

Análisis de la consistencia de los datos de un sistema de vigilancia sanitario empleando Statechart y R.

Data consistence analysis of a health surveillance system using Statechart and R.

Presentación: 17/10/2023

Camila Kerps

Universidad Tecnológica Nacional, Facultad Regional Santa Fe - Argentina
Camilakerps@hotmail.com

Resumen

El objetivo de esta investigación es la evaluación de la consistencia de los datos proporcionados por el Sistema Nacional de Vigilancia de la Salud (SNVS 2.0). Para realizar tal evaluación se propone el diseño de una máquina de estados que especifique los distintos estados por lo que puede transcurrir un paciente en el sistema SNVS 2.0. La máquina será implementada empleando el lenguaje R, a partir de su implementación se simularán las máquinas asociadas a cada paciente para la identificación de problemas en la calidad de los datos.

Palabras clave: Tuberculosis, sistema de salud pública, sistema de vigilancia sanitario, máquina de estados

Abstract

The main goal of this research is the evaluation of the consistency of the data provided by the Health Surveillance System SNVS 2.0. To carry out such an evaluation, the design of a state machine is proposed. This machine specifies the different states through which a patient can pass in the SNVS 2.0 system. The machine will be implemented using the R language, from its implementation the machines associated with each patient will be simulated to identify problems in the quality of the data.

Keywords: Tuberculosis, public health system, health surveillance system, state machine

Introducción

La tuberculosis (TB) es una enfermedad infecciosa causada por un tipo de bacteria que suele afectar a los pulmones. La Organización Mundial de la Salud (World Health Organization, 2023) estima que alrededor de una cuarta parte de la población mundial se ha infectado con el bacilo de la TB. Entre el 5 y el 10% de estas personas acabarán presentando síntomas y enfermando de TB. La TB suele tratarse con antibióticos y puede ser mortal si no se trata. Antes de la pandemia de coronavirus (COVID-19), la TB era la principal causa de muerte por un único agente infeccioso, superando al VIH/Sida (World Health Organization, 2022).

Sin tratamiento, la tasa de fallecimiento causada por la enfermedad de la tuberculosis es alta (cerca del 50%). Con los tratamientos actualmente recomendados (de 4-6 meses), alrededor del 85% de los pacientes pueden ser curados (World Health Organization, 2022).

El tratamiento precoz y la adherencia al tratamiento de la TB es fundamental para lograr la curación del paciente, evitar la aparición de resistencia a los medicamentos, y para suprimir las fuentes de infección para la población. La interrupción del tratamiento se puede prevenir, o limitar, mediante el desarrollo de estrategias y herramientas tecnológicas de monitorización de la adherencia digital, basadas en los sistemas de vigilancia de la salud pública. El registro y reporte de datos es necesario para monitorear tendencias en la epidemia, el progreso en el tratamiento de pacientes individuales y grupos de pacientes y asegurar la continuidad del cuidado aún en situaciones de traslado (World Health Organization, 2012). A partir de tales registros es posible la identificación de la interrupción del tratamiento de manera temprana a nivel del paciente, alertar a los equipos de salud y facilitar las acciones para la continuidad del tratamiento. Este es uno de los motivos por los cuales el correcto seguimiento de los datos relacionada a los casos de TB es tan importante.

