PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Evaluating the UK and European Referral and Investigation Recommendations for Children with Short Stature
AUTHORS	White, Gemma Cosier, Shakira Andrews, Afiya Martin, Lee Willemsen, Ruben Savage, Martin Storr, Helen

VERSION 1 – REVIEW

REVIEWER	Reviewer name: Dr. Peter Flom Institution and Country: Peter Flom Consulting New York, United States Competing interests: None
REVIEW RETURNED	06-Jan-2022

GENERAL COMMENTS	I confine my remarks to statistical aspects of this paper. I am confused. The statistical methods seem to be appropriate, but I'm unclear on exactly what was done and why. Some specific examples are listed below, but I think the whole paper needs a lot more explanation of what each measure was, how often it was measured, and how it was used. Also, I'm kind of unclear on what the whole purpose was. Sorry .... and maybe the confusion is due to the fact that I am a statistician and not a subject matter expert, but, if you are aiming for a somewhat general audience, I think more explanation and detail is needed. Lines 38-41 Why compare the pathological group to the non-pathological group on the measures that define the pathology? Doesn't this wind up being "short people were shorter"? Line 73 Not a statistics question, but why just girls? Line 122 but actually a general comment. I don't see a description of how often, or at what intervals, measurements were used. Here (line 122) it looks like there were 2 measurements, at an interval of 6 to 12 months. But this should be described more completely General: All the variables need to be defined. What is target height, for instance? Line 165 How was this division into groups done? (This is shown in Fig 1, but I think it needs to be in the text).

REVIEWER	Reviewer name: Dr. Alan Rogol Institution and Country: University of Nottingham School of Medicine,

	United Kingdom of Great Britain and Northern Competing interests: None
REVIEW RETURNED	12-Jan-2022

GENERAL COMMENTS	The authors investigate the "ability" of the UK and European Referral algorithms as well as investigation recommendations for the evaluation of children with short stature. They review the construction of both algorithms and how they were deployed. Here the emphasis is on the sensitivity and specificity of the evaluation using patients at two academic clinics in London. The background is well stated and the subjects summarized in Table 2 for the diagnoses (pathological) noted in table 1. 1. The methods are noted, but I suspect that they measure the children more accurately than most clinics—that will come back to be noted in the discussion and thus I believe that a couple of sentences about the precise protocol would be important. With a readily available electronic medical record or growth curve one measurement of height may be enough if it "fits" with previous measurements; otherwise, a second measurement to be sure the two are within perhaps 4 mm would be appropriate. 2. Table 2 notes that those with pathological diagnoses are the shorter with the greater deflection from Target Height. The issue of multiple height measurements may underlie the statistical insignificance of the HDSDS. 3. The sensitivities and specificities noted in table 3 are indicated to be "poor" but are actually better than those studies that used the "shotgun" approach as in their reference 20 and the editorial that went along with that. 4. Significantly more testing was done than shown in reference 20, but I am not sure how cost effective that was for new (unsuspected) diagnoses, which a decade ago were more than $100,000 per diagnosis. 5. Having noted the problems with sensitivity and specificity it would be a great service for these investigators to speculate how these two algorithms might be "tweaked" to have a better yield, knowing that a change in sensitivity leads to a change in the specificity. Additionally, whatever is proposed has to fit in with the time (personnel) and monetary constraints of the clinic and health care system

REVIEWER	Reviewer name: Malcolm Donaldson Institution and Country: United Kingdom of Great Britain and Northern Ireland Competing interests: None
REVIEW RETURNED	01-Feb-2022

