

When Is a Positive Test for Pediatric Growth Hormone Deficiency a True-Positive Test?

George M. Bright^a Peter A. Morris^b Ron G. Rosenfeld^c

^aPediatric Clinical Trial Consultant, Seabrook, TX, USA; ^bDepartment of Management Science and Engineering, VMN Group LLC, Stanford University, Stanford, CA, USA; ^cOregon Health & Science University, Professor of Pediatrics, Stanford University, Stanford, CA, USA

Keywords

Childhood · Diagnosis · Growth hormone · Growth hormone deficiency · Diagnosis

Abstract

Background: In most cases, the growth hormone stimulation test is a necessary component for the diagnosis of growth hormone deficiency (GHD) in children. Diagnostic testing can lead to unnecessary treatment of children with false-positive test results and omission of treatment in children with false-negative results. False-positive results are suggested by the absence of typical growth responses in treated children and false-negative results are suggested by continued growth failure in those left untreated. **Summary:** The probability that a positive test result indicates the presence of the condition (true positive) depends on the prevalence of that condition in the test population and the false positive rate of the test. This probability has been estimated using published data on the prevalence of GHD in children and the false positive rates estimated from performance of stimulation tests in normally growing children and from repeated testing in short children. Because of the low prevalence of GHD and the substantial false positive rate of the

test, the probability of a true-positive result in a child with short stature is 0.028, or about 1 in 36 cases. **Key Messages:** In children with short stature, most positive growth hormone stimulation test results will be false-positive results, resulting in growth hormone treatment of children misdiagnosed as growth hormone deficient. Additional information is required for accurate diagnosis and prediction of successful treatment outcomes in children. Improvements in diagnostic accuracy and treatment outcome predictions can be anticipated from the use of additional predictive enrichment markers identified and evaluated in broadly based studies of growth hormone treatment in children.

© 2021 S. Karger AG, Basel

General Considerations for Evaluating a Test

It is of value to the clinician and, perhaps, of even greater importance to the patient, to know when a positive test result indicates the presence of a medical condition. Each medical condition has a prevalence and probability of occurrence in the general population. A condition with a prevalence of 50 per 100,000 persons has an occurrence probability of 0.0005. A positive result from a

well-developed diagnostic test can increase the probability that the condition is present. Positive tests, however, are not always true-positive results; that is, a positive test may not always indicate the presence of the condition (false-positive result).

Here, we review the theoretical construct that relates test results to the presence or absence of a condition. We demonstrate, using general examples, that the probability that a positive test result is, in fact, a true-positive result is significantly impacted by the prevalence rate of the condition and the false positive rate of the test. Additional general examples are generated to demonstrate the pros and cons of using multiple tests to diagnose the presence of a condition. We then examine the limitations in our current ability to discriminate between positive tests and true-positive tests in the diagnosis of growth hormone deficiency (GHD) in children. With these limitations in mind, we propose a new approach to the study of the diagnosis of GHD in children.

The relationship between test results and the presence or absence of a condition is traditionally explained in a 2 × 2 construct:

	Condition positive	Condition negative
Test positive	True positive	False positive
Test negative	False negative	True negative

For our first examples, the probability that a positive test is a true positive (P_{ptp}) is the ratio of all true-positive tests to the sum of all positive tests defined as:

$$P_{ptp} = \frac{\# \text{ true positive tests}}{\# \text{ true positive} + \# \text{ false positive tests}}$$

Dependence of P_{ptp} on Prevalence

The first example demonstrates how, in general, P_{ptp} is dependent on the prevalence of the condition. For this example, it is assumed that all persons with the condition will have a positive test result (no false negatives) and that the false positive rate is 0.01 (1 per 100). For prevalences ranging from 50 to 50,000 cases per 100,000 persons, examples of the calculations for P_{ptp} are shown in Table 1 and the dependence of P_{ptp} on prevalence is shown in Figure 1.

