

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)	Social stigmatization in late identified patients with disorders of sex development in Indonesia
AUTHORS	Dessens, Arianne; Ediati, Annastasia; Juniarto, A.; Birnie, Erwin; Okkerse, Jolanda; Wisniewski, Amy; Drop, Stenvert; Faradz, Sultana

VERSION 1 - REVIEW

REVIEWER	Meyer Bahlburg, Heino Dept. Psychiatry, College of Physicians & Surgeons of Columbia University, USA Competing interests: None.
REVIEW RETURNED	19-Jun-2017

GENERAL COMMENTS	<p>This manuscript reports a unique study of social stigma in 2 East-Asian DSD samples, one of children and adolescents, the other of adults, most of whom untreated by hormones and/or surgery prior to the study, i.e., samples that are very difficult to obtain in Western countries - thus, presenting very important data. I have a number of suggestions to improve the MS.</p> <ol style="list-style-type: none"> 1. One of the American or European co-authors who are used to write English-language publications should thoroughly review and edit the MS for phrasing and grammar. (Some examples are included among the points below). 2. P. 2 of 22, Abstract, lines 11-13: This is not a sentence. 3. Same page, lines 13-14: That sentence is unclear and awkward. 4. P. 3, lines 6-9: Add how clinical management tries to achieve the reduction etc. 5. P. 3, lines 20-23: Use the present tense. 6. P. 7, Table 3: Explain "Social gender role change" in a footnote: Is this physician-imposed in infancy or early childhood due to medical considerations, or due to emerging cross-gender identity, or is this patient-initiated gender change at later ages? 7. P. 11, lines 20-22: Same as point (6) above. 8. P. 15, line 3, "hard to cope with": The study does not provide data on coping per se; needs re-phrasing. 9. P. 15, lines 5-6: How do the authors conceptualize the causal direction: stigma \square stress, or stress vulnerability \square higher sensitivity to potential stigma \square reports of greater stigma frequency? 10. P. 15, line 9, "qualitative data": The Methods section must provide details on the collection and evaluation of qualitative data, and the qualitative findings should first be reported in the Results section. 11. P. 15, lines 17-18: Unclear meaning, awkward formulation. 12. P. 16, line 3, "collective society": Means what? 13. P. 16, lines 8-10: What happened to the other 4 patients? 14. P. 16, lines 15-16: Unclear meaning, awkward formulation.
-------------------------	--

	<p>15. P. 17, lines 10-13: The last 2 sentences of the conclusion are not addressed by data from this study and should, therefore, not be included in the Conclusion.</p> <p>16. P. 20: Ref. 17 repeats Ref. 2.</p>
--	---

REVIEWER	<p>Zainuddin, Ani Amelia</p> <p>The National University of Malaysia (UKM)</p> <p>Malaysia</p> <p>Competing interests: There are no competing interests.</p>
REVIEW RETURNED	20-Jun-2017

GENERAL COMMENTS	<p>Very interesting, important and well written paper. The paper should be accepted after minor revisions are made.</p> <p>The minor revisions recommended are;</p> <ol style="list-style-type: none"> 1. Need to include more details on how the SSS-DSD was constructed, i.e.. were the questions derived from the researchers' past clinical experiences, or from a prior study or from focus groups or experts panel? Any pretesting to test for face validity conducted? Any pilot study done? What language was used? 2. Detail the different domains of the SSS-DSD scale and the total maximum score of each domain and total maximum score of all the domains combined in a paragraph in the Materials section so that Table 4a and Table 4b is easier to comprehend. What does a low score mean? What does a high score mean? For example, does a high score mean higher degree of stigmatisation / stress? And vice versa? 3. On page 6, line 25 continuing into line 1 page 7, the authors wrote, "... the majority had attended high school and worked in the lower -income sector or were unemployed." please explain who does the majority in this sentence refer to? The parents or the participants? 4. Under the Discussion section, page 15, lines from 14 to 18, there were contradicting statements. First it was stated that " Overall, many patients did not give high rates for experienced social stigmatisation". Then line 17-18: "...patients with an atypicaldeal with incomprehension and social stigmatisation". Make your points in this paragraph clearer. 5. Same page as above (page 15, lines 20 to 23), the authors state " Does the study also support medical interventions.....46, XY DSD?" Where does this point come in in the overall study? This is not clear, this sentence (and the following one) is best removed completely. 6. Another recommendation to reduce feelings of isolation and also maybe to empower the DSD patients and parents better is to have more training of more psychologists that can counsel them. 7. Page 16, lines 15 to 16, the authors wrote " Our study focused on the relationship between local stigmatisation and atypical consequently to delay of treatment." I think a word is missing. Perhaps " Our study focused on the relationship between local stigmatisation and atypical appearances as a consequence of delay of treatment".
-------------------------	--

