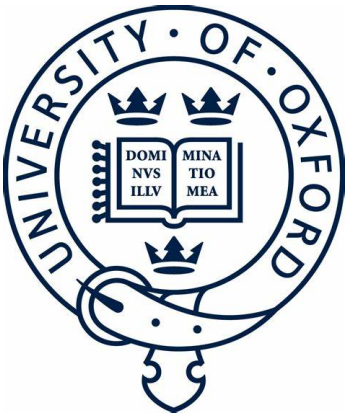
Improving efficacy endpoints in clinical trials for outbreak-prone infectious diseases

Josephine Bourner



A thesis submitted for the degree of

Doctor of Philosophy (D.Phil)

Green Templeton College

Contents

List of Figures

Abstract

Background: Selecting clinically meaningful and measurable efficacy endpoints for trials of outbreak-prone infections presents multiple challenges, including poor clinical characterisation, heterogeneity in disease presentation, and the absence of standardised methodologies to evaluate patient outcomes.

Methods: This thesis addresses these issues through three case studies: Lassa fever, mpox, and bubonic plague. A mixed-methods approach was used, including systematic reviews to examine outcome selection challenges, validity and reliability studies to evaluate measurement tools, and a retrospective cohort study. In the absence of sufficient data to inform endpoint selection for Lassa fever, stakeholder consultations were conducted. Additionally, data from an ongoing randomised controlled trial were used to assess the risks of including problematic components in endpoint definitions.

Results: Two primary challenges were identified: (1) limited robust evidence in the literature characterising disease outcomes and key patient events, and (2) the widespread use of unvalidated or inadequate measurement tools in current trials. These issues undermine the ability to evaluate interventions effectively in the context of outbreak-prone infections.

Conclusion: This work made direct contributions to clinical trials for Lassa fever and bubonic plague. For Lassa fever, the primary endpoint developed in this project has been adopted in the INTEGRATE trial. For plague, the removal of bubo measurement from the composite endpoint addressed a major source of bias, improving the trial's validity. In addition to these contributions, the thesis proposes a structured framework for endpoint

selection, providing both immediate and long-term value to researchers working on clinical trials for outbreak-prone infectious diseases.

Nomenclature

(AI) Artificial Intelligence

(CAR) Central African Republic

(CI) Confidence Interval

(ClinROs) Clinician-Reported Outcome measures

(CRF) Case Report Form

(Ct) Cycle threshold

(CTU) Clinical Trials Unit

(D) Day

(EAP) Expanded Access Programme

(ECDC) European Centre for Disease Control

(HIV) Human Immunodeficiency Virus

(IPB) Institut Pasteur de Bangui

(IPM) Institut Pasteur de Madagascar

(IQR) Inter-quartile range

(ISARIC) International Severe Acute Respiratory and Emerging Infections Consortium

(MedDRA) Medical Dictionary for Regulatory Activities

(ML) Machine learning

(MPXV) Mpox virus

(PCR) Polymerase Chain Reaction

(POCUS) Point-of-care ultrasound

(PT) Preferred Term

(QA) Quality assurance

(R&D) Research and Development

(RECIST) Response Evaluation Criteria in Solid Tumors

(SAE) Serious Adverse Event

(SOC) System Organ Class

(TEC) Téchnicien d'étude clinique; clinical study technician

(VZV) Varicella zoster virus

Acknowledgements

During the course of my DPhil I have had the privilege to meet so many wonderful people who I am truly grateful to know, particularly those who started as colleagues and who I can now call friends. They are all too numerous to list here, but (if they were to ever read this, which as a friend, I don't recommend) I would want them to know the impact they have had not only on my academic journey but by broadening my perspectives and bringing light to the dark, endless, imposter syndrome-y tunnel of fear, frustration and self-doubt that is academia.

My parents are the first people who I want to thank for supporting every ambition I have, regardless of how questionable it might be. They are the reason my deluded self-belief has got me this far. I have struggled for at least two days to express the depth of my gratitude and love for Ashleigh Cheyne and Elise Pesonel. It's unsummarisable, so I hope they know what they mean to me; I'm so happy we found each other. I would also like to thank my sister, Hayley Probert, for her unwavering support, encouragement and grounding. This project is also dedicated to my closest friends Kitty Chan, Gita Parmer, Lakshmi Manoharan, and Asia Mace, all of whom are responsible for any success I achieve in my life.

I would also like to thank my colleagues at the Pandemic Sciences Institute for their support, collaboration and kindness throughout my DPhil, in particular Emmanuelle Denis, Muriel Lunn, Nina Jamieson and Wajeeha Malik. A special thanks goes to the wonderful colleagues who I've had the pleasure to collaborate with through my projects, including Ravaka Randriamparany, Tsinjo Fehizoro, Théodora Mayouya-Gamana, Mihaja Rabera-

hona, Rindra Randremanana, Lovarivelo Andriamarohasina, Festus Mbrenga, Yap Boum II, Emmanuel Nakouné, and to all of those who have given time to me.

I would like to express my sincere gratitude to my supervisors, Piero Olliaro, Amanda Rojek and Alex Salam, for their invaluable guidance, support, and encouragement throughout the course of my DPhil. Their expertise and insight have been instrumental in shaping this thesis, and I am truly thankful for the opportunity to work under their supervision.

Preface

My PhD started in January 2021, almost a year after the first cases of COVID-19 were detected and the world plunged in to a series of lockdowns that may have extended well in to my PhD (although their memory has since been erased in the haze of the pandemic). COVID-19 could have presented the perfect scenario for an eager early-career pandemic researcher working on trials for emerging infections, but as the pandemic wore on my focus remained with the diseases that have lived among humans for much longer — thousands of years longer for some — and for which no meaningful therapeutic advances have been made.

Having previously worked as a Trial Manager in an oncology Clinical Trials Unit (CTU), I was aware that there would be some methodological, operational and regulatory differences between the machine that is cancer research and trials for emerging infections, some of which are probably responsible for the little progress in therapeutic advancement made to date for diseases like plague, mpox and Lassa fever. Indeed, research on emerging infections is far from the industry of cancer research that I had previously been used to, which operated within a vast NHS-supported research infrastructure and was underpinned by long-established methodological frameworks. In this world, the trials could take ten years to complete recruitment — and they sometimes did.

The world of emerging infectious disease research, by comparison, appeared disorganised and chaotic. There was no established network of trial sites, because predicting the location of a disease outbreak is rarely possible, and if any of my new trials took ten years to complete, whatever outbreak being investigated would likely have ended years before. But the biggest challenge appeared to be methodological. Specifically, unlike

trials for therapeutics targeting solid tumours, there was no methodological framework telling researchers what to measure and how to measure it.

