

## Módulo 4: Análisis de datos

### Unidad 1: Calidad de los datos

1. Recolección de datos: herramientas de capturas de información de la vigilancia en salud pública.
2. Calidad de los datos: importancia y problemas en la calidad de datos, depuración de bases de datos
3. Tipo de errores comunes en el registro de los datos
4. Causas e impacto de la mala calidad de los datos y ejemplos
5. Pasos para mejorar la calidad de los datos

### Objetivos de aprendizaje

Al final de esta sesión, usted estará en capacidad de:

- Identificar los diferentes procedimientos de recolección de datos.
- Identificar la importancia y problemas en la calidad de datos
- Identificar el proceso de depuración de bases de datos
- Conocer los tipos de errores comunes en el registro de los datos
- Valorar causas e impacto de la mala calidad de los datos
- Establecer los pasos para mejorar la calidad de los datos

### Introducción

Los datos son las piezas individuales o recolección de hechos, cantidades, caracteres, símbolos y en general elementos crudos de conocimiento; que pueden ser persistidos y relacionados de alguna manera por la institución, ya sea en medio físico o electrónico, y que no es necesario que hayan tenido un procesamiento, cálculos o estructuras elaboradas previas en su proceso de construcción” (1).

El análisis de datos resulta fundamental en el propósito de generar información clave para la toma de decisiones, contribuyendo esto a la formulación de políticas basadas en la evidencia, así como a la comprensión de fenómenos intervinientes en los procesos de salud enfermedad y la afectación de los determinantes de la salud, entre otros (2).

El artículo 19 de la Ley 1751 de 2015 plantea: “Con el fin de alcanzar un manejo veraz, oportuno, pertinente y transparente de los diferentes tipos de datos generados por todos los actores, en sus diferentes niveles y su transformación en información para la toma de decisiones, se implementará una política que incluya un sistema único de información en salud, que integre los componentes demográficos, socioeconómicos, epidemiológicos, clínicos, administrativos y financieros. Los agentes del Sistema deben suministrar la información que requiera el Ministerio de Salud y Protección Social, en los términos y condiciones que se determine” (3).

---

#OrgullosamenteINS

 @INS Colombia  @insaludColombia  @INS Colombia

## 1. Recolección de datos: herramientas de capturas de información (3)

La recolección de datos es el componente más costoso y difícil de un sistema de vigilancia. Se pueden emplear diferentes metodologías de recolección de datos con el fin de garantizar la calidad, uniformidad y fiabilidad de los datos recolectados (4).

- Facilidad en la recolección: utilizar formularios estándar, que cumplan con características de claridad, sencillez, exigencia de solo información importante y ausencia de ambigüedad.
- Definiciones: definiciones normalizadas y específicas para recopilar información precisa.
- Puntualidad: formularios de recolección enviados a intervalos regulares, por ejemplo, diaria, semanal o mensual.
- Exhaustividad: la necesidad de que la detección de casos sea completa varía en función de la incidencia de la enfermedad vigilada.

Las actividades de recolección de datos son la detección, la notificación y la confirmación de los datos del evento de salud bajo vigilancia:

- Para la detección de casos se requiere aplicar una definición de caso estandarizada, así como definir los datos mínimos a recolectar y ubicar las fuentes de dichos datos.
- Para la notificación de casos se requiere identificar la red local de unidades notificadoras y el personal notificador, así como elaborar y difundir los procedimientos de notificación, incluidos los formularios y registros, la periodicidad de la notificación y el tipo de vigilancia que se pone en marcha.
- Para la clasificación de casos (sospechoso, probable y confirmado) se requiere contar con un procedimiento básico de seguimiento de los casos.
- Para la validación de los datos, debe existir un protocolo básico de control de calidad de los datos, incluyendo la integridad, consistencia, uniformidad y confiabilidad de los datos de vigilancia.

