Revista esprint Investigación

https://rei.esprint.tech
Vol. 5 N° Esp.2, Edición Especial 2026 (380-392) ISSN: 2960-8317
Salud y Ciencias de la Vida

Artículo de revisión

Relación entre pruebas de estimulación de hormona del crecimiento y biomarcadores séricos en talla baja pediátrica

Relationship between growth hormone stimulation tests and serum biomarkers in paediatric short stature

Jhoel Stalin Ortiz Avila*
Universidad Nacional de Chimborazo
Riobamba - Ecuador
jhoel.ortiz@unach.edu.ec
https://orcid.org/0009-0006-6438-8691

Ivan Marcelo Cantuña Vallejo
Universidad Nacional de Chimborazo
Riobamba - Ecuador
ivan.cantuna@unach.edu.ec
https://orcid.org/0000-0003-4975-1596

***Correspondencia:**
jhoel.ortiz@unach.edu.ec

Cómo citar este artículo:
Ortiz, J., & Cantuña, I. (2026). Relación entre pruebas de estimulación de hormona del crecimiento y biomarcadores séricos en talla baja pediátrica. *Esprint Investigación, 5*(2), 380-392. https://doi.org/10.61347/ei.v5iEsp.2.410

Recibido: *18 de septiembre de 2026*
Aceptado: *23 de octubre de 2026*
Publicado: *4 de noviembre de 2026*

Copyright: Derechos de autor 2026 Jhoel Stalin Ortiz Avila, Ivan Marcelo Cantuña Vallejo.

Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial 4.0.

Resumen: La deficiencia de hormona del crecimiento (GHD) representa un desafío diagnóstico debido a la secreción pulsátil de esta hormona y las limitaciones de las pruebas de estimulación y los biomarcadores séricos. El objetivo fue evaluar la exactitud diagnóstica del factor de crecimiento similar a la insulina tipo 1 (IGF-1), la proteína de unión IGFBP-3 y su cociente frente a las pruebas de estimulación de hormona del crecimiento (GH) en niños con talla baja. Se realizó una revisión sistemática siguiendo las directrices PRISMA 2020, mediante búsquedas en PubMed/MEDLINE, ScienceDirect y DOAJ, considerando publicaciones entre enero de 2021 y abril de 2026. El riesgo de sesgo se evaluó mediante QUADAS-2. De 168 registros identificados, se incluyeron 10 estudios con 2993 participantes. Se analizaron indicadores de sensibilidad, especificidad, valores predictivos, área bajo la curva (AUC) y razones de verosimilitud. El IGF-1 aislado presentó sensibilidades del 42,2 % al 79,0 % y especificidades del 41,7 % al 90,1 %, con una capacidad discriminativa limitada en una cohorte (AUC = 0,517). La IGFBP-3 alcanzó una sensibilidad del 53,1 % y una especificidad del 81,3 %, mientras que el cociente IGF-1/IGFBP-3 registró valores del 87,5 % y 83,0 %, respectivamente. La combinación de IGF-1 bajo con clonidina mostró especificidades elevadas, aunque con sensibilidades variables según el punto de corte. Se concluyó que los biomarcadores basales presentaron un rendimiento diagnóstico heterogéneo y constituyeron herramientas complementarias, sin sustituir las pruebas de estimulación de GH.

Palabras clave: Factor I del crecimiento similar a la insulina, hormona de crecimiento humana, proteína 3 de unión a factor de crecimiento similar a la insulina, trastornos del crecimiento.

Abstract: *The diagnosis of growth hormone deficiency (GHD) remains challenging due to the pulsatile secretion of growth hormone and the limitations of stimulation tests and serum biomarkers. The objective was to evaluate the diagnostic accuracy of insulin-like growth factor 1 (IGF-1), insulin-like growth factor-binding protein 3 (IGFBP-3), and their ratio compared with growth hormone (GH) stimulation tests in children with short stature. A systematic review was conducted following PRISMA 2020 guidelines, using PubMed/MEDLINE, ScienceDirect, and DOAJ to identify studies published between January 2021 and April 2026. Risk of bias was assessed using QUADAS-2. Of 168 records identified, 10 studies involving 2,993 participants were included. Sensitivity, specificity, predictive values, area under the curve (AUC), and likelihood ratios were analyzed. IGF-1 alone showed sensitivity ranging from 42.2% to 79.0% and specificities from 41.7% to 90.1%, with limited discriminatory ability in one cohort (AUC = 0.517). IGFBP-3 achieved a sensitivity of 53.1% and a specificity of 81.3%, whereas the IGF-1/IGFBP-3 ratio yielded corresponding values of 87.5% and 83.0%. Combining low IGF-1 levels with a clonidine stimulation test resulted in high specificities, although sensitivities varied according to the cutoff value. In conclusion, basal biomarkers demonstrated heterogeneous diagnostic performance and served as complementary tools rather than replacements for GH stimulation tests.*

Keywords*: Growth disorders, human growth hormone, insulin-like growth factor I, insulin-like growth factor-binding protein 3 I.*

1. Introducción

La talla baja, definida como una estatura inferior a −2 desviaciones estándar (DE) para la edad y el sexo, comprende desde variantes normales del crecimiento hasta alteraciones endocrinas, como la deficiencia de hormona del crecimiento (GHD, por sus siglas en inglés). Esta condición afecta aproximadamente a uno de cada 4000 a 10 000 niños, por lo que su identificación constituye un desafío diagnóstico que requiere diferenciar a los pacientes afectados de aquellos con variantes normales del crecimiento, evitando procedimientos invasivos y tratamientos innecesarios (Hage et al., 2021; Penco et al., 2021).

La hormona del crecimiento (GH) presenta una secreción pulsátil regulada por la hormona liberadora de GH (GHRH) y la somatostatina, lo que limita el valor diagnóstico de una medición aislada. Por ello, se utilizan pruebas de estimulación farmacológica para evaluar la respuesta secretora de GH, complementadas con la determinación del factor de crecimiento similar a la insulina tipo 1 (IGF-1) y su proteína de unión IGFBP-3. Aunque estos biomarcadores presentan menor variabilidad durante el día, su capacidad para identificar la deficiencia hormonal sigue siendo objeto de debate (Yau & Rapaport, 2022; Yuen et al., 2023).