As this project will explore, uncertainty about meaningful patient outcomes and how they should be measured is particularly problematic, not only for clinical trials – regardless of the disease being targeted – but for clinical decision-making and public health policy. Select the "wrong" endpoint — either the outcome or the measurement tool being used — and you risk several potentially serious problems in your trial results, including: lacking a sufficient number of outcome events to detect statistically significant differences between treatment arms; misunderstanding the therapeutic effect of the drug; risking measurement bias; and selecting an endpoint that may not be meaningful to clinicians or patients. All of which could have a knock-on effect when developing national treatment algorithms and when making critical life- or disease-altering treatment decisions for individual patients.

The risk of not identifying the most meaningful, measurable patient outcome to evaluate in a trial seemed like a pretty fundamental flaw, particularly because its repercussions have the potential to extend beyond the bounds of the trial itself. The first clinical trial for an emerging infection that I worked on was a randomised controlled trial comparing two treatment regimens for bubonic plague. [1] [2] This would be the first clinical trial ever conducted for plague that is powered to detect significant treatment effects between two groups; it could therefore be the first to generate meaningful clinical data about an effective intervention for plague.

The trial defined a composite primary endpoint evaluating therapeutic response based on five criteria: survival/mortality, absence of fever, a 25% decrease in bubo size, no co-administration of another effective therapy for plague, and no clinical decision to continue treatment. The third component, relating to the reduction in bubo size, however was a source of long-term confusion that would follow me in to this DPhil. Firstly, there is very little data available about what a bubo is, how it evolves over time and its role in bubonic plague. There is in fact no evidence to suggest bubo size has any relationship with the overall clinical status of a patient. Secondly, it was unclear why a reduction of 25% was significant in the course of illness over any other level of reduction. Finally we were measuring bubo size with a digital caliper and logic alone told me this would probably

generate a decent amount of measurement error. I was confused both about what the evaluation of this component meant in the context of the trial and how it came to be part of the endpoint.

While these concerns were eventually justified (as we will see in later chapters), I now realise the naivety of the judgement I passed about a field of research that is so utterly complex and frustrating (which is evident having chosen to be here now for five years), in which the effects of historic under-investment in research for the diseases that pose the greatest global health security threat are being felt by researchers today. Evaluating patient outcomes and selecting endpoints in trials for outbreak-prone infections isn't straightforward and there is no single solution to progressing research for therapeutics in this field, but the point remains that patients need treatment, so what choice do the decision-makers have but to do the best they can with the little data that is available?

My DPhil is potentially an exploration and attempt to make sense of an inherently chaotic research arena and manoeuvre it to conform to some kind of organised framework. The problem with this desire however is that the diseases that this project deals with are not like other diseases. They are neither consistent nor predictable, so they can't be treated like other areas of research that has historically benefited from regular methodological and clinical research investment – although parallels with better-studied conditions such as sepsis or ARDS may offer a valuable starting point for adapting frameworks and approaches. I believe however that the small efforts – often addressing the hard (and sometimes boring) topics – that are made quietly in the background can make meaningful contributions to optimise the evaluations we make of interventions and ensure our findings are reasonable and relevant. This, I think, is the contribution of my DPhil to this research field.

Introduction

Clinical trials function as the primary gateway for new interventions to generate the required evidence for regulatory licensing, followed by their subsequent integration into clinical care. Their design should follow robust, quality-driven and evidence-based scientific principles that support clinical, regulatory and policy decision-making based on the safety and effectiveness of an intervention. [3]

Understanding the safety and effectiveness of an intervention is achieved in a clinical trial by defining specific, measurable endpoints that provide clear criteria to assess whether a treatment has a meaningful impact on patient health, typically focusing on how a patient feels, functions, or survives. [4] Patients may however experience a diverse range of clinical features over the course of an illness, so identifying the important clinical features and outcomes that could be targeted for evaluation is vital. The outcomes selected for evaluation in a trial should however represent the clinical features of a disease that have the potential to be modified by the intervention and are considered to be meaningful both from a clinical and regulatory standpoint, as well as from the perspective of the patient.

Endpoints can be classified into different types, such as primary endpoints, which address the main question of the trial, and secondary or exploratory endpoints, which gather additional information on the effects of an intervention. Typically, the most important outcome(s) that establish the effectiveness of the intervention are evaluated as the "primary" endpoint – how impactful the intervention is on how a patient feels, functions or survives. The remaining outcomes of interest, including those related to safety, are then evaluated as "secondary" or "exploratory" endpoints. Endpoint selection – particularly that of the primary endpoint – forms a critical part of the trial design process, as it directly

influences trial design, sample size estimation, and overall interpretation of the effect of the intervention.

Specifically, an endpoint and the rationale behind the assumptions made to evaluate the endpoint impact the conclusions that can be drawn from the results (**Figure 1**). Poor choice of endpoint risks researchers over- or under-stating the efficacy of a drug – or risks the wrong conclusion being drawn about its effect on the disease. Overstating the efficacy of a drug may lead to individual patient harm and additional costs to the health system; understating efficacy may in turn lead to the abandonment of effective therapeutics having wasted substantial research and development (R&D) investments, resulting in a missed opportunity for people in need.

