

# Automatização da codificação e seleção das causas de óbitos no Peru: estudo descritivo, 2016-2019

Javier Vargas-Herrera<sup>1</sup>, Janet Miki<sup>2</sup>, Liliana López Wong<sup>3</sup>, Jorge Miranda Monzón<sup>3</sup>, Rodolfo Villanueva<sup>4</sup>

<sup>1</sup>Universidad Nacional Mayor de San Marcos, Unidad de Telesalud, Lima, Peru

<sup>2</sup>Vital Strategies, Programa de Registro Civil y Estadísticas Vitales, Lima, Peru

<sup>3</sup>Ministerio de Salud del Perú, Oficina General de Tecnologías de la Información, Lima, Peru

<sup>4</sup>Universidad Alas Peruanas, Escuela de Ingeniería de Sistemas e Informática, Lima, Peru

## RESUMO

**Objetivo:** descrever o desempenho de um *software* na seleção automática da causa básica de morte no Peru, entre 2016 e 2019. **Métodos:** estudo descritivo do desempenho de um *software* na seleção automatizada da causa básica de morte ao longo dos anos (teste qui-quadrado de tendência) e da correlação entre a forma de declaração de óbito e o desempenho do *software* (coeficientes de correlação e determinação). **Resultados:** foram analisadas 446.217 declarações de óbito; a proporção de declarações com causa básica de morte aumentou de 69,6%, em 2016, para 78,8%, em 2019 (p-valor < 0,001); observou-se correlação linear direta entre as declarações eletrônicas de óbito e o desempenho do *software* (coeficiente de correlação = 0,95; R<sup>2</sup> = 0,89). **Conclusão:** o *software* apresentou bom desempenho na seleção automática da causa básica de morte, com aumento significativo entre 2016 e 2019.

**Palavras-chave:** Causas de Morte; Registros de Mortalidade; Classificação Internacional de Doenças; Sistemas de Informação em Saúde; Tecnologia da Informação; Epidemiologia Descritiva.

## INTRODUÇÃO

A informação sobre a mortalidade é útil para medir o impacto das intervenções em saúde.<sup>1</sup> A Organização Mundial da Saúde (OMS) estabeleceu um conjunto de diretrizes e regras para a codificação e seleção da causa básica da morte com vistas à certificação médica; não obstante, há uma preocupação com a confiabilidade dessa codificação.<sup>2-5</sup> A concordância da codificação, entre codificadores, é maior nos países onde foi implementada uma política de engajamento e retenção de codificadores específicos, atingindo quase 80%,<sup>2,6</sup> semelhantemente ao desempenho dos *softwares* de codificação automática.

A aplicação incorreta da codificação e definição da causa básica de morte afeta a qualidade dos dados e sua comparabilidade entre países. Existem *softwares* para automatizar essas notificações, utilizados por cada vez mais países,<sup>7-10</sup> a apontar uma tendência para o uso de inteligência artificial nesse processo.<sup>11</sup> Na América Latina, há experiências de implementação de *softwares* com esse propósito, do México ao Brasil, Chile, Colômbia e Peru.<sup>12</sup>

Precisamente no Peru, em 2017, foi implementado o Sistema Informático Nacional de Defunciones (SINADEF), permitindo que as declarações de óbito se realizem via formulários eletrônicos, em tempo real, melhorando a qualidade e oportunidade das notificações.<sup>13,14</sup> O Ministério da Saúde do Peru (MINSA) também decidiu adotar o *software* Iris, desenvolvido por um consórcio internacional liderado pelo Instituto Alemão de Informação e Documentação Médica, que utiliza um algoritmo baseado nas regras da Décima Revisão da Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde (CID-10), visando automatizar a codificação e seleção da causa básica de morte.<sup>15</sup>

O objetivo desta nota de pesquisa foi descrever o desempenho de um *software* na seleção automática da causa básica de morte no Peru, entre 2016 e 2019.

