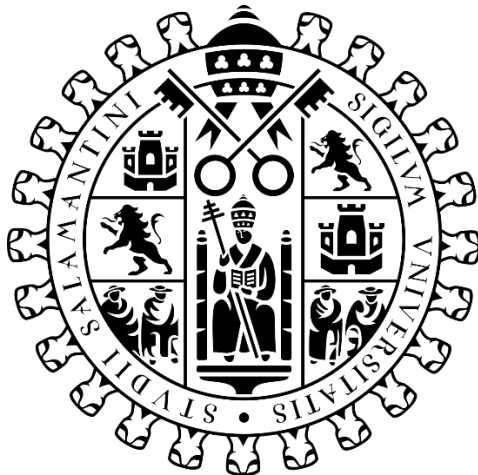


**UNIVERSIDAD DE SALAMANCA**  
DEPARTAMENTO DE ESTADÍSTICA

---

**MÉTODOS MULTIVARIANTES PARA EVALUAR  
TUMORES NEUROENDOCRINOS**



**TESIS DOCTORAL**

**KARIME MONTES ESCOBAR**

**2022**



# MÉTODOS MULTIVARIANTES PARA EVALUAR TUMORES NEUROENDOCRINOS

---

Memoria para optar al Grado de  
Doctor, por el Departamento de  
Estadística de la Universidad de  
Salamanca, presenta:

***Karime Montes Escobar***

Salamanca

2022



## *Agradecimientos*

---

Me gustaría agradecer a mi director de tesis, José Luis Vicente Villardón, Profesor del Departamento de Estadística de la Universidad de Salamanca, por haberme dado la oportunidad de realizar la tesis bajo su supervisión. Agradezco la confianza depositada en mí, y con la esperanza de haber cumplido todas las expectativas. Su dedicación, tiempo y experiencia fueron primordiales para llegar hasta aquí.

Mi especial gratitud a la Profesora Doctora M<sup>a</sup> Purificación Galindo Villardón por su ayuda, siempre estuvo allí cuando más lo necesitaba, también a todos los profesores del Departamento de Estadística de la Universidad de Salamanca, por sus conocimientos y confianza en mi trabajo y siempre dispuestos a ayudar.

Agradezco también a mis compañeros del doctorado, junto a ellos se pudo consolidar lo que es el sentido de la amistad, me queda muy claro que el valor de esta es incalculable. También quiero agradecer a mi compañero Javier de la Hoz Maestre, por el gran apoyo brindado durante todo este tiempo.

A mi familia y compañero de vida por darme la motivación que necesitaba para culminar con este reto propuesto.



# ÍNDICE

<b>1. Solicitud presentación de tesis</b>	<b>10</b>
1.1 Autorización del director	14
1.2 Compendio de trabajos Publicados	18
1.3 Aceptación por escrito y con firma original de los coautores para que el doctorando presente el trabajo	23
<b>2. Introducción</b>	<b>29</b>
<b>3. Estado del arte</b>	<b>35</b>
<b>3.1. Tumores Neuroendocrinos</b>	<b>37</b>
3.1.1. Clasificación de tumores neuroendocrinos	39
3.1.2. Clasificación de neoplasias endócrinas	40
<b>3.2. Técnicas estadística</b>	<b>43</b>
<b>3.3. HJ-Biplot</b>	<b>43</b>
3.3.1 Propiedades de los marcadores en el HJ Biplot	44
<b>3.4. Biplot Logístico</b>	<b>49</b>
3.4.1. Algoritmo general	50
3.4.2. Interpretación Biplot Logístico	52
3.4.3. Calidad de representación del Biplot Logístico	52
3.4.4. Regiones de predicción	52
3.4.5. Contribuciones Al Biplot	53
<b>3.5. Método LDA</b>	<b>55</b>
3.5.1. Frecuencia de Término - Frecuencia Inversa de Documento (TF-IDF)	56
3.5.2. Análisis Semántico Latente (LSA)	56
3.5.3. Análisis Semántico Latente Probabilístico (PLSA)	58
3.5.4. Latent Dirichlet Allocation (LDA)	60
3.5.5. LDA en el entorno R	65
<b>4. Referencias</b>	<b>68</b>
<b>5. Coherencia y relación directa entre los artículos presentados</b>	<b>81</b>
<b>6. Publicaciones originales que conforman la Tesis Doctoral</b>	<b>86</b>
6.1 <i>Frequency of Neuroendocrine Tumor Studies: Using Latent Dirichlet Allocation and HJ-Biplot</i>	88
6.2 <i>Neuroendocrine Tumors: An Analysis of Prevalence, Incidence, and Survival a Hospitals-Based Study in Ecuador</i>	106
6.3 <i>Clinical Related Factors to Neuroendocrine Tumors in Ecuadorian Patients: A Logistic Biplot Approach</i>	122
<b>7. Resúmenes de los Artículos</b>	<b>138</b>

<b>7.1</b>	<b><i>Frequency of Neuroendocrine Tumor Studies: Using Latent Dirichlet Allocation and Hj-Biplot Statistical Methods</i></b>	<b>140</b>
<b>7.2</b>	<b><i>Neuroendocrine Tumors: An Analysis of Prevalence, Incidence, and Survival a Hospitals-Based Study in Ecuador</i></b>	<b>149</b>
<b>7.3</b>	<b><i>Clinical Related Factors to Neuroendocrine Tumors in Ecuadorian Patients: A Logistic Biplot Approach</i></b>	<b>157</b>





## ***1. SOLICITUD PRESENTACIÓN DE TESIS***



***Estimada Coordinadora del Programa de Doctorado:***

*Karime Montes Escobar con pasaporte número: AR297158, y alumno del programa "DOCTORADO EN ESTADÍSTICA MULTIVARIANTE APLICADA" matriculado en el plan de estudios: ESTADÍSTICA MULTIVARIANTE APLICADA (R.D.99/2011) y con número de expediente 79:*

Solicita que se tenga en consideración la información aportada en este documento con el objetivo de poder presentar la tesis con título "*MÉTODOS MULTIVARIANTES PARA EVALUAR TUMORES NEUROENDOCRINOS*" mediante el formato de compendio de artículos/publicaciones. La información aportada se corresponde con lo establecido en el procedimiento para la presentación de la tesis doctoral en la *Universidad de Salamanca en el Formato de Compendio de Artículos/Publicaciones:*

A continuación, se detallan los documentos adjuntos en esta solicitud:

- Página Inicial especificando que la tesis corresponde a un compendio de trabajos previamente publicados detallando para cada uno de ellos: referencia de la revista, editorial, DOI y afiliaciones de cada uno de los miembros autores.
- Autorización del director o codirector para la presentación de la tesis mediante el formato de compendio de artículos/publicaciones.
- Tema objeto de estudio (Introducción).
- Estado del Arte.
- Apartado en castellano que refleje la coherencia y relación directa entre los artículos/publicaciones presentadas. Dicho apartado incluirá al menos una introducción con los antecedentes del tema objeto de estudio, la hipótesis de trabajo y los objetivos, así como las principales conclusiones.
- Copia completa de las publicaciones originales que conformarán la Tesis Doctoral (artículos, capítulos de libro, libro o libros aceptados o publicados).

Para cada uno de los 3 artículos presentados, un resumen en castellano en el cual se especifican: los objetivos de la investigación, la metodología utilizada, los resultados alcanzados, y las conclusiones.



## *1.1 Autorización del director*





**VNiVERSIDAD  
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

**DEPARTAMENTO DE ESTADÍSTICA**

---

***DR. JOSE LUIS VICENTE VILLARDÓN***

*Profesor Titular del Departamento de Estadística de la Universidad de  
Salamanca*

CERTIFICA: Que **D. Karime Montes Escobar** ha realizado bajo su dirección el trabajo para optar el Grado de Doctor, presenta mediante compendio de publicaciones con el título: *Métodos multivariantes para evaluar tumores neuroendocrinos*, y para que conste, firman el presente certificado en Salamanca, en septiembre de 2022.

Fdo.: Jose Luis Vicente Villardón





## *1.2 Compendio de trabajos Publicados*



La tesis con título “*MÉTODOS MULTIVARIANTES PARA EVALUAR TUMORES NEUROENDOCRINOS*” corresponde a un compendio de trabajos previamente publicados en revistas científicas indexadas en el *Journal Citation Reports*. A continuación, se detalla el nombre y afiliación de cada uno de los autores, la referencia completa de la revista o editorial y el DOI de cada uno de ellos.

---

**AUTORES (P.O. DE FIRMA):** KARIME MONTES ESCOBAR <sup>1,2</sup>, JOSÉ LUIS VICENTE VILLARDON <sup>2</sup>, JAVIER DE LA HOZ-M <sup>3</sup>, LELLY MARÍA USECHE CASTRO <sup>1</sup>, DANIEL FABRICIO ALARCÓN CANO <sup>4</sup> AND ALINE SITENESKI <sup>5</sup>. \*

**TÍTULO:** FREQUENCY OF NEUROENDOCRINE TUMOR STUDIES: USING LATENT DIRICHLET ALLOCATION AND HJ-BIPLLOT STATISTICAL METHODS

**REFERENCIA:** DOI: [10.3390/math9182281](https://doi.org/10.3390/math9182281). VOL. 9. ISSUE 18, PP. 2281. ISSN: 2227-7390

**REVISTA:** MATHEMATICS

**EDITORIAL:** MDPI

**AFILIACIONES:**

1. DEPARTMENT OF MATHEMATICS AND STATISTICS, INSTITUTE OF BASIC SCIENCES, TECHNICAL UNIVERSITY OF MANABÍ, PORTOVIEJO 130105, ECUADOR; KARIME.MONTES@GMAIL.COM (K.M.E.); LELLY.USECHE@UTM.EDU.EC (L.M.U.C.)
2. STATISTICS DEPARTMENT, UNIVERSITY OF SALAMANCA, 37007 SALAMANCA, SPAIN; VILLARDON@USAL.ES
3. FACULTAD DE INGENIERÍA, UNIVERSIDAD DEL MAGDALENA, SANTA MARTA 470004, COLOMBIA; JDELAHOZ@UNIMAGDALENA.EDU.CO
4. TEACHING AND RESEARCH, SOLCA, MANABÍ, PORTOVIEJO 130105, ECUADOR; ALARCON.DANIEL.MD@GMAIL.COM
5. RESEARCH INSTITUTE AND FACULTY OF HEALTH SCIENCES, TECHNICAL UNIVERSITY OF MANABÍ, PORTOVIEJO 130105, ECUADOR

**LUGAR DE PUBLICACION:** BASEL, SWITZERLAND

**FECHA DE PUBLICACION:** 2021

**FACTOR DE IMPACTO:** PUBLICADO (JCR-2021) 2.592

REVISION POR ALMENOS DOS EXPERTOS INDEPENDIENTES



---

**AUTORES (P.O. DE FIRMA):** KARIME MONTES ESCOBAR<sup>1,2</sup>, JOSE LUIS VICENTE VILLARDON<sup>2</sup>, ROSA ELISA VILLACÍS GONZALEZ<sup>3</sup>, PAUL HUMBERTO CASTILLO COR-DOVA<sup>4</sup>, JOHANNA MABEL SÁNCHEZ RODRÍGUEZ<sup>5,6</sup>, MELINA DE LA CRUZ-VELEZ<sup>7</sup> AND ALINE SITENESKI <sup>8</sup>. \*

**TÍTULO:** NEUROENDOCRINE TUMORS: AN ANALYSIS OF PREVALENCE, INCIDENCE, AND SURVIVAL A HOSPITALS-BASED STUDY IN ECUADOR

**REFERENCIA:** DOI: ISSN: 2227-9032

**REVISTA:** HEALTHCARE

**EDITORIAL:** MDPI

**AFILIACIONES:**

1. DEPARTMENT OF MATHEMATICS AND STATISTICS, INSTITUTE OF BASIC SCIENCES, UNIVERSIDAD TÉCNICA DE MANABÍ, PORTOVIEJO 130105, ECUADOR; KARIME.MONTES@GMAIL.COM
2. STATISTICS DEPARTMENT, UNIVERSITY OF SALAMANCA, 37007 SALAMANCA, SPAIN; VILLARDON@USAL.ES
3. TEACHING AND RESEARCH, SOLCA, EL ORO, MACHALA 070206, ECUADOR; DRAROSAVILLACIS@GMAIL.COM
4. TEACHING AND RESEARCH, SOLCA, LOJA, LOJA 110150, ECUADOR; PCASTILLO@SOLCALOJA.MED.EC
5. FACULTAD DE MEDICINA, UNIVERSIDAD LAICA ELOI ALFARO DE MANABÍ, MANTA 130203, ECUADOR; JOHANNA.SANCHEZ@ULEAM.EDU.EC
6. FACULTAD DE CIENCIAS DE LA SALUD, UNIVERSIDAD ESTATAL DEL SUR DE MANABI, JIPIJAPA 130650; JOHANNA.SANCHEZ@UNESUM.EDU.EC
7. FACULTY OF HEALTH SCIENCES, MEDICINE CAREER, UNIVERSIDAD TÉCNICA DE MANABÍ, PORTOVIEJO 130105, ECUADOR; MELINADELACRUZ100@GMAIL.COM
8. RESEARCH INSTITUTE, UNIVERSIDAD TÉCNICA DE MANABÍ, PORTOVIEJO 130105, ECUADOR. [ALINE.SITENESKI@GMAIL.COM](mailto:ALINE.SITENESKI@GMAIL.COM)

**LUGAR DE PUBLICACION:** BASEL, SWITZERLAND

**FECHA DE PUBLICACION:** 2022

**FACTOR DE IMPACTO:** ACEPTADO (JCR-2022) 3.160

REVISION POR ALMENOS DOS EXPERTOS INDEPENDIENTES





**AUTORES (P.O. DE FIRMA):** KARIME MONTES ESCOBAR <sup>a,b</sup>, JOSE LUIS VICENTE VILLARDON <sup>b</sup>, DANIEL FABRICIO ALARCON CANO <sup>c</sup>, ALINE SITENESKI <sup>d,e</sup>

**TITULO:** CLINICAL RELATED FACTORS TO NEUROENDOCRINE TUMORS IN ECUADORIAN PATIENTS: A LOGISTIC BIPLLOT APPROACH

**REFERENCIA: DOI:** 10.54817/IC.v63n1a02. VOL. 63 (1), PP. 19 – 31. ISSN: 0535-5133

**REVISTA:** INVESTIGACION CLINICA

**AFILIACIONES:**

- a. DEPARTMENT OF MATHEMATICS AND STATISTICS. INSTITUTE OF BASIC SCIENCES. TECHNICAL UNIVERSITY OF MANABÍ, PORTOVIEJO, 130105, ECUADOR.
- b. STATISTICS DEPARTMENT. UNIVERSITY OF SALAMANCA, SALAMANCA, 37007, SPAIN.
- c. TEACHING AND RESEARCH, SOLCA MANABÍ, ECUADOR
- d. RESEARCH INSTITUTE UNIVERSITY OF MANABÍ, PORTOVIEJO, 130105, ECUADOR.
- e. HEALTH SCIENCES, MEDICINE CAREER, TECHNICAL UNIVERSITY OF MANABÍ, PORTOVIEJO, 130105, ECUADOR.

**LUGAR DE PUBLICACION:** VENEZUELA

**FECHA DE PUBLICACION:** 2022

**FACTOR DE IMPACTO: PUBLICADO** (JCR-2021) 0.683 (4º CUARTIL)

REVISION POR ALMENOS DOS EXPERTOS INDEPENDIENTES

---



*1.3 Aceptación por escrito y con firma original de los coautores para que el doctorando presente el trabajo*





**AUTORES (P.O. DE FIRMA):** KARIME MONTES ESCOBAR <sup>1,2</sup>, JOSÉ LUIS VICENTE VILLARDON <sup>2</sup>, JAVIER DE LA HOZ-M <sup>3</sup>, LELLY MARÍA USECHE CASTRO <sup>1</sup>, DANIEL FABRICIO ALARCÓN CANO <sup>4</sup> AND ALINE SITENESKI <sup>5</sup>. \*

**TÍTULO:** FREQUENCY OF NEUROENDOCRINE TUMOR STUDIES: USING LATENT DIRICHLET ALLOCATION AND HJ-BIPLLOT STATISTICAL METHODS

**REFERENCIA:** DOI: [10.3390/math9182281](https://doi.org/10.3390/math9182281). VOL. 9. ISSUE 18, PP. 2281. ISSN: 2227-7390

**REVISTA:** MATHEMATICS

**EDITORIAL:** MDPI

**AFILIACIONES:**



1. DEPARTMENT OF MATHEMATICS AND STATISTICS, INSTITUTE OF BASIC SCIENCES, TECHNICAL UNIVERSITY OF MANABÍ, PORTOVIEJO 130105, ECUADOR; KARIME.MONTES@GMAIL.COM (K.M.E.); LELLY.USECHE@UTM.EDU.EC (L.M.U.C.)
2. STATISTICS DEPARTMENT, UNIVERSITY OF SALAMANCA, 37007 SALAMANCA, SPAIN; VILLARDON@USAL.ES
3. FACULTAD DE INGENIERÍA, UNIVERSIDAD DEL MAGDALENA, SANTA MARTA 470004, COLOMBIA; JDELAHOZ@UNIMAGDALENA.EDU.CO
4. TEACHING AND RESEARCH, SOLCA, MANABÍ, PORTOVIEJO 130105, ECUADOR; ALARCON.DANIEL.MD@GMAIL.COM
5. RESEARCH INSTITUTE AND FACULTY OF HEALTH SCIENCES, TECHNICAL UNIVERSITY OF MANABÍ, PORTOVIEJO 130105, ECUADOR

**LUGAR DE PUBLICACION:** BASEL, SWITZERLAND

**FECHA DE PUBLICACION:** 2021

**FACTOR DE IMPACTO: PUBLICADO** (JCR-2021) 2.592

REVISION POR ALMENOS DOS EXPERTOS INDEPENDIENTES

---

Los coautores del trabajo titulado “*FREQUENCY OF NEUROENDOCRINE TUMOR STUDIES: USING LATENT DIRICHLET ALLOCATION AND HJ-BIPLLOT STATISTICAL METHODS*”, con DOI: [10.3390/math9182281](https://doi.org/10.3390/math9182281), hacen constar que Doña Karime Montes Escobar con pasaporte número: AR297159 es el autor principal de la investigación recogida en este artículo y que conocen y dan su conocimiento para que el artículo sea utilizado para que D. Karime Montes Escobar presente su tesis doctoral denominada *MÉTODOS MULTIVARIANTES PARA EVALUAR TUMORES NEUROENDOCRINOS* en el formato de compendio de artículos:

D. José Luis Vicente Villardon

D. Javier De La Hoz Maestre

D. Lelly María Useche Castro

D. Daniel Fabricio Alarcón Cano

D. Aline Siteneski

**AUTORES (P.O. DE FIRMA):** KARIME MONTES ESCOBAR<sup>1,2</sup>, JOSE LUIS VICENTE VILLARDON<sup>2</sup>, ROSA ELISA VILLACÍS GONZALEZ<sup>3</sup>, PAUL HUMBERTO CASTILLO CORDOVA<sup>4</sup>, JOHANNA MABEL SÁNCHEZ RODRÍGUEZ<sup>5,6</sup>, MELINA DE LA CRUZ-VELEZ<sup>7</sup> AND ALINE SITENESKI<sup>8,\*</sup>

**TÍTULO:** NEUROENDOCRINE TUMORS: AN ANALYSIS OF PREVALENCE, INCIDENCE, AND SURVIVAL A HOSPITALS-BASED STUDY IN ECUADOR

**REFERENCIA:** DOI: ISSN: 2227-9032

**REVISTA:** HEALTHCARE

**EDITORIAL:** MDPI

**AFILIACIONES:**



1. DEPARTMENT OF MATHEMATICS AND STATISTICS, INSTITUTE OF BASIC SCIENCES, UNIVERSIDAD TÉCNICA DE MANABÍ, PORTOVIEJO 130105, ECUADOR; KARIME.MONTES@GMAIL.COM
2. STATISTICS DEPARTMENT, UNIVERSITY OF SALAMANCA, 37007 SALAMANCA, SPAIN; VILLARDON@USAL.ES
3. TEACHING AND RESEARCH, SOLCA, EL ORO, MACHALA 070206, ECUADOR; DRAROSAVILLACIS@GMAIL.COM
4. TEACHING AND RESEARCH, SOLCA, LOJA, LOJA 110150, ECUADOR; PCASTILLO@SOLCALOJA.MED.EC
5. FACULTAD DE MEDICINA, UNIVERSIDAD LAICA ELOI ALFARO DE MANABÍ, MANTA 130203, ECUADOR; JOHAN-NA.SANCHEZ@ULEAM.EDU.EC
6. FACULTAD DE CIENCIAS DE LA SALUD, UNIVERSIDAD ESTATAL DEL SUR DE MANABI, JIPIJAPA 130650; JOHAN-NA.SANCHEZ@UNESUM.EDU.EC
7. FACULTY OF HEALTH SCIENCES, MEDICINE CAREER, UNIVERSIDAD TÉCNICA DE MANABÍ, PORTOVIEJO 130105, ECUADOR; MELINADELACRUZ100@GMAIL.COM
8. RESEARCH INSTITUTE, UNIVERSIDAD TÉCNICA DE MANABÍ, PORTOVIEJO 130105, ECUADOR. [ALINE.SITENESKI@GMAIL.COM](mailto:ALINE.SITENESKI@GMAIL.COM)

**LUGAR DE PUBLICACION:** BASEL, SWITZERLAND

**FECHA DE PUBLICACION:** 2022

**FACTOR DE IMPACTO: ACEPTADO** (JCR-2022) 3.160

REVISION POR ALMENOS DOS EXPERTOS INDEPENDIENTES

---

Los coautores del trabajo titulado “*NEUROENDOCRINE TUMORS: AN ANALYSIS OF PREVALENCE, INCIDENCE, AND SURVIVAL A HOSPITALS-BASED STUDY IN ECUADOR*”, hacen constar que Doña Karime Montes Escobar con pasaporte número: AR297159 es el autor principal de la investigación recogida en este artículo y que conocen y dan su conocimiento para que el artículo sea utilizado para que D. Karime Montes Escobar presente su tesis doctoral denominada *MÉTODOS MULTIVARIANTES PARA EVALUAR TUMORES NEUROENDOCRINOS* en el formato de compendio de artículos:

Jose Luis Vicente Villardon

Rosa Elisa Villacís González

Paul Humberto Castillo Cordova

Johanna Mabel Sánchez Rodríguez

Melina De La Cruz Vélez

Aline Siteneski

**AUTORES (P.O. DE FIRMA):** KARIME MONTES ESCOBAR <sup>a,b</sup>, JOSE LUIS VICENTE VILLARDON <sup>b</sup>, DANIEL FABRICIO ALARCON CANO <sup>c</sup>, ALINE SITENESKI <sup>d,e</sup>

**TITULO:** CLINICAL RELATED FACTORS TO NEUROENDOCRINE TUMORS IN ECUADORIAN PATIENTS: A LOGISTIC BIPLLOT APPROACH

**REFERENCIA: DOI:** 10.54817/IC.v63n1a02. VOL. 63 (1), PP. 19 – 31. ISSN: 0535-5133

**REVISTA:** INVESTIGACION CLINICA

**AFILIACIONES:**

- a. DEPARTMENT OF MATHEMATICS AND STATISTICS. INSTITUTE OF BASIC SCIENCES. TECHNICAL UNIVERSITY OF MANABÍ, PORTOVIEJO, 130105, ECUADOR.
- b. STATISTICS DEPARTMENT. UNIVERSITY OF SALAMANCA, SALAMANCA, 37007, SPAIN.
- c. TEACHING AND RESEARCH, SOLCA MANABÍ, ECUADOR
- d. RESEARCH INSTITUTE UNIVERSITY OF MANABÍ, PORTOVIEJO, 130105, ECUADOR.
- e. HEALTH SCIENCES, MEDICINE CAREER, TECHNICAL UNIVERSITY OF MANABÍ, PORTOVIEJO, 130105, ECUADOR.

**LUGAR DE PUBLICACION:** VENEZUELA

**FECHA DE PUBLICACION:** 2022

**FACTOR DE IMPACTO: PUBLICADO** (JCR-2021) 0.683 (4º CUARTIL)

REVISION POR ALMENOS DOS EXPERTOS INDEPENDIENTES



---

Los coautores del trabajo titulado “*CLINICAL RELATED FACTORS TO NEUROENDOCRINE TUMORS IN ECUADORIAN PATIENTS: A LOGISTIC BIPLLOT APPROACH*”, hacen constar que Doña Karime Montes Escobar con pasaporte número: AR297159 es el autor principal de la investigación recogida en este artículo y que conocen y dan su conocimiento para que el artículo sea utilizado para que D. Karime Montes Escobar presente su tesis doctoral denominada *MÉTODOS MULTIVARIANTES PARA EVALUAR TUMORES NEUROENDOCRINOS* en el formato de compendio de artículos:

D. Jose Luis Vicente Villardon

D. Daniel Fabricio Alarcon Cano

D. Aline Siteneski



## ***2. INTRODUCCIÓN***



Los tumores neuroendocrinos (TNE) fueron reconocidos por primera vez como un tipo específico de crecimiento diferente a mediados de 1800, el primero en describir un tumor carcinoide fue Langhans en el año de 1867, sin embargo, quien lo registro por primera vez como un nuevo tumor fue Lubarsch en 1888, así dos años después Ransom fue quien explico el cuadro clásico de síntomas del síndrome carcinoide (Modlin et al., 2003).

Los TNE constituyen un grupo de tumores malignos que surgen de las células neuroendocrinas de todo el cuerpo que se definen como neoplasias epiteliales con una diferenciación predominante neuroendocrina (Klimstra, 2010). Comúnmente se localizan en el páncreas, en el tracto digestivo y en el pulmón, aunque pueden aparecer en cualquier órgano del cuerpo (Hernando & Capdevilla, 2020). Su incidencia anual es menor a 10 casos por cada 100.000 habitantes (de Miguel Novoa et al., 2014), aunque gracias al desarrollo de las técnicas diagnósticas se ha aumentado la incidencia en los últimos 30 años (Jiao et al., 2015).

Los tumores neuroendocrinos son poco frecuentes, aunque han sido motivo de estudio para muchos investigadores, ya que la dificultad para una clasificación, diagnóstico y tratamiento para el paciente se debe a:

- Su heterogeneidad, tanto en la localización como en la histología, lo que dificulta disponer de una clasificación uniforme.
- La actividad hormonal que a veces pueden presentar algunas neoplasias endocrinas, ocasionando síndromes clínicos por hipersecreción hormonal.
- La baja correlación existente entre la morfología y el comportamiento biológico, que dificulta predecir su pronóstico.
- Su baja frecuencia, que hace que el conocimiento y experiencia sobre ellos sean limitados

La obtención de datos sobre incidencia para TNE ha sido siempre difícil debido a la baja frecuencia de estas neoplasias, su largo curso clínico, diagnóstico poco específico, escasos registros de tumores y como ya se ha explicado anteriormente, la falta de una clasificación clara y objetiva para dichos tumores. En el registro norteamericano del programa SEER (Surveillance, Epidemiology and End Results) se recogen datos desde 1973 y los datos existentes para la población caucásica muestran una incidencia de 2,5-5 nuevos casos por 100.000 habitantes al año (Carcinoides GI y Broncopulmonares), representando el 0,5% del total de tumores (Taal & Visser, 2004).

Aunque se consideran tumores de baja frecuencia su incidencia ha aumentado notablemente en los últimos 30 años. Además, la prevalencia de estos tumores es

relativamente alta debido a la elevada supervivencia que puede llegar a ser del 35% al 60% a 5 años para pacientes con enfermedad avanzada. Como ya se expuso anteriormente los TNE más frecuentes son los “carcinoides clásicos” que pueden aparecer en diversas localizaciones corporales, siendo los pulmonares lo más frecuentes (31% del total de TNE), aunque la mayoría de las lesiones están localizadas en el tracto Gastro Intestinal con un (66%), siendo los sitios predominantes por orden decreciente, intestino delgado, recto, apéndice, colon y estómago.

Con base en lo anterior, se ha demostrado que en la investigación para los tumores neuroendocrinos se usan aplicaciones repetitivas de métodos estadísticos, sin que se explique por qué se han elegido y sin proporcionar detalles de su relación con la naturaleza y propiedades de los datos valorados para cada estudio, empleando rutinariamente las mismas técnicas estadísticas independientemente del tipo de datos.

De acuerdo con lo anterior, se proponen algunas técnicas multivariantes, que permitieron analizar a los Tumores Neuroendocrinos, como primera alternativa para este estudio o en cualquier estudio que se obtienen datos binarios, (Vicente-Villardón et al., 2006; Vicente-Villardón et al., 2001) propone un nuevo enfoque Biplot, llamado Biplot Logístico. El modelo se basa en las coordenadas de los individuos y las variables están previstas para obtener respuestas de tipo logístico (Vicente-Villardón et al., 2004). Formulando la metodología, se define la matriz de datos  $\mathbf{X}_{I \times J}$ , en donde las filas corresponden a los individuos y las columnas se asocian a las variables binarias (ausencia o presencia).

También se presenta una técnica que utiliza una elección adecuada de marcadores para las filas y columnas, en los años 1985 – 1986 Galindo propuso una variante de las técnicas estándar de biplot, el HJ-Biplot, permite la representación simultánea en el mismo sistema de coordenadas. Además, demostró que esta alternativa asegura una alta calidad de representación tanto para las filas como para las columnas. De hecho, esta es la principal ventaja de este método de ordenación que proporciona mejores resultados que otras técnicas multivariantes convencionales (Galindo, 1986; Galindo & Cuadras, 1986).

Finalmente, como una tercera opción de análisis estadístico para este estudio, se consideró el modelado de tópicos Asignación Latente de Dirichlet (LDA) que tienen como referente a un grupo de algoritmos y técnicas estadísticas de aprendizaje, extracción y reconocimiento de la información, donde su objetivo es analizar estructuras ocultas de un conjunto de documentos para determinar los tópicos, como estos tienen una conexión entre sí y, además, los cambios que tienen durante el tiempo. Su ventaja se encuentra en no requerir de ningún etiquetado o anotación previa de los documentos, estos tópicos emergen del análisis de textos originales (Blei, 2012).



La idea en los modelados de tópicos es hallar un marco adecuado de modelado para datos discretos, donde su objetivo es disminuir la dimensionalidad de los datos al igual que numerosas técnicas multivariantes. Algunos investigadores propusieron como resultado disminuir cada documento a un vector de números reales, representando por proporciones de recuentos. Los modelos de tópicos se basan en suposiciones básicas, en donde, cada tópico se refiere a una colección de palabras y cada documento hace como referencia una mezcla de tópicos donde se puede tener en cuenta la clasificación de texto, la recuperación de información y resumen de texto, entre otras (Kao & Poteet, 2007).

El modelado de tópicos tiene como objetivo principal la organización y el resumen de archivos electrónicos en diferentes formatos (redes sociales, artículos científicos, páginas web, libros, imágenes, sonido y videos), logrando una escala que a nivel de notación humana sería imposible de hallar (Blei, 2012).

Actualmente, el logaritmo Asignación Latente de Dirichlet (LDA) propuesto por (Blei et al., 2003) se destaca dentro de los algoritmos más populares para el modelado de tópicos (Blei, 2012; DiMaggio et al., 2013; Grimmer, 2010; Jacobi et al., 2016), los algoritmos Análisis Semántico Latente (LSA) (Deerwester et al., 1990) y el Análisis Semántico Latente Probabilístico (PLSA) (Hofmann, 1999).

Esta tesis, presentada por el formato de compendio de publicaciones, consta de un primer capítulo "*Solicitud Presentación de Tesis*", que engloba toda la documentación necesaria para poder presentar la tesis en este formato. A continuación, el capítulo "Introducción" presenta una introducción acerca del tema de estudio y la problemática a resolver.

En el capítulo "Estado de arte", se presenta una revisión teórica de las principales técnicas multivariantes involucradas en este trabajo de tesis doctoral, mostrando especial atención a su aplicación en Tumores Neuroendocrinos. En el capítulo "Coherencia y relación directa entre los artículos" se enseña un resumen con los objetivos generales y la coherencia entre las tres publicaciones expuestas. En el capítulo "Publicaciones Originales" se muestran las técnicas multivariantes propuesta en el marco de este trabajo de tesis doctoral. Las técnicas estadísticas multivariantes y los resultados obtenidos se describen a través del compendio de tres trabajos de investigación publicados en revistas indexadas en el índice ISI JCR. En el capítulo "Resúmenes de los Artículos" se presenta un resumen para cada uno de los artículos, poniendo especial atención a los objetivos, a la propuesta realizada en cada uno de los artículos y a los resultados obtenidos



### ***3. ESTADO DEL ARTE***



### 3.1. Tumores Neuroendocrinos

Las células neuroendocrinas tienen propiedades endocrinas y neuronales y están ampliamente distribuidas en el cuerpo (Rindi & Wiedenmann, 2020). Muchos órganos contienen este fenotipo celular y pueden verse afectados por los Tumores Neuroendocrinos (TNE). El tumor tiene una característica morfofuncional particular, produce hormonas y presenta un patrón de tinción inmunohistoquímica con cromogranina, sinaptofisina y enolasa específica de neuronas (Chen, 2017). Además, la Organización Mundial de la Salud (OMS) recomienda caracterizar los TNE por la fracción de proliferación medida por el recuento mitótico o (más fácilmente) por el porcentaje positivo de Ki67 (índice Ki67) (Rindi et al., 2019). Cabe mencionar que los estudios epidemiológicos indican que la fuente más frecuente de metástasis en los TNE es el intestino delgado, pulmones y colon (Riihimäki et al., 2016).

Existe un retraso significativo entre la aparición de los síntomas y el diagnóstico de los TNE y la mayoría de los pacientes son diagnosticados en una etapa tardía de la enfermedad (J. C. Yao et al., 2008). Trabajos anteriores consideran a los TNE como un cáncer de bajo grado, generalmente no funcional, sin una supervivencia prolongada de los pacientes con pronóstico según el grado y el estadio (Vanoli et al., 2019). De hecho, la OMS en 2017 clasifica los TNE con un Ki67 > 20% como TNE de grado 3 (bien diferenciados) o ECN de grado 3 (poco diferenciados) sobre la base de características morfológicas como necrosis y diferenciación (Lloyd et al., 2017). Los TNE representan aproximadamente el 0,5% de todas las neoplasias malignas conocidas actualmente (Taal & Visser, 2004b). De hecho, el origen y el grado de clasificación de los TNE pueden definir la supervivencia del paciente, si se informaron previamente a los 103 meses de la enfermedad que se origina en el intestino delgado y solo 14 meses para los cánceres que se originan en el colon (Dasari et al., 2017b). Un estudio de gravedad de varios centros a nivel nacional en Corea informó un alto porcentaje de G1 del 92,3% (Cho et al., 2012). Por otro lado, la investigación en Kentucky encontró una población con 67% en G3 (Chauhan et al., 2018). Reforzando esta noción, nuestro estudio reporta la mayoría de los pacientes en G3 - 68%. Sin embargo, los TNE de menor grado no indican un mejor resultado en términos de supervivencia (Chauhan et al., 2018). Está bien establecido que el diagnóstico temprano de cualquier tipo de cáncer puede mejorar la calidad de vida y la supervivencia, y los pacientes con síndrome carcinoide tienen una calidad de vida significativamente peor que los pacientes con solo TNE (Beaumont et al., 2012).

En respuesta a los avances farmacológicos, las nuevas técnicas de diagnóstico y el diagnóstico temprano de la enfermedad, la supervivencia de los pacientes con TNE ha mejorado con el tiempo (Dasari et al., 2017b). Se informaron índices más altos de supervivencia en pacientes con TNE gastrointestinales distantes y TNE pancreáticos distantes (Dasari et al., 2017b). El TNE pancreático es una neoplasia maligna poco frecuente con un comportamiento biológico

relativamente indolente en comparación con el adenocarcinoma pancreático. Se informó anteriormente que la media de supervivencia en pacientes con TNE pancreático avanzado fue de alrededor de 27 meses (J. C. Yao et al., 2008). Sin embargo, los TNE son responsables de aproximadamente el 10% de todos los casos por la prevalencia de tumor pancreático y solo el 1% de todos los casos por la incidencia (J. C. Yao et al., 2007, 2010; J. Yao & Phan, 2012). Aproximadamente el 64% de los pacientes con NET pancreáticos se presentan con enfermedad metastásica y están presentes en una etapa avanzada de la enfermedad y es probable que tengan mala (J. Yao & Phan, 2012). El NET pancreático suele afectar a personas de entre 40 y 69 años, pero a un número significativo de pacientes menores de 35 años (Moore et al., 2003; J. C. Yao et al., 2007). Se estima que entre el 40% y el 91% de los NET pancreáticos no son funcionales, es decir, sin síntomas hormonales clínicamente evidentes (Klimstra et al., 2010). Por tanto, el manejo terapéutico del NET pancreático depende del grado de diferenciación del tumor que puede diferenciarse bien o mal. De hecho, se considera que son la presencia de síntomas causados por la hipersecreción de hormonas y el estadio al momento del diagnóstico de la enfermedad (J. Yao & Phan, 2012). Según un análisis de la base de datos SEER (Surveillance, Epidemiology and End Results) entre 1973 y 2000, el NET pancreático tuvo una incidencia anual por millón fue de 1.8 en mujeres y 2.6 en hombres (Klimstra et al., 2010).

En el caso de los NET de pulmones, carcinomas de células grandes que representan aproximadamente el 3% de todos los cánceres de pulmón (Fasano et al., 2015). La cirugía no presenta eficacia ya que la mayoría de los pacientes mueren por recidiva (Fournel et al., 2013). El pronóstico de supervivencia aumenta cuando se respeta el límite de 3 cm de tumor para cirugía, por encima de este límite, la recurrencia es frecuente y la supervivencia desciende significativamente (Kujtan et al., 2018; Rossi et al., 2005; Sarkaria et al., 2011). Más allá de eso, la guía de la Red Nacional Integral del Cáncer (NCCN) para el carcinoma neuroendocrino de células grandes de células no pequeñas y en estadio específico en el cáncer de pulmón, destaca la necesidad de quimioterapia adyuvante o quimior radioterapia después de la resección en algunos casos que incluyen un tamaño de tumor > 4 cm. Morfológicamente idéntico al carcinoma de células pequeñas de pulmón, los TNE de mama con células poco diferenciadas (Rosen & Gattuso, 2017). Los tumores neuroendocrinos de mama representan menos del 1% de los cánceres de mama (Bussolati & Badve, 2012). La terapia endocrina y la radioterapia son capaces de aumentar la supervivencia de las pacientes con TNE de mama (Brask et al., 2014; B. Wei et al., 2010), pero aparentemente, se ha llegado al consenso sobre el pronóstico, la mayoría de los estudios sugieren un mal pronóstico (Rosen & Gattuso, 2017).

Por otro lado, en los TNE del intestino medio, la guía oficial afirma que el papel de la resección del tumor primario solo en el contexto de la enfermedad metastásica irreseccable no está bien definido y respalda la cirugía solo si se puede lograr una resección casi completa de los tumores (Howe et al., 2017). Sin embargo, el beneficio de supervivencia de la resección del tumor primario

en el contexto de enfermedad metastásica se demostró en estudios con pacientes con NET gastrointestinales (Lewis et al., 2019). En el caso de metástasis hepáticas neuroendocrinas, la resección quirúrgica presenta el mayor potencial de curación (DiNorcia et al., 2010; Mayo et al., 2010; Musunuru et al., 2006). Desafortunadamente, el papel curativo solo es factible en el 10-25% de los pacientes después de la resección en metástasis hepáticas (Martinez et al., 2007) y la recurrencia de los TNE es prevalente, siendo la principal causa de muerte en la mayoría de estos pacientes (Mayo et al., 2010). De hecho, los datos sobre la supervivencia general de los TNE son bien reportados en estudios de países desarrollados (Dasari et al., 2017b). Cabe mencionar que, los informes sobre la recurrencia pueden informar a los pacientes sobre la probabilidad de éxito del tratamiento y el riesgo de recurrencia después de la intervención quirúrgica (Merath et al., 2018).

A pesar del aumento en la prevalencia, incidencia y severidad en el momento del diagnóstico de los TNE reportado en los estudios, los datos reportados son principalmente de países desarrollados (Dasari et al., 2017; Rindi et al., 2007). El tema es particularmente importante y estudios como se propone el presente trabajo son importantes para llenar las lagunas de las evidencias científicas en los países en desarrollo.

### **3.1.1. Clasificación de tumores neuroendocrinos**

La reciente clasificación de la OMS incluye un esquema general que se aplica a todos los tumores endocrinos gastrointestinales, independientemente de su localización anatómica.

La nueva clasificación de la OMS incluye un esquema general aplicable a todas las neoplasias endócrinas gastroenteropancreáticas, independientemente de su localización anatómica.

Incluye de cinco categorías principales que diferencian a las neoplasias con fenotipo endócrino puro de las neoplasias mixtas exócrino-endócrinas (con fenotipo divergente) y las lesiones pseudotumorales (Rindi et al., 2019).

1. Tumor endócrino bien diferenciado (de comportamiento benigno/incierto).
2. Carcinoma endócrino bien diferenciado (bajo grado de malignidad).
3. Carcinoma endócrino pobremente diferenciado (small cell), (alto grado de malignidad).
4. Carcinoma mixto endócrino-exócrino (el más agresivo dicta el pronóstico).
5. Lesiones tumor - like.

Entre los tumores endocrinos con un fenotipo endocrino puro, la distinción entre tumores endocrinos bien y mal diferenciados se basa en el tamaño del tumor, índice mitótico y tasa de proliferación, invasión local y/o invasión linfática y presencia de metástasis.

En el segundo paso, se refinó la clasificación introduciendo la ubicación anatómica, el tipo celular de origen, la biología, la actividad hormonal y la asociación con síndromes clínicos, para abordar la clasificación. El tipo se asocia con el pronóstico (Zamora, 2009).