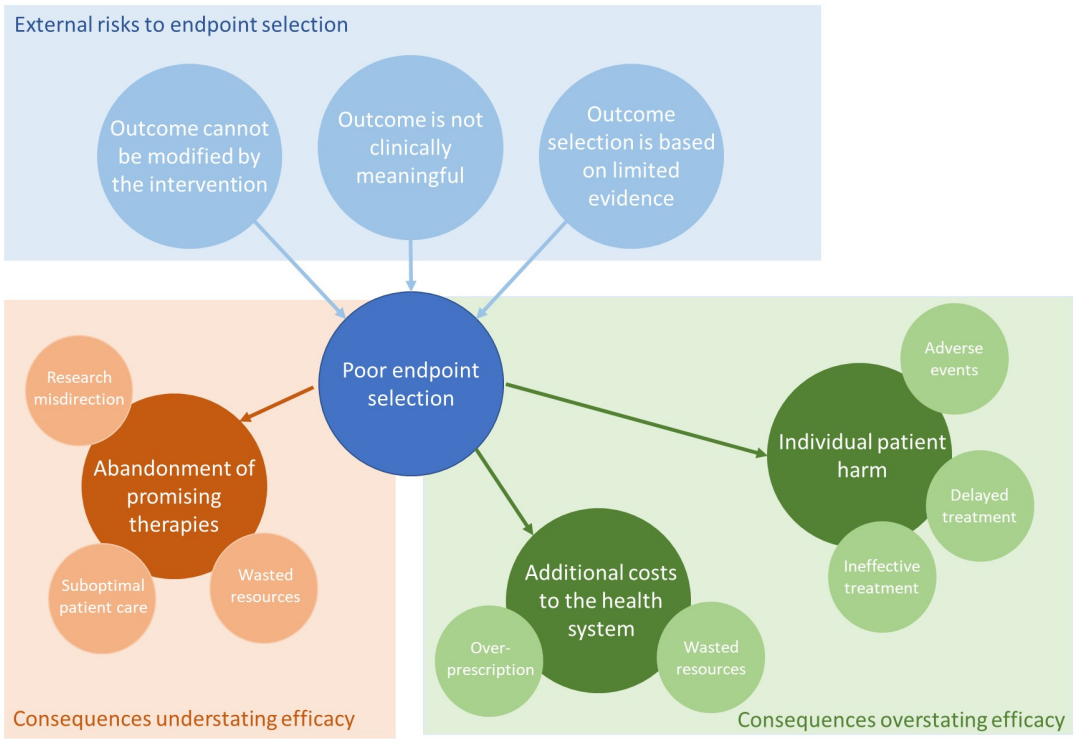


Figure 1: Potential consequences of poor endpoint selection

Ultimately, for a drug to receive regulatory approval for licensing and begin circulating in mainstream clinical care, conclusive evidence must be generated in a trial to demonstrate that it in fact has the effect on the disease that it purports to have. [4] Selecting an endpoint that adequately reflects the important effects of the intervention is therefore vital.

The definition of appropriate endpoints and subsequent generation of robust clinical evidence in a clinical trial is, however, easier for some diseases than it is for others. The limited number of therapeutic interventions that have received regulatory approval for outbreak-prone infections is a symptom of the challenges that are faced in this research area – and in fact where some interventions have been granted approval, this has been achieved via non-conventional, non-clinical regulatory pathways. Of the diseases shown in **Table 1**, which represent outbreak-prone infectious diseases that result in acute illness, no clinical trials have ever been conducted for 11 diseases. In the 17 trials that have taken place to date for the remaining eight diseases, 2778 participants have been enrolled in total, but only one (6%) for orthoebolaviruses (specifically Ebola virus (EBOV)), which enrolled 25% of all included participants, had a conclusive outcome. [5] Yet despite the lack of evidence that has been generated, drugs have been licensed for four outbreak-prone infectious diseases – anthrax, Ebola virus (EBOV), mpox and plague (**Figure 2**).

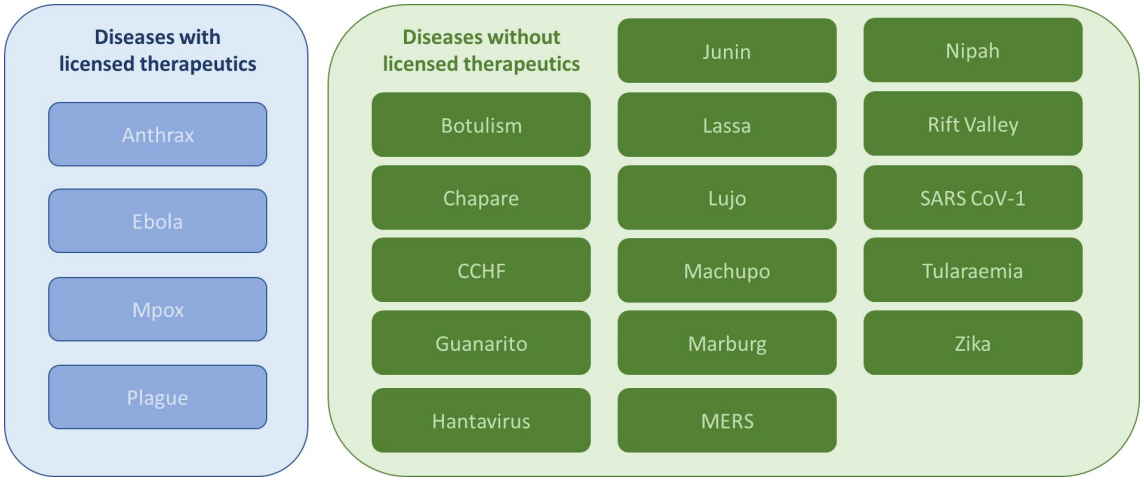


Figure 2: Outbreak-prone infectious diseases with and without licensed therapeutics

While therapeutics to treat Ebola virus (EBOV) have been licensed based on robust trial data, the drugs that have been licensed for the treatment of anthrax, mpox and plague have all been approved either under emergency measures, or based on data collected in small, experimental studies or under the U.S. FDA's "Animal Rule"(which permits the licensing of therapeutics based on animal data alone where clinical trials in humans are neither feasible nor ethical). [6]

FDA Animal Rule

The FDA's Animal Rule [7] provides a regulatory pathway for approving drugs when it is not ethical or feasible to conduct trials in humans, typically in cases involving serious or life-threatening conditions, and the product under development is likely to provide clinical benefit. Under this rule, approval can be based on well-controlled animal studies, provided certain conditions are met. These include a reasonably well-understood mechanism of action, evidence of efficacy in more than one predictive animal species, and animal models that closely replicate the human disease. Although efficacy comes from animal data, safety must still be evaluated in humans. The endpoints used in the animal studies must also clearly relate to meaningful human outcomes, such as survival or prevention of severe disease. This approach allows potentially life-saving treatments to be made available when traditional human trials are not possible.

Even where approvals are not in place for certain drugs shown in Table 1, such as ribavirin for Lassa fever and gentamicin for plague, these drugs are nevertheless recommended in national or international treatment guidelines without efficacy having been substantiated in a clinical trial.

The other 16 diseases shown in **Table 1** have no recommended or licensed treatments.

Table 1: Summary of research investments and regulatory drug approvals for selected NIAID priority pathogens [8]

Family	**Pathogen**	**Recommended therapeutic interventions**	**Trials completed**	**Participants enrolled**	**Licensed drugs**
Arenaviruses	Chapare	-	-	-	-
	Guanarito	-	-	-	-
	Junin	-	-	-	-
	Lassa	Ribavirin [9]	1	441	-
	Lujo	-	-	-	-
	Machupo	-	-	-	-
Bunyaviruses	CCHF	-	2	354	-
	Hantavirus	-	1	38	-
	Rift Valley	-	-	-	-
Coronaviruses	MERS	-	1	95	-
	SARS CoV-1	-	-	-	-
Filoviruses	Ebola virus (EBOV)	Atoltivimab/ maftivimab/ odesivimab; ansuvimab [10]	8	1567	Atoltivimab/ maftivimab/ odesivimab; ansuvimab [11] [12]
	Bombali virus (BOMV)	-	-	-	-
	Bundibugyo virus (BDBV)	-	-	-	-