### Contribuições do estudo

<b>Principais resultados</b>	Observou-se bom desempenho do <i>software</i> na seleção automática da causa básica de morte, que passou de 69,6%, em 2016, para 78,8%, em 2019. Este resultado correlacionou-se com o uso de declarações de óbito <i>online</i> pelos médicos.
<b>Implicações para os serviços</b>	A codificação e a seleção automática das causas de morte melhoram a produtividade e a oportunidade das informações, contribuindo para a qualidade do sistema de informações de país.
<b>Perspectivas</b>	É necessário analisar a concordância entre os termos médicos nos dicionários dos <i>softwares</i> utilizados nos países da América do Sul, em nome de uma melhora na padronização e comparabilidade das informações sobre causas de morte.

## MÉTODOS

### Delineamento

Trata-se de estudo descritivo do desempenho de um *software* na seleção automática da causa básica de morte no Peru, entre os anos de 2016 e 2019. Esse desempenho foi definido como a capacidade do *software* para obter a causa básica da morte

### Contexto

Até 2016, todos os óbitos contavam com declarações de óbito em formato de papel, transcritas para um *software-desktop* denominado Eventos Vitais e enviadas para o MINSA na forma de arquivos. Em 2017, o SINADEF foi implantado na Web. O sistema permite o registro das declarações de óbito em duas modalidades: digitadas *online* diretamente pelos médicos; ou transcritas em formatos de papel. Em 2018, adotou-se

o *software* Iris, e, à partir de 2016, os bancos de dados de mortalidade passaram a ser processados utilizando-se desse aplicativo para obter as causas básicas de morte. O dicionário Iris foi adaptado com 12.246 termos médicos em linguagem natural, tendo por referência as causas de óbito preenchidas diretamente pelos profissionais médicos.

### Participantes

O presente estudo incluiu os óbitos ocorridos no Peru entre os anos de 2016 e 2019.<sup>16</sup> Foram excluídos os óbitos não declarados e os que não estavam disponíveis no momento do processamento pelo *software*.

### Variáveis

As variáveis investigadas foram: declaração do óbito processada (com causa básica de morte; sem causa básica de morte); termos médicos registrados (com código CID-10; sem código CID-10); tipo de erro na declaração rejeitada pelo *software* (de sintaxe; de código; de sistema); forma de declaração do óbito (em papel; *online*); e ano do óbito (2016 a 2019).

### Fontes de dados e mensuração

A fonte dos dados constituiu-se dos bancos de dados das declarações de óbito referentes ao período de 2016 a 2019, fornecidos pelo MINSA em planilhas. Os dados foram processados pelo Iris nas seguintes datas: banco de dados de mortalidade de 2016, 01/06/2018; banco de 2017, 26/04/2019; banco de 2018, 20/06/2020; e banco de 2019, 22/06/2021.

### Controle de viés

Os registros dos bancos de mortalidade passaram por controle de qualidade, para eliminar eventuais registros duplicados ou modificar registros com dados inconsistentes.

### Métodos estatísticos

As variáveis obtidas após o processamento com o Iris foram disponibilizadas em tabelas de distribuição de frequência simples. Para análise de tendência, foi utilizado o teste qui-quadrado de tendência. As variáveis dependentes consideradas foram o índice de desempenho do Iris (número de declarações de óbito com causa básica de morte dividido pelo número total de declarações de óbito) e o índice de desempenho do Iris na codificação CID-10 (número de termos médicos com códigos CID-10 dividido pelo número total de termos médicos); e a variável independente, o ano de ocorrência do óbito. Para analisar a correlação linear entre a forma de declaração do óbito e o desempenho do Iris, foram utilizados o coeficiente de correlação de Pearson e o coeficiente de determinação ( $R^2$ ). O nível de significância utilizado foi de 5%. O *software* Excel® 2016 foi utilizado nas análises.

### Aspectos éticos

O estudo baseou-se na análise de variáveis contempladas nos bancos de dados de mortalidade do MINSA, também disponíveis na Plataforma Nacional de Dados Abertos <https://www.datosabiertos.gob.pe/> e que não incluem informações capazes de permitir a identificação das pessoas falecidas.