**Tabla 1.** Directrices para el tratamiento de los tumores neuroendocrinos gastroenteropancreáticos (incluidas las neoplasias broncopulmonares y tímicas) (Group: et al., 2004).

	Tamaño	Bordes	Profundidad de invasión	Invasión vascular	Atipía celular	Índice KI-67
<b>Tumor Endócrino bien diferenciado</b>	< 2 cm	Bien delimitados	Mucosa submucosa	ausente	células diferenciadas	< 2 %
<b>Carcinoma Endócrino bien diferenciado</b>	> 2 cm	Poco delimitados	crecimiento invasivo musc. propia	presente	células diferenciadas	2-15%
<b>Carcinoma Endócrino pob. diferenciado</b>	> 2 cm	Poco delimitados	crecimiento invasivo musc. propia	presente	severa atipía	> 15%

### 3.1.2. Clasificación de neoplasias endócrinas

Extraída de (Hamilton & Aaltonen, 2000):

#### 3.3.2.1. Esófago

- Neoplasia endócrina bien diferenciada
- Carcinoma endócrino pobremente diferenciado (carcinoma de células pequeñas)
- Carcinoma mixto exócrino-endócrino.

#### 3.3.2.2. Estómago

#### **Neoplasia endócrina bien diferenciada**

- ECL-C (*de células entero-cromafin like*):
  - Tipo I (asociado a gastritis crónica atrófica)
  - Tipo II (asociado a Zollinger-Ellison, MEN I)
  - Tipo III (esporádico)
- EC-C (de células enterocromafines, productoras de serotonina) G-C (de células productoras de gastrina)
- Otros



**Neoplasia endócrina pobremente diferenciada**

- Carcinoma de células pequeñas
- Carcinoma neuroendócrino de células grandes

**Carcinoma mixto exócrino-endócrino**

**Lesiones tumor-like**

- Hiperplasia
- Displasia

**3.3.2.3. Duodeno-Intestino Delgado**

**Neoplasia endócrina bien diferenciada**

- Gastrinoma funcionantes o no funcionantes
- Somatoestatinoma
- EC-C (de células enterocromafines, productor de serotonina)
- L-C (tumor productor de péptido glucagon-like y PP/PYY)
- Paraganglioma gangliocítico

**Carcinoma mixto exócrino-endócrino**

**Neoplasia endócrina pobremente diferenciada**

- Carcinoma de células pequeñas
- Otros

**3.3.2.4. Apéndice**

**Neoplasia endócrina bien diferenciada**

- EC-C (de células enterocromafines, productor de serotonina)
- L-C (tumor productor de péptido glucagon-like y PP/PYY)
- Otros

**Neoplasia endócrina pobremente diferenciada**

- Carcinoma de células pequeñas

**Carcinoma mixto exócrino-endócrino**

- *Globet cell tumors*
- Otros

### 3.3.2.5. Íleon-Ciego-Colon-Recto

#### Neoplasia endócrina bien diferenciada

- EC-C (productor de serotonina)
- L-C (tumor productor de péptido glucagon-like y PP/PYY)
- Otros

#### Neoplasia endócrina pobremente diferenciada

- Carcinoma de células pequeñas
- Carcinoma neuroendócrino de células grandes

#### Carcinoma mixto exócrino-endócrino

### 3.3.2.6. Páncreas

#### Tumor NE bien diferenciado

Funcionantes:

- Insulinoma            Gastrinoma
- Glucagonoma        Vipoma
- Somatoestatinoma    Otros

No funcionantes:

- Adenoma (< 0.5 cm) Otros

#### Carcinoma NE bien diferenciado

Funcionantes:

- Insulinoma            Gastrinoma
- Glucagonoma        Productor de serotonina con síndrome carcinoide
- Somatoestatinoma    Productor de ACTH con síndrome de *Cushing*

No funcionantes:

#### Carcinoma NE pobremente diferenciado

- (Small cell)

#### Carcinoma mixto exócrino-endócrino

#### Lesiones pseudotumorales

- Hiperplasia de los islotes

- Nesidioblastosis
- Hiperplasia endócrina

No funcionantes:

**Carcinoma TNE pobremente diferenciado**

**Carcinoma mixto exócrino-endócrino**

En conclusión, la heterogeneidad de los tumores neuroendocrinos, tanto en aspectos diagnósticos como terapéuticos, debe ser considerada individualizada y jerarquizada en el marco de un equipo multidisciplinar. La clasificación actual de la OMS brinda acceso a diagnósticos relativamente uniformes y en la mayoría de los casos, determina su probabilidad de malignidad.

### **3.2. Técnicas estadística**

Para lograr los objetivos propuestos, se aplicaron los siguientes métodos estadísticos: HJ-Biplot, Biplot logístico y el método de Asignación Latente de Dirichlet (LDA).

### **3.3. HJ-Biplot**

El método biplot puede verse como el análogo multivariante de los gráficos de dispersión, donde las muestras (o casos) se grafican como puntos en dos variables. Este es un análisis de datos exploratorio, método que busca patrones ocultos en una matriz de datos. Los biplot fueron introducidos por (Gabriel, 1971) en el contexto de los Análisis de Componentes Principales (ACP) (Hotelling, 1933; Pearson, 1901).

La característica fundamental que hace diferenciar las distintas representaciones gráficas asociadas a los métodos clásicos de reducción de dimensionalidad es el hecho de que, en un biplot, es posible obtener una representación conjunta de filas y columnas de una matriz de datos. Por lo tanto, con los biplot, la distribución multivariante de un conjunto de variables se puede aproximar en un espacio de baja dimensión (generalmente dos dimensiones) dando una visualización útil de toda la estructura.

Desde el punto de vista algebraico, el biplot se basa en los mismos principios que apoyan la mayoría de las técnicas factoriales de reducción de dimensionalidad, es decir, utiliza la Descomposición de Valores Singulares (DVS), mediante esta técnica es posible encontrar una buena representación en un espacio de baja dimensión, además, DVS proporciona las coordenadas de las dimensiones (o direcciones en el espacio), que generalmente se denominan (en sentido matemático) vectores singulares.

Las dimensiones están dispuestas de tal manera que son ortogonales y representan sucesivamente la mayor cantidad de variación posible, además, la técnica proporciona medidas (valores singulares) que, si se elevan al cuadrado, indican la cantidad de variabilidad explicada por cada dimensión (Kroonenberg, 2008). Entonces, para mostrar la variabilidad principal en un gráfico bidimensional, se deben usar las dos primeras dimensiones.

La característica fundamental de los métodos biplots, es que los datos se reproducen y luego se incorporan en una representación conjunta de filas y columnas. Se designaron las dos factorizaciones biplot más importantes propuestas por (Gabriel, 1971): GH-Biplot y JK-Biplot. GH-Biplot obtiene una representación de alta calidad para las  $J$  columnas (variables) y no tan alta para las  $I$  filas (individuos), mientras que JK-Biplot logra una representación de alta calidad para las filas y no tan alta para las columnas.

Utilizando una elección adecuada de marcadores para las filas y columnas, Galindo en los años 1985 - 1986 propuso una variante de las técnicas estándar de biplot, el HJ-Biplot, que permite la representación simultánea en el mismo sistema de coordenadas, también ha demostrado que esta alternativa asegura una alta calidad de representación tanto para las filas como para las columnas.

De hecho, esta es la principal ventaja de este método de ordenación que proporciona mejores resultados que otras técnicas multivariantes convencionales (Galindo, 1986; Galindo & Cuadras, 1986) HJ-Biplot se aplica en varias investigaciones, ver por ejemplo (Martínez-Ruiz et al., 2001; Mendes et al., 2009; Rivas-Gonzalo et al., 1993; Santos et al., 1991). Por lo tanto, se presentan brevemente los antecedentes teóricos de la técnica HJ-Biplot.

### **3.3.1 Propiedades de los marcadores en el HJ Biplot**

El HJ-Biplot es una técnica de representación simétrica simultánea que, hasta cierto punto, la extensión se asemeja al Análisis de Correspondencia (AC) (Benzecri, 1973), pero no se limita a los datos de frecuencia. Es una representación conjunta, en un espacio vectorial de baja dimensión (generalmente en dos dimensiones), de las filas ( $I$ ) y columnas ( $J$ ) de la matriz de datos bidireccional  $\mathbf{X}$ , utilizando marcadores (puntos / vectores), por sus filas ( $\mathbf{a}_1, \dots, \mathbf{a}_I$ ) y por sus columnas ( $\mathbf{b}_1, \dots, \mathbf{b}_J$ ). Los marcadores se obtienen de la DVS habitual de la matriz de datos.

La matriz inicial  $\mathbf{X}_{I \times J}$  se puede escribir de acuerdo con la DVS definida como:

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \quad (1)$$

Donde  $\mathbf{U}$  es una matriz cuyos vectores columna son ortonormales y corresponden a los vectores propios de  $\mathbf{X}\mathbf{X}^T$ ,  $\mathbf{V}$  es una matriz cuyos vectores columna son también ortonormales y se

corresponden con los vectores propios de  $\mathbf{X}^T\mathbf{X}$ ,  $\mathbf{\Lambda}$  es una matriz diagonal que contiene los valores singulares de la matriz  $\mathbf{X}$ , que son las raíces cuadradas no negativas de los valores propios de  $\mathbf{X}^T\mathbf{X}$  ordenados de manera decreciente, para las matrices  $\mathbf{U}$  y  $\mathbf{V}$  sean ortonormales debe cumplirse que  $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$ , esta propiedad asegura la unicidad de la factorización. La elección adecuada de marcadores para las filas ( $I$ ) y para las columnas ( $J$ ) en un subespacio de baja dimensión proporciona la representación biplot.

Para el caso de HJ-Biplot, se elige la siguiente factorización:

$$\mathbf{J} = \mathbf{U}\mathbf{\Lambda} \quad (2)$$

y

$$\mathbf{H} = \mathbf{V}\mathbf{\Lambda} \quad (3)$$

Donde  $\mathbf{J}$  son los marcadores de las columnas de la matriz  $\mathbf{X}$  y  $\mathbf{H}$  los marcadores de las filas de la matriz  $\mathbf{X}$ . Ambos marcadores se pueden representar en el mismo sistema de referencia. De hecho, teniendo en cuenta la relación entre  $\mathbf{U}$  y  $\mathbf{V}$ , es decir,  $\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{\Lambda}^{-1}$ :

$$\mathbf{U}\mathbf{\Lambda} = (\mathbf{X}\mathbf{V}\mathbf{\Lambda}^{-1})\mathbf{\Lambda} = \mathbf{X}\mathbf{V} \quad (4)$$

Los marcadores elegidos para las filas coinciden con la proyección de los puntos  $I$  en el espacio de los componentes principales de las columnas.

Similar,  $\mathbf{V} = \mathbf{X}^T\mathbf{U}\mathbf{\Lambda}^{-1}$ , por o tanto:

$$\mathbf{V}\mathbf{\Lambda} = (\mathbf{X}^T\mathbf{U}\mathbf{\Lambda}^{-1})\mathbf{\Lambda} = \mathbf{X}^T\mathbf{U} \quad (5)$$

Es decir, los marcadores elegidos para las columnas coinciden con la proyección de los puntos  $J$  en el espacio de los componentes principales de las filas.

Ambas representaciones están relacionadas. Permite,  $\mathbf{A} = \mathbf{X}\mathbf{V}$  y  $\mathbf{B} = \mathbf{X}^T\mathbf{U}$ , luego:

$$\mathbf{B} = \mathbf{X}^T\mathbf{U} = \mathbf{X}^T\mathbf{X}\mathbf{V}\mathbf{\Lambda}^{-1} = \mathbf{X}^T\mathbf{A}\mathbf{\Lambda}^{-1} \quad (6)$$

y

$$\mathbf{A} = \mathbf{X}\mathbf{V} = \mathbf{X}\mathbf{X}^T\mathbf{U}\mathbf{\Lambda}^{-1} = \mathbf{X}\mathbf{B}\mathbf{\Lambda}^{-1} \quad (7)$$

Una vez que se tiene eso, la coordenada de la  $J$  columna se puede expresar como una función de la  $h$  coordenadas de las  $I$  filas.

La bondad de ajuste para las dos primeras dimensiones de HJ-Biplot se logra mediante:

$$\frac{\lambda_1^2 + \lambda_2^2}{\sum_{s=1}^S \lambda_s^2} \quad (8)$$

donde  $\lambda_s$  son los valores propios.

Y así, el HJ-Biplot (como se mencionó anteriormente), logra una calidad de representación óptima tanto para filas como para columnas, ya que ambas están representadas en el mismo sistema de referencia. Además, está estrechamente relacionado con el Análisis de Componentes Principales (ACP), ya que las matrices de varianza y covarianza se grafican en un plan que explica la mayor parte de la inercia (Galindo, 1986; Galindo & Cuadras, 1986).

Las reglas para la interpretación del HJ-Biplot son una combinación de las reglas utilizadas en otras técnicas de escalado multidimensional (como AC, Análisis Factorial y biplot clásico) se basan en conceptos geométricos muy simples. A continuación, se presentan esas reglas interpretativas básicas (Galindo, 1986; Mendes et al., 2009):

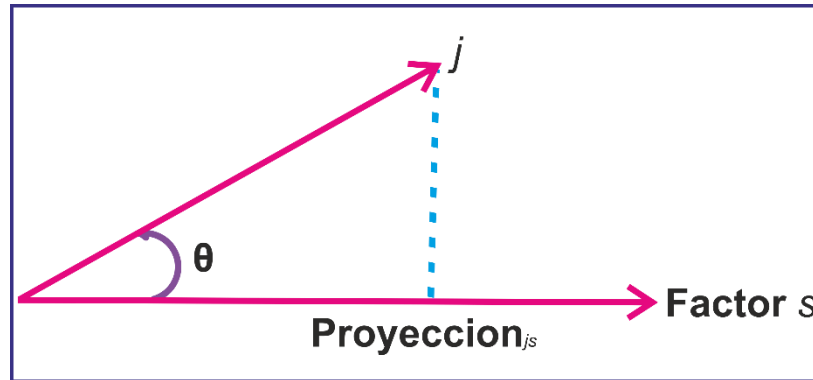
- Las distancias entre los marcadores de fila se interpretan como una función inversa de similitudes, de tal manera que los marcadores más cercanos (individuos) son más similares. Esta propiedad permite la identificación de grupos de individuos con perfiles similares;
- Las longitudes de los marcadores de columna (vectores) se aproximan a la desviación estándar de las variables;
- Los cosenos de los ángulos entre los vectores columna aproximan las correlaciones entre variables de tal manera que pequeños ángulos agudos se asocian con variables con altas correlaciones positivas; los ángulos obtusos cercanos al ángulo recto se asocian con variables con altas correlaciones negativas y los ángulos rectos se asocian con variables no correlacionadas. De la misma forma, los cosenos de los ángulos entre los marcadores de variable y los ejes (Componentes Principales) aproximan las correlaciones entre ellos. Para datos estandarizados, estos se aproximan a las cargas factoriales en el análisis factorial;
- El orden de las proyecciones ortogonales de los marcadores de fila (puntos) sobre un marcador de columna (vector) se aproxima al orden de los elementos de fila (valores) en esa columna. Cuanto mayor es la proyección de un punto individual sobre un vector variable, más se desvía este individuo del promedio de esa variable.
- Varias mediciones son esenciales para una correcta interpretación de HJ-Biplot (Galindo; 1986; Galindo & Cuadras, 1986).

Por ejemplo:

- La contribución absoluta del elemento de columna  $j$  (variable) al factor  $s$

$$CAE_j F_s = b_{js}^2 \quad (9)$$

Donde  $b_{js}^2$  denota la proyección de la variable  $j$  en la dimensión  $s$  (Figura 1)



**Figura 1.** Representación geométrica de la proyección de la variable  $j$  en la dimensión  $s$

- La contribución relativa del elemento de columna  $j$  (variables) al factor  $s$ .

$$CRE_j F_s = \frac{b_{js}^2}{\lambda_s} \quad (10)$$

Expresa la importancia de la variable para explicar un factor o dimensión.

- La contribución relativa del factor  $s$  al elemento de columna  $j$  (variables).

$$CRF_s E_j = \frac{b_{js}^2}{\sum_s b_{js}^2} \quad (11)$$

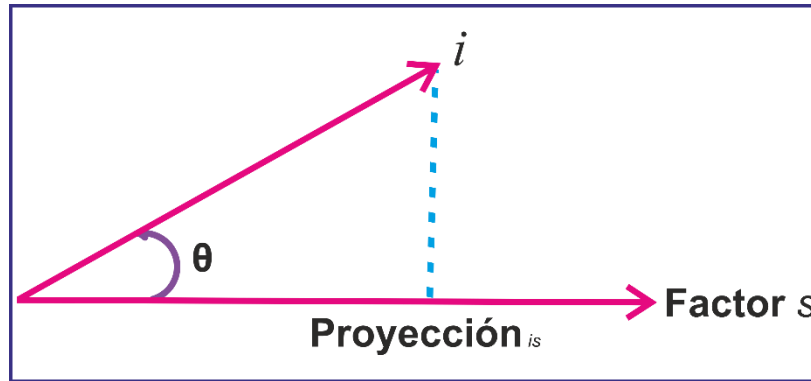
Expresa la variabilidad relativa de la variable explicada por un factor o dimensión.

Por analogía, para los elementos de fila

- La contribución absoluta del elemento de fila  $i$  (individuos) al factor  $s$ .

$$CAE_i F_s = a_{is}^2 \quad (12)$$

Donde  $a_{is}^2$  denota la proyección de la fila  $i$  en la dimensión  $s$  (Figura 2).



**Figura 2.** Representación geométrica de la proyección de la variable  $i$  en la dimensión  $s$

- La contribución relativa del elemento de columna  $i$  (individuos) al factor  $s$

$$CRE_{iF_s} = \frac{a_{is}^2}{\lambda_s} \quad (13)$$

Expresa la importancia de la variable para explicar un factor o dimensión

- La contribución relativa del factor  $s$  al elemento de columna  $i$  (individuos)

$$CRF_s E_i = \frac{a_{is}^2}{\sum_s a_{is}^2} \quad (14)$$

Expresa la variabilidad relativa de la variable explicada por un factor o dimensión.

Además, para un punto (fila o columna) en un plan factorial, la calidad de representación (QR) se puede definir sumando las dos contribuciones relativas de estos factores al elemento, es decir:

- Para el elemento de la columna  $j$

$$QRE_{j_{(1-2)}} = CRF_1 E_j + CRF_2 E_j = \cos^2 \theta_{j(1-2)} \quad (15)$$

Para el elemento fila  $i$

$$QRE_{i_{(1-2)}} = CRF_1 E_i + CRF_2 E_i = \cos^2 \theta_{i(1-2)} \quad (16)$$



Solo los puntos con una alta calidad de representación pueden interpretarse adecuadamente (Galindo, 1986; Galindo & Cuadras, 1986).

### 3.4. Biplot Logístico

Sea  $\mathbf{X}_{I \times J}$  una matriz de datos donde las filas corresponden a  $I$  individuos y las columnas a  $J$  caracteres binarios. Sea  $\pi_{ij} = E(x_{ij})$  la probabilidad observada donde el carácter  $j$  esté presente en el individuo  $i$ , y  $x_{ij}$  la probabilidad observada. Usualmente  $x_{ij}$  es 0 o 1, lo que da como resultado una matriz de datos binaria. El Biplot Logístico se formula como:

$$\pi_{ij} = \frac{e^{b_{j0} + \sum_s b_{js} a_{is}}}{1 + e^{b_{j0} + \sum_s b_{js} a_{is}}} \quad (17)$$

Donde  $a_{is}$  y  $b_{js}$  ( $i = 1, \dots, I; j = 1, \dots, J; s = 1, \dots, S$ ) son los parámetros del modelo utilizados como marcadores de fila y columna respectivamente

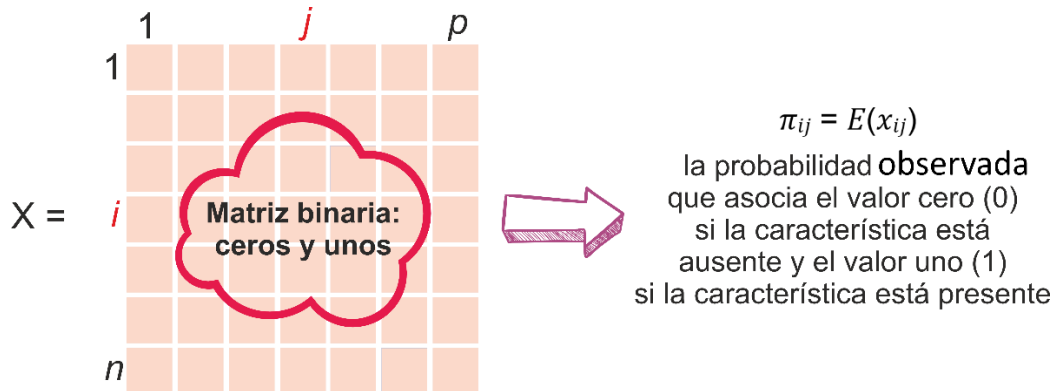


Figura 3. Matriz binaria

El modelo en (17) es un modelo (bi) lineal generalizado que tiene el logit como función de enlace.

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = b_{j0} + \sum_{s=1}^S b_{js} a_{is} = b_{j0} + \mathbf{a}_i^T \mathbf{b}_j \quad (18)$$

donde  $\mathbf{a}_i = (a_{i1}, \dots, a_{is})^T$  y  $\mathbf{b}_j = (b_{j1}, \dots, b_{js})^T$

Matricialmente:

$$\text{logit}(\mathbf{\Pi}) = \mathbf{1}_I \mathbf{b}_0^T + \mathbf{A} \mathbf{B}^T \quad (19)$$

donde:

- $\Pi$  • Matriz de probabilidades esperadas
- $\mathbf{1}_I$  • Vector de unos
- $\mathbf{b}_0 = (b_{j0})$  • Término que contiene las constantes
- $\mathbf{A}$  y  $\mathbf{B}$  • Matrices que contienen los marcadores para las filas y columnas de la matriz  $\mathbf{X}$

El modelo (19) es similar a los modelos de la teoría del rasgo latente o de la respuesta al ítem, donde los ejes de ordenación son considerados como variables latentes, estos explican la asociación entre las variables observadas. En ese sentido, se supone que los individuos responden de forma independiente a las variables y que las variables son independientes para valores dados de los rasgos latentes.

De acuerdo con los supuestos, la función de verosimilitud es:

$$Prob(x_{ij} | (\mathbf{b}_0, \mathbf{A}, \mathbf{B})) = \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{x_{ij}} (1 - \pi_{ij})^{1-x_{ij}} \quad (20)$$

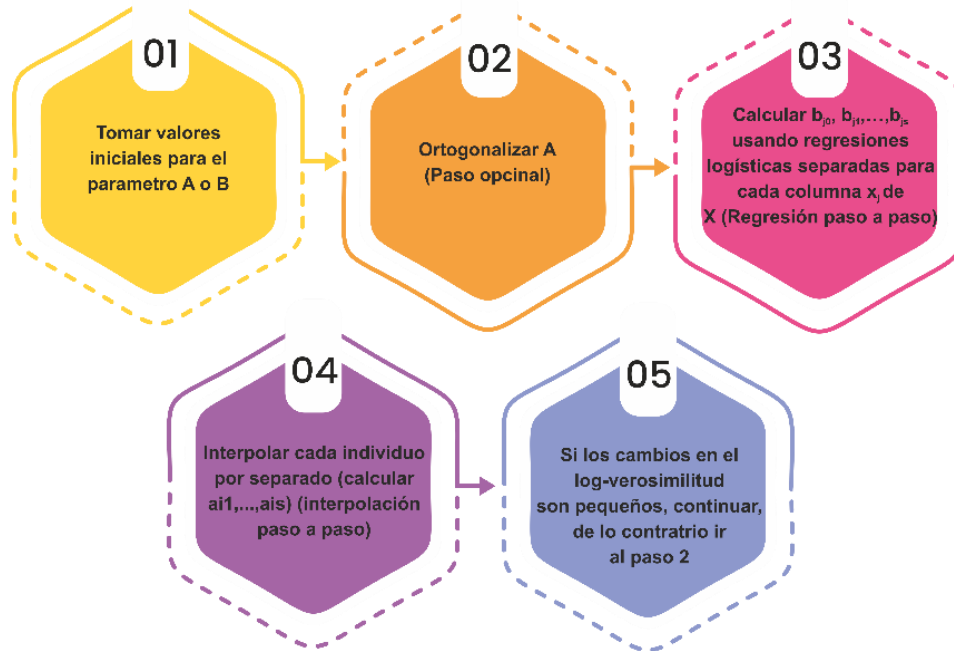
Tomando el logaritmo de la función de verosimilitud se obtiene:

$$L = \log Prob(x_{ij} | (\mathbf{b}_0, \mathbf{A}, \mathbf{B})) = \sum_{i=1}^I \sum_{j=1}^J [x_{ij} \log(\pi_{ij}) + (1 - x_{ij}) \log(1 - \pi_{ij})] \quad (21)$$

Para obtener las estimaciones es necesario tomar las derivadas de  $L$  con respecto a todos los parámetros, igualarlos a cero y resolver las ecuaciones simultáneas  $3J + 2I$ . (Vicente-Villardón et al., 2006) precisan los parámetros del Biplot Logístico en una representación que cambia  $A$  y  $B$ . Un conjunto de parámetros es introducido, mientras que el otro se mantiene fijo, donde la dinámica se va repitiendo hasta que la probabilidad converge a un grado de precisión esperado.

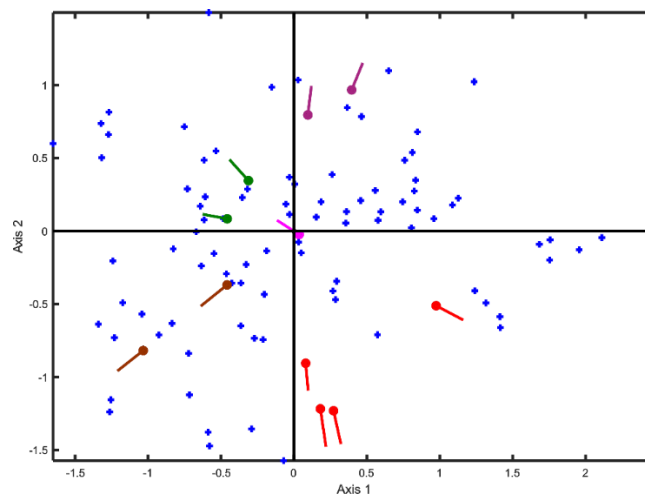
### 3.4.1. Algoritmo general

El Biplot Logístico representa a las variables e individuos, este no es referente a las puntuaciones del individuo en la variable, este método representa a la probabilidad que el individuo tenga la característica que define la variable. Por lo que su resultado se ve en técnicas de la interpretación de en el plano resultante (Figura 4).



**Figura 4.** Aplicación del Biplot Logístico

El conjunto de puntos que predicen el 0.5 es entonces una línea recta perpendicular al segmento que pasa por el origen del vector; esa línea divide la representación en dos regiones, que predicen respectivamente la presencia y ausencia de una característica particular (Figura 5), sin embargo, no todas las variables estarán asociadas a la clasificación, es por ello, que solo se proyectan aquellas que están con una mejor calidad de representación.



**Figura 5.** Representación bidimensional del Biplot Logístico

### **3.4.2. Interpretación Biplot Logístico**

Para determinar qué tipo de (variable) se relaciona más con que (Individuo) de acuerdo con el gradiente o factor latente, al calcularse las cargas de los factores múltiples. Las variables están más relacionadas a un gradiente particular, cuando el valor absoluto de la carga del factor es mayor que el coseno para los gradientes y cercano a la unidad; lo anterior es interpretado como una correlación entre los tipos de (variables) y el gradiente (factor latente). Esto puede ser interpretado de la siguiente manera:

- La longitud del vector predice el poder discriminante de la variable; a menor longitud, mayor poder discriminante, y viceversa
- La dirección a la que apunta el vector predice la zona positiva del gradiente; pudiendo determinarse que variables se asocian con cada eje.
- El ángulo entre vectores predice el grado de asociación entre esas variables.
- El ángulo entre cada vector y el eje predice el grado de relación entre el vector y el eje (gradiente).

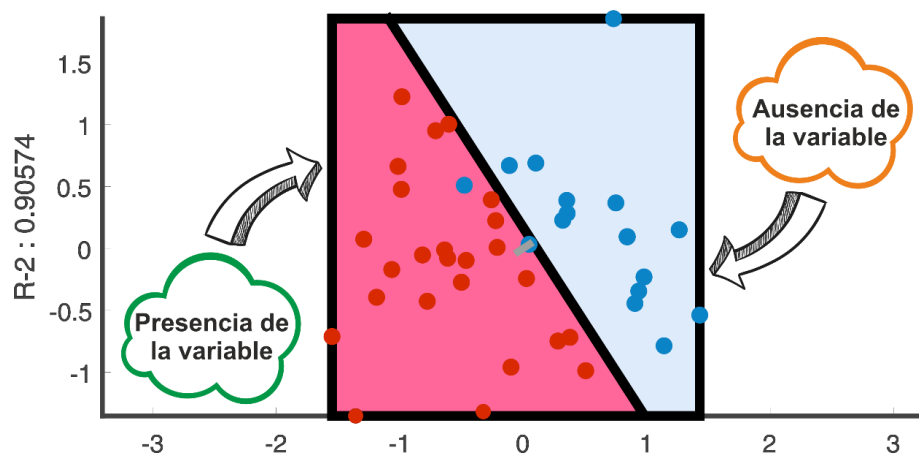
### **3.4.3. Calidad de representación del Biplot Logístico**

Al obtener los resultados del Biplot Logístico, es posible medir la calidad de cada una de las variables, teniendo en cuenta tres elementos fundamentales:

- a. El p-valor a partir del cual se determina qué variables son significativas en el modelo.
- b.  $r^2$  de Nagelkerke para obtener la capacidad explicativa del modelo logístico.
- c. El porcentaje de variables bien clasificadas obtenidas de las probabilidades esperadas, teniendo en cuenta como referencia el percentil 0.5 como indicador de corte para la predicción de presencia y ausencia (menor de 0.5 indica ausencia).

### **3.4.4. Regiones de predicción**

Asimismo, es posible adquirir para cada variable un diagrama de ordenación, este predice la presencia y ausencia del atributo de acuerdo con la división de la variable en dos regiones (Vicente-Villardón, 2010). Las dos regiones son separadas por una recta perpendicular al vector que representa a la variable y corta al vector en el punto 0.5. El comienzo de cada vector pertenece a la probabilidad de 0.5 que indica la presencia del individuo, y lo que corresponde al otro extremo de la flecha pertenece a la probabilidad de 0.75.



**Figura 6.** Regiones de predicción del Biplot Logístico

### 3.4.5. Contribuciones Al Biplot

Desde su comienzo, el método biplot (Gabriel, 1971) se ha convertido en un excelente método de representación gráfica. De esta forma, en la estadística multivariantes, cada día se agregan nuevas investigaciones, lo que hace un mayor aporte al desarrollo y aplicación del método biplot teniendo en cuenta sus diferentes variantes.

El método biplot tradicionalmente se ha utilizado en diagnósticos de modelos y con fines descriptivos (Bradú & Gabriel, 1974). Actualmente, los estudios sobre la metodología biplot se sigue utilizando por investigadores y estos combinan está, con técnicas clásicas, como el análisis de componentes principales, el análisis de varianza, entre otros.

Desde la perspectiva de la teoría y la práctica, han surgido nuevas técnicas analíticas, por tanto, los resultados de la nueva ruta de investigación se recogen en un gráfico biplot. En paralelo a este desarrollo, se manifiestan nuevas opciones de análisis de datos, mediante diversas aplicaciones informáticas.

La siguiente, es una tabla resumen que muestra las contribuciones más destacadas del método biplot, durante un periodo superior a cuarenta años desde sus inicios.

**Tabla 2.** Contribuciones Biplot

Referencia Contribución	Referencia Contribución
(Gabriel, 1971)	<ul style="list-style-type: none"> <li>• Creador del Biplot Clásico</li> <li>• Plantea dos tipos de modelos: GH-Biplot y JK-Biplot</li> </ul>
(Greenacre, 1984)	<ul style="list-style-type: none"> <li>• Presenta una nueva forma de nombrar el GH Biplot y JK Biplot</li> </ul>

(Kempton, 1984)	<ul style="list-style-type: none"> <li>• Les llama CMP y RMP correspondientemente</li> <li>• Demuestra que los métodos biplot aumentan la información obtenida con otros métodos.</li> </ul>
(M. P. G. Villardón, 1986)	<ul style="list-style-type: none"> <li>• Demuestra la conveniencia de representar filas y columnas sobre un mismo sistema de coordenadas</li> <li>• Define el HJ-Biplot</li> </ul>
(Ter Braak, 1986)	<ul style="list-style-type: none"> <li>• Utiliza los biplot en el contexto del Análisis Directo del Gradiente</li> <li>• Muestra la utilidad del biplot en la relación entre el análisis de redundancia y la correlación canónica</li> </ul>
(Blázquez, 1998)	<ul style="list-style-type: none"> <li>• Incluye información externa de filas y columnas</li> <li>• Utiliza modelos lineales generalizados alternos</li> </ul>
(Gauch Jr, 1988)	<ul style="list-style-type: none"> <li>• Utiliza el biplot para la validación de modelos y estudia la relación genotipo ambiente.</li> </ul>
(J C Gower & Harding, 1988)	<ul style="list-style-type: none"> <li>• Definen los biplot no lineales</li> </ul>
(Tavera, 1992)	<ul style="list-style-type: none"> <li>• Realiza una clasificación jerárquica ascendente utilizando el HJ-Biplot</li> <li>• Basa esta relación en el concepto de inercia</li> </ul>
(John C Gower, 1992)	<ul style="list-style-type: none"> <li>• Señala que en el biplot la factorización no está basada en DVS</li> </ul>
(Braak & Looman, 1994)	<ul style="list-style-type: none"> <li>• Incorpora el biplot de la matriz de coeficientes de la regresión y el biplot basado en regresión de rango reducido.</li> </ul>
(J. L. V. Villardón, 1992)	<ul style="list-style-type: none"> <li>• Define los biplot generalizados</li> </ul>
(Gómez, 1995)	<ul style="list-style-type: none"> <li>• Propone el HJ-Biplot como una alternativa del Análisis Canónico de correspondencia</li> </ul>
(J C Gower & Hand, 1996)	<ul style="list-style-type: none"> <li>• Definen los biplot de interpolación y predicción</li> </ul>
(Rodríguez, 1996)	<ul style="list-style-type: none"> <li>• Desarrolla los Meta-biplot</li> <li>• Define una estructura consenso para estudiar la relación</li> </ul>
(Carlier & Kroonenberg, 1996)	<ul style="list-style-type: none"> <li>• En base a los modelos de Tucker y Tuckals3 proponen el biplot interactivo y conjunto para tablas de 3 vías.</li> </ul>

(Cárdenas & Galindo Villardón, 2001)	<ul style="list-style-type: none"> <li>• Estudian el biplot desde el punto de vista inferencial bajo la metodología de los modelos bilineales generalizados</li> </ul>
(J. L. V. Villardón et al., 2001)	<ul style="list-style-type: none"> <li>• Generaliza el MANOVA Biplot de una vía al caso de dos factores de variación</li> </ul>
(Crossa et al., 2002; Yan et al., 2001; Yan & Hunt, 2002; Yan & Kang, 2002)	<ul style="list-style-type: none"> <li>• Utiliza el biplot como modelos bilineales para el análisis de la interacción genotipo ambiente</li> </ul>
(Baccala, 2004)	<ul style="list-style-type: none"> <li>• Desarrolla y hace una propuesta sobre el biplot de múltiples vías</li> </ul>
(Vairinhos, 2003)	<ul style="list-style-type: none"> <li>• Desarrolla un sistema de minería de datos a través de los métodos biplot</li> </ul>
(González & Villardón, 2013)	<ul style="list-style-type: none"> <li>• Profundiza en el estudio de la metodología biplot en presencia de valores atípicos</li> </ul>
(Vicente-Villardón et al., 2006)	<ul style="list-style-type: none"> <li>• Desarrolla el biplot logístico</li> </ul>
(Demey et al., 2008)	<ul style="list-style-type: none"> <li>• Desarrolla la teoría del biplot logístico externo</li> </ul>
(Egido & Egido, 2014)	<ul style="list-style-type: none"> <li>• Desarrolla el Biplot dinámico</li> </ul>
(Librero, 2015), (Nieto-Librero et al., 2017)	<ul style="list-style-type: none"> <li>• Implementa el Clustering Disjoint Biplot</li> </ul>
(Librero, 2015)	<ul style="list-style-type: none"> <li>• Realiza el análisis biplot de forma inferencial, combinándolo con los métodos Bootstrap.</li> </ul>
(Hernández Sánchez, 2016); (Sanchez et al., 2015; Sánchez & Vicente-Villardón, 2013)	<ul style="list-style-type: none"> <li>• Desarrolla la teoría del biplot para datos de tipo nominal y ordinal</li> </ul>

### 3.5. Método LDA

Actualmente, el logaritmo Asignación Latente de Dirichlet (LDA) propuesto por (Blei et al., 2003) se destaca dentro de los algoritmos más populares para el modelado de tópicos (Blei, 2012; DiMaggio et al., 2013; Grimmer, 2010; Jacobi et al., 2016), los algoritmos Análisis Semántico Latente (LSA) (Deerwester et al., 1990) y el Análisis Semántico Latente Probabilístico (PLSA) (Hofmann, 1999), Serán descritos a continuación junto con frecuencia de término – frecuencia inversa de documento TF-IDF (Salton et al., 1975).

Para las secciones posteriores se utilizarán las definiciones y nomenclaturas comunes entre las técnicas o algoritmos de modelación de tópicos (Blei et al., 2003):

- **Palabra:** unidad básica para datos discretos, es definida como un componente de un vocabulario indexado por  $\{1, \dots, V\}$ . Se representan palabras utilizando vectores unitarios que poseen un solo componente igual a uno y el restante igual a cero. Se utilizarán superíndices

para indicar componentes. La palabra  $v$ -ésima en el vocabulario es representada por un vector  $V$  tal que  $w^v = 1$  y  $w^u = 0$  para  $u \neq v$ ;

- **Documento:** Secuencia de  $N$  palabras expresadas por  $w = (w_1, w_2, \dots, w_N)$ , donde  $w_n$  es la  $n$ -ésima palabra en la serie;
- **Corpus:** Subconjunto de  $M$  documentos, formado de acuerdo con una serie de criterios de diseño claros para una intención específica (Atkins et al., 1992) denotados por  $D = \{w_1, w_2, \dots, w_M\}$ ;
- **Tópico:** Variable latente expresada por  $Z = \{z_1, z_2, \dots, z_k\}$  donde el número  $k$  es un parámetro que se debe especificar.

### 3.5.1. Frecuencia de Término - Frecuencia Inversa de Documento (TF-IDF)

El modelo algebraico TF-IDF (Salton et al., 1975), representa documentos de texto como vectores. Existen dos supuestos (Sebastiani, 2002). El primer supuesto, representa la cantidad de veces que una palabra sale en el documento, es decir, es más representativa si su frecuencia en el documento es mayor, a esto se le da como nombre Frecuencia de Término (TF). El segundo supuesto, no resalta aquellas palabras que son más frecuentes en el documento y resalta aquellas que aparecen con poca frecuencia, a esto se le conoce con el nombre de Frecuencia de Documento Inversa (IDF) tiene como definición:

$$idf_i = \log \frac{|D|}{|\{d: t_i \in d\}|} \quad (22)$$

donde  $|D|$  representa el número total de documentos en el corpus y  $|\{d: t_i \in d\}|$  es el número de documentos que aparecen en el término  $t_i$ . Si el término  $t_i$  aparece en todos los documentos del corpus,  $idf_i$  es igual a 0. Si  $idf_i$  es mayor al término  $t_i$  aparecerán pocos documentos. Se asume el valor TF-IDF de la manera siguiente:

$$t_d = f_{t,d} * \log \frac{|D|}{|\{d: t_i \in d\}|} \quad (23)$$

donde  $f_{t,d}$  es igual a la frecuencia que el término  $t$  aparece en el documento  $d$ .

### 3.5.2. Análisis Semántico Latente (LSA)

Quien introdujo el Análisis Semántico Latente (LSA) fue (Deerwester et al., 1990). El algoritmo LSA tiene como objetivo, tomar fragmentos textuales y representarlos como una ecuación lineal, esto corresponde a la suma de los significados de cada palabra que lo conforman y las frecuencias en que estas co-ocurren en aquel fragmento (Landauer et al., 1998).



El algoritmo LSA corresponde a un análisis de componentes principales aplicado, que determina una matriz dimensionalmente constituida por  $n$  (documentos) por  $m$  (palabras), donde cada  $a_{ij}$  referente al elemento que se encuentra usualmente definido por una frecuencia ponderada  $i$  en el documento  $j$ . El algoritmo LSA utiliza DVS para encontrar un subespacio de baja dimensión, la mayor parte de la variación en la colección de documentos se encuentran en el espacio de las características TF – IDF.

LSA está compuesto por las etapas siguientes (Evangelopoulos et al., 2012):

1. Generación de la matriz de entrada. La matriz generada de acuerdo con los términos del documento, en donde las celdas son utilizadas para la representación de lo importante de las palabras en las oraciones. Hay enfoques existentes para los valores puestos en las celdas. Los más utilizados son:

- Las Frecuencias de las palabras: Completa la celda con las frecuencias de cada palabra en la oración;
- Representación binaria: cada celda se determina con 0/1 teniendo en cuenta la presencia o ausencia de una palabra en la oración;
- TF – IDF: cada celda se determina con el valor correspondiente TF – IDF de la palabra. Cuando el valor TF – IDF es alto tiene como significado que la palabra tiene mayor frecuencia en la oración, pero poco frecuente en el documento. Cuando los valores de la palabra aumentan, significa que esa palabra tiene mayor representatividad en la oración;
- Log entropía (Shannon, 1948): se completa la celda con el valor log – entropía de la palabra, esta suministra información sobre la palabra con respecto a la oración.