Table 1: Summary of research investments and regulatory drug approvals for selected NIAID priority pathogens [8]

Family	**Pathogen**	**Recommended therapeutic interventions**	**Trials completed**	**Participants enrolled**	**Licensed drugs**
	Reston virus (RESTV)	-	-	-	-
	Sudan virus (SUDV)	-	-	-	-
	Taï Forest virus (TAFV)	-	-	-	-
	Marburg virus (MARV)	-	-	-	-
Henipaviruses	Nipah	-	1	194	-
Mosquito-borne viruses	Zika	-	-	-	-
Poxviruses	Mpox	Brincidofovir; cidofovir; tecovirimat [13]	1	19	Tecovirimat [14]

Table 1: Summary of research investments and regulatory drug approvals for selected NIAID priority pathogens [8]

Family	**Pathogen**	**Recommended therapeutic interventions**	**Trials completed**	**Participants enrolled**	**Licensed drugs**
Bacteria	Anthrax	Doxycycline; minocycline; ciprofloxacin; levofloxacin; amoxicillin; penicillin [15]	-	-	Obiltoxaximab; ciprofloxacin; doxycycline; levofloxacin; penicillin; raxibacumab [16]
	Botulism	-	-	-	-
	Plague	Ciprofloxacin; levofloxacin; moxifloxacin; gentamicin; streptomycin; doxycycline; chloramphenicol; ofloxacin; gemifloxacin; amikacin; tobramycin; plazomicin; tetracycline; omadacycline [17]	2	70	Ciprofloxacin[1]; doxycycline[1]; levofloxacin[1]; moxifloxacin[1]; streptomycin[1] [18]

Table 1: Summary of research investments and regulatory drug approvals for selected NIAID priority pathogens [8]

Family	**Pathogen**	**Recommended therapeutic interventions**	**Trials completed**	**Participants enrolled**	**Licensed drugs**
	Tularaemia	-	-	-	-

The limited actionable evidence generated for outbreak-prone infections through clinical trials makes it difficult to identify interventions that have a clinically meaningfully impact on how patients feel, function, or survive. Some of these diseases, such as plague, Ebola virus (EBOV), and Lassa fever, affect hundreds or thousands of people annually during periods of "normal" transmission, and even more during periods of large outbreaks. The lack of effective, evidence-based therapies poses a substantial and avoidable risk to large populations in endemic areas. However, given the potential for these diseases to trigger larger outbreaks that may spread across borders and into non-endemic regions, it is possible that the impact of a lack of research into viable therapeutics could be felt much more widely. It is therefore critical that the factors that limit research from generating meaningful outputs are urgently characterised and addressed.

Origins of this project

The origins of this project are two-fold. First, as shown in Table 1 patients seeking treatment for outbreak-prone infections face several challenges: many diseases lack trials evaluating therapeutic interventions, leading to the absence of licensed treatments or the use of drugs without a clinical evidence base. Consequently, more unlicensed drugs have been incorporated in to treatment guidelines than approved drugs, and more drugs have been licensed without robust clinical data than from adequately-powered trials in humans (**Figure 3**). This sequence of events, in the absence of clinical trials, leads to unquantified patient risks.

Figure 3: Consequences of the lack of clinical trials for outbreak-prone infections

Second, it is clear that trials for other diseases can successfully generate robust clinical data supported by standardised methodologies that result in evidence-based licensing decisions. This progress is driven by investment in research methodologies. In trials for solid tumours, for example, endpoint selection is supported by a standardised guideline (Response Evaluation Criteria in Solid Tumors or RECIST) that has been developed to support accurate, consistent evaluation of therapeutic efficacy [19] and which has resulted in regulatory approval of new therapeutics. [20] Methodological research has also led to variations of RECIST, such as mRECIST and iRECIST, tailored to different trial types while preserving consistency in tumor evaluation. [21] [22] Similar regulatory-endorsed efficacy endpoints have been established for conditions like diabetes, HIV, and vaccines. [23] [24] [25]

Outbreak-prone infections

There are a number of terms that can be used to characterise diseases that are transmitted to (and sometimes between) humans as a result of contact with a pathogen and that have the potential to cause sudden increases in disease among a population that are in excess of its typical pattern. In this thesis, they will be referred to as outbreak-prone infectious diseases, but elsewhere some of the same diseases have been grouped under terms such

as high consequence infectious diseases [26], emerging infectious diseases [27] and priority pathogens [28].

The diseases explored in this thesis are a subset of the outbreak-prone infections that are often found on these overlapping lists brought together by a common theme: their history of being difficult to study.

Outbreak-prone infectious diseases present a number of challenges to research. The characteristics that make an emerging infectious disease difficult to study and their impact on the generation of evidence in clinical trials are numerous, complex and inconsistent. A summary of some of these characteristics are explored in greater detail below.

Low and sporadic annual case numbers

Generally, the number of confirmed cases of outbreak-prone infections in defined geographic regions is relatively low during a "normal" transmission season compared to the incidence of other diseases, such as cancers, diarrhoeal diseases and cardiovascular disorders. [29] In the last year that complete data were available, 4060 new cases of cervical cancer (2022) [30] were diagnosed in Madagascar compared to 104 cases of plague (2018). [31]

Small numbers of cases pose a challenge to clinical trials as they limit the number of cases that can be enrolled, risking insufficient data being collected to precisely estimate the effect of the trial intervention. In addition, fluctuations in the annual number of confirmed cases of a disease -– e.g. widely varying numbers of reported cases from year-to-year — create an additional operational complexity for a clinical trial. As trials have a finite amount of funding and receive ethical approval for a defined period of time – which is typically based on an informed estimation of the number of years it would take to reach the target sample size – fluctuating case numbers can impede a) the accurate estimation of the trial duration and b) predicted enrolment, risking early termination due to insufficient funds or low confidence in the trial's ability to generate results.

Annual numbers of plague cases in Madagascar, where the highest global burden can be found [31], can vary substantially from year-to-year. In 2017, for example, 661 cases of plague were reported in Madagascar. The next year however 104 cases were reported,

representing a decrease of 81% (**Figure 4**). Equally the number of deaths is prone to annual fluctuation from 13% (2017) to 32% (2018).

Plague is an extreme example of this challenge, but it is not an isolated example. Nipah has also been reported in low, fluctuating case numbers (**Figure 4**) primarily in three countries – Bangladesh, India and Malaysia. [32] Nipah has however only been consistently reported each year in Bangladesh, where the median number of annual cases is 12.5 and the median case fatality ratio is approximately 70%. [33] Similar numbers of cases are seen in India, but several years can pass without any cases being identified. [33] Based on a rudimentary sample size calculation [34] [35], a clinical trial that uses a conservative effect size of 0.3, power of 0.8 and significance level of 0.05 would require a sample of 174 participants, which could mean a trial for Nipah would take almost 14 years (along with a substantial amount of funding) to complete.