## RESULTADOS

Entre 2016 e 2019, foram analisados 446.217 óbitos de residentes em todas as regiões do Peru, registrados no banco de dados de mortalidade do MINSA. Esse quantitativo correspondeu a 67% das mortes estimadas para o período de estudo. Foram excluídos os óbitos que não se encontravam cadastrados no sistema de mortalidade, no momento do processamento (Figura 1).

Observou-se aumento progressivo do índice de desempenho do *software*, cujo percentual de declarações de óbito processadas passou de 69,6%, em 2016, para 78,8%, em 2019 ( $p$ -valor < 0,001) (Tabela 1).

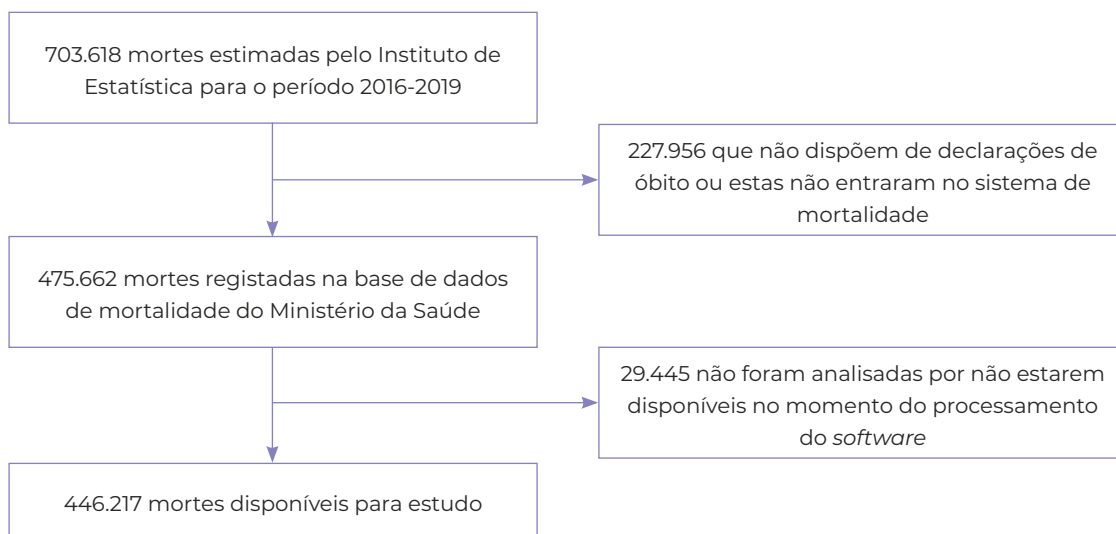


Figura 1 – Processo de seleção das mortes estudadas, Peru, 2016-2019

Tabela 1 – Desempenho do software na seleção da causa básica de morte e processamento de declarações de óbito (n = 446.217), Peru, 2016-2019

Desempenho do software	2016		2017		2018		2019		p-valor <sup>a</sup>
	N	%	N	%	N	%	N	%	
<b>Declarações de óbito processadas</b>									< 0,001
Com causa básica de morte	67.697	69,6	86.848	71,8	79.979	74,8	95.405	78,8	
Sem causa básica de morte	29.542	30,4	34.174	28,2	26.926	25,2	24.653	21,2	
Total	97.241	100,0	121.024	100,0	106.905	100,0	121.047	100,0	
<b>Termos médicos processados</b>									< 0,001
Com código CID-10	222.446	87,2	321.904	89,8	299.988	91,1	346.635	92,7	
Sem código CID-10	32.641	12,8	36.623	10,2	29.401	8,9	27.297	7,3	
Total	255.087	100,0	358.527	100,0	329.389	100,0	373.932	100,0	
<b>Declarações de óbito rejeitadas segundo o tipo de erro</b>									< 0,001
Erro de sintaxe	-	-	744	2,3	545	2,1	582	2,4	
Erro de código	26.819	90,8	30.337	94,0	24.616	94,9	23.130	93,8	
Erro de sistema	2.725	9,2	1.185	3,7	790	3,0	941	3,8	
Total	29.544	100,0	32.266	100,0	25.951	100,0	24.653	100,0	
<b>Forma de declaração de óbito</b>									< 0,001
Em papel	96.605	100,0	85.986	71,0	32.227	30,1	25.397	21,0	
Eletrônico – online	636	0,0	35.038	29,0	74.678	69,9	95.650	79,0	
Total	97.241	100,0	121.024	100,0	106.905	100,0	121.047	100,0	