The arrows in Figure 1 indicate that for the example of a test with a zero false negative rate and a 0.01 false positive rate (1 per 100), an even money bet that a positive test is a true-positive result ($P_{ptp} \geq 0.5$) can only occur when disease prevalence is greater than 1,000 cases/100,000 persons (1 per hundred). The calculations in Table 1 also

indicate that the position of the probability curve will depend on the false positive rate. This dependence is examined next.

Dependence of P_{ptp} on False Positive Rate

The second example demonstrates that the probability of a positive test being a true-positive test also depends on the false positive rate. This example is for a condition with a prevalence of 50/100,000 persons. It is assumed that all 50 persons with the disease will have a positive result (no false negatives) and in the remaining 99,950 people, the rate of false-positive results ranges from 0.0001 to 0.1. Table 2 shows examples of the calculations for P_{ptp} for various false positive rates and Figure 2 shows the graphical representation for the range of false positive rates from 0.00001 to 1.0.

Table 2 and Figure 2 indicate the probability that a positive test is a true-positive result decreases with increasing false positive rate. The arrows in Figure 2 show that for the example where prevalence is 50 cases/100,000 persons, an even money bet that a positive test is a true-positive result ($P_{ptp} \geq 0.5$) can only occur when the false positive rate is ≤ 0.0005 (5 per 10,000); a false positive rate that is rarely achieved with clinical testing. A general strategy to overcome less than optimal false positive rates is to perform multiple tests and require that both tests be positive. As is demonstrated in the next section, this strategy can reduce the false positive rate, but only at the expense of increasing the false negative rate.

In the clinical setting, a patient with a false-positive result may receive an unnecessary treatment and a patient with a false-negative result will not receive a treatment that may have been beneficial. In the research setting evaluating a new therapy, false negatives and false positives can skew the treatment population and yield inaccurate impressions of the safety and efficacy of the new therapy. As demonstrated in our third example, the rates of false positives and false negatives are interrelated. For this example, we consider the use of multiple tests to improve diagnostic accuracy. Arbitrarily, the test population has 10,000 subjects; 10% (1,000 subjects) have the condition and 90% do not. All subjects receive 4 different diagnostic tests. Again, arbitrarily, the positive rates of each test are assumed to be 0.8 (80%) in those with the condition and 0.2 (20%) in those without the condition. In this example, Table 3 shows the calculations to determine the number of true positives, false positives, false positive rates, and false negatives for 1, 2, 3, and 4 tests. It should be mentioned that these calculations can be anticipated to be influ-

