado para prever casos maternos mensais com AIC de 117,02. Os casos de mortalidade previstos foram de 10 a 11 mensais para os anos de 2012 a 2013, indicando que a meta do Objetivo de Desenvolvimento do Milênio (ODM) 5, que preconizava a redução significativa das taxas de mortalidade materna, não poderia ser alcançada até 2015, e que as taxas de mortalidade materna continuariam aumentando<sup>36</sup>. Outro estudo semelhante, que utilizou ARIMA (1,0,1) para modelar e prever o número mensal de casos de

dengue hemorrágica (DHF) no sul da Tailândia no período de 1994-2005, mostrou curvas de previsão consistentes com o padrão dos valores reais<sup>37</sup>.

Pelo que se tem conhecimento até o presente, este é o primeiro estudo a sugerir a aplicação do modelo ARIMA, para ajustar e prever à médio prazo a incidência do carcinoma *in situ* do colo do útero - D06 com até 13 observações e a mortalidade por câncer do colo do útero - C53 com até 38 observações. Sabe-se que os modelos ARIMA são mais eficientes em situações que envolvem um pequeno número de observações - comprimento entre 50 (mínimo) e 100 observações<sup>38</sup>.

Uma limitação que precisa ser considerada ao interpretar os resultados, é o intervalo temporal trabalhado nas três séries do câncer do colo do útero, que foi anual. Logo, não foi possível observar se as flutuações se repetiam, sistematicamente, dentro dos anos investigados. Isso significa que, não foi analisada a característica sazonal nas séries temporais estudadas. Assim, em estudos posteriores, poderão ser utilizados dados mensais para prever a incidência e a mortalidade do câncer do colo do útero, a fim de se obter padrões sazonais e previsões com maiores precisões. Entretanto, este estudo



consegue identificar os modelos compatíveis com as séries estudadas, estimando-as e validando-as, com os menores erros para as previsões obtidas.

## CONCLUSÃO

Foi concluído que o modelo ARIMA de Box-Jenkins foi o mais adequado para a taxa de incidência do carcinoma *in situ* do colo do útero e da taxa de mortalidade por câncer do colo do útero. Enquanto que o modelo de Holt foi o mais indicado para a taxa de incidência do câncer do colo do útero invasor, com razoável precisão e simplicidade, pois a série temporal exibe apenas tendência linear e nenhuma sazonalidade. Sugere-se assim, a utilização do modelo de suavização exponencial para o câncer invasor do colo do útero, a fim de estimar o número de casos diagnosticados ou, na medida que, outros casos sejam estimados com a incorporação de novos dados, permitindo que o modelo seja testado novamente, podendo confirmar a existência ou não da sazonalidade e que seja estável e eficaz. Assim, nos três casos analisados, no presente estudo, as taxas previstas correspondem de perto aos níveis observados, sugerindo que os modelos encontrados foram uma boa

representação dos dados da incidência e a mortalidade por câncer do colo do útero na região da Grande Cuiabá.

A aplicação dos modelos também evidenciou um declínio na incidência do câncer invasor e um pequeno aumento no câncer *in situ*, enquanto na mortalidade se observou uma estabilidade nos resultados e nas previsões que refletem os resultados das ações de controle para o câncer nessa localidade.

Acredita-se que o estudo possa contribuir com as ações tomadas pelos gestores, uma vez que, os achados indicam possíveis falhas de comunicação da rede, bem como, indicando que ações mais eficientes devem ser implementadas para que todas as mulheres tenham acesso às ações de prevenção, seguimento e tratamento do câncer do colo do útero.

## REFERÊNCIAS

1. Claro IB, Lima LD, Almeida PF. Diretrizes, estratégias de prevenção e rastreamento do câncer do colo do útero: as expectativas do Brasil e do Chile. *Ciênc Saúde Coletiva*. 2021; 26(10):4497-4509.
2. Bray F, Ferlay J, Soerjomataram I, Siegel R, Torre LA, Jemal A. Global

- câncer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68:394-424.
3. Ministério da Saúde (BR). Instituto Nacional do Câncer José Alencar Gomes da Silva. Estimativa | 2020. Incidência de câncer no Brasil. - Rio de Janeiro: INCA; 2019. [acesso em 2022 dez. 10]. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/estimativa-2020-incidencia-de-cancer-no-brasil.pdf>.
  4. Nakagawa JT, Espinosa MM, Barbieri M, Schirmer J. Carcinoma do colo do útero: taxa de sobrevivência e fatores prognósticos em mulheres no Estado de Mato Grosso. *Acta Paul Enferm.* 2011; 24(5):631-637.
  5. Stanley M. HPV - immune response to infection and vaccination. *Infect Agents Cancer.* 2010; 5(19).
  6. Ministério da Saúde (BR). Instituto Nacional do Câncer José Alencar Gomes da Silva. Controle do câncer do colo do útero: conceito e magnitude. Rio de Janeiro: INCA; 2020.
  7. Ministério da Saúde (BR). Instituto Nacional de Câncer José Alencar Gomes da Silva. Diretrizes brasileiras para o rastreamento do câncer do colo do útero. Rio de Janeiro: INCA; 2016.
  8. Fundação Oswaldo Cruz. Instituto Nacional de Saúde da Mulher, da Criança e do Adolescente Fernandes Figueira. Portal de Boas Práticas em Saúde da Mulher, da Criança e do Adolescente. Postagens: Rastreamento do Câncer do Colo do útero: cobertura, periodicidade e população-alvo. Rio de Janeiro; 2018. [acesso em 2022 dez 11] Disponível em: <https://portaldeboaspraticas.iff.fiocruz.br/atencao-mulher/rastreamento-do-cancer-do-colo-do-uterocobertura-periodicidade-e-populacao-alvo/>.
  9. Ministério da Saúde (BR). Instituto Nacional de Câncer José Alencar Gomes da Silva. Ações de controle: prevenção do câncer do colo do útero. Rio de Janeiro: INCA; 2020.
  10. Duavy LM, Batista FLR, Jorge MSB, Santos JBF. A percepção da mulher sobre o exame preventivo do câncer cérvico-uterino: estudo de caso. *Ciênc Saúde Coletiva.* 2007; 12(3):733-742.
  11. Casarin MR, Piccoli JCE. Educação em saúde para prevenção do câncer de colo do útero em mulheres do município de Santo Ângelo/RS. *Ciênc*

- Saúde Coletiva. 2021; 16(9):3925-3932.
12. World Health Organization (WHO). International statistical classification of diseases and related health problems 10th Revision; 2019.
  13. Ministério da Saúde (BR). Instituto Nacional de Câncer José Alencar Gomes da Silva. Registros de Câncer de Base Populacional. Tabulador das informações dos RCBP. Rio de Janeiro: INCA; 2020.
  14. Ministério da Saúde (BR). Instituto Nacional do Câncer José Alencar Gomes da Silva. Atlas on-line de mortalidade por câncer. Rio de Janeiro: INCA; 2020.
  15. Rede Interagencial de Informações para a Saúde. Indicadores básicos para a saúde no Brasil: conceitos e aplicações. 2ª ed. Brasília: RIPSAs; 2008.
  16. Morettin PA, Toloi CMC. Análise de séries temporais. Ed Edgard Blucher Ltda. 3ª Ed. São Paulo; 2018.
  17. Alves CC, Hoepers E, Corazza EJ, et al. Aplicação de métodos estatísticos com suavização exponencial dupla e tripla para previsão de demanda na gestão de estoques. Rev Prod Online. 2019; 19(3):1001-1026.
  18. López AR. Análise de intervenção de uma ação do programa de rastreamento do câncer de mama nas séries de mortalidade no Brasil e regiões [tese]. Rio de Janeiro: UFRJ/COPPE/Programa de Engenharia Biomédica; 2018.
  19. Mi Domenech. Identificação de um modelo ARIMA (p,d,q). Maringá; 2019 [acesso em 2020 dez. 10]. Disponível em: <https://www.midomenech.com.br/identificacao-de-um-modelo-arimadq/>.
  20. Coelho Júnior LM, Melquíades TF, Martins KLC, Santos Júnior EP, Freitas GP. Previsão do consumo de eletricidade no nordeste brasileiro. Rev Engevista. 2018; 20(3):408-423.
  21. Cai T, Li X. Forecasting methods to reduce inventory level in supply chain. J Appl Math Phys. 2022; 10(2):301-310.
  22. Moraes AC, Soares TS, Cruz ES. Height, volume and form factor equations for tectona grandis l. f. in Alta Floresta (MT). Sci Agrar Parana. 2020; 19(1):27-37.
  23. Holanda Filho JD. Métodos quantitativos de previsão de vendas: estudo de caso em uma indústria cearense de alimentos. [monografia]. Fortaleza Universidade Federal do Ceará, Curso de