La evidencia científica señala limitaciones en la interpretación de estos indicadores, debido a que las concentraciones de IGF-1 están condicionadas por factores nutricionales, hepáticos y puberales, además de presentar valores superpuestos entre pacientes con y sin deficiencia hormonal (Iwayama et al., 2021; Yuen et al., 2023). Asimismo, Lee et al. (2023) identificaron diferencias analíticas entre plataformas de medición, mientras que Haj-Ahmad et al. (2023) examinaron la utilidad diagnóstica del cociente molar IGF-1/IGFBP-3 como indicador complementario de la deficiencia de GH.

Las pruebas de estimulación también presentan limitaciones relacionadas con la variabilidad analítica y la ausencia de puntos de corte universalmente aceptados. Mientras que Yuen et al. (2023) describieron diferencias en los umbrales diagnósticos, Arlien-Søborg et al. (2025) señalaron que un pico estimulado de GH inferior a 5 μg/L es compatible con deficiencia grave en pacientes pediátricos. Adicionalmente, Abawi et al. (2021) encontraron que el incremento de una DE en el índice de masa corporal se relaciona con una reducción del 11,6 % en el pico de GH.

Estas limitaciones pueden generar discrepancias entre los resultados de los biomarcadores séricos y las pruebas de estimulación, dificultando la clasificación diagnóstica de los pacientes con talla baja. Investigaciones recientes han examinado la utilidad del IGF-1 como herramienta de cribado y su combinación con pruebas dinámicas para mejorar la identificación de la deficiencia hormonal (Fatani, 2023; Khadilkar et al., 2025; Qiu et al., 2025). Sin embargo, resulta necesario integrar estos hallazgos y examinar las diferencias metodológicas que condicionan su interpretación clínica.

En este contexto, el objetivo del estudio fue evaluar, mediante una revisión sistemática de la literatura publicada entre 2021 y 2026, la exactitud diagnóstica del IGF-1, la IGFBP-3 y su cociente, utilizados individualmente o combinados con pruebas dinámicas, tomando como referencia las pruebas de estimulación de GH en niños con talla baja. Se consideraron indicadores como sensibilidad, especificidad, valores predictivos, razones de verosimilitud y área bajo la curva (AUC), así como los factores analíticos y clínicos que influyen en su rendimiento diagnóstico.

Para orientar la investigación, se formularon las siguientes preguntas: (P1) ¿Qué sensibilidad, especificidad, valores predictivos y AUC presentan el IGF-1, la IGFBP-3 y su cociente frente a las pruebas de estimulación de GH? (P2) ¿Cómo varía el rendimiento diagnóstico al combinar concentraciones bajas de IGF-1 con una sola prueba de estimulación? (P3) ¿Qué influencia tienen los puntos de corte de GH, los métodos de inmunoensayo y el índice de masa corporal sobre el rendimiento diagnóstico de estas pruebas?

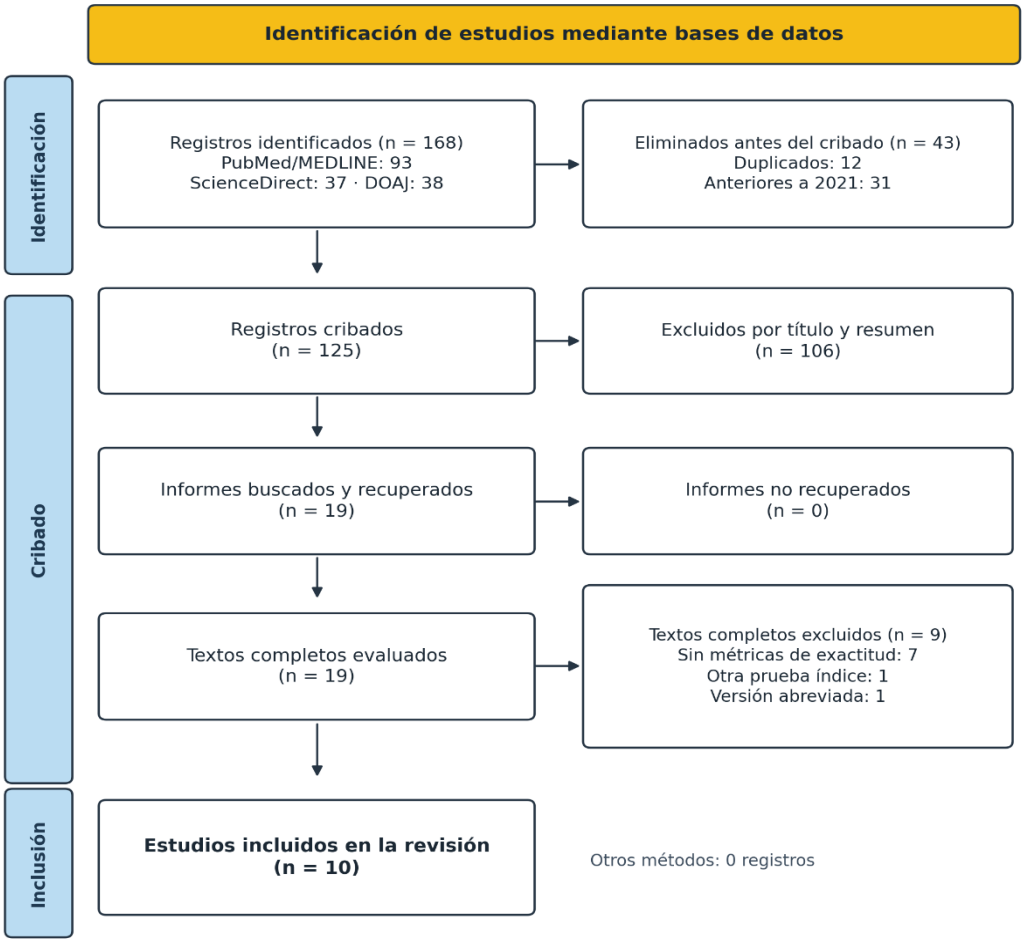
2. Metodología

Diseño de investigación

Se realizó una revisión sistemática de la literatura con síntesis narrativa, siguiendo las directrices PRISMA 2020 propuestas por Page et al. (2021). El proceso de identificación, cribado, elegibilidad e inclusión de los estudios se representó mediante un diagrama de flujo (figura 1). No se efectuó un metaanálisis debido a la heterogeneidad de las pruebas de referencia, los puntos de corte de hormona del crecimiento (GH) y las plataformas analíticas utilizadas en las investigaciones seleccionadas.

