Full title: Challenges in Clinical Diagnosis of Clade I Mpox: Highlighting the Need for Enhanced Diagnostic Approaches

Short title: Challenges in Clinical Diagnosis of Clade I Mpox

Authors: Josephine Bourner¹, Esteban Garcia¹, Festus Mbrenga², Yap Boum II², Amy Paterson¹, Benjamin Jones¹, Piero Olliaro¹, Emmanuel Nakouné², Amanda Rojek¹

Affiliations:

¹ ISARIC, Pandemic Sciences Institute, University of Oxford, Oxford, UK

² Institut Pasteur de Bangui, Bangui, Central African Republic

Corresponding author: josephine.bourner@ndm.ox.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

**Background:** Due to limited diagnostic capacity and availability of point-of-care tests, diagnosis of Clade I mpox in the regions most affected by the disease is usually on clinical grounds and may be complicated due to the similarity between mpox and varicella (chickenpox) lesions. Clinical assessment of lesions is also used for determining clinical progress and has been used to assess patient outcomes in clinical trials. However, there has been no investigation into whether clinicians can (i) identify Clade I mpox compared to other viral lesions (ii) differentiate between Clade I mpox lesion stages.

**Methodology/Principle findings:** The objective of this study was to evaluate inter-rater reliability and agreement between clinicians assessing lesions in patients with Clade I mpox. We presented clinicians with 17 images of Clade I mpox or varicella and asked them to independently indicate the most likely diagnosis – mpox or varicella – and to categorise the lesions according to their stage. When selecting the most likely diagnosis, accuracy varied across all images, the inter-rater reliability was poor ($\kappa = 0.223; z = 10.1$) and agreement was moderate ($P_o = 68\%$). When categorising lesions according to their type, if a single lesion type was present in the image, inter-rater reliability was moderate ($\kappa = 0.671, z = 40.6$) and agreement was good ($P_o = 78\%$), but when multiple lesion types were shown in an image, both inter-rater reliability ($\kappa = 0.153, z = 10.5$) and agreement ($P_o = 29\%$) decreased substantially.

**Conclusions:** This study demonstrates that there are presently limitations in using clinical assessment to diagnose Clade I mpox and evaluate lesion stage and treatment outcomes, which have an impact on clinical practice, public health and clinical trials. More robust indicators and tools are required to inform clinical, public-health, and research priorities, but these must be implementable in countries affected by mpox.
Author summary

Mpox is a zoonotic illness caused by the monkeypox virus (MPXV), for which there are two distinct sub-clades. Clade I is typically found in central Africa and is associated with worse patient outcomes than Clade II. Diagnosis of mpox is most commonly performed using PCR, but in settings with limited laboratory capacity diagnosis is usually performed on clinical grounds taking lesion presentation in to account. Lesion presentation is also used to assess patient outcomes in both clinical and research settings. However, there has been no investigation into whether clinicians can (i) identify Clade I mpox compared to other viral lesions (ii) differentiate between Clade I mpox lesion stages, which has important implications for clinical practice, research and public health. Our study, which presented 16 clinicians with 17 sets of images of Clade I mpox or varicella and asked them to i) provide the most likely diagnosis and ii) categorise the lesions in to their stages, demonstrates that there are presently limitations in using clinical assessment to diagnose Clade I mpox and evaluate lesion stage and treatment outcomes. Alternative methods and tools are therefore required that can be easily implemented in affected countries.
Introduction

Mpox is a zoonotic illness caused by the monkeypox virus (MPXV), for which there are two distinct genetic sub-types, referred to as Clades. (1) Clade I mpox is found primarily in central Africa and is associated with worse patient outcomes than Clade II, which is further divided into Clade IIa (historically reported in west Africa) and Clade IIb, which caused a Public Health Emergency of International Concern (PHEIC) in 2022. (1)

