

KNO-016 · Cómo se construye la evidencia: integrando las herramientas

KNO · Formas de conocer

Juan Segurola

v0.6 · ES

Todos los derechos reservados

Formas de conocer

Perfil epistémico: Interpretación científica actual **Licencia al publicar:** Todos los derechos reservados

Resumen del artículo

La evidencia no es una sustancia que aparezca automáticamente cuando se acumulan datos. ¿Qué debe conectarse para que una observación sea relevante para una afirmación científica? Este artículo sigue esa pregunta desde la formulación de afirmaciones y la medición hasta el diseño de estudios, la modelización, la identificación causal, la incertidumbre, la replicación, la reproducibilidad, la triangulación, la directitud, la transparencia y el escrutinio colectivo. Pregunta qué vínculos son empíricos, cuáles dependen de supuestos, cuáles pueden probarse de forma independiente y cuáles siguen siendo vulnerables a errores compartidos. También pregunta por qué más estudios no significan necesariamente evidencia más fuerte si repiten las mismas debilidades y por qué la evidencia puede volverse más fiable sin convertirse en certeza. El objetivo es hacer explícita la arquitectura del apoyo probatorio para que otro lector vea no solo la conclusión propuesta, sino cómo debería sostenerla cada paso.

En 60 segundos

La evidencia científica es un argumento estructurado con anclajes empíricos. Una afirmación gana credibilidad cuando las mediciones son trazables, el diseño aborda las alternativas relevantes, el modelo es apropiado para la pregunta, la incertidumbre se representa honestamente y análisis o estudios independientes exponen la afirmación a modos de fallo diferentes.

Ninguna etiqueta como «revisado por pares», «estadísticamente significativo», «replicado» o «aleatorizado» garantiza corrección. La fuerza de la evidencia depende del ajuste entre pregunta y método, de la magnitud y dirección de los posibles sesgos, de la consistencia entre estudios relevantes, de la precisión, de la directitud y de la transparencia. La idea práctica es separar el resultado observado de los supuestos que lo conectan con la interpretación. Cada herramienta reduce algunas ambigüedades y deja otras abiertas; por eso el análisis debe declarar qué puede identificar, qué depende de condiciones adicionales y qué observación podría contradecir la conclusión. La precisión formal no elimina por sí sola esos límites.

Por qué importa esta pregunta

Los científicos rara vez observan una conclusión directamente. Observan huellas, medidas o resultados y los conectan con una afirmación mediante diseño, modelos, supuestos y conocimiento de fondo. Cuando esos vínculos quedan implícitos, el desacuerdo es difícil de diagnosticar y los

fallos de replicación son difíciles de interpretar. Hacer explícita la cadena probatoria aclara dónde un resultado es fuerte, dónde depende de supuestos y qué nueva observación podría cambiarlo. Eso importa para la práctica científica y para decidir cuánta confianza merece una afirmación.

1. La evidencia empieza con una afirmación suficientemente precisa para poder fallar

Antes de preguntar si la evidencia es fuerte, hay que saber qué afirmación se está apoyando.

[C01] Una afirmación científica debe identificar su objetivo con suficiente claridad para que observaciones o análisis puedan contar en su contra. Las narrativas amplias que cambian de significado después de cada resultado son difíciles de someter a prueba porque el objetivo evidencial se desplaza. [S01, S02]

Las definiciones operacionales no son mero trámite. Conectan el lenguaje teórico con procedimientos observables. Una misma palabra puede referirse a estimandos, mediciones o clasificaciones diferentes. Si el objetivo cambia entre diseño, análisis e interpretación, la evidencia aparente puede responder a una pregunta distinta de la planteada inicialmente.

En qué punto estamos

Antes de preguntar si la evidencia es fuerte, hay que saber qué afirmación se está apoyando. Las definiciones operacionales no son mero trámite. Conectan el lenguaje teórico con procedimientos observables. Una misma palabra puede referirse a estimandos, mediciones o clasificaciones diferentes. La sección siguiente examina la idea de que las observaciones se producen, no simplemente se encuentran.

2. Las observaciones se producen, no simplemente se encuentran

KNO-008 mostró que los instrumentos amplían los sentidos; KNO-011 mostró que los números dependen de convenciones de medida. Aquí esas lecciones pasan a formar parte de la cadena de evidencia.

[C02] El valor evidencial de una medición depende de cómo se definió el mensurando o variable, cómo se calibró el instrumento o procedimiento, qué correcciones se aplicaron y qué incertidumbre permanece. Una salida numérica no se valida a sí misma. [S03]

Una fotografía, un espectro, un ensayo analítico o una respuesta de encuesta ya son el resultado de decisiones sobre muestreo, resolución, procesamiento y clasificación. La evidencia empieza antes del análisis estadístico.

En qué punto estamos

KNO-008 mostró que los instrumentos amplían los sentidos; KNO-011 mostró que los números dependen de convenciones de medida. Aquí esas lecciones pasan a formar parte de la cadena de evidencia. Una fotografía, un espectro, un ensayo analítico o una respuesta de encuesta ya son el resultado de decisiones sobre muestreo, resolución, procesamiento y clasificación. La sección siguiente examina la idea de que el diseño determina qué alternativas pueden excluirse.

3. El diseño determina qué alternativas pueden excluirse

Preguntas distintas requieren diseños distintos.

[C03] Controles experimentales, aleatorización, cegamiento, medición repetida y estructuras comparativas observacionales abordan sesgos potenciales diferentes. Ninguno es un marcador universal de calidad independiente de la pregunta científica. [S04]

Para preguntas causales, el diseño determina qué contrastes contrafactuales pueden identificarse plausiblemente. Para preguntas descriptivas, un muestreo representativo puede importar más que la asignación aleatoria de tratamientos. En ciencias históricas, huellas convergentes y consistencia mecanística pueden soportar la evidencia cuando intervenir es imposible.

En qué punto estamos

Preguntas distintas requieren diseños distintos. Para preguntas descriptivas, un muestreo representativo puede importar más que la asignación aleatoria de tratamientos. En ciencias históricas, huellas convergentes y consistencia mecanística pueden soportar la evidencia cuando intervenir es imposible. El diseño limita qué explicaciones siguen siendo compatibles. La sección siguiente examina la idea de que los modelos son puentes entre observaciones y afirmaciones.

4. Los modelos son puentes entre observaciones y afirmaciones

Los datos rara vez hablan directamente de la proposición de interés. Un modelo traduce observaciones en cantidades, mecanismos o predicciones.

[C04] Toda afirmación evidencial basada en modelos está condicionada en algún grado por elecciones de modelización. Una buena práctica distingue lo directamente observado de lo inferido a través del modelo y prueba la sensibilidad a alternativas plausibles. [S06]

No es un defecto exclusivo de las simulaciones complejas. Incluso un ajuste lineal, una curva de calibración o un umbral diagnóstico constituyen modelos.

En qué punto estamos

Los datos rara vez hablan directamente de la proposición de interés. No es un defecto exclusivo de las simulaciones complejas. Incluso un ajuste lineal, una curva de calibración o un umbral diagnóstico constituyen modelos. Lo importante aquí es la relación declarada entre evidencia, método e inferencia, no una promesa genérica de certeza. La sección siguiente examina la idea de que la evidencia causal exige más que asociación.

5. La evidencia causal exige más que asociación

KNO-014 fijó la frontera causal central.

[C05] Una asociación observada puede apoyar una afirmación causal solo mediante supuestos y características del diseño que aborden confusión, causalidad inversa, selección y otras explicaciones competidoras. Por tanto, la fuerza de la evidencia causal depende de la estrategia de identificación, no solo del tamaño o significación estadística de la asociación. [S04, S07]

Pueden combinarse métodos causales distintos, pero múltiples análisis que compartan el mismo sesgo oculto no constituyen confirmación independiente.

En qué punto estamos

KNO-014 fijó la frontera causal central. Pueden combinarse métodos causales distintos, pero múltiples análisis que compartan el mismo sesgo oculto no constituyen confirmación independiente. La conclusión resulta útil solo en el papel definido por esta sección, manteniendo visibles sus condiciones probatorias. La sección siguiente examina la idea de que la incertidumbre pertenece dentro de la afirmación.

6. La incertidumbre pertenece dentro de la afirmación

KNO-015 trató la incertidumbre como objeto metodológico y no como un intervalo decorativo.

[C06] La evaluación de evidencia debe distinguir incertidumbre de medida, muestreo, modelización y otras fuentes reconocidas. Un intervalo estrecho dentro de un modelo no establece que el error sistemático no modelizado sea pequeño. [S03, S08]

Una afirmación fuerte especifica por ello no solo una estimación, sino también las condiciones bajo las cuales se espera que sea fiable.

En qué punto estamos

KNO-015 trató la incertidumbre como objeto metodológico y no como un intervalo decorativo. Una afirmación fuerte especifica por ello no solo una estimación, sino también las condiciones bajo las cuales se espera que sea fiable. Su valor está en hacer inspeccionable este paso inferencial concreto, no en tratar el método como autojustificado. La sección siguiente examina la idea de que replicación y reproducibilidad prueban vínculos diferentes.