### Definición de caso de un evento de interés en salud pública

Una definición de caso es una serie de criterios estandarizados para decidir si una persona tiene una enfermedad particular u otra condición relacionada con la salud. Al utilizar una definición de caso estandarizada, se asegura que cada caso es diagnosticado de la misma manera, independientemente de cuándo o dónde ocurrió o quien lo identificó. Se puede comparar el número de casos de la enfermedad ocurridos en un tiempo y lugar dados con el número de casos ocurridos en otro tiempo y otro lugar (5).

La definición de caso es fundamental en el desarrollo de un sistema de vigilancia; debe ser simple y aceptable; es el instrumento básico para las actividades de recolección de datos de vigilancia, de esta depende la detección, la notificación y la clasificación de casos. La definición de caso para fines de vigilancia no es exactamente igual a la definición clínica del evento. La definición de caso debe ser lo suficientemente sensible para captar los casos verdaderos de forma sencilla y rápida y lo suficientemente específica para evitar que el número de casos falsos positivos sea excesivo.

Con fines epidemiológicos, el diagnóstico de un caso depende de la evidencia disponible, por lo cual la definición de caso debe distinguir niveles con criterios específicos a distintos grados de certeza diagnóstica como, por ejemplo, los siguientes:

---

**#OrgullosamenteINS**

 @INS Colombia  @insaludColombia  @INS Colombia

- Caso sospechoso: signos y síntomas compatibles con la enfermedad, sin evidencia alguna de laboratorio (ausente, pendiente o negativa).
- Caso probable: signos y síntomas compatibles con la enfermedad, sin evidencia definitiva de laboratorio.
- Caso confirmado: evidencia definitiva de laboratorio, con o sin signos y/o síntomas compatibles con la enfermedad.
- Caso por nexo o vínculo epidemiológico: antecedente de riesgo de contagio, por ejemplo, contacto con un caso confirmado de infección durante su periodo de transmisibilidad o haber estado en una zona de transmisión comprobada (6) .

### Procedimientos de recolección de datos para la vigilancia

Es importante reconocer que la recolección de datos para la vigilancia en salud pública no siempre se basa única o exclusivamente en la notificación rutinaria de casos atendidos en la consulta de todos los servicios de salud de una entidad territorial, que sigue siendo el modelo más aplicado por las unidades de epidemiología y los programas de control de enfermedades. En realidad, la vigilancia puede y debe proporcionar información relevante para la acción en salud a partir de la recolección de datos de distintas fuentes. Ello tiene un doble propósito: hacer más eficiente el proceso de recolección de datos y controlar la calidad de los datos. La necesidad de tener datos más detallados para vigilar un evento de salud dependerá de los objetivos específicos del sistema de vigilancia, los diferentes sistemas de recopilación de datos están resumidos en la siguiente tabla.

Procedimiento	Descripción
Vigilancia pasiva	El destinatario de los datos debe esperar a que los proveedores de datos informen
Vigilancia activa	Los datos se obtienen mediante la búsqueda de casos
Vigilancia centinela	Se basa en una muestra preestablecida de fuentes notificadoras que acuerdan notificar todos los casos
Vigilancia basada en el análisis de datos secundarios	Datos de los organismos sanitarios, disponibles con fines de vigilancia
Encuestas e investigaciones especiales de vigilancia	Pueden ser investigaciones epidemiológicas en terreno

Cada evento de salud bajo vigilancia, además de la definición de caso, tiene que estar apropiadamente identificado en función de un conjunto mínimo de datos sobre variables relacionadas al tiempo, lugar y persona, no solamente para los fines del análisis, sino fundamentalmente para facilitar la identificación de grupos de población objetivo de las medidas de control e intervenciones en salud pública que se emprenderán. Debe evitarse en todo momento la recolección de datos superfluos; es decir, sólo deberá recolectarse datos para los cuales se ha previsto una utilización específica y relevante para los propósitos de la vigilancia, no de una investigación exhaustiva.