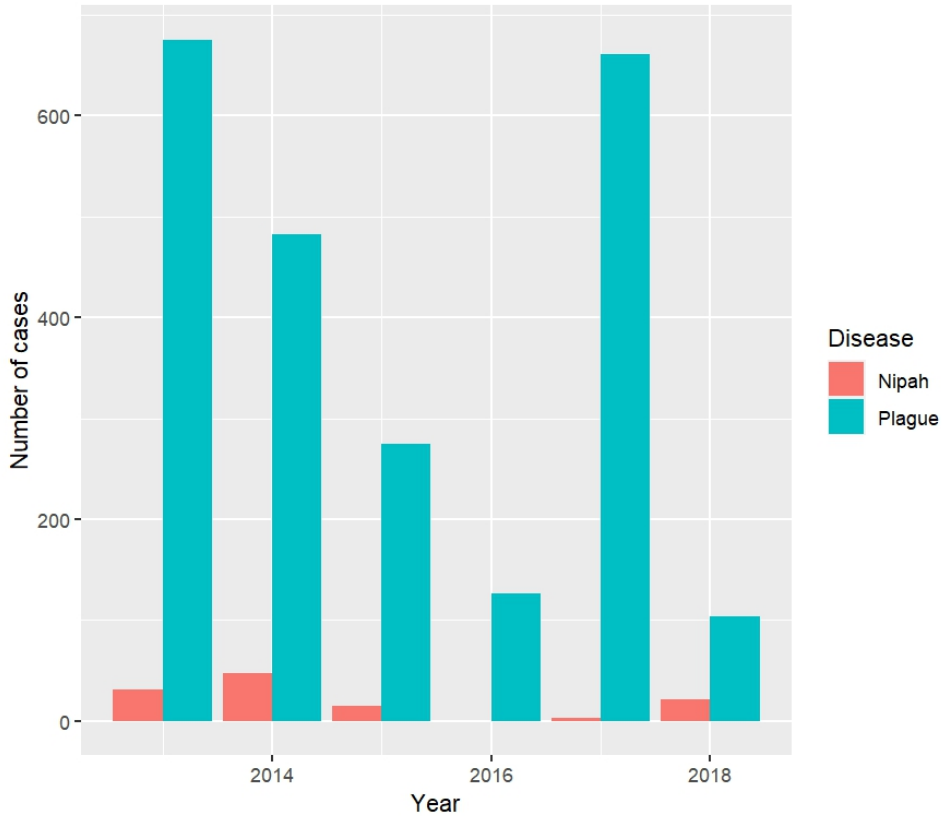


Figure 4: Annual case numbers of plague and Nipah 2013 to 2018

Bayesian designs offer a solution to this challenge through flexible design (e.g. sample size re-estimation and response-adaptive randomisation). [36] For example, during the

2014–2016 Ebola virus outbreak in the West Africa, the PREVAIL II trial used a Bayesian adaptive design to evaluate multiple treatments, using probability-based stopping rules allowing for early identification of effective therapies. [37] Bayesian designs are especially valuable in outbreak settings as they support rapid decision-making. However, their use in trials remains limited due to several challenges, including the requirement for specialised statistical expertise, design complexity, general uncertainty around using limited data to obtain prior distribution, and regulatory bodies being more familiar with traditional, frequentist approaches. [36]

Wide geographic dispersion of cases

In addition to low, fluctuating case numbers, diseases that occur over widespread geographic areas pose a number of operational challenges for trials, which risk limiting recruitment capacity. The limited financial and human resources afforded to many clinical trials mean a restricted number of trial sites must be carefully selected based on predicted case numbers. This selection is typically based on historic national reporting data, with health centres in regions consistently reporting higher case numbers chosen as trial sites.

The majority of cases of Lassa fever in Nigeria for example derive from four states [38], but over the course of the last five years cases have arisen in moderate to high numbers in most states [39] – covering an area of $> 900,000 km^2$. To optimise the use of limited resources and funding, it would be logical to select Lassa fever treatment centres in these four high-incidence states as clinical trial sites. While this strategy would capture approximately 85% of confirmed cases in 2022/3, this figure may vary from year-to-year. Furthermore, there is no guarantee that all cases would present to the participating Lassa fever treatment centres or that they would be either eligible or willing to participate, further reducing the potential participant pool.

Identifying potential participants is one challenge, but in the event that a multi-centre study covers a wide geographic area in one or multiple countries, the other challenge is heterogeneity in the standard and quality of care that patients may receive. In areas that consistently see high numbers of cases – for example at specialist treatment centres located in endemic regions – the standard of care for frequently-seen outbreak-prone in-

fections is likely to be high, due to greater investment in training, specialist staff and equipment to support local public health needs. Outside these areas, where cases may be sporadically reported in lower numbers, the standard of care may be lower, due to the prioritisation of other conflicting local public health challenges. Differences in the type of care available (e.g. quality of facilities, available equipment, staff training, case management guidelines) to patients at different potential clinical trial sites may have an impact on patient outcomes. This poses a challenge to the design of most clinical trials, which aim to minimise external factors that might affect the evaluation of the intervention. While statistical methods can help identify and adjust for treatment heterogeneity, these tools have limitations and cannot be applied with a one-size-fits-all approach [40].

Limited research capacity at health centres in endemic regions

There are a myriad of issues that have systematically eroded clinical research capacity in low- and middle-income countries (LMICs) [41] – the countries that carry the greatest burden of outbreak-prone infections [29]. While each of these reasons are important and have a meaningful impact on the reason why so few clinical trials are undertaken in LMICs, other resources are better placed to fully explore these issues in greater detail. As well as systemic barriers, there are also several operational barriers that persist at the site-level and that may obstruct the implementation of clinical trials in endemic regions.

Implementing and conducting a clinical trial is a substantial operational undertaking, resulting in an increased workload for staff working at participating trial sites. [42] Often however the initiation of a clinical trial does not result in additional human resources being provided to trial sites to help carry out an increased volume of work. In some areas where outbreak-prone infections are most prevalent, there are already chronic staff shortages that are perpetuated by under-funded health systems, [43] [44] [45] and pressure on the workforce is only exacerbated by additional activities and documentation requirements of a clinical trial.

In addition to adequate clinical staffing to perform the trial protocol is the need for a wider research workforce, which is often lacking in under-resourced health facilities. An exercise conducted by the West Africa Lassa fever Consortium (WALC) to understand

research capacity gaps among potential Lassa fever research sites demonstrated that only 58% of the 19 sites assessed had an existing data manager or research manager (some of whom had no formal research experience and GCP training) [46].