a) Teste qui-quadrado de tendência.

Observou-se, também, tendência crescente do índice de desempenho do *software* na codificação CID-10, com aumento progressivo da proporção desse desempenho: de 87,2%, em 2016, para 92,7%, em 2019 ( $p$ -valor < 0,001). Entretanto, a maior proporção de erros nos registros que o *software* não conseguiu processar são erros de codificação (Tabela 1).

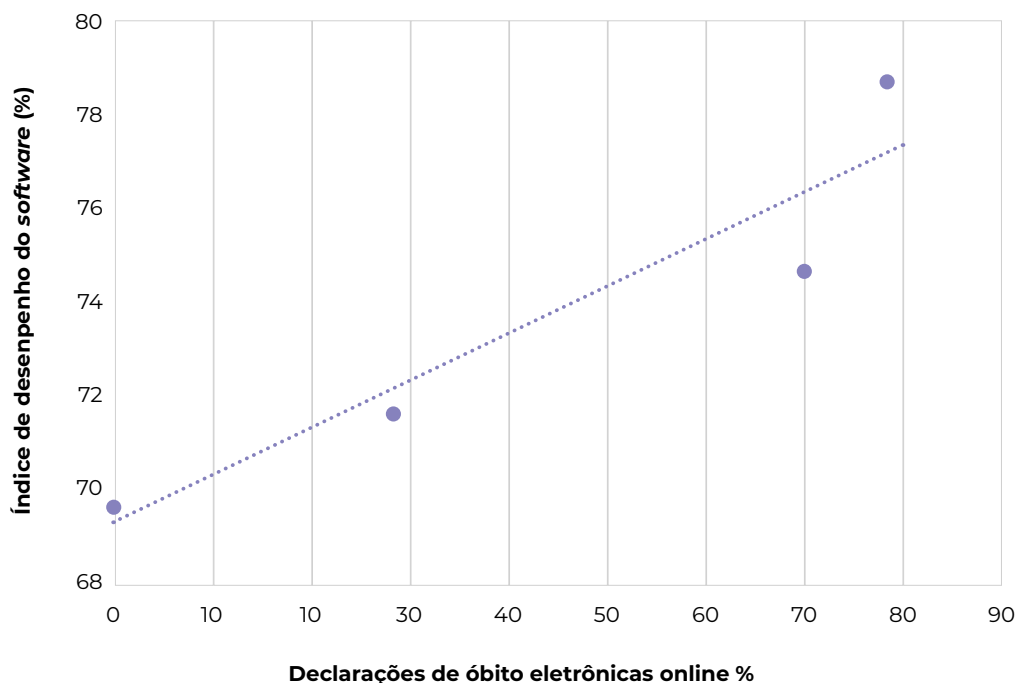
Observou-se correlação linear direta entre a proporção de declarações de óbito preenchidas diretamente por médicos no SINADEF e o desempenho do Iris: coeficiente de correlação de Pearson = 0,95;  $R^2 = 0,89$  (Figura 2).