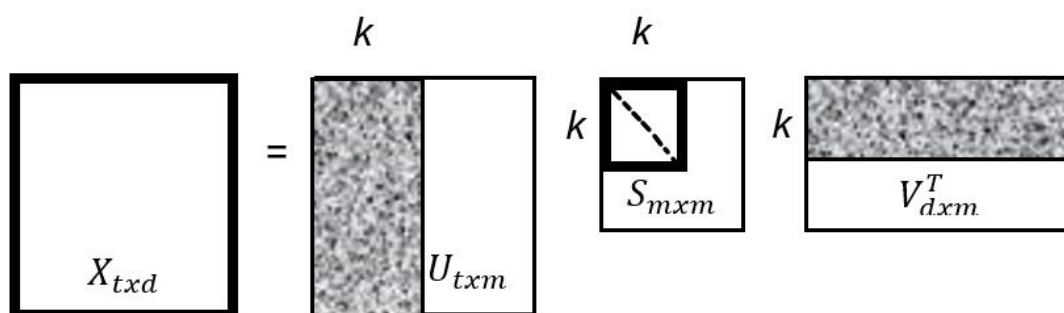
La dimensionalidad para las matrices en este paso tiende a ser grandes y con porcentajes altos en los alores de las celdas iguales a cero, he ahí el nombre de *matriz sparse* (matrices dispersas). Consecuentemente, se destaca que LSA toma las palabras más frecuentes y las elimina, esto se conoce como *stopword* que consiste en la eliminación de aquellas palabras que no aportan significado.

2. Reducción de la dimensionalidad: para este paso se tiene en cuenta la generación SDV en la matriz termino – documento. La idea principal es que exista una representación de puntos en un espacio euclidiano de la matriz de los términos del documento, donde los vectores son utilizados para revelar los documentos u oraciones en este espacio.

La dinámica de la DVS es descomponer la matriz términos – documento  $X_{txd}$ , en el producto de otras tres matrices: la matriz  $U_{txm}$  columna-ortogonal con  $m$  que representa la

dimensionalidad, la matriz diagonal  $S_{m \times m}$  con valores singulares dispuestos en orden decreciente, y, una matriz transpuesta  $V_{d \times m}$  columna-ortogonal (Lochbaum & Streeter, 1989), donde  $t$  indica el número de términos y  $d$  indica el número de documentos. Subsiguientemente, las matrices se truncan en un número  $k$  arbitrario de dimensiones, para suprimir parte del ruido existente en la matriz original y así extraer la relación semántica latente en la colección (Figura 7).

3. Análisis cuantitativo: finalizando, cada termino y documento que se representa en el espacio definido por  $k$  dimensiones retenidas mediante la SDV se determina el análisis de acuerdo con el método analítico específico según el objetivo buscado.



**Figura 7.** Descomposición y truncamiento de la matriz términos - documento (adoptado de (Berry & Kogan, 2010))

### 3.5.3. Análisis Semántico Latente Probabilístico (PLSA)

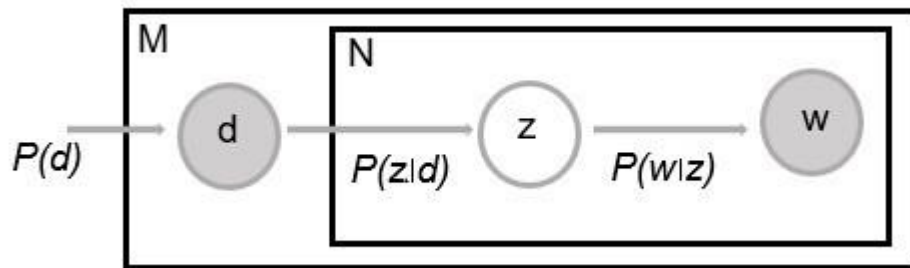
Anteriormente el modelo consistía en un modelo determinista, sin embargo, es posible determinar un modelo probabilístico teniendo en cuenta el espacio determinado según documento y palabras. En ese sentido, este modelo es llamado Análisis Semántico Probabilístico (PLSA) definido por (Hofmann, 1999). Diferenciado con su antecesor LSA, el algoritmo PLSA se caracteriza por tener una base consolidada en inferencia estadística. A diferencia de LSA que se deriva del álgebra lineal y ejecuta una SVD sobre la matriz términos – documentos.

PLSA es un modelo que genera clases latentes y descompone la probabilidad de matrices o matriz como mezcla de distribuciones multinomiales condicionadas de forma independiente que se modelan utilizando el algoritmo de maximización de expectativas (EM). Este algoritmo, es el método que encuentra las estimaciones de los parámetros con mayor probabilidad para el modelo que depende de variables latentes no observadas (en este caso, los tópicos).

No de los supuestos de los modelos de tópicos consiste en que estos se mezclan y cada colección de palabras es un tópico. El modelo PLSA añade un giro probabilístico a este supuesto y consiste en:

- Teniendo en cuenta un documento  $d$ , el tema  $z$  se halla presente en el documento con probabilidad  $P(z/d)$ ;
- Teniendo en cuenta un tema  $z$ , la palabra  $w$  se extrae de  $z$  con probabilidad  $P(w/z)$ .

La figura 2 muestra de manera gráfica el modelo con la notación “placas” esta se utiliza frecuentemente para representar modelos gráficos probabilísticos. Los cuadrados son “placas” estas representan replicas, que son identidades repetidas. En ese sentido, la placa superficial representa los documentos ( $M$ ), sin embargo, la placa interna representa todas las posiciones de las palabras que se repiten en un documento ( $N$ ); las posiciones se asocian con la elección del tópico ( $z$ ) y de la palabra ( $w$ ). Consecuentemente los nodos vienen siendo etiquetados utilizando la variable que representan, los nodos pintados en color gris son la representación de las variables aleatorias observadas del modelo, sin embargo, los que no se encuentran pintados, son la representación de las variables aleatorias latentes de este.



**Figura 8.** Notación de placa representado en el modelo asimétrico del Análisis Semántico Latente Probabilístico (adoptado (Blei et al., 2003)).

Para determinar la probabilidad de un conjunto de observaciones en una palabra y un documento es:

$$P(D, W) = P(D) \sum_z P(Z|D)P(W|Z) \quad (24)$$

$$P(D, W) = \sum_z P(Z)P(D|Z)P(W|Z) \quad (25)$$

Si se observa la ecuación tres, esta indica la probabilidad de ver cierto documento, de esta manera y teniendo en cuenta el cómo se distribuye el tópico en el documento, se puede determinar la probabilidad de encontrar cierta palabra en el interior del documentó. Para la ecuación  $P(D)$ ,  $P(Z|D)$  y  $P(W|Z)$  estos representan los parámetros del modelo.  $P(D)$  se puede

determinar teniendo en cuenta el corpus, sin embargo, los parámetros restantes se modelizan como distribuciones multinomiales considerando el algoritmo EM.

La ecuación (4) muestra una de las parametrizaciones, esta se utiliza como un conjunto de tres parámetros, en esta se puede observar un paralelismo de manera directa entre los modelos PLSA y LSA donde:

$$P(Z) = S, P(D|Z) = U \text{ y } P(W|Z) = V^T \quad (26)$$

Según lo anterior se puede observar que el modelo PLSA es un modelo más flexible, sin embargo, este tiene algunos problemas  $P(D)$ , ya que no se sabe el cómo asignar probabilidad a otros documentos. Asimismo, la parametrización para PLSA tiene un comportamiento de crecimiento lineal con el número de documentos que se tiene, este comportamiento tiende a sobre ajustar el modelo.

#### 3.5.4. Latent Dirichlet Allocation (LDA)

El método LDA consiste en la versión bayesiana de PLSA basada en el supuesto de la bolsa de palabras (Blei et al., 2003). Lo anterior tiene como significado que las palabras en un documento se pueden intercambiar y estos se representan como series de palabras individuales. Quien aplico el modelo LDA a corpus textuales inicialmente fue (Blei et al., 2003), sin embargo, este ha sido usado mediante imágenes (Iwata et al., 2007) y videos (Wang et al., 2007).

El modelo LDA es aquel que estudia el cómo se producen datos, y de acuerdo con la generación del modelo, se cuestiona que variable objeto los genera. Este modelo pide que la caracterización de los tópicos y documentos se extraigan de la distribución Dirichlet, esta consiste en la generalización multivariante de la matriz beta.

Teniendo en cuenta a (Blei et al., 2003), el modelo generativo se puede resumir de la siguiente manera:

1. A cada tópico  $K$ , se le extrae una distribución según las palabras  $\phi_k \sim Dir(\alpha)$ ;
2. Teniendo en cuenta cada documento  $d$ ,
  - a. Se extrae un vector de proporción de tópico  $\theta_d \sim Dir(\beta)$ ;
  - b. Se extrae una palabra  $w_{d,i} \sim Mult(\theta_{zd,i}), w_{d,i} \in \{1, \dots, V\}$ .

La nomenclatura que se utiliza para la descripción del modelo generativo, esta se encuentra en la tabla 2.

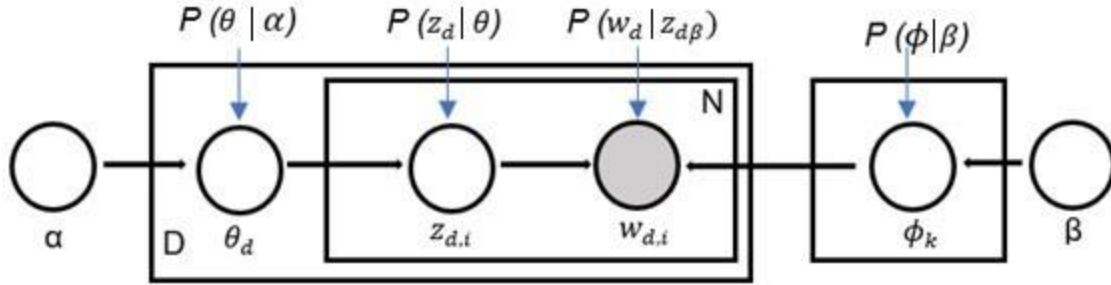
**Tabla 3.** Significado de cada variable para el modelo LDA

Variable	Definición
K	Número de tópico
V	Número de palabras en el vocabulario
M	Número de documentos
$N_{d=1...M}$	Número de palabras en el documento $d$
N	El número total de palabras en todos los documentos; suma de todos los valores, es decir, $N_d N = \sum_{d=1}^M N_d$
$\alpha_{K=1...K}$	Hiperparámetro a priori del tópico k en un documento; por lo general el mismo para todos los temas; normalmente un número menor que 1, por ejemplo 0.1, a preferir distribuciones tema dispersos, es decir, pocos tópicos por documento
$\alpha$	Colección de todos los valores del hiperparámetro, visto como un único vector $\alpha_k$
$\beta_{w=1...v}$	Hiperparámetro a priori de la palabra w en un tema; suele ser el mismo para todas las palabras; normalmente un número mucho menor que 1, por ejemplo, 0.001, a preferir fuertemente distribuciones de palabras dispersas, es decir, pocas palabras por tema
$\beta$	Colección de todos los valores del hiperparámetro, visto como un único vector $\beta_w$
$\phi_{k=1...K, \omega=1...V}$	Probabilidad de que palabra w se encuentren el en tópico k
$\phi_{k=1...K}$	La distribución de las palabras en el tópico k
$\theta_{d=1...M, K=1...K}$	Probabilidad de tópico k se produzca en el documento $d$
$\theta_{d=1...M}$	Distribución de los tópicos en el documento $d$
$\mathcal{Z}_{d=1...M, k=1...K}$	Identidad del tópico de la palabra w en el documento $d$
z	Identidad del tópico de todas las palabras en todos los documentos
$w_{d=1...M, w=1...N_d}$	La identidad de la palabra w en el documento $d$
w	La identidad de todas las palabras en todos los documentos

Fuente: (Blei et al., 2003)

El método LDA tiene como objetivo inferir o estimar las variables latentes, esto quiere decir, que su distribución es calculada de acuerdo con los documentos. En la figura 8, se encuentra una representación gráfica teniendo en cuenta la notación de placas, en donde se puede deducir la distribución conjunta de la siguiente manera:

$$P(w, z, \theta, \phi | \alpha, \beta) = P(\theta | \alpha) P(z | \theta) P(\phi | \beta) P(w | z, \phi) \quad (27)$$



**Figura 9.** Notación de placa del modelo de Asignación Latente de Dirichlet. Adoptado de (Blei et al., 2003).

A la derecha de la ecuación 24 se encuentra:

- $P(\vartheta|\alpha)$ : Esta representa la distribución de tópicos por documento, teniendo en cuenta el parámetro Dirichlet, el cual es un vector-K con componentes  $\alpha_k > 0$  (el operador punto ( $\alpha \cdot$ ) se introduce en la ecuación 25, es la abreviatura en el índice de las variables para la suma de todos los valores.

$$P(\theta|\alpha) = \frac{\Gamma(\alpha \cdot)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_k^{\alpha_k - 1} \quad (28)$$

- $P(z|\theta)$ : corresponde a la distribución del tópico  $z$  en el corpus, este depende de la distribución anteriormente mencionada. Consecuentemente, a cada palabra  $w_i$  en el documento de  $N$  palabras, se le determina un valor de  $1, \dots, K$ .

$$P(z|\theta) = \prod_{d=1}^D \prod_{k=1}^K \theta_{d,k}^{n_{d,k}} \quad (29)$$

Esta distribución  $P(z|\vartheta)$  expresa la probabilidad de  $z$  para todos los documentos y también los tópicos en termino de numero de palabras  $n_{d,k}$ , esta es la cantidad de veces en el cual se ha asignado un tema  $k$  a cualquier palabra que se encuentra en el documento  $d$ .

- $P(\phi|\beta)$ : Representa las distribuciones de los tópicos según el termino, teniendo en cuenta todo el corpus  $\phi_k$ . Estas se obtienen de una distribución de Dirichlet con parámetro  $\beta$ .  $\phi_{k,v}$  proporciona la probabilidad en donde el termino  $v$  se adquiriera de acuerdo con el tópico elegido, este se enuncia para todos los tópicos y cada una de las palabras del vocabulario así:

$$P(\phi|\beta) = \prod_{k=1}^K \frac{\Gamma(\beta_{k,\cdot})}{\prod_{v=1}^V \Gamma(\beta_{k,v})} \prod_{v=1}^V \phi_{k,v}^{\beta_{k,v} - 1} \quad (30)$$

- Finiquitando, la probabilidad de un corpus  $w$  teniendo en cuenta  $z$  y  $\phi$  en el modelo grafico es:

$$P\left(w|z, \phi\right) = \prod_{k=1}^K \prod_{v=1}^V \phi_{k,v}^{n_{k,v}} \quad (31)$$

La ecuación 24 se podría reescribir apartando las variables latentes, con el fin de considerar un modelo de probabilidad dado, un corpus  $w$  y los hiperparámetros  $(\alpha, \beta)$ . Para poder realizar una “estimación máximo-verosímil de los parámetros del modelo y para inferir la distribución de las variables latentes (Blei et al., 2003) es necesaria la probabilidad anterior mente nombrada.

$$P\left(\frac{w}{z}, \phi\right) = \int_{\phi} \int_{\theta} \sum_z \left( \prod_{d=1}^D \frac{\Gamma(\alpha)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_{d,k}^{\alpha_k + n_{d,k} - 1} \right) \left( \prod_{k=1}^K \frac{\Gamma(\beta_{k,\cdot})}{\prod_{v=1}^V \Gamma(\beta_{k,v})} \prod_{v=1}^V \phi_{k,v}^{\beta_{k,v} + n_{k,v} - 1} \right) d\theta d\phi \quad (32)$$

La adición sobre todas las composiciones posibles de asignaciones de tópicos hace que esta probabilidad sea computacionalmente intratable (Blei et al., 2003) y es por ello que se debe hacer uso de algoritmos de *Machine Learning* para hallar aproximaciones de la probabilidad marginal. Ya que la probabilidad a posteriori no puede calcularse de manera exacta, esta puede hallarse mediante una aproximación lo suficientemente cerca al verdadero valor teniendo en cuenta inferencia estadística.

Seguidamente se puede diferenciar dos tipos de algoritmos inferenciales: Los algoritmos basados en variaciones, estos consisten en la creación de una familia de distribuciones más cercanas (estas distancias se miden con la divergencia de Kullback-Leibler (KL) teniendo en cuenta el verdadero valor posteriori. Es necesario saber que algoritmos basados en la variación y en el muestreo proveen resultados con igualdad de precisión (Asuncion et al., 2012).

Medir el rendimiento en un modelo de tópicos es necesario para dar garantía de la generalización a partir de los datos de manera útil. Anteriormente se mencionó que los modelos de tópicos son modelos de variables latentes donde son utilizados para hacer colecciones entre palabras y los tópicos semánticos latentes se utilizan para las colecciones de documentos (Griffiths & Steyvers, 2004) y (Blei & Lafferty, 2007). Por definición el número de tópicos se debe establecer a priori.

Seguidamente, se puede decir, que no es trivial elegir un número adecuado de tópicos para una colección de documentos, ya que debe buscarse un balance donde la necesidad de una gran cantidad de tópicos y la necesidad de un número limitado de estos sea más fácil para que el experto pueda verificarlos y comprenderlos.

Teniendo en cuenta que objetivos y que medios tienen a disponibilidad computacionalmente el investigador, este puede emplear distintas métricas de rendimiento (Buntine, 2009; Chang & Blei, 2009; Wallach et al., 2009). Existen medidas que evalúan el desempeño de los modelos de tópicos probabilísticos en minería de texto, procesamiento, lenguaje natural y recuperación de información teniendo en cuenta otras áreas, se encuentran:

- La *perplejidad* definida (para un conjunto de prueba de  $M$  documentos) por (Blei et al., 2003) como:

$$perplejidad(D_{test}) = \left\{ -\frac{\sum_{d=1}^M \log P(w_d)}{\sum_{d=1}^M N_d} \right\} \quad (33)$$

$N_d$  es el equivalente al número de palabras en el documento  $d$ -ésimo del corpus de prueba  $D_{test}$  y  $W_d$  es el equivalente a las palabras en el corpus. El valor más bajo de perplejidad indica que el modelo es mejor.

- Para la validación cruzada en algunos casos, el modelo se puede evaluar mediante una validación cruzada del error de una tarea externa en cuestión, mediante la clasificación de los documentos o teniendo en cuenta la recuperación de la información (Titov & McDonald, 2008; X. Wei & Croft, 2006).
- Verosimilitud empírica (L. Wei & McCallum, 2006).
- Verosimilitud marginal, se aproxima por:
  - Media armónica (Griffiths & Steyvers, 2004; Newton & Raftery, 1994)
  - Chib-style (Chib, 1995)
  - Left-to-right samplers, (Del Moral et al., 2006);

Posteriormente, encontramos la *coherencia* (Röder et al., 2015), esta está basada en la hipótesis de distribución (Harris, 1954), asegura que las palabras con significados equivalentes tienden a coexistir en contextos equivalentes.

Para que se llegue a una puntuación general para el tópico la coherencia utiliza cuatro etapas (Röder et al., 2015): (1) división de las  $N$  palabras importantes del tema en pares; (2) cálculo de la probabilidad para cada palabra de forma individual o teniendo en cuenta pares de palabras; (3) calcular la medida de confirmación que captura el acuerdo de pares; y finalizando, (4) la incorporación de medidas de confirmación individuales en una puntuación general de coherencia del tema.



### 3.5.5. LDA en el entorno R

En la actualidad existen software de modelados de tópicos abierto y cerrado, sin embargo, quien implemento el primer software LDA fue David Blei en 2004, este efectúa inferencia variacional, esta fue explicada por (Blei et al., 2003). A continuación, se ofrecerá una visión sobre el cómo se implementa en el software R quien permite además del modelad de tópicos, además del manejo de datos textuales.

La red completa de archivos de R (Comprehensive R Archive Network-CRAN), por sus siglas en inglés, provee orientación sobre los paquetes que son relevantes para determinadas tareas. Allí se puede hallar una lista de paquetes afines con el procesamiento natural del lenguaje. En total se encuentran 59 paquetes afines con la minería de texto y modelado de tópicos, de los cuales ocho realizan el modelado de tópicos LDA (Tabla 3).

**Tabla 4.** Paquetes del software R para Asignación Latente Dirichlet

Paquete	Título	Cita
<b>lda</b>	Collapsed Gibbs Sampling Methods for Topic Models	Jonathan Chang (2015). lda: Collapsed Gibbs Sampling Methods for Topic Models. R package version 1.4.2. <a href="https://CRAN.R-project.org/package=lda">https://CRAN.R-project.org/package=lda</a>
<b>lda.svi</b>	Fit Latent Dirichlet Allocation Models using Stochastic Variational Inference	Nicholas Erskine (2019). lda.svi: Fit Latent Dirichlet Allocation Models using Stochastic Variational Inference. R package version 0.1.0. <a href="https://CRAN.R-project.org/package=lda.svi">https://CRAN.R-project.org/package=lda.svi</a>
<b>ldaPrototype</b>	Prototype of Multiple Latent Dirichlet Allocation Runs	Jonas Rieger (2020). ldaPrototype: Prototype of Multiple Latent Dirichle Allocation Runs. R package version 0.1.1. <a href="https://CRAN.R-project.org/package=ldaPrototype">https://CRAN.R-project.org/package=ldaPrototype</a>
<b>LDATS</b>	Latent Dirichlet Allocation Coupled with Time Series Analyses	Juniper L. Simonis, Erica M. Christensen, David J. Harris, Renata M. Diaz, Hao Ye, Ethan P. White and S.K. Morgan Ernest (2020). LDATS: Latent Dirichlet Allocation Coupled with Time Series Analyses. R package version 0.2.7. <a href="https://CRAN.R-project.org/package=LDATS">https://CRAN.R-project.org/package=LDATS</a>
<b>ldatuning</b>	Tuning of the Latent Dirichlet Allocation Models Parameters	Murzintcev Nikita (2019). ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters. R package version 1.0.0. <a href="https://CRAN.R-project.org/package=ldatuning">https://CRAN.R-project.org/package=ldatuning</a>
<b>LDavis</b>	Interactive Visualization of Topic Models	Carson Sievert and Kenny Shirley (2015). LDavis: Interactive Visualization of Topic Models. R package version 0.3.2. <a href="https://CRAN.R-project.org/package=LDavis">https://CRAN.R-project.org/package=LDavis</a>
<b>topicdoc</b>	Topic-Specific Diagnostics for LDA and CTM Topic Models	Doug Friedman (2019). topicdoc: Topic-Specific Diagnostics for LDA and CTM Topic Models. R package version 0.1.0. <a href="https://CRAN.R-project.org/package=topicdoc">https://CRAN.R-project.org/package=topicdoc</a>

<b>topicmodels</b>	Topic Models	Grün B, Hornik K (2011). "topicmodels: An R Package for Fitting Topic Models." <i>Journal of Statistical Software</i> , *40*(13), 1-30. doi: 10.18637/jss.v040.i13 (URL: <a href="https://doi.org/10.18637/jss.v040.i13">https://doi.org/10.18637/jss.v040.i13</a> ).
<b>LDAShiny</b>	An R package for exploratory review of scientific literature based on a Bayesian probabilistic model and machine learning tools	De la Hoz-M, J., Fernández-Gómez, M. J., & Mendes, S. (2021). LDAShiny: An R package for exploratory review of scientific literature based on a Bayesian probabilistic model and machine learning tools. <i>Mathematics</i> , 9(14), 1671.

---



## ***4. REFERENCIAS***



- Asuncion, A., Welling, M., Smyth, P., & Teh, Y. W. (2012). On smoothing and inference for topic models. *ArXiv Preprint ArXiv:1205.2662*.
- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing*. <https://doi.org/10.1093/lc/7.1.1>
- Baccala, N. (2004). *Contribuciones al analisis de matrices de datos multivaria: tipologia de las variables*. Phd Thesis.
- Beaumont, J. L., Cella, D., Phan, A. T., Choi, S., Liu, Z., & Yao, J. C. (2012). Comparison of health-related quality of life in patients with neuroendocrine tumors with quality of life in the general US population. *Pancreas*, 41(3), 461–466.
- Benzecri, J. P. (1973). *L'analyse des donnees, vol 2 Dunod*. Paris.
- Berry, M. W., & Kogan, J. (2010). Text Mining: Applications and Theory. In *Text Mining: Applications and Theory*. <https://doi.org/10.1002/9780470689646>
- Blázquez, A. (1998). *Análisis biplot basado en modelos lineales generalizados*. Tesis Doctoral, Universidad de Salamanca, España.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84.
- Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of Science. *The Annals of Applied Statistics*. <https://doi.org/10.1214/07-aos114>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*. <https://doi.org/10.1016/b978-0-12-411519-4.00006-9>
- Braak, C. J. F. Ter, & Looman, C. W. N. (1994). Biplots in reduced-rank regression. *Biometrical Journal*, 36(8), 983–1003.
- Bradu, D., & Gabriel, K. R. (1974). Simultaneous statistical inference on interactions in two-way analysis of variance. *Journal of the American Statistical Association*, 69(346), 428–436.
- Brask, J. B., Talman, M. M., & Wielenga, V. T. (2014). Neuroendocrine carcinoma of the breast—a pilot study of a Danish population of 240 breast cancer patients. *Apmis*, 122(7), 585–592.
- Buntine, W. (2009). Estimating likelihoods for topic models. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [https://doi.org/10.1007/978-3-642-05224-8\\_6](https://doi.org/10.1007/978-3-642-05224-8_6)
- Bussolati, G., & Badve, S. (2012). Carcinomas with neuroendocrine features. *WHO Classification of Tumours of the Breast*, 4, 62e63.
- Cárdenas, O. C., & Galindo Villardón, M. P. (2001). *Biplot con información externa basado en modelos bilineales generalizados*.

- Carlier, A., & Kroonenberg, P. M. (1996). Decompositions and biplots in three-way correspondence analysis. *Psychometrika*, *61*(2), 355–373.
- Chang, J., & Blei, D. M. (2009). Relational topic models for document networks. *Journal of Machine Learning Research*.
- Chauhan, A., Yu, Q., Ray, N., Farooqui, Z., Huang, B., Durbin, E. B., Tucker, T., Evers, M., Arnold, S., & Anthony, L. B. (2018). Global burden of neuroendocrine tumors and changing incidence in Kentucky. *Oncotarget*, *9*(27), 19245.
- Chen, L. C. (2017). An effective LDA-based time topic model to improve blog search performance. *Information Processing and Management*. <https://doi.org/10.1016/j.ipm.2017.08.001>
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.1995.10476635>
- Crossa, J., Cornelius, P. L., & Yan, W. (2002). *Biplots of linear-Bilinear models for studying crossover genotype X environment interaction [Erratum: Sept/Oct 2002, v. 42 (5), p. 1761.]*.
- Dasari, A., Shen, C., Halperin, D., Zhao, B., Zhou, S., Xu, Y., Shih, T., & Yao, J. C. (2017a). Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncology*. <https://doi.org/10.1001/jamaoncol.2017.0589>
- Dasari, A., Shen, C., Halperin, D., Zhao, B., Zhou, S., Xu, Y., Shih, T., & Yao, J. C. (2017b). Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol.* 2017; 3 (10): 1335–42. *PubMed PubMed Central Article*.
- de Miguel Novoa, M. P., Fernández Capel, F., Redondo Sedano, J. V., Sellers Carrera, M., Aranda Jiménez, V., Ortiz Pereira, P., Sastre Valera, J., Ortega Medina, L., & Díaz Pérez, J. Á. (2014). Tumores neuroendocrinos gastroenteropancreáticos: Características clínicas, proceso diagnóstico y pronóstico en el Hospital Universitario Clínico San Carlos (Madrid). *Endocrinología y Nutrición*. <https://doi.org/10.1016/j.endonu.2013.12.001>
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASI1>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9)
- Del Moral, P., Doucet, A., & Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*. <https://doi.org/10.1111/j.1467-9868.2006.00553.x>
- Demey, J. R., Vicente-Villardón, J. L., Galindo-Villardón, M. P., & Zambrano, A. Y. (2008). Identifying molecular markers associated with classification of genotypes by External Logistic Biplots. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btn552>

- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*. <https://doi.org/10.1016/j.poetic.2013.08.004>
- DiNordia, J., Lee, M. K., Reavey, P. L., Genkinger, J. M., Lee, J. A., Schrope, B. A., Chabot, J. A., & Allendorf, J. D. (2010). One hundred thirty resections for pancreatic neuroendocrine tumor: evaluating the impact of minimally invasive and parenchyma-sparing techniques. *Journal of Gastrointestinal Surgery*, *14*(10), 1536–1546.
- Egido, J., & Egido, M. J. (2014). *Package 'dynBiplotGUI.'*
- Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012). Latent semantic analysis: Five methodological recommendations. *European Journal of Information Systems*. <https://doi.org/10.1057/ejis.2010.61>
- Fasano, M., della Corte, C. M., Papaccio, F., Ciardiello, F., & Morgillo, F. (2015). Pulmonary large-cell neuroendocrine carcinoma: from epidemiology to therapy. *Journal of Thoracic Oncology*, *10*(8), 1133–1141.
- Fournel, L., Falcoz, P. E., Alifano, M., Charpentier, M.-C., Boudaya, M.-S., Magdeleinat, P., Damotte, D., & Régnard, J.-F. (2013). Surgical management of pulmonary large cell neuroendocrine carcinomas: a 10-year experience. *European Journal of Cardio-Thoracic Surgery*, *43*(1), 111–114.
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*. <https://doi.org/10.1093/biomet/58.3.453>
- Galindo, M. P., & Cuadras, C. M. (1986). Una extensión del método Biplot y su relación con otras técnicas. *Publicaciones de Bioestadística y Biomatemática*, *17*.
- Gauch Jr, H. G. (1988). Model selection and validation for yield trials with interaction. *Biometrics*, 705–715.
- Gómez, M. J. F. (1995). *Contribuciones al análisis multivariante directo del gradiente mediante estudio combinado de configuraciones espaciales*. Universidad de Salamanca.
- González, S. H., & Villardón, M. P. G. (2013). BIPROB: UN MÉTODO PARA OBTENER UN BIPLLOT ROBUSTO. *Investigación Operacional*, *27*(3), 287–299.
- Gower, J. C. (1992). Generalized biplots. *Biometrika*, *79*(3), 475–493.
- Gower, J. C., & Hand, D. J. (1996). *Biplots*. Chapman&Hall, London, UK.
- Gower, J. C., & Harding, S. A. (1988). Nonlinear biplots. *Biometrika*, *75*(3), 445–455.
- Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*.



- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0307752101>
- Grimmer, J. (2010). A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis*. <https://doi.org/10.1093/pan/mpp034>
- Group, M. of N. N. T., Öberg, K., Astrup, L., Eriksson, B., Falkmer, S. E., Falkmer, U. G., Gustafsen, J., Haglund, C., Knigge, U., & Vatn, M. H. (2004). Guidelines for the management of gastroenteropancreatic neuroendocrine tumours (including bronchopulmonary and thymic neoplasms) Part II—Specific NE Tumour Types. *Acta Oncologica*, 43(7), 626–636.
- Hamilton, S. R., & Aaltonen, L. A. (2000). *Pathology and genetics of tumours of the digestive system* (Vol. 2). IARC press Lyon:
- Harris, Z. S. (1954). Distributional Structure. WORD. <https://doi.org/10.1080/00437956.1954.11659520>
- Hernández Sánchez, J. C. (2016). *Biplot logístico para datos nominales y ordinales*.
- Hernando, J., & Capdevilla, J. (2020). *Tumores neuroendocrinos*. Sociedad Española de Oncología Médica. <https://seom.org/info-sobre-el-cancer/tumor-neuroendocrino?showall=1>.
- Hofmann, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1999*. <https://doi.org/10.1145/312624.312649>
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417.
- Howe, J. R., Cardona, K., Fraker, D. L., Kebebew, E., Untch, B. R., Wang, Y.-Z., Law, C. H., Liu, E. H., Kim, M. K., & Menda, Y. (2017). The surgical management of small bowel neuroendocrine tumors: consensus guidelines of the North American Neuroendocrine Tumor Society (NANETS). *Pancreas*, 46(6), 715.
- Iwata, T., Saito, K., Ueda, N., Stromsten, S., Griffiths, T. L., & Tenenbaum, J. B. (2007). Parametric embedding for class visualization. *Neural Computation*. <https://doi.org/10.1162/neco.2007.19.9.2536>
- Jacobi, C., Van Attevelde, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*. <https://doi.org/10.1080/21670811.2015.1093271>
- Jiao, X., Li, Y., Wang, H., Liu, S., Zhang, D., & Zhou, Y. (2015). Clinicopathological features and survival analysis of gastroenteropancreatic neuroendocrine neoplasms: a retrospective study in a single center of China. *Chinese Journal of Cancer Research*, 27(3), 258.

- Kao, A., & Poteet, S. R. (2007). Natural language processing and text mining. In *Natural Language Processing and Text Mining*. <https://doi.org/10.1007/978-1-84628-754-1>
- Kempton, R. A. (1984). The use of biplots in interpreting variety by environment interactions. *The Journal of Agricultural Science*, 103(1), 123–135.
- Klimstra, D. S. (2010). Neuroendocrine neoplasms of the pancreas. *WHO Classification of Tumours of the Digestive System*, 322–326.
- Klimstra, D. S., Modlin, I. R., Coppola, D., Lloyd, R. v, & Suster, S. (2010). The pathologic classification of neuroendocrine tumors: a review of nomenclature, grading, and staging systems. *Pancreas*, 39(6), 707–712.
- Kroonenberg, P. M. (2008). *Applied multiway data analysis* (Vol. 702). John Wiley & Sons.
- Kujtan, L., Muthukumar, V., Kennedy, K. F., Davis, J. R., Masood, A., & Subramanian, J. (2018). The role of systemic therapy in the management of stage I large cell neuroendocrine carcinoma of the lung. *Journal of Thoracic Oncology*, 13(5), 707–714.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*. <https://doi.org/10.1080/01638539809545028>
- Lewis, A., Raoof, M., Ituarte, P. H. G., Williams, J., Melstrom, L., Li, D., Lee, B., & Singh, G. (2019). Resection of the primary gastrointestinal neuroendocrine tumor improves survival with or without liver treatment. *Annals of Surgery*. <https://doi.org/10.1097/SLA.0000000000002809>
- Librero, A. B. N. (2015). *Versión inferencial de los métodos biplot basada en remuestreo bootstrap y su aplicación a tablas de tres vías*. Universidad de Salamanca.
- Lloyd, R. v, Osamura, Y. R., Kloppel, G., & Rosai, J. (2017). WHO classification of tumours of endocrine organs. In *WHO Press*.
- Lochbaum, K. E., & Streeter, L. A. (1989). Comparing and combining the effectiveness of latent semantic indexing and the ordinary vector space model for information retrieval. *Information Processing and Management*. [https://doi.org/10.1016/0306-4573\(89\)90100-3](https://doi.org/10.1016/0306-4573(89)90100-3)
- Martinez, L., Puig, I., & Valls, C. (2007). Colorectal liver metastases: Radiological diagnosis and staging. *European Journal of Surgical Oncology (EJSO)*, 33, S5–S16.
- Martínez-Ruiz, C., Fernández-Santos, B., & Gómez-Gutiérrez, J. M. (2001). Effects of substrate coarseness and exposure on plant succession in uranium-mining wastes. *Plant Ecology*, 155(1), 79–89.
- Mayo, S. C., de Jong, M. C., Pulitano, C., Clary, B. M., Reddy, S. K., Gamblin, T. C., Celinksi, S. A., Kooby, D. A., Staley, C. A., & Stokes, J. B. (2010). Surgical management of hepatic

- neuroendocrine tumor metastasis: results from an international multi-institutional analysis. *Annals of Surgical Oncology*, 17(12), 3129–3136.
- Mendes, S., Fernández-Gómez, M. J., Galindo-Villardón, M. P., Morgado, F., Maranhão, P., Azeiteiro, U. M., & Bacelar-Nicolau, P. (2009). The study of bacterioplankton dynamics in the Berlengas Archipelago (West coast of Portugal) by applying the HJ-biplot method. *ARQUIPÉLAGO-Life and Marine Sciences*, 25–35.
- Merath, K., Bagante, F., Beal, E. W., Lopez-Aguilar, A. G., Poultides, G., Makris, E., Rocha, F., Kanji, Z., Weber, S., Fisher, A., Fields, R., Krasnick, B. A., Idrees, K., Smith, P. M., Cho, C., Beems, M., Schmidt, C. R., Dillhoff, M., Maithel, S. K., & Pawlik, T. M. (2018). Nomogram predicting the risk of recurrence after curative-intent resection of primary non-metastatic gastrointestinal neuroendocrine tumors: An analysis of the U.S. Neuroendocrine Tumor Study Group. *Journal of Surgical Oncology*. <https://doi.org/10.1002/jso.24985>
- Modlin, I. M., Lye, K. D., & Kidd, M. (2003). A 5-decade analysis of 13,715 carcinoid tumors. *Cancer*. <https://doi.org/10.1002/cncr.11105>
- Moore, F. D., Scoinski, M. A., & Joste, N. E. (2003). Endocrine tumors and malignancies. *Atlas of Diagnostic Oncology. 3rd Ed. Philadelphia: Elsevier Science Limited*.
- Musunuru, S., Chen, H., Rajpal, S., Stephani, N., McDermott, J. C., Holen, K., Rikkers, L. F., & Weber, S. M. (2006). Metastatic neuroendocrine hepatic tumors: Resection improves survival. *Archives of Surgery*. <https://doi.org/10.1001/archsurg.141.10.1000>
- Newton, M. A., & Raftery, A. E. (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society: Series B (Methodological)*. <https://doi.org/10.1111/j.2517-6161.1994.tb01956.x>
- Nieto-Librero, A. B., Sierra, C., Vicente-Galindo, M. P., Ruíz-Barzola, O., & Galindo-Villardón, M. P. (2017). Clustering Disjoint HJ-Biplot: A new tool for identifying pollution patterns in geochemical studies. *Chemosphere*, 176, 389–396.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572.
- Riihimäki, M., Hemminki, A., Sundquist, K., Sundquist, J., & Hemminki, K. (2016). The epidemiology of metastases in neuroendocrine tumors. *International Journal of Cancer*, 139(12), 2679–2686.
- Rindi, G., Klersy, C., Albarello, L., Baudin, E., Bianchi, A., Buchler, M. W., Caplin, M., Couvelard, A., Cros, J., de Herder, W. W., Delle Fave, G., Doglioni, C., Federspiel, B., Fischer, L., Fusai, G., Gavazzi, F., Hansen, C. P., Inzani, F., Jann, H., ... Falconi, M. (2019). Competitive testing of the WHO 2010 versus the WHO 2017 grading of pancreatic neuroendocrine neoplasms: Data from a large international cohort study. *Neuroendocrinology*. <https://doi.org/10.1159/000494355>

- Rindi, G., Klöppel, G., Couvelard, A., Komminoth, P., Körner, M., Lopes, J. M., McNicol, A.-M., Nilsson, O., Perren, A., & Scarpa, A. (2007). TNM staging of midgut and hindgut (neuro) endocrine tumors: a consensus proposal including a grading system. *Virchows Archiv*, 451(4), 757–762.
- Rindi, G., & Wiedenmann, B. (2020). Neuroendocrine neoplasia of the gastrointestinal tract revisited: Towards precision medicine. *Nature Reviews Endocrinology*, 16(10), 590–607.
- Rivas-Gonzalo, J. C., Gutiérrez, Y., Polanco, A. M., Hebrero, E., Vicente, J. L., Galindo, P., & Santos-Buelga, C. (1993). Biplot analysis applied to enological parameters in the geographical classification of young red wines. *American Journal of Enology and Viticulture*, 44(3), 302–308.
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408.
- Rodríguez, J. M. (1996). *Contribuciones a la Integración de Subespacios desde una Perspectiva Biplot*. Universidad de Salamanca.
- Rosen, L. E., & Gattuso, P. (2017). Neuroendocrine tumors of the breast. In *Archives of Pathology and Laboratory Medicine*. <https://doi.org/10.5858/arpa.2016-0364-RS>
- Rossi, G., Cavazza, A., Marchioni, A., Longo, L., Migaldi, M., Sartori, G., Bigiani, N., Schirosi, L., Casali, C., & Morandi, U. (2005). Role of chemotherapy and the receptor tyrosine kinases KIT, PDGFR $\alpha$ , PDGFR $\beta$ , and Met in large-cell neuroendocrine carcinoma of the lung. *Journal of Clinical Oncology*, 23(34), 8774–8785.
- Salton, G., Wong, A., & Yang, C. S. (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM*. <https://doi.org/10.1145/361219.361220>
- Sánchez, J. C. H., & Vicente-Villardón, J. L. (2013). Logistic biplot for nominal data. *ArXiv Preprint ArXiv:1309.5486*.
- Sanchez, J. C. H., Vicente-Villardón, J. L., & Sanchez, M. J. C. H. (2015). *Package 'OrdinalLogisticBiplot.'*
- Santos, C., Muñoz, S. S., Gutiérrez, Y., Hebrero, E., Vicente, J. L., Galindo, P., & Rivas, J. C. (1991). Characterization of young red wines by application of HJ biplot analysis to anthocyanin profiles. *Journal of Agricultural and Food Chemistry*, 39(6), 1086–1090.
- Sarkaria, I. S., Iyoda, A., Roh, M. S., Sica, G., Kuk, D., Sima, C. S., Pietanza, M. C., Park, B. J., Travis, W. D., & Rusch, V. W. (2011). Neoadjuvant and adjuvant chemotherapy in resected pulmonary large cell neuroendocrine carcinomas: a single institution experience. *The Annals of Thoracic Surgery*, 92(4), 1180–1187.

- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys (CSUR)*, 34(1), 1–47.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Taal, B. G., & Visser, O. (2004a). Epidemiology of neuroendocrine tumours. *Neuroendocrinology*. <https://doi.org/10.1159/000080731>
- Taal, B. G., & Visser, O. (2004b). Epidemiology of neuroendocrine tumours. *Neuroendocrinology*, 80(Suppl. 1), 3–7.
- Tavera, S. V. (1992). *Las técnicas de representación de datos multidimensionales en el estudio del Índice de Producción Industrial (IPI) en la CEE*. Universidad de Salamanca.
- Ter Braak, C. J. F. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67(5), 1167–1179.
- Titov, I., & McDonald, R. (2008). Modeling online reviews with multi-grain topic models. *Proceeding of the 17th International Conference on World Wide Web 2008, WWW'08*. <https://doi.org/10.1145/1367497.1367513>
- Vairinhos, V. M. (2003). *Desarrollo de un sistema para minería de datos basado en los métodos BIPLLOT*.
- Vanoli, A., Albarello, L., Uncini, S., Fassan, M., Grillo, F., di Sabatino, A., Martino, M., Pasquali, C., Milanetto, A. C., & Falconi, M. (2019). Neuroendocrine Tumors (NETs) of the Minor Papilla/Ampulla. *The American Journal of Surgical Pathology*, 43(6), 725–736.
- Vicente-Villardón. (2010). MULTBIPLLOT: A package for Multivariate Analysis using Biplots. In *Departamento de Estadística. Universidad de Salamanca*.
- Vicente-Villardón, J. L., Galindo-Villardón, M. P., & Blázquez-Zaballos, A. (2004). Constrained Logistic Biplots. *SALAMANCA STATISTICS SEMINAR V*, 34.
- Vicente-Villardón, J. L., Galindo-Villardón, M. P., & Blázquez-Zaballos, A. (2006). Logistic biplots. *Multiple Correspondence Analysis and Related Methods*. London: Chapman & Hall, 503–521.
- Villardón, J. L. V. (1992). *Una alternativa a las técnicas factoriales clásicas basada en una generalización de los métodos Biplot*. Universidad de Salamanca.
- Villardón, J. L. V., Amaro, R., Galindo-Villardón, P., & Ramírez, G. (2001). Manova Biplot para diseños con varios factores basado en modelos lineales generales multivariantes. *XXVI Congreso Nacional de Estadística e Investigación Operativa: Úbeda, 6-9 de Noviembre de 2001*, 123.

- Galindo, M. P. (1986). Una alternativa de representacion simultanea: HJ-Biplot. *Qüestió: Quaderns d'estadística i Investigació Operativa*, 13–23.
- Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/1553374.1553515>
- Wang, Y., Sabzmejdani, P., & Mori, G. (2007). Semi-latent dirichlet allocation: A hierarchical model for human action recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [https://doi.org/10.1007/978-3-540-75703-0\\_17](https://doi.org/10.1007/978-3-540-75703-0_17)
- Wei, B., Ding, T., Xing, Y., Wei, W., Tian, Z., Tang, F., Abraham, S., Nayeemuddin, K., Hunt, K., & Wu, Y. (2010). Invasive neuroendocrine carcinoma of the breast: A distinctive subtype of aggressive mammary carcinoma. *Cancer*. <https://doi.org/10.1002/cncr.25352>
- Wei, L., & McCallum, A. (2006). Pachinko allocation: DAG-structured mixture models of topic correlations. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/1143844.1143917>
- Wei, X., & Croft, W. B. (2006). LDA-based document models for ad-hoc retrieval. *Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. <https://doi.org/10.1145/1148170.1148204>
- Yan, W., Cornelius, P. L., Crossa, J., & Hunt, L. A. (2001). Two types of GGE biplots for analyzing multi-environment trial data. *Crop Science*, 41(3), 656–663.
- Yan, W., & Hunt, L. A. (2002). Biplot analysis of diallel data. *Crop Science*, 42(1), 21–30.
- Yan, W., & Kang, M. S. (2002). *GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists*. CRC press.
- Yao, J. C., Eisner, M. P., Leary, C., Dagohoy, C., Phan, A., Rashid, A., Hassan, M., & Evans, D. B. (2007). Population-based study of islet cell carcinoma. *Annals of Surgical Oncology*, 14(12), 3492–3500.
- Yao, J. C., Hassan, M., Phan, A., Dagohoy, C., Leary, C., Mares, J. E., Abdalla, E. K., Fleming, J. B., Vauthey, J.-N., & Rashid, A. (2008). One hundred years after “carcinoid”: epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *Journal of Clinical Oncology*, 26(18), 3063–3072.
- Yao, J. C., Lombard-Bohas, C., Baudin, E., Kvols, L. K., Rougier, P., Ruzniewski, P., Hoosen, S., Peter, J. S., Haas, T., & Lebwohl, D. (2010). Daily oral everolimus activity in patients with metastatic pancreatic neuroendocrine tumors after failure of cytotoxic chemotherapy: a phase II trial. *Journal of Clinical Oncology*, 28(1), 69.

Yao, J., & Phan, A. T. (2012). Optimising therapeutic options for patients with advanced pancreatic neuroendocrine tumours. *Eur Oncol Haematol*, 8(4), 217–223.

Zamora, V. (2009). La nueva clasificación de la OMS. Nuevos desafíos. *Acta Gastroenterológica Latinoamericana*, 39(1), 15–17.





## ***5. COHERENCIA Y RELACIÓN DIRECTA ENTRE LOS ARTÍCULOS PRESENTADOS***



Este capítulo describe la coherencia de cada uno de los artículos que componen esta tesis doctoral. Los artículos se encuentran alineados con un trabajo de investigación bien definido y planificado orientado a métodos estadísticos multivariantes para el análisis de Tumores Neuroendocrinos (TNE).

---

La presencia de TNE es relativamente rara, presenta una incidencia anual de menos de 10 casos por 100.000 habitantes. Sin embargo, la prevalencia de datos de los TNE ha aumentado en los últimos años debido al desarrollo y la amplia aplicación de tecnologías modernas de imagenología y endoscopia. En los EUA, la prevalencia de duración limitada a 20 años aumentó drásticamente del 0,006% en 1993 al 0,048% en 2012 y los TNE se observaron con más frecuencia en el recto, seguidos de los pulmones y el intestino delgado.

Según datos SEER (Surveillance, Epidemiology and End Results) de los EUA la prevalencia pasó de 3.9 en 1995 a 6.61 en 2012. A pesar del aumento en la prevalencia reportado en los estudios, cabe mencionar que aún faltan datos sobre los TNE en los países subdesarrollados. El tema es particularmente importante, así el presente trabajo se propuso a generar el primer reporte de los NET en Ecuador. Los datos fueron cuidadosamente evaluados. Se postulo la hipótesis de que la incidencia y prevalencia de los NET aumentó desde los primeros registros a lo largo de los años de acuerdo con datos publicados en otros países.

Teniendo en cuenta estos antecedentes, los estudios retrospectivos en diferentes poblaciones son importantes como una alerta para el diagnóstico precoz de la enfermedad por parte de los médicos, lo que determina la sobrevida del paciente, además de proporcionar datos que destacan el órgano más acometido con la enfermedad. Una vez publicado los artículos se determinó los primeros reportes sobre los TNE en Ecuador, los síntomas más relevantes de acuerdo con la ubicación del tumor y los temas más estudiados a nivel mundial sobre los TNE y de esta manera se contribuye significativamente con los datos a nivel mundial sobre esta neoplasia. A continuación, se detallan las técnicas estadísticas propuestas en cada una de las publicaciones realizadas.

El artículo *FREQUENCY OF NEUROENDOCRINE TUMORS STUDIES: USING LATENT DIRICHLET ALLOCATION AND HJ-BIPLLOT STATISTICAL METHODS* propone una alternativa de análisis estadístico con los métodos HJ-Biplot y Latent Dirichlet Allocation (LDA), el método LDA analiza la literatura publicada a nivel mundial acerca de los Tumores Neuroendocrinos (TNE) desde una probabilidad bayesiana e identifica los principales tópicos a través del tiempo evaluado (40 años). El segundo método HJ-Biplot describe patrones y tendencias temporales de los tópicos sobre tumores neuroendocrinos, HJ-Biplot ofrece hacer una evaluación más precisa poniendo de

manifiesto las relaciones existentes entre las partes, esta técnica mantiene una representación gráfica simultánea de los tópicos y los años en estudio, además con la longitud de los vectores se puede observar la variabilidad que representa los tópicos y la similitud que hay entre estos.

El artículo *NEUROENDOCRINE TUMORS: AN ANALYSIS OF PREVALENCE, INCIDENCE, AND SURVIVAL IN A HOSPITAL-BASED STUDY IN ECUADOR*, propone una alternativa HJ-Biplot para determinar la relación existente entre las variables, grado de diferenciación y los meses de supervivencia, según ubicación del tumor de los pacientes en estudio, también se evaluó el sexo y la edad según la localización del tumor, esto teniendo en cuenta la información generada en los hospitales de la Sociedad de Lucha Contra el Cáncer (SOLCA) durante los periodos 2005 - 2020. Caber resaltar que este método no había sido utilizado anteriormente para analizar tumores neuroendocrinos desde la prevalencia, incidencia y supervivencia.

El artículo *CLINICAL RELATED FACTORS TO NEUROENDOCRINE TUMORS IN ECUADORIAN PATIENTS: A LOGISTIC BIPLLOT APPROACH* propone una alternativa para identificar los síntomas más frecuentes en pacientes con TNE utilizando variables binarias (presencia/ausencia) con distribución binomial. Esta metodología permitió obtener una representación gráfica de los tipos de tumores neuroendocrinos y los síntomas principales de los pacientes. En este sentido el enfoque proporciona una mejor interpretación para relacionar el tipo de tumor con sus síntomas, en donde una gráfica simultánea representa las columnas (síntomas) y las filas (pacientes) de la matriz de datos. La propuesta se basa en el hecho de que el Biplot Logístico define la calidad de la representación de los individuos, grupos de individuos y variables (tipo de tumor o síntoma principal).

Con base en lo expuesto, se puede observar, que las técnicas propuestas para este estudio ofrecen un nuevo enfoque sobre el análisis de Tumores Neuroendocrinos, estas facilitan la interpretación de los análisis de resultados en comparación con otros métodos tradicionales de agrupamiento.



***6. PUBLICACIONES ORIGINALES QUE CONFORMAN LA TESIS  
DOCTORAL***



## *6.1 Frequency of Neuroendocrine Tumor Studies: Using Latent Dirichlet Allocation and Hj-Biplot*





## Article

# Frequency of Neuroendocrine Tumor Studies: Using Latent Dirichlet Allocation and HJ-Biplot Statistical Methods

Karime Montes Escobar <sup>1,2</sup>, José Luis Vicente-Villardón <sup>2</sup>, Javier de la Hoz-M <sup>3</sup>, Lelly María Useche-Castro <sup>1</sup>, Daniel Fabricio Alarcón Cano <sup>4</sup> and Aline Siteneski <sup>5,\*</sup>

<sup>1</sup> Department of Mathematics and Statistics, Institute of Basic Sciences, Technical University of Manabí, Portoviejo 130105, Ecuador; karime.montes@gmail.com (K.M.E.); lelly.useche@utm.edu.ec (L.M.U.-C.)

<sup>2</sup> Department of Statistics, University of Salamanca, 37008 Salamanca, Spain; villardon@usal.es

<sup>3</sup> Facultad de Ingeniería, Universidad del Magdalena, Santa Marta 470004, Colombia; jdelahoz@unimagdalena.edu.co

<sup>4</sup> Teaching and Research, SOLCA, Manabí, Portoviejo 130105, Ecuador; alarcon.daniel.md@gmail.com

<sup>5</sup> Research Institute, Faculty of Health Sciences, Technical University of Manabí, Portoviejo 130105, Ecuador

\* Correspondence: aline.siteneski@gmail.com; Tel.: +593-99129-2040

**Abstract:** Background: Neuroendocrine tumors (NETs) are severe and relatively rare and may affect any organ of the human body. The prevalence of NETs has increased in recent years; however, there seem to be more data on particular types, even though, despite the efforts of different guidelines, there is no consensus on how to identify different types of NETs. In this review, we investigated the countries that published the most articles about NETs, the most frequent organs affected, and the most common related topics. Methods: This work used the Latent Dirichlet Allocation (LDA) method to identify and interpret scientific information in relation to the categories in a set of documents. The HJ-Biplot method was also used to determine the relationship between the analyzed topics, by taking into consideration the years under study. Results: In this study, a literature review was conducted, from which a total of 7658 abstracts of scientific articles published between 1981 and 2020 were extracted. The United States, Germany, United Kingdom, France, and Italy published the majority of studies on NETs, of which pancreatic tumors were the most studied. The five most frequent topics were  $t_{21}$  (clinical benefit),  $t_{11}$  (pancreatic neuroendocrine tumors),  $t_{13}$  (patients one year after treatment),  $t_{17}$  (prognosis of survival before and after resection), and  $t_3$  (markers for carcinomas). Finally, the results were put through a two-way multivariate analysis (HJ-Biplot), which generated a new interpretation: we grouped topics by year and discovered which NETs were the most relevant for which years. Conclusions: The most frequent topics found in our review highlighted the severity of NETs: patients have a poor prognosis of survival and a high probability of tumor recurrence.

**Keywords:** neuroendocrine tumor 1; pancreatic neuroendocrine tumor 2; prevalence 3; Latent Dirichlet Allocation 3; HJ-Biplot 4



**Citation:** Escobar, K.M.; Vicente-Villardón, J.L.; de la Hoz-M, J.; Useche-Castro, L.M.; Alarcón Cano, D.F.; Siteneski, A. Frequency of Neuroendocrine Tumor Studies: Using Latent Dirichlet Allocation and HJ-Biplot Statistical Methods. *Mathematics* **2021**, *9*, 2281. <https://doi.org/10.3390/math9182281>

Academic Editor: Seifedine Kadry

Received: 13 August 2021

Accepted: 6 September 2021

Published: 16 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Neuroendocrine tumors (NETs) are so called because the organs they affect present cells with neural and endocrine properties [1]. NETs are malignant epithelial neoplasms with a predominant neuroendocrine differentiation [2].

NET severity may be graded from 1 to 3, but some well-differentiated NETs have a distinctive feature related to Ki67 index marker counts higher than 20% or a mitotic count higher than 20 per 2 mm<sup>2</sup> [3]. In 2010, the World Health Organization (WHO) created a three-tier grading scheme and pointed out that the grade 3 group can indeed be heterogeneous, comprising both poorly differentiated and well-differentiated forms [4]. Recently, a new guideline proposed a change in the grade 1–2 cutoff, which was implemented (grade 1: Ki67 ≤ 3%). In fact, the NET precocious diagnostic is essential to increasing patient survival [5]. NETs may appear in any organ of the body, but they are more frequently

the source of metastasis in the small intestine followed by the pancreas, which causes metastases in the lungs and colon [6].

The incidence of NETs has increased in different countries over the years, probably because of advances in diagnostic techniques, such as endoscopy and computerized tomography, higher awareness among clinicians, and an aging population [7]. Several studies over the years have shown an increase in the prevalence and incidence of NETs [7,8], but the majority were conducted in the United States and Europe [8], so little is known in undeveloped countries about NET incidence and prevalence or which organs are the most affected. The aim of this review was to analyze which countries contributed the best scientific evidence about NETs. In addition, we investigated which related topics were the most studied between 1981 and 2020.

In this work, we used a Latent Dirichlet Allocation (LDA) model to identify and interpret NTE topics and research trends in our set of documents [9]. Thus, from this collection, called the “corpus,” the Bayesian-based LDA generates a probabilistic extension of the latent semantic analysis [9,10]. In addition, it assumes an a priori sparse Dirichlet distribution on topics in the documents using Gibbs sampling [11], which determines a topic’s probability from documents that combine various topics in different proportions. In contrast, the HJ-biplot method [12] was used to obtain a more precise data evaluation by revealing relationships between the analyzed data.

## 2. Materials and Methods

First, the basic concepts of the Latent Dirichlet Allocation (LDA) are introduced. Then, the procedure is used to identify latent topics and research trends in neuroendocrine tumors. Next, the quantitative indices are defined so the results can be explained. Finally, the HJ-biplot provides a graphic multivariate representation of rows and columns of a data matrix in a low-dimensional subspace where the results’ relative positions are interpretable.

### 2.1. Latent Dirichlet Allocation (LDA)

The LDA is a method of unsupervised text mining in which themes or topics of certain documents can be identified from a larger collection, the corpus. It is based on Bayesian models and is considered to be an extension of probabilistic latent semantic analysis [10,13]. The LDA adds a sparse Dirichlet prior to distribution of the items to a document, using a Gibbs sampler [11] to assign probabilities to topics for each term. The documents are then grouped into the topics to which they belong, assuming that the documents exhibit multiple mixtures of subjects in different proportions. The LDA allows the inference or estimation of latent variables; that is, it calculates their conditional distribution in documents. Equation (1) shows the statistical assumptions behind the generative process of the LDA.

$$p(\beta_K, \theta_D, z_D, w_D) = \prod_{k=1}^K p(\beta_k|\eta) \prod_{m=1}^M p(\theta_m|\alpha) \prod_{n=1}^N p(z_{m,n}|\theta_m) P(w_{m,n}|z_{m,n}, \beta_{m,k}) \quad (1)$$

Here,  $K$ ,  $m$ , and  $n$  denote, respectively, the number of topics, the number of articles, and the number of words in a given document;  $\alpha$  and  $\eta$  are Dirichlet hyperparameters of the prior distributions over  $\theta$  and  $\beta$ , respectively;  $\theta_m$  is the distribution of topics for article  $m$  (real vector of length  $K$ );  $z_{m,n}$  is the topic for the  $n$ th word in the  $m$ th article;  $w_{m,n}$  is the  $n$ th word of the  $m$ th document; and  $\beta_k$  is the distribution of words for topic  $k$ . We had to focus on the only observable variable (i.e., the words within the documents) to infer the hidden structure with the statistical inference methods. The conditional probability, also known as the posteriori probability, is expressed by Equation (2).

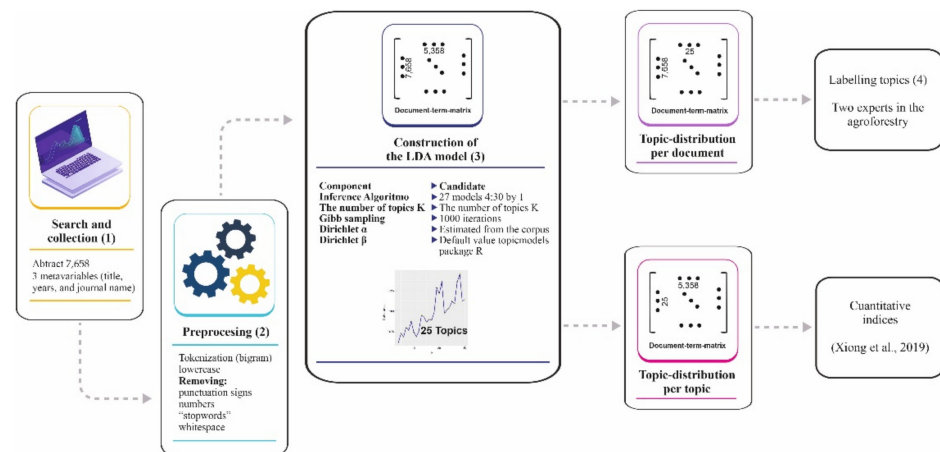
$$p(\beta_K, \theta_M, z_M | w_M) = \frac{p(\beta_K, \theta_M, z_M, w_M)}{p(w_M)} \quad (2)$$

Although the posterior probability cannot be exactly computed due to the term in the denominator [13], an approximation to its true value can be achieved with statistical

posterior inference methods. Two main types of inference techniques can be discerned: variation-based [13] and sampling-based algorithms [14]. An example of the latter is the Gibbs sampler [9]. Both algorithms provide similarly accurate results [15].

## 2.2. Identifying Research Topics

To identify the research topics in NETs using the LDA, the study was divided in four stages: Literature Search, Preprocessing, Selection of the number of LDA topics and model construction, and Labeling the topics (Figure 1).



**Figure 1.** Scheme of the methodological process used in the procedure for the identification of topics through Latent Dirichlet Allocation.

### 2.2.1. Literature Search

The PubMed database was used as source for this paper. The inclusion criteria focused on the selection of research articles, published in English, from 1981 to December 2020. The search was conducted on 10 July 2021 using the following query:

“neuroendocrine tumors [Title/Abstract] AND Journal Article [PT] AND 1981:2020[DP]”

Four variables were used: author, title, year of publication, and author affiliation. Books, reviews, book chapters, gray literature, and reports were not included to avoid noise in the results. After executing the search query, the preliminary database returned 8216 documents, which were filtered to remove those that were repeated, misclassified, or contained no abstract. Based on the inclusion criteria, 7658 articles published between 1981 and 2020 were included.

### 2.2.2. Preprocessing

After reviewing the full texts of the eligible studies, LDA was used—a “bag of words” model in which documents are represented as a sequence of individual words. Each document was tokenized, a necessary process for obtaining individual words (also known as unigrams) from sentences. Lowercase text, punctuation marks, hyphens, square brackets, blank spaces, and other characters were eliminated. In addition, a standard list of words called “stop words” was also removed to ensure that each sentence was grammatically correct. As a result of this preprocessing, a matrix of document terms was obtained where each article was represented in a  $V$ -dimensional vector of a collection of words. Data processing in this part of the study was carried out using LDAShiny [16], an open-source package for R programming language (R Development Core Team 2019), which contains a tool that provides a web-based graphical user interface to perform a review of the scientific literature under the Bayesian approach of Latent Dirichlet Allocation (LDA) and machine learning algorithms.

### 2.2.3. Selection of the Number of LDA Topics and Construction of the Model

Topics are latent variables that use correlations between words and latent semantic themes in a collection of documents [17]. The definition assumes that the expected number of topics  $k$  (i.e., latent variables) must be established a priori. Simulations were carried out varying  $k$  from 4 to 30 in incremental steps of one, and an inference algorithm with 500 iterations was used, namely Gibbs sampling [11].

A quality LDA model was determined by using a topic coherence measure [18], which is a measure of a topic model from the perspective of human interpretability and is considered a more adequate measure than a computational metric such as perplexity [19].

### 2.2.4. Labeling of Topics

A naive labeling algorithm based on probable bigrams provided by the package `textmineR` was used [20]. However, given that algorithms have a very limited ability to understand the latent meaning of human language [21], it was also decided to use manual labelling, which is considered standard in topic modeling, from three sources of information: the 20 most frequent word lists (most likely), a sample of the titles, and the three most loaded articles.

In addition, to improve the labeling, we visualized the topics in a two-dimensional area by computing the distance between them [22]. This method showed the area representing the prevalence of the topic and indicated how widespread a topic was within all documents. This analysis displayed the similarity between topics with respect to their probability distribution over words [23].

### 2.3. HJ-Biplot

The HJ-biplot [12], an extension of the classical biplots introduced by Gabriel [24], is an exploratory data analysis method that looks for hidden patterns in a data matrix, and then graphically represents the information contained in the rows (years) and columns (topics) [12]. This multivariate statistical technique was chosen since it offers a more precise data evaluation in which the relationships between the parts, years, and topics are highlighted. For this analysis, `Multibplot` software [25] was used because it provides a fast and easy way to incorporate our tables from an Excel format.

The data representation consisted of visualizing a matrix of multivariate data  $X_{n \times p}$  using vectors as points called markers  $g_1, g_2, \dots, g_n$  for each row, and vectors called markers  $h_1, h_2, \dots, h_p$  for each column. Each row represented a subject, and each column a variable, such that both marker sets could be superimposed onto the same reference system with the maximum quality of representation. If the rows of matrix  $A$  were described as markers  $g_1, g_2, \dots, g_n$  and matrix  $B$  as markers  $h_1, h_2, \dots, h_p$ , the result was  $X = AB^T$ .

The markers were obtained from the usual singular value decomposition (SVD) of the data matrix. The SVD of matrix  $X$  was defined by  $X = UDV^T$ , where  $U$  is the matrix where columns are the eigenvectors of  $XX^T$ ,  $V$  is the matrix where columns are the eigenvectors of  $X^T X$ , and  $D$  is the matrix diagonal of the singular values  $\lambda_i$  of  $X$ . Let  $A$  and  $B$  be the matrices of the first two columns of  $UD$  and  $VD$ , respectively (Figure 2).

### 2.4. Quantitative Indices Used to Analyze the Trend of Topic

The number of documents and words found were quite large, so it was not possible to understand the topic trend intuitively. Therefore, we used some quantitative indices [26] to analyze the main outputs of the topic modeling algorithm: the collection of terms with associated frequencies that characterized a topic, and the percentage composition for each document. The topic distribution over time  $(\theta_k^y)$  was:

$$(\theta_k^y) = \frac{\sum_{m \in y} \theta_{mk}}{n^y} \quad (3)$$

where  $m \in y$  represents articles published in a given year,  $\theta_{mk}$  is the proportion the  $k$ th topic in each article, and  $n^y$  is the total number of articles published in the year.

To facilitate the characterization of the topics by their tendency, simple regression slopes for each topic were used, in which the year was a dependent variable and the proportion of the topics in the corresponding year was the response variable [9]. The topics obtained by regression were positive or negative at a statistical significance level of 0.01 and classified as positive or negative trends, respectively (Equation (3)).

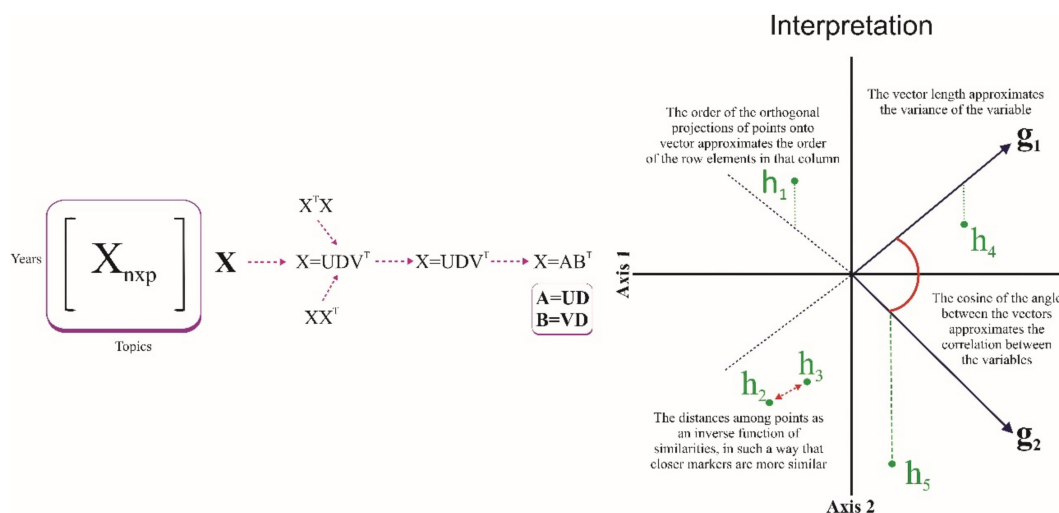


Figure 2. Description of the HJ-Biplot method.

### 3. Results

Reference lists of the articles were reviewed to identify additional studies. The detailed search strategy is presented in Table 1.

Table 1. Electronic search strategy for PubMed.

Database	Search	Filters/Limits
PubMed	Tumors neuroendocrine	Humans English

The full search strategy for each database (Mesh: Medical Subject Headings) and the results are available from the authors.

The initial search yielded 8216 articles. After applying the inclusion and exclusion criteria and searching the reference lists of included studies, we identified a total of 7658 studies (Figure 3).

Figure 4 shows the countries that published the majority of the scientific research on NETs: the United States (2271), Germany (246), the United Kingdom (237), France (216), Italy (177), Japan (166), Spain (156), The Netherlands (151), Canada (150), and finally, China (131). Other countries did not represent the same frequency of publication or have the same impact.

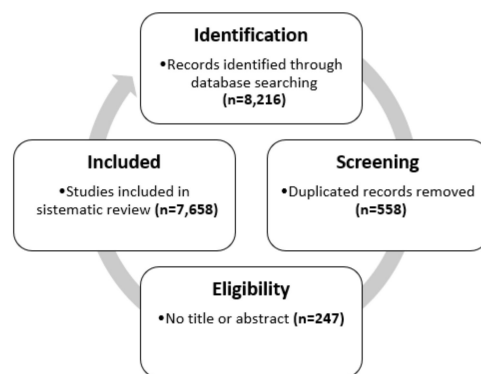


Figure 3. Gathering, filtering, and selection the articles for studies on neuroendocrine tumors.

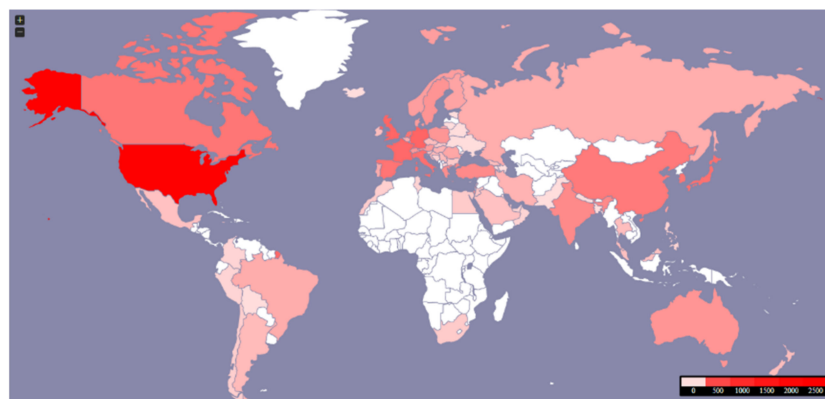


Figure 4. Worldwide distribution of 7658 articles published on NETs between 1981–2020.

The coherence scores for all LDA models are shown in Figure 5. The results suggested that the LDA model with the optimum coherence score contained 25 topics ( $k = 25$ ). Distribution by document  $\theta_m$  was added to calculate average probability  $\theta_k^y$ , all the articles published in a given year, and to identify trends (Figure 6). We found that the probability of some topics increased gradually over time (red): t\_2 (cost and effect on quality of life), t\_3 (Markers for carcinomas), t\_6 (GEP-NETs), t\_10 (solid-pseudopapillary tumor of the pancreas), t\_11 (Pancreatic neuroendocrine tumors), t\_13 (Patients one year after treatment), t\_15 (Radionuclide therapy), t\_17 (Prognosis of survival before and after resection), t\_20 (pheochromocytoma and paraganglioma), t\_21 (Clinical benefit), t\_22 (AJCC staging system), and t\_24 (long term NF-pNETs).

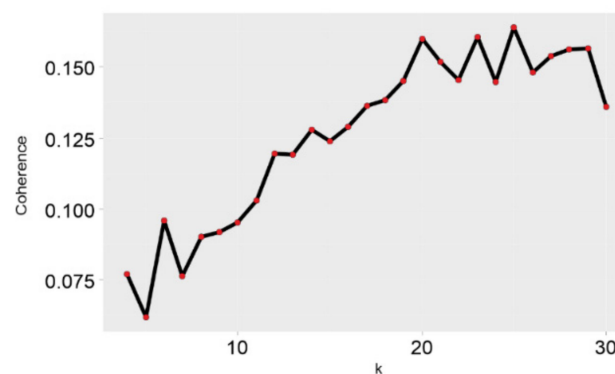


Figure 5. Coherence.



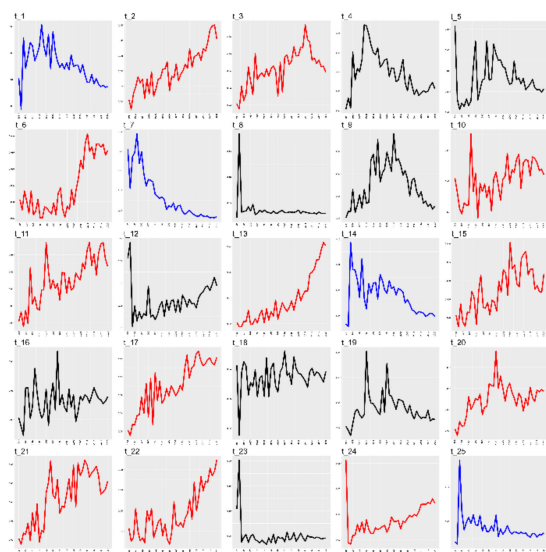


Figure 6. Trend of topics.

The topics that showed decreasing probability (blue) were t<sub>1</sub> (gene expression in cell line normal tissue), t<sub>7</sub> (Merkel cell carcinoma), t<sub>14</sub> (Large-cell, small-cell, carcinoma), and t<sub>25</sub> (chromogranin A levels); topics where there was no observed trend (black) were t<sub>4</sub> (Ectopic ACTH syndrome), t<sub>5</sub> (liver transplantation for hepatic metastases), t<sub>8</sub> (primary tumor well-differentiated), t<sub>9</sub> (somatostatin receptor subtypes), t<sub>12</sub> (lymph node metastasis), t<sub>16</sub> (Fine needle aspiration), t<sub>18</sub> (grade neoplasm), t<sub>19</sub> (multiple endocrine neoplasia), and t<sub>23</sub> (Bile duct: a case report).

The heat map (Figure 7) shows how the topics were distributed by years. Each pixel represents the probability that a topic was found in a specific year. A cluster analysis was performed on the years and topics to calculate the Euclidean distance between each pair of topics. The dendrogram represents a greater distance between the years, which means that they differed significantly from each other. Smaller distances implied that the years were similar. The cluster analysis showed built hierarchies, so it was possible to divide the components into several groups. Our study found five groups, through bootstrap analysis, that had higher probability proportions for certain years: Topic 7 (Merkel cell carcinoma) of group 5 (1981–1992); topics 23 (bile duct: a case report) and 8 (primary tumor well-differentiated) of group 4 (1982); and topic 25 (chromogranin A levels) of group 3 (1983).

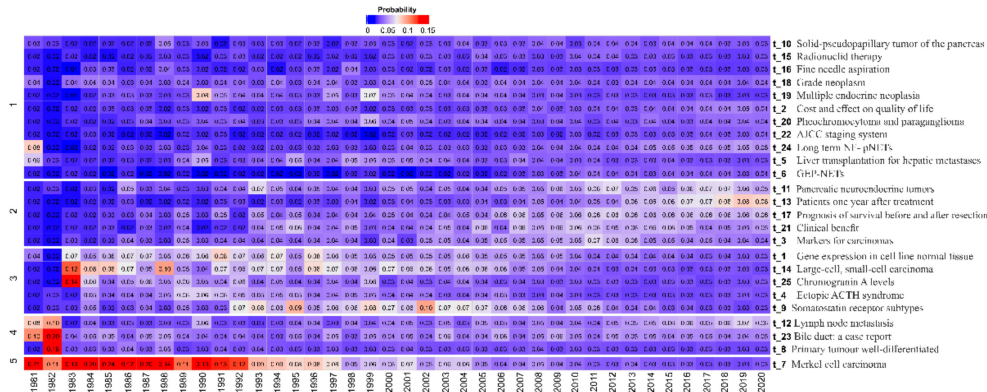


Figure 7. Representation of the distribution of topics by year through the heat map.

Table 2 shows the topic names generated from the words with the highest number of repetitions and ranked for relevance. After searches within the articles according to each topic, 5 words with the highest number of repetitions generated the prevalence rankings.



The 5 main ones were: t\_21 (Clinical benefit), t\_11 (Pancreatic neuroendocrine tumors), t\_13 (Patients one year after treatment), t\_17 (Prognosis of survival before and after resection), and t\_3 (Markers for carcinomas).

Our study also carried out an analysis using the LDA method where relationships between topics were observed. The existence of a possible relationship was obtained by the statistical analysis of pancreatic cancer according to the matrix generated in the LDA model (Figure 8).

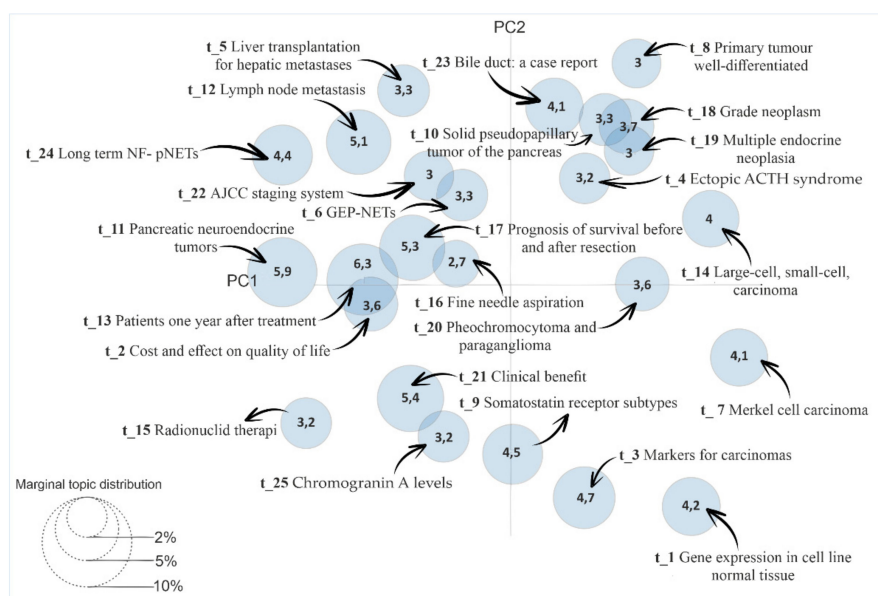


Figure 8. Identification of relationships among topics from the LDAvis Method.

The grouping of the topics and years was also analyzed using multivariate HJ-Biplot analysis with the theta matrix from the results. The matrix was composed of values between 0 and 1 according to the topics and years under study. We found coherent groupings as shown in Figure 9, with 79.9% data variability.

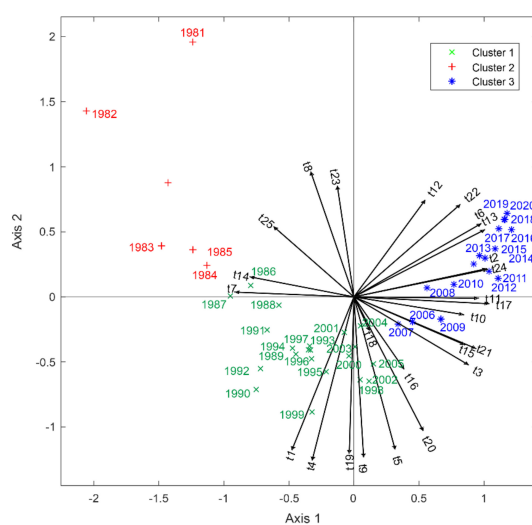


Figure 9. Association between topics by year using the HJ-Biplot method. A small acute angle means a positive relation; a plane angle (Formed by two right angles 180°), negative relation; and a straight angle (Formed by two perpendicular lines 90°), no relation. The segments (lines) represent each of the variables (topics); the colors red, blue, and green represent groups of years with respect to the topics.

**Table 2.** Frequent topics, where: P = Prevalence; PR = Prevalence Ranking; ART. = Article number.

TOPIC	TOPIC NAMES	P	PR	ART.	TOPICS TERMS				
t_21	Clinical benefit	5.4	1	630	net	therapi	tumor	treatment	clinic
t_11	Pancreatic neuroendocrine tumors	5.9	2	561	pancreat	tumor	neuroendocrin	present	report
t_13	Patients one year after treatment	6.3	3	552	patient	year	ag	median	diagnosi
t_17	Prognosis of survival before and after resection	5.3	4	501	surviv	patient	prognost	resection	analysi
t_3	Markers for carcinomas	5.1	5	483	tumor	cell	neuroendocrin	ne	carcinoid
t_23	Bile duct: a case report	4.1	6	458	case	report	rare	present	diagnosi
t_9	Somatostatin receptor subtypes	4.5	7	439	receptor	somatostatin	tumor	somatostatin_receptor	analog
t_12	Lymph node metastasis	4.7	8	432	resect	surgic	node	recurr	surgeri
t_7	Merkel cell carcinoma	4.1	9	349	cell	tumor	posit	marker	stain
t_14	Large-cell, small-cell, carcinoma	4.0	10	340	carcinoma	cell	carcinoid	lung	small
t_20	Pheochromocytoma and paraganglioma	3.6	11	315	gene	mutat	pheochromocytoma	tumor	paraganglioma
t_1	Gene expression in cell line normal tissue	4.2	12	310	express	cell	tissu	protein	tumor
t_5	Liver transplantation for hepatic metastases	3.3	13	281	liver	metastas	hepat	metastat	primari
t_25	Chromogranin A levels	3.2	14	240	level	cga	serum	chromogranin	measur
t_19	Multiple endocrine neoplasia	3.0	15	235	men	endocrin	type	tumor	multipl
t_4	Ectopic ACTH syndrome	3.2	16	224	carcinoid	symptom	hormon	tumor	syndrom
t_10	Solid pseudopapillary tumor of the pancreas	3.3	17	219	pancreat	pancreat_tumor	pancrea	tumor	neoplasm
t_15	Radionuclide therapy	3.2	18	197	therapi	dose	prrt	lu	radionuclid
t_24	Long term NF- pNETs	4.4	19	170	pnet	year	group	rang	term
t_16	Fine needle aspiration	2.7	20	152	diagnosi	eu	lesion	fna	biopsi
t_2	Cost and effect on quality of life	3.6	21	140	clinic	data	base	manag	develop
t_18	Grade neoplasm	3.7	22	130	tumor	featur	malign	grade	neoplasm
t_22	AJCC staging system	3.0	23	125	stage	cancer	small	incid	system
t_6	GEP-NETs	3.3	24	89	net	tumor_net	gep	gep_net	Gastroentero
t_8	Primary tumor well-differentiated	3.0	25	86	differenti	tumor	well	primari	pancreat site

The topics that had greater relevance for 1981–1985 were t\_23 (Bile duct: a case report), t\_8 (Primary tumor well-differentiated), t\_25 (Chromogranin A levels), t\_14 (Large-cell, small-cell, carcinoma), and t\_7 (Merkel cell carcinoma).