A partir del año 2018 la Secretaría de Gobierno de Salud en Argentina implementa el Sistema Nacional de Vigilancia de la Salud (SNVS 2.0) como sistema único de notificación oficial de Eventos de Notificación Obligatoria (Secretaría de Salud, 2018). Este sistema recopila información sobre los eventos de salud para ponerla a disposición de quienes deben tomar decisiones de salud pública o tienen a su cargo las acciones de control o prevención. La TB constituye un evento de notificación obligatoria. En el manual Instrucciones para la notificación de Tuberculosis a través del SNVS 2.0 (Secretaría de Salud, 2018) se indica el procedimiento para la notificación del evento empleando tal sistema. No obstante, las evaluaciones epidemiológicas anuales de la situación de la TB en Argentina reflejan que muchas veces la información obtenida a partir del registro nacional resulta incompleta. Por ejemplo, en el año 2020, se notificaron 11380 nuevos casos de TB, el 41,4 % de los casos no contaron con información respecto a su resultado de tratamiento y un 83,2% de los registros no contó con información respecto a condición de VIH (Dirección de Respuesta al VIH, ITS, Hepatitis Virales y Tuberculosis, 2022). German et al. (2001) resaltan que la completitud y validez de los datos disponibles en todo Sistemas de Vigilancia en Salud es fundamental para la aceptabilidad del sistema tanto a nivel operativo como a nivel de definición de políticas públicas. La calidad de la información es un factor clave en el desarrollo de las actividades y es el objetivo principal de este proyecto. Si los datos disponibles son de alta calidad, es posible tomar acciones correctivas para los problemas identificados, planear y evaluar un procedimiento adecuado, y formular hipótesis de investigación (German et al., 2001).

El Instituto Nacional de Enfermedades Respiratorias Dr. Emilio Coni (INER) lleva a cabo la vigilancia epidemiológica y operativa de la TB. Para realizar estas actividades ha implementado en el lenguaje R diversas herramientas de análisis de los eventos reportados en SNVS 2.0 (<http://www.anlis.gov.ar/iner/>). Como parte de estas herramientas, desde INER han desarrollado en R algoritmos para detectar problemas de completitud y validez. Sin embargo, la lógica del problema bajo análisis está especificada en un lenguaje de programación y dificulta su estudio y mantenimiento. Con el desarrollo de este proyecto se pretende el desarrollo de herramientas de un nivel mayor de abstracción, donde los distintos participantes puedan comprender, analizar y mantener las reglas definidas para la completitud y validez de los datos. En primer lugar, se aplicarán estas herramientas en la medición y evaluación de la consistencia de los datos sobre los eventos de TB proporcionados por el SNVS 2.0. Eventualmente, las herramientas desarrolladas podrán ser generalizadas para poder aplicarlas a distintos conjuntos de datos de notificaciones sanitarias. En la siguiente sección se describe la metodología de trabajo propuesta. Luego se presentan los primeros resultados y las conclusiones alcanzadas.

Metodología

Como se mencionó anteriormente, este proyecto surge de la necesidad de evaluar la consistencia de los datos de los eventos sobre TB notificados en el SNVS 2.0. En el manual de normas y procedimientos de vigilancia y control de eventos de notificación

obligatoria (Ministerio de Salud, 2022) se definen 151 eventos de notificación obligatoria. En particular, se define un tipo de evento para TB, donde se indica que se notificarán en este evento todos los casos de TB, sean estos bacteriológicamente confirmados o clínicamente diagnosticados. Todos estos casos deben ser notificados, independientemente de si iniciaron tratamiento o no. Un caso de TB clínicamente diagnosticado es aquel que no cumple con los criterios para la confirmación bacteriológica, pero ha sido diagnosticado con TB activa por un médico, quien ha decidido dar al paciente un ciclo completo de tratamiento de TB; esta definición incluye casos diagnosticados sobre la base de anomalías a los rayos X o histología sugestiva y casos extrapulmonares sin confirmación de laboratorio.

SNVS 2.0 contiene registro de los eventos para los distintos tipos. A partir del mismo se genera un archivo filtrado con eventos vinculados a TB. El archivo trabajado contiene un registro de eventos con información acerca de cada caso, incluyendo datos como la fecha de diagnóstico, el estado de tratamiento inicial y resultado del tratamiento que, al analizar el documento, se encontró que podrían aportar al desarrollo de los estados y a su vez ser causantes de las transiciones entre los mismos. Por tal motivo se analiza la potencialidad de las máquinas de estados finitos de UML para modelar y evaluar la consistencia de los datos.

