

Boletines Epidemiológicos con R *

Carlos Andrés Barco Rojas

Escuela Superior de Administración Pública - Territorial Caldas

Jorge Mario Estrada Alvarez

Secretaria de Salud Pública y Seguridad Social de Pereira - Comfamiliar

Risaralda

El objetivo primordial de este estudio fue desarrollar un flujo de trabajo destinado a mejorar el procesamiento de datos en el ámbito de la vigilancia de salud pública. Para este propósito, se empleó el software R, con el fin de agilizar la recolección, análisis y presentación de datos epidemiológicos, permitiendo la generación automática de informes. El diseño metodológico abarcó la selección de conjuntos de datos procedentes del Sistema de Información en Salud Pública (SIVIGILA), que contenían información relativa a más de 116 eventos de interés en salud pública. Se aplicó un código de preprocesamiento de datos que excluyó casos inadecuados y duplicados, y se categorizaron los casos según la naturaleza de la enfermedad y su ubicación, ya sea en el lugar de infección o residencia. Los resultados se consolidaron en tablas con codificación de colores específica para una semana epidemiológica determinada. También se implementaron herramientas epidemiológicas, como los canales endémicos, para detectar comportamientos inusuales y brotes en la ciudad. Este flujo de trabajo en R Software conlleva una reducción sustancial en el tiempo dedicado al preprocesamiento de datos y acelera la producción de informes automatizados en el campo de la salud pública. Como resultado, los epidemiólogos pueden emplear herramientas de análisis avanzado y disponer de más tiempo para interpretar resultados y detectar patrones inusuales. La adopción de esta metodología tiene el potencial de mejorar la eficiencia en la vigilancia de salud pública y de facilitar la toma de decisiones basadas en información sólida.

Keywords: Vigilancia en salud pública , R Markdown, Tidyverse, Boletines epidemiológicos

Introducción

La Vigilancia en Salud Pública es un proceso sistemático y constante que involucra la recolección, análisis, interpretación y difusión de datos específicos relacionados con la salud [Brookmeyer and Stroup \(2003\)](#). Estos datos se utilizan en la planificación, ejecución y evaluación de prácticas en salud pública, convirtiéndola en un componente esencial para el monitoreo y control de enfermedades, así como para la toma de decisiones informadas en el ámbito de la salud.

En Colombia, el Instituto Nacional de Salud desempeña un papel central como entidad rectora en la vigilancia de eventos de interés en salud pública a través del Sistema de Información para la Vigilancia (SIVIGILA). Este sistema agiliza la notificación de enfermedades catalogadas como de interés en salud pública debido a su impacto y capacidad de generar brotes y epidemias que afectan la salud colectiva.

Dentro de la vigilancia en salud pública, se incluye un componente de divulgación de resultados, que se lleva a cabo principalmente a través de la emisión de publicaciones periódicas que informan sobre el comportamiento de los eventos notificados en el sistema. Esto se realiza mediante la aplicación de la triada analítica epidemiológica, denominada tiempo, lugar y persona. Este

*Autores de contacto: carlos.barco@esap.edu.co, jmario.estrada.alvarez@gmail.com

enfoque implica el uso de diversos métodos analíticos y representaciones gráficas para resumir, presentar y analizar los resultados obtenidos a través de la vigilancia.

La complejidad de estos reportes se debe a la diversidad de fuentes de información que pueden ser utilizadas para un mismo reporte, así como a las estructuras de datos diferentes, algunas a nivel individual (sujetos) y otras a nivel agregado. Esto implica el procesamiento de múltiples bases de datos con estructuras distintas para generar un solo reporte. Esta diversidad ha dificultado la práctica del análisis en salud pública en las diversas entidades territoriales de salud, como las Secretarías de Salud departamentales y municipales. Ha generado desafíos en la gestión de varios formatos de datos y ha ralentizado la producción de informes de vigilancia, así como la implementación de métodos estadísticos más sensibles pero también más complejos.

En este contexto, el uso de herramientas tecnológicas se ha vuelto fundamental para agilizar la recopilación, análisis y presentación de datos epidemiológicos [Khan \(2013\)](#). Hasta el momento, a nivel nacional, Microsoft Excel ha sido una herramienta destacada entre los equipos de vigilancia en salud pública, pero es propenso a errores y ofrece a los usuarios promedio una funcionalidad analítica limitada. Del mismo modo, el software de “apuntar y hacer clic”, como Tableau y PowerBI, ofrece un aprendizaje rápido, pero limita las posibilidades de análisis personalizado [Neale Batra et al. \(2021\)](#).