For the years 1986–2005, the topics were t\_1 (Gene expression in cell line normal tissue), t\_4 (Ectopic ACTH syndrome), t\_19 (Multiple endocrine neoplasia), t\_9 (Somatostatin receptor subtypes), t\_5 (Liver transplantation for hepatic metastases), t\_20 (Pheochromocytoma and paraganglioma), and t\_16 (Fine needle aspiration).

Finally, for the years 2006–2020, the most relevant topics were t\_12 (Lymph node metastasis), t\_22 (AJCC staging system), t\_6 (GEP-NETs), t\_13 (Patients one year after treatment), t\_2 (Cost and effect on quality of life), t\_24 (Long term NF-NETs), t\_11 (Pancreatic neuroendocrine tumors), t\_17 (Prognosis of survival before and after resection), t\_10 (Solid-pseudopapillary tumor of the pancreas), t\_21 (Clinical benefit), t\_15 (Radionuclide therapy), and t\_3 (Markers for carcinomas).

#### 4. Discussion

The presence of NETs is relatively rare, with fewer than 10 cases per 100,000 inhabitants annually [27]. However, the data prevalence of NETs has increased in recent years due to the development and wide application of modern imaging and endoscopic technology [28]. In the U.S., the 20-year limited-duration prevalence increased dramatically from 0.006% in 1993 to 0.048% in 2012, with NETs more frequently observed in the rectum, followed by the lungs and small intestine [7]. According to the SEER database in the U.S., it went from 3.9 in 1995 to 6.61 in 2012 [7]. Because this topic is particularly important, and because there is a lack of data on NETs from undeveloped countries [8], this work investigated countries that produced the most scientific studies on NETs between 1981 and 2020: the United States, Germany, the United Kingdom, France, and Italy. To the best of our knowledge, this is the first study to assess which countries are the most frequent publishers of scientific works about NETs, as well as the most frequent publishers of related topics.

Because most NETs are misdiagnosed, proper diagnosis and treatment after the first symptoms takes years [29]. In fact, our statistics showed that from 1981 to 2020, the most frequently published topic was “markers for carcinomas.” Classical neuroendocrine markers include the expression of synaptophysin, considered more sensitive, and chromogranin A, considered more specific [30]. Serum chromogranin A also may be used as an immunohistochemical biomarker to assess NETs in addition to serving as a treatment monitor [31].

NETs may be classified into two subtypes due to clinical and genetic differences. The well-differentiated form is defined as a NET, and the poorly differentiated form is defined as a neuroendocrine carcinoma (NEC) [1]. Following the WHO, the main factor when characterizing NETs is proliferation fraction grading, measured by either mitotic count or by the Ki67-positive percentage [32]. Nevertheless, it seems a great difficulty exists in diagnosing NETs due to the heterogeneous nature of tumors and different patient symptoms [7,29]. Indeed, this fact explains our results, at least in part, because the most frequently studied topic for NETs was “markers for carcinomas.”

In line with these facts, our work showed that the second- and third-most frequently studied topics were “clinical benefit” and “carcinoma treatment, tumor response, and survival,” respectively, because physicians look for patients’ increasing survival and improved quality of life [33]. Currently, NET treatments include therapies in isolation or in combination with others [34]. Common therapies are the multi-targeted receptor tyrosine kinase inhibitor Sunitinib [35], the radiolabeled somatostatin analogue lutetium-177 (177Lu)-dotatate [36], the mechanistic target of rapamycin inhibitor everolimus [36], and the vascular endothelial growth factor antibody bevacizumab [37]. In response to pharmacological advances, new diagnostic techniques, and early-stage diagnosis, the survival of patients with NETs has improved over time [7]. Those with distant gastrointestinal or pancreatic NETs reported higher indices of survival [7]. Unsurprisingly, our study found pancreatic NETs to be the most widely studied between 1981 and 2020.

A pancreatic NET is a rare malignancy with relatively non-painful biologic behavior compared with a pancreatic adenocarcinoma [38]. A median survival for patients with advanced pancreatic NET of around 27 months was reported [39]. However, the prevalence of pancreatic tumors represents 10% of NETs and only 1% of all cancer cases [39–41]. Approximately 64% of patients with pancreatic NETs present with metastatic disease and are diagnosed at the advanced stage; consequently, they have a poor prognosis [39]. Pancreatic NET commonly affects people between 40 and 69, but there is a significant number of patients younger than 35 that are diagnosed [40,42]. An estimated 40–91% of pancreatic NETs are nonfunctional, having no clinically evident hormonal symptoms [2,43]. Thus, therapeutic management of pancreatic NETs depends on the degree to which the tumor is well- or poorly differentiated. Symptoms caused by hypersecretion of hormones and the disease diagnosis stage must be considered [44]. Based on an analysis of the SEER database from 1973 to 2000, the annual incidence of pancreatic NETs was 1.8 and 2.6 per million for women and men, respectively, [2,43]. Given the severity and prevalence of this tumor, our results showed the highest number of pancreatic NET studies compared with those for other organs.

Furthermore, we found “prognosis of survival before and after resection” and “patients one year after treatment” to be the fourth- and fifth-most frequent topics. In clinical practice, common, low-grade tumors are managed with surgical resection [45]; in the case of pancreatic NETs, American and European guidelines recommend resection for NETs > 2 cm [46,47]. However, there is controversy over resecting pancreatic NETs ≤ 2 cm. Some physicians prefer observation and surveillance [48–50], even though several lines of evidence recommend resection based on the potential for malignant differentiation, lymph node metastasis, and distant metastasis, even among patients with small, non-functional pancreatic NETs [51–54]. A study of 487 patients showed that the occurrence of pancreatic NET is rare and based on tumor size, nodal metastasis, grading, and vascular invasion. Thus, patients with a grade 1 pancreatic NET without nodal metastasis and vascular invasion may be cured by surgery [55].

Increasing prognosis of survival, preventing tumor recurrence, and promoting a better quality of life of patients are the basic objectives of NET research, and the more frequently studied topics in our review reflect these concerns. The guidelines are specific to each NE-affected organ. For example, in lung NETs, a large cell carcinoma represents approximately 3% of all cancers [56], and surgical procedures are not effective since most patients die from recurrence [57]. The prognosis for patients increases if surgery is performed when the tumor is <3 cm [58–60]. Furthermore, the National Comprehensive Cancer Network (NCCN) guidelines for large-cell neuroendocrine stage-specific carcinoma and non-small-cell lung carcinoma highlight the need for adjuvant chemotherapy or chemoradiotherapy after resection in some cases, including those where the tumor is >4 cm. Morphologically identical to small cell carcinoma of the lung, breast NETs have poorly differentiated cells [61] and represent less than 1% of breast cancers [62]. Endocrine therapy and radiotherapy are able to increase survival [63,64], but no consensus has been reached on prognosis; most studies suggest a poor outcome [61].

On other hand, in mid-gut NETs, the guideline proposes primary tumor resection alone in the setting of unrespectable metastatic disease because surgery can completely remove the tumor [49]. Nevertheless, studies of patients with gastrointestinal NETs have shown increased survival after resection in the setting of metastatic disease [65]. In the case of neuroendocrine liver metastases, resection represents the major potential for a cure [66–68]. Unfortunately, the curative role of surgery seems to be only feasible in 10–25% of patients after resection in liver metastases [69]. The recurrence of a NET represents the major cause of death in most patients after resection of metastatic tumors [67]. In our review, the topics “prognosis of survival before and after resection” and “patients one year after treatment” are among the five most-studied topics due to their importance for survival and quality of life patients with NETs. Indeed, data on overall NET survival are the most frequently reported in the literature [7]. It is worth mentioning that data on

recurrence may inform patients about the probability of treatment success and the risk of recurrence following a surgical intervention [70].

## 5. Conclusions

In summary, our review found that the countries with the most research about NETs were the United States, Germany, the United Kingdom, France, and Italy, and statistical analyses of pancreatic NETs were more prevalent between 1981 and 2020. Furthermore, our results confirmed that “clinical benefits,” “patients one year after treatment,” “prognosis of survival before and after resection,” and “markers for carcinomas” were the most frequent topics in the scientific literature during the last 39 years.

Finally, the LDA method employed in this review grouped each subject in a category based on the high probability subjective observations of words. Thus, the methods showed effectiveness in generating responses about the more common topics studied in the NETs. In contrast, the HJ-Biplot method was integral for grouping topics by year and finding which NETs were the most relevant and for which years.

**Author Contributions:** Conceptualization, K.M.E.; methodology, K.M.E.; software, K.M.E. and J.d.l.H.-M.; validation, K.M.E. and D.F.A.C.; formal analysis, K.M.E.; investigation, K.M.E.; resources, K.M.E.; data curation, K.M.E. and J.d.l.H.-M.; writing original draft preparation, A.S.; writing review and editing, A.S.; visualization, K.M.E. and L.M.U.-C.; supervision, J.L.V.-V.; project administration, K.M.E. All authors have read and agreed to the published version of the manuscript.

**Funding:** Not applicable.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** This study was supported by the Department of Mathematics and Statistics, Institute of Basic Sciences and Research Institute of the Technical University of Manabí.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Rindi, G.; Wiedenmann, B. Neuroendocrine neoplasia of the gastrointestinal tract revisited: Towards precision medicine. *Nat. Rev. Endocrinol.* **2020**, *16*, 590–607. [[CrossRef](#)]
2. Klimstra, D.S.; Modlin, I.R.; Coppola, D.; Lloyd, R.V.; Suster, S. The Pathologic Classification of Neuroendocrine Tumors: A review of nomenclature, grading, and staging systems. *Pancreas* **2010**, *39*, 707–712. [[CrossRef](#)]
3. Hijioka, S.; Hosoda, W.; Mizuno, N.; Hara, K.; Imaoka, H.; Bhatia, V.; Mekky, M.A.; Tajika, M.; Tanaka, T.; Ishihara, M.; et al. Does the WHO 2010 classification of pancreatic neuroendocrine neoplasms accurately characterize pancreatic neuroendocrine carcinomas? *J. Gastroenterol.* **2015**, *50*, 564–572. [[CrossRef](#)] [[PubMed](#)]
4. Vanoli, A.; Albarello, L.; Uncini, S.; Fassan, M.; Grillo, F.; Di Sabatino, A.; Martino, M.; Pasquali, C.; Milanetto, A.C.; Falconi, M.; et al. Neuroendocrine Tumors (NETs) of the Minor Papilla/Ampulla: Analysis of 16 Cases Underlines Homology With Major Ampulla NETs and Differences From Extra-Ampullary Duodenal NETs. *Am. J. Surg. Pathol.* **2019**, *43*, 725–736. [[CrossRef](#)]
5. Pavel, M.E.; Singh, S.; Strosberg, J.R.; Bubuteishvili-Pacaud, L.; Degtyarev, E.; Neary, M.P.; Carnaghi, C.; Tomasek, J.; Wolin, E.; Raderer, M.; et al. Health-related quality of life for everolimus versus placebo in patients with advanced, non-functional, well-differentiated gastrointestinal or lung neuroendocrine tumours (RADIANT-4): A multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **2017**, *18*, 1411–1422. [[CrossRef](#)] [[PubMed](#)]
6. Riihimäki, M.; Hemminki, A.; Sundquist, K.; Sundquist, J.; Hemminki, K. The epidemiology of metastases in neuroendocrine tumors. *Int. J. Cancer* **2016**, *139*, 2679–2686. [[CrossRef](#)]
7. Dasari, A.; Shen, C.; Halperin, D.M.; Zhao, B.; Zhou, S.; Xu, Y.; Shih, T.; Yao, J.C. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients with Neuroendocrine Tumors in the United States. *JAMA Oncol.* **2017**, *3*, 1335–1342. [[CrossRef](#)] [[PubMed](#)]
8. Rindi, G.; Klöppel, G.; Couvelard, A.; Komminoth, P.; Körner, M.; Lopes, J.M.; McNicol, A.-M.; Nilsson, O.; Perren, A.; Scarpa, A.; et al. TNM staging of midgut and hindgut (neuro) endocrine tumors: A consensus proposal including a grading system. *Virchows Arch.* **2007**, *451*, 757–762. [[CrossRef](#)]
9. Griffiths, T.L.; Steyvers, M. Finding scientific topics. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5228–5235. [[CrossRef](#)] [[PubMed](#)]
10. Hornik, K.; Grün, B. Topicmodels: An R package for fitting topic models. *J. Stat. Softw.* **2011**, *40*, 1–30.
11. Geman, S.; Geman, D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *6*, 721–741. [[CrossRef](#)]



12. Galindo, M.P.; Cuadras, C.M. *Una Extensión del Método Biplot y su Relación Con otras Técnicas*; Publicaciones de Bioestadística y Biomatemática, Universidad de Barcelona no: Barcelona, Spain, 1986.
13. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
14. Porteous, I.; Newman, D.; Ihler, A.; Asuncion, A.; Smyth, P.; Welling, M. Fast collapsed gibbs sampling for latent dirichlet allocation. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD 08*; ACM Press: New York, NY, USA, 2008; pp. 569–577.
15. Asuncion, A.; Welling, M.; Smyth, P.; Teh, Y.W. On smoothing and inference for topic models. *arXiv* **2012**, arXiv:1205.2662.
16. De la Hoz-M, J.; Fernández-Gómez, M.; Mendes, S. LDAShiny: An R Package for Exploratory Review of Scientific Literature Based on a Bayesian Probabilistic Model and Machine Learning Tools. *Mathematics* **2021**, *9*, 1671. [[CrossRef](#)]
17. Blei, D.M.; Lafferty, J.D. A correlated topic model of Science. *Ann. Appl. Stat.* **2007**, *1*. [[CrossRef](#)]
18. Röder, M.; Both, A.; Hinneburg, A. Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*; ACM Press: New York, NY, USA, 2015; pp. 399–408.
19. Chang, J.; Blei, D.M. Relational topic models for document networks. In *Proceedings of the Artificial Intelligence and Statistics, Clearwater Beach, FL, USA, 16–18 April 2009*; Volume 5, pp. 81–88.
20. Jones, T. *textmineR: Functions for Text Mining and Topic Modeling, R Package Version 3.0. 4*; R Foundation for Statistical Computing: Vienna, Austria, 2019; Volume 3, Available online: <https://cran.r-project.org/web/packages/textmineR/index.html> (accessed on 1 January 2021).
21. Lau, J.H.; Grieser, K.; Newman, D.; Baldwin, T. Automatic labelling of topic models. In *Proceedings of the ACL-HLT 2011-the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011*; pp. 1536–1545.
22. Sievert, C.; Shirley, K. LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Baltimore, MD, USA, 27 June 2014*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014; pp. 63–70.
23. Chuang, J.; Ramage, D.; Manning, C.D.; Heer, J. Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012*; pp. 443–452.
24. Gabriel, K.R. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **1971**, *58*, 453–467. [[CrossRef](#)]
25. Vicente-Villardón, J.L. *MULTBILOT: A Package for Multivariate Analysis Using Biplots*; Departamento de Estadística. Universidad de Salamanca: Salamanca, Spain, 2010.
26. Xiong, H.; Cheng, Y.; Zhao, W.; Liu, J. Analyzing scientific research topics in manufacturing field using a topic model. *Comput. Ind. Eng.* **2019**, *135*, 333–347. [[CrossRef](#)]
27. Novoa, M.P.D.M.; Capel, F.F.; Sedano, J.V.R.; Carrera, M.S.; Jiménez, V.A.; Pereira, P.O.; Valera, J.S.; Medina, L.O.; Pérez, J.; Ángel, D. Tumores neuroendocrinos gastroenteropancreáticos: Características clínicas, proceso diagnóstico y pronóstico en el Hospital Universitario Clínico San Carlos (Madrid). *Endocrinología y Nutrición* **2014**, *61*, 234–241. [[CrossRef](#)]
28. Jiao, X.; Li, Y.; Wang, H.; Liu, S.; Zhang, N.; Zhou, Y. Clinicopathological features and survival analysis of gastroenteropancreatic neuroendocrine neoplasms: A retrospective study in a single center of China. *Chin. J. Cancer Res.* **2015**, *27*, 258–266. [[CrossRef](#)]
29. Chauhan, A.; Kohn, E.; Del Rivero, J. Neuroendocrine Tumors-Less Well Known, Often Misunderstood, and Rapidly Growing in Incidence. *JAMA Oncol.* **2020**, *6*, 21. [[CrossRef](#)] [[PubMed](#)]
30. Bellizzi, A.M. Immunohistochemistry in the diagnosis and classification of neuroendocrine neoplasms: What can brown do for you? *Hum. Pathol.* **2020**, *96*, 8–33. [[CrossRef](#)]
31. Ahmed, M. Gastrointestinal neuroendocrine tumors in 2020. *World J. Gastrointest. Oncol.* **2020**, *12*, 791–807. [[CrossRef](#)]
32. Lloyd, R.V.; Osamura, Y.R.; Kloppel, G.; Rosai, J. *WHO Classification of Tumours of Endocrine Organs*; World Health Organization: Geneva, Switzerland, 2017.
33. Pearman, T.P.; Beaumont, J.L.; Cella, D.; Neary, M.P.; Yao, J. Health-related quality of life in patients with neuroendocrine tumors: An investigation of treatment type, disease status, and symptom burden. *Support. Care Cancer* **2016**, *24*, 3695–3703. [[CrossRef](#)]
34. Pavel, M.E.; Hainsworth, J.D.; Baudin, E.; Peeters, M.; Hörsch, D.; Winkler, R.E.; Klimovsky, J.; Lebowitz, D.; Jehl, V.; Wolin, E.M.; et al. Everolimus plus octreotide long-acting repeatable for the treatment of advanced neuroendocrine tumours associated with carcinoid syndrome (RADIANT-2): A randomised, placebo-controlled, phase 3 study. *Lancet* **2011**, *378*, 2005–2012. [[CrossRef](#)]
35. Raymond, E.; Dahan, L.; Raoul, J.-L.; Bang, Y.-J.; Borbath, I.; Lombard-Bohas, C.; Valle, J.; Metrakos, P.; Smith, D.; Vinik, A.; et al. Sunitinib Malate for the Treatment of Pancreatic Neuroendocrine Tumors. *N. Engl. J. Med.* **2011**, *364*, 501–513. [[CrossRef](#)] [[PubMed](#)]
36. Yao, J.C.; Fazio, N.; Singh, S.; Buzzoni, R.; Carnaghi, C.; Wolin, E.; Tomasek, J.; Raderer, M.; Lahner, H.; Voi, M.; et al. Everolimus for the treatment of advanced, non-functional neuroendocrine tumours of the lung or gastrointestinal tract (RADIANT-4): A randomised, placebo-controlled, phase 3 study. *Lancet* **2016**, *387*, 968–977. [[CrossRef](#)]
37. Yao, J.C.; Guthrie, K.A.; Moran, C.; Strosberg, J.R.; Kulke, M.H.; Chan, J.A.; LoConte, N.; McWilliams, R.R.; Wolin, E.M.; Mattar, B.; et al. Phase III Prospective Randomized Comparison Trial of Depot Octreotide Plus Interferon Alfa-2b Versus Depot Octreotide Plus Bevacizumab in Patients With Advanced Carcinoid Tumors: SWOG S0518. *J. Clin. Oncol.* **2017**, *35*, 1695–1703. [[CrossRef](#)]

38. Luo, M.; Delaplane, S.; Jiang, A.; Reed, A.; He, Y.; Fishel, M.; Nyland, R.L.; Borch, R.F.; Qiao, X.; Georgiadis, M.M.; et al. Role of the Multifunctional DNA Repair and Redox Signaling Protein Ape1/Ref-1 in Cancer and Endothelial Cells: Small-Molecule Inhibition of the Redox Function of Ape1. *Antioxid. Redox. Signal.* **2008**, *10*, 1853–1867. [[CrossRef](#)]
39. Yao, J.C.; Hassan, M.M.; Phan, A.T.; Dagohoy, C.G.; Leary, C.C.; Mares, J.E.; Abdalla, E.K.; Fleming, J.B.; Vauthey, J.-N.; Rashid, A.; et al. One Hundred Years After “Carcinoid”: Epidemiology of and Prognostic Factors for Neuroendocrine Tumors in 35,825 Cases in the United States. *J. Clin. Oncol.* **2008**, *26*, 3063–3072. [[CrossRef](#)] [[PubMed](#)]
40. Yao, J.C.; Eisner, M.P.; Leary, C.; Dagohoy, C.; Phan, A.; Rashid, A.; Hassan, M.; Evans, D.B. Population-Based Study of Islet Cell Carcinoma. *Ann. Surg. Oncol.* **2007**, *14*, 3492–3500. [[CrossRef](#)] [[PubMed](#)]
41. Yao, J.C.; Lombard-Bohas, C.; Baudin, E.; Kvols, L.K.; Rougier, P.; Ruzsniwski, P.; Hoosen, S.; Peter, J.S.; Haas, T.; Lebwohl, D.; et al. Daily Oral Everolimus Activity in Patients with Metastatic Pancreatic Neuroendocrine Tumors After Failure of Cytotoxic Chemotherapy: A Phase II Trial. *J. Clin. Oncol.* **2010**, *28*, 69–76. [[CrossRef](#)]
42. Moore, F.D.; Scoinski, M.A.; Joste, N.E. Endocrine tumors and malignancies. In *Atlas of Diagnostic Oncology*, 3rd ed.; Elsevier Sci. Ltd.: Philadelphia, PA, USA, 2003.
43. Halfdanarson, T.R.; Rabe, K.G.; Rubin, J.; Petersen, G.M. Pancreatic neuroendocrine tumors (PNETs): Incidence, prognosis and recent trend toward improved survival. *Ann. Oncol.* **2008**, *19*, 1727–1733. [[CrossRef](#)] [[PubMed](#)]
44. Yao, J.; Phan, A.T. Optimising Therapeutic Options for Patients with Advanced Pancreatic Neuroendocrine Tumours. *Eur. Oncol. Haematol.* **2012**, *8*. [[CrossRef](#)]
45. Schurr, P.G.; Strate, T.; Rese, K.; Kaifi, J.T.; Reichelt, U.; Petri, S.; Kleinhans, H.; Yekebas, E.F.; Izbicki, J.R. Aggressive Surgery Improves Long-term Survival in Neuroendocrine Pancreatic Tumors: An institutional experience. *Ann. Surg.* **2007**, *245*, 273–281. [[CrossRef](#)]
46. Falconi, M.; Eriksson, B.; Kaltsas, G.; Bartsch, D.K.; Capdevila, J.; Caplin, M.; Kos-Kudla, B.; Kwekkeboom, D.; Rindi, G.; Klöppel, G.; et al. ENETS Consensus Guidelines Update for the Management of Patients with Functional Pancreatic Neuroendocrine Tumors and Non-Functional Pancreatic Neuroendocrine Tumors. *Neuroendocrinology* **2016**, *103*, 153–171. [[CrossRef](#)]
47. Kattan, M.W.; Hess, K.R.; Amin, M.B.; Lu, Y.; Moons, K.G.; Gershenwald, J.E.; Gimotty, P.A.; Guinney, J.H.; Halabi, S.; Lazar, A.J.; et al. American Joint Committee on Cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. *CA A Cancer J. Clin.* **2016**, *66*, 370–374. [[CrossRef](#)]
48. Lee, L.C.; Grant, C.S.; Salomao, D.R.; Fletcher, J.G.; Takahashi, N.; Fidler, J.L.; Levy, M.J.; Huebner, M. Small, nonfunctioning, asymptomatic pancreatic neuroendocrine tumors (PNETs): Role for nonoperative management. *Surgery* **2012**, *152*, 965–974. [[CrossRef](#)] [[PubMed](#)]
49. Sallinen, V.; Le Large, T.Y.; Galeev, S.; Kovalenko, Z.; Tiefrunk, E.; Araujo, R.; Ceyhan, G.O.; Gaujoux, S. Surveillance strategy for small asymptomatic non-functional pancreatic neuroendocrine tumors—a systematic review and meta-analysis. *HPB* **2017**, *19*, 310–320. [[CrossRef](#)]
50. Partelli, S.; Cirocchi, R.; Crippa, S.; Cardinali, L.; Fendrich, V.; Bartsch, D.K.; Falconi, M. Systematic review of active surveillance versus surgical management of asymptomatic small non-functioning pancreatic neuroendocrine neoplasms. *BJS* **2016**, *104*, 34–41. [[CrossRef](#)]
51. Dong, D.; Zhang, X.; Poultsides, G.; Rocha, F.; Weber, S.; Fields, R.; Idrees, K.; Cho, C.; Maithel, S.K.; Pawlik, T.M.; et al. Impact of tumor size and nodal status on recurrence of nonfunctional pancreatic neuroendocrine tumors  $\leq 2$  cm after curative resection: A multi-institutional study of 392 cases. *J. Surg. Oncol.* **2019**, *120*, 1071–1079. [[CrossRef](#)]
52. Haynes, A.; Deshpande, V.; Ingkakul, T.; Vagefi, P.A.; Szymonifka, J.; Thayer, S.P.; Ferrone, C.R.; Wargo, J.A.; Warshaw, A.L. Implications of Incidentally Discovered, Nonfunctioning Pancreatic Endocrine Tumors: Short-term and long-term patient outcomes. *Arch. Surg.* **2011**, *146*, 534–538. [[CrossRef](#)]
53. Gratian, L.; Pura, J.; Dinan, M.; Roman, S.; Reed, S.; Sosa, J.A. Impact of Extent of Surgery on Survival in Patients with Small Nonfunctional Pancreatic Neuroendocrine Tumors in the United States. *Ann. Surg. Oncol.* **2014**, *21*, 3515–3521. [[CrossRef](#)]
54. Sharpe, S.M.; In, H.; Winchester, D.J.; Talamonti, M.S.; Baker, M.S. Surgical Resection Provides an Overall Survival Benefit for Patients with Small Pancreatic Neuroendocrine Tumors. *J. Gastrointest. Surg.* **2015**, *19*, 117–123. [[CrossRef](#)] [[PubMed](#)]
55. Marchegiani, G.; Landoni, L.; Andrianello, S.; Masini, G.; Cingarlini, S.; D’Onofrio, M.; De Robertis, R.; Davì, M.; Capelli, P.; Manfrin, E.; et al. Patterns of Recurrence after Resection for Pancreatic Neuroendocrine Tumors: Who, When, and Where? *Neuroendocrinology* **2019**, *108*, 161–171. [[CrossRef](#)]
56. Fasano, M.; Della Corte, C.M.; Papaccio, F.; Ciardiello, F.; Morgillo, F. Pulmonary Large-Cell Neuroendocrine Carcinoma: From Epidemiology to Therapy. *J. Thorac. Oncol.* **2015**, *10*, 1133–1141. [[CrossRef](#)] [[PubMed](#)]
57. Fournel, L.; Falcoz, P.E.; Alifano, M.; Charpentier, M.-C.; Boudaya, M.-S.; Magdeleinat, P.; Damotte, D.; Régnard, J.-F. Surgical management of pulmonary large cell neuroendocrine carcinomas: A 10-year experience. *Eur. J. Cardiothorac. Surg.* **2013**, *43*, 111–114. [[CrossRef](#)] [[PubMed](#)]
58. Sarkaria, I.S.; Iyoda, A.; Roh, M.S.; Sica, G.; Kuk, D.; Sima, C.S.; Pietanza, M.C.; Park, B.J.; Travis, W.D.; Rusch, V. Neoadjuvant and Adjuvant Chemotherapy in Resected Pulmonary Large Cell Neuroendocrine Carcinomas: A Single Institution Experience. *Ann. Thorac. Surg.* **2011**, *92*, 1180–1187. [[CrossRef](#)] [[PubMed](#)]
59. Rossi, A.; Cavazza, A.; Marchioni, A.; Longo, L.; Migaldi, M.; Sartori, G.; Bigiani, N.; Schirosi, L.; Casali, C.; Morandi, U.; et al. Role of Chemotherapy and the Receptor Tyrosine Kinases KIT, PDGFR $\alpha$ , PDGFR $\beta$ , and Met in Large-Cell Neuroendocrine Carcinoma of the Lung. *J. Clin. Oncol.* **2005**, *23*, 8774–8785. [[CrossRef](#)] [[PubMed](#)]

60. Kujtan, L.; Muthukumar, V.; Kennedy, K.F.; Davis, J.R.; Masood, A.; Subramanian, J. The Role of Systemic Therapy in the Management of Stage I Large Cell Neuroendocrine Carcinoma of the Lung. *J. Thorac. Oncol.* **2018**, *13*, 707–714. [[CrossRef](#)] [[PubMed](#)]
61. Rosen, L.E.; Gattuso, P. Neuroendocrine Tumors of the Breast. *Arch. Pathol. Lab. Med.* **2017**, *141*, 1577–1581. [[CrossRef](#)]
62. Bussolati, G.; Badve, S. Carcinomas with neuroendocrine features. *WHO Classif. Tumours Breast* **2012**, *4*, 62–63.
63. Brask, J.B.; Talman, M.-L.M.; Wielenga, V.T. Neuroendocrine carcinoma of the breast—a pilot study of a Danish population of 240 breast cancer patients. *APMIS* **2014**, *122*, 585–592. [[CrossRef](#)] [[PubMed](#)]
64. Wei, B.; Ding, T.; Xing, Y.; Wei, W.; Tian, Z.; Tang, F.; Abraham, S.C.; Nayeemuddin, K.M.; Hunt, K.K.; Wu, Y. Invasive neuroendocrine carcinoma of the breast: A distinctive subtype of aggressive mammary carcinoma. *Cancer* **2010**, *116*, 4463–4473. [[CrossRef](#)]
65. Lewis, A.; Raoof, M.; Ituarte, P.H.G.; Williams, J.; Melstrom, L.; Li, D.; Lee, B.; Singh, G. Resection of the Primary Gastrointestinal Neuroendocrine Tumor Improves Survival With or Without Liver Treatment. *Ann. Surg.* **2019**, *270*, 1131–1137. [[CrossRef](#)]
66. DiNorcia, J.; Lee, M.K.; Reavey, P.L.; Genkinger, J.M.; Lee, J.A.; Schrope, B.A.; Chabot, J.A.; Allendorf, J.D. One Hundred Thirty Resections for Pancreatic Neuroendocrine Tumor: Evaluating the Impact of Minimally Invasive and Parenchyma-Sparing Techniques. *J. Gastrointest. Surg.* **2010**, *14*, 1536–1546. [[CrossRef](#)] [[PubMed](#)]
67. Mayo, S.C.; de Jong, M.C.; Pulitano, C.; Clary, B.M.; Reddy, S.K.; Gamblin, T.C.; Celinksi, S.A.; Kooby, D.A.; Staley, C.A.; Stokes, J.B.; et al. Surgical Management of Hepatic Neuroendocrine Tumor Metastasis: Results from an International Multi-Institutional Analysis. *Ann. Surg. Oncol.* **2010**, *17*, 3129–3136. [[CrossRef](#)] [[PubMed](#)]
68. Musunuru, S.; Chen, H.; Rajpal, S.; Stephani, N.; McDermott, J.C.; Holen, K.; Rikkers, L.F.; Weber, S.M. Metastatic Neuroendocrine Hepatic Tumors: Resection improves survival. *Arch. Surg.* **2006**, *141*, 1000–1004. [[CrossRef](#)]
69. Martinez, L.; Puig, I.; Valls, C. Colorectal liver metastases: Radiological diagnosis and staging. *Eur. J. Surg. Oncol.* **2007**, *33*, S5–S16. [[CrossRef](#)]
70. Merath, K.; Bagante, F.; Beal, E.; Lopez-Aguilar, A.G.; Poultides, G.; Makris, E.; Rocha, F.; Kanji, Z.; Weber, S.; Fisher, A.; et al. Nomogram predicting the risk of recurrence after curative-intent resection of primary non-metastatic gastrointestinal neuroendocrine tumors: An analysis of the U.S. Neuroendocrine Tumor Study Group. *J. Surg. Oncol.* **2018**, *117*, 868–878. [[CrossRef](#)]





*6.2 Neuroendocrine Tumors: An Analysis of Prevalence,  
Incidence, and Survival a Hospitals- Based Study in  
Ecuador*



## Article

# Neuroendocrine Tumors: An Analysis of Prevalence, Incidence, and Survival in a Hospital-Based Study in Ecuador

Karime Montes Escobar <sup>1,2</sup>, Jose Luis Vicente-Villardón <sup>2</sup>, Rosa Elisa Villacís Gonzalez <sup>3</sup>, Paul Humberto Castillo Cordova <sup>4</sup>, Johanna Mabel Sánchez Rodríguez <sup>5,6</sup>, Melina De la Cruz-Velez <sup>7</sup> and Aline Siteneski <sup>8,\*</sup>

<sup>1</sup> Department of Mathematics and Statistics, Institute of Basic Sciences, Universidad Técnica de Manabí, Portoviejo 130105, Ecuador

<sup>2</sup> Statistics Department, University of Salamanca, 37007 Salamanca, Spain

<sup>3</sup> Teaching and Research, SOLCA, El Oro, Machala 070206, Ecuador

<sup>4</sup> Teaching and Research, SOLCA, Loja 110150, Ecuador

<sup>5</sup> Facultad de Medicina, Universidad Laica Eloi Alfaro de Manabí, Manta 130203, Ecuador

<sup>6</sup> Facultad de Ciencias de la Salud, Universidad Estatal del Sur de Manabí, Jipijapa 130650, Ecuador

<sup>7</sup> Faculty of Health Sciences, Medicine Career, Universidad Técnica de Manabí, Portoviejo 130105, Ecuador

<sup>8</sup> Research Institute, Universidad Técnica de Manabí, Portoviejo 130105, Ecuador

\* Correspondence: aline.siteneski@gmail.com; Tel.: +593-959266284



**Citation:** Escobar, K.M.; Vicente-Villardón, J.L.; Villacís Gonzalez, R.E.; Castillo Cordova, P.H.; Sánchez Rodríguez, J.M.; De la Cruz-Velez, M.; Siteneski, A. Neuroendocrine Tumors: An Analysis of Prevalence, Incidence, and Survival in a Hospital-Based Study in Ecuador. *Healthcare* **2022**, *10*, 1569. <https://doi.org/10.3390/healthcare10081569>

Academic Editor: Hermann Nabi

Received: 30 June 2022

Accepted: 16 August 2022

Published: 18 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Neuroendocrine tumors (NETs) represent a heterogeneous malignancy group of neoplasms, with a limited amount of data from Latin America. Thus, this observational study aimed to provide data about the prevalence, incidence, and survival rates for NET in Ecuadorian hospitals. The study was conducted using data from the Society for the Fight Against Cancer (SOLCA). We evaluated patients with NETs (2000–2020) using the HJ-Biplot method and Cox proportional hazards. Annual age-adjusted incidence and limited-duration prevalence in multivariable analyses as well as hazard ratios (HRs) for mortality and survival were obtained. In the years 2000–2020, the age-adjusted incidence rate increased by 9-fold in the stomach and by 7-fold in the breast. The incidence rates were 1.38 per 100,000 persons in the lung and at 1.79 per 100,000 persons in gastroenteropancreatic sites (rectum, stomach, and pancreas). The prevalence increased from 0.0027% in 2000 to 0.0736% in 2019 and 0.0245% in 2020. Overall survival was worse for metastatic NETs (HR, 4.061; 95% CI, 1.932–8.540;  $p < 0.001$ ) and advanced local NETs (HR, 2.348; 95% CI, 1.007–5.475  $p < 0.048$ ) than for localized NETs. In conclusion, the NET incidence increased in the last 20 years and survival decreased over time, especially for metastatic tumors in the pancreas and the nostril.

**Keywords:** incidence; neuroendocrine tumor; population; prevalence; survival; HJ-Biplot

## 1. Introduction

Neuroendocrine tumors (NETs) are relatively rare tumors that are considered potentially malignant, with primary tumors being derived from neuroendocrine cells; they are found in several organs throughout the body [1]. From a distance, neuroendocrine cells exert a regulatory function on target organs that have an endocrine function; besides their paracrine function at the local organ level, the role is often of unknown specific significance [2]. Frequently, NETs originate in the gastrointestinal tract, lungs, pancreas, and thyroid [3–6], and patients above the age of 65 present a higher incidence [7]. NETs are a heterogeneous group of neuroendocrine neoplasms that differ from neuroendocrine carcinomas, and the prognosis significantly varies among subgroups [8]. Commonly, NET recognition comes from pathologic classification, in grades based on their morphologic features [8–10]; the low grades signify less risk of distant metastases, and the high grades are more aggressive with a poor prognosis [11].

Recently, the incidence and prevalence of neuroendocrine tumors (NETs) have grown, owing to, among other aspects, the increased diagnosis rates of currently available diagnostic techniques such as computed tomography and endoscopy [12,13] for early-stage tumors [14]. Globally, the highest rates of increase have occurred in countries such as the United States, Canada, and Norway [15]. Over the last four decades, the incidence of NETs has increased more than sixfold in the United States alone, with a tendency toward localized tumors rather than metastatic tumors. Approximately 170,000 people were diagnosed with NETs in the United States in 2012 [16,17]. Gastroenteropancreatic sites represent the most common NET subtype, comprising 55–70% of all NETs [14].

Particularly noteworthy is the fact that the evidence of NET prevalence and incidence in Latin American countries is very scarce [18,19]. In Argentina, an observational study evaluated a hospital database and showed that among the 532 patients diagnosed with NETs, 86.6% were gastroenteropancreatic NETs and 13.4% were bronchial NETs, with a median age of 53.2 months and an overall survival time for gastroenteropancreatic NET patients of 65.1 months [19]. A similar study was conducted in Chile, and the report evaluated 166 NET patients; they observed the prevalence of NETs among males, with an average age of 53 years and an overall survival time of 110 months. The NET primary tumor sites were found in the gastroenteropancreatic system and the small bowel [20]. Recently, two cohorts were evaluated in Panamá between 2016 and 2019, with NET occurrence at a mean age of 60 years, with a diagnosis grade of G1, and with the liver as the organ most affected [21]. Finally, a Brazilian registry of 1000 patients was evaluated over 33.7 months; the survival rate in the study was 29.3%, and the prevalence was 71.6% for thoracic NETs and 20.2% for gastroenteropancreatic tumors [18].

The heterogeneity of NETs has resulted in many different rates of prevalence and incidence, determined by the predominant site of the tumor, the vulnerable population, ethnic differences, and lifestyle. These varied aspects demonstrate the need for national and regional databases [18–20,22]. Ecuador is one of South America's smallest nations, lying on the west coast, with a total size of 283,561 km<sup>2</sup> (Central Intelligence Agency, The World Factbook, South America, 2018). It is geographically divided into four large sections: the Amazon, the Highlands, the Coast, and the Galapagos Islands. It is politically divided into 24 provinces with 224 cantons (Instituto Nacional de Estadística y Censos. Población y Demografía, 2018). In Ecuador, the Cancer Fighting Society (SOLCA) has maintained registry data of cancer prevalence and incidence since 1984 (Ministerio de Salud Pública). There are six hospitals distributed throughout the main provinces that care for approximately 31% of Ecuador's oncological cases [23]. Evidence of NET prevalence or incidence in Ecuador is scarce; for example, the global surveillance of cancer survival for the period of 1995–2009 that analyzed 67 countries, including Ecuador, did not report NET cases, probably because of the low incidence or prevalence rates [24].

In this context, the aim of our study was to evaluate the prevalence, incidence, and survival rate of NETs among Ecuadorian patients. In addition, we investigated the predominant organs affected by NETs and the impact of the histological differentiation proposed by WHO in 2010, the classifications into grades G1/G2/G3 (proliferative index Ki-67), and the primary sites of the NETs. For this purpose, 20 years' worth of data of registered patients at SOLCA hospitals were analyzed with classical and distinct statistical Cox proportional hazards and the HJ-Biplot method.

## 2. Materials and Methods

The retrospective cohort study was accomplished with a clinical story assay of NET patients. The Cancer Fighting Society (SOLCA) databases of the Manabí, Cuenca, Loja, and El Oro provinces were analyzed. All NET patients who had been diagnosed and attended to at the Oncology Service of SOLCA centers between January 2000 and December 2020 were included. Patients who did not have a complete clinical story, abandoned treatment, or did not follow treatment in the hospital were excluded. Here, we assessed age, sex, population, stage and differentiation grade of tumor, year of diagnosis, and survival of the NET patients

(measured in months or until death). The clinical and pathological characteristics of the study population are detailed in Table 1.

**Table 1.** Characteristics of study population.

Characteristics	All Patients N = 281	% 100
<b>Gender</b>		
Men	155	55
Women	126	45
<b>Age (range)</b>		
	<b>Median = 61.6</b>	
10–19	5	2
20–29	6	2
30–39	18	6
40–49	18	6
50–59	57	20
60–69	56	20
70–79	64	23
80+	57	20
<b>Localization</b>		
Lung	51	18.1
Stomach	34	12.1
Breast	26	9.3
Nostril	20	7.1
Rectum	18	6.4
Cervix	16	5.7
Unknown	15	5.3
Pancreas	14	5
Appendix	13	4.6
Skin	11	3.9
Duodenum	9	3.2
Anus	8	2.8
Cecum	8	2.8
Colon	8	2.8
Mediastinum	8	2.8
Lymph Nodes	5	1.8
Liver	5	1.8
Retroperitoneum	3	1.1
Pyloric Antrum	1	0.4
Coledocus	1	0.4
Endometrium	1	0.4
Exocervix	1	0.4

**Table 1.** *Cont.*

Characteristics	All Patients N = 281	% 100
<b>Localization</b>		
Ileum	1	0.4
Pelvis	1	0.4
Prostate	1	0.4
Soft tissues	1	0.4
Bladder	1	0.4
<b>WHO 2010</b>		
G1	56	20
G2	69	25
G3	156	56
<b>Stage</b>		
Local	52	19
Advanced Local	78	28
Metastatic	151	54
<b>Survival</b>		
Alive	67	23.8
Dead	214	76.2
<b>SG</b> (Overall survival in months) Total cases	Median: 28.70	7.5
<b>SG</b> (Overall survival in months) 5 years	Median: 21	16

### 2.1. NET Classification

The NETs were analyzed according to the WHO (2010) histological grades: Grade 1 (G1, Ki-67 index <2% and mitotic count <2/10 high-power fields (HPFs)), Grade 2 (G2, Ki-67 index 3–20% and/or mitotic count 2–20/10 HPF), and Grade 3 (G3, Ki-67 index 3–20% and/or mitotic count 2–20/10 HPF), and G3 (G3 neuroendocrine carcinoma were not differentiated), notwithstanding tumor differentiation [8,25]. In addition, the tumor stages of NET patients were assayed as local, advanced local, and metastatic. The local tumor was within the limits of the organ of origin; the advanced local tumor expanded beyond the organ of origin into adjacent tissues or organs or infiltrated regional lymph nodes; finally, the metastatic tumor was one that extended to remote sites outside the organ of origin.