7. Replicación y reproducibilidad prueban vínculos diferentes

Los términos se confunden con frecuencia.

[C07] Las National Academies distinguen reproducibilidad computacional, obtener resultados computacionales consistentes usando los mismos datos, código y métodos, de replicabilidad, obtener resultados consistentes en estudios que abordan la misma pregunta científica con nuevos datos. [S01]

[C08] Una replicación exitosa aumenta la confianza pero no demuestra que una afirmación original sea universalmente correcta; una replicación fallida tampoco la refuta por sí sola de forma concluyente, porque las diferencias pueden proceder de heterogeneidad real, variables no controladas, defectos metodológicos u otras causas. [S01]

Esta asimetría importa. La repetición aporta evidencia sobre robustez, no funciona como un operador lógico de demostración.

En qué punto estamos

Los términos se confunden con frecuencia. La repetición aporta evidencia sobre robustez, no funciona como un operador lógico de demostración. Lo importante aquí es la relación declarada entre evidencia, método e inferencia, no una promesa genérica de certeza. La sección siguiente examina la idea de que la triangulación es más fuerte cuando los errores son distintos.

8. La triangulación es más fuerte cuando los errores son distintos

Evidencia independiente no significa simplemente «más estudios».

[C09] La triangulación gana fuerza cuando métodos diferentes abordan la misma pregunta y poseen fuentes clave de sesgo diferentes. La convergencia es más informativa cuando no se espera que los sesgos de todos los enfoques empujen el resultado en la misma dirección. [S07]

Diez estudios con el mismo instrumento, mecanismo de selección y modelo pueden reproducir un sesgo compartido. Un conjunto menor con diseños genuinamente diferentes puede tener mayor valor diagnóstico.

En qué punto estamos

Evidencia independiente no significa simplemente «más estudios». Un conjunto menor con diseños genuinamente diferentes puede tener mayor valor diagnóstico. La conclusión resulta útil solo en el papel definido por esta sección, manteniendo visibles sus condiciones probatorias. La sección siguiente examina la idea de que la evidencia es acumulativa, pero acumular exige estructura.

9. La evidencia es acumulativa, pero acumular exige estructura

Meta-análisis, revisiones sistemáticas y marcos de evidencia pueden combinar estudios, aunque la combinación solo tiene sentido cuando las preguntas subyacentes son suficientemente compatibles.

[C10] Marcos como GRADE examinan explícitamente riesgo de sesgo, inconsistencia, indirectitud, imprecisión y sesgo de difusión/publicación en lugar de tratar por sí solo el número de estudios como certeza. [S09]

No se afirma que GRADE gobierne toda la ciencia. Ilustra una arquitectura general: los cuerpos de evidencia deben evaluarse por error sistemático, coherencia, relevancia y precisión, no solo agregarse.

En qué punto estamos

Meta-análisis, revisiones sistemáticas y marcos de evidencia pueden combinar estudios, aunque la combinación solo tiene sentido cuando las preguntas subyacentes son suficientemente compatibles. No se afirma que GRADE gobierne toda la ciencia. Ilustra una arquitectura general: los cuerpos de evidencia deben evaluarse por error sistemático, coherencia, relevancia y precisión, no solo agregarse. La sección siguiente examina la idea de que la directitud importa.

10. La directitud importa

Una evidencia puede ser exacta y, aun así, responder a la pregunta equivocada.

Un sustituto de laboratorio puede no predecir un resultado clínico. Un resultado en una población puede no transportarse a otra. Una variable proxy puede medir solo una parte del constructo previsto.

[C11] El concepto de indirectitud formaliza desajustes entre la pregunta de interés y las poblaciones, intervenciones o exposiciones, comparadores y resultados realmente estudiados. [S09, S10]

La validez externa y la transportabilidad pertenecen, por tanto, a la evaluación de evidencia y no a un añadido posterior a la validez interna.

En qué punto estamos

Una evidencia puede ser exacta y, aun así, responder a la pregunta equivocada. Un sustituto de laboratorio puede no predecir un resultado clínico. Un resultado en una población puede no transportarse a otra. Una variable proxy puede medir solo una parte del constructo previsto. La sección siguiente examina la idea de que la transparencia cambia lo que otros pueden comprobar.

11. La transparencia cambia lo que otros pueden comprobar

La evidencia es más fuerte cuando puede inspeccionarse el camino desde el material bruto hasta la conclusión.

[C12] Una comunicación clara de métodos, pasos de procesamiento de datos, elecciones analíticas e incertidumbre facilita la comprobación computacional y ayuda a otros investigadores a identificar errores o probar robustez. La reproducibilidad se favorece mediante acceso a los artefactos digitales pertinentes y métodos suficientemente completos. [S01, S02]

La transparencia no vuelve correcto un análisis defectuoso. Hace que sus defectos sean más descubribles.

En qué punto estamos

La evidencia es más fuerte cuando puede inspeccionarse el camino desde el material bruto hasta la conclusión. La transparencia no vuelve correcto un análisis defectuoso. Hace que sus defectos sean más descubribles. La inspección externa depende de esa trazabilidad. La sección siguiente examina la idea de que revisión por pares y consenso son filtros, no sustitutos de la evidencia.

12. Revisión por pares y consenso son filtros, no sustitutos de la evidencia

La revisión por pares puede detectar errores, pedir aclaraciones y mejorar la comunicación, pero no reproduce de forma independiente el experimento subyacente. El consenso puede resumir un panorama evidencial maduro, pero unos supuestos compartidos también pueden propagarse por un campo.

Estos procesos institucionales pertenecen plenamente a KNO-020. Aquí la frontera es más simple: la validación social puede reforzar el control de calidad, pero la cadena evidencial debe seguir conectando observaciones y afirmaciones.

En qué punto estamos

La revisión por pares puede detectar errores, pedir aclaraciones y mejorar la comunicación, pero no reproduce de forma independiente el experimento subyacente. El consenso puede resumir un panorama evidencial maduro, pero unos supuestos compartidos también pueden propagarse por un campo. Estos procesos institucionales pertenecen plenamente a KNO-020. La sección siguiente examina una auditoría práctica de evidencia.

13. Una auditoría práctica de evidencia

Antes de aceptar una conclusión científica fuerte, preguntar:

1. **¿Cuál es exactamente la afirmación?** ¿Descriptiva, predictiva, causal, mecánica o histórica?

2. **¿Qué observaciones la anclan?** ¿Qué se midió o registró directamente?
3. **¿Cómo se construyeron las variables?** ¿Qué clasificaciones, umbrales o transformaciones intervinieron?
4. **¿Qué diseño aborda las alternativas?** ¿Controles, aleatorización, emparejamiento, orden temporal, variación natural, casos comparativos?
5. **¿Qué modelo conecta datos y afirmación?** ¿Qué supuestos son esenciales?
6. **¿Qué incertidumbres están cuantificadas?** ¿Cuáles siguen reconocidas pero no cuantificadas?
7. **¿Qué sesgos podrían desplazar la respuesta?** ¿En qué dirección?
8. **¿Se ha reproducido el cálculo?** ¿Puede ejecutarse de nuevo el camino analítico?
9. **¿Se ha replicado o triangulado el hallazgo?** ¿Los nuevos modos de fallo son realmente distintos?
10. **¿Qué tan directa es la evidencia para esta población, cantidad o mecanismo objetivo?**
11. **¿Es consistente con un cuerpo de evidencia más amplio?** Si no, ¿es informativo el desacuerdo?
12. **¿Qué conclusión está realmente justificada?** No convertir «compatible con» en «demostrado».

En qué punto estamos

Antes de aceptar una conclusión científica fuerte, preguntar: ¿Qué incertidumbres están cuantificadas? ¿Cuáles siguen reconocidas pero no cuantificadas? ¿Qué observaciones la anclan? ¿Qué se midió o registró directamente? ¿Cómo se construyeron las variables? ¿Qué clasificaciones, umbrales o transformaciones intervinieron? La sección siguiente examina la idea de que la evidencia puede fortalecerse sin convertirse en certeza.

14. La evidencia puede fortalecerse sin convertirse en certeza

La ciencia acumula restricciones. Las mediciones estrechan valores posibles; los experimentos eliminan alternativas; los modelos organizan implicaciones; métodos independientes prueban vulnerabilidades distintas; la replicación explora robustez; la síntesis muestra dónde los resultados convergen y dónde no.

El punto final rara vez es la certeza. Es una afirmación cuyas alternativas restantes resultan progresivamente más difíciles de sostener.

En qué punto estamos

Las mediciones estrechan valores posibles; los experimentos eliminan alternativas; los modelos organizan implicaciones; métodos independientes prueban vulnerabilidades distintas; la replicación explora robustez; la síntesis muestra dónde los resultados convergen y dónde no. Esto delimita el alcance del artículo y prepara la síntesis final sin ampliar la afirmación más allá de lo establecido en las secciones anteriores.