R software, desarrollado por el equipo principal de [R R Core Team \(2019\)](#), es una de las herramientas más utilizadas y referenciadas en el campo de la ciencia de datos. Es la más empleada en la implementación de análisis de datos epidemiológicos. La diversidad de librerías disponibles facilita el pre-procesamiento, análisis y presentación de resultados, además de permitir la combinación de diversas fuentes y estructuras de datos en un solo lugar. Esto, junto con la reproducibilidad deseada y la automatización de reportes, supone una ventaja significativa para la vigilancia en salud pública al ahorro de tiempo y la rápida disponibilidad de resultados para los tomadores de decisiones.

El objetivo principal de este trabajo consistió en estructurar un flujo de trabajo que abarcara desde la entrada de datos de la vigilancia en salud pública, su consolidación, pre-procesamiento y la producción de un reporte estandarizado de boletín epidemiológico. Este informe contendría los resultados sobre el comportamiento en tiempo, lugar y persona de eventos de interés en salud pública.

Materiales y metodos

Se diseñó un flujo de trabajo en R software que se ejecuta de manera semanal y abarca las siguientes etapas: la selección de conjuntos de datos recopilados del Sistema de Información en Salud Pública, SIVIGILA. Estos conjuntos de datos se componen de dos tipos de estructuras. En primer lugar, múltiples datasets que contienen información a nivel de casos notificados por parte de las instituciones de salud, abarcando más de 116 eventos de interés en salud pública de notificación obligatoria. En segundo lugar, se encuentran datasets con casos agregados, correspondientes a lo que se conoce bajo los protocolos del Instituto Nacional de Salud como notificación colectiva. Esto se aplica principalmente a eventos como la infección respiratoria aguda y la infección diarreica aguda.

Posteriormente, se estructuró un código de pre-procesamiento de datos que sigue reglas de depuración preestablecidas. Estas reglas incluyen la eliminación de casos descartados, casos que no cumplen con los criterios del protocolo de notificación, la supresión de duplicados basada en el documento, la fecha de consulta y el evento. Además, de acuerdo a la naturaleza de la enfermedad, se realiza la selección de casos por municipio de procedencia y/o residencia. Para

eventos de tipo transmisible, el análisis se enfoca en el lugar de posible infección, mientras que para eventos no transmisibles, el análisis se basa en el lugar de residencia del afectado.

Por último, se elaboró un informe en R-Markdown con una plantilla de LaTeX que incluye la representación de datos en tablas, gráficos y la implementación de funciones para la aplicación de algoritmos de detección de comportamientos inusuales en series de tiempo, específicamente en relación a los conteos semanales de los eventos trazadores de interés en salud pública.

Se emplearon algunas funciones de R base y librerías de R para el preprocesamiento, como Tidyverse de [Wickham et al. \(2019\)](#), y lubridate [Grolemund and Wickham \(2011\)](#). Además, se utilizó la librería gtsummary de [Sjoberg et al. \(2021\)](#) para la construcción de tablas de informe, y se emplearon funciones específicas para el análisis de datos epidemiológicos, como incidence, desarrollada por [Kamvar et al. \(2019\)](#).

Resultados

Se ha desarrollado un flujo de trabajo que reduce el procesamiento de datos y la producción de reportes a cuestión de minutos, logrando esto mediante la creación de un código estandarizado en la fase de preprocesamiento, como se muestra a continuación:

```
# Filtrado de casos D, R, 6
datos<-filter(datos2, ajuste_=="0" | ajuste_=="3" |
              ajuste_=="4" | ajuste_=="5" | ajuste_=="7")
#Filtardo de procedencia Pereira
proce<-filter(datos2, ndep_proce=="RISARALDA" & nmun_proce=="PEREIRA")
proce<-filter(proce, cod_eve=="580" | cod_eve=="210" | cod_eve=="220" |
              cod_eve=="217" | cod_eve=="895" |
              cod_eve=="465" | cod_eve=="420" |
              cod_eve=="430" | cod_eve=="440" |
              cod_eve=="205" | cod_eve=="310")
#Eliminado duplicados por identificación y fecha consulta
procedencia<-distinct(proce, num_ide_, cod_eve, fec_con_, .keep_all = TRUE)
```

con una visualización agradable y presentación estándar ver figura 1.