This same exercise also identified critical equipment shortages – including for basic supportive care – a lack of drug storage facilities, and diagnostic tests. These items were completely unavailable at some sites, but would be required to support clinical trial delivery and avoid heterogeneity in clinical care.

Poor clinical characterisation

For some outbreak-prone infectious diseases, while there has been little progress made in the way of randomised controlled trials, there has also been little progress made in large-scale observational cohort studies. These types of studies describe disease evolution and patient outcomes, summarise disease incidence, and assess causality [47], which are critical to inform trial design, particularly in the case of unknown or evolving pathogens.

Selection of trial endpoints specifically is driven by evidence on what happens during the course of a disease, when the important events happen and how frequently they happen. This in turn determines the outcomes of interest for evaluation in the trial, i.e. the important events that the intervention needs to modify for it to be declared useful. The outcome selected for evaluation as the primary endpoint also has important implications for the sample size calculation, which uses the expected event rate in the control arm and effect size to determine how many participants are needed to detect significant differences between groups. Where limited robust observational data exist, substantial uncertainty can be generated around the most appropriate outcome that could represent a clinically-meaningful response to treatment, how frequently it occurs and the scale of reduction that could reasonably be expected with the intervention. It also impacts understanding potential confounders, methods of stratification and potential meaningful sub-group analyses.

Low numbers of important events

Where important patient events are known, their incidence can act as the next barrier to endpoint selection. To be a viable endpoint in a trial, the outcome of interest must be detectable in a sufficient number of cases, at least in the control arm, for statistically significant differences to be observed between treatment groups.

Mortality is often used as the primary endpoint in trials for outbreak-prone infections. Of the 11 ongoing and completed trials for outbreak-prone infections that have defined a primary endpoint (not all have), most have used a survival/mortality endpoint (**Table 2**). Mortality/survival is an example of a "hard" endpoint – the type that are more acceptable to regulators as they demonstrate a direct clinical effect that is unambiguous and objectively measured. [48] While using a single hard endpoint like survival/mortality may appear to be a straightforward option, for outbreak-prone infections it is often not a viable or the most clinically-relevant endpoint. Only three of the seven trials in Table 2 using this endpoint have reached their target sample size and only one has generated sufficient evidence to move the trial intervention to licensing. While mortality is certainly an important event in general, in these cases it may not be the most clinically-relevant and there may be other outcomes that are considered to be more reflective of the impact of the disease on patient health and well-being, such as those directly impacting quality of life and daily functioning.

This challenge is however complicated when the absolute number of mortality events is relatively low and the overall case fatality ratio (CFR) is high. Nipah, for example, has a high CFR of approximately 60% (and can be as high as 100% in some outbreaks), but on average 13 Nipah deaths are recorded each year. [33] These deaths have occurred across multiple countries, often in very rural areas, sometimes in clusters of less than 10 cases, which would make accruing sufficient numbers of these events to detect statistically significant differences challenging.

In other cases, the overall proportion of mortality events may be low, such as in bubonic plague where globally approximately 16% of plague cases result in death. [31] In this case, the low overall proportion of mortality events may be too small for statistically significant differences to be detected between arms, but there may also be more clinically-meaningful

outcome events – those affecting higher proportions of patients – for bubonic plague than mortality alone.

For this reason, some clinical trials opt to define a composite primary endpoint, constructed from multiple outcomes that each appear to be significant in the course of illness and combined would increase the expected event rate, thus improving the trial's chances of detecting differences between treatment groups and improving precision. [49] [50]

To date, one completed and two ongoing clinical trials of therapeutics for outbreak-prone infections have used a composite primary endpoint (Table 2).

Table 2: Composite primary endpoints used in ongoing and completed clinical trials for outbreak-prone infectious diseases.

Trial	**Disease**	**Primary endpoint**
A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection [51]	Ebola virus (EBOV)	Survival/ mortality
Experimental Treatment of Ebola Virus Disease with Brincidofovir [52]	Ebola virus (EBOV)	Survival/ mortality
Experimental Treatment of Ebola Virus Disease with TKM-130803: A Single-Arm Phase 2 Clinical Trial [53]	Ebola virus (EBOV)	Survival/ mortality
Interferon $\beta - 1a$ for the treatment of Ebola virus disease: A historically controlled, single-arm proof-of-concept trial	Ebola virus (EBOV)	Reduction in blood viremia, resolution of clinical symptoms; improvement in survival; safety of IFN $\beta - 1a$ treatment [54]
A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics [5]	Ebola virus (EBOV)	Survival/ mortality

Table 2: Composite primary endpoints used in ongoing and completed clinical trials for outbreak-prone infectious diseases.

Trial	**Disease**	**Primary endpoint**
Experimental Treatment with Favipiravir for Ebola Virus Disease (the JIKI Trial): A Historically Controlled, Single-Arm Proof-of-Concept Trial in Guinea [55]	Ebola virus (EBOV)	Survival/ mortality
Evaluation of Convalescent Plasma for Ebola Virus Disease in Guinea [56]	Ebola virus (EBOV)	Survival/ mortality
Treatment of acute Nipah encephalitis with ribavirin [57]	Nipah	Survival/ mortality
Treatment of plague with gentamicin or doxycycline in a randomized clinical trial in Tanzania [58]	Plague	Cure or improvement in condition defined as: • Cure: resolution of fever and painful bubo swelling, and, if initially present, recovery from pneumonia or any other symptoms of plague. • Improvement of condition: the persistence of any sign or symptom in a milder degree.

Table 2: Composite primary endpoints used in ongoing and completed clinical trials for outbreak-prone infectious diseases.

Trial	**Disease**	**Primary endpoint**
ISTH/ANRS 0409s INTEGRATE Lassa Fever Study [59]	Lassa fever	• Death • New onset of kidney failure • New onset of acute respiratory failure • New onset of shock
An open-label, randomized, non-inferiority trial of the efficacy and safety of ciprofloxacin versus an aminoglycoside + ciprofloxacin in the treatment of bubonic plague (IMASOY) [1] [2]	Plague	The proportion of patients with bubonic plague with a therapeutic response (assessed on Day (D) 11). Therapeutic response is defined as follows for subjects with a visible and measurable bubo: • Alive • Resolution of fever (uncorrected axillary temperature <37.5° C) • Has not received alternative treatment for plague • No clinical decision to continue anti-plague antibiotics beyond day 10

The use of composite primary endpoints is however not so straightforward. Aggregated data from multiple outcomes can create challenges interpreting the true effect of the drug, particularly when the importance of each individual outcome component may not be equally weighted from the perspective of both patients and clinicians. [60] Determining the overall effect size associated with the composite may also be challenging, as data rarely exist on the frequency of all combined components of the endpoints; rather, data exist on the frequency of each individual outcome. [61] The effect size is a key factor in calculating sample size, and the uncertainty caused by the lack of empirical evidence on outcome frequencies raises doubts about the accuracy of the sample size estimation. [61] Finally, the interpretation of the composite result is also complicated by patients having combinations of each endpoint component, some of which may not be affected by the study drug as much as others and may occur at lower frequencies – further limiting accurate interpretation of how the study drug interacts with the disease.