Figura 1

Diagrama de flujo PRISMA 2020 del proceso de identificación, selección e inclusión de estudios



Nota. Diagrama adaptado de PRISMA 2020 (Page et al., 2021). El filtro temporal se aplicó durante la depuración de los registros, excepto en la estrategia combinada de PubMed/MEDLINE, en la que se aplicó directamente en la base de datos.

Criterios de elegibilidad

La pregunta de investigación se estructuró mediante el formato PIRD, que comprendió la población, la prueba índice, la prueba de referencia y los desenlaces diagnósticos. Los criterios de inclusión y exclusión se establecieron considerando las características de los participantes, las pruebas evaluadas, los indicadores de exactitud diagnóstica y el diseño de los estudios. Asimismo, se admitieron investigaciones que compararon pruebas de estimulación entre sí o con la respuesta clínica al

tratamiento, por su contribución a la evaluación de las pruebas de referencia (tabla 1).

Tabla 1

Componentes PIRD y criterios de elegibilidad

Componente	Criterios de inclusión	Criterios de exclusión
Población (P)	Niños, adolescentes o pacientes en transición con talla baja evaluados por sospecha de deficiencia de GH.	Estudios realizados exclusivamente en adultos o animales.
Prueba índice (I)	IGF-1, IGFBP-3 o cociente IGF-1/IGFBP-3, evaluados individualmente o combinados con pruebas dinámicas.	Estudios que utilizaron exclusivamente radioinmunoensayos u otros marcadores distintos de los establecidos.
Prueba de referencia (R)	Pruebas de estimulación de GH mediante clonidina, glucagón, arginina, L-dopa, ITT o protocolos combinados, comparadas entre sí o con la respuesta clínica al tratamiento.	Estudios centrados exclusivamente en la reducción de los tiempos de muestreo de una misma prueba.
Desenlace (D)	Estudios que reportaron sensibilidad, especificidad, valores predictivos o área bajo la curva ROC.	Investigaciones que no proporcionaron métricas de exactitud diagnóstica.
Diseño y publicación	Estudios originales y guías publicados en inglés o español entre enero de 2021 y abril de 2026.	Reportes y series de casos, editoriales, cartas al editor y revisiones narrativas.

Nota. PIRD = población, prueba índice, prueba de referencia y desenlace; GH = hormona del crecimiento; IGF-1 = factor de crecimiento similar a la insulina tipo 1; IGFBP-3 = proteína de unión al factor de crecimiento similar a la insulina tipo 1 de tipo 3; ITT = prueba de tolerancia a la insulina; ROC = característica operativa del receptor.

Fuentes de información y estrategia de búsqueda

Se consultaron las bases de datos PubMed/MEDLINE, ScienceDirect y Directory of Open Access Journals (DOAJ), mediante descriptores MeSH y términos libres organizados en tres bloques: condición clínica (B1), biomarcadores (B2) y pruebas dinámicas (B3). Se realizaron búsquedas independientes y combinadas, según las características de cada plataforma, y se registraron las cadenas utilizadas y el número de publicaciones recuperadas. Las estrategias de búsqueda y los resultados correspondientes se presentan en la tabla 2.

Las búsquedas se exportaron en octubre de 2026, mientras que el periodo de elegibilidad se delimitó previamente entre el 1 de enero de 2021 y el 30 de abril de 2026. El filtro temporal se aplicó directamente en la estrategia combinada de PubMed/MEDLINE y durante la depuración de las demás exportaciones. Como resultado, se eliminaron 31 registros anteriores a 2021 antes del cribado, mientras que los posteriores a abril de 2026 se descartaron durante la revisión de títulos y resúmenes.

Tabla 2

Cadenas de búsqueda y registros obtenidos por base de datos

Base de datos	Cadena de búsqueda	Aplicación del filtro temporal	Registros (n)
PubMed/MEDLINE	B1: ("Growth Hormone Deficiency" OR "Short Stature")	Durante la depuración	26
PubMed/MEDLINE	B2: ("IGF-1" OR "IGFBP-3")	Durante la depuración	17
PubMed/MEDLINE	B3: ("Dynamic Testing" OR "Stimulation test" OR "Clonidine" OR "Glucagon")	Durante la depuración	8
PubMed/MEDLINE	(IGF-1 OR IGFBP-3) AND (growth hormone stimulation test) AND children AND diagnos*	En la base de datos, 2021–2026	42
ScienceDirect	B1: ("Growth Hormone Deficiency" OR "Short Stature")	Durante la depuración	24
ScienceDirect	B2: ("IGF-1" OR "IGFBP-3")	Durante la depuración	1
ScienceDirect	B3: ("Dynamic Testing" OR "Stimulation test" OR "Clonidine" OR "Glucagon")	Durante la depuración	12
DOAJ	B1 AND B2 AND B3	Durante la depuración	38

Total	168

Nota. Se recuperaron 93 registros de PubMed/MEDLINE, 37 de ScienceDirect y 38 de DOAJ. Los bloques B1, B2 y B3 se exportaron por separado en las búsquedas correspondientes. El filtro temporal se aplicó durante la depuración de registros, excepto en la estrategia combinada de PubMed/MEDLINE.