- Graduação em Administração; 2018. 74 p.
24. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria; 2022 [acesso em 2022 out. 12]. Disponível em: <https://www.R-project.org/>.
  25. Ministério da Saúde (BR). Conselho Nacional de Saúde. Resolução nº. 466, de 12 de dezembro de 2012. Aprova diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. Brasília: Diário Oficial da União; 2012.
  26. Ministério da Saúde (BR). Instituto Nacional do Câncer José Alencar Gomes da Silva. Controle do Câncer do Colo do Útero: Histórico das ações. Rio de Janeiro: INCA; 2020.
  27. Ayres ARG, Silva GA, Guimarães RM. Tendência da incidência de câncer do colo do útero invasor em quatro capitais brasileiras: dados dos registros de câncer de base populacional, 1990-2004. Cad Saúde colet. 2013; 21(3):289-95.
  28. Silva VM, Vasconcelos KP, Diniz DDS, Farias GM, Oliveira AEA. Fatores que influenciam a não adesão da mulher ao exame Papanicolau: revisão de literatura. Rev Inter Saúde, 2021; 8(único):326-340.
  29. Ministério da Saúde (BR). Instituto Nacional de Câncer. Plano de ação para redução da incidência e mortalidade por câncer do colo do útero. Rio de Janeiro; 2010. [acesso em 2020 dez. 10]. Disponível em: [bvsms.saude.gov.br/bvs/publicacoes/plano\\_acao\\_reducao\\_cancer\\_colo.pdf](https://bvsms.saude.gov.br/bvs/publicacoes/plano_acao_reducao_cancer_colo.pdf).
  30. Liu Q, Liu X, Jiang B, Yang W. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. BMC Infect Dis. 2011; 11(218).
  31. Ministério da Saúde (BR). Instituto Nacional de Câncer. Controle do Câncer do Colo do Útero: Histórico das ações. - Rio de Janeiro: INCA; 2020.
  32. Ministério da Saúde (BR). Instituto Nacional do Câncer José Alencar Gomes da Silva. Estimativa | 2018 incidência de câncer no Brasil. Coordenação de Prevenção e Vigilância. - Rio de Janeiro: INCA; 2017.
  33. Ministério da Saúde (BR). Instituto Nacional do Câncer. Dados e números sobre câncer do colo do útero. Relatório anual 2022. - Rio de

- Janeiro: INCA; 2022. [acesso em 2022 dez. 14]. Disponível em: <https://www.inca.gov.br/publicacoes/relatorios/dados-e-numeros-sobre-cancer-do-colo-do-utero-relatorio-anual-2022>.
34. Anokye R, Acheampong E, Owusu I, et al. Tie series analysis of malária in Kumasi: using ARIMA models to forecast future incidence. *Cogent Soc Sciences*. 2018; 4(1):1461544.
35. Tian CW, Wang H, Luo XM. Time-series modelling and forecasting of hand, foot and mouth disease cases in China from 2008 to 2018. *Epidemiol Infect*. 2019; 147:e82.
36. Adedia D, Nanga S, Appiah SK, Lotsi A, Abaye DA. Box-Jenkins methodology in predicting maternal mortality records from a public health facility in Gana. *Open J Appl Sciences*. 2018; 8(6):189-202.
37. Promprou S, Jaroensutasinee M, Jaroensutasinee K. Forecasting dengue haemorrhagic fever cases in southern Thailand using ARIMA models. *Dengue Bulletin*. 2006; 30:99-106.
38. Montgomery DC, Jennings CL, Kulahci M. Introduction to time series analysis and forecasting. 2nd Edition. Wiley: Hoboken; 2015.

**Financiamento:** Os autores declaram que não houve financiamento.

**Conflito de interesses:** Os autores declaram não haver conflito de interesses.

**Participação dos autores:**

- **Concepção:** Anffe RCM, Espinosa MM, Souza PCF, Galvão ND.
- **Desenvolvimento:** Anffe RCM, Espinosa MM, Souza PCF, Galvão ND.
- **Redação e revisão:** Anffe RCM, Espinosa MM, Souza PCF, Galvão ND.

**Como citar este artigo:** Anffe RCM, Espinosa MM, Souza PCF, Galvão ND. Modelos para previsão das taxas de incidência e mortalidade do câncer do colo do útero. *Journal Health NPEPS*. 2022; 7(2):e10446.

Submissão: 20/09/2022

Aceito: 01/12/2022