For this reason, regulators recommend that each component of a composite endpoint be analysed separately to ensure that claims of efficacy are not disproportionately driven by certain components, and to confirm that clinically important outcomes are not negatively impacted – a detail that might be overlooked when grouping multiple outcomes under a single measure. [62]

Objectives and structure

It is clear that researchers identifying and defining clinically-meaningful and measurable efficacy endpoints for outbreak-prone infections are confronted with several challenges, but the need for clinical evidence remains urgent considering the diminutive number of available therapeutics.

The objectives of this thesis are therefore to:

- Understand the source(s) of bias and limitations affecting endpoint selection in clinical trials for outbreak-prone infectious diseases
- Understand how these challenges affect the reliability of trial results

- Propose ways in which clinical trials can improve the reliability and usefulness of evidence around therapeutics for outbreak-prone infectious diseases through endpoint selection

These objectives will be investigated by examining real-world challenges three outbreak-prone infections: Lassa fever, clade I mpox and bubonic plague. While these diseases have some differences – Lassa fever and mpox are both viral infections, and plague is a bacterial infection – they exhibit some fundamental similarities, particularly when it comes to the way in which interventions are evaluated (**Table 3**).

Table 3: Characteristics of Lassa fever, Clade I mpox and plague

Characteristic	**Lassa fever**	**Clade I mpox**	**Plague**
Pathogen type	Virus	Virus	Bacteria
Primary endemic region	West Africa	Central/West Africa	Madagascar/ DRC
Number of licensed therapeutics	0	1	5
Number of clinical trials	1	1	3
Number of therapeutics licensed as a result of trial data	0	0	0

Specifically, all three diseases are examples of acute infections that disproportionately affect some of the most economically vulnerable communities in the world. Cases are limited to a few endemic regions – 98% of plague cases are reported from Madagascar and the Democratic Republic of Congo (DRC) [31], Lassa fever is reported predominantly in

Nigeria and Sierra Leone [63] and clade I mpox cases are confined to the DRC, Central African Republic (CAR) and Cameroon [64] (**Figure 5**).

As shown in Figure 5, the burden of these three diseases lies primarily in Sub-Saharan Africa where cases often arise in rural settings with limited healthcare infrastructure. [65] [66] [67] However cases of plague also arise in small numbers across North and South America, and Central and East Asia.

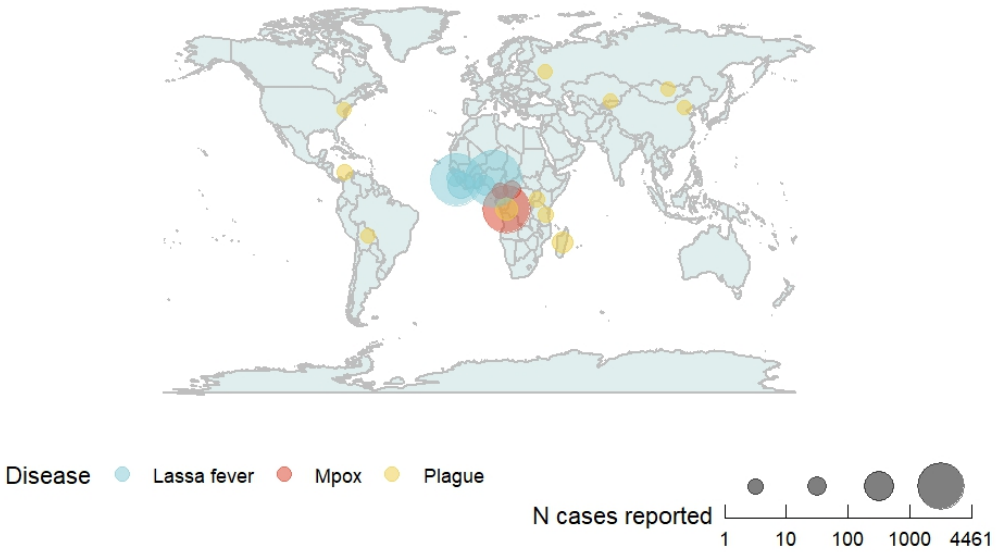


Figure 5: Global distribution of reported cases of Lassa fever, mpox (clade I), and plague

While therapeutic interventions are available for each disease, treatments that actively target the pathogens are limited and based on poor quality clinical evidence, the generation of which is further complicated by substantial methodological challenges – particularly in relation to the identification of clinically meaningful and measurable endpoints.

Chapter 1 will explore through systematic reviews gaps in the clinical characterisation of bubonic plague and Lassa fever that limit the definition of endpoints in trials. As systematic reviews summarising the clinical characterisation of bubonic plague and Lassa fever have not been published to date, these reviews also aim to explore whether endpoint selection can be informed by existing literature – or whether further prospective studies would be required.

Chapter 2 will focus on the IMASOY trial and investigate the risks and biases of evaluating bubo size reduction as a component of its composite primary endpoint. This chapter will explore measurement error through validity, inter- and intra-rater analyses of the measurement of a large dataset of artificial buboes and a small cohort of real buboes in clinical cases of plague.

Chapter 3 will explore potential endpoints that have been used (time to lesion resolution) and considered (time to PCR negativity) for clinical trials for mpox, focusing on Clade I mpox. To understand the potential endpoint limitations and ways forward for clade I mpox three studies were conducted investigating: 1) the reliability of clinical evaluation of mpox lesions, 2) assessment of different specimen types via PCR and 3) characterising treatment response in a small cohort of clade I mpox patients.

Chapter 4 explores strategies to advance clinical research on outbreak-prone infections. It presents the results of the IMASOY trial on bubonic plague, summarises the outcomes of a stakeholder consultation aimed at defining clinical endpoints for Lassa fever, and proposes a practical framework to guide endpoint selection across a broad range of outbreak-prone infectious diseases.

Introduction to the diseases included in this thesis

Bubonic plague

Plague is a zoonotic infectious disease caused by *Yersinia pestis*, a Gram-negative bacteria with persistent foci in parts of Africa, North and South America, and Asia. [68] [69]

Between 2013 and 2018 – the last year for which global data exist – 2886 cases of plague, resulting in 504 deaths, were reported across 11 countries. [31] Madagascar bears the highest disease burden, accounting for over 80% of the global cases. (Figure 6).