Selección de estudios y extracción de datos

Se identificaron inicialmente 168 registros, de los cuales se eliminaron 12 duplicados y 31 publicaciones anteriores al periodo establecido. Posteriormente, se examinaron los títulos y resúmenes de 125 registros, y se excluyeron 106 por incumplimiento de los criterios de elegibilidad. Se evaluaron 19 textos completos, de los cuales nueve fueron excluidos por los motivos detallados en la figura 1. Finalmente, se incluyeron 10 estudios en la revisión sistemática.

De los estudios incluidos se extrajeron el país de realización, el diseño metodológico, el tamaño de la muestra, la prueba de referencia, el punto de corte de GH y las métricas de rendimiento diagnóstico. También se registraron los umbrales utilizados para los biomarcadores y las características analíticas relevantes. No se imputaron datos ausentes y la información no reportada se identificó mediante la abreviatura NR. Las decisiones adoptadas durante la selección se conservaron como material suplementario.

Evaluación del riesgo de sesgo

El riesgo de sesgo se evaluó mediante la herramienta QUADAS-2, desarrollada por Whiting et al. (2011), que comprendió cuatro dominios: selección de pacientes, prueba índice, prueba de referencia, y flujo y tiempos. Asimismo, se examinó la aplicabilidad de los tres primeros dominios. Cada apartado se clasificó según un nivel de riesgo bajo, alto o incierto, considerando las características metodológicas y la información proporcionada por los estudios incluidos.

Se asignó un riesgo alto cuando el punto de corte se obtuvo de la misma muestra utilizada para evaluar su rendimiento o cuando la prueba índice formó parte de la referencia diagnóstica, debido al posible sesgo de incorporación. El riesgo se consideró incierto cuando no se informó el umbral de GH empleado. Dos autores evaluaron de manera independiente los estudios seleccionados y resolvieron las discrepancias mediante consenso, conforme a los criterios establecidos para la valoración metodológica.

Análisis y síntesis de los datos

Se extrajeron los indicadores de sensibilidad (Se), especificidad (Sp), valor predictivo positivo (VPP), valor predictivo negativo (VPN) y área bajo la curva ROC (AUC). Asimismo, se calcularon las razones de verosimilitud positiva (RV+) y negativa (RV−), mediante las expresiones $RV+ = Se/(1 - Sp)$ y $RV- = (1 - Se)/Sp$. Se consideraron indicadores de elevada capacidad discriminativa una RV+ superior a 10 y una RV− inferior a 0,1, siguiendo los criterios de interpretación establecidos.

Los indicadores no informados se reconstruyeron a partir de tablas de contingencia de 2 × 2 cuando se dispuso de los recuentos correspondientes a la cohorte completa. Cuando un estudio presentó varias pruebas o puntos de corte, sus resultados se analizaron por separado, sin calcular promedios entre estimaciones de sensibilidad y especificidad. Debido a la heterogeneidad metodológica y analítica de las investigaciones, se realizó una síntesis narrativa. Los valores de p se conservaron según lo reportado por los autores originales, considerando un nivel de significación de $\alpha = 0,05$.

3. Resultados

Características de los estudios incluidos

Se identificaron 168 registros procedentes de PubMed/MEDLINE (n = 93), ScienceDirect (n = 37) y DOAJ (n = 38). Tras eliminar 12 duplicados y 31 registros anteriores a 2021, se examinaron 125 publicaciones, de las cuales 106 fueron excluidas durante el cribado. Posteriormente, se evaluaron 19 textos completos y se descartaron nueve: siete por ausencia de métricas de exactitud diagnóstica, uno por utilizar una prueba índice diferente y otro por comparar versiones abreviadas de una misma prueba. Finalmente, se incluyeron diez estudios (figura 1).

Los estudios comprendieron un total de 2993 participantes, aunque no todos correspondieron a poblaciones evaluadas específicamente por sospecha de deficiencia de hormona del crecimiento (GHD). Predominaron los diseños observacionales retrospectivos y las investigaciones realizadas en Asia y Oriente Medio, mientras que un estudio procedió de Europa y ninguno de América Latina. Las pruebas de referencia incluyeron diferentes agentes estimulantes, con umbrales de GH que variaron según el protocolo empleado. Estas diferencias limitaron la comparabilidad directa de los resultados.

La tabla 3 presenta las características metodológicas de los diez estudios incluidos, considerando el país de realización, el diseño, el tamaño de la muestra y las pruebas de referencia utilizadas. También se especifican los participantes diagnosticados con GHD y los umbrales empleados para interpretar la respuesta de GH. Esta información permitió identificar la heterogeneidad de los procedimientos diagnósticos y contextualizar las estimaciones de exactitud reportadas por los autores.

Tabla 3

Características de los estudios incluidos (n = 10)

Estudio	País	Diseño	n (GHD)	Prueba de referencia y umbral de GH
Iwayama et al. (2021)	Japón	Observacional	298 (60)	Clonidina, arginina y L-dopa; NR
Penco et al. (2021)	Italia	Retrospectivo	206 (26)*	ATT o ITT, pico de GH inferior a 8 ng/mL; ATT + GHRH, inferior a 20 ng/mL
Oron et al. (2022)	Israel	Retrospectivo	515 (NR)	Arginina + clonidina frente a clonidina y prueba confirmatoria; NR
Fatani (2023)	Arabia Saudita	Retrospectivo	724 (187)	Clonidina y glucagón; pico inferior a 7 ng/mL
Haj-Ahmad et al. (2023)	Jordania	Transversal	235 (64)	Glucagón e ITT; clonidina ante resultados discordantes; pico inferior a 6,25 ng/mL
Chang y Kim (2024)	Corea del Sur	Retrospectivo	273 (135)	L-dopa, ITT y arginina; NR
Khadilkar et al. (2025)	India	Retrospectivo	75 (38)	Prueba de estimulación; NR
Qiu et al. (2025)	China	Transversal	352 (117)	Clonidina y arginina; pico inferior a 10 μg/L
Ozdemir Dilek y Özgüç Comlek (2026)	Turquía	Retrospectivo	101 (44)	Clonidina y L-dopa o ITT; pico inferior a 7 ng/mL y VC inferior a −2 DE
Srilanchakon (2026)	Tailandia	Cohorte retrospectiva	214 (63)	Clonidina y una segunda prueba; pico inferior a 7 ng/mL