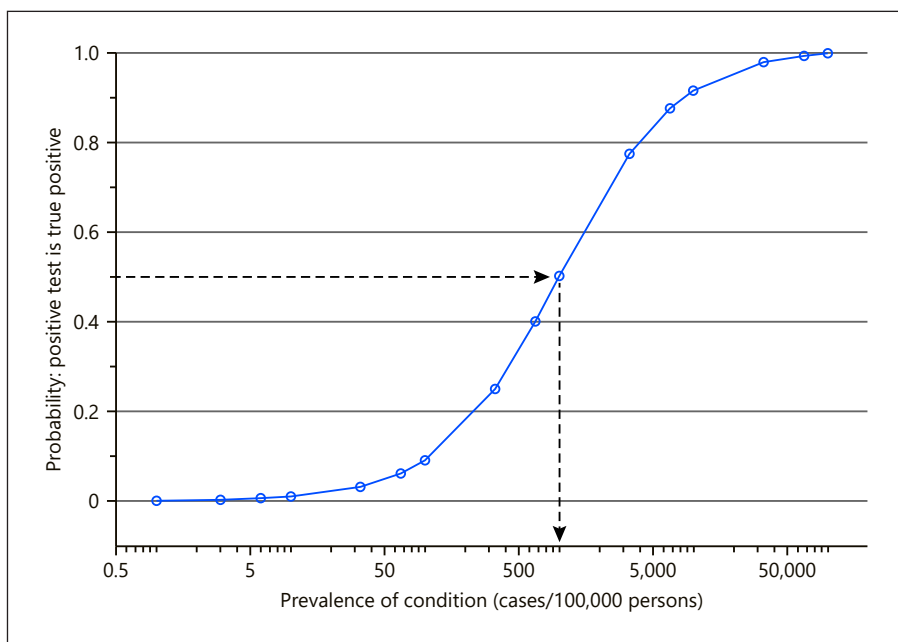


Fig. 1. The general relationship for the dependence of P_{ptp} on prevalence of the condition. The arrows indicate that for a test with a 0 false negative rate and a 0.01 (1 in 100) false positive rate, the equal probability that a positive test result is either a true positive or a false positive ($P_{ptp} = 0.5$) occurs at a prevalence of about 1,000 cases/100,000 persons (1 in 100). P_{ptp} decreases rapidly with lower prevalences.

Table 1. Sample calculations for the dependence of P_{ptp} on the prevalence of the condition

Prevalence (cases/100,000 persons)	True positives	False positives 0.01 × (100,000 – TP)	Total positives	P_{ptp}
50	50	999.5	1,049.5	0.048
500	500	995	1,495	0.33
5,000	5,000	950	5,950	0.84
50,000	50,000	500	50,500	0.99

This example assumes a zero false negative rate and a false positive rate of 0.01 (1 per 100). The number of false positives increases and P_{ptp} decreases rapidly as the prevalence decreases.

enced by any correlations among the results of multiple diagnostic tests.

Table 3 clearly demonstrates that the use of multiple tests with a requirement that all tests be positive can theoretically reduce the false positive rate. In this arbitrary example, the false positive rate decreases from 0.69 to 0.03. This example also illustrates the downside of the multiple test approach, namely, that as the false positive rate declines, the number of false negatives increases.

Diagnostic Accuracy of GHD in Childhood

With the theoretical background provided above, let us now turn to the diagnosis of GHD in childhood. In this section, we consider 2 topics: What limitations currently

Table 2. Sample calculations for the dependence of P_{ptp} on false positive rate

False positive rate (FPR)	True positives	False positives (FPR × 99,950)	Total positives	P_{ptp}
0.0001	50	9.99	59.99	0.83
0.001	50	99.95	149.95	0.33
0.01	50	999.5	1,049.5	0.048
0.1	50	9,995	10,045	0.005

This example shows that for a prevalence of 50/100,000 and a zero false negative rate, P_{ptp} decreases rapidly as the false positive rate increases.

Fig. 2. The general relationship for the dependence of P_{ptp} on the false positive rate. The arrows indicate that for the example where prevalence is 50 cases/100,000 persons, the equal probability ($P_{\text{ptp}} = 0.5$) that a positive test is either a true positive or a false positive occurs when the false positive rate is approximately 0.0005 (5 per 10,000). P_{ptp} decreases rapidly at higher false positive rates.