Las máquinas de estados finitos de UML (OMG, 2017) son una extensión de statechart propuesto por Harel (1987). Estas máquinas son una herramienta útil para la representación de los distintos estados por los que puede pasar un objeto bajo análisis en función de su estado y los eventos que suceden. La representación obtenida con máquinas de estados es de mayor nivel de abstracción que una implementación en un lenguaje de programación. En consecuencia, es posible especificar una máquina de estados finito UML para definir los estados en los que debe estar un paciente de TB registrado en el sistema SNVS 2.0. Esta especificación permitirá luego evaluar si la secuencia de eventos registradas permite la evolución del estado de cada paciente a un estado válido definido o si en el sistema SNVS 2.0 no se ha registrado un evento que por el estado en que se encuentra, debería suceder. De esta manera, se emplearía la especificación de la máquina y su simulación a partir de los eventos registrados en SNVS 2.0 para ver si los pacientes transcurren adecuadamente por los estados, indicando que la información es consistente.

Para la especificación de la máquina de estados finitos se toma como guía las distintas propiedades definidas en cada evento de TB y sus respectivos valores. A partir de los mismos, se delinearon algunos estados. Esto se hizo teniendo en cuenta qué datos los condicionarían y cuáles serían los motivos causantes de una transición entre ellos. En un primer análisis se identificaron los estados definidos en la Tabla 1.

Estado	Descripción
Cargado en el sistema	Los datos del paciente son cargados en el sistema
En tratamiento	El paciente inicia un tratamiento.
Perdida de seguimiento	Se pierde contacto y/o seguimiento con el paciente al final de su tratamiento más reciente.
Curado	El tratamiento más reciente se considera exitoso.
Fracaso	El tratamiento más reciente fracasa.
En traslado	El paciente es derivado a otro efector de salud.
Fallecido	Se registra el fallecimiento del paciente.

Tabla 1 – Estados Identificados

Siguiendo los estados definidos se implementará una máquina de estados mediante la cual se realizaría la evaluación de la consistencia de los datos, analizando para cada paciente los eventos registrados en el SNVS 2.0. Para concluir que la base es consistente y está completa, es necesario que para cada paciente se inicie y finalice en un estado válido. Caso contrario, estaría existiendo un problema y no se estaría cumpliendo para alguno de los mismos el comportamiento considerado como correcto.

Resultados y discusión

Para realizar el análisis de los datos se utilizó el lenguaje R (<https://www.r-project.org/>), que facilita la manipulación de datos y el cálculo y visualización de los mismos (Riza, 2020). Se dispone de un archivo en formato csv con los registros de eventos de TB de SNVS 2.0 de los años 2019-2022. La base contiene 562.354 registros de eventos, de 50.597 casos (pacientes) distintos. Cada evento es descrito por una entrada del archivo, con 126 campos. En la Tabla 2 se incluye la descripción de los principales campos de interés.

Una vez cargada la base de datos se generaron distintas estructuras proporcionadas por el lenguaje que incluían parte de la información que se consideró relevante para los estados. Esto sirvió para analizar los valores que se repetían y en qué situaciones variaban, permitiendo también una mejor comprensión del sistema.

Campo de registro de evento	Descripción	Tipo	Posibles valores
ID_SNVS_EVENTO_CASO	Identificador único del evento	numérico	
FECHA_APERTURA	Fecha apertura del caso automático del sistema	Date	
FECHA_PAPEL	Fecha registrada del caso en el formulario papel u otro (fecha de notificación)	Date	
CLASIF_INICIO_TRAT	Clasificación al Inicio del tratamiento	character	Fracaso Nuevo Otros Pérdida del seguimiento Recuperado Recaída Traslado
RESULTADO_TRATAMIENTO	Resultado del tratamiento	character	Curado En tratamiento Fallecido Fracaso del tratamiento Pérdida de seguimiento Traslado Tratamiento completo
FECHA_INICIO_TRAT	Fecha de inicio del tratamiento	Date	
FECHA_FIN_TRAT	Fecha de fin del tratamiento	Date	
FECHA_FALLECIMIENTO	Fecha de defunción	Date	
CAUSA_FALLECIMIENTO	Causa de defunción relacionada con TB	character	
FALLECIDO	Indica el fallecimiento del paciente	character	Si No