There is a growing epidemic of Clade I mpox in the Democratic Republic of the Congo (DRC) – with a doubling in the number of cases over last year’s. (2) There are several concerning features to this outbreak, including an expansion in geographical areas affected, new introduction to dense urban populations (including Kinshasa, the capital city of DRC), and the first descriptions of sexual transmission of this clade. Both the World Health Organization (WHO) and European Centre for Disease Prevention and Control risk assessments (ECDC) highlight that improving awareness and support for clinicians to diagnose cases is a key response priority. (2, 3) While Clade II mpox is rarely fatal – 156 death out of over 91,000 cases reported outside Africa since May 2002 –, mortality associated with Clade I mpox is 1-12% (1) and there have been 581 deaths reported out of over 12,500 cases in DRC since February 2023. (2)

However, diagnosis of mpox is difficult. Confirmation of mpox diagnosis is primarily by PCR, (4, 5) but, concerningly, new evidence suggests that the Clade I-specific RT-PCR test recommended by the US CDC is impacted due to genetic mutations in the virus causing this outbreak. (6) There are no point-of-care or rapid diagnostic tests alternatives at present. Laboratory diagnosis is impacted by operational challenges and limited laboratory capacity –in the present DRC epidemic, only 9% of suspected mpox cases have been tested by PCR. (7)

Therefore, diagnosis of Clade I mpox in the regions most affected is usually on clinical grounds. This is reported as difficult due to the similarity between mpox and varicella (chickenpox) lesions. Clinical
assessment of lesions is also used for determining clinical progress. Whether a lesion is ‘active’, ‘inactive’ or ‘resolved’ is used for determination of infectivity and infection-prevention and control requirements, decisions around whether to offer a patient access to potential treatments, and declarations of cure. The assessment of lesions has in addition become the focal area of evaluation in clinical trials. Several trials for Clade I and Clade II mpox use time to lesion resolution as the primary endpoint to indicate treatment success. (8-12)

Despite the myriad uses of clinical diagnosis of Clade I mpox, there has been no investigation into whether clinicians can (i) identify Clade I mpox compared to other viral lesions (ii) differentiate between Clade I mpox lesion stages – although some suspected limitations of this endpoint have already been identified. (13) A previous exercise undertaken by these authors on Clade IIb lesions demonstrated only moderate agreement among different assessors. (14)

This study evaluates agreement between clinicians on a differential diagnosis between Clade I mpox and varicella, and classification of lesion stages. Whether mpox Clade I lesions can be reliably classified on clinical grounds has important implications both in the low-resource settings where the virus circulates and in the event of spread to historically non-endemic regions – for decisions on treatment, public health control, and for the robustness of clinical research.

Methods

The objective of this study was to evaluate inter-rater reliability and agreement between clinicians assessing lesions in patients with Clade I mpox. Two focal areas of agreement and reliability were evaluated: 1) differential diagnosis between Clade I mpox and varicella based on lesion presentation; 2) categorisation of lesion stages. The design, conduct and results of this study are reported according to the Guidelines for Reporting Reliability and Agreement Studies (GRAAS). (15) Ethical approval for this study was obtained from the University of Oxford Medical Science Interdivisional Research Ethics Committee (R84355/RE001).
Participants

The participants in this study were clinicians who had experience treating and managing patients with Clade I mpox.

Data collection

Using the RedCap survey tool (16, 17), participants were shown 17 sets of images of patients with either Clade I mpox or varicella lesions confirmed by PCR. The anonymised images selected for use in the survey were derived from the Institut Pasteur de Bangui clinical image library and showed lesions of different stages and located on different areas of the body.

Based on the lesion presentation in the images, the participants were asked to independently indicate the most likely diagnosis – mpox or varicella – and to categorise the lesions as being either active, scabbed or resolved, or they had the option to state they were unable to classify the lesions.

Before starting the questionnaire, participants were shown the World Health Organisation’s working definition of each lesion stage. (18)

The survey also captured each clinician’s country of practice, average number of patients they manage with Clade I mpox in a year, and their self-rated confidence evaluating mpox lesions.

Confidence evaluating mpox lesions was assessed on an ordinal scale from one to ten, with one representing no confidence and ten representing complete confidence.

The questionnaire used in this study can be found in S1 – Appendix 1 and was provided in both English and French.

Data analysis

All analyses completed in this study were conducted with the ‘irr’ package in R Statistical Software (v4.3.2) (19) and validated using ‘statsmodels’ package in Python (v3.11.5) by two analysts.

Accuracy identifying mpox or varicella was evaluated using the percentage of raters selecting the correct diagnosis for each image. Spearman's rank correlation coefficient (Spearman’s $\rho$) was
employed to assess the relationship between experience—assessed by self-reported expertise and
the number of patients managed—and each evaluator’s accuracy percentage. Spearman’s method,
chosen for its suitability with small samples and ordinal data without assuming a normal distribution,
necessitates a cautious interpretation of results due to the reduced statistical power and potential
influence of noise and outliers inherent to limited datasets.

Inter-rater reliability was evaluated using Fleiss’ kappa coefficient ($\kappa$) and inter-rater agreement was
evaluated using proportion of partial agreement ($P_o$), and proportion of exact agreement ($P_{oe}$). Partial
agreement refers to the agreement between any one of the participants’ selections in a multiple-
choice answer and exact agreement refers to a complete match between participants’ selections in a
multiple-choice answer.

Questions related to lesion classification assessed only the images in which the patient had a
confirmed diagnosis of mpox. The first analysis evaluates responses to the images where only a
single or homogenous lesions were present and for which the $\kappa$ and $P_{oe}$ are reported. The second
analysis evaluates responses to the images where multiple types of lesions (combinations of active,
scabbed and resolved) are present, for which the $\kappa$, the $P_o$ the $P_{oe}$ are reported.

Results

The questionnaire was sent to 38 potential participants in total, of whom 17 accessed the
questionnaire and completed the eligibility check. One respondent was not eligible to participate as
they were not directly involved in the clinical management of patients with Clade I mpox.

All 16 participants who started the questionnaire completed it in full. A summary of the participants’
country of work, confidence assessing an mpox lesion and number of mpox patients they have
personally managed is summarised in Table 1.
Table 1 – Description of study participants

<table>
<thead>
<tr>
<th>Country of work, n (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Democratic Republic of Congo</td>
<td>8 (50%)</td>
</tr>
<tr>
<td>Central African Republic</td>
<td>2 (13%)</td>
</tr>
<tr>
<td>France</td>
<td>2 (13%)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>2 (13%)</td>
</tr>
<tr>
<td>Nigeria</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>Belgium</td>
<td>1 (6%)</td>
</tr>
</tbody>
</table>

Number of mpox patients the respondents has personally managed, n (%):

<table>
<thead>
<tr>
<th>Number of mpox patients</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;5</td>
<td>0</td>
</tr>
<tr>
<td>5-10</td>
<td>2 (13%)</td>
</tr>
<tr>
<td>10-20</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>20-50</td>
<td>5 (31%)</td>
</tr>
<tr>
<td>&gt;50</td>
<td>8 (50%)</td>
</tr>
</tbody>
</table>

Confidence assessing mpox lesions, median (IQR):

<table>
<thead>
<tr>
<th>Confidence score</th>
<th>Median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>7.5 - 8</td>
</tr>
</tbody>
</table>

Table 2 – Correlation between raters’ diagnostic accuracy, self-rated confidence and experience using Spearman’s correlation (ρ)

<table>
<thead>
<tr>
<th>Image</th>
<th>Self-rated confidence</th>
<th>Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>Self-rated confidence</td>
<td>0.3</td>
<td>1</td>
</tr>
<tr>
<td>Experience</td>
<td>0.168</td>
<td>-0.042</td>
</tr>
</tbody>
</table>