REVIEWER	<p>Zucker, Kenneth</p> <p>University of Toronto</p> <p>Canada</p> <p>Competing interests: None</p>
REVIEW RETURNED	20-Jun-2017

<p>GENERAL COMMENTS</p>	<p>I have now read the unmasked manuscript entitled “Social stigmatization in late identified patients with disorders of sex development in Indonesia.” The authors (five PhDs and three MD/PhDs) are all experts in the DSD area.</p> <p>The English and proper use of punctuation could be improved. I suggest that one of the authors whose first language is English work on this.</p> <p>1. Abstract: “Rejection elicited depression in females and children and adolescents...”: This sentence does not read well. It is not entirely clear to me what “In text analysis...” means.</p> <p>2. If the “outcome” measure is “social stigmatization,” why not use multiple regression to identify significant predictors? Or, let us say that the outcome measure is depression, why not see if visibility of the condition functions as a mediator of stigma? In general, I feel like the analytic plan could be improved upon.</p> <p>3. Introduction: p. 3, para 1: “These interventions are criticized...” By whom?</p> <p>4. See the new paper by Meyer-Bahlburg et al. (2017) online first in Archives of Sexual Behavior.</p> <p>5. Method: What was the participation rate? How many patients were excluded? Were participants compensated for their participation?</p> <p>6. P. 6, para 1: I would change orally to verbally.</p> <p>7. In the Method section, the number of items on the SSS-DSD should be given.</p> <p>8. Eigen values should be eigenvalues.</p> <p>9. Results: The order should be reliability and validity. One can have reliability, but not validity. One can’t have validity without reliability.</p> <p>10. For each item (as in Table 3a), I would like to see the mean ratings and SD (in Table 4a and 4b, the authors report the median value and the range for the factor scores, but I would like to see the mean and SD—also, in these tables, the number of items/factor should be noted). The English translation of some of the items reads poorly. There are some oddities that should be clarified. Let us take Question 1a as an example: Can other people see that your child got (sic) a genital (sic) that is (slightly) different from that of other children? Suppose the parent answers with a 1, meaning “not at all.” Question 2a then asks “How stressful is this to you?” If the parent answers 1a with “not at all” one presumes that Question 2a could only be answered as “not at all.” I find it a bit weird that the “how stressful” questions are embedded with questions about stigmatization. Personally, I would recommend removing all of the “how stressful” questions and re-do the factor analysis. The factor loadings should be expressed like this: 0.60 (not with the “period” elevated like it is a bullet point—same when reporting p values). The same issues apply to Table 3b. For each of the factors on the SSS-DSD, one should see a mean score along with a SD. It is really not</p>
--------------------------------	--

	<p>clear how the authors performed the correlations on p. 10.</p> <p>11. Tables 4a and b: The authors report p values but do not provide the actual values from the Mann-Whitney U test. These values must be provided. It is rather bizarre, for example, to see that median values are, say for Atypical appearance/behaviour identical for boys and girls, yet the p value is $<.001$. Without seeing the U values, one become skeptical. Given the numerous tests conducted, one wonders about using a correction factor.</p> <p>12. I prefer delaying comment on the Discussion until a revised ms is prepared.</p> <p>13. Reference 2 = Reference 17.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author

This manuscript reports a unique study of social stigma in 2 East-Asian DSD samples, one of children and adolescents, the other of adults, most of whom untreated by hormones and/or surgery prior to the study, i.e., samples that are very difficult to obtain in Western countries - thus, presenting very important data. I have a number of suggestions to improve the MS.

1. One of the American or European co-authors who are used to write English-language publications should thoroughly review and edit the MS for phrasing and grammar. (Some examples are included among the points below).

Response: English writing and grammar were reviewed and rephrased accordingly.

2. P. 2 of 22, Abstract, lines 11-13: This is not a sentence.

Response: We adapted the sentence.

3. Same page, lines 13-14: That sentence is unclear and awkward.

Response: We adapted the sentence.

4. P. 3, lines 6-9: Add how clinical management tries to achieve the reduction etc.

Response: Sorry I don't understand the comment. To which sentence does the comment refers?

5. P. 3, lines 20-23: Use the present tense.

Response: We adapted the sentence.

6. P. 7, Table 3: Explain “Social gender role change” in a footnote: Is this physician-imposed in infancy or early childhood due to medical considerations, or due to emerging cross-gender identity, or is this patient-initiated gender change at later ages?

Response: We adapted Table 2. For detailed information on social gender change we refer to our paper on gender development

7. P. 11, lines 20-22: Same as point (6) above.

Response: For detailed information on social gender change we refer to our paper on gender development

8. P. 15, line 3, “hard to cope with”: The study does not provide data on coping per se; needs re-phrasing.

Response: We adapted the sentence.

9. P. 15, lines 5-6: How do the authors conceptualize the causal direction: stigma \diamond stress, or stress vulnerability \diamond higher sensitivity to potential stigma \diamond reports of greater stigma frequency?

Response: This study design basically is a cross sectional study. This means that the study can only produce associations between determinant and outcome as result, but causality cannot be determined. While it is assumed that stigma results in stress, causality remains unproven.

10. P. 15, line 9, “qualitative data”: The Methods section must provide details on the collection and evaluation of qualitative data, and the qualitative findings should first be reported in the Results section.

Response: We provided information on applied methods and findings in the Methods and Results sections (lines 269,270,272,273,293-296,416-420).

11. P. 15, lines 17-18: Unclear meaning, awkward formulation.

Response: We adapted the sentence.

12. P. 16, line 3, “collective society”: Means what?

Response: In Western cultures the interests (needs, rights) of the individual person is highly valued. In many non-western cultures, the collective needs of the community, the ethnic group or tribe, is highly valued and individuals may be asked to accommodate personal needs to that of the community. Also see: Meyer Bahlburg H.F.L. Arch. Sex Behav. 2017;46:337-339 and references 2,14,17-21,24&25

13. P. 16, lines 8-10: What happened to the other 4 patients?

Response: See our paper in ref 24: we added the omitted information.

14. P. 16, lines 15-16: Unclear meaning, awkward formulation.

Response: We adapted the sentence.

15. P. 17, lines 10-13: The last 2 sentences of the conclusion are not addressed by data from this study and should, therefore, not be included in the Conclusion.

Response: We agree with the reviewer. We adapted the sentences.

16. P. 20: Ref. 17 repeats Ref. 2.

Response: We adapted the references.

Reviewer: 2

Comments to the Author

Very interesting, important and well written paper. The paper should be accepted after minor revisions are made.

The minor revisions recommended are;

1. Need to include more details on how the SSS-DSD was constructed, i.e. were the questions derived from the researchers' past clinical experiences, or from a prior study or from focus groups or experts panel? Any pretesting to test for face validity conducted? Any pilot study done? What language was used?

Response: We adapted the introduction (lines80-84). The questions in the SSS-DSD were based on patients' and parents' spontaneous reports on their experiences with social stigmatization. Three of the authors, Annastasia Ediat, Arianne Dessens and Jolanda Okkerse developed and constructed this the scale, Saskia Wieringa gave advices. During this process, they communicated in English. The SSS-DSD was written in Bahasa Indonesia and in English. Saskia Wieringa, who speaks Bahasa Indonesia well and who got a degree in English literature, checked translations.

The rating scale was piloted in a small group of 20 patients and parents with DSD. After a few adaptations, the SSS-DSD seemed suitable for application in this study. Apart from this small scale pilot, no large scale psychometric pretesting of the measurement instrument was applied. We were aware that formal psychometric pretesting is desirable to optimize the instrument. However, this was considered unfeasible; large patient groups (or sizable numbers of patient-parents) were not available – regrettably a problem often encountered research in patients with rare genetic conditions.

We added the following to the Methods section (see "Materials" paragraph):

"The rating scale was piloted in a small group of 20 patients and parents with DSD. After a few adaptations, the SSS-DSD seemed suitable for application in this study. Formal large scale psychometric pretesting among sizable numbers of patients or their parents was considered unfeasible in view of the limited numbers of patients with rare genetic conditions." (see lines 133-138).

2. Detail the different domains of the SSS-DSD scale and the total maximum score of each domain and total maximum score of all the domains combined in a paragraph in the Materials section so that Table 4a and Table 4b is easier to comprehend. What does a low score mean? What does a high

score mean? For example, does a high score mean higher degree of stigmatization / stress? And vice versa?

Response: We made some adaptations in Tables 3a+b and 4a+b. Scales and scale score ranges are better indicated now, making it easier to understand the data.