### 2.2. Statistical Analysis

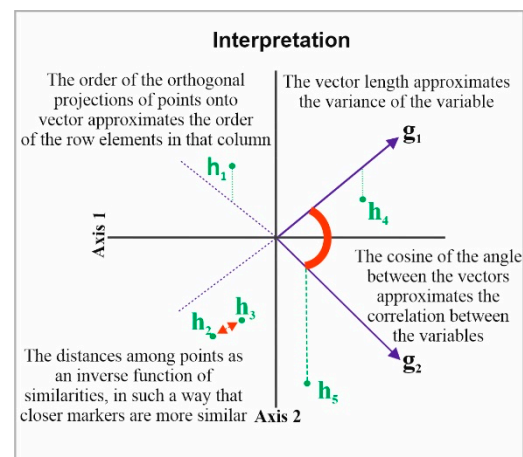
The examination of survival at 20 years according to site was performed using the Kaplan–Meier method and the log-rank test. In addition, we evaluated the data from the Sociedad de Lucha Contra el Cancer (SOLCA) hospital registry to extract the median overall survival (OS) by site, stage, and grade. To evaluate the most recent trends in survival, we conducted multivariable survival analyses. Five-year overall survival (OS) time and the Cox proportional hazards model were used in the multivariable analysis, with censoring applied at 5 years. Covariates for this analysis included factors known to influence the prognosis of NETs, including grade, age, stage, site, and time interval from diagnosis. The overall model was significant at  $p < 0.001$ .

Incidence (including annual percentage change) and limited-duration prevalence rates (20 years) were calculated using the annual percentage change by fitting a least-squares regression to the rates, using the calendar year as a regression variable. Age-adjusted incidence rates were computed using weighted proportions of the corresponding age

groups in the Ecuadorian standard population for each province: El Oro (715,751), Manabí (1,562,079), Lora (521,184), and Azuay (881,394), with a total of 3,670,408 (Instituto Nacional de Estadística y Censos, 2020).

### 2.3. HJ-Biplot

Next, the HJ-Biplot [26] was used, which is an extension of the classic biplot introduced by Gabriel [27] and is an exploratory data analysis method that searches for hidden patterns in the data matrix. The HJ-Biplot has the advantage of simultaneously being a representation and achieving an optimal rendering quality for both rows and columns, with both represented in the same reference system (Figure 1). The data matrix then graphically shows the information contained in the rows (sample location) and columns (grade and months). The Biplot is a multivariate statistical technique that was performed in order to allow for a more detailed data analysis, highlighting the links between parts (sample location) and the gender and age of the individuals.



**Figure 1.** Description the HJ-Biplot method.

Statistical analyses were performed using IBM SPSS version 25.0 (IBM, Armonk, NY, USA), R software version 3.3.1 (<http://www.r-project.org>; accessed on 30 June 2022), and the MultiBiplot freely available software [28]. Comparative differences were considered significant at  $p < 0.01$ .

## 3. Results

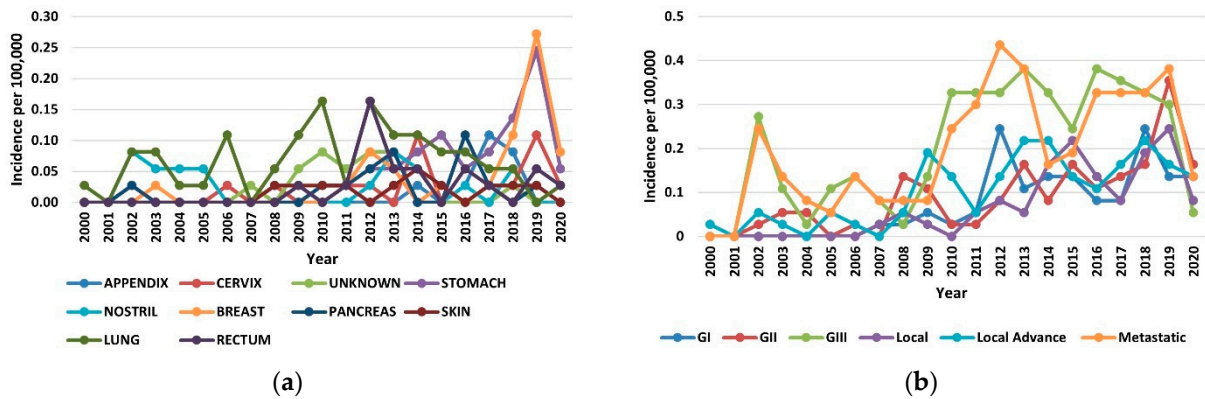
Between 2000 and 2020, 281 admissions corresponding to patients affected by NETs were registered in the database of SOLCA clinics in four provinces of Ecuador. We observed that the most common NETs were those of the lung (18.1%), stomach (12.1%), breast (9.3%), nasal cavity (7.1%), and rectum (6.4%). The database permitted the analysis of tumor grade according to the WHO classification; most of the cases were at the advanced stage G3 (56%), followed by G2 (25%), and G1 (20%). The primary tumor sites that most frequently presented with stage III disease were the lung (21%), stomach (8.9%), pancreas (8.9%), nostril (7%), unknown (7%), and finally, the breast (6.4%) (Table 1).

### 3.1. Annual Incidence

The increase in the incidence of NETs in the period of 2000–2020 occurred across all organs, stages, and grades. In the year 2019, the increases in incidence for various sites ranged from 8-fold in the breast to 5-fold in the stomach (Figure 2a). Among the stage groups, the incidence increased the most for metastatic NETs, from 0.24 per 100,000 persons in 2000 to 0.38 per 100,000 persons in 2020 ( $p < 0.001$ ) (Figure 2b). The highest incidences were 1.38 per 100,000 persons for the lung, 1.79 per 100,000 persons for gastroenteropancreatic sites (including 0.92 per 100,000 persons for the stomach, 0.49 per 100,000 persons for



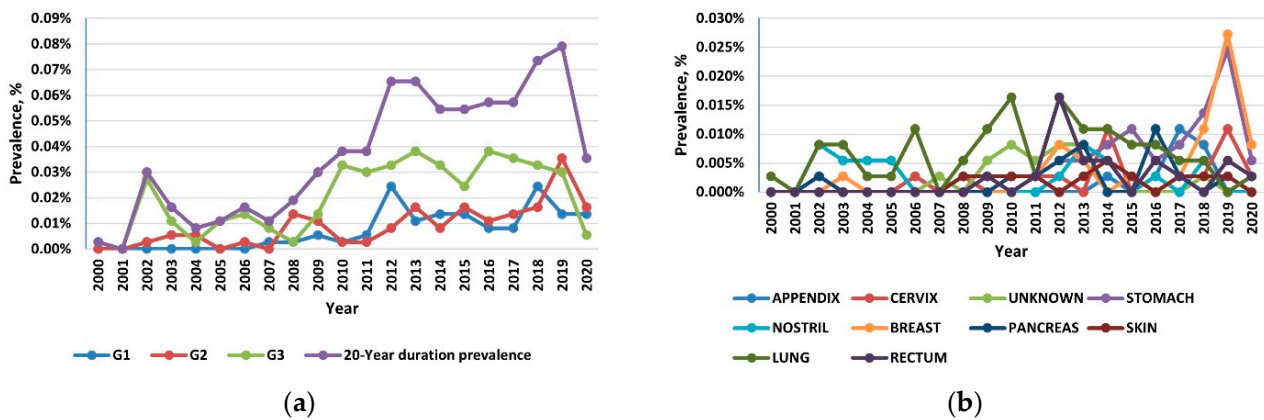
the rectum, and 0.38 per 100,000 persons for the pancreas), and 0.40 per 100,000 persons for NETs with an unknown primary site of origin.



**Figure 2.** Incidence trends of neuroendocrine tumors (NETs) from 2000 to 2020: (a) Annual age-adjusted incidence of NETs by site; (b) Annual age-adjusted incidence of NETs by stage and grade.

### 3.2. Prevalence

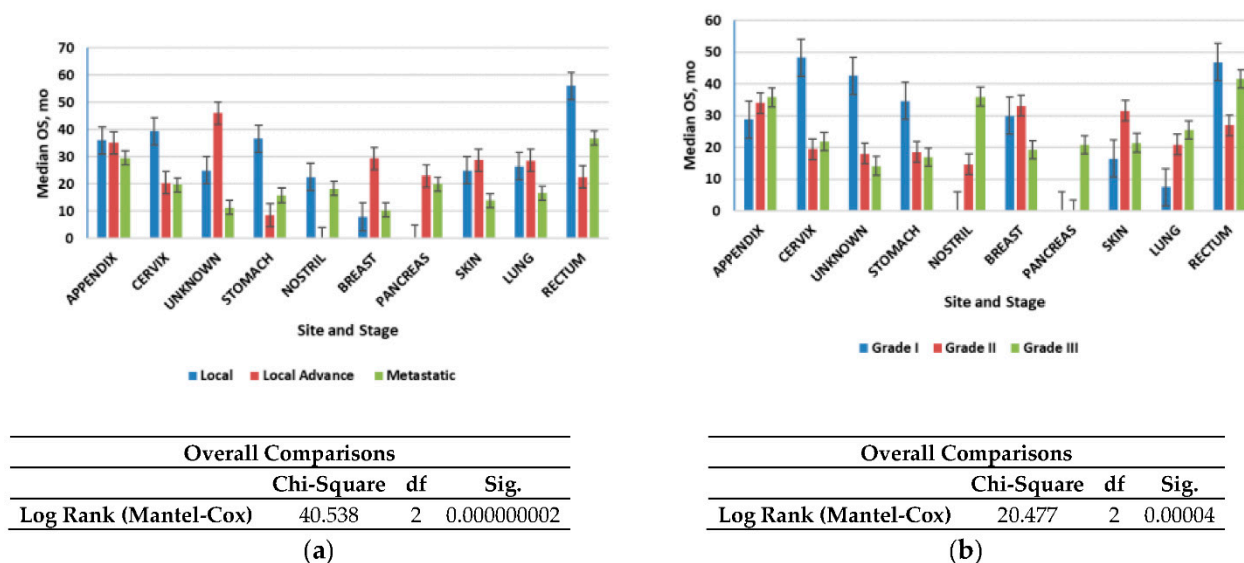
Reflecting the rising incidence and indolent nature of NETs, the 20-year limited-duration prevalence increased substantially, from 0.0027% in 2000 to 0.0736% in 2019 and 0.0245% in 2020 ( $p < 0.001$ ) (Figure 3a). Among the grade groups, the prevalence increased the most for the G3 NETs, and among the sites, prevalence was the highest in the lung, followed by the stomach and the pancreas (Figure 3b).



**Figure 3.** (a) The 20-year limited-duration prevalence of all neuroendocrine tumors (NETs), according to grade; (b) The 20-year limited-duration prevalence of neuroendocrine tumors by site.

### 3.3. Survival

Figure 4a shows that the median OS time for all patients was 2.4 years (28.8 months); localized NETs had a better median OS at 2.30 years (26.6 months) as compared with advanced local NETs at 2.14 years (25.7 months) and metastatic NETs at 1.60 years (19.2 months) ( $p < 0.001$ ). Of those with known grades, G1 NETs had the highest median OS (2.12 years) (25.4 months) among the grade groups, G2 NETs had a worse OS at 2.07 years (24.9 months), while G3 NETs had the worst OS (1.8 years). NETs in the rectum (3.6 years) and appendix (2.7 years) had the best median OS among the site groups, while NETs in the pancreas (1.19 years) and nostril (1.14 years) had the worst median OS. All these differences in survival were significant (log-rank  $p < 0.001$ ).



**Figure 4.** Median overall survival (OS) of neuroendocrine tumors (NETs): (a) Median OS of all patients included in the study according to stage; (b) Median OS of all patients included in the study according to grade. Error bars indicate 95% CI.

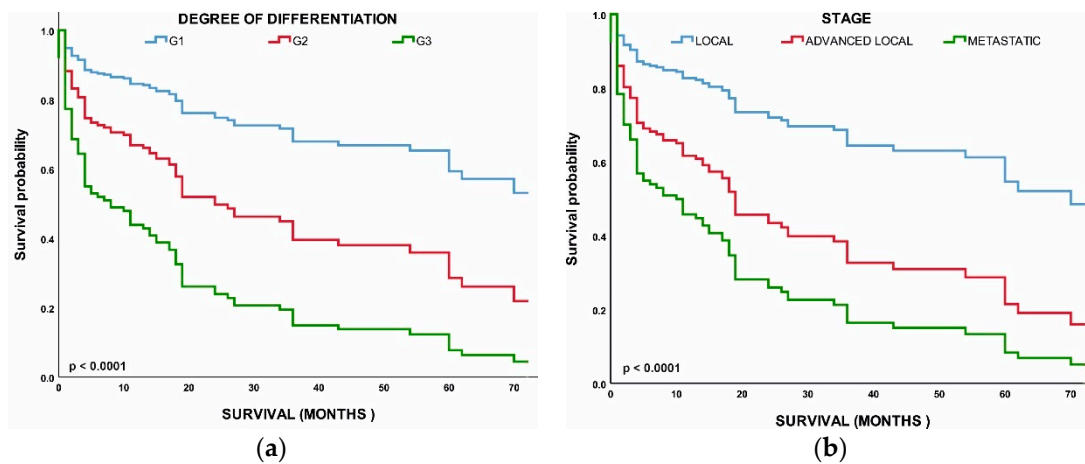
We then evaluated the survival patterns according to site and stage (Figure 4b). In the localized NETs, the median OS ranged from 3 years in the small intestine to more than 4.6 years in the appendix. In advanced local NETs, the median OS ranged from 2.9 years for NETs in the appendix to more than 30 years in the unknown primary site. For distant NETs, those in the rectum had the best median OS (3 years); NETs in the appendix (2.4 years) and pancreas (1.6 months) had the worst median OS. All of these differences in OS were significant (log-rank  $p < 0.001$ ) (Figure 4b).

### 3.4. Multivariable Analysis of OS

Next, we performed a multivariable analysis with hazard ratios (HRs) calculated for 5-year mortality hazard rates (Table 2). We found that patients with G2 (HR, 2.407; 95% CI, 0.941–6.154;  $p < 0.037$ ) and G3 (HR, 5.003; 95% CI, 2.221–11.273;  $p < 0.001$ ) NETs had worse OS than those with G1 NETs. Age, stage, and grade were all found to have a significant correlation with survival. Overall survival was worse in metastatic NETs (HR, 4.061; 95% CI, 1.932–8.540;  $p < 0.048$ ) and advanced local NETs (HR, 2.348; 95% CI, 1.007–5.475;  $p < 0.001$ ) than in the localized NETs. All of the above comparisons were significant at  $p < 0.001$  (Figure 5).

**Table 2.** Cox proportional hazards regression model for overall survival (5 years).

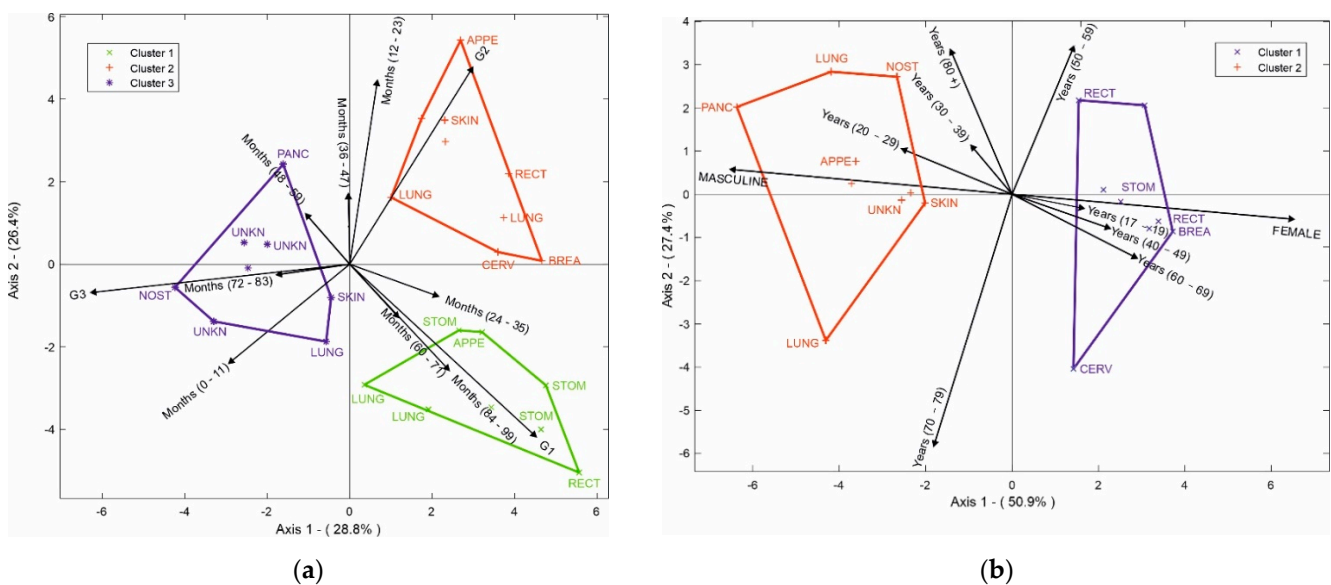
Variable	Estimate	Hazard Ratio	95% CI		p Value
Age (years)	0.023	1.023	1.004	1.042	0.015
<b>WHO 2010</b>					
G1	-	Ref	Ref	Ref	0.000
G2	0.878	2.407	0.941	6.154	0.037
G3	1.610	5.003	2.221	11.273	0.001
<b>Stage</b>					
Local	-	Ref	Ref	Ref	0.001
Advanced Local	0.853	2.348	1.007	5.575	0.048
Metastatic	1.402	4.061	1.932	8.540	0.001



**Figure 5.** (a) Survival of patients by tumor grade: grade 1 (G1), grade 2 (G2), grade 3 (G3); (b) Survival of patients by tumor stage: local, advanced local, and metastatic.

3.5. HJ-Biplot

In addition, the HJ-Biplot method was used (Figure 6a) to determine the relationship among variables, considering the 10 most prevalent NETs. The degree of differentiation, months of survival, and the location of the tumor were evaluated. In the degree of differentiation for G1, the highest survival rates occurred between 84 and 99 months; tumors in the stomach and appendix had a high correlation with 24–35 months of survival, while tumors in the rectum and lung had an even higher survival rate at 84–99 months. When we evaluated the degree of differentiation for G2, the highest survival was observed between 12 and 23 months in NETs in the appendix, lung, breast, and cervix. Finally, we evaluated the degree of differentiation for G3; the skin had the worst prognosis, with a survival range between 0 and 11 months, a survival range between 48 and 59 months for the pancreas, and between 72 and 83 months for nasal cavity tumors together with NETs of unknown origin.



**Figure 6.** (a) HJ-Biplot relating the degree of differentiation and months of survival, according to tumor location; (b) HJ-Biplot relating gender and age according to tumor location.

Figure 6b shows age, gender, and location of the NETs where two groups were observed. The first group represents the female gender, which had a significant correlation with the cervix, breast, stomach, and rectum tumors. The second group was the male

gender, which was correlated with tumors of the pancreas, unknown site, skin, lung, nostril, and appendix. It was possible to determine that the rectum NETs for the female gender were related to ages between 50 and 59 years, while the stomach, breast, and cervix NETs had a higher connection with women between 60 and 69 years. For men, the nasal fossa tumor appeared in individuals over 80 years old, NETs of the pancreas, appendix, skin, and unknown sites had a greater connection with young people between 20 and 29 years old, and lung tumors appeared between the ages of 70 and 79 years.

#### 4. Discussion

In this study of NETs using data from an Ecuadorian center, we assessed 20 years' worth of data and found that the age-adjusted annual incidence increased for metastatic NETs from 0.24 per 100,000 persons in 2000 to 0.38 per 100,000 persons in 2020. Prevalence rates grew from 0.0027% in 2000 to 0.0736% in 2019 and 0.0245% in 2020. A decrease in NET prevalence in 2020 as compared with 2019 may have been a consequence of undercounting due to the COVID-19 pandemic. Unfortunately, our data showed that advanced-grade NETs showed the highest prevalence, with main sites in the lung, stomach, and pancreas. Additionally, our NET survival rate showed a low median at 2.4 years for all patients; the rates were better for localized NETs and the worst for metastatic NETs. These data on patients' survival may reflect the public and private structure of the health system in Latin America, where public policies encouraging health care and access to cancer treatment are still scarce [29].

This retrospective cohort study examined the general incidence increases—9-fold for the stomach and 7-fold for the breast—between 2000 and 2020. Prior studies carried out worldwide have also reported that the NET incidence increased 6.4-fold in 39 years; their data reported an increase from 1.09 per 100,000 in 1973 to 6.98 per 100,000 in 2012 in the United States [14]. According to a Kentucky study, a 20-year database showed the rise from 3.1 in 1995 to 10.3 in 2015 per 100,000 cases of NET patients [22]. In fact, differences in population rates may result from a combination of environmental factors and biological differences owing to differing national demographics [30–32]. Ecuador is positioned on the equatorial landline, has territory in both hemispheres, and is traversed from north to south by the Andes Mountains. It has approximately 17,511,000 inhabitants, with a predominant mestizo (descendants of Indigenous Americans and Spaniards) population. In Ecuador, the Quito SOLCA database revealed that between 2011 and 2015, NETs represented 1% of total cancer cases per 100,000 inhabitants, and the Guayaquil SOLCA data reported no specific NETs in their published reports [33]. Our study considered the SOLCA database of NETs patients from four large centers of the country, which were assessed for a period of 20 years. In the Ecuadorian population, the most common site of origin for NETs was the lung, followed by the gastroenteropancreatic system.

Previous data demonstrated that mostly gastroenteropancreatic NET tumor sites surged, as reported in the EUA database, from 1.09 per 100,000 in 1973 to 6.98 per 100,000 in 2012 [14]. It is important to highlight that the growth in incidence data can be attributed to modern imaging techniques that favor appropriate diagnoses and an improved recognition of neuroendocrine histology [14,15]. In fact, gastroenteropancreatic NETs represented the most common site of origin when standardizing our data. Worth noting is the fact that gastroenteropancreatic NETs represented the second most prevalent digestive cancer [5]. Pancreatic NETs originating in Langerhans islets or alternative origins in precursors in the ductal epithelium [34–36] because of malignance represented the most widely studied NETs between 1981 and 2020 [37]. Our data also suggested poor survival prognosis among pancreatic NET patients, at only 1.19 years after diagnosis. On the other hand, the NET site of origin prone to the earliest diagnosis was the stomach.

A relevant consideration of our incidence analysis was the high rate of lung NETs, with 1.38 per 100,000 inhabitants. In this study, we observed that the lungs represented the most prevalent site for NETs. In the lungs, the NETs originate in the amine precursor uptake and decarboxylation of neuroendocrine cells from Kulchitsky cells [38] and account



for around 20% of all lung malignancies [39]. In addition, pulmonary tumor morphologies are different when compared to other NETs, with the presence of rosette-like structures, organoid nesting, and peripheral palisading patterns [39]. A recent retrospective analysis showed the lung as the second most likely NET site in the USA, with the highest rates from 2000 to 2012 at 1.49 per 100,000 [14]. An increase of up to 6% per year was observed for pulmonary NETs in the USA [40]. Another study performed in the USA showed a 7% yearly rise in the incidence of lung NETs between 2004 and 2014 [41]. In particular, the primary tumor sites that most frequently presented with severe stage (G3) disease in our data were the lungs. The overall prognosis of our data was not favorable for NET patients, with over half of tumor diagnoses being in the advanced stage.

NETs are malignant heterogeneous tumors originating from cells of the diffuse neuroendocrine system that can show both nerve and endocrine cell features and are found in many organs in the body [42]. Nevertheless, several risk factors are associated with tumor incidence; for example, a family history of neuroendocrine tumors of endocrine neoplasia [43,44]. More frequently found are clinical symptoms associated with NETs such as cough, dyspnea, weight loss, diarrhea, constipation, abdominal pain, dyspepsia, hypertensive crisis, distended abdomen, and intestinal obstruction [45]. NETs are often identified based on phenotypical and morphological features consistent with the WHO classification [39], which is commonly revealed by employing immunostaining with neuroendocrine markers and Ki-67. In the current WHO categorization, synaptophysin, chromogranin A, and CD56 are recommended as neuroendocrine indicators, with synaptophysin and chromogranin A being recommended as first-line options [39]. NETs have distinct clinical and biological properties depending on where they originate; for instance, those originating from enterochromaffin cells characterized as G1 NETs are part of the diffuse neuroendocrine system and may present a prevalence of 39.2% for GI NETs of the rectum and 27.8% for the small intestine [46]. In this context, our results suggest that 56% of NETs are diagnosed in the G3 stage, most frequently in the lung, followed by the stomach and pancreas. Nowadays, predictive and prognostic markers have been extensively researched in order to determine a better clinical management strategy for NETs [47].

General radiology methods such as more sensitive computed tomography and magnetic resonance imaging, in addition to the best therapies for gastrointestinal NETs, were previously reported as responsible for the increased survival indices [12,13]. Unfortunately, in our data, the low survival rate found among NET patients studied for 2.4 years could represent the advanced stage of tumor diagnosis. However, our data represent a low median at 2.12 years (25.4 months) for G1 NETs and 1.8 years (21.6 months) for G3 NETs as compared with other population studies that have survival rates of >30 years for regional diseases and 12 months for patients with distant metastases [14]. Evidently, the survival rate was reduced in the more advanced stage of the disease and diagnosis at a high grade; indeed, we observed small survival rates of 1.19 years for the pancreas and 1.14 years for the nostril.

Our study also has some limitations. First, the characteristics found in our sample may not be representative of the performance in other centers of the country. However, in Ecuador, the Instituto Nacional de Estadística y Censos (INEC) collects, analyzes, and reports statistical information on health, including cancer deaths, but it does not report on NET specifically. Second, INEC uses the International Classification of Diseases, tenth revision (ICD-10), to report the causes of death. This classification provides codes for causes of cancer death, but not for NET specifically. There is a possibility that the ICD-10 includes NET in a wide range of endocrine, nutritional, and metabolic diseases, which makes the use of such database in our evaluation difficult. Third, we collected the SOLCA data from four important regions of Ecuador, and we acknowledge that this may not show the real scenario of NET patients. SOLCA data from Quito and Guayaquil were missing because representatives from these sites refused to participate in our research. For these reasons, our results on prevalence, incidence, and survival may underestimate the real number of NET patients in Ecuador. Our data represent an upper ceiling to the high proportion of

G3; in fact, this may have contributed to the median survival time of 2.4 years, which is significantly worse than what is known for the classical distribution of NET G1/G2. Finally, the current WHO (2019) classification defines NETs as well-differentiated and poorly differentiated. Well-differentiated NETs are based on the mitotic rate and/or Ki-67 labeling index and the grades of G1/low, G2/intermediate, and G3/high. Poorly differentiated ones are G3 NETs in neuroendocrine carcinomas of small or large cell carcinoma [48]. The WHO 2019 classification did not use in our database evaluation between 2010 and 2020. New studies are necessary to provide the NET current WHO classification in the Ecuadorian population.

The findings of this study have important implications for NET patients, for communities in the country, healthcare systems, and future research. Based on the results of this study, the NETs are defined as a serious type of cancer with low survival times and with its prevalence increased in the last two decades. Especially in low-income and mid-income countries, understanding the prevalence of NETs is crucial in changing public policies and setting priorities in order to improve patient care. Therefore, the estimates of survival time in this study and the identification of the characteristics of patients most affected by NETs as well as the more common sites of the disease may help to guide physicians in making early diagnoses among Ecuadorian populations. Although relatively rare, our results highlight the severity of NETs, and the data presented may contribute to changes in public policies in the country. This is the first study on NETs in Ecuador, and the data on incidence and prevalence must be considered as preliminary. More research is necessary at a national scale to ensure that the whole population is represented.

## 5. Conclusions

The incidence of NETs increased in Ecuadorian hospitals between 2000 and 2020, and we provided primary data on their prevalence. Our data showed that the drastic reduction in diagnosed cases of NETs in 2020 reflected the drop in the number of people seeking health services owing to the worldwide COVID-19 pandemic. It is possible to assert that the statistical HJ-Biplot method we used confirmed our results, and the classical evaluation and proportional hazards model effectively determined the survival rates. Unfortunately, the advanced stages of NETs during the diagnosis of patients in Ecuador contributed to the low survival time observed in our data. The survival was greater for those with localized NETs, while metastatic tumors presented a worse prognosis, particularly in the pancreas and nostril NETs. It is possible that for the Ecuadorian population, standard diagnosis procedures with a focus on the main NET sites found in our study may be implemented.

**Author Contributions:** Data curation, K.M.E. and M.D.I.C.-V.; formal analysis, K.M.E.; investigation, K.M.E. and A.S.; methodology, A.S.; supervision, J.L.V.-V.; visualization, R.E.V.G., P.H.C.C. and J.M.S.R.; writing—original draft, K.M.E. and A.S.; writing—review and editing, K.M.E. and A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Due to nature of the data extracted, all patients were previously anonymized. This is considered a low-risk investigation because it is only a review of medical records. Data management was carried out under the Habeas Data law for security of information management. We did not have access to the personal data of the patients.

**Acknowledgments:** This study was supported by the Department of Mathematics and Statistics, Institute of Basic Sciences and Research Institute of the Technical University of Manabí. The authors would like to thank SOLCA Manabí, Cuenca, Loja, and El Oro for their support and for facilitating the data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Caldarella, A.; Crocetti, E.; Paci, E. Distribution, incidence, and prognosis in neuroendocrine tumors: A population based study from a cancer registry. *Pathol. Oncol. Res.* **2011**, *17*, 759–763. [[CrossRef](#)]
2. Rindi, G.; Inzani, F. Neuroendocrine neoplasm update: Toward universal nomenclature. *Endocr. Relat. Cancer* **2020**, *27*, R211–R218. [[CrossRef](#)] [[PubMed](#)]
3. Kulke, M.H.; Mayer, R.J. Carcinoid tumors. *N. Engl. J. Med.* **1999**, *1*, 858–868. [[CrossRef](#)] [[PubMed](#)]
4. Modlin, I.M.; Oberg, K.; Chung, D.C.; Jensen, R.T.; de Herder, W.W.; Thakker, R.V.; Caplin, M.; Delle Fave, G.; Kaltsas, G.A.; Krenning, E.P. Gastroenteropancreatic neuroendocrine tumours. *Lancet Oncol.* **2008**, *9*, 61–72. [[CrossRef](#)]
5. Yao, J.C.; Hassan, M.; Phan, A.; Dagohoy, C.; Leary, C.; Mares, J.E.; Abdalla, E.K.; Fleming, J.B.; Vauthey, J.N.; Rashid, A.; et al. One hundred years after “carcinoid”: Epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J. Clin. Oncol.* **2008**, *26*, 3063–3072. [[CrossRef](#)]
6. Frenkel, S.; Pe’er, J. Choroidal metastasis of adenocarcinoma of the lung presenting as pigmented choroidal tumor. *Case Rep. Ophthalmol.* **2012**, *3*, 311–316. [[CrossRef](#)]
7. Van Der Zwan, J.M.; Trama, A.; Otter, R.; Larrañaga, N.; Tavilla, A.; Marcos-Gragera, R.; Dei Tos, A.P.; Baudin, E.; Poston, G.; Links, T.; et al. Rare neuroendocrine tumours: Results of the surveillance of rare cancers in Europe project. *Eur. J. Cancer* **2013**, *49*, 2565–2578. [[CrossRef](#)]
8. Rindi, G. 3. The ENETS guidelines: The new TNM classification system. *Tumori J.* **2010**, *96*, 806–809. [[CrossRef](#)]
9. Rindi, G.; Klöppel, G.; Alhman, X.X.H.; Caplin, M.; Couvelard, A.; De Herder, W.W.; Eriksson, B.; Falchetti, A.; Falconi, M.; Komminoth, P. TNM staging of foregut (neuro) endocrine tumors: A consensus proposal including a grading system. *Virchows Arch.* **2006**, *449*, 395–401. [[CrossRef](#)]
10. Khanna, L.; Prasad, S.R.; Sunnapwar, A.; Kondapaneni, S.; Dasyam, A.; Tammiseti, V.S.; Salman, U.; Nazarullah, A.; Katabathina, V.S. Pancreatic Neuroendocrine Neoplasms: 2020 Update on Pathologic and Imaging Findings and Classification. *Radiographics* **2020**, *40*, 1240–1262. [[CrossRef](#)]
11. Oronsky, B.; Ma, P.C.; Morgensztern, D.; Carter, C.A. Nothing But NET: A Review of Neuroendocrine Tumors and Carcinomas. *Neoplasia* **2017**, *19*, 991–1002. [[CrossRef](#)] [[PubMed](#)]
12. Hess, E.P.; Haas, L.R.; Shah, N.D.; Stroebel, R.J.; Denham, C.R.; Swensen, S.J. Trends in computed tomography utilization rates. *J. Patient Saf.* **2014**, *10*, 52–58. [[CrossRef](#)] [[PubMed](#)]
13. Pelc, N.J. Recent and future directions in CT imaging. *Ann. Biomed. Eng.* **2014**, *42*, 260–268. [[CrossRef](#)] [[PubMed](#)]
14. Dasari, A.; Shen, C.; Halperin, D.; Zhao, B.; Zhou, S.; Xu, Y.; Shih, T.; Yao, J.C. Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol.* **2017**, *3*, 1335–1342. [[CrossRef](#)] [[PubMed](#)]
15. Leoncini, E.; Boffetta, P.; Shafir, M.; Aleksovska, K.; Boccia, S.; Rindi, G. Increased incidence trend of low-grade and high-grade neuroendocrine neoplasms. *Endocrine* **2017**, *58*, 368–379. [[CrossRef](#)]
16. Chauhan, A.; Kohn, E.; Del Rivero, J. Neuroendocrine Tumors—Less Well Known, Often Misunderstood, and Rapidly Growing in Incidence. *JAMA Oncol.* **2020**, *6*, 21–22. [[CrossRef](#)]
17. Cives, M.; Strosberg, J.R. Gastroenteropancreatic neuroendocrine tumors. *CA Cancer J. Clin.* **2018**, *68*, 471–487. [[CrossRef](#)]
18. Younes, R.N. Tumores neuroendócrinos: Registro de 1000 pacientes. *Rev. Assoc. Med. Bras.* **2008**, *54*, 305–307. [[CrossRef](#)]
19. O’Connor, J.M.; Marmissolle, F.; Bestani, C.; Pesce, V.; Belli, S.; Dominichini, E.; Mendez, G.; Price, P.; Giacomi, N.; Pairola, A. Observational study of patients with gastroenteropancreatic and bronchial neuroendocrine tumors in Argentina: Results from the large database of a multidisciplinary group clinical multicenter study. *Mol. Clin. Oncol.* **2014**, *2*, 673–684. [[CrossRef](#)]
20. Pinto, M.P.; Muñoz Medel, M.; Carrillo, D.; Retamal, I.N.; Bravo, M.L.; Valenzuela, Y.; Nervi, B.; Sánchez, C.; Galindo, H.; Ibañez, C. Chilean registry for neuroendocrine tumors: A Latin American perspective. *Horm. Cancer* **2019**, *10*, 3–10. [[CrossRef](#)]
21. Cukier, M.; Vergara, R.; Mendez-Rios, J.D.; Castillo, O.; Barrera, I.; Tello, E.; El Achar, O.; Loo, Y.; Tapia, H.; Perez, G. Neuroendocrine tumors in Panama: A nationwide database analysis. *Mol. Clin. Oncol.* **2021**, *15*, 1–8. [[CrossRef](#)] [[PubMed](#)]
22. Chauhan, A.; Yu, Q.; Ray, N.; Farooqui, Z.; Huang, B.; Durbin, E.B.; Tucker, T.; Evers, M.; Arnold, S.; Anthony, L.B. Global burden of neuroendocrine tumors and changing incidence in Kentucky. *Oncotarget* **2018**, *9*, 19245–19254. [[CrossRef](#)] [[PubMed](#)]
23. Corral, F.; Cueva, P.; Yépez, J.; Tarupi, W. Tendencias en incidencia y mortalidad por cáncer durante tres décadas en Quito—Ecuador. *Colomb. Med.* **2018**, *49*, 35–41. [[CrossRef](#)]
24. Allemani, C.; Weir, H.K.; Carreira, H.; Harewood, R.; Spika, D.; Wang, X.-S.; Bannon, F.; Ahn, J.V.; Johnson, C.J.; Bonaventure, A. Global surveillance of cancer survival 1995–2009: Analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* **2015**, *385*, 977–1010. [[CrossRef](#)]
25. Assarzaghan, N.; Montgomery, E. What is new in the 2019 World Health Organization (WHO) classification of tumors of the digestive system: Review of selected updates on neuroendocrine neoplasms, appendiceal tumors, and molecular testing. *Arch. Pathol. Lab. Med.* **2021**, *145*, 664–677. [[CrossRef](#)]
26. Villardón, M.P.G. Una alternativa de representacion simultanea: HJ-Biplot. *Qüestió Quad. d’estadística Investig. Oper.* **1986**, *10*, 13–23.
27. Gabriel, K.R. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **1971**, *58*, 453–467. [[CrossRef](#)]
28. Vicente-Villardón, J.L. *MULTBILOT: A Package for Multivariate Analysis Using Biplots*; Departamento de Estadística, Universidad de Salamanca: Salamanca, Spain, 2010.