## DISCUSSÃO

No período estudado, o desempenho do *software* aumentou em função do progresso de sua capacidade para codificar, com a CID-10, os termos usados pelos médicos na notificação das causas de morte. A proporção de declarações preenchidas diretamente pelos

médicos correlacionou-se com o desempenho do *software*. A implementação do SINADEF teve papel importante nesse processo, ao permitir a construção de um dicionário adaptado ao contexto peruano e, a cada ano, contribuir com um maior número de termos médicos. Estudos têm demonstrado as melhoras alcançadas após o treinamento dos médicos no preenchimento das declarações de óbito, e na qualidade dos dados por eles registrados,<sup>17,18</sup> além da possibilidade de aprimorar o próprio desempenho do *software*.

Uma limitação do estudo reside na falta de uma análise da concordância entre o *software* e a aplicação de regras por codificadores experientes. No Peru, esse tipo de análise é difícil porque, antes da implementação do SINADEF, a maior parte das declarações de óbito era codificada por pessoal sem treinamento formal na CID-10. Uma segunda limitação encontra-se no fato de, neste estudo, o desempenho do Iris não ser analisado por variáveis sociodemográficas ou grupo de doenças.



**Figura 2 – Correlação entre a proporção de declarações de óbito produzidas em formato eletrônico *online* e o índice de desempenho do *software* na seleção da causa básica de óbito, Peru, 2016-2019**

Notas: Correlação linear de Pearson = 0,95; Coeficiente de determinação  $R^2 = 0,89$ .

O maior desempenho do *software* foi próximo de 80%. Um estudo realizado em São Paulo, no ano de 2010, em uma amostra de 666 óbitos com o propósito de testar o dicionário do *software* em português, constatou desempenho de 95%.<sup>19</sup> Na Holanda, durante um estudo de implementação da automatização, com dados de 134.262 óbitos ocorridos em 2009, verificou-se um aumento de desempenho de 17% no primeiro lote para 69% no último lote, após uma série de melhorias no dicionário.<sup>20</sup> Na Espanha, um estudo para avaliar o impacto da automatização dos registros da causa de morte na mortalidade na comunidade autônoma (macrorregião geopolítica) de Navarra, sobre 5.060 óbitos ocorridos em 2014, identificou um desempenho de 90%.<sup>21</sup> Ao se avaliar a usabilidade do Iris em uma pequena amostra de mortes em Burkina Faso, encontrou-se um desempenho de 90%.<sup>10</sup>

No presente estudo, a codificação automática de termos médicos de causas de morte atingiu 93%, enquanto na Itália, em 2016, essa proporção foi de 78%.<sup>22</sup>

A maioria dos erros que levaram o *software* a rejeitar uma declaração de óbito aconteceram na codificação: erros de digitação, de ortografia ou com caracteres estranhos.<sup>19</sup> Entretanto, o

*software* também apresenta dificuldades para codificar adequadamente as causas externas, porque os médicos legistas utilizam uma gama bastante variada de causas de óbito, o que afeta a eficiência do dicionário.<sup>20</sup>

Há um movimento global orientado à seleção automatizada da causa básica de morte. Quase todos os países da União Europeia já utilizam o Iris. Na América Latina, o *software* encontra-se em processo de implantação, em vários países. No Brasil, ele foi incorporado a um aplicativo móvel para médicos, com o objetivo de melhorar o preenchimento da Declaração de Óbito (DO).<sup>23,24</sup>

Conclui-se que há uma tendência de melhoria no desempenho do *software* para a seleção da causa básica de morte no Peru. Aparentemente, isso se deve à implantação do SINADEF e à otimização do dicionário de termos médicos. Mais estudos sobre o Iris são necessários, no sentido de se avaliar o impacto do *software* nas estatísticas de mortalidade. Considerando-se que sua implementação na região melhorará a comparabilidade dos dados, é mister estudar a concordância entre os termos médicos dos dicionários utilizados nos países da América do Sul.

### CONTRIBUIÇÃO DOS AUTORES

Vargas-Herrera J e Miki J contribuíram na concepção e delineamento do estudo, análise e interpretação dos resultados, redação e revisão crítica do conteúdo do manuscrito. Wong LL, Monzón JM e Villanueva R contribuíram na análise e interpretação dos dados, e revisão do conteúdo do manuscrito. Todos os autores aprovaram a versão final do manuscrito e são responsáveis por todos os seus aspectos, incluindo a garantia de sua precisão e integridade.