**#OrgullosamenteINS**

 @INS Colombia
  @insaludColombia
  @INS Colombia

## Fuentes de datos para la vigilancia

Un variado número de fuentes de datos puede ser usado para la vigilancia en salud pública. En general, los datos pueden ser obtenidos a partir de reportes de registro rutinario, por esfuerzos especiales de investigación o a partir de bases de datos recolectados con otro propósito. Las fuentes de datos varían de lugar a lugar, dependiendo del nivel de desarrollo de los servicios de salud y otras instituciones, la calidad y cobertura de laboratorio, la disponibilidad de computadoras, redes informáticas y otros recursos y las características locales de las enfermedades (4).

Las fuentes de datos más comunes para vigilancia en salud pública son:

a) Notificación de casos: es el procedimiento medular de la vigilancia por medio del cual los servicios de salud informan rutinaria y obligatoriamente a la autoridad sanitaria sobre la atención de eventos sujetos a vigilancia.

b) Registros: son sistemas permanentes de consignación de eventos ejecutados por instituciones públicas o privadas donde se consigna regularmente la ocurrencia de ciertos eventos (nacimientos, defunciones, hospitalizaciones, inmunizaciones, accidentes de tránsito, contaminación ambiental, asistencia escolar y laboral, etc.). Los más comunes son:

- Registro civil (nacimientos, defunciones)
- Censos y anuarios estadísticos
- Informes de laboratorio
- Historias clínicas hospitalarias
- Informes de consulta externa y servicios de urgencia (públicos y privados)
- Registro de enfermedades de declaración obligatoria
- Registro de cáncer y de otras enfermedades crónicas
- Certificados médicos de defunción
- Protocolos de necropsia hospitalarios y forenses
- Monitoreo ambiental y climático
- Registros policiales de denuncias de hechos violentos
- Registros de asistencia y ausentismo escolar y laboral
- Registros veterinarios de reservorios animales
- Registros de venta y utilización de medicamentos y productos biológicos

c) Investigación de casos y brotes: es el procedimiento estandarizado de búsqueda activa y exhaustiva de información complementaria sobre uno o más casos asociados a determinado evento, usualmente como respuesta organizada ante la sospecha de epidemia, sea originada por rumores, vigilancia o análisis de registros.

d) Encuestas: son procedimientos de recolección de información por los cuales se obtiene información en un punto específico de tiempo sobre determinadas características de interés, generalmente no disponibles en otras fuentes de datos. Las encuestas más frecuentes son las serológicas, entomológicas, de morbilidad, socioeconómicas, etnográficas y las llamadas encuestas de demografía y salud.

---

#OrgullosamenteINS

 @INS Colombia  @insaludColombia  @INS Colombia

e) Rumores: son opiniones espontáneas y no confirmadas originadas en la comunidad y divulgadas por sus líderes y/o a través de los medios de comunicación de masas, asociadas al incremento de casos o muertes por una determinada causa.

## 2. Calidad de los datos: problemas en la calidad de datos, depuración de bases de datos

La calidad de un sistema de vigilancia se mide a menudo por la calidad de los datos recolectados. Desde la perspectiva de la vigilancia en salud pública, la calidad de los datos se puede definir como el nivel de precisión y completitud de la información de los datos notificados, recopilados y analizados (5). La calidad de los datos está influenciada por el desempeño de las pruebas de diagnóstico, la claridad de los formularios, la calidad de la capacitación y supervisión de las personas que completan estos formularios (7).

Se debe promover que los actores del sistema mejoren la calidad del registro de datos. Las variables que permiten la identificación, la localización y las características de personas son claves para la caracterización de los eventos (8).

Si los datos recolectados y notificados son lo suficientemente completos, precisos y oportunos, su análisis permite:

- Describir de forma precisa el comportamiento epidemiológico de la enfermedad
- Evaluar el estado de salud de la población
- Definir prioridades en salud pública
- Evaluar programas

### Problemas de la calidad de datos (9) (10)

Los problemas de calidad de los datos pueden ocurrir en cualquier paso del ciclo de vigilancia de la salud pública, desde la atención inicial del paciente en el sistema de salud hasta el análisis de datos a nivel central.