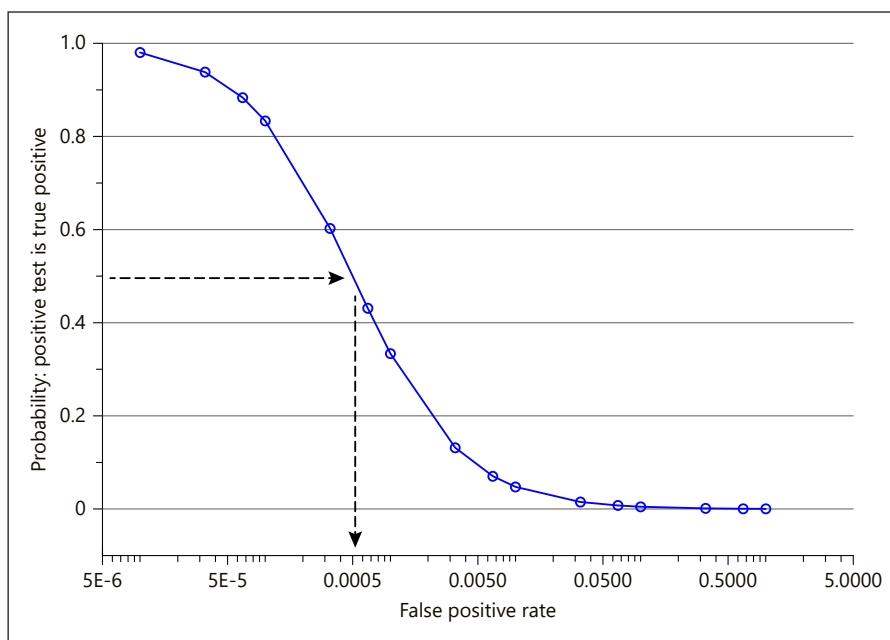


Table 3. Effect of using multiple tests on the relationship between false-positive and false-negative test results

	True positives (TP)	False positives (FP)	Total positives (TP + FP)	False positive rate FP/(FP + TP)	False negatives 1,000 - TP
1 test	800	1,800	2,600	0.69	200
2 tests	640	360	1,000	0.36	360
3 tests	512	72	584	0.12	488
4 tests	410	14	424	0.03	590

This example assumes a population of 10,000 subjects – 1,000 have the condition and 9,000 do not – and that 80% of all subjects with the condition will have positive tests and 20% of all subjects without the condition will have positive tests. If multiple tests are used and all must be positive for a true diagnosis, the false positive rate decreases but the number of false negatives increases as additional tests are added.

exist for the diagnostic accuracy of GHD in childhood? Given any limitations, what further research might improve diagnostic accuracy?

One very substantial limitation is the low prevalence of GHD during childhood, which has been estimated to occur in only 18–27 per 100,000 or approximately 1 in 3,800–5,500 children [1–4]. Even with the very generous assumption of a diagnostic test with an false positive rate of only 1 per hundred, the probability that a positive test is a true-positive test for a prevalence of 25 per 100,000 can be calculated as in the example of Table 2. This probability is 0.024 or about 1 in 40 chances of a positive test being a true-positive result. If the false positive rate was actually 0.2 (20%), the probability falls to 0.0012 or about 1 in 830.

Diagnostic accuracy is degraded by the false positive rate. False positive rates for GH stimulation tests have been estimated in normally growing children and by repeat testing in children with short stature. In 472 normally growing children with a normal IGF-I, the false positive rate to a single test was 0.33 [5]. The false positive rate decreased to 0.10 when 2 GH stimulations tests were used. In another study, the false positive rate was 0.61 in 84 normally growing, prepubertal children [6]. Test-retest concordance is only 40–60 percent when GH stimulation tests are repeated within 1.5 years [7–9]. Using an incidence of 20 cases/100,000 persons, a zero false negative rate, and a false positive rate of 0.3, the probability that a positive growth hormone stimulation test indicates

growth hormone deficiency in that population is 0.0007 (about 7 in 10,000).

Given the restraints imposed by a low prevalence rate and substantial false positive rates, how has the diagnosis of growth hormone deficiency in childhood been improved? One strategy has been to limit the use of growth hormone testing to children with higher probabilities of having growth hormone deficiency; for example, testing only children with significant short stature or children whose heights differ greatly from what is expected given the heights of parents or siblings. Per Figure 1, the probability of a positive growth hormone test indicating growth hormone deficiency is increased because this strategy effectively raises the “prevalence rate.” In the US population, there are approximately 73,000,000 persons with ages ≤ 17 years [10]. With a prevalence of 20 cases per 100,000 persons, 1 would expect 14,600 children with growth hormone deficiency. Of these 73,000,000 children, about 2.38% or 1,737,400 children might be expected to have a height standard deviation score ≤ -2.0 . If all GHD cases were confined to children with height standard deviation scores ≤ -2.0 , the anticipated incidence of GHD in short children would be anticipated to be $14,600/1,737,400 = 0.0084$ or about 840 cases per 100,000 persons with short stature. At this prevalence and a false positive rate of 0.3, the probability that conventional growth hormone stimulation testing indicates growth hormone deficiency in the short stature population, given a positive test result, is 0.028 (about 1 in 36). Although the use of multiple stimulation tests and addition of other tests (e.g., IGF-I, bone age, and height velocity) can be anticipated to improve this low probability, the false positive rates for such tests are currently unknown and require further study.