GENERAL COMMENTS	Thank you for asking me to read this paper. The authors have retrospectively reviewed the case notes of 143 subjects from their two centres who were referred with short stature and applied three auxological parameters to determine their sensitivity and specificity in predicting pathological as opposed to non-pathological short stature. The parameters were: height standard deviation score (HtSDS), height deflection SDS (HtDefSDSand HtSDS minus target height (TH) SDS (Ht-THSDS) and the recommended cut-offs for both UK and Netherlands were applied. Sensitivity and specificity were low for single parameters but increased when HtSDS and Ht-THSDS were combined. The authors' conclusion is essentially that growth assessment prior to referral should take account of parental height although they do not specifically state this. This message might seem rather obvious but is nevertheless

	important, since insufficient attention to the importance of parental height measurement, and the calculation of mid-parental height, persist in the UK today. General comments A) I found the manuscript quite confusing and had to read the paper several times before I felt that I could give the above summary (which I hope is reasonably accurate). I have made some suggestions for clarification, which I hope will be helpful. B) I would personally find the paper easier to read if height standard deviation score, height deflection SDS and Height SDS– TH SDS were written: HtSDS; HtDefSDS; and Ht-TH SDS. I am more used to height velocity being used than height deflection and it took me time to adjust to the two 'D's in HDSDS. I accept that this issue is a matter of style, and authors should not feel under pressure to comply with my suggestions, but at least consider these. C) The definitions of the three auxological cut-offs need to be crystal clear. At present, height deflection SDS, height velocity and centiles are being mixed up. D) Given the importance which is (rightly) being attached to parental height measurement, authors need to state how often this was measured in both parents, measured in only one, reported in both; or unavailable. What correction factor was used for men and women – 13 cm? Specific comments Title: This is bland, refers to Europe rather than the Netherlands, and does not in my view encompass the aim of the study. Could authors consider something like "Sensitivity and specificity of UK and Dutch growth criteria in predicting the diagnosis of pathological short stature"? Abstract Objective: this is woolly and should state the aims of the study more clearly Study design The three parameters being tested, with their UK and Netherlands cut-offs, should be given. Demographics and referral route are not a priority in this abstract (word count always being a challenge). The four study groups should also be defined. Results Sensitivity and specificity for both UK and Netherlands are needed. Conclusion Aren't the authors showing that the three growth parameters have poor sensitivity and specificity when taken in isolation but that combining HtSDS with Ht-THSDS improves this, and highlighting the importance of parental height measurement prior to referral? Introduction Page 4, line 59. This sentence should be rephrased to reflect the benefits of early diagnosis. Consider a girl with coarctation of aorta and gonadal dysgenesis owing to Turner syndrome. Early diagnosis does not avoid these complications. Line 62. Perhaps 'serial height measurements…' Line 64. Perhaps '…(including chromosomal disorders)…. Line 68. Consider adding psychosocial problems to the list of conditions causing growth failure. Line 72. Is the age at diagnosis of 15.1 years in Turner syndrome reported in Denmark representative of other European countries and North America? Page 4, line 97 onwards. This paragraph needs some revision to make it clearer. The three growth parameters addressed in UK and Dutch guidelines are: short stature; discordance between child's height status and parental heights; and slow growth rate as indicated by low height velocity, centile crossing or change in HtSDS over time. These parameters should not be mentioned together in the same sentence (as Ht-TH SDS and Ht Def SDS are at present). The introduction should end with a statement to make it clear to the reader as to exactly what the paper is about. Methods Page 6 line 116: I suggest amending to read, "…..a consecutive