La mala calidad de la información puede generar consecuencias como:

- No identificar casos o eventos que tienen importancia para la salud pública.
- Análisis inadecuado del comportamiento de los eventos
- Las partes interesadas son informadas de manera incorrecta y tardíamente
- Pérdida de confianza, apoyo y financiamiento
- Toma de decisiones incorrectas
- Oportunidades no aprovechadas de identificar fortalezas o carencias en las actividades del programa
- Uso inadecuado de recursos al tener que dedicarlos a corregir los errores

Problemas	Descripción
-----------	-------------

#OrgullosamenteINS



@INS Colombia



@insaludColombia



@INS Colombia

Forma en que los datos están siendo consumidos por la entidad	Los datos deben tener la capacidad de poder ser utilizados bajo diferentes sistemas de información dentro de la entidad, se debe evaluar si esto es posible, aprovechando su potencial para futuros análisis.  <u>Ejemplo:</u> los datos pueden identificar cuales casos de sarampión no tenían esquema de vacunación completo para la edad, e informar al programa de vacunación.
Forma como se procesan los datos	Es importante verificar que después de la recolección, la forma o metodología con que se manipula y procesan los datos son los indicados.
Calidad a la hora de identificar el dato	Algunas veces los datos pueden ser similares, es por eso por lo que se debe asegurar que cada objeto contenga llaves o identificadores que los represente y distinga de otros.
Calidad del dato que es recolectado	Es importante verificar la forma con que se recolectan los datos en la entidad.  <ul style="list-style-type: none"> <li>• ¿Será que las herramientas usadas son las indicadas?</li> <li>• ¿La información recolectada es la correcta?</li> <li>• ¿Se están utilizando los adecuados filtros para la recolección?</li> <li>• ¿Los sistemas de información cuentan con atributos que permiten la trazabilidad de la información bajo estándares de seguridad?</li> </ul> Ejemplo: cuando se recolecta la edad de los casos de eventos de interés en salud pública, es fundamental para ayudar a orientar políticas de atención y prevención en salud.
Ciclo de vida del dato (confiabilidad de la información al pasar del tiempo.)	Los datos tienen un ciclo de vida y es muy importante identificar hasta qué punto es confiable la información después de haber pasado un cierto tiempo.
Calidad en cómo se está usando el dato	Igual que en el punto anterior es importante que los datos sean interpretados de la forma correcta.

Fuente: Alcaldía Mayor de Bogotá, Política de Calidad de los Datos y Plan de Calidad de la Información (11)

Algunos ejemplos de problemas con la calidad de los datos usados en la vigilancia en salud pública son:

Relacionados con los registros individuales:

- Datos perdidos
- Datos incorrectos
- Datos ilegibles

Relacionados con el sistema de información:

- Registros faltantes

**#OrgullosamenteINS**



@INS Colombia



@insaludColombia



@INS Colombia

- Registros duplicados
- Registros retrasados

### Depuración de bases de datos

La depuración de bases permite analizar los datos de manera fidedigna, fijando elementos aptos para la generación de información, para esto se requiere verificar los casos notificados por evento iniciando con las variables de procedencia/ocurrencia, lo que permite identificar los casos importados de los diferentes eventos notificados al Sivigila, esto aplica para país, departamento o municipio, según el caso (8).

Existen diversas estrategias para depurar o purificar una base de datos antes de su análisis. El enfoque principal radica en detectar registros duplicados, información repetida o registros incompletos que deben ser excluidos, todo ello con el objetivo de asegurar la calidad de los datos antes de su análisis.

### 3. Tipo de errores comunes en el registro de los datos

Existen variedad de razones relacionadas con la deficiente calidad del dato, resulta valioso conocer el motivo que genera esto e identificar dónde se crea este dato “erróneo” para ir a la raíz del problema y aplicar correctivos que reduzcan su ocurrencia (12).

Tipo de error	Descripción	Ejemplo
<b>Transposición</b>	Intercambio de los datos causado por error de digitación	Ingresar “39” como “93”
<b>Copiado</b>	Copiar o ingresar el dato incorrecto por similar apariencia	Cero “0” ingresado como letra “O”
<b>Codificación</b>	Uso de código incorrecto	“1” (SI) codificado como “2” (NO)
<b>Consistencia</b>	Dos o más respuestas contradictorias	Fecha de salida es anterior a la fecha de ingreso
<b>Intervalo</b>	El dato esta por fuera del intervalo de valores probables o posibles	Las opciones son de 1 a 5, pero se ingresó 7.

Fuente: INS – Curso Frontline, programa de entrenamiento en epidemiología de campo (10)

### 4. Causas e impacto de la mala calidad de los datos y ejemplos

#### Causas de la mala calidad de los datos

Los errores en la calidad del dato pueden ocurrir durante la recolección, al ingresarlos al sistema o al realizar su depuración y/o análisis. Las causas más frecuentes de los errores son:

#### Durante la recolección de datos

#OrgullosamenteINS

 @INS Colombia
  @insaludColombia
  @INS Colombia

- Formularios mal diligenciados
- Información suministrada incompleta (paciente)
- Información faltante en los registros
- Recolección tardía
- Notificación tardía
- Barreras de lenguaje

#### **Durante la entrada, depuración y análisis de datos**

- Errores de transcripción
- Errores de cálculo
- Errores de manejo de datos (datos perdidos, archivos perdidos, archivos incorrectos)

Uno de los principales motivos de una mala calidad de datos es el proceso de entrada manual de datos. Interviene el factor humano y como es normal puede haber errores de comunicación e interpretación, errores tipográficos, equivocaciones y otros factores externos. (12)

#### **Impacto de la mala calidad de los datos de vigilancia**

La mala calidad de la información de la vigilancia en salud pública puede tener un impacto negativo y grave ocasionando:

- Distorsión del comportamiento/aparición de la enfermedad
- No detección de eventos de interés en salud pública
- Brotes perdidos o no detectados
- Reducción en los recursos asignados
- Mal direccionamiento de los recursos
- Seguimiento y evaluación inadecuada de la eficiencia de los programas y/o intervenciones en salud pública

### **5. Pasos para mejorar la calidad de los datos**

- Protocolos, procedimientos y registros estandarizados: promover su uso y divulgación a los actores que intervienen en el proceso de vigilancia de la salud pública
- Garantizar capacitación y entrenamiento permanente en la recolección, registro, análisis e interpretación de los datos
- Realizar auditoría y seguimiento periódico a la calidad del dato en los diferentes niveles (local, distrital, departamental, nacional)
- Realimentación permanente a los actores del sistema
- Fomentar la precritica a los datos a notificar, antes de cargarlos al aplicativo en los diferentes niveles

---

**#OrgullosamenteINS**

 @INS Colombia  @insaludColombia  @INS Colombia

---

## Como realizar una realimentación constructiva

- Comience y termine la conversación señalando lo que está funcionando bien
- Explique los datos / sea fáctico
- Realice preguntas para comprender mejor el problema
- Pregúntele a la persona qué preguntas puede tener para usted.
- Asegúrele que el objetivo en común es encontrar la solución adecuada.
- Exprese la disponibilidad de guiar y ayudar incluso después de que termine la conversación.

## Resumen

- La recolección de datos es el componente más costoso y difícil de un sistema de vigilancia en salud pública.
- La calidad de los datos afecta todo el sistema de vigilancia
- Muchos tipos de errores pueden afectar la calidad de datos.
- Usted puede implementar medidas para garantizar la recolección, la consolidación y el almacenamiento de datos de calidad.

Proporcionar realimentación a quienes recopilan los datos, informándoles sobre los errores identificados y cómo evitarlos en el futuro. Esto crea conciencia y responsabilidad en la mejora de la calidad de los datos. ¡La realimentación funciona!