Improving Diagnostic Accuracy

The low probability of GHD in a child with short stature and positive growth hormone stimulation tests requires that improvements be made in our diagnostics scheme. Diagnostic accuracy can be defined as the maximizing of true results and minimizing of false results. Diagnostic accuracy of GHD in childhood, however, has a limitation imposed by the nature of GHD itself. The clinical sequelae of growth hormone deficiency (e.g., short stature, altered body composition) are believed to result from a chronic deficiency state of low production and secretion of growth hormone and IGF-I during the growth years. This chronic deficiency state is not, as far as we know, a condition that can be verified with “1-time tests” such as tissue biopsy or available imaging techniques. A

possible, but rare, exception is the finding of complete pituitary aplasia with pituitary imaging or the discovery of a genetic defect resulting in complete or near-complete GHD. Unless a new technology emerges that permits monitoring GH and IGF-I production over several years, we will not have a gold standard to document that chronic deficiency state. Yet, the appraisal of diagnostic accuracy requires separation of test results according to the presence/absence of the condition and positive/negative test results. For GHD in childhood, what will serve as the “gold standard” for the diagnosis? How do we define “true positive” and “false positive”?

In the absence of a practical and affordable technology to document a years-long deficiency state, the remaining option is to define GHD by its response to treatment. One approach is to define growth hormone deficiency as that for which the clinical sequelae of GHD are ameliorated by treatment with a physiological replacement dose of growth hormone. This treatment response approach would allow post hoc evaluations of the sensitivity and specificity of the various tests, alone and in combination, and an estimate of prevalence of that condition. Given the state of our current knowledge, however, the treatment response approach to diagnosis is also problematic. Limitations to the adoption of a treatment response definition of GHD include important questions: What constitutes “physiologic replacement”? What constitutes “amelioration of clinical sequelae”?

Currently approved dosing schemes for recombinant growth hormone has resulted in GH exposures that are pharmacological, not physiological. Exposures are pharmacological in terms of both quantity and timing. Growth hormone is secreted into the circulation as several discrete bursts throughout a 24-h period [11], while the daily and weekly injections of rhGH result in single large changes in serum GH that last, variably, for the better part of a day or a week, respectively [12–14]. Some decades ago, researchers used isotope dilution techniques and radiolabeled rhGH to determine daily GH production rates in children [15]. Deconvolution of frequently sampled serum GH can also estimate production rate [11]. The possibility existed that the appropriate daily dose of rhGH would mimic the actual daily production rate. We have since learned 3 limitations to this approach:

- measured GH production rates had nontrivial between-subject variability [11];
- a production rate-based treatment algorithm would require a GH formulation with 100% bioavailability (or a dose of a formulation with incomplete bioavailability scaled to mimic the production rate);

- currently approved formulations of rhGH do not mimic pulsatile release [12, 14].

Perhaps more problematic to the treatment response approach to diagnosis is the definition of a positive treatment response. If defined as an increase in height velocity or an increase in height standard deviation score, increase of what segregates positive from negative responses? The broad range of adult heights in reference populations argues strongly that what might be a treatment success for 1 person might be an inadequate response for another.