29. Uema, D.; Alves, C.; Mesquita, M.; Nuñez, J.E.; Siepmann, T.; Angel, M.; Rego, J.F.M.; Weschenfelder, R.; Rocha Filho, D.R.; Costa, F.P. Carcinoid heart disease and decreased overall survival among patients with neuroendocrine tumors: A retrospective multicenter Latin American cohort study. *J. Clin. Med.* **2019**, *8*, 405. [[CrossRef](#)]
30. Das, S.; Dasari, A. Epidemiology, incidence, and prevalence of neuroendocrine neoplasms: Are there global differences? *Curr. Oncol. Rep.* **2021**, *23*, 1–7. [[CrossRef](#)]
31. Ren, H.M.; Liao, M.Q.; Tan, S.X.; Cheng, C.; Zhu, S.; Zheng, L.; Ma, J.R.; Mu, Y.J.; Li, W.L.; Zhang, S.W.; et al. Global, Regional, and National Burden of Cancer in Children Younger Than 5 Years, 1990–2019: Analysis of the Global Burden of Disease Study 2019. *Front. Public Health* **2022**, *10*, 910641. [[CrossRef](#)]
32. Kim, H.I.; Lim, H.; Moon, A. Sex Differences in Cancer: Epidemiology, Genetics and Therapy. *Biomol. Ther. (Seoul)* **2018**, *26*, 335–342. [[CrossRef](#)] [[PubMed](#)]
33. Jaramillo-Feijo, L.; Real-Cotto, J.; Tanca-Campo, J.; Puga-Peña, G.; Quino-Briones, R. Incidencia y mortalidad del cáncer, en Hospital Solca—Guayaquil. *J. Health Med. Sci.* **2020**, *6*, 227–232.
34. Vortmeyer, A.O.; Huang, S.; Lubensky, I.; Zhuang, Z. Non-islet origin of pancreatic islet cell tumors. *J. Clin. Endocrinol. Metab.* **2004**, *89*, 1934–1938. [[CrossRef](#)] [[PubMed](#)]
35. Grande, E.; Capdevila, J.; Barriuso, J.; Antón-Aparicio, L.; Castellano, D. Gastroenteropancreatic neuroendocrine tumor cancer stem cells: Do they exist? *Cancer Metastasis Rev.* **2012**, *31*, 47–53. [[CrossRef](#)]
36. Klöppel, G.; Anlauf, M.; Perren, A. Endocrine precursor lesions of gastroenteropancreatic neuroendocrine tumors. *Endocr. Pathol.* **2007**, *18*, 150–155. [[CrossRef](#)]
37. Escobar, K.M.; Vicente-Villardón, J.L.; de la Hoz-M, J.; Useche-Castro, L.M.; Alarcón Cano, D.F.; Siteneski, A. Frequency of Neuroendocrine Tumor Studies: Using Latent Dirichlet Allocation and HJ-Biplot Statistical Methods. *Mathematics* **2021**, *9*, 2281. [[CrossRef](#)]
38. Wolin, E.M. Advances in the diagnosis and management of well-differentiated and intermediate-differentiated neuroendocrine tumors of the lung. *Chest* **2017**, *151*, 1141–1146. [[CrossRef](#)]
39. Travis, W.D.; Brambilla, E.; Nicholson, A.G.; Yatabe, Y.; Austin, J.H.M.; Beasley, M.B.; Chirieac, L.R.; Dacic, S.; Duhig, E.; Flieder, D.B. The 2015 World Health Organization classification of lung tumors: Impact of genetic, clinical and radiologic advances since the 2004 classification. *J. Thorac. Oncol.* **2015**, *10*, 1243–1260. [[CrossRef](#)]
40. Modlin, I.M.; Lye, K.D.; Kidd, M. A 5-decade analysis of 13,715 carcinoid tumors. *Cancer* **2003**, *97*, 934–959. [[CrossRef](#)]
41. Broder, M.S.; Cai, B.; Chang, E.; Neary, M.P. Incidence and prevalence of neuroendocrine tumors of the lung: Analysis of a US commercial insurance claims database. *BMC Pulm. Med.* **2018**, *18*, 1–11. [[CrossRef](#)]
42. Modlin, I.M.; Champaneria, M.C.; Bornschein, J.; Kidd, M. Evolution of the diffuse neuroendocrine system—clear cells and cloudy origins. *Neuroendocrinology* **2006**, *84*, 69–82. [[CrossRef](#)] [[PubMed](#)]
43. Strosberg, J.R.; Halfdanarson, T.R.; Bellizzi, A.M.; Chan, J.A.; Dillon, J.; Heaney, A.P.; Kunz, P.L.; O’Dorisio, T.M.; Salem, R.; Segelov, E. The North American Neuroendocrine Society (NANETS) consensus guidelines for surveillance and medical management of midgut neuroendocrine tumors. *Pancreas* **2017**, *46*, 707. [[CrossRef](#)] [[PubMed](#)]
44. Rinzivillo, M.; Capurso, G.; Campana, D.; Fazio, N.; Panzuto, F.; Spada, F.; Cicchese, N.; Partelli, S.; Tomassetti, P.; Falconi, M. Risk and protective factors for small intestine neuroendocrine tumors: A prospective case-control study. *Neuroendocrinology* **2016**, *103*, 531–537. [[CrossRef](#)] [[PubMed](#)]
45. Escobar, K.M.; Villardón, J.L.V.; Cano, D.F.A.; Siteneski, A. Clinical related factors to neuroendocrine tumors in Ecuadorian patients: A logistic biplot approach. *Investig. Clin.* **2022**, *63*, 19–31.
46. Sackstein, P.E.; O’Neil, D.S.; Neugut, A.I.; Chabot, J.; Fojo, T. Epidemiologic trends in neuroendocrine tumors: An examination of incidence rates and survival of specific patient subgroups over the past 20 years. *Seminars Oncol.* **2018**, *45*, 249–258. [[CrossRef](#)] [[PubMed](#)]
47. Zatelli, M.C.; Grossrubatscher, E.M.; Guadagno, E.; Sciammarella, C.; Faggiano, A.; Colao, A. Circulating tumor cells and miRNAs as prognostic markers in neuroendocrine neoplasms. *Endocr. Relat. Cancer* **2017**, *24*, R223–R237. [[CrossRef](#)] [[PubMed](#)]
48. Nagtegaal, I.D.; Odze, R.D.; Klimstra, D.; Paradis, V.; Rugge, M.; Schirmacher, P.; Washington, K.M.; Carneiro, F.; Cree, I.A. WHO Classification of Tumours Editorial Board. The 2019 WHO classification of tumours of the digestive system. *Histopathology* **2020**, *76*, 182–188. [[CrossRef](#)]





### *6.3 Clinical Related Factors to Neuroendocrine Tumors in Ecuadorian Patients: A Logistic Biplot Approach*



---

---

## Clinical related factors to neuroendocrine tumors in Ecuadorian patients: a logistic biplot approach.

*Karime Montes Escobar*<sup>1,2</sup>, *José Luis Vicente Villardón*<sup>2</sup>, *Daniel Fabricio Alarcón Cano*<sup>3</sup>  
and *Aline Siteneski*<sup>4,5</sup>

<sup>1</sup>Departamento de Matemática y Estadística. Instituto de Ciencias Básicas.  
Universidad Técnica de Manabí, Portoviejo, Ecuador.

<sup>2</sup>Departamento de Estadística. Universidad de Salamanca, Salamanca, España.

<sup>3</sup>Departamento de Docencia e Investigación. SOLCA, Manabí, Ecuador

<sup>4</sup>Instituto de Investigación, Universidad Técnica de Manabí, Portoviejo, Ecuador.

<sup>5</sup>Facultad de Ciencias de la Salud, Carrera de Medicina. Universidad Técnica de Manabí,  
Portoviejo, Ecuador.

**Key words:** neuroendocrine tumors; cancer symptoms; statistical method; logistic biplot.

**Abstract.** Neuroendocrine tumors (NETs) are relative rare, affecting neuroendocrine cells throughout the body. Most tumors are diagnosed at advanced stages. NETs prevalence has increased in the last years but there is little data available in developing countries. The aim of this study was to describe symptoms associated with NETs in patients of the Society for the Fight Against Cancer (SOLCA) in Ecuador from 2005 to 2020; using logistic biplots, in a hospital database, generating binary responses (presence/absence) relevant to this study. The results showed that the mean age was 59 and the study showed no difference in prevalence between genders. NETs were mainly found in lungs (19%), followed by stomach (18%) and skin (9%). Most patients had pathological diagnosis G2 and G3 (30% and 70%, respectively). Symptoms as cough, dyspnea, weight loss, diarrhea, constipation, abdominal pain, dyspepsia, hypertensive crisis, distended abdomen and intestinal obstruction had p values <0.05. Additionally, the statistical analysis showed that cough and intestinal obstruction were also common, bearing in mind that patients had most frequent NETs in the lungs and skin. In summary, our results indicate that symptoms of NETs patients were positively associated with lung and skin. Further investigation is needed focusing on the type of NETs and their symptoms in order to establish an early marker for diagnosis.

## **Factores clínicos asociados a tumores neuroendocrinos en pacientes ecuatorianos: un análisis biplot logístico.**

*Invest Clin* 2022; 63 (1): 19 – 31

**Palabras clave:** tumores neuroendocrinos; síntomas de cáncer; método estadístico; biplot logístico.

**Resumen.** Los tumores neuroendocrinos (TNE) son relativamente raros y afectan a las células neuroendocrinas de todo el cuerpo. La mayoría de los tumores se diagnostican en etapas avanzadas. La prevalencia de los TNE ha aumentado en los últimos años, pero hay pocos datos en los países en desarrollo. El objetivo de este estudio fue determinar los síntomas asociados a los TNE en pacientes de la Sociedad de Lucha contra el Cáncer (SOLCA) en Ecuador entre 2005 y 2020, utilizando biplots logísticos en una base de datos hospitalaria, generando respuestas binarias (presencia / ausencia) relevantes para este estudio. Los resultados mostraron que la edad promedio era de 59 años y el estudio no encontró diferencias en la prevalencia entre géneros. Los TNE se encontraron con mayor frecuencia en los pulmones (19%), seguidos del estómago (18%) y piel (9%). La mayoría de los pacientes tenían diagnóstico patológico G2 y G3 (30% y 70% respectivamente). Los síntomas como tos, disnea, pérdida de peso, diarrea, estreñimiento, dolor abdominal, dispepsia, crisis hipertensiva, abdomen distendido y obstrucción intestinal tuvieron valores de  $p < 0,05$ . Además, el análisis estadístico mostró que la tos y la obstrucción intestinal también eran comunes, teniendo en cuenta que los pacientes tenían TNE más frecuentes en los pulmones y la piel. En resumen, nuestros resultados indican que los síntomas de los pacientes con TNE se asociaron positivamente con los pulmones y la piel. Se necesitan más investigaciones que se centren en el tipo de TNE y sus síntomas a fin de establecer un marcador más temprano para el diagnóstico.

*Received:* 13-08-2021    *Accepted:* 03-09-2021

### **INTRODUCTION**

Neuroendocrine tumors (NETs) are a group of malignant tumors appearing from neuroendocrine cells throughout the body that are defined as epithelial neoplasms with predominantly neuroendocrine differentiation<sup>1</sup>. NETs are commonly located in the pancreas, digestive tract, and the lungs although they can appear in any organ<sup>2</sup>. NETs grading is probably the most useful instrument for providing prognostic infor-

mation for clinicians<sup>3</sup>. Currently, the classification recognizes that neuroendocrine cancer at any anatomical site is either well-differentiated defined as NETs, and poorly differentiated defined as neuroendocrine carcinoma<sup>3</sup>. Severity may be evaluated by grading the tumor between 1 and 3 and per proliferation fraction as measured by mitotic count by the Ki67-positive percentage<sup>4,5</sup>. Approximately one-third of the high-grade group are the most aggressive NETs fraction, predicting either a worst event-free

survival or overall survival compared with grade 2 NETs<sup>6</sup>. Unfortunately, most patients are diagnosed in late stage of NETs<sup>7</sup>.

NETs symptoms vary depending on biological aggressiveness and the anatomical location of the tumor<sup>8</sup>. Vague and nonspecific symptoms may cause considerable delay and difficulty in diagnosis<sup>9</sup>. The clinical presentation of NETs is generally related to the secretion of hormones that may indicate carcinoid syndrome or hormonal symptoms<sup>10</sup>. Carcinoid syndrome commonly cause diarrhea, flushing, fatigue<sup>9</sup>, abdominal pain and intermitted bowel obstruction<sup>11</sup>. It is worth mentioning that the number of bowel movements and flushing episodes significantly affect the quality of life of NETs patients<sup>10</sup>. The intensity of these symptoms can vary widely, depending primarily on the site of disease and grade of tumor. More studies are necessary for clinicians to become familiar with the natural history and patterns of disease progression, which are characteristic of NETs<sup>7</sup>. Therefore, the precocious diagnosis of tumor may increase survival and improve NETs patient's quality of life.

In Ecuador, the Cancer Fighting Society (SOLCA) Quito Nucleus and National Tumor Registry exists since 1984 (Minist Salud Pública) and cares for approximately 31% of the oncological cases of the country<sup>12</sup>. In Ecuador, the clinical management of NETs is significantly lower, compared to that of the United States and Europe and there are few specialized centers managing a multidisciplinary environment<sup>13</sup>. Despite the high prevalence and severity of NETs, this tumor has not been previously studied in people from Ecuador to the best of our knowledge. Consequently, the aim of this research was to investigate the presence of NETs patients in the hospital database over fifteen years. Based on the above, a logistic biplot is proposed as an alternative to identify the most frequent symptoms in patients with NET using binary variables (presence / absence) with binomial distribution<sup>14</sup>. This analysis allows us to obtain a graphic representation of the NETs types and the main associated

symptoms. In this sense, the approach that provides a better interpretation to relate the type of tumor with its symptoms, taking into account the comparison with the classic classification / ordering techniques, is provided by the biplot methods<sup>15,16</sup>, where a simultaneous graph represents the columns (variables) and the rows (individuals) of the given data matrix.

## MATERIAL AND METHODS

### Collected database

The study was done using the hospital database of SOLCA Portoviejo, in the Province of Manabí, Ecuador, analyzing patients with NETs diagnosis between 2005 and 2020.

### Statistical analysis

The main problem with our data is the very limited number of patients and the high number of variables (symptoms, types of tumor, locations, etc...) that prevents from using standard methods as logistic regression. An exploratory multivariate technique would be more adequate to explore possible patterns before trying to apply a more formal model using a relevant subset of symptoms obtained from it. This kind of techniques have also the advantage of showing the interrelations among all the variables.

For the exploratory analysis of our data, we used a multivariate technique called "logistic biplot"<sup>14</sup> that was specifically designed to treat binary data, as well as more traditional analyses. Although logistic biplots have not been extensively used with medical data, some authors<sup>17</sup> propose these kind of multivariate techniques as a mean to perform an exploratory analysis previous to a more formal analysis. Other authors<sup>18</sup> also propose the technique as an analysis of the residuals of a model, to check for the goodness of fit or the violation of some previous hypothesis about the data.

Because logistic biplots are less known, we include here a brief description. A biplot is a graphical representation of a data

matrix, containing the measures subjects on numerical variables<sup>15</sup>, that jointly represents subjects and variables on the same plot. We have patients with different NETs and related symptoms (presence or absence) organized in a binary data matrix (0 for presence, 1 for absence).

The adequate method for binary data is the logistic biplot, originally proposed by Vicente-Villardón *et al*<sup>14</sup> and later extended by Demey *et al*<sup>19</sup> to binary matrices where the number of variables is high compared to the number of individuals. This extension was applied to the investigation of the single nucleotide polymorphisms related to different racial groups of people with data from the HAPMAP project<sup>20</sup>. The method has been applied in different fields as sustainability<sup>21</sup>, dermatology<sup>22</sup> and lately it has also been extended to deal with nominal data<sup>23</sup>.

Logistic biplots represent individuals as points on a Euclidean map (scattergram) and variables as directions (arrows) on the same map. Patients with similar combinations of symptoms tend to group together while distinct individuals tend to pull apart. The vectors for the variables on the plot are the directions more correlated to the presence of the symptom, so we can infer which symptoms are responsible for the grouping of the patients. To interpret the symptoms associated to the grouping of patients we project the points onto the direction of the vectors - the further (in the direction of the arrow) the point is projected, the higher is the probability of the symptom. The origin of the vector is the point predicting probability 0.5 and the arrow indicates the direction of increasing probabilities.

Angles among vectors can help with the identification of the relationships among the symptoms: small acute angles mean strong positive relations; plane angles (Formed by two right angles 180°), negative relations; and straight angles (formed by two perpendicular lines 90°), are not related. The position of the vectors can also help identifying the main gradients (or latent traits) summarizing the variation of the symptoms and

its relation to the different locations of the cancer. It is called logistic biplot because the relation of the symptoms and the traits (dimensions) is logistic, it is like a Principal Component Analysis (or biplot) for binary data in which the relation to the components is logistic rather than linear. When data are binary, like those obtained in the analysis of neuroendocrine tumors, Classical Linear Biplots and Principal Components Analysis (PCA) are not suitable because the response along the dimensions is linear. This is the same reason why linear regression is not appropriate for binary or categorical data<sup>14, 21</sup>.

The MULTBILOT<sup>24</sup> and MultBiplotR<sup>25</sup> packages were used to perform the calculations and obtain the graphical representations. For this study, we first performed a Principal Coordinates Analysis (Classical Multidimensional Scaling), with the simple matching coefficient, and then fitting separated logistic regressions to represent the directions on the Euclidean map<sup>19</sup>. The interpretation of the relevance of the symptoms to classify the patients was evaluated by traditional indicators for logistic regressions as Wald tests, pseudo R-squared coefficients or percentages of accurate classification.

## RESULTS

### Population characteristics and NETs symptoms

A total of 94 adult patients (aged 19 to 95 years-old) were diagnosed with NETs between 2005 and 2020 and included in our research. The most frequent primary location was the lungs (19%), followed by stomach (18%), rectum (8%) and skin (8%). Unknown primary location represents 8% of the tumors studied. The set symptoms present in our patients with different types of NETs are shown in Table 1.

Table 2 represent sociodemographic characteristic, mortality, primary site and histologic grade of NETs patients. The mean age in this study was 59 years-old and the prevalence was higher in men (57% versus

**Table 1**

Represent total symptoms found in NETs patients among eleven years in SOLCA, Manabí.

Primary location	Patient	Percentage
LUNG	18	19%
STOMACH	17	18%
RECTAL	8	9%
SKIN	8	9%
UNKNOWN PRIMARY	8	9%
PANCREAS	5	5%
GANGLIA	4	4%
LIVER	4	4%
DUODENUM	3	3%
APPENDIX	2	2%
CECAL APPENDIX	2	2%
CERVIX	2	2%
COLON	2	2%
RETROPERITONEUM	2	2%
RIGHT COLON	2	2%
COLEDOCO	1	1%
ENDOMETRIC	1	1%
ILEUM	1	1%
LEFT COLON	1	1%
SOFT TISSUES	1	1%
UTERINE NECK	1	1%
VESICAL TUMOR	1	1%
TOTAL	94	100%

43%). The most prevalent stage in NETs patients was G3 (70%) while G1 and G2 represent 30% of the patients.

**Logistic Biplot in types of NETs and our symptoms in patients of SOLCA**

Fig. 1 shows the result of applying the logistic biplot to the binary matrix of 94 patients and 37 symptoms. Each patient (points) has been labeled using the primary location of its tumor and each symptom (arrow) using the name of the symptom. Only the symptoms significantly related to the patient’s configuration were retained on the plot (See Table

**Table 2**

Represent sociodemographic characteristic, mortality, primary site and histologic grade of NETs patients.

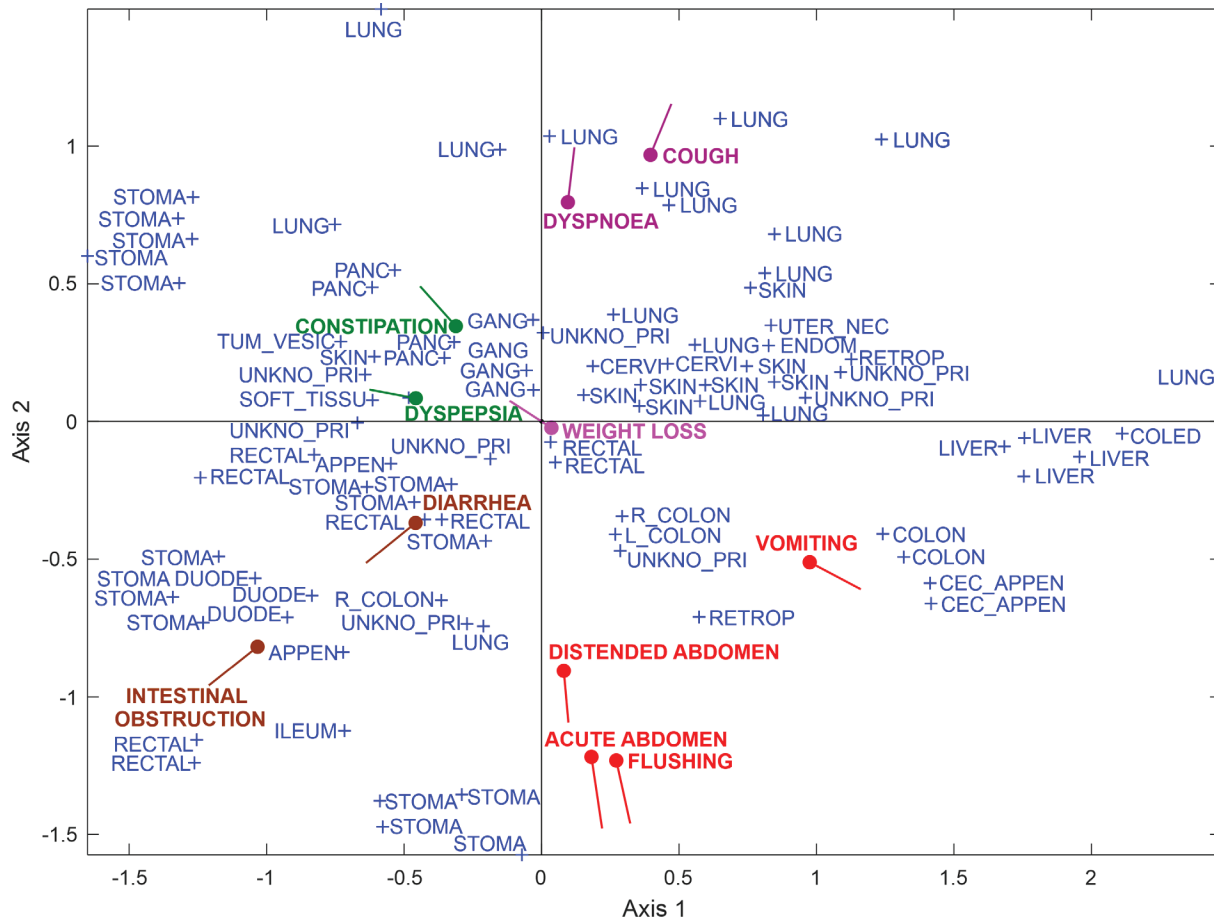
Variable	Parameters	No. Patients (n)	Percentage
Age	<40	11	12%
	40-54	22	23%
	55-70	30	32%
	> 70	31	33%
Sex	Man	54	57%
	Woman	40	43%
Mortality	Alive	38	40%
	Dead	56	60%
Histologic grade	Type 1 y 2	28	30%
	Type 3	66	70%

3). We observe that most patients having lungs as the primary location are on the top of the plot, while tumors related to the digestive system (rectum, stomach, right colon, ileum and pancreas) are on the bottom part and well separated from the “lung” group, so we can conclude that the analysis has some power to classify the tumors.

When projecting the patients’ points onto the directions for the symptoms, we obtain the expected probability that the patient has the symptom. In order to predict the expected presence when the expected probability is higher than 0.5. For each arrow, we only represent the point predicting 0.5 and the direction of increasing probabilities. Thus, we can correctly predict 95.75% of the presences and absences. All the individual symptoms have percentages of correct predictions over 94%. The associated R-squared values are all over 70% and the information of each variable can be accurately interpreted. We can conclude that the representation is a good picture of our data matrix and can be used for exploration.

To search for the symptoms that characterize each group, we can look at the di-





**Fig. 1.** Biplot representation of the relations among the type of disease and the symptom that determine the groups structure in the simulation. Small acute angles mean strong positive relations; plane angles (Formed by two right angles  $180^\circ$ ), negative relations; and straight angles (Formed by two perpendicular lines  $90^\circ$ ), are not related. Sign (+) represents each of the individuals. Segments (lines) represent each of the variables (symptoms). Point represents the value of 0.05. Direction of the vector predicts the increase in the probability. The colors: red, brown, green, purple represents group the variables with respect to the individuals.

rection they point. For example, *Dyspnea* and *Cough* are present in patients for which the primary location is *lungs*, while *Diarrhea*, *Stomach pains*, *Distended abdomen*, *Hypertensive Crisis* and *Intestinal obstruction* point to the direction of the tumors related to the digestive system. Both sets of symptoms point in opposite direction meaning that they are inversely related, that is, the symptoms present in one group are absent in the other. Together, they de-

fine the gradient that separate both types of tumors. *Dyspnea* and *Cough* point to the same direction, meaning that they are highly positively correlated and appear together in some patients. *Weight Loss* has a very good fit but is not clearly characterized in any of the groups. The same is true to *Dyspepsia* and *Constipation*. The three are closely related to the first dimension, but they are less important to separate between groups of tumors.

**Table 3**  
Goodness of adjunct for the main symptoms in NETs patients.

Primary symptom	Deviance	p-value	R2	% Correct
Vomiting	5.326	2,22045E-14	0,9390	100
Dyspepsia	6.467	0	0,8622	91.489
Flushing	4.791	1,78129E-07	0,8474	100
Distended Abdomen	5.922	1,69207E-11	0,8167	97.872
Diarrhea	4.848	0	0,8136	94.681
Weight Loss	6.146	0	0,8129	94.681
Cough	5.591	8,95759E-10	0,7999	97.872
Intestinal Obstruction	4.953	1,40092E-09	0,7847	94.681
Dyspnoea	5.574	6,51029E-11	0,7410	94.681
Acute Abdomen	4.306	1,96949E-07	0,7350	98.936
Constipation	5.730	1,26565E-14	0,7266	92.553

## DISCUSSION

Neuroendocrine cells have endocrine and neural properties and are widely distributed in the body<sup>3</sup>. Many organs contain this cell phenotype and may be affected by NETs. The tumor has a particular morphofunctional characteristic, such as the hormones production, and the immunohistochemical features staining pattern with chromogranin, synaptophysin, and neuron-specific enolase<sup>26</sup>. Moreover, the World Health Organization (WHO) advises characterizing NETs by proliferation fraction as measured by either mitotic count or (more easily) by the Ki67-positive percentage (Ki67 index)<sup>6</sup>. It is worth mentioning that, epidemiological studies indicate that the most frequent source of metastasis in NETs are the small intestine, lungs and colon<sup>27</sup>. In the present retrospective databased study, we found lungs and stomach as more frequent NETs followed by rectum, and skin. Here, the diagnosis was through synaptophysin, chromagranin or Ki67 by biopsy, and most patients had several pathological diagnostics. Despite the malignancies of the tumor, to the best of our knowledge, this is the first evidence about NETs in patients from the Ecuadorian coast.

An approach that facilitates the genetic interpretation, compared to the classic techniques of classification/ordination, is provided by the Biplot methods<sup>28, 15, 16</sup>, that is, a simultaneous graphical representation of the rows (individuals) and the columns (variables) of a given data matrix. The main uses are exploratory, although it has also been used as a graphical representation for more formal models<sup>29</sup>. The biplot can be fitted by performing alternating regressions and interpolations<sup>30-32</sup>. However, when data are binary, like those obtained in the analysis of molecular information, Classical Linear Biplots and Principal Components Analysis (PCA) are not suitable because the response along the dimensions is linear. This is the same reason why linear regression is not appropriate for binary or categorical data.

There is a significant delay between the symptoms onset and diagnosis of NETs and, consequently, most patients are diagnosed at a late stage of the disease<sup>7</sup>. Our results demonstrated that most patients were diagnosed in histologic grade 3, metastatic or advanced NETs and do not survive. Previous works consider NETs as a low-grade, usually non-functioning, malignant cancer characterized by long survival of pa-

tients with prognosis depending on grade and stage<sup>33</sup>. In fact, WHO in 2017 classified NETs with a Ki67 >20% as either grade 3 NET (well-differentiated) or as grade 3 NEC (poorly differentiated) on the basis of morphological characteristics as necrosis and differentiation<sup>34</sup>. NETs account for about 0.5% of all currently diagnosed malignancies<sup>35</sup>. Given this scenario, our work presents most NETs cases registered in the advanced stage, which could explain, at least in part, the death rates from NETs found in the present data-based study. Furthermore, the retrospective characteristic of the study with a duration of fifteen years should be considered. Indeed, NETs origin and ranking grade can define patient survival if previously reported 103 months from disease originating in the small intestine and only 14 months for cancers originating in the colon<sup>36</sup>. A severity nation-wide multi-center study in Korea reported a high G1 percentage of 92.3%<sup>37</sup>. However, research in Kentucky found a population with 67% in G3<sup>38</sup>. Reinforcing this notion, our study reports most patients in G3 - 70%. Nevertheless, of special importance is that lower grade NETs do not indicate a better outcome in terms of survival<sup>38</sup>.

Given this background, retrospective studies in different populations are important as an alert for the early diagnosis of disease. In the present study, we decided to investigate NETs more frequently and its associated symptoms. Lungs were the most frequently NETs reported in our study and it was associated with the symptoms of cough and dyspnea. In line with our results, previous research as a nation-wide study from the Netherlands and SEER database reported pulmonary NETs as being the most prevalent<sup>36</sup>. The NETs are often misdiagnosed before the cause of the patient's symptoms is finally diagnosed and managed appropriately<sup>9</sup>.

Our results demonstrated that after lung and stomach, the most incidence of NETs was

in the rectum and skin. We found a significantly positive relationship between the most frequent symptoms and the NETs found during or study. Therefore, symptoms such as diarrhea, constipation, dyspepsia, and intestinal obstruction in addition to weight loss were more prevalent. We should note that our research found that NETs in the stomach and rectum were associated with syndromes such as carcinoid syndrome. It has been previously described that carcinoid syndrome occurs in 8% to 35% of NETs patients<sup>39</sup>. This event appears when vasoactive hormones secreted by metastases (i.e. serotonin, histamine, or tachykinins), are no longer metabolized and inactivated by the liver and reach the general circulation<sup>40</sup>. The classic symptom of episodic diarrhea was previously described by more than 70% of patients as consequence of carcinoid syndrome<sup>41</sup>. This data point is similar with what we found in our study. Additionally, it is important to note that the clinical presence of diarrhea and/or abdominal pain can often lead to misdiagnosis, confusing carcinoid syndrome as small bowel obstructions or irritable bowel disease<sup>42</sup>. Unfortunately, there is a delay between the onset of symptoms and the diagnosis of carcinoid tumor; the median time reported range from 2 to 20 years<sup>43</sup>. It is well established that early diagnosis of any type of cancer can improve the quality of life and survival and patients with carcinoid syndrome have a significantly worse quality of life than patients with only NETs<sup>44</sup>. Supporting this fact, our findings underscore the need to further identify all symptoms associated with NETs in an effort to aid physicians to speed up the diagnosis.

Unsurprisingly, the lungs are the most frequently found NET during the fifteen years studied. Importantly, we demonstrated that skin was the fourth most frequently NETs discovered. Based on previous work, NETs in the skin or soft tissue delay late manifestations of disease when represent metastases from other organs<sup>45</sup>. However, notably, according to the World Health Organization

books on skin and soft tissue tumors stated that primary cutaneous and soft tissue NETs are extremely rare, except for Merkel cell carcinoma of the skin <sup>46</sup>. Therefore, only a few case studies of soft tissue NETs are available in the literature <sup>47</sup>. Our findings show that 8% of all cases in the fifteen years studied were skin NET, mostly in elderly men. Supporting this finding, previous studies showed that Merkel cell carcinoma are most commonly found in elderly and in sun-exposed skin <sup>48, 49</sup>, patients typically have fair-skin and the median age of diagnosis of 65 years <sup>49</sup>. It is interesting to note that the most common symptom reported in our work is skin damage with unknown or asymptomatic causes. This evidence is corroborated by other studies. Despite the diversity of skin manifestations, the tumor is typically painless and develops as a small, less than 2 centimeters, violaceous papule, plaque, cyst, or infiltrative nodule <sup>50</sup>. The tumor has a multifactorial cause and the exact mechanism by which it acts in the organism it is not fully elucidated, but Merkel cell polyomavirus is thought to be a major contributor to the pathogenesis of this malignancy <sup>51</sup>.

Altogether, the results provided suggest that the most frequent clinical symptoms (cough and intestinal obstruction) were positively related to the type of NETs. When looking at the prevalence of NETs, the lungs were the most frequent site, which reinforces the hypothesis of the vulnerability of this organ. Further research is needed and should focus on identifying symptoms of NETs with the objective of early diagnosis. Furthermore, the prevalence of NETs specific in each population must be underscored in future research.

The logistic biplot technique allows the generation of relevant information based on clinical antecedents, as it was for this study on the presence or absence of the disease, taking into account patients with different neuroendocrine tumors.

## ACKNOWLEDGMENTS AND FUNDING

This study was supported by the Department of Mathematics and Statistics. Institute of Basic Sciences and Research Institute of the Technical University of Manabí. The authors would like to thank the SOLCA Portoviejo for support and facilitate data and Robin Siteneski for English corrections.

### Declaration of competing interest

The authors declare that no financial support or compensation has been received from any individual or corporate entity over the past three years for research or professional service and there is no personal financial holding that could be perceived as constituting a potential conflict of interest.

### Authors' ORCID numbers

- Karime Montes Escobar (KME)  
0000-0002-9555-0392
- José Luis Vicente Villardon (JLVV)  
0000-0003-1416-6813
- Daniel Fabricio Alarcón Cano (DFAC) 0000-0002-9680-1050
- Aline Sitenesky (AS)  
0000-0001-6692-7253

### Authors' contribution

- Interpretation of data, analysis and writing the manuscript (KME)
- Supervision and writing (JLVV)
- Data collected (DFAC)
- Writing the manuscript and prepared the final draft (AS).
- All authors have read and approved the final version of submitted manuscript.

## REFERENCES

1. Klimstra DS, Modlin IR, Coppola D, Lloyd RV, Suster S. The pathologic classification of neuroendocrine tumors: a review of nomenclature, grading, and staging systems *Pancreas* 2010; 39(6):707-12. doi: 10.1097/MPA.0b013e3181ec124e.
2. Zandee WT, de Herder WW. The evolution of neuroendocrine tumor treatment reflected by ENETS Guidelines. *Neuroendocrinol* 2018;106(4):357-365. doi: 10.1159/000486096.
3. Rindi G, Wiedenmann B. Neuroendocrine neoplasia of the gastrointestinal tract revisited: towards precision medicine. *Nat Rev Endocrinol* 2020;16(10):590-607. doi: 10.1038/s41574-020-0391-3.
4. Rindi G, Klimstra DS, Abedi-Ardekani B, Asa SL, Bosman FT, Brambilla E, Busam KJ, de Krijger RR, Dietel M, El-Naggar AK, Fernandez-Cuesta L, Klöppel G, McCluggage WG, Moch H, Ohgaki H, Rakha EA, Reed NS, Rous BA, Sasano H, Scarpa A, Scoazec JY, Travis WD, Tallini G, Trouillas J, van Krieken JH, Cree IA. A common classification framework for neuroendocrine neoplasms: an International Agency for Research on Cancer (IARC) and World Health Organization (WHO) expert consensus proposal. *Mod Pathol* 2018;31(12):1770-1786. doi:10.1038/s41379-018-0110-y.
5. Hijioka S, Hosoda W, Mizuno N, Hara K, Imaoka H, Bhatia V, Mekky MA, Tajika M, Tanaka T, Ishihara M, Yogi T, Tsutumi H, Fujiyoshi T, Sato T, Hieda N, Yoshida T, Okuno N, Shimizu Y, Yatabe Y, Niwa Y, Yamao K. Does the WHO 2010 classification of pancreatic neuroendocrine neoplasms accurately characterize pancreatic neuroendocrine carcinomas? *J Gastroenterol* 2015;50(5):564-72. doi: 10.1007/s00535-014-0987-2.
6. Rindi G, Klersy C, Albarello L, Baudin E, Bianchi A, Buchler MW, Caplin M, Couvelard A, Cros J, de Herder WW, Delle Fave G, Doglioni C, Federspiel B, Fischer L, Fusai G, Gavazzi F, Hansen CP, Inzani F, Jann H, Komminoth P, Knigge UP, Landoni L, La Rosa S, Lawlor RT, Luong TV, Marinoni I, Panzuto F, Pape UF, Partelli S, Perren A, Rinzivillo M, Rubini C, Ruszniewski P, Scarpa A, Schmitt A, Schinzari G, Scoazec JY, Sessa F, Solcia E, Spaggiari P, Toumpanakis C, Vanoli A, Wiedenmann B, Zamboni G, Zandee WT, Zerbi A, Falconi M. Competitive testing of the WHO 2010 versus the WHO 2017 grading of pancreatic neuroendocrine neoplasms: data from a Large International Cohort Study. *Neuroendocrinol* 2018;107(4):375-386. doi: 10.1159/000494355.
7. Yao JC, Hassan M, Phan A, Dagohoy C, Leary C, Mares JE, Abdalla EK, Fleming JB, Vauthey JN, Rashid A, Evans DB. One hundred years after "carcinoid": epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol* 2008; 26(18):3063-72. doi:10.1200/JCO.2007.15.4377.
8. Oronsky B, Ma PC, Morgensztern D, Carter CA. Nothing but NET: a review of neuroendocrine tumors and carcinomas. *Neoplasia* 2017;19(12):991-1002. doi: 10.1016/j.neo.2017.09.002.
9. Chauhan A, Kohn E, Del Rivero J. Neuroendocrine tumors-less well known, often misunderstood, and rapidly growing in incidence. *JAMA Oncol* 2020;6(1):21-22. doi: 10.1001/jamaoncol.2019.4568.
10. Pearman TP, Beaumont JL, Cella D, Neary MP, Yao J. Health-related quality of life in patients with neuroendocrine tumors: an investigation of treatment type, disease status, and symptom burden. *Support Care Cancer* 2016;24(9):3695-703. doi:10.1007/s00520-016-3189-z.
11. Kulke MH, Mayer RJ. Carcinoid tumors. *N Engl J Med* 1999;340(11):858-68. doi:10.1056/NEJM199903183401107.
12. Corral Cordero F, Cueva Ayala P, Yépez Maldonado J, Tarupi Montenegro W. Trends in cancer incidence and mortality over three decades in Quito - Ecuador. *Colomb Med* 2018;49(1):35-41. doi: 10.25100/cm.v49i1.3785.
13. Rindi G, Klöppel G, Couvelard A, Komminoth P, Körner M, Lopes JM, McNicol AM, Nilsson O, Perren A, Scarpa A, Scoazec JY, Wiedenmann B. TNM staging of midgut and hindgut (neuro) endocrine tumors: a



- consensus proposal including a grading system. *Virchows Arch* 2007;451(4):757-62. doi:10.1007/s00428-007-0452-1.
14. **Vicente-Villardón JL, Galindo-Villardón MP, Blázquez-Zaballos A.** Logistic biplots. Multiple correspondence analysis and related methods. London: Chapman & Hall. 2006;503-21.
  15. **Gabriel KR.** The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 1971; 58(3), 453-467. doi.org/10.1093/biomet/58.3.453.
  16. **Sharov AA, Dudekula DB, Ko MSH.** A web-based tool for principal component and significance analysis of microarray data. *Bioinformatics* 2005; 21(10), 2548-2549 doi:10.1093/bioinformatics/bti343.
  17. **Bravata DM, Shojania KG, Olkin I, Ravah A.** CoPlot: a tool for visualizing multivariate data in medicine. *Stat Med* 2008;27(12):2234-47. doi: 10.1002/sim.3078.
  18. **Sepúlveda R, Vicente-Villardón JL, Galindo MP.** The Biplot as a diagnostic tool of local dependence in latent class models. A medical application. *Stat Med* 2008;27(11):1855-69. doi: 10.1002/sim.3194.
  19. **Demey JR, Vicente-Villardón JL, Galindo-Villardón MP, Zambrano AY.** Identifying molecular markers associated with classification of genotypes by External Logistic Biplots. *Bioinformatics* 2008; 24(24), 2832-2838. doi:10.1093/bioinformatics/btn552.
  20. **Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu FL, Yang HM, et al.** The international HapMap project. 2003. doi:10.1038/nature02168.
  21. **Gallego-Alvarez I, Ortas E, Vicente-Villardón JL, Álvarez Etxebarria I.** Institutional constraints, stakeholder pressure and corporate environmental reporting policies. *Bus Stratég Environ* 2017; 26(6), 807-825. doi.org/10.1002/bse.1952.
  22. **Cañueto J, Cardeñoso-Álvarez E, García-Hernández JL, Galindo-Villardón P, Vicente-Galindo P, Vicente-Villardón JL, Alonso-López D, De Las Rivas J, Valero J, Moyano-Sanz E, Fernández-López E, Mao JH, Castellanos-Martín A, Román-Curto C, Pérez-Losada J.** MicroRNA (miR)-203 and miR-205 expression patterns identify subgroups of prognosis in cutaneous squamous cell carcinoma. *Br J Dermatol* 2017;177(1):168-178. doi: 10.1111/bjd.15236.
  23. **Hernández-Sánchez JC, Vicente-Villardón JL.** Logistic biplot for nominal data. *Adv Data Anal Classif* 11. 2017; 11(2), 307-326. doi.org/10.1007/s11634-016-0249-7.
  24. **Vicente-Villardón, J.L.** MULTBILOT: A package for Multivariate Analysis using Biplots. Departamento de Estadística. Universidad de Salamanca. 2010. (<http://biplot.usal.es/ClassicalBiplot/index.html>).
  25. **Vicente-Villardón, J.L.** MultBiplotR: Multivariate Analysis using Biplots. R package version 0.1. 0. Website <http://biplot.usal.es/multbiplot/multbiplot-in-r/> [accessed 20 November 2019].
  26. **Chen TY, Morrison AO, Susa J, Cockerell CJ.** Primary low-grade neuroendocrine carcinoma of the skin: An exceedingly rare entity. *J Cutan Pathol* 2017;44(11):978-81. doi: 10.1111/cup.13028.
  27. **Riihimäki M, Hemminki A, Sundquist K, Sundquist J, Hemminki K.** The epidemiology of metastases in neuroendocrine tumors. *Int J Cancer* 2016; 139(12):2679-2686. doi:10.1002/ijc.30400.
  28. **Chapman S, Schenk P, Kazan K, Manners, J.** Using biplots to interpret gene expression patterns in plants. *Bioinformatics* 2002; 18(1), 202-204. doi.org/10.1093/bioinformatics/18.1.202.
  29. **Gabriel, K. R.** Generalised bilinear regression. *Biometrika* 1998; 85(3), 689-700. doi.org/10.1093/biomet/85.3.689
  30. **Gabriel KR, Zamir S.** Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics* 1979; 21 : 489-498.
  31. **Gower JC, Hand DJ.** *Biplots* 1995; 54. CRC Press.
  32. **Jongman, E, & Jongman, S R R** *Data analysis in community and landscape ecology.* Cambridge University Press, 1995.
  33. **Vanoli A, Albarello L, Uncini S, Fassan M, Grillo F, Di Sabatino A, Martino M, Pasquali C, Milanetto AC, Falconi M, Partelli S, Doglioni C, Schiavo-Lena M, Brambilla T, Pietrabissa A, Sessa F, Capella C, Rindi G, La Rosa S, Solcia E, Paulli M.** Neuroendocrine tumors (NETs) of the minor papilla/ampulla: analysis of 16 cases underlines

- homology with major ampulla NETs and differences from extra-ampullary duodenal NETs. *Am J Surg Pathol* 2019;43(6):725-736. doi:10.1097/PAS.0000000000001234.
34. **Lloyd R V, Osamura YR, Kloppel G, Rosai J.** WHO classification of tumours of endocrine organs. WHO Press 2017; WHO Classification of Tumours, *4th Edition, Volume 10*.
  35. **Taal BG, Visser O.** Epidemiology of neuroendocrine tumours. *Neuroendocrinol* 2004;80 (Suppl 1):3-7. doi: 10.1159/000080731.
  36. **Dasari A, Shen C, Halperin D, Zhao B, Zhou S, Xu Y, Shih T, Yao JC.** Trends in the incidence, prevalence, and survival outcomes in patients with neuroendocrine tumors in the United States. *JAMA Oncol* 2017;3(10):1335-1342. doi: 10.1001/jamaoncol.2017.0589.
  37. **Cho MY, Kim JM, Sohn JH, Kim MJ, Kim KM, Kim WH, Kim H, Kook MC, Park DY, Lee JH, Chang H, Jung ES, Kim HK, Jin SY, Choi JH, Gu MJ, Kim S, Kang MS, Cho CH, Park MI, Kang YK, Kim YW, Yoon SO, Bae HI, Joo M, Moon WS, Kang DY, Chang SJ.** Current trends of the incidence and pathological diagnosis of gastroenteropancreatic neuroendocrine tumors (GEP-NETs) in Korea 2000-2009: Multicenter Study. *Cancer Res Treat* 2012;44(3):157-165. doi: 10.4143/crt.2012.44.3.157.
  38. **Chauhan A, Yu Q, Ray N, Farooqui Z, Huang B, Durbin EB, Tucker T, Evers M, Arnold S, Anthony LB.** Global burden of neuroendocrine tumors and changing incidence in Kentucky. *Oncotarget* 2018;9(27):19245-19254. doi: 10.18632/oncotarget.24983.
  39. **Rorstad O.** Prognostic indicators for carcinoid neuroendocrine tumors of the gastrointestinal tract. *J Surg Oncol* 2005;89(3):151-160. doi: 10.1002/jso.20179.
  40. **Rupp AB, Ahmadjee A, Morshedzadeh JH, Ranjan R.** Carcinoid syndrome-induced ventricular tachycardia. *Case Rep Cardiol* 2016;2016:9142598. doi: 10.1155/2016/9142598.
  41. **McCormick D.** Carcinoid tumors and syndrome. *Gastroenterol Nurs* 2002;25(3):105-11. doi: 10.1097/00001610-200205000-00004.
  42. **Boudreaux JP, Klimstra DS, Hassan MM, Woltering EA, Jensen RT, Goldsmith SJ, Nutting C, Bushnell DL, Caplin ME, Yao JC.** North American Neuroendocrine Tumor Society (NANETS). The NANETS consensus guideline for the diagnosis and management of neuroendocrine tumors: well-differentiated neuroendocrine tumors of the jejunum, ileum, appendix, and cecum. *Pancreas* 2010;39(6):753-766. doi:10.1097/MPA.0b013e3181ebb2a5.
  43. **Toth-Fejel S, Pommier RF.** Relationships among delay of diagnosis, extent of disease, and survival in patients with abdominal carcinoid tumors. *Am J Surg* 2004;187(5):575-579. doi: 10.1016/j.amjsurg.2004.01.019.
  44. **Beaumont JL, Cella D, Phan AT, Choi S, Liu Z, Yao JC.** Comparison of health-related quality of life in patients with neuroendocrine tumors with quality of life in the general US population. *Pancreas* 2012;41(3):461-466. doi: 10.1097/MPA.0b013e3182328045.
  45. **Jedrych J, Pulitzer M.** Primary carcinoid tumor of the skin: a literature review. *Int J Surg Pathol* 2014;22(2):129-135. doi: 10.1177/1066896913516672.
  46. **Terada T.** Primary cutaneous neuroendocrine tumor (atypical carcinoid) expressing KIT and PDGFRA with myoepithelial differentiation: a case report with immunohistochemical and molecular genetic studies. *Int J Clin Exp Pathol* 2013;6(4):802-809.
  47. **Koo HS, Sohn YM, Park YK.** Sonographic appearance of a neuroendocrine tumor arising in the axilla: case report and literature review. *J Clin Ultrasound* 2014;42(1):30-32. doi: 10.1002/jcu.22024.
  48. **Santos-Juanes J, Fernández-Vega I, Fuentes N, Galache C, Coto-Segura P, Vivanco B, Astudillo A, Martínez-Cambor P.** Merkel cell carcinoma and Merkel cell polyomavirus: a systematic review and meta-analysis. *Br J Dermatol* 2015;173(1):42-49. doi: 10.1111/bjd.13870.
  49. **Daoud MA, Mete O, Al Habeeb A, Ghazarian D.** Neuroendocrine carcinoma of the skin--an updated review. *Semin Diagn Pathol* 2013;30(3):234-244. doi:10.1053/j.semdp.2013.07.002.