### CONFLITOS DE INTERESSE

Os autores declaram não haver conflitos de interesse.

### AGRADECIMENTO

Agradecemos ao doutor Benjamin Clapham, da organização Vital Strategies, pelo apoio na preparação do artigo, e aos doutores Juan Cortez-Escalante e Jaqueline García de Almeida Ballesterero pela revisão da tradução do artigo para o português.

**Correspondência:** Javier Vargas-Herrera | [jvargash@unmsm.edu.pe](mailto:jvargash@unmsm.edu.pe)

**Recebido em:** 29/03/2023 | **Aprovado em:** 07/07/2023

**Editor associado:** Taís Freire Galvão

## REFERÊNCIAS

1. Suthar AB, Khalifa A, Yin S, Wenz K, Ma Fat D, Mills SL, et al. Evaluation of approaches to strengthen civil registration and vital statistics systems: a systematic review and synthesis of policies in 25 countries. *PLoS Med.* 2019;16(9):e1002929. doi: 10.1371/journal.pmed.1002929
2. Antini C, Rajs D, Muñoz-Quezada MT, Mondaca BAL, Heiss G. Reliability of cause of death coding: an international comparison. *Cad Saude Publica.* 2015;31(7):1473-82. doi: 10.1590/0102-311X00099814
3. Минаева А, Вайсман К. The peculiarities of coding and the determination of the primary cause of death from the diseases induced by the human immunodeficiency virus in accordance with ICD-10. *Sud Med Ekspert.* 2015;58(2):27-9. doi: 10.17116/sudmed201558227-29
4. Winkler V, Ott JJ, Becher H. Reliability of coding causes of death with ICD-10 in Germany. *Int J Public Health.* 2010;55(1):43-8. doi: 10.1007/s00038-009-0053-7
5. Gamage USH, Adair T, Mikkelsen L, Mahesh PKB, Hart J, Chowdhury H, et al. The impact of errors in medical certification on the accuracy of the underlying cause of death. *PLoS One.* 2021;16(11):e0259667. doi: 10.1371/journal.pone.0259667
6. Harteloh P, Bruin K, Kardaun J. The reliability of cause-of-death coding in The Netherlands. *Eur J Epidemiol.* 2010;25(8):531-8. doi: 10.1007/s10654-010-9445-5
7. Eckert O. Electronic coding of death certificates. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.* 2019;62(12):1468-75. doi: 10.1007/s00103-019-03045-2
8. Rey G. Death certificate data in France: production process and main types of analyses. *Rev Med Interne.* 2016;37(10):685-93. doi: 10.1016/j.revmed.2016.01.011
9. Barro SG, Rey G, Staccin P. Study of the usability of an automated coding software for causes of death in an African context. *Stud Health Technol Inform.* 2019;264:1978-9. doi: 10.3233/SHTI190743
10. Rey G, Bounebache K, Rondet C. Causes of deaths data, linkages and big data perspectives. *J Forensic Leg Med.* 2018;57:37-40. doi: 10.1016/j.jflm.2016.12.004
11. Falissard L, Morgand C, Roussel S, Imbaud C, Ghosn W, Bounebache K, et al. A deep artificial neural network-based model for prediction of underlying cause of death from death certificates: algorithm development and validation. *JMIR Med Inform.* 2020;8(4):e17125. doi: 10.2196/17125
12. Pan American Health Organization. sistema de Codificación Automatizada de causa de muerte "Iris" [Internet]. Washington: Pan American Health Organization; 2015 [cited 2023 May 29]. Available from: <https://www3.paho.org/relacsis/index.php/es/biblioteca/1126-documentos-sistema-de-codificacion-automatizada-de-causa-de-muerte-iris/>
13. Vargas-Herrera J, Ruiz KP, Nuñez GG, Ohno JM, Pérez-Lu JE, Huarcaya WV, et al. Resultados preliminares del fortalecimiento del sistema informático nacional de defunciones en Perú. *Med Exp Salud Publica.* 2018;35(3):505-14. doi: 10.17843/rpmesp.2018.353.3913
14. Vargas-Herrera J, Monzón MJ, Wong LL, Ohno JM. La cobertura de muertes con certificación médica en el Perú, 2012-2019. *An Fac med.* 2022;83(2):123-9. doi: 10.15381/anales.v83i2.23011