---

**#OrgullosamenteINS**

 @INS Colombia  @insaludColombia  @INS Colombia

## Bibliografía

1. Ministerio de Salud y Protección Social, Plan de calidad de los componentes de la información-2020.pdf [Internet]. [citado el 10 de abril de 2023]. Disponible en: <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/OT/plan-calidad-componentes-informacion-minsalud-2020.pdf>
2. Vinculación de datos de NCHS: darse cuenta del poder de los datos [Internet]. 2021 [citado el 9 de abril de 2023]. Disponible en: <https://www.cdc.gov/nchs/data-linkage/datalinkagestory.htm>
3. Congreso de Colombia, Ley Estatutaria N° 1751 de 2015 [Internet]. [citado el 9 de abril de 2023]. Disponible en: [https://www.minsalud.gov.co/Normatividad\\_Nuevo/Ley%201751%20de%202015.pdf](https://www.minsalud.gov.co/Normatividad_Nuevo/Ley%201751%20de%202015.pdf)
4. Declich S, Carter AO. Public health surveillance: historical origins, methods and evaluation. Bull World Health Organ. 1994;72(2):285–304.
5. U.S. Department of health and human services, Centers for Disease Control and Prevention (CDC). Principles of Epidemiology in Public Health Practice, Third Edition: An Introduction to Applied Epidemiology and Biostatistics [Internet]. [citado el 9 de abril de 2023]. Disponible en: <https://www.cdc.gov/csels/dsepd/ss1978/ss1978.pdf>
6. Organización Panamericana de la Salud (OPS). COVID-19 Glosario sobre brotes y epidemias. Un recurso para periodistas y comunicadores [Internet]. 2020 [citado el 3 de agosto de 2023]. Disponible en: <https://www.paho.org/es/documentos/covid-19-glosario-sobre-brotes-epidemias-recurso-para-periodistas-comunicadores>
7. German RR, Horan JM, Lee LM, Milstein B, Pertowski CA. Updated guidelines for evaluating public health surveillance systems; recommendations from the Guidelines Working Group. Centers for Disease Control and Prevention (U.S.), editor. el 27 de julio de 2001; Disponible en: <https://stacks.cdc.gov/view/cdc/13376>
3. Instituto Nacional de Salud, Colombia. Lineamientos nacionales de vigilancia en salud pública 2023.pdf [Internet]. [citado 9 de abril de 2023]. Disponible en: <https://www.ins.gov.co/BibliotecaDigital/lineamientos-nacionales-2023.pdf>
9. FETPF-Curriculum-Guide2.0\_11-2020-508.pdf [Internet]. [citado el 9 de abril de 2023]. Disponible en: [https://www.cdc.gov/globalhealth/healthprotection/fetp/pdf/FETPF-Curriculum-Guide2.0\\_11-2020-508.pdf](https://www.cdc.gov/globalhealth/healthprotection/fetp/pdf/FETPF-Curriculum-Guide2.0_11-2020-508.pdf)
10. Instituto Nacional de Salud. Curso Frontline, Programa de entrenamiento en epidemiología de campo, Módulo 1, unidad 5 calidad de los datos.
11. Alcaldía Mayor de Bogotá. Política de Calidad de los Datos y Plan de Calidad de la Información [Internet]. [citado el 9 de abril de 2023]. Disponible en: [https://www.umv.gov.co/sisgestion2019/Documentos/ESTRATEGICOS/EGTI/EGTI-DI-015-V1\\_Politica\\_Calidad\\_de\\_Datos\\_y\\_Plan\\_de\\_Calidad\\_de\\_la\\_Informacion.docx](https://www.umv.gov.co/sisgestion2019/Documentos/ESTRATEGICOS/EGTI/EGTI-DI-015-V1_Politica_Calidad_de_Datos_y_Plan_de_Calidad_de_la_Informacion.docx)
12. Balagueró T. Deusto. 2018 [citado el 10 de abril de 2023]. 4 causas de mala calidad de los datos en Big Data. Disponible en: <https://www.deustoformacion.com/blog/gestion-empresas/4-causas-mala-calidad-datos-big-data>

---

#OrgullosamenteINS



@INS Colombia



@insaludColombia



@INS Colombia