	series of new patient referrals of children and adolescents with short stature….” Line 118. Between 2016 and 2021 inclusive’? Line 122. Describe how height was measured, giving the make of stadiometer, and perhaps confirming that this was calibrated before each clinic. A separate sentence should state that height was remeasured between 6 and 12 months later, to allow growth rate to be assessed. I would personally leave the definition of height deflection until the next section, and not confuse the reader by referring to SD’s and velocity (cm/yr). Line 125. Please state the year of growth data (e.g. 1990) and give a reference for UK data used. At present, there is no specific mention of measured as opposed to reported parental height. Were parental heights measured by a trained observer? Were reported heights recorded when one or both parents were not present? Is it valid to take mother’s height as target height when for example she might measure 152 cm, and the father 180 cm? Please define target height and state how it was calculated. What correction factor was used – 13 cm? Why were two bone age methods used? A single method would make for a more consistent study. How useful was bone age? Line 133. ‘UK referral criteria’ This section needs careful revision - I had to make my own little table up with the three criteria as rows and the two countries as columns to get my head round this bit. For a start, there needs to be mention of the Dutch criteria to read “UK and Dutch referral criteria” I suggest a caption ‘UK and Dutch referral criteria; and diagnostic criteria. Authors can then give the UK criteria as three bullet-pointed rows, one for each criterion. Height deflection should be clearly defined in a single term - change in Ht SDS over a minimum time interval, and not as a mixture of change in HtSDS, centile crossing, and decreased in height velocity - this is confusing for the reader. The Dutch referral criteria can be either given in tandem – row by row; or following this. The four diagnostic groups can then be defined. Authors must remember that actual data, such as follow up period for the various groups, constitute results and not methods and should be transferred to the results section. Page 7 line 158. Should this read, “Selected individuals referred with short stature…”? Results There should be an introductory paragraph stating how many of the 143 individuals had parental heights available and in how many both or one parents had been measured. We also need to know in how many patients Height deflection SDS had been determined. Page 8, line 187. The caption, “Referrals and source of referrals” should state “See Figure 1”. Figure 1 constitutes data, not methods, and belongs here. Line 191 onwards might read: “Data were available in the remaining 143 individuals in whom mean (range) age was 8.7(0.5-19.9) years.” No need to cite Figure 1 here, because it should already have been flagged up. Line 192. When mean is given, it is usual to give SD and when median is given it is usual to give range, depending on whether the data are normally distributed or not. Can authors look at this? Page 9 line 195. I would be wary of the term ‘height deficit’. This implies that short stature represents a deficiency, yet we know that most short individuals are normal. Authors could write, “height status” Line 205. For Group 4, the median (range) HtSDS is needed. Presumably this was around – 2 SDS but not as low as -2.6 SDS? The tables need some attention. Table 1 needs a stand-alone title so that the reader can understand it without consulting the text. Karyotypes can be written

	conventionally as 46, XY and 45,X. Table 2 can be sacrificed and incorporated in the text. Line 216. This should read, "Referral criteria (See Tables 2 and 3)". Supplemental Table 2, which is central to the paper, should replace the current Table 2. Line 220: 'to correctly identify…" Diagnostic investigations The authors report that only 4 out of 143 children had no investigation. How many children had bone age assessment only? Given that normal familial short stature and CDGP are clinical diagnoses, was there overreliance on investigation? Discussion This is long, extending to five pages, and can be reduced. Page 9 Lines 245-251 duplicate the introduction. Does the male predominance of short stature referrals reflect the differing attitudes to normal variant short stature in UK culture? Does the limited helpfulness of height deflection reflect the logistic difficulties of obtaining serial measurements and the time needed to demonstrate a trend, rather than the consistency and quality of height measurements? With appropriate training, accurate height measurement should be within the scope of all health care professionals, including those in primary care. If short stature is causing the family and the boy or girl distress, then referral to a specialist centre is appropriate, even if the degree of short stature is not as severe as the very short UK and Dutch cut-offs. What is the take home message from this study? This needs to come over more clearly both at the end of the abstract and the end of this discussion. Conclusion This paper has valuable information. The take home message, that combining degree of short stature and height in relation to parental heights is more valuable than any single parameter in discriminating between normal variant and pathological short stature, is important. However, the authors need to be clearer about what they are trying to say, and submit a more focused manuscript with a shorter and more succinct discussion. Given the importance of parental height measurement, this needs to be more clearly and thoroughly dealt with in the methods section. I hope that the authors will be able to come up with a tighter and clearer revision and would be glad to review this. Malcolm Donaldson

VERSION 1 – AUTHOR RESPONSE

Manuscript bmjpo-2021-001385: "EVALUATING THE UK AND EUROPEAN REFERRAL AND INVESTIGATION RECOMMENDATIONS FOR CHILDREN WITH SHORT STATURE"

Responses to Editor in Chief and the reviewer's comments

The authors have addressed each of the reviewer's comments in detail (below). The changes made to the manuscript are highlighted in the revised manuscript.