15. Iris Institute. Iris user reference manual V5.8.1S2 [Internet]. Bonn: Iris Institute; 2022 [update 2022 Feb 3; cited 2023 Feb 15]. Available from: [https://www.bfarm.de/SharedDocs/Downloads/EN/Code-Systems/iris-institute/manuals/iris-user-reference-manual-v5-8-1s2\\_.pdf.html;jsessionid=A6DF42A5C4F65BD129823DB7BCDE50E5.internet281?nn=922496&cms\\_dIDConfirm=true&cms\\_calledFromDoc=922496](https://www.bfarm.de/SharedDocs/Downloads/EN/Code-Systems/iris-institute/manuals/iris-user-reference-manual-v5-8-1s2_.pdf.html;jsessionid=A6DF42A5C4F65BD129823DB7BCDE50E5.internet281?nn=922496&cms_dIDConfirm=true&cms_calledFromDoc=922496)
16. Instituto Nacional de Estadística e Informática. Boletín Especial n° 24 - Perú: estimaciones y proyecciones de la población nacional, por año calendario y edad simple, 1950-2050. Lima: Instituto Nacional de Estadística e Informática; 2019 [cited 2023 May 23]. Available from: [https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1681/](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1681/)
17. Miki J, Rampatige R, Richards N, Adair T, Cortez-Escalante J, Vargas-Herrera J. Saving lives through certifying deaths: assessing the impact of two interventions to improve cause of death data in Perú. *BMC Public Health*. 2018;18(1):1329. doi: 10.1186/s12889-018-6264-1
18. Vargas-Herrera J, Meneses G, Cortez-Escalante J. Physicians' perceptions as predictors of the future use of the national death information system in Peru: cross-sectional study. *J Med Internet Res*. 2022;24(8):e34858. doi: 10.2196/34858
19. Martins RC, Buchalla CM. Automatic coding and selection of causes of death: an adaptation of Iris software for using in Brazil. *Rev Bras Epidemiol*. 2015;18(4):883-93. doi: 10.1590/1980-5497201500040016
20. Harteloh P. The implementation of an automated coding system for cause-of-death statistics. *Inform Health Soc Care*. 2020;45(1):1-14. doi: 10.1080/17538157.2018.1496092
21. Floristán YF, Osinaga JD, Prieto JC, Perez JA, Moreno-Iribas C. Coding Causes of Death with IRIS Software. Impact in Navarre Mortality Statistic. *Rev Esp Salud Publica*. 2016;90:e1-e9.
22. Orsi C, Navarra S, Frova L, Grande E, Marchetti S, Pappagallo M, et al. Impact of the implementation of ICD-10 2016 version and Iris software on mortality statistics in Italy. *Epidemiol Prev*. 2019;43(2-3):161-70. doi: 10.19191/EP19.2-3.P161.055
23. Suárez LC. Primer bienio de estadísticas de mortalidad con el codificador automático Iris de causas de muerte. *Gac Sanit*. 2018;32(1):5-7. doi: 10.1016/j.gaceta.2016.11.009
24. Ishitani LH, Cunha CCD, Ladeira RM, Corrêa PRL, Santos MRD, Rego MAS, et al. Evaluation of a smartphone application to improve medical certification of the cause of death. *Rev Bras Epidemiol*. 2019;22(Suppl 3):e190014.supl.3. doi: 10.1590/1980-549720190014.supl.3