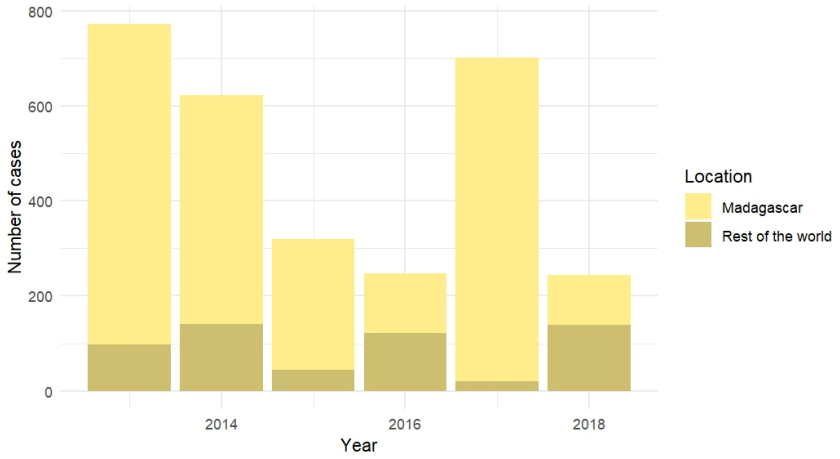


Figure 6: Reported distribution of plague cases in Madagascar versus the rest of the world

Confirmation of plague currently relies on laboratory testing through the detection of *Y. pestis* by polymerase chain reaction (PCR), culture or a seroconversion or 4-fold increase in anti-F1 antibody titer in serum samples. [70] While rapid diagnostic tests are available and offer high sensitivity when compared to culture and PCR performed on bubo aspirates, the poor specificity of these tests means that laboratory methods are still needed to generate a confirmed diagnosis. [71] The limited availability and accessibility of these services in the locations in which plague is most frequently reported means there is a possibility that many cases could go undetected.

Diagnostic challenges also arise prior to laboratory testing as clinical suspicion of plague may be limited by the non-specific clinical manifestations of the disease. Other than characteristic swollen lymph nodes (called "buboes") in cases of bubonic plague, patients may present with only a limited number of other non-discriminatory signs and symptoms, including fever, headache, chills and weakness. [68] [72]

Estimates of the case fatality ratio (CFR) in clinical settings are wide-ranging, extending from 24% to 60% [68] [73], but a lower CFR of $\approx$ 5% has been reported in research settings [58].

While several antibiotics for the treatment of plague have been approved under the FDA's "Animal Rule" [6] – including levofloxacin, ciprofloxacin and moxifloxacin [74] – others such as doxycycline and streptomycin have received approval outside the "Animal Rule", but the clinical evidence base on which approval was granted is unclear. [75]

Other treatment recommendations have also been made for drugs that have demonstrated effectiveness in small numbers of patients when used off-label for the treatment of plague. [76] Large-scale clinical evidence demonstrating the safety and efficacy of any drug used in the treatment of plague is therefore lacking.

Generating the evidence required to support the use of existing drugs (as well as any novel drugs that may be developed) is complicated by a number of operational and methodological factors. Firstly, enrolling a sufficient number of patients in a trial would require substantial resource investment, as a large number of trial sites would need to be activated over a wide geographic area to account for the sporadic nature of case detection year-on-year. Secondly, a number of drugs that lack robust clinical testing are currently used in the treatment of plague; the acceptability of any future trial comparing an active plague treatment to supportive care alone would therefore likely be limited, making establishing the true effect of any drug challenging. Finally, there are no established methodologies, including the identification of clinically-meaningful and measurable patient outcomes, that could support the generation of quality evidence in a trial setting.

To date, the results of only two small-scale randomised controlled trials in humans evaluating the safety and efficacy of plague therapeutics have been published [58] [77], and the results of a third, larger trial [1] [2] have demonstrated the non-inferiority of oral ciprofloxacin compared to standard treatment (results publication pending), but its findings do not resolve the broader question of optimal treatment, as many other regimens are recommended and have been licensed for plague. Moreover, uncertainty remains regarding the most appropriate outcomes to use when evaluating plague treatments, including around the original primary endpoint used for the IMASOY trial [1], leaving critical gaps when it comes to understanding whether treatments work.

Lassa fever

Lassa fever is an acute viral haemorrhagic illness endemic to parts of West Africa. While cases are reported annually in Nigeria, Sierra Leone, Liberia and Guinea, [78] [79] [80] sporadic cases have been reported in Benin and Togo with the potential to extend further in areas of Sub-Saharan Africa where the animal reservoir, Mastomys natalensis, is endemic. [81]

Lassa fever is estimated to cause 500,000 new infections and 10,000 deaths per year, [82] [83]. During outbreaks, human-to-human transmission is a particular concern for healthcare workers treating suspected cases of Lassa fever for whom challenges in adhering to infection prevention measures have been noted due to suboptimal provision of protective equipment, limited laboratory capacity and inadequate emergency preparedness. [84] [85]

Clinical suspicion of Lassa fever is also challenging as symptoms are often non-specific, most commonly including fever alongside headache, vomiting and abdominal pain, with a small proportion of patients presenting visible haemorrhage or other severe symptoms, and long-term complications such as hearing loss. [9] [63] In clinical settings, the case fatality rate (CFR) has been reported between 24–27% [86] [87] overall – although higher in pregnant women (34%) [88] and children (63%) [89]. In research settings however the CFR has been reported as low as 13% for adults and 6% for children. [90] Since Lassa fever was first described in 1970 [91], no drug has received approval for the treatment of Lassa fever. Ribavirin, in conjunction with supportive care, has been incorporated in to national treatment guidelines [9], but this recommendation is based on weak evidence. Only one clinical trial has been conducted to date to evaluate the safety and efficacy of ribavirin for Lassa fever [92], which has since generated concern about its methods, analysis and the safety of ribavirin when used to treat mild cases of Lassa fever. [93] [94]

While further evidence from robust clinical trials is required to re-evaluate the safety and efficacy of ribavirin for Lassa fever, there are also a small number of novel drug candidates that may be ready for clinical trials in the coming years. [95] [96] [97] As there has been limited investment to date to generate clinical evidence for Lassa fever therapeutics, methodological frameworks to evaluate new treatments in a robust and efficient manner are currently lacking.

Clade I mpox

Mpox, caused by the monkeypox virus (MPXV), is a zoonotic disease with two distinct genetic subtypes known as Clades. [98] Clade I, primarily found in Central Africa, is associated with more severe disease outcomes than Clade II, which includes Clade IIa (historically reported in West Africa) and Clade IIb disease – the latter being responsