



**Baishideng
Publishing
Group**

7041 Koll Center Parkway, Suite
160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: office@baishideng.com
https://www.wjgnet.com

Peer-review report(s). Authors must resolve all issues in the manuscript that are raised in the peer-review report(s) and provide point-by-point responses to each of the issues raised in the peer-review report(s):

Reviewer #1:

Scientific Quality: Grade A (Excellent)

Language Quality: Grade A (Priority publishing)

Conclusion: Accept (High priority)

Specific Comments to Authors:

Given the novelty and standardization of using an AI tool to study the relationship between radiographic findings and clinically significant patient perception, this is a study that should be published as soon as possible. The methods are scientifically sound and demonstrate quality retrospective validation of the AI tool. It is critical for academic institutions to investigate and clarify the clinical utility of medical AI devices out to market as per [Chouffani El Fassi, S., Abdullah, A., Fang, Y. et al. Not all AI health tools with regulatory authorization are clinically validated. Nat Med (2024). <https://doi-10.1038/s41591-024-03203-3>]. However, as the manuscript is currently written, it was difficult to understand the value of the findings in clinical context. **More specifically, the introduction does not contain enough references or reflection about the correlation between radiologic measures and the clinical experience of patients with hip dysplasia.** The introduction should answer the following questions and contain the following information: - Has the correlation between manual radiographic measures of hip dysplasia and PROMs already been studied (you only stated that your study was the first to evaluate the relationship between AI measures and PROMs)? **If the present study is the first to investigate the manual measure-PROM relationship, state it clearly.** If not, cite the prior studies. - **Are PROMs and radiologic measures standard of care in hip dysplasia management (endorsed by academic societies for example)?** - **References to prior studies of HIPPO AI clinical validation should be included (specific published papers if there are any)** As for the discussion section, the following questions should be answered: - **Why might the AI (experimental) group have found statistically significant correlations between CCD and iHOT-12, SF-12 versus the manual (control) group?** - How do your findings support or contradict previous studies? - How do your clinical validation findings compare to previous clinical validation studies of HIPPO AI? - How long did it take the MSK radiologist to conduct manual measurements? AI performance should be compared to end-user performance which would be an MSK radiologist, not trained research raters. (reference radiologist timing from other papers if needed) Some other minor points: - Consider moving explanation of the benefits of using an AI tool for radiologic measures and the potential for correlations to PROMs (why the project matters) to the first paragraph of the introduction. - Consider mentioning the benefit of



**Baishideng
Publishing
Group**

7041 Koll Center Parkway, Suite
160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: office@baishideng.com
https://www.wjgnet.com

prospective clinical validation studies (device tested with real patient data after implementation in patient care and/or data collected after study begins) and randomized controlled trials (experimental group that uses device and control group that does not use devices are compared after randomized assignment) as future projects in the discussion section.

Reviewer #2:

Scientific Quality: Grade B (Very good)

Language Quality: Grade B (Minor language polishing)

Conclusion: Minor revision

Specific Comments to Authors:

While the abstract provides a good summary, consider adding more detailed explanations or interpretations of the weak correlations found. Clarify any potential biases or limitations in the study, such as the predominance of female patients or the age range, and how these factors might impact the generalizability of the results. Discuss the practical application of AI in clinical settings, particularly how it might complement but not replace the insights gained from PROMs.

Reviewer 1: We greatly appreciate your feedback and excellent comments. Please see responses and changes below.

-Has the correlation between manual radiographic measures of hip dysplasia and PROMs already been studied (you only stated that your study was the first to evaluate the relationship between AI measures and PROMs)? If the present study is the first to investigate the manual measure-PROM relationship, state it clearly. If not, cite the prior studies.

Response: We clarified that this manuscript is the first to examine the PROM and AI radiographic measurements, and that one prior study has examined manual measurements with PROMs in HD patients. "One prior study evaluated the by Takegami et al evaluated the relationship between manual individual radiographic parameters with the patient-reported outcome measurements in Japanese patients¹⁸. However, it is not yet known how these standardized deep-learning software generated



**Baishideng
Publishing
Group**

7041 Koll Center Parkway, Suite
160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: office@baishideng.com
https://www.wjgnet.com

measurements obtained in the US population correlate with their PROMs data. Additionally, it is not known if a validated AI tool can assist in predicting PROMs data and providing comprehensive evaluation for HD patients.”

AND

“This is the first study to evaluate manual and AI measures of radiographs in patients with hip dysplasia and associate radiographic findings with preoperative PROMs data.”

-References to prior studies of HIPPO AI clinical validation should be included (specific published papers if there are any)

Response: A reference and clarification was added to the end of this sentence “To that end, AP radiographic measurements are auto-evaluated by HIPPO software, which is a validated AI hip measurement tool validated in a European study and Conformite Europeenne (CE) certified [ImageBiopsy Lab Inc. (Vienna, Austria)].¹⁷”

-Are PROMs and radiologic measures standard of care in hip dysplasia management (endorsed by academic societies for example)?

Response: We added the following sentences to the introduction “Hip radiographs are the current gold standard for the initial screening and assessment of HD.⁶” and clarified the current perception of PROMs and need for additional evaluation “Despite their common use in the clinical evaluation of patients with HD and pain, the International Hip-related Pain Research Network meeting in 2018 ruled that more studies are needed to further evaluate the usefulness of PROMS.”