Diagnosis

Accuracy selecting the most likely diagnosis between mpox and varicella based on lesion presentation varied across all images, with between 25% and 100% of the respondents correctly identifying the disease and a median = 75% (Q1: 62.5%, Q3: 93.75%). No correlation was detected between accuracy and either self-rated confidence (Spearman’s ρ = 0.3) or experience (Spearman’s ρ = 0.17) assessing lesions (Table 2).
When asked to select the most likely diagnosis of the patient between varicella and mpox based on the lesion presentation shown in the images, the inter-rater reliability was poor ($\kappa = 0.223; z = 10.1$) and agreement was moderate ($P_o = 68\%$). (Fig 1)

**Fig 1. Dendrogram of the suspected diagnosis associated with the images based on lesion presentation**

In addition to the reported results, in the supplementary appendices we show that there appears to be improved diagnostic accuracy if a single lesion type is present in the image, compared to images in which lesions appeared to be in multiple concurrent stages of evolution (S2 – Appendix 2). We did not observe any relationship between diagnosis and accuracy (S2 – Appendix 2). However, our sample size is too small to draw any generalisable conclusions.

**Lesion classification**

When a single lesion type was present in the image, inter-rater reliability was moderate ($\kappa = 0.671, z = 40.6$) and agreement was good ($P_o = 78\%$) (Table 3, Fig 2).

**Fig 2. Dendrogram of the respondents’ lesion assessments for which a single lesion type was present in the image**

However, when multiple lesion types were shown in an image, both inter-rater reliability ($\kappa = 0.153, z = 10.5$) and agreement ($P_o = 29\%$) decreased substantially. (Table 3)

Where partial agreement was assessed, inter-rater reliability was moderate ($\kappa = 0.447, z = 25.9$) and agreement was good ($P_o = 73\%$). (Fig 3) (Table 2)

**Fig 3. Dendrogram of the respondents’ lesion assessments for which multiple lesion types were present in the image**
Table 3 – Summary of inter-rater reliability and agreement results

<table>
<thead>
<tr>
<th>Diagnosis of patients in images</th>
<th>Response selection formats</th>
<th>Lesion types present in image</th>
<th>Fleiss’ kappa</th>
<th>Z score</th>
<th>Percentage of agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpx &amp; Varicella</td>
<td>Single choice</td>
<td>Single types and multiple types</td>
<td>0.223</td>
<td>10.1</td>
<td>Percentage of exact agreement 67.745</td>
</tr>
<tr>
<td>Mpx</td>
<td>Single choice</td>
<td>Single types</td>
<td>0.671</td>
<td>40.6</td>
<td>Percentage of exact agreement 78.14</td>
</tr>
<tr>
<td>Mpx</td>
<td>Multiple choice</td>
<td>Multiple types</td>
<td>0.447</td>
<td>25.9</td>
<td>Percentage of partial agreement 73.18</td>
</tr>
<tr>
<td>Mpx</td>
<td>Multiple choice</td>
<td>Multiple types</td>
<td>0.153</td>
<td>10.5</td>
<td>Percentage of exact agreement 29.04</td>
</tr>
</tbody>
</table>

Discussion

This study demonstrates that there are presently limitations in using clinical assessment to diagnose Clade I mpox and evaluate lesion stage and treatment outcomes. We found moderate accuracy, poor reliability, and moderate agreement among clinicians deciding between mpox and varicella diagnosis based on lesion presentation. In settings where laboratory diagnosis is not available, or delayed, there is a risk that patients could be managed according to an incorrect care pathway (e.g. provision of an incorrect antiviral). Misclassification of cases can have implications for public health activities as varied as disease surveillance, allocation of vaccinations, drug procurement, and reporting of notifiable diseases obligations. This has consequences also for the evaluation of treatment effects, both in clinical practice and research, with incorrect or inconsistent classification of outcomes.
We found that for classification of mpox lesion status there was moderate reliability and good agreement if a single lesion type is present – although this may not be sufficient to state that there is good agreement beyond chance (20, 21) – but it appears to be most challenging to obtain consistent assessments when multiple lesion types are present. As lesions are unlikely to follow the same evolution pattern over time, the poor reliability and agreement between clinicians when multiple lesion types are present creates a substantial challenge. This includes clinical trials, which rely on a single overall assessment of a patient’s lesion presentation at a specific timepoint to act as an indicator of a patient’s overall outcome and response to treatment.