Se obtuvieron resultados consolidados en tablas con resaltados en color, de acuerdo a una parametrización de incremento para una semana epidemiológica específica. Además, se implementaron herramientas epidemiológicas, como los canales endémicos, que posibilitan la vigilancia de comportamientos inusuales y brotes a nivel de la ciudad ver figura 2.

Conclusión

Se ha desarrollado un flujo de trabajo utilizando R Software que combina diversas fuentes de información, así como múltiples estructuras de datos recopiladas de manera sistemática a partir de eventos de vigilancia en salud pública. Esto conlleva a una reducción significativa del tiempo dedicado al pre-procesamiento de datos y acelera la producción de reportes automatizados. Como resultado, los epidemiólogos pueden emplear herramientas avanzadas de análisis y, al mismo tiempo, disponer de más tiempo para la interpretación de resultados y la detección e investigación de comportamientos inusuales.

Boletín Epidemiológico Municipio de Pereira

Índice

1 Eventos Trazadores	1
1.1 Infección respiratoria aguda	1
1.2 Circulación virus respiratorios	5
1.3 Enfermedades transmitidas por vectores	6
1.4 Infección Diarreica Aguda	8
2 Eventos de interés en salud pública	8
2.1 Mortalidad materno-perinatal	8
2.2 Agresiones por animales potencialmente transmisores de rabia	8
2.3 Tuberculosis	8
2.4 VIH/SIDA	8
2.5 Sífilis gestacional y congénita	9
2.6 Cáncer	9
2.7 Defectos congénitos y bajo peso al nacer	9
2.8 Inmunoprevenibles	9
2.9 Virus nuevo (COVID-19)	9
3 Vigilancia de exantemática	11
3.1 Vigilancia Varicela en distintos grupos poblacionales	11

Figura 1: Portada boletín epidemiológico

De igual manera en el canal endémico para la notificación por IRA hospitalizados en Unidad de Cuidados Intensivos (UCI) (figura 3), con descenso para esta semana epidemiológica notificada. La proporción acumulada de hospitalizaciones por IRA en UCI alcanza un 2.99%.

Edad (años)	Hosp. UCI IRA	Hosp. UCI todas las causas	% de Hosp. UCI
< 1	0	41	0.0
1	2	5	40.0
2 a 4	0	6	0.0
5 a 19	0	22	0.0
20 a 39	0	33	0.0
40 a 59	1	38	2.6
mayor e igual 60	0	91	0.0

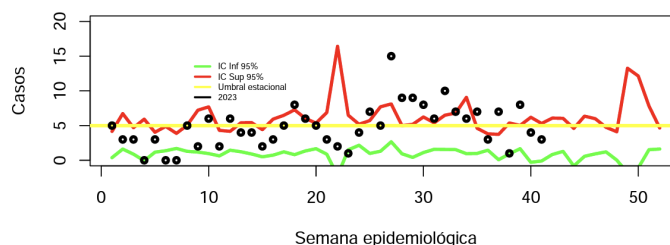


Figura 3: Canal endémico Hospitalizaciones por IRA en UCI, Pereira 2023

Figura 2: Salidas de tablas y graficas en boletín epidemiológico

References

- Brookmeyer, R. and D.F. Stroup. 2003. *Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance*. Oxford University Press.
URL: <https://books.google.com.co/books?id=7f3Y9M26StAC>
- Grolemund, Garrett and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40(3):1–25.
URL: <https://www.jstatsoft.org/v40/i03/>
- Kamvar, Zhian N., Jun Cai, Juliet R.C. Pulliam, Jakob Schumacher and Thibaut Jombart. 2019. "Epidemic curves made easy using the R package incidence [version 1; referees: awaiting peer review]." *F1000Research* 8(139).
URL: <https://doi.org/10.12688/f1000research.18002.1>
- Khan, Amir Maroof. 2013. "R-software: A Newer Tool in Epidemiological Data Analysis." *Indian J Community Med* 38:56–58.
URL: <https://doi.org/10.4103/0970-0218.106630>
- Neale Batra et al. 2021. *The Epidemiologist R Handbook*. Applied epi.
URL: <https://epirhandbook.com/en/index.html>
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
URL: <https://www.R-project.org/>
- Sjoberg, Daniel D., Karissa Whiting, Michael Curry, Jessica A. Lavery and Joseph Larmarange. 2021. "Reproducible Summary Tables with the gtsummary Package." *The R Journal* 13:570–580.
URL: <https://doi.org/10.32614/RJ-2021-053>
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan and Romain François et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4(43):1686.