We made the following changes to the manuscript:

"For all sum scores, a higher score indicates a relatively higher level of stigma, atypicality, social exclusion emotional problems." (See Methods, Data Analysis paragraph, lines 183-5; and below Tables 4a and 4b .)

3. On page 6, line 25 continuing into line 1 page 7, the authors wrote, "... the majority had attended high school and worked in the lower -income sector or were unemployed." please explain who does the majority in this sentence refer to? The parents or the participants?

Response: In table 2, frequencies for numbers of fathers, mothers and adult patients for attained education level and type of employment are separately given.

4. Under the Discussion section, page 15, lines from 14 to 18, there were contradicting statements. First it was stated that " Overall, many patients did not give high rates for experienced social stigmatization". Then line 17-18: "...patients with an atypicaldeal with incomprehension and social stigmatization". Make your points in this paragraph clearer.

Response: We adapted the text

5. Same page as above (page 15, lines 20 to 23), the authors state" Does the study also support medical interventions.....46, XY DSD?" Where does this point come in in the overall study? This is not clear, this sentence (and the following one) is best removed completely.

Response: Thank you for this valid suggestion. We adapted the text

6. Another recommendation to reduce feelings of isolation and also maybe to empower the DSD patients and parents better is to have more training of more psychologists that can counsel them.

Response: We agree with the reviewer and adapted the text

7. Page 16, lines 15 to 16, the authors wrote" Our study focused on the relationship between local stigmatization and atypical consequently to delay of treatment." I think a word is missing. Perhaps "Our study focused on the relationship between local stigmatization and atypical appearances as a consequence of delay of treatment".

Response: This correct, indeed a word is missing, for which we apologise. We corrected and adapted the text accordingly.

Reviewer: 3

Comments to the Author

I have now read the unmasked manuscript entitled “Social stigmatization in late identified patients with disorders of sex development in Indonesia.” The authors (five PhDs and three MD/PhDs) are all experts in the DSD area.

The English and proper use of punctuation could be improved. I suggest that one of the authors whose first language is English work on this.

1. Abstract: “Rejection elicited depression in females and children and adolescents...”: This sentence does not read well.

Response: We adapted the abstract

It is not entirely clear to me what “In text analysis...” means.

Response: We adapted the abstract. The open-ended questions are now described in the Methods section (lines 127-129)

2. If the “outcome” measure is “social stigmatization,” why not use multiple regression to identify significant predictors? Or, let us say that the outcome measure is depression, why not see if visibility of the condition functions as a mediator of stigma? In general, I feel like the analytic plan could be improved upon.

Response: We agree with the reviewer that a multivariable model would have been desirable, and that the univariate analyses presented in Tables 4A+4B are somewhat unsatisfactory.

Considering the limited sample size (n=81 for Table 4A and n=34 for Table 4B) and the considerable number of possible determinants (to name a few: sex of rearing, treatment status, social gender role, visibility of DSD and adult/parent, but several others can be hypothesized; e.g. see Table 2) as well as the multiple outcomes measures, a multiple regression model was deemed unfeasible, leading to unclear results with unclear validity. That is why we focused the analysis on the results of the univariate analysis.

3. Introduction: p. 3, para 1: “These interventions are criticized...” By whom?

Response: By authors from references 2-6

4. See the new paper by Meyer-Bahlburg et al. (2017) online first in Archives of Sexual Behavior.

Response: Can the reviewer specify which paper published by Meyer Bahlburg and coworkers in Arch Sex Behav in 2017? There are three papers on this subject, we selected the one in reference 2.

5. Method: What was the participation rate? How many patients were excluded? Were participants compensated for their participation?

Response: The participation rate was 78%. More extensive information on patient inclusion are provided in papers refs 14, 24, 25. We adapted this in the manuscript

6. P. 6, para 1: I would change orally to verbally.

We adapted the manuscript

7. In the Method section, the number of items on the SSS-DSD should be given.

Response: The items are given in Tables 3a+b

8. Eigen values should be eigenvalues.

We adapted the manuscript

9. Results: The order should be reliability and validity. One can have reliability, but not validity. One can't have validity without reliability.

Response: We agree that it is useless to consider validity when the reliability of the measurement scale is low or weak. Our paper does not include direct measures of reliability (e.g. test-retest reliability), except Cronbach's alpha as a measure of internal consistency.