- 
50. Goto K, Anan T, Fukumoto T, Kimura T, Misago N. Carcinoid-like/labyrinthine pattern in sebaceous neoplasms represents a sebaceous mantle phenotype: immunohistochemical analysis of aberrant vimentin expression and cytokeratin 20-positive Merkel cell distribution. *Am J Dermatopathol* 2017;39(11):803-810. doi:10.1097/DAD.0000000000000806.
51. Leventhal JS, Braverman IM. Skin manifestations of endocrine and neuroendocrine tumors. *Semin Oncol* 2016;43(3):335-340. doi: 10.1053/j.seminoncol.2016.02.022.





## ***7. RESÚMENES DE LOS ARTÍCULOS***



*7.1 Frequency of Neuroendocrine Tumor Studies: Using  
Latent Dirichlet Allocation and Hj-Biplot Statistical  
Methods*



Frequency of Neuroendocrine tumors studies: Using Latent Dirichlet Allocation and HJ-Biplot statistical methods

---

**Resumen:** Los tumores neuroendocrinos (TNE) son graves y relativamente raros y pueden afectar a cualquier órgano del cuerpo humano. La prevalencia de los TNE ha aumentado en los últimos años, sin embargo, parece haber más datos sobre tipos particulares, aunque a pesar de los esfuerzos de diferentes guías, no hay consenso sobre cómo identificar diferentes tipos de TNE. En esta revisión, investigamos los países que publicaron más artículos sobre los TNE, los órganos afectados con mayor frecuencia y los temas relacionados más comunes. Este trabajo utilizó el método Latent Dirichlet Allocation (LDA) para identificar e interpretar información científica en relación con las categorías en un conjunto de documentos. También se utilizó el método HJ-Biplot para determinar la relación entre los temas analizados, tomando en consideración los años en estudio. En este estudio se realizó una revisión de la literatura de la cual se extrajeron un total de 7658 resúmenes de artículos científicos publicados entre 1981 y 2020. Estados Unidos, Alemania, Reino Unido, Francia e Italia publicaron la mayoría de los estudios sobre TNE, de los cuales los tumores pancreáticos fueron los más estudiados. Los cinco temas más frecuentes fueron t\_21 (beneficio clínico), t\_11 (tumores neuroendocrinos pancreáticos), t\_13 (pacientes un año después del tratamiento), t\_17 (pronóstico de supervivencia antes y después de la resección) y t\_3 (marcadores de carcinomas). Finalmente, los resultados obtenidos mediante la aplicación de un análisis multivariante bidireccional (HJ-biplot) permitieron una nueva interpretación: se agruparon los temas por año y se descubrió qué TNE eran los más relevantes y para qué años. Los temas más frecuentes encontrados en nuestra revisión destacaron la gravedad de los TNE: los pacientes tienen un mal pronóstico de supervivencia y una alta probabilidad de recidiva tumoral.

Los **objetivos** que se persiguen en este artículo son los siguientes:

Mediante técnicas multivariantes:

- Encontrar los países que publicaron más artículos sobre Tumores neuroendocrinos (TNE) durante el periodo en estudio.
- Detectar que órganos son los más afectados con mayor frecuencia en pacientes con TNE.
- Determinar que temas sobre tumores neuroendocrines se relacionan y durante que años.

**Metodología:** En primer lugar, se introducen los conceptos básicos de la Latent Dirichlet Allocation (LDA). Luego, el procedimiento se utiliza para identificar temas latentes y tendencias de investigación en tumores neuroendocrinos. A continuación, se definen los índices cuantitativos para poder explicar los resultados. Finalmente, HJ-biplot proporciona una

representación gráfica multivariante de filas y columnas de una matriz de datos en un subespacio de baja dimensión donde las posiciones relativas de los resultados son interpretables.

### **Identificación de temas de investigación:**

Para identificar los temas de investigación en NETs utilizando el método LDA, el estudio se dividió en cuatro etapas: Búsqueda de literatura, Preprocesamiento, Selección del número de temas o tópicos LDA y construcción del modelo y Etiquetado de los temas.

### **Búsqueda de literatura**

La base de datos PubMed se utilizó como fuente para este documento. Los criterios de inclusión se centraron en la selección de artículos de investigación, publicados en inglés, desde 1981 hasta diciembre de 2020. La búsqueda se realizó el 10 de julio de 2021 mediante la siguiente consulta:

“tumores neuroendocrinos [Título/Resumen] Y Artículo de revista [PT] Y 1981:2020[DP]”

Se utilizaron cuatro variables: autor, título, año de publicación y afiliación del autor. No se incluyeron libros, reseñas, capítulos de libros, literatura gris e informes para evitar ruido en los resultados. Después de ejecutar la consulta de búsqueda, la base de datos preliminar arrojó 8216 documentos, que se filtraron para eliminar aquellos que estaban repetidos, mal clasificados o que no contenían resumen. Con base en los criterios de inclusión, se incluyeron 7658 artículos publicados entre 1981 y 2020.

### **Preprocesamiento**

Después de revisar los textos completos de los estudios elegibles, se utilizó LDA, un modelo de "bolsa de palabras" en el que los documentos se representan como una secuencia de palabras individuales. Cada documento fue tokenizado, un proceso necesario para obtener palabras individuales (también conocidas como unigramas) a partir de oraciones. Se eliminaron el texto en minúsculas, los signos de puntuación, los guiones, los corchetes, los espacios en blanco y otros caracteres. Además, también se eliminó una lista estándar de palabras llamadas "palabras vacías" para garantizar que cada oración fuera gramaticalmente correcta. Como resultado de este preprocesamiento se obtuvo una matriz de términos del documento donde cada artículo se representaba en un vector V-dimensional de una colección de palabras. El procesamiento de datos en esta parte del estudio se llevó a cabo utilizando LDAShiny, un paquete de código abierto para el lenguaje de programación R (R Development Core Team 2019), que contiene una herramienta que proporciona una interfaz gráfica de usuario basada en la web para realizar una revisión de la literatura científica bajo el enfoque bayesiano de Latent Dirichlet Allocation (LDA) y algoritmos de aprendizaje automático.

## Selección del Número de Temas LDA y Construcción del Modelo

Se realizaron simulaciones variando  $k$  de 4 a 30 en pasos incrementales de uno y se utilizó un algoritmo de inferencia con 500 iteraciones, a saber, el muestreo de Gibbs.

Se determinó un modelo LDA de calidad utilizando una medida de coherencia de tema, que es una medida de un modelo de tema desde la perspectiva de la interpretabilidad humana y se considera una medida más adecuada que una métrica computacional como la perplejidad.

## Etiquetado de temas

Se utilizó un algoritmo de etiquetado ingenuo basado en bigramas probables proporcionados por el paquete `textmineR`. Sin embargo, dado que los algoritmos tienen una capacidad muy limitada para comprender el significado latente del lenguaje humano, también se decidió utilizar el etiquetado manual, que se considera estándar en el modelado de temas, a partir de tres fuentes de información: las 20 palabras más frecuentes listas (lo más probable), una muestra de los títulos y los tres artículos más cargados.

Además, para mejorar el etiquetado, visualizamos los temas en un área bidimensional calculando la distancia entre ellos. Este método mostró el área que representaba la prevalencia del tema e indicó qué tan extendido estaba un tema dentro de todos los documentos. Este análisis mostró la similitud entre los temas con respecto a su distribución de probabilidad sobre las palabras.

## HJ-Biplot

El biplot HJ, una extensión de los biplots clásicos introducidos por Gabriel, es un método de análisis de datos exploratorio que busca patrones ocultos en una matriz de datos y luego representa gráficamente la información contenida en las filas (años) y columnas (temas). Se optó por esta técnica estadística multivariada ya que ofrece una evaluación de datos más precisa en la que se destacan las relaciones entre las partes, años y temas. Para este análisis se utilizó el software `Multibiplot` ya que proporciona una forma rápida y sencilla de incorporar nuestras tablas desde un formato Excel.

La representación de los datos consistió en visualizar una matriz de datos multivariados  $X_{n \times p}$  utilizando como puntos vectores llamados marcadores  $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n$  para cada fila, y vectores llamados marcadores  $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_p$  para cada columna. Cada fila representaba un tema y cada columna una variable, de modo que ambos conjuntos de marcadores podían superponerse en un mismo sistema de referencia con la máxima calidad de representación. Si las filas de la matriz



A se describieran como marcadores  $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n$  y la matriz B como marcadores  $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_p$ , el resultado fue  $\mathbf{X} = \mathbf{AB}^T$ .

Los marcadores se obtuvieron a partir de la habitual descomposición en valores singulares (SVD) de la matriz de datos. La SVD de la matriz X fue definida por  $\mathbf{X} = \mathbf{UDV}^T$ , donde U es la matriz donde las columnas son los vectores propios de  $\mathbf{XX}^T$ , V es la matriz donde las columnas son los vectores propios de  $\mathbf{X}^T\mathbf{X}$  y D es la matriz diagonal del singular valor  $\lambda_i$  de X. Sean A y B las matrices de las dos primeras columnas de UD y VD, respectivamente.

### Índices cuantitativos utilizados para analizar la tendencia del tema

La cantidad de documentos y palabras encontradas fue bastante grande, por lo que no fue posible comprender la tendencia del tema de manera intuitiva. Por lo tanto, utilizamos algunos índices cuantitativos para analizar los principales resultados del algoritmo de modelado de temas: la colección de términos con frecuencias asociadas que caracterizan un tema y la composición porcentual de cada documento. La distribución del tema en el tiempo  $\theta_k^y$  estaba:

$$(\theta_k^y) = \frac{\sum_{m \in y} \theta_{mk}}{n^y} \quad (1)$$

Donde  $m \in y$  representa los artículos publicados en un año determinado;  $\theta_{mk}$  es la proporción del k-ésimo tema en cada artículo; y  $n^y$  es el número total de artículos publicados en el año y.

Para facilitar la caracterización de los temas por su tendencia, se utilizaron pendientes de regresión simple para cada tema, en la que el año fue una variable dependiente y la proporción de los temas en el año correspondiente fue la variable respuesta. Los temas obtenidos por regresión fueron positivos o negativos a un nivel de significancia estadística de 0.01 y clasificados como tendencias positivas o negativas, respectivamente.

**Resultados:** La búsqueda inicial arrojó 8216 artículos. Después de aplicar los criterios de inclusión y exclusión y buscar en las listas de referencias de los estudios incluidos, se identificaron un total de 7658 estudios.

Los países que publicaron la mayoría de las investigaciones científicas sobre NET fueron Estados Unidos (2271), Alemania (246), Reino Unido (237), Francia (216), Italia (177), Japón (166), España (156), Países Bajos (151), Canadá (150) y, por último, China (131). Otros países no representaron la misma frecuencia de publicación ni tuvieron el mismo impacto.

Se encontró que la probabilidad de algunos temas aumentó gradualmente con el tiempo: t\_2 (costo y efecto en la calidad de vida), t\_3 (Marcadores de carcinomas), t\_6 (GEP-NET), t\_10

(tumor sólido-pseudopapilar del páncreas), t\_11 (Tumores neuroendocrinos pancreáticos), t\_13 (Pacientes un año después del tratamiento), t\_15 (Terapia con radionucleidos), t\_17 (Pronóstico de supervivencia antes y después de la resección), t\_20 (Feocromocitoma y paraganglioma), t\_21 (Beneficio clínico), t\_22 ( sistema de estadificación AJCC) y t\_24 (NF-pNET a largo plazo).

Los temas que mostraron una probabilidad decreciente fueron t\_1 (expresión génica en tejido normal de línea celular), t\_7 (carcinoma de células de Merkel), t\_14 (carcinoma de células grandes, células pequeñas) y t\_25 (niveles de cromogranina A); los temas en los que no se observó una tendencia fueron t\_4 (síndrome de ACTH ectópica), t\_5 (trasplante de hígado por metástasis hepática), t\_8 (tumor primario bien diferenciado), t\_9 (subtipos de receptores de somatostatina), t\_12 (metástasis en ganglios linfáticos), t\_16 (aspiración con aguja fina), t\_18 (neoplasia de grado), t\_19 (neoplasia endocrina múltiple) y t\_23 (vía biliar: reporte de un caso).

También se analizó la agrupación de los temas y años mediante análisis multivariado HJ-Biplot con la matriz theta de los resultados. La matriz estuvo compuesta por valores entre 0 y 1 según los temas y años en estudio. Encontramos agrupaciones coherentes con una variabilidad de datos del 79,9%.

Los temas que tuvieron mayor relevancia durante el periodo 1981–1985 fueron t\_23 (Conducto biliar: reporte de un caso), t\_8 (Tumor primario bien diferenciado), t\_25 (Niveles de cromogranina A), t\_14 (Células grandes, células pequeñas, carcinoma), y t\_7 (carcinoma de células de Merkel).

Durante los años 1986–2005, los temas fueron t\_1 (expresión génica en tejido normal de línea celular), t\_4 (síndrome de ACTH ectópica), t\_19 (neoplasia endocrina múltiple), t\_9 (subtipos de receptores de somatostatina), t\_5 (trasplante de hígado para metástasis hepáticas), t\_20 (feocromocitoma y paraganglioma) y t\_16 (aspiración con aguja fina).

Finalmente, para los años 2006–2020, los temas más relevantes fueron t\_12 (metástasis en ganglios linfáticos), t\_22 (sistema de estadificación AJCC), t\_6 (GEP-NET), t\_13 (pacientes un año después del tratamiento), t\_2 (costo y efecto sobre calidad de vida), t\_24 (TNE-NF a largo plazo), t\_11 (Tumores neuroendocrinos pancreáticos), t\_17 (Pronóstico de supervivencia antes y después de la resección), t\_10 (Tumor sólido pseudopapilar del páncreas), t\_21 (Beneficio clínico), t\_15 (Terapia con radionúclidos) y t\_3 (Marcadores de carcinomas).

**Conclusiones:** Nuestra revisión encontró que los países con más investigación sobre NET fueron los Estados Unidos, Alemania, el Reino Unido, Francia e Italia, y que los TNE pancreáticos fueron más frecuentes entre 1981 y 2020. Además, Los resultados confirmaron que “beneficios clínicos”, “pacientes un año después del tratamiento”, “pronóstico de supervivencia antes y después de la

resección” y “marcadores de carcinomas” fueron los temas más frecuentes en la literatura científica durante los últimos 39 años.

Finalmente, el método LDA empleado en esta revisión agrupó a cada sujeto en una categoría basado en las observaciones subjetivas de alta probabilidad de las palabras. Así, los métodos mostraron efectividad en la generación de respuestas sobre los temas más comunes estudiados en las NET. En contraste, el método HJ-Biplot fue integral para agrupar temas por año y encontrar qué NET fueron los más relevantes y para qué años.



*7.2 Neuroendocrine Tumors: An Analysis of Prevalence, Incidence, and Survival a Hospitals-Based Study in Ecuador*



Neuroendocrine Tumors: An Analysis of Prevalence, Incidence, and Survival in a Hospital-Based Study in Ecuador

---

**Resumen:** Los tumores neuroendocrinos (TNE) representan un grupo heterogéneo de neoplasias malignas, con una cantidad limitada de datos de América Latina. Por lo tanto, este estudio observacional tuvo como objetivo proporcionar datos sobre prevalencia, incidencia y supervivencia de NET en hospitales ecuatorianos. El estudio se realizó utilizando datos de la Sociedad para la Lucha Contra el Cáncer (SOLCA). Evaluamos pacientes con TNE (2000-2020) utilizando el método HJ-Biplot y los riesgos proporcionales de Cox. Incidencia anual ajustada por edad y prevalencia de duración limitada en análisis multivariable, así como el riesgo se obtuvieron los cocientes (HR) de mortalidad y supervivencia. En los años 2000–2020, la edad ajustada la tasa de incidencia aumentó 9 veces en el estómago y 7 veces en la mama. Las tasas de incidencia fueron 1,38 por 100.000 personas en el pulmón y en 1,79 por 100.000 personas en gastroenteropancreático (recto, estómago y páncreas). La prevalencia aumentó de 0,0027% en 2000 a 0,0736% en 2019 y 0,0245 % en 2020. La supervivencia global fue peor para los TNE metastásicos (HR, 4,061; IC 95 %, 1.932–8.540;  $p < 0,001$ ) y TNE locales avanzados (HR, 2,348; IC 95 %, 1,007–5,475  $p < 0,048$ ) que para NET localizados. En conclusión, la incidencia de TNE aumentó en los últimos 20 años y la supervivencia disminuyó con el tiempo, especialmente para tumores metastásicos en el páncreas y la fosa nasal.

Los **objetivos** que se persiguen en este artículo son los siguientes:

Mediante técnicas multivariantes:

- Determinar la incidencia, prevalencia y supervivencia en Tumores Neuroendocrinos (NET).
- Mediante el método HJ-Biplot agrupar variables para determinar relación entre estas.

**Metodología:** Estudio de cohorte retrospectivo se realizó con un ensayo de historia clínica de pacientes con NET. Se analizaron las bases de datos de la Sociedad de Lucha Contra el Cáncer (SOLCA) de las provincias de Manabí, Cuenca, Loja y El Oro. Se incluyeron todos los pacientes con TNE que habían sido diagnosticados y atendidos en el Servicio de Oncología de los centros SOLCA entre enero de 2000 y diciembre de 2020. Se excluyeron los pacientes que no tenían historia clínica completa, abandonaron el tratamiento o no siguieron el tratamiento en el hospital. Aquí se evaluó la edad, el sexo, la población, el estadio y el grado de diferenciación del tumor, el año de diagnóstico y la supervivencia de los pacientes con NET (medida en meses o hasta la muerte).

## Clasificación neta

Los NET se analizaron según los grados histológicos de la OMS (2010): Grado 1 (G1, índice Ki-67 <2% y recuento mitótico <2/10 campos de alta potencia (HPF)), Grado 2 (G2, Ki-67 índice 3–20 % y/o recuento mitótico 2–20/10 HPF), y Grado 3 (G3, índice Ki-67 3–20 % y/o recuento mitótico 2–20/10 HPF) y G3 (G3 neuroendocrino carcinoma no se diferenciaron), a pesar de la diferenciación tumoral. Además, los estadios tumorales de los pacientes con NET se evaluaron como local, local avanzado y metastásico. El tumor local estaba dentro de los límites del órgano de origen; el tumor local avanzado se expandió más allá del órgano de origen hacia los tejidos u órganos adyacentes o infiltró los ganglios linfáticos regionales; finalmente, el tumor metastásico era aquel que se extendía a sitios remotos fuera del órgano de origen.

## Análisis estadístico

El examen de la supervivencia a los 20 años según el sitio se realizó mediante el método de Kaplan-Meier y la prueba de rangos logarítmicos. Además, se evaluaron los datos del registro hospitalario de la Sociedad de Lucha Contra el Cáncer (SOLCA) para extraer la mediana de supervivencia general (SG) por sitio, estadio y grado. Para evaluar las tendencias más recientes en la supervivencia, realizamos análisis de supervivencia multivariable. El tiempo de supervivencia general (SG) a cinco años y el modelo de riesgos proporcionales de Cox se utilizaron en el análisis multivariante con censura aplicada a los 5 años. Las covariables para este análisis incluyeron factores que se sabe que influyen en el pronóstico de los TNE, incluido el grado, la edad, el estadio, el sitio y el intervalo de tiempo desde el diagnóstico. El modelo general fue significativo a  $p < 0,001$ .

La incidencia (incluyendo el cambio porcentual anual) y las tasas de prevalencia de duración limitada (20 años) se calcularon usando el cambio porcentual anual ajustando una regresión de mínimos cuadrados a las tasas, usando el año calendario como variable de regresión. Las tasas de incidencia ajustadas por edad se calcularon usando proporciones ponderadas de los grupos de edad correspondientes a la población estándar ecuatoriana para cada provincia: El Oro (715.751), Manabí (1.562.079), Lora (521.184) y Azuay (881.394), con un total de 3.670.408 (Instituto Nacional de Estadística y Censos, 2020).

## HJ-Biplot

A continuación, se utilizó el HJ-Biplot, que es una extensión del clásico biplot introducido por Gabriel y es un método exploratorio de análisis de datos que busca patrones ocultos en la matriz de datos. El HJ-Biplot tiene la ventaja de ser simultáneamente una representación y lograr una calidad de representación óptima tanto para filas como para columnas, con ambos representados en el mismo sistema de referencia. Luego, la matriz de datos muestra



gráficamente la información contenida en las filas (ubicación de la muestra) y columnas (grado y meses). El Biplot es una técnica estadística multivariante que se realizó para permitir un análisis de datos más detallado, destacando los vínculos entre las partes (ubicación de la muestra) y el sexo y la edad de los individuos.

Los análisis estadísticos se realizaron con IBM SPSS versión 25.0 (IBM, Armonk, NY, EE. UU.), software R versión 3.3.1 (<http://www.r-project.org>; consultado el 30 de junio de 2022) y software MultiBplot disponible gratuitamente. Las diferencias comparativas se consideraron significativas a  $p < 0,01$ .

**Resultados:** Entre 2000 y 2020 se registraron en la base de datos de las clínicas SOLCA en cuatro provincias del Ecuador 281 ingresos correspondientes a pacientes afectados por TNE. Observamos que los TNE más frecuentes fueron los de pulmón (18,1%), estómago (12,1%), mama (9,3%), cavidad nasal (7,1%) y recto (6,4%). La base de datos permitió el análisis del grado tumoral según la clasificación de la OMS; la mayoría de los casos se encontraban en el estadio avanzado G3 (56%), seguido de G2 (25%) y G1 (20%). Los sitios de tumor primario que con mayor frecuencia se presentaron con enfermedad en estadio III fueron pulmón (21 %), estómago (8,9 %), páncreas (8,9 %), fosa nasal (7 %), desconocido (7 %) y, por último, mama (6,4%).

### **Incidencia Anual**

El aumento en la incidencia de TNE en el período 2000-2020 ocurrió en todos los órganos, etapas y grados. En el año 2019, los aumentos en la incidencia para varios sitios variaron de 8 veces en el seno a 5 veces en el estómago. Entre los grupos de etapa, la incidencia fue la que más aumentó para los TNE metastásicos, de 0,24 por 100.000 personas en 2000 a 0,38 por 100.000 personas en 2020 ( $p < 0,001$ ). Las incidencias más altas fueron 1,38 por 100.000 personas para el pulmón, 1,79 por 100 000 personas para los sitios gastroenteropancreáticos (incluido 0,92 por 100 000 personas para el estómago, 0,49 por 100 000 personas para el recto y 0,38 por 100 000 personas para el páncreas) y 0,40 por 100.000 personas para NET con un sitio primario de origen desconocido.

### **Prevalencia**

Como reflejo de la incidencia creciente y la naturaleza indolente de los TNE, la prevalencia de duración limitada de 20 años aumentó sustancialmente, de 0,0027 % en 2000 a 0,073 % en 2019 y 0,0245 % en 2020 ( $p < 0,001$ ). Entre los grupos de grado, la prevalencia aumentó más para los TNE G3, y entre los sitios, la prevalencia fue más alta en el pulmón, seguido del estómago y el páncreas.

## Supervivencia

La mediana del tiempo de Supervivencia Global (SG) para todos los pacientes fue de 2,4 años (28,8 meses); los NET localizados tuvieron una mejor mediana de SG a los 2,30 años (26,6 meses) en comparación con los NET locales avanzados a los 2,14 años (25,7 meses) y los NET metastásicos a los 1,60 años (19,2 meses) ( $p < 0,001$ ). De aquellos con grados conocidos, los G1 NET tuvieron la mediana de SG más alta (2,12 años) (25,4 meses) entre los grupos de grado, los G2 NET tuvieron una peor SG a los 2,07 años (24,9 meses), mientras que los G3 NET tuvieron la peor SG (1,8 años). Los NET en el recto (3,6 años) y el apéndice (2,7 años) tuvieron la mejor SG media entre los grupos de sitio, mientras que los NET en el páncreas (1,19 años) y las fosas nasales (1,14 años) tuvieron la peor SG media. Todas estas diferencias en la supervivencia fueron significativas (log-rank  $p < 0,001$ ).

Luego evaluamos los patrones de supervivencia según el sitio y la etapa. En los TNE localizados, la mediana de SG varió de 3 años en el intestino delgado a más de 4,6 años en el apéndice. En las NET locales avanzados, la mediana del sistema operativo osciló entre 2,9 años para las NET en el apéndice y más de 30 años en el sitio primario desconocido. Para NET distantes, aquellos en el recto tuvieron la mejor mediana de SG (3 años); Los NET en el apéndice (2,4 años) y el páncreas (1,6 meses) tuvieron la peor OS media. Todas estas diferencias en la SG fueron significativas (log-rank  $p < 0,001$ ).

## Análisis multivariable de OS

Realizamos un análisis multivariable con cocientes de riesgo (HR) calculados para tasas de riesgo de mortalidad a 5 años. Encontramos que los pacientes con TNE G2 (HR, 2,407; IC 95 %, 0,941–6,154;  $p < 0,037$ ) y G3 (HR, 5,003; IC 95 %, 2,221–11,273;  $p < 0,001$ ) tenían peor SG que aquellos con G1 RED. Se encontró que la edad, el estadio y el grado tenían una correlación significativa con la supervivencia. La supervivencia global fue peor en los TNE metastásicos (HR, 4,061; IC 95 %, 1,932–8,540;  $p < 0,048$ ) y TNE locales avanzados (HR, 2,348; IC 95 %, 1,007–5,475;  $p < 0,001$ ) que en los TNE localizados. Todas las comparaciones anteriores fueron significativas a  $p < 0,001$ .

## HJ-Biplot

Se utilizó el método HJ-Biplot quien determina la relación entre variables, considerando los 10 TNE más prevalentes. Se evaluó el grado de diferenciación, los meses de supervivencia y la localización del tumor. En el grado de diferenciación para G1, las mayores tasas de supervivencia se dieron entre los 84 y 99 meses; los tumores en el estómago y el apéndice tenían una alta correlación con una supervivencia de 24 a 35 meses, mientras que los tumores en el recto y el pulmón tenían una tasa de supervivencia aún mayor de 84 a 99 meses. Cuando evaluamos el grado de diferenciación para G2, la mayor supervivencia se observó entre 12 y 23 meses en TNE

de apéndice, pulmón, mama y cuello uterino. Finalmente, evaluamos el grado de diferenciación para G3; la piel tuvo el peor pronóstico, con un rango de supervivencia entre 0 y 11 meses, un rango de supervivencia entre 48 y 59 meses para el páncreas.

Por otra parte, el HJ-Biplot muestra la edad, el sexo y la ubicación de los NET donde se observaron dos grupos. El primer grupo representa el género femenino, que tuvo una correlación significativa con los tumores de cuello uterino, mama, estómago y recto. El segundo grupo fue el género masculino, que se correlacionó con tumores de páncreas, sitio desconocido, piel, pulmón, fosa nasal y apéndice. Se pudo determinar que los TNE de recto para el género femenino se relacionaron con edades entre 50 y 59 años, mientras que los TNE de estómago, mama y cuello uterino tuvieron mayor conexión con mujeres entre 60 y 69 años. Para los hombres, el tumor de fosa nasal apareció en mayores de 80 años, los TNE de páncreas, apéndice, piel y sitios desconocidos tuvieron mayor conexión con jóvenes entre 20 y 29 años, y los tumores de pulmón aparecieron entre los 70 años y 79 años.

**Conclusiones:** La incidencia de TNE aumentó en los hospitales ecuatorianos entre 2000 y 2020, y proporcionamos datos primarios sobre su prevalencia. Nuestros datos mostraron que la drástica reducción en casos diagnosticados de TNE en 2020 reflejó la caída en el número de personas que buscan servicios de salud debido a la pandemia mundial de COVID-19. Es posible afirmar que el método estadístico HJ-Biplot que utilizamos confirmó nuestros resultados, y la evaluación clásica y el modelo de riesgos proporcionales determinaron efectivamente las tasas de supervivencia. Desafortunadamente, los estadios avanzados de los TNE durante el diagnóstico de pacientes en Ecuador contribuyeron al bajo tiempo de supervivencia observado en nuestros datos. La supervivencia fue mayor para aquellos con lesiones localizadas. Los TNE, mientras que los tumores metastásicos presentan peor pronóstico, sobre todo en el páncreas y redes nasales. Es posible que, para la población ecuatoriana, el diagnóstico estándar se pueden implementar procedimientos con un enfoque en los principales sitios NET encontrados en nuestro estudio.



### *7.3 Clinical Related Factors to Neuroendocrine Tumors in Ecuadorian Patients: A Logistic Biplot Approach*



Clinical related factors to neuroendocrine tumors in Ecuadorian patients: a logistic biplot approach

---

**Resumen:** Los tumores neuroendocrinos (TNE) son relativamente raros y afectan a las células neuroendocrinas de todo el cuerpo. La mayoría de los tumores se diagnostican en etapas avanzadas. La prevalencia de los TNE ha aumentado en los últimos años, pero hay pocos datos sobre los países en desarrollo. El objetivo de este estudio fue determinar los síntomas asociados a los TNE en pacientes de la Sociedad de Lucha contra el Cáncer (SOLCA) en Ecuador entre 2005 y 2020 utilizando biplots logísticos en una base de datos hospitalaria generando respuestas binarias (presencia / ausencia) relevantes para este estudio. Los resultados mostraron que la edad promedio era de 59 años y el estudio no encontró diferencias en la prevalencia entre géneros. Los TNE se encontraron con mayor frecuencia en los pulmones (19%), seguidos del estómago (18%). La mayoría de los pacientes tenían diagnóstico patológico G2 y G3 (30% y 70% respectivamente). Los síntomas como tos, disnea, pérdida de peso, diarrea, estreñimiento, dolor abdominal, dispepsia, crisis hipertensiva, abdomen distendido y obstrucción intestinal tuvieron valores de  $p < 0,05$ . Además, el análisis estadístico mostró que la tos y la obstrucción intestinal también eran comunes teniendo en cuenta que los pacientes tenían TNE más frecuentes en los pulmones y la piel. En resumen, nuestros resultados indican que los síntomas de los pacientes con TNE se asociaron positivamente con los pulmones y la piel. Se necesitan más estudios que se centren en el tipo de TNE y sus síntomas para prevenir y diagnosticar el tumor a tiempo.

Los **objetivos** que se persiguen en este artículo son los siguientes:

Mediante técnicas multivariantes:

- Mediante el método Biplot logístico encontrar la relación existente entre el síntoma principal del paciente con la localización del tumor.

**Metodología:** El estudio se realizó utilizando la base de datos del hospital de SOLCA Portoviejo, en la Provincia de Manabí, Ecuador, analizando pacientes con diagnóstico de Tumor Neuroendocrino (TNE) entre el periodo 2005 y 2020.

### **Análisis estadístico**

El principal problema con nuestros datos es el número limitado de pacientes y el alto número de variables (síntomas, tipos de tumor, ubicaciones, etc...) que impide el uso de los métodos estándar como la regresión logística. Una técnica multivariante exploratoria puede ser más adecuada para explorar posibles patrones antes de intentar aplicar una forma más formal de modelo usando un subconjunto relevante de síntomas obtenido de ella. Este tipo de técnicas tienen también la ventaja de mostrar las interrelaciones entre todas las variables.

Para el análisis exploratorio de nuestros datos, utilizamos una técnica multivariante llamada “biplot logístico” que fue diseñada específicamente para tratar datos binarios. Aunque los biplots logísticos no han sido ampliamente utilizados con datos médicos, algunos autores proponen este tipo de técnicas multivariadas como un medio para realizar un análisis exploratorio previo a un análisis más formal. Otros autores también proponen la técnica como un análisis de los residuos de un modelo, para comprobar la bondad de idoneidad o la violación de alguna anterior hipótesis sobre los datos.

Los ángulos entre vectores pueden ayudar con la identificación de las relaciones entre los síntomas: pequeños ángulos agudos significan fuerte relaciones positivas, ángulos planos (formados por dos ángulos rectos  $180^\circ$ ), relaciones negativas y ángulos rectos (formados por dos perpendiculares líneas  $90^\circ$ ), no están relacionados. La posición de los vectores también puede ayudar a identificar los principales gradientes (o rasgos latentes) que resumen la variación de los síntomas y su relación con las diferentes ubicaciones de los tipos de cáncer.

Se llama Biplot Logístico porque la relación de los síntomas y los rasgos (dimensiones) es logístico, es como un Análisis de componentes Principales (o biplot) para datos binarios en los que la relación con los componentes es logístico más que lineal. cuando los datos son binarios, como los obtenidos en el análisis de Tumores Neuroendocrinos, el biplot clásico y análisis de componentes principales (ACP) no son adecuados porque la respuesta a lo largo de las dimensiones es lineal. Este es la misma razón por la que la regresión lineal no es apropiada para datos binarios o categóricos.

Se utilizaron paquetes MULTIBILOT y MultiBiplotR para realizar los cálculos y obtener las representaciones gráficas. Para este estudio, primero se realizó un Análisis de Coordenadas Principales (Clásico Escalamiento multidimensional), con el simple coeficiente de coincidencia y luego el ajuste separado teniendo en cuenta las regresiones logísticas para representar las direcciones en el mapa euclidiano. La interpretación de la relevancia de los síntomas para clasificar a los pacientes se evaluó por métodos tradicionales de indicadores para regresiones logísticas como pruebas de Wald, pseudo coeficientes R-cuadrado y porcentajes de clasificación precisa.

**Resultados:** Características de la población y TNE síntomas: Un total de 94 pacientes adultos (de 19 a 95 años) fueron diagnosticados con TNE entre 2005 y 2020 e incluidos en esta investigación. La ubicación primaria más frecuente fue el pulmón (19%), seguido del estómago (18%), recto (8%) y piel (8%). La ubicación Desconocido representa el 8% de los tumores estudiados.

La edad en este estudio fue de 59 años y la prevalencia fue mayor en los hombres 57% versus 43% en mujeres. La etapa más prevalente en pacientes con TNE fue G3 (70%) mientras que G1 y G2 representan 30% de los pacientes.



### **Biplot logístico en tipos de NETs y síntomas en pacientes de SOLCA**

Se aplicó el método Biot Logístico a la matriz binaria de 94 pacientes y 37 síntomas. Cada paciente (puntos) ha sido etiquetado usando la ubicación principal de su tumor y cada síntoma (flecha) usando el nombre del síntoma. Solo los síntomas significativamente relacionados con la configuración del paciente fueron retenidos en el Biplot Logístico.

La mayoría de los pacientes que tienen los pulmones como la ubicación principal están en la parte superior del Biplot Logístico, mientras que los tumores relacionados con el aparato digestivo (recto, estómago, colon derecho, íleon y páncreas) están en la parte inferior y bien separados del grupo "pulmón", por lo que podemos concluir que el análisis tiene algún poder para clasificar los tumores.

Al proyectar los puntos de los pacientes en las instrucciones para los síntomas, se obtiene la probabilidad esperada de que el paciente tiene el síntoma. Para predecir la presencia esperada cuando la esperada probabilidad es superior a 0,5. Para cada flecha, solo representamos el punto prediciendo 0.5 y la dirección de aumento probabilidades. Así, podemos correctamente predecir el 95,75% de las presencias y ausencias. Todos los síntomas individuales tienen porcentajes de predicciones correctas sobre 94%. Los valores de R-cuadrado asociados son sobre el 70% y la información de cada variable puede interpretarse con precisión. Se puede concluir que la representación en la imagen es buena teniendo en cuenta la matriz de datos y puede ser utilizada para la exploración.

Buscar los síntomas que caracterizan cada grupo, se puede observar en la dirección que señalan. Por ejemplo, disnea y tos están presentes en pacientes para los cuales la ubicación principal son los pulmones, mientras que la diarrea, dolores de estómago, abdomen distendido, Crisis Hipertensiva y Obstrucción Intestinal señalar la dirección de los tumores relacionados con el sistema digestivo. Ambos conjuntos de los síntomas apuntan en dirección opuesta lo que significa que están inversamente relacionados, es decir, los síntomas presentes en un grupo están ausentes en el otro. Juntos, definen el gradiente que separa ambos tipos de tumores La disnea y la tos apuntan a la misma dirección, lo que significa que están alta y positivamente correlacionados y aparecen juntos en algunos pacientes. La pérdida de peso tiene un muy buen ajuste, pero no está claramente caracterizado en cualquiera de los grupos. Lo mismo que con los síntomas dispepsia y estreñimiento. El tres están íntimamente relacionados con la primera dimensión, pero son menos importantes para separar entre grupos de tumores.

Conclusiones: Los resultados proporcionados sugieren que los síntomas clínicos más frecuentes (tos y obstrucción intestinal) fueron positivamente relacionados con el tipo de NET. al mirar en la prevalencia de TNE, los pulmones fueron el sitio más frecuente, lo que refuerza la hipótesis de

la vulnerabilidad de este órgano. Se necesita más investigación y debe centrarse en identificar los síntomas de los TNE con el objetivo de diagnóstico precoz. Además, la prevalencia de TNE específicos en cada población debe subrayarse en futuras investigaciones.

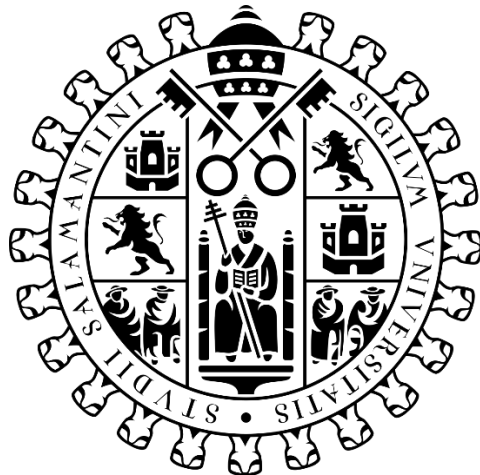
La técnica Biplot Logístico permite la generación de información relevante basada en antecedentes clínicos, como lo fue para este estudio de la presencia o ausencia de la enfermedad, teniendo en cuenta pacientes con diferentes tumores neuroendocrinos.



**UNIVERSIDAD DE SALAMANCA**  
**DEPARTAMENTO DE ESTADÍSTICA**

---

**MÉTODOS MULTIVARIANTES PARA EVALUAR  
TUMORES NEUROENDOCRINOS**



**TESIS DOCTORAL**

**KARIME MONTES ESCOBAR**

**2022**