Nota. GHD = deficiencia de hormona del crecimiento; ATT = prueba de tolerancia a la arginina; ITT = prueba de tolerancia a la insulina; GHRH = hormona liberadora de GH; VC = velocidad de crecimiento; DE = desviaciones estándar; NR = no reportado. *En Penco et al. (2021) se analizaron 278 pruebas realizadas a 206 pacientes, de los cuales 86 fueron evaluados por sospecha de GHD. La cifra de 26 corresponde a este subgrupo.*

La evaluación mediante QUADAS-2 mostró que ninguno de los estudios presentó riesgo de sesgo bajo en los cuatro dominios examinados. Se identificó riesgo alto en la prueba índice de tres investigaciones que determinaron sus puntos de corte utilizando la misma muestra de evaluación (Iwayama et al., 2021; Khadilkar et al., 2025; Qiu et al., 2025). Asimismo, se señaló posible sesgo de incorporación en tres estudios que utilizaron la clonidina tanto en la prueba evaluada como en la definición de referencia

(Fatani, 2023; Ozdemir Dilek & Özgüç Comlek, 2026; Srilanchakon, 2026).

Rendimiento diagnóstico de los biomarcadores basales frente a las pruebas de estimulación

El rendimiento diagnóstico del IGF-1 aislado presentó diferencias entre las investigaciones analizadas, con sensibilidades comprendidas entre el 42,2 % y el 79,0 %, y especificidades entre el 41,7 % y el 90,1 %. Iwayama et al. (2021) observaron una capacidad discriminativa limitada, con un área bajo la curva (AUC) de 0,517 y ausencia de diferencias significativas entre pacientes con y sin GHD (p = 0,23). Por su parte, Penco et al. (2021) reportaron una sensibilidad del 79,0 % y una especificidad del 52,0 % para un punto de corte de −1,5 DE.

Khadilkar et al. (2025) identificaron un punto de corte de −2,41 DE para el IGF-1, con sensibilidad del 74,3 %, especificidad del 89,2 % y razón de verosimilitud positiva (RV+) de 6,88. En contraste, Haj-Ahmad et al. (2023) comunicaron una sensibilidad del 42,2 % y una especificidad del 90,1 % para el IGF-1 aislado. Estos resultados evidenciaron diferencias en el rendimiento de los puntos de corte evaluados, cuya interpretación estuvo condicionada por las características de las poblaciones y los procedimientos diagnósticos utilizados.

Respecto a los biomarcadores complementarios, Haj-Ahmad et al. (2023) reportaron una sensibilidad del 53,1 % y una especificidad del 81,3 % para la IGFBP-3 aislada. El cociente IGF-1/IGFBP-3 alcanzó una sensibilidad del 87,5 %, una especificidad del 83,0 % y un valor predictivo negativo (VPN) del 94,7 %. Por el contrario, la exigencia simultánea de valores bajos en los tres indicadores incrementó la especificidad hasta el 97,7 %, pero redujo la sensibilidad al 29,7 %.

Qiu et al. (2025) encontraron concentraciones inferiores de IGF-1 e IGFBP-3 en los niños diagnosticados con GHD. El modelo multivariable que incorporó ambos biomarcadores, sus cocientes y el índice de masa corporal alcanzó un AUC de 0,936, con sensibilidad del 95,7 % y especificidad del 79,6 %. Sin embargo, dado que el modelo se desarrolló y evaluó utilizando la misma muestra, estos resultados correspondieron al rendimiento estimado en la población estudiada y no a una validación externa.

La tabla 4 presenta las pruebas índices y de referencia utilizadas en los estudios, junto con sus respectivos puntos de corte y estimaciones de rendimiento diagnóstico. Se incluyen sensibilidad, especificidad, valores predictivos, AUC y razones de verosimilitud, cuando estuvieron disponibles. Las investigaciones que evaluaron diferentes pruebas o umbrales se presentan mediante registros independientes, con el propósito de conservar las estimaciones originales y facilitar su comparación descriptiva.

Rendimiento de las combinaciones diagnósticas y factores asociados

La combinación de IGF-1 con una prueba dinámica mostró variaciones importantes según el punto de corte empleado. Fatani (2023) reportó que un IGF-1 ≤ 0 DE combinado con clonidina identificó a 146 de los 187 pacientes con GHD, con una sensibilidad reconstruida del 78,1 % y una especificidad del 94,0 %. Al utilizar un umbral de IGF-1 ≤ −2 DE, la sensibilidad disminuyó al 18,7 %, mientras que la especificidad aumentó al 98,5 %. Estos resultados mostraron que los puntos de corte más restrictivos redujeron la detección de casos.

Srilanchakon (2026) encontró resultados similares al combinar concentraciones bajas de IGF-1 con una prueba de clonidina. Con un punto de corte de −1,5 DE, la sensibilidad alcanzó el 60,5 % y la especificidad el 95,9 %, mientras que el umbral de −2,5 DE redujo la sensibilidad al 31,6 % y elevó la

especificidad al 98,6 %. Aunque este último criterio presentó una RV+ de 22,57, su RV− de 0,69 indicó una capacidad limitada para descartar la deficiencia hormonal.

El índice de masa corporal también se relacionó con diferencias en la respuesta a las pruebas dinámicas. Ozdemir Dilek y Özgüç Comlek (2026) encontraron valores inferiores del pico de GH tras la administración de clonidina en pacientes con sobrepeso u obesidad, respecto a aquellos con peso normal (4,5 frente a 8,2 ng/mL; p = 0,043). La especificidad disminuyó del 84,8 % al 50,0 % (p = 0,008), aunque aumentó al 75,0 % al incorporar concentraciones de IGF-1 y velocidad de crecimiento.

Las pruebas de estimulación también presentaron diferencias en su rendimiento según el protocolo utilizado. Oron et al. (2022) encontraron proporciones similares de GHD mediante una prueba combinada de arginina y clonidina y dos pruebas separadas (12,7 % frente a 14,4 %; p = 0,609). Por otra parte, Chang y Kim (2024) comunicaron especificidades del 50,3 %, 72,3 % y 64,7 % para L-dopa, ITT y arginina, respectivamente, al considerar la respuesta clínica al tratamiento con GH.