Also, additional references were added after the sentence “PROMs have become increasingly important in evaluating indications for treatment and prognosis for HD patients”

- Why might the AI (experimental) group have found statistically significant correlations between CCD and iHOT-12, SF-12 versus the manual (control) group? **Response:** We added the following sentences “While the exact reason for this significance is not known, the authors hypothesize that the difficulty of



measuring CCD among manual readers compared to standardized AI tool introduced sufficient variation to prevent an observed association.³¹ These findings further highlight the importance of standardization in measurement and interpretation of radiographic measurements.”

- How do your clinical validation findings compare to previous clinical validation studies of HIPPO AI?

Response: Only one previous validation study was performed which showed moderate to strong correlation with manual measurements. The following sentences were included as additional evaluation:

“The authors do not believe that these weak correlations are due to inaccuracies in the AI measurement tool, which was previously validated by Archer et al revealing moderate to strong associations with trained manual readers.¹⁷ Additionally, the vast majority of observed correlations were nonsignificant and contained similar results to the manual readers, with exception of CCA angle and certain PROMs on AI reads, thus suggesting a similar radiographic accuracy between groups as described previously.”

- How do your findings support or contradict previous studies?

Response: There was one previous study that was added in these sentences “The findings of this study differ from those of Takegami et al., where the LCEA angle in 108 Japanese HD patients was independently associated with the Japanese Orthopaedic Association's hip disease questionnaire. However, the PROMs examined in our study were different and applied to a heterogeneous U.S. population, limiting direct comparison.”

-How long did it take the MSK radiologist to conduct manual measurements?

Response: The following was added “and trained radiologists on average require 83 seconds per AP hip radiograph”

- Consider moving explanation of the benefits of using an AI tool for radiologic measures and the potential for correlations to PROMs (why the project matters) to the first paragraph of the introduction.

Response: We have added additional paragraphs per your excellent suggestions above, and find the current flow to be easier to follow.



-Consider mentioning the benefit of prospective clinical validation studies (device tested with real patient data after implementation in patient care and/or data collected after study begins) and randomized controlled trials (experimental group that uses device and control group that does not use devices are compared after randomized assignment) as future projects in the discussion section.

Response: We added the following sentences for future directions:

“Future studies should also incorporate prospective clinical validation studies to assess AI tools against traditional radiographic measurements, post-implementation in patient care settings. Additionally, randomized controlled trials comparing patient outcomes using AI-derived data with those using manual radiographic assessments are critical to establish the effectiveness of AI in clinical decision-making for hip dysplasia”

Reviewer #2:

-While the abstract provides a good summary, consider adding more detailed explanations or interpretations of the weak correlations found.

Response: Thank you for your comments. We have attempted to expand on the interpretation of the correlations found with the following paragraphs “While the exact reason for this significance is not known, the authors hypothesize that the difficulty of measuring CCD among manual readers compared to standardized AI tool introduced sufficient variation to prevent an observed association.³¹ These findings further highlight the importance of standardization in assessment and interpretation of radiographic measurements. The findings of this study differ from those of Takegami et al., where the LCEA angle in 108 Japanese HD patients was independently associated with the Japanese Orthopaedic Association's hip



**Baishideng
Publishing
Group**

7041 Koll Center Parkway, Suite
160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: office@baishideng.com
https://www.wjgnet.com

disease questionnaire. However, the PROMs examined in our study were different and applied to a heterogeneous U.S. population, limiting direct comparison. Despite the potential for AI to streamline clinical workflow, our study highlights the difficulty and current unfeasibility of correlating radiographic findings with patient-centric outcomes such as PROMs. Although HIPPO is efficient at measuring, it may require more training to recognize patterns that better match patients experience. This highlights an area where AI can develop to become more clinically meaningful.”

-Clarify any potential biases or limitations in the study, such as the predominance of female patients or the age range, and how these factors might impact the generalizability of the results.

Response: Thank you for you comment, we have created a dedicated limitations paragraph to address this

“Our study has several limitations. The gender distribution in our study was predominantly female, reflecting the higher incidence of HD in women³². This distribution may influence the correlations observed and thus may not be generalizable to a male population. Additionally, most participants were middle-aged adults, so our results might not reflect the bone density and joint health variations found in older patients, and thus may affect generalizability of this study³³. Finally, the manual measurements, while performed by medical students under the supervision of an MSK radiologist, are not immune to human error. Anatomical variability might have led to inaccuracies; however, extensive training aimed to mitigate such errors, and their impact on the study's validity is considered minimal.”



**Baishideng
Publishing
Group**

7041 Koll Center Parkway, Suite
160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: office@baishideng.com
https://www.wjgnet.com

-Discuss the practical application of AI in clinical settings, particularly how it might complement but not replace the insights gained from PROMs

Response: The following paragraph was added when discussing the clinical application of AI: “While AI can rapidly provide quantitative data valuable for initial screenings and monitoring disease progression, it should complement—not replace—PROMs, which encapsulate the patient’s subjective experience and the functional impact of the disease. PROMs remain essential for capturing the holistic impact on quality of life, guiding more personalized treatment approaches.”