1. Editor in Chief Comments to Author:

Title add "a single centre audit" and replace "European" with "Dutch"

We thank the Editor in Chief for their comments. The title has been changed accordingly.

Introduction last para. Avoid use of the word "first" to describe your study (see instructions to authors)

Thank you for highlighting this, this has been removed (line 114).

Supple table 1 should be in the main paper

This has been added to the main paper (Table 1).

2. Reviewers' Comments to Author:

<u>Reviewer 1: Dr. Peter Flom, Peter Flom Consulting</u>

Comments to Author

I confine my remarks to statistical aspects of this paper.

I am confused. The statistical methods seem to be appropriate, but I'm unclear on exactly what was done and why. Some specific examples are listed below, but I think the whole paper needs a lot more explanation of what each measure was, how often it was measured, and how it was used. Also, I'm kind of unclear on what the whole purpose was.

Sorry .... and maybe the confusion is due to the fact that I am a statistician and not a subject matter expert, but, if you are aiming for a somewhat general audience, I think more explanation and detail is needed.

The authors are very grateful for your suggestions. We had addressed these individually below.

Lines 38-41 Why compare the pathological group to the non-pathological group on the measures that define the pathology? Doesn't this wind up being "short people were shorter"?

We thank the reviewer for this comment. The groups were not just defined by height measurements. They were categorised by clinical diagnoses. Height alone is an insufficient marker of whether a child has a disorder of growth. Due to the genetic determination of height, children may be a normal height compared to the rest of the population but have a pathological process meaning they could fail to reach their genetically determined height potential. Similarly, children may appear to be 'short' but this is a normal growth trajectory within a short family. Hence the combination of screening rules, clinical assessment and a standardised protocol of investiagtions are necessary. For example, Group 1 which encompassed primary and secondary growth disorders, were defined by their abnormal growth, clinical assessment and the investigations which confirmed the specific condition causing the short stature.

Line 73 Not a statistics question, but why just girls?

This refers to individuals with Turner syndrome which is a sex chromosomal disorder affecting females. Hence the use of 'girls' in this sentence (line 73).

Line 122 but actually a general comment. I don't see a description of how often, or at what intervals, measurements were used. Here (line 122) it looks like there were 2 measurements, at an interval of 6 to 12 months. But this should be described more completely.

This has been clarified in the auxology section of the methods (lines 126-143).

General: All the variables need to be defined. What is target height, for instance?

This has been clarified in the auxology section of the methods (lines 126-143).

Line 165 How was this division into groups done? (This is shown in Fig 1, but I think it needs to be in the text).

This has been clarified in the diagnostic classification section of the methods (lines 159-174).

Reviewer 2: Dr. Alan Rogol, University of Virginia

Comments to the Author

The authors investigate the "ability" of the UK and European Referral algorithms as well as investigation recommendations for the evaluation of children with short stature. They review the construction of both algorithms and how they were deployed. Here the emphasis is on the sensitivity and specificity of the evaluation using patients at two academic clinics in London. The background is well stated and the subjects summarized in Table 2 for the diagnoses (pathological) noted in table 1.

The authors are very grateful for your comments. We have addressed these individually below.

1. The methods are noted, but i suspect that they measure the children more accurately than most clinics--that will come back to be noted in the discussion and thus i believe that a couple of sentences about the precise protocol would be important. With a readily available electronic medical record or growth curve one measurement of height may be enough if it "fits" with previous measurements; otherwise, a second measurement to be sure the two are within perhaps 4 mm would be appropriate.