We adapted our paper as follows:

1) Title "Validity and Reliability of..." is rephrased into "Reliability and Validity of..." (see Results line 214); and

2) In this paragraph, we first address the Reliability and then Validity (see changes made in lines 215-223, and in lines 235-237.

10. A. For each item (as in Table 3a), I would like to see the mean ratings and SD (in Table 4a and 4b, the authors report the median value and the range for the factor scores, but I would like to see the mean and SD

Response: This suggestion surprised us somewhat. Mean and SD of the rating is very helpful when the SSS-DSD data approximately follow a Gaussian distribution. However, our data are often skewed which means that they clearly do not follow a Gaussian distribution. Reporting means and SDs of skewed data is usually discouraged because data that are skewed to the right tend to produce inflated means and SDs, while skewness to the right tends to produce deflated means and SDs. In both cases, skewness tends to produce incorrect descriptive data to the reader. That is why we summarized our data and scores in medians and ranges (see for example Table 4A+B), which is generally considered the appropriate alternative presentation for means and SDs when data are skewed. We are happy to add the means and SDs to Tables 4b+b. Since that would produce very "busy" tables, we refrained from this. When the reviewer and editor explicitly requests us to add the means and SDs to the tables, we are still happy to do so.

B. also, in these tables, the number of items/factor should be noted).

Response: The number of items for each factor is provided in Tables 3a+b.

The English translation of some of the items reads poorly. There are some oddities that should be clarified. Let us take Question 1a as an example: Can other people see that your child got (sic) a genital (sic) that is (slightly) different from that of other children? Suppose the parent answers with a 1, meaning "not at all." Question 2a then asks "How stressful is this to you?" If the parent answers 1a with "not at all" one presumes that Question 2a could only be answered as "not at all." I find it a bit weird that the "how stressful" questions are embedded with questions about stigmatization. Personally, I would recommend removing all of the "how stressful" questions and re-do the factor analysis.

Response: We added the b. questions “How stressful is this to you” to investigate the emotional impact of atypical physical appearance and related social stigmatization. Parents and patients who rated “1” (“not at all”), on the a. question, could rate 1 or 2 on the b. question. It worked this way, as we observed high correlations between the a. and b. questions.

The content of the questions was based on parents and patients reports during visits to the outpatient clinic (lines 80-84). As can be seen in Table 2, the majority of parents and patients came from rural and suburban areas. In Indonesia, as well as any other Asian countries, there are large socioeconomic differences between citizens and people from rural and suburban areas regarding access to facilities. Many households in rural areas do not have a bathroom, people take a bath in the river. Villages in rural areas are small, collective communities. Community members know other each other well, are involved and share. Under such circumstances it is difficult to keep private parts private, or to keep anything else private.

C. The factor loadings should be expressed like this: 0.60 (not with the “period” elevated like it is a bullet point—same when reporting p values). The same issues apply to Table 3b.

Response: We will adapt this in accordance to the journal’s guidelines / editor’s preference.

D. For each of the factors on the SSS-DSD, one should see a mean score along with a SD. It is really not clear how the authors performed the correlations on p. 10.

Response: The correlations on were calculated as unweighted sumscores. Regarding the means and SDs of the sum scores, see this reviewer’s comment #10 A.

11. Tables 4a and b: The authors report p values but do not provide the actual values from the Mann-Whitney U test. These values must be provided. It is rather bizarre, for example, to see that median values are, say for Atypical appearance/behaviour identical for boys and girls, yet the p value is <.001. Without seeing the U values, one become skeptical. Given the numerous tests conducted, one wonders about using a correction factor.

Response: One should bear in mind that the Mann-Whitney U test does not present the test of two medians (there is no generally accepted statistic that tests the median values between two independent groups), but a test on the ordering (ranking) of scores between two independent groups. Despite the equality of medians, the orderings between groups can still be significantly different when there are more high or extreme values in one of the groups. The ranges provided suggest this is indeed the case. With that information, the Mann-Whitney U values are superfluous in our opinion.

Regarding the second comment (“Given the numerous tests conducted, one wonders about using a correction factor.”): This is correct and the reviewer seems to hint to adjustment for multiple testing (Bonferroni-correction or otherwise). Considering the limited numbers of patients (n=34) and parents (n=81), the multiple outcome measures (five measures in Table 4A and four measures in Table 4B), and the four subgroup analyses, some correction for multiple testing would have been desirable and justified.

12. I prefer delaying comment on the Discussion until a revised ms is prepared.

13. Reference 2 = Reference 17.

Response: We adapted the references.