These findings occurred despite the self-rated confidence among the participants being high (IQR: 7.5-8). We found no correlation between confidence and experience defined according to the number of patients with mpox the respondent had managed (Fig 4). Due to the sporadic and widespread reporting of Clade I mpox across a large geographic area, it is possible that confidence, and agreement among real-world clinicians, who may only occasionally manage a patient with mpox or varicella, may be lower than reported in this study where we sampled a relatively expert group.

Fig 4. Scatterplot of respondents’ self-rated confidence plotted against experience (measured according to the number of patients managed)

A parallel study found higher agreement and reliability among clinicians assessing Clade IIb mpox lesions. (14) This may be due to differences in lesion presentation between the two diseases, with Clade IIb lesions often being more localised and fewer than Clade I lesions. The surveys were not combined due to the significant differences in clinical presentation between the clades, and that clinical experience spanning both clades is limited to a very small number of highly-expert clinicians.

There is a continued requirement for resources (such as the WHO lesion assessment atlas) to assist clinicians and clinical triallists working in low-resource environments to improve their assessments, alongside support for improved diagnostic capacity. There is an urgent need to work to improve and find other correlates of disease progression (e.g. how much PCR status reflects ongoing infectivity).
Considering the lack of correlation between clinician experience and their ability to diagnose Clade I Mpox, this study underscores the potential of Artificial Intelligence (AI) and Machine Learning (ML) in improving Mpox diagnosis, especially in resource-constrained settings where access to doctors may be limited and initial assessments are often conducted by nurses or community health workers. The application of AI and ML technologies, capable of capturing and analysing images, presents a significant opportunity to enhance the diagnostic accuracy of clinicians and nurses in areas lacking PCR testing facilities. This advancement is particularly crucial before the availability of rapid diagnostic tests. Although these technologies have demonstrated their efficacy during the pandemic, specifically with the Clade IIb variant of Mpox, data on their effectiveness with Clade I remain scarce. By leveraging AI, we can aim to bridge the diagnostic gap, ensuring timely and accurate detection of Mpox, even in the most challenging environments, thereby improving patient outcomes and controlling the spread of the disease more effectively. (22, 23)

**Limitations**

This study included a relatively low number of participants, because few clinicians worldwide have current or recent experience managing patients with Clade I Mpox. The response rate was also challenged by difficulty identifying and reaching clinicians who work in remote areas with limited internet connection. Assessments made by visual inspection of images may not be reflective of all the information used to make assessments in a clinical scenario (such as patient’s descriptions of evolution of lesions, associated characteristics such as pain). We used reliability and agreement for assessment of mpox lesions (compared to accuracy) because there is no ‘gold standard’ classification to which to compare. We did not provide images where there was mpox and varicella coinfection.

**Conclusion**

It is difficult for experienced clinicians to distinguish clade I Mpox from varicella, and to reliably assess disease stage for clade I Mpox. More robust indicators and tools are required to inform
clinical, public-health, and research priorities, but these must be implementable in countries affected by Mpxx.

Acknowledgements

This work was supported by the UK Foreign, Commonwealth and Development Office and Wellcome [215091/Z/18/Z] and the Bill & Melinda Gates Foundation [OPP1209135]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References


Supporting information

S1 Appendix 1 – Questionnaire

S2 Appendix 2 – Summary of question characteristics and accuracy score

S3 Appendix 3 – Full dataset