Penco et al. (2021) informaron resultados patológicos en la primera prueba de estimulación de 34 de los 86 niños evaluados por sospecha de GHD. La segunda prueba confirmó el diagnóstico en 26 pacientes, lo que correspondió a un valor predictivo positivo del 76,5 %. Estos resultados mostraron diferencias entre las pruebas dinámicas y la importancia de considerar los procedimientos confirmatorios al interpretar la respuesta de GH

Tabla 4

Rendimiento diagnóstico de los biomarcadores y las pruebas dinámicas en los estudios incluidos

Estudio	Prueba índice y punto de corte	Se (%)	Sp (%)	VPP (%)	VPN (%)	AUC	RV+	RV−
Iwayama et al. (2021)	IGF-1 ≤ −1,493 DE	68,5	41,7	25,0	82,3	0,517	1,17	0,76
Penco et al. (2021)	IGF-1 ≤ −1,5 DE	79,0	52,0	NR	NR	NR	1,65	0,40
Oron et al. (2022)	Arginina + clonidina frente a dos pruebas	NR	NR	NR	NR	NR	NR	NR
Fatani (2023)	IGF-1 ≤ 0 DE + clonidina	78,1	94,0	82,0	92,5	0,609	13,10	0,23
Fatani (2023)	IGF-1 ≤ −2 DE + clonidina	18,7	98,5	81,4	77,7	0,662	12,56	0,83
Haj-Ahmad et al. (2023)	IGF-1 bajo	42,2	90,1	61,4	80,6	NR	4,26	0,64
Haj-Ahmad et al. (2023)	IGFBP-3 baja	53,1	81,3	51,5	82,2	NR	2,84	0,58
Haj-Ahmad et al. (2023)	Cociente IGF-1/IGFBP-3 bajo	87,5	83,0	65,9	94,7	NR	5,15	0,15
Haj-Ahmad et al. (2023)	Tres indicadores bajos	29,7	97,7	82,6	78,8	NR	12,91	0,72
Chang y Kim (2024)	L-dopa	NR	50,3	NR	NR	NR	NR	NR
Chang y Kim (2024)	ITT	NR	72,3	NR	NR	NR	NR	NR
Chang y Kim (2024)	Arginina	NR	64,7	NR	NR	NR	NR	NR
Khadilkar et al. (2025)	IGF-1 ≤ −2,41 DE	74,3	89,2	NR	NR	NR	6,88	0,29
Qiu et al. (2025)	Cociente IGF-1/edad	NR	NR	NR	NR	0,921	NR	NR
Qiu et al. (2025)	Modelo multivariable, corte 0,232	95,7	79,6	70,0	97,4	0,936	4,69	0,05
Ozdemir Dilek y Özgüç Comlek (2026)	Clonidina inferior a 5 ng/mL, peso normal	90,5	84,8	79,2	93,3	NR	5,95	0,11
Ozdemir Dilek y Özgüç Comlek (2026)	Clonidina inferior a 5 ng/mL, sobrepeso u obesidad	82,6	50,0	61,3	75,0	NR	1,65	0,35
Ozdemir Dilek y Özgüç Comlek (2026)	Clonidina + IGF-1 inferior a −1,5 DE y VC inferior a −2 DE	84,2	75,0	84,2	75,0	NR	3,37	0,21
Srilanchakon (2026)	IGF-1 ≤ −1,5 DE + clonidina	60,5	95,9	NR	NR	NR	14,76	0,41

Srilanchakon (2026)	IGF-1 ≤ −2,5 DE + clonidina	31,6	98,6	92,3	NR	NR	22,57	0,69

Nota. Se = sensibilidad; Sp = especificidad; VPP = valor predictivo positivo; VPN = valor predictivo negativo; AUC = área bajo la curva ROC; RV+ y RV− = razones de verosimilitud positiva y negativa; VC = velocidad de crecimiento; DE = desviaciones estándar; ITT = prueba de tolerancia a la insulina; NR = no reportado. Se conservaron las estimaciones proporcionadas en el documento, incluidas las reconstruidas para Fatani (2023), que requieren comprobación con los recuentos de la cohorte analizada.

4. Discusión

Los resultados de esta revisión sistemática evidenciaron que la exactitud diagnóstica de los biomarcadores basales frente a las pruebas de estimulación de GH fue limitada y heterogénea. El IGF-1 aislado presentó sensibilidades insuficientes para descartar de manera fiable la deficiencia de hormona del crecimiento (GHD), mientras que su especificidad varió según los criterios diagnósticos utilizados. Estos hallazgos coincidieron con Marran et al. (2026), quienes, en un estudio sudafricano de 357 niños publicado fuera del periodo de elegibilidad, reportaron sensibilidades del 18,8 % al 44,1 % y una concordancia baja con las pruebas dinámicas ($\kappa = 0{,}055–0{,}234$).

La heterogeneidad observada estuvo relacionada, en parte, con las diferencias metodológicas entre los estudios. Los umbrales de GH variaron entre 6,25 y 10 ng/mL, mientras que Marran et al. (2026) identificaron cambios en la sensibilidad del IGF-1 según el punto de corte utilizado. Asimismo, Penco et al. (2021) encontraron que aproximadamente una cuarta parte de las primeras pruebas patológicas no fue confirmada, mientras que Chang y Kim (2024) reportaron especificidades del 50,3 % al 72,3 % frente a la respuesta clínica. Estas diferencias evidenciaron las limitaciones de las pruebas dinámicas como referencia diagnóstica uniforme.

La variabilidad analítica constituyó otro factor relevante para interpretar las diferencias de rendimiento diagnóstico. Lee et al. (2023) identificaron discrepancias entre plataformas de medición de IGF-1, mientras que Arlien-Søborg et al. (2025) señalaron la persistencia de controversias sobre los puntos de corte de GH. A ello se sumó el posible sesgo de incorporación cuando la clonidina formó parte tanto de la prueba índice como de la referencia diagnóstica. Estas condiciones limitaron la comparabilidad de las estimaciones de exactitud entre estudios y reforzaron la importancia de considerar las características de cada método analítico.