Tabla 2 – Principales campos de registro de eventos

Con ayuda del análisis de los valores de cada una de las variables y qué podrían significar para el evento, se fueron delineando posibles transiciones entre los estados. En la Figura 1 se ilustra una primera aproximación a la máquina de estados.

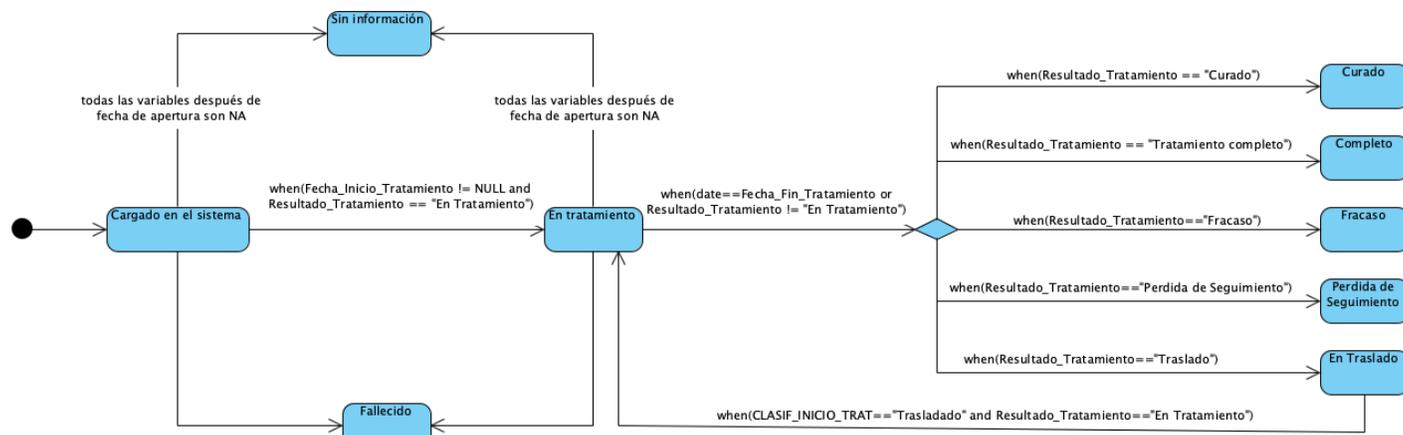


Figura 1 – Primer propuesta de máquina de estado de Paciente

La máquina de estados de la Figura 1 será implementada en R. Su implementación será utilizada para procesar los archivos csv con los registros de eventos e identificar problemas de completitud y validez de los datos disponibles. En la Figura 2 se ilustra una primera propuesta del algoritmo a ser implementado.

```

analisisCalidadDatos(archivoCSV) {
  maquinasSimuladas <- list()
  Para cada registro del archivoCSV {
    paciente <- obtenerPaciente(registro)
    Si esNuevo(paciente) {
      maquina <- inicializarMáquinaPaciente(paciente)
      maquinasSimuladas.incorporar(maquina)
      maquina.aplicarRegistro(registro)
    } sino {
      maquina <- maquinasSimuladas.obtenerMáquinaPaciente(paciente)
      maquina.aplicarRegistro(registro)
    }
    Si detectaProblema(maquina)
      informarProblema(maquina)
  }
  Para cada maquina de maquinasSimuladas {
    Si estadoNoValido(maquina)
      informarProblema(maquina)
  }
}
    
```

Figura 2 – Algoritmo para la Evaluación de la Calidad

Conclusiones

La propuesta se basa en la hipótesis de que el conjunto de registros de cada paciente se debe encontrar en todo momento en un estado definible por una máquina de estados finitos y que la ejecución de la secuencia de registros de cada paciente debería evolucionar la máquina a un estado válido. En caso de presentarse problemas de completitud o invalidez, la máquina no debería presentar este comportamiento. Se presentarían eventos que no serían aplicables al estado posible, o debería vencerse el tiempo que puede permanecer en cada estado. Para estos casos, se podría extender la máquina propuesta en la Figura 1 con estados trampas (Hopcroft et al., 2007), representando así explícitamente los problemas identificados.