Despite these shortcomings of a treatment response approach to the diagnosis of GHD in children, further research with this method might yield substantial improvements in diagnostic accuracy. There seems to be no question that GHD has a low prevalence in children (references), but there is little contemporary information on true and false positive rates for single tests or multiple tests used in combination. Accordingly, what might be done to maximize true positive and true negative rates while minimizing false positive and false negative rates? Much can be learned from a study of GH treatment in children with a wide range of height deficits. A wide range of height deficits ensures that the population represents a broad range of probabilities for the presence of growth hormone deficiency. The pretreatment growth patterns of all subjects would be observed for 6 months, all subjects would have multiple diagnostic tests and then all subjects would receive rhGH treatment for 1 year. Initially, a well-characterized dose of rhGH might be used; a large enough study might accommodate several doses. The distribution of the growth responses could then be defined and permit identification and evaluation of predictive enrichment markers, to be used alone or in combination, to enrich the population for positive responses for various levels of treatment response (e.g., the 10th, 25th, and 50th percentile growth response). As has been recently done for an oral growth hormone secretagogue, cutoff values for each of the predictive enrichment markers would be chosen on their ability to maximize true-positive and true-negative diagnoses while minimizing false-negative

and false-positive diagnoses [16]. These exercises can be anticipated to define the markers, and the cutoff value of the markers, that optimize the prediction of a positive growth response to rhGH and, by inference, improve the diagnostic accuracy of growth hormone deficiency in childhood.

Much has been learned about the variables that influence growth responses to rhGH in children diagnosed with GHD by conventional methods. The growth of children is influenced by the dose of GH [17], age at treatment onset [18], severity of GHD [19], and other pretreatment characteristics [20–22]. Despite knowledge of these factors controlling rhGH treatment outcomes, the range of growth responses is broad [18] and a substantial fraction of GHD children do not reach normal adult heights when treated with rhGH [23]. Accordingly, it is possible that these studies contained subjects in which the diagnosis of GHD was a “false positive.”

Acknowledgements

This manuscript was produced solely by the authors and was not sponsored by other institutions or individuals.

Conflict of Interest Statement

G.M.B. and P.A.M. have no conflicts to disclose. R.G.R. is an active consultant for Biomarin, OPKO, Ipsen, and Lumos Pharma Inc.

Funding Sources

None.

Author Contributions

G.M.B. initiated the concept and drafted the initial version. All authors contributed equally to the final version.

References

- 1 Lindsay R, Feldkamp M, Harris D, Robertson J, Rallison M. Utah growth study: growth standards and the prevalence of growth hormone deficiency. *J Pediatr*. 1994;125:29–35.
- 2 Vimpani GV, Vimpani AF, Lidgard GP, Cameron EH, Farquhar JW. Prevalence of severe growth hormone deficiency. *Br Med J*. 1977;2:427–30.
- 3 Thomas M. Prevalence and demographic features of childhood growth hormone deficiency in Belgium during the period 1986–2001. *Eur J Endocrinol*. 2004;151:67–72.
- 4 Hilken J. Uk audit of childhood growth hormone prescription, 1998. *Arch Dis Child*. 1998;84:387–9.

- 5 Ghigo E, Bellone J, Aimaretti G, Bellone S, Loche S, Cappa M, et al. Reliability of provocative tests to assess growth hormone secretory status. Study in 472 normally growth children. *J Clin Endocrinol*. 1996;81:3323–7.
- 6 Marin G, Domené HM, Barnes KM, Blackwell BJ, Cassorla FG, Cutler GB. The effects of estrogen priming and puberty on the growth hormone response to standardized treadmill exercise and arginine-insulin in normal girls and boys. *J Clin Endocrinol Metab*. 1994;79:537–41.
- 7 Tassoni P, Cacciari E, Cau M, Colli C, Tosi M, Zucchini S, et al. Variability of growth hormone response to pharmacological and sleep tests performed twice in short children. *J Clin Endocrinol Metab*. 1990;71:230–4.
- 8 Ben Dori E, Ziv CA, Auerbach A, Greenberg Y, Zaken H, Levy-Khademi F. The inter-test variability of growth hormone stimulation tests and factors affecting this variability. *Growth Horm IGF Res*. 2020;55:1–5.
- 9 Cacciari E, Tassoni P, Cicognani A, Pirazolli P, Salardi S, Balsamo A, et al. Value and limits to pharmacological and physiological tests to diagnose growth hormone (GH) deficiency and predict therapy response: first and second retesting during replacement therapy of patients defined as GH deficient. *J Clin Endocrinol Metab*. 1994;79(6):1663–9.
- 10 <https://childstats.gov/americaschildren/tables/pop1.asp>.
- 11 Giustina A, Veldhuis JD. Pathophysiology of the neuroregulation of growth hormone secretion in experimental animals and the human. *Endocr Rev*. 1998;19(6):717–97.
- 12 Bidlingmaier M, Kim J, Savoy C, Kim MJ, Ebrecht N, de la Motte S, et al. Comparative pharmacokinetics and pharmacodynamics of a new sustained release growth hormone (GH), LB03003, versus daily GH in adults with GH deficiency. *J Clin Endocrinol Metab*. 2006;91(8):20006–2015.
- 13 Juul RV, Rasmussen MH, Agerso H, Overgaard RV. Pharmacokinetics and pharmacodynamics of once-weekly somapacitan in children and adults: supporting dosing rationales with a model-based analysis of three Phase 1 trials. *Clin Pharmacokinet*. 2019;58:63–75.
- 14 Walvoord EC, de la Pena A, Park S, Silverman B, Cuttler L, Rose SR, et al. Inhaled and subcutaneous GH in children compared with subcutaneous GH in children with GH deficiency: pharmacokinetics, pharmacodynamics and safety. *J Clin Endocrinol Metab*. 2009;94:2052–9.
- 15 Kowarski A, Thompson RG, Migeon CJ, Blizzard RM. Determination of integrated plasma concentrations and true secretion rates of human growth hormone. *J Clin Endocrinol Metab*. 1971;32:356–60.
- 16 Bright GM, Do M-HT, McKew JC, Blum WF, Thorner MO. Development of a predictive enrichment marker for the oral GH secretagogue LUM-201 in pediatric growth hormone deficiency. *J Endocrine Society*. 2021;5(6):1–10.
- 17 CohenBright PGM, Rogol AD, Kappelgaard A-M, Rosenfeld RG. Effects of dose and gender on the growth and growth factor response to GH in GH-deficient children: implications for efficacy and safety. *J Clin Endocrinol Metab*. 2002;87:90–8.
- 18 Bakker B, Frane J, Anhalt H, Lippe B, Rosenfeld RG. Height velocity targets from the national cooperative growth study for first-year growth hormone responses in short children. *J Clin Endocrinol Metab*. 2008;93:352–7.
- 19 Ranke MB, Lindberg A. Observed and predicted growth responses in prepubertal children with growth disorders: guidance of growth hormone treatment by empirical variables. *J Clin Endocrinol Metab*. 2010;95:1229–37.
- 20 Lindberg A, Chatelain P, Wilton P, Cutfield W, Albertsson-Price DA. Derivation and validation of a mathematical model for predicting the responses to exogenous recombinant human growth hormone (GH) in prepubertal children with idiopathic GH deficiency. *J Clin Endocrinol Metab*. 1999;84:1174–83.
- 21 Blethen SL, Compton P, Lippe BM, Rosenfeld RG, August GP, Johanson A. Factors predicting the response to growth hormone (GH) therapy in prepubertal children with GH deficiency. *J Clin Endocrinol Metab*. 1993;76:574–9.
- 22 Kristrom B, Wikland KA. Growth prediction models. *Horm Res*. 2002;57(2):66–70.
- 23 Ranke MB. Short and long-term effects of growth hormone in children and adolescents with GH deficiency. *Front Endocrinol*. 2021;12:720419.