Thank you for this important comment. We have clarified the measurement protocol in the auxology section of the methods (lines 126-143). The measurements were undertaken as part of the routine clinic assessment and as such were single measurements. They were done by a trained auxologist using calibrated equipment and therefore we agree that our measurements are likely to be more accurate than some clinics. This is also highlighted in the discussion (line 302).

We agree that ideally, our study protocol would include second measurements to validate the accuracy of the values. However, as this study was a retrospective observational study, this was not possible. In terms of assessing whether the first measurement 'fits' with previous measurements, this is usually not possible in the UK, as children's heights are not routinely measured in primary care and NCMP measurement are not recorded in the medical records.

2. Table 2 notes that those with pathological diagnoses are the shorter with the greater deflection from Target Height. The issue of multiple height measurements may underlie the statistical insignificance of the HDSDS.

As addressed above, we agree with the reviewer that multiple height measurements would have improved the accuracy and we would expect abnormal height deflectioin children with pathological short stature. However, particularly low sensitivities and limited helpfulness are observed in several published studies. This suggests there are more universal issues which are complex and likely to be multifactorial. We have highlighted these discussion points further in lines 292-293 and 298-305.

3. The sensitivities and specificities noted in table 3 are indicated to be "poor" but are actually better than those studies that used the "shotgun" approach as in their reference 20 and the editorial that went along with that.

We have changed 'poor' to 'suboptimal' in the abstract (line 48). The overarching message of this manuscript is that a combined approach using the HtSDS and Ht-THSDS criteria enhances the identification of pathological short stature. We have clarified this throughout the revised manuscript. We do agree that this approach is superior to the one adopted in reference 20 where growth rates were only documented in 37% patients. It would be difficult to compare our study to reference 20 as children who were at higher risk of pathology (with symptoms or low height velocity) were excluded at the outset.

The title has been changed accordingly.

Abstract

Objective: this is woolly and should state the aims of the study more clearly

The abstract objective has been modified to state the aims more clearly (lines 30-32).

Study design: The three parameters being tested, with their UK and Netherlands cut-offs, should be given. Demographics and referral route are not a priority in this abstract (word count always being a challenge). The four study groups should also be defined.

The abstract study design has been modified and now includes the 3 parameters being tested and the associated UK and Dutch cut-offs (lines 35-38). As we have now exceeded the word count limitation, we have been unable to define the four study groups in the abstract. However, these have been clarified in the methods section.

Results: Sensitivity and specificity for both UK and Netherlands are needed.

These have been added to the abstract results section (lines 39-47).

Conclusion: Aren't the authors showing that the three growth parameters have poor sensitivity and specificity when taken in isolation but that combining HtSDS with Ht-THSDS improves this, and highlighting the importance of parental height measurement prior to referral?

The abstract conclusion has been modified to clearly state the key findings (lines 48-51).

Introduction

Page 4, line 59. This sentence should be rephrased to reflect the benefits of early diagnosis. Consider a girl with coarctation of aorta and gonadal dysgenesis owing to Turner syndrome. Early diagnosis does not avoid these complications.

Line 62. Perhaps 'serial height measurements…'

Added to the text (line 63).

Line 64. Perhaps '…(including chromosomal disorders)….

Added to the text (line 65).

Line 68. Consider adding psychosocial problems to the list of conditions causing growth failure.

Added to the text (line 67).

Line 72. Is the age at diagnosis of 15.1 years in Turner syndrome reported in Denmark representative of other European countries and North America?

Thank you for pointing this out. For balance, we have added the data from the UK and US (lines 73-78) and added an additional reference (reference 3).

Page 4, line 97 onwards. This paragraph needs some revision to make it clearer. The three growth parameters addressed in UK and Dutch guidelines are: short stature; discordance between child's height status and parental heights; and slow growth rate as indicated by low height velocity, centile crossing or change in HtSDS over time. These parameters should not be mentioned together in the same sentence (as Ht-TH SDS and Ht Def SDS are at present).

This has been clarified (lines 100-105).

The introduction should end with a statement to make it clear to the reader as to exactly what the paper is about.

This has been clarified (lines 114-117).

Methods

Page 6 line 116: I suggest amending to read, "…..a consecutive series of new patient referrals of children and adolescents with short stature…."

This has been amended (line 121-122).

Line 118. Between 2016 and 2021 inclusive'?

This has been amended (line 123).

Line 122. Describe how height was measured, giving the make of stadiometer, and perhaps confirming that this was calibrated before each clinic. A separate sentence should state that height was remeasured between 6 and 12 months later, to allow growth rate to be assessed. I would personally leave the definition of height deflection until the next section, and not confuse the reader by referring to SD's and velocity (cm/yr).

This has been clarified in the auxology section of the methods (lines 126-143).

Line 125. Please state the year of growth data (e.g. 1990) and give a reference for UK data used.

The details of the reference data used has been added (line 133).

At present, there is no specific mention of measured as opposed to reported parental height.

Were parental heights measured by a trained observer? Were reported heights recorded when one or both parents were not present? Is it valid to take mother's height as target height when for example she might measure 152 cm, and the father 180 cm?

As now stated in the manuscript (lines 207-208), both parents heights were available for 133 individuals and maternal height alone in 3 individuals (total 136). These were measured by a trained observer and were not reported heights. If only one parent attended the first visit, the other parent was measured at a subsequent appointment, this has been clarified in the relevant method section (lines 129-130). Mother's height alone was only available for only 3 individuals, the methodology for calculating TH is now fully described (138-140)

Please define target height and state how it was calculated. What correction factor was used – 13 cm?

We have clarified the definition and method for calculating target height (lines 134-137).

Why were two bone age methods used? A single method would make for a more consistent study. How useful was bone age?

The bone ages reported in this manuscript were calculated by the Greulich and Pyle method. We have corrected this (line 147).

Line 133. 'UK referral criteria'

This section needs careful revision - I had to make my own little table up with the three criteria as rows and the two countries as columns to get my head round this bit.

For a start, there needs to be mention of the Dutch criteria to read "UK and Dutch referral criteria"

I suggest a caption 'UK and Dutch referral criteria; and diagnostic criteria.

Authors can then give the UK criteria as three bullet-pointed rows, one for each criterion. Height deflection should be clearly defined in a single term - change in Ht SDS over a minimum time interval, and not as a mixture of change in HtSDS, centile crossing, and decreased in height velocity - this is confusing for the reader.

The Dutch referral criteria can be either given in tandem – row by row; or following this.

The four diagnostic groups can then be defined.

Thank you for these comments, we have rewritten this section as suggested to make the UK and Dutch criteria clearer (lines 150-157).

Authors must remember that actual data, such as follow up period for the various groups, constitute results and not methods and should be transferred to the results section.

We have moved the follow up periods from methods to the relevant results sections (lines 224 and 230)

Page 7 line 158. Should this read, "Selected individuals referred with short stature…"?

We have corrected this (line 172).

Results

There should be an introductory paragraph stating how many of the 143 individuals had parental heights available and in how many both or one parents had been measured. We also need to know in how many patients Height deflection SDS had been determined.

This has been clarified in the 'referrals and source of referrals' section (lines 205-217).

Page 8, line 187. The caption, "Referrals and source of referrals" should state "See Figure 1". Figure 1 constitutes data, not methods, and belongs here.

We have corrected this (line 205).

Line 191 onwards might read: "Data were available in the remaining 143 individuals in whom mean (range) age was 8.7(0.5-19.9) years." No need to cite Figure 1 here, because it should already have been flagged up.

We have corrected this (line 205-206).

Line 192. When mean is given, it is usual to give SD and when median is given it is usual to give range, depending on whether the data are normally distributed or not. Can authors look at this?

We have corrected this in line 211 and throughout this and subsequent results sections.

Page 9 line 195. I would be wary of the term 'height deficit'. This implies that short stature represents a deficiency, yet we know that most short individuals are normal. Authors could write, "height status"

We have corrected this in line 215 and elsewhere in the manuscript.

Line 205. For Group 4, the median (range) HtSDS is needed. Presumably this was around – 2 SDS but not as low as -2.6 SDS?

As above, we have added the mean and SDs for all 4 groups (lines 219-231).

The tables need some attention. Table 1 needs a stand-alone title so that the reader can understand it without consulting the text. Karyotypes can be written conventionally as 46, XY and 45,X.

The title of this Table (now Table 2) has been changed and the karyotypes have been amended as suggested.

Table 2 can be sacrificed and incorporated in the text.

This table has been removed from the main manuscript and placed in the supplemental document (Supplemental table 1). The relevant data has been added the results sections in lines lines 219-231.

Line 216. This should read, "Referral criteria (See Tables 2 and 3)".

We have corrected this (line 239). These tables are now 3 and 4.

Supplemental Table 2, which is central to the paper, should replace the current Table 2.

We have moved this to the main manuscript (Table 3).

Line 220: 'to correctly identify…"

We have corrected this (line 243).

Diagnostic investigations

The authors report that only 4 out of 143 children had no investigation. How many children had bone age assessment only? Given that normal familial short stature and CDGP are clinical diagnoses, was there overreliance on investigation?

No patients had bone age alone, this has been clarified in the manuscript (line 257). Clearly there is a degree of overreliance on investigations as even though Group 4 (normal stature and normal growth trajectory) had the least testing

compared to the other groups, they still had mean 82% of the recommended tests and 41% had all the investigations. We have highlighted this in the discussion (lines 344-346).

Discussion

This is long, extending to five pages, and can be reduced. Page 9 Lines 245-251 duplicate the introduction.

We agree and have significantly reduced this to avoid duplication with the introduction.

Does the male predominance of short stature referrals reflect the differing attitudes to normal variant short stature in UK culture?

The predominance of male referrals is striking in our cohort and is discussed in detail (lines 269-280). The authors agree that attitudes and social pressures may in part explain the well-recognised gender biases that exist in short stature referral and treatment. An additional sentence has been added to emphasise this (lines 277-278).

Does the limited helpfulness of height deflection reflect the logistic difficulties of obtaining serial measurements and the time needed to demonstrate a trend, rather than the consistency and quality of height measurements? With appropriate training, accurate height measurement should be within the scope of all health care professionals, including those in primary care.

We agree with the reviewer that the difficulties with height deflection are complex and likely to be multifactorial. We have highlighted these discussion points in lines 295-297 and 303-305.

If short stature is causing the family and the boy or girl distress, then referral to a specialist centre is appropriate, even if the degree of short stature is not as severe as the very short UK and Dutch cut-offs.

We agree with the reviewer and our data show that the strict UK criteria for height SDS would miss a significant proportion (59%) of children with pathological stature. However, there is also a need to reduce unnecessary referrals as this can result in unnecessary anxiety and NHS resource wastage. We have tried to clarify this message in the revised manuscript.

What is the take home message from this study? This needs to come over more clearly both at the end of the abstract and the end of this discussion.

This has been clarified at the end of the abstract and the discussion.

Conclusion

This paper has valuable information. The take home message, that combining degree of short stature and height in relation to parental heights is more valuable than any single parameter in discriminating between normal variant and pathological short stature, is important.

However, the authors need to be clearer about what they are trying to say, and submit a more focused manuscript with a shorter and more succinct discussion. Given the importance of parental height measurement, this needs to be more clearly and thoroughly dealt with in the methods section.

I hope that the authors will be able to come up with a tighter and clearer revision and would be glad to review this.

Thank you for your in-depth review and helpful comments. We think the manuscript has been significantly improved by the revisions.