La propuesta es un work in progress, se está realizando con el soporte del INER, quienes asisten en la validación de los estados identificados y sus posibles transiciones. Una vez finalizada la validación, se procederá a la implementación de la propuesta en R y a la evaluación de los resultados alcanzados. En caso de que sean favorables, se definirán los algoritmos necesarios empleados por el INER para sus actividades de mejora de la calidad de los datos del sistema de vigilancia.

Cabe destacar que otro de los propósitos contemplados es la posibilidad de generalizar este proceso de evaluación para que pueda ser aplicado a distintos conjuntos de datos de notificaciones sanitarias para otras enfermedades.

Referencias bibliográficas

Dirección de Respuesta al VIH, ITS, Hepatitis Virales y Tuberculosis (2022). “Boletín N° 5. Tuberculosis y Lepra en la Argentina”, Año V, Marzo de 2022. Disponible en <https://bancos.salud.gob.ar/recurso/boletin-ndeg-5-tuberculosis-y-lepra-en-la-argentina>.

German, R., Lee, L., Horan, J., Milstein, R., Pertowski, C., Waller, M. (2001). Updated guidelines for evaluating public health surveillance systems; recommendations from the Guidelines Working Group. *MMWR Recomm Rep.* 2001 Jul 27;50(RR-13):1-35. PMID: 18634202.

Harel, D. (1987). “Statecharts: A Visual Formalism for Complex Systems”. *Science of Computer Programming*, 8, 231-274.

Hopcroft, J., Motwani, R., Ullman, J. (2007). *Introducción a la teoría de autómatas lenguajes y computación*, 3ra. edición. Pearson Addison Wesley:Madrid.

Ministerio de Salud (2022). “Manual de normas y procedimientos de vigilancia y control de eventos de notificación obligatoria”. Ministerio de Salud de la Nación. Argentina. Disponible en <https://bancos.salud.gob.ar/recurso/manual-de-normas-y-procedimientos-de-vigilancia-y-control-de-eventos-de-notificacion>.

OMG (2017). “Unified Modeling Language. 2.5.1”. Disponible en <https://www.omg.org/spec/UML/2.5.1>.

Riza, L.S. (2020). “Data Science in R: Software and Implementations : Keynote Speakers 2”. Proceedings of the 6th International Conference on Science in Information Technology (ICSITech), doi: 10.1109/ICSITech49800.2020.9392050.

Secretaría de Salud (2018). “Instrucciones para la notificación de Tuberculosis al Sistema Nacional de Vigilancia de la Salud SNVS2.0”. Secretaría de Salud, Ministerio de Salud y Desarrollo Social de la Nación. Argentina. Disponible en <https://bancos.salud.gob.ar/recurso/instrucciones-para-la-notificacion-de-tuberculosis-al-snvs20>.

World Health Organization (2012). “Electronic recording and reporting for tuberculosis care and control”. Disponible en <https://apps.who.int/iris/handle/10665/44840>.

World Health Organization (2022). “Global tuberculosis report 2022”. Disponible en [who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022](https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022).

World Health Organization (2023). “Tuberculosis”. Disponible en <https://www.who.int/es/news-room/fact-sheets/detail/tuberculosis>.

Agradecimientos

Este trabajo fue realizado en el ámbito del PID SITCBFE0008464TC financiado por la UTN. Se agradece el aporte brindado a los miembros del INER para la realización de este trabajo.