Entre los biomarcadores basales, el cociente IGF-1/IGFBP-3 presentó un equilibrio favorable entre sensibilidad y especificidad, con un valor predictivo negativo del 94,7 % (Haj-Ahmad et al., 2023). No obstante, este resultado procedió de una única investigación y su interpretación dependió de la prevalencia de GHD en la población evaluada. Por su parte, Qiu et al. (2025) obtuvieron una razón de verosimilitud negativa de 0,05 mediante un modelo multivariable. Aunque este resultado sugirió una elevada capacidad para descartar la enfermedad en la muestra estudiada, la ausencia de validación externa limitó su generalización.

La combinación de biomarcadores con pruebas dinámicas mostró una mayor especificidad, aunque su sensibilidad dependió considerablemente del punto de corte seleccionado. Fatani (2023) y Srilanchakon (2026) reportaron especificidades superiores al 90 % al combinar concentraciones bajas de IGF-1 con una prueba de clonidina. Sin embargo, los umbrales más restrictivos redujeron la detección de pacientes con GHD. En Fatani (2023), la especificidad observada estuvo relacionada principalmente con la prueba de clonidina. Por tanto, la evidencia disponible no permitió establecer que la incorporación del IGF-1 mejorara sistemáticamente el rendimiento de una prueba dinámica aislada.

El índice de masa corporal también influyó en la interpretación de las pruebas de estimulación.

Abawi et al. (2021) identificaron una disminución del pico de GH asociada con el incremento del IMC, mientras que Ozdemir Dilek y Özgüç Comlek (2026) observaron una reducción de la especificidad de la clonidina en niños con sobrepeso u obesidad. La incorporación del IGF-1 y la velocidad de crecimiento mejoró la clasificación diagnóstica en esta población. Estos hallazgos sugirieron que la adiposidad puede modificar la respuesta hormonal y generar resultados que requieren una valoración clínica complementaria antes de establecer el diagnóstico.

Desde la perspectiva del laboratorio clínico, los resultados respaldaron la necesidad de interpretar los biomarcadores y las pruebas dinámicas de manera conjunta, considerando sus limitaciones analíticas y clínicas. En este contexto, se consideró pertinente expresar las concentraciones de IGF-1 e IGFBP-3 mediante puntuaciones de desviación estándar ajustadas a las características del método, interpretar el pico de GH según el ensayo y el IMC del paciente, y valorar las discrepancias entre pruebas como hallazgos que requieren evaluación adicional. Estas recomendaciones derivaron de la síntesis narrativa y no constituyeron criterios diagnósticos validados.

Limitaciones

Esta revisión presentó limitaciones relacionadas con el número reducido de estudios incluidos, sus diseños predominantemente observacionales y retrospectivos, y la heterogeneidad de los criterios diagnósticos utilizados. Estas diferencias impidieron realizar un metaanálisis y obtener estimaciones agrupadas de exactitud diagnóstica. Además, la búsqueda se limitó a tres fuentes de información y publicaciones en español e inglés, sin incluir Scopus ni Web of Science. La ausencia de investigaciones latinoamericanas y de validaciones externas suficientes restringió la aplicabilidad de los puntos de corte a otras poblaciones, incluida la ecuatoriana.

5. Conclusiones

Los resultados de esta revisión sistemática evidenciaron que el IGF-1 y la IGFBP-3 presentaron una exactitud diagnóstica limitada y heterogénea frente a las pruebas de estimulación de GH en niños con talla baja. El IGF-1 aislado no mostró un rendimiento suficiente para confirmar o descartar de manera fiable la deficiencia de hormona del crecimiento (GHD). Aunque el cociente IGF-1/IGFBP-3 y los modelos multivariables alcanzaron mejores indicadores diagnósticos, la evidencia procedió de estudios individuales sin validación externa suficiente, lo que limitó la generalización de sus resultados.

La combinación de concentraciones bajas de IGF-1 con una sola prueba de estimulación alcanzó especificidades elevadas en determinadas cohortes. Sin embargo, su sensibilidad fue variable y estuvo condicionada por los puntos de corte de los biomarcadores, los umbrales de GH y los criterios diagnósticos empleados. En consecuencia, la evidencia disponible no permitió establecer esta combinación como un procedimiento confirmatorio de aplicación general, aunque sugirió su posible utilidad complementaria en contextos clínicos específicos.

La heterogeneidad de los umbrales de GH, las diferencias entre métodos analíticos y la influencia del índice de masa corporal constituyeron factores relevantes para interpretar las variaciones del rendimiento diagnóstico. Estos hallazgos destacaron la necesidad de considerar intervalos de referencia específicos para cada método y las características clínicas del paciente. Asimismo, se identificó la necesidad de desarrollar investigaciones prospectivas, con validación externa y participación de poblaciones latinoamericanas, que permitan establecer criterios diagnósticos más consistentes y evaluar su aplicabilidad en el contexto ecuatoriano.

Referencias

Abawi, O., Augustijn, D., Hoeks, S., de Rijke, Y., & van den Akker, E. (2021). Impact of body mass index on growth hormone stimulation tests in children and adolescents: A systematic review and meta-analysis. *Critical Reviews in Clinical Laboratory Sciences, 58*(8), 576–595. https://doi.org/10.1080/10408363.2021.1956423

Arlien-Søborg, M., Radovick, S., Boguszewski, M., Bidlingmaier, M., Johannsson, G., Grimberg, A., Ho, K., Biller, B., Choong, C., Hoffman, A., Backeljauw, P., Boguszewski, C., Bollerslev, J., Brue, T., Chanson, P., Christ, E., Cianfarani, S., Clayton, P., Cohen, P., . . . Jorgensen, J. (2025). Consensus and controversies about diagnosing GH deficiency: A Delphi survey by the GH research society. *Pituitary, 28*(3), Article 57. https://doi.org/10.1007/s11102-025-01526-z