Nota de frontera

KNO-016 es deliberadamente un **artículo de síntesis**. No sustituye los métodos detallados de KNO-008 a KNO-015 y no se adelanta a KNO-020 sobre revisión y consenso ni a KNO-021 sobre ciencia abierta. Su función es mostrar cómo medición, diseño experimental, modelización, causalidad, incertidumbre, reproducibilidad, replicación y triangulación se combinan en un argumento evidencial auditable.

Lo que sabemos y cómo lo sabemos

La evidencia se construye conectando afirmaciones precisas con observaciones mediante medida, diseño, modelos y supuestos inferenciales. Replicación, reproducibilidad, triangulación, directitud y transparencia refuerzan vínculos distintos, sobre todo cuando sus modos de fallo no son compartidos.

Lo que todavía no sabemos

Ninguna auditoría garantiza que se hayan imaginado todas las alternativas relevantes, que todas las fuentes de sesgo sean independientes o que un modelo bien apoyado sea la única explicación posible. La evidencia puede fortalecerse sin convertirse en conclusión lógica final.

Qué podría hacernos cambiar de opinión

La evaluación probatoria debe cambiar si falla la calibración de una medida clave, el diseño permite una explicación alternativa, se rompe una predicción del modelo, una replicación comparable fracasa o un método con estructura de error diferente contradice el resultado existente.

Si solo recuerdas una cosa

La evidencia fuerte no es una pila de resultados, sino una cadena de apoyo cuyos supuestos, errores y pruebas pueden inspeccionarse.

Artículos relacionados

- KNO-012 · Experiments and Variable Control
- KNO-013 · Models, Maps and Reality
- KNO-014 · Causality and Correlation
- KNO-015 · Uncertainty and Probability
- KNO-020 · Review, Consensus and Disagreement

Referencias

[S01] National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press, 2019. DOI: 10.17226/25303.

[S02] National Academies of Sciences, Engineering, and Medicine. “Understanding Reproducibility and Replicability,” chapter 3 in *Reproducibility and Replicability in Science*, pp. 39–54. Washington, DC: The National Academies Press, 2019. DOI: 10.17226/25303.

- [S03] Joint Committee for Guides in Metrology (JCGM). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*. JCGM 100:2008(E). DOI: 10.59161/JCGM100-2008E.
- [S04] Hernán, Miguel A. “A Definition of Causal Effect for Epidemiological Research.” *Journal of Epidemiology and Community Health* 58(4), 2004, 265–271. DOI: 10.1136/jech.2002.006361.
- [S05] Schulz, Kenneth F.; Altman, Douglas G.; Moher, David; CONSORT Group. “CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials.” *BMJ* 340, 2010, c332. DOI: 10.1136/bmj.c332.
- [S06] Frigg, Roman; Hartmann, Stephan. “Models in Science.” *The Stanford Encyclopedia of Philosophy*, edición de verano de 2025; revisión sustantiva del 2 de abril de 2025. <https://plato.stanford.edu/archives/sum2025/entries/models-science/>
- [S07] Lawlor, Debbie A.; Tilling, Kate; Davey Smith, George. “Triangulation in aetiological epidemiology.” *International Journal of Epidemiology* 45(6), 2016, 1866–1886. DOI: 10.1093/ije/dyw314.
- [S08] Mastrandrea, Michael D.; Field, Christopher B.; Stocker, Thomas F.; Edenhofer, Ottmar; Ebi, Kristie L.; Frame, David J.; Held, Hermann; Kriegler, Elmar; Mach, Katharine J.; Matschoss, Patrick R.; Plattner, Gian-Kasper; Yohe, Gary W.; Zwiers, Francis W. *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*. IPCC, 2010. <https://www.ipcc.ch/site/assets/uploads/2018/05/uncertainty-guidance-note.pdf>
- [S09] Neumann, Ignacio; Brennan, Sue; Meerpohl, Joerg; Davoli, Marina; Alonso Coello, Pablo; Akl, Elie; Dahm, Philipp; Skoetz, Nicole; Xia, Jun; Brozek, Jan; Schünemann, Holger J. “Overview of the GRADE approach.” *GRADE Book*. Last modified 12 May 2026; consulted 5 September 2026. <https://book.gradepro.org/guideline/overview-of-the-grade-approach>
- [S10] “Indirectness.” *GRADE Book*. Last modified 12 May 2026; consulted 5 September 2026. <https://book.gradepro.org/guideline/indirectness>. Author display currently contains an apparent duplicated Holger J. Schünemann entry; author chain therefore remains P pending human confirmation and is not normalised in this bibliography.