Chang, S., & Kim, C. (2024). Comparison of growth hormone stimulation tests in prepubertal children with short stature according to response to growth hormone replacement. *Annals of Pediatric Endocrinology & Metabolism, 29*(5), 316–324. https://doi.org/10.6065/apem.2346234.117

Fatani, T. (2023). Diagnostic Value of IGF-1 in Growth Hormone–Deficient Children: Is a Second Growth Hormone Stimulation Test Necessary? *Journal of the Endocrine Society, 7*(4), Article bvad018. https://doi.org/10.1210/jendso/bvad018

Hage, C., Gan, H., Ibba, A., Patti, G., Dattani, M., Loche, S., Maghnie, M., & Salvatori, R. (2021). Advances in differential diagnosis and management of growth hormone deficiency in children. *Nature Reviews Endocrinology, 17*(10), 608–624. https://doi.org/10.1038/s41574-021-00539-5

Haj-Ahmad, L., Mahmoud, M., Sweis, N., Bsisu, I., Alghrabli, A., Ibrahim, A., & Zayed, A. (2023). Serum IGF-1 to IGFBP-3 molar ratio: A promising diagnostic tool for growth hormone deficiency in children. *The Journal of Clinical Endocrinology & Metabolism, 108*(4), 986–994. https://doi.org/10.1210/clinem/dgac609

Iwayama, H., Kitagawa, S., Sada, J., Miyamoto, R., Hayakawa, T., Kuroyanagi, Y., Muto, T., Kurahashi, H., Ohashi, W., Takagi, J., & Okumura, A. (2021). Insulin-like growth factor-1 level is a poor diagnostic indicator of growth hormone deficiency. *Scientific Reports, 11*(1), Article 16159. https://doi.org/10.1038/s41598-021-95632-0

Khadilkar, V., Karguppikar, M., Bhalerao, A., Bhade, K., & Khadilkar, A. (2025). Insulin like growth factor-1 as a screening tool to differentiate between idiopathic short stature and growth hormone deficiency. *The Indian Journal of Pediatrics, 92*(10), 1098–1100. https://doi.org/10.1007/s12098-025-05687-4

Lee, J., Cradic, K., Singh, R., Jones, J., & Li, J. (2023). Discordance of insulin-like growth factor-1 results and interpretation on four different platforms. *Clinica Chimica Acta, 539*, 130–133. https://doi.org/10.1016/j.cca.2022.11.034

Marran, K., Segal, D., Kamerman, P., & Saksenberg, D. (2026). Value of IGF-1 for childhood growth hormone deficiency. *Endocrine and Metabolic Science, 22*, Article 100322. https://doi.org/10.1016/j.endmts.2026.100322

Oron, T., Krieger, A., Yakobovich-Gavan, M., Tenenbaum, A., Diamant, R., Phillip, M., & Lazar, L. (2022). Diagnosing growth hormone deficiency: Can a combined arginine and clonidine stimulation test replace 2 separate tests? *Endocrine Practice, 28*(1), 36–43. https://doi.org/10.1016/j.eprac.2021.08.004

Ozdemir Dilek, S., & Özgüç Comlek, F. (2026). Obesity-related changes in growth hormone stimulation test performance under pediatric growth hormone deficiency. *Children, 13*(2), Article 299. https://doi.org/10.3390/children13020299

Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., Mulrow, C., Shamseer, L., Tetzlaff, J., Akl, E., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., Lalu, M., Li, T., Loder, E., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Penco, A., Bossini, B., Giangreco, M., Vidonis, V., Vittori, G., Grassi, N., Pellegrin, M., Faleschini, E., Barbi, E., & Tornese, G. (2021). Should pediatric endocrinologists consider more carefully when to perform a stimulation test? *Frontiers in Endocrinology, 12*, Article 660692. https://doi.org/10.3389/fendo.2021.660692

Qiu, S., Zhao, T., Lian, Y., & Liu, C. (2025). Constructing a precise prediction model for screening growth hormone deficiency in children in a single laboratory setting: Using the serum IGF-1-to-age ratio as a promising predictive marker. *Translational Pediatrics, 14*(12), 3231–3243. https://doi.org/10.21037/tp-2025-481

Srilanchakon, K. (2026). Low IGF-1 combined with a single clonidine stimulation test as a confirmatory tool for pediatric growth hormone deficiency. *Growth Hormone & IGF Research, 84*, Article 101698. https://doi.org/10.1016/j.ghir.2026.101698

Whiting, P., Rutjes, A., Westwood, M., Mallett, S., Deeks, J., Reitsma, J., Leeflang, M., Sterne, J., & Bossuyt, P. (2011). QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine, 155*(8), 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

Yau, M., & Rapaport, R. (2022). Growth hormone stimulation testing: To test or not to test? That is one of the questions. *Frontiers in Endocrinology, 13*, Article 902364. https://doi.org/10.3389/fendo.2022.902364

Yuen, K., Johannsson, G., Ho, K., Miller, B., Bergada, I., & Rogol, A. (2023). Diagnosis and testing for growth hormone deficiency across the ages: A global view of the accuracy, caveats, and cut-offs for diagnosis. *Endocrine Connections, 12*(7), Article e220504. https://doi.org/10.1530/EC-22-0504

Transparencia

Conflicto de interés

Los autores declaran que no existen conflictos de interés de naturaleza alguna como parte de la presente investigación.

Fuente de financiamiento

Los autores financiaron completamente la investigación.

Contribución de autoría

Jhoel Stalin Ortiz Avila: Conceptualización, metodología, software, validación, análisis formal, investigación, visualización, redacción - preparación del borrador original, redacción - revisión y edición, financiamiento, recursos.

Ivan Marcelo Cantuña Vallejo: Conceptualización, metodología, validación, análisis formal, investigación, gestión de datos, redacción - preparación del borrador original, redacción - revisión y edición, financiamiento, administración del proyecto, supervisión.

Los autores contribuyeron activamente en el análisis de los resultados, revisión y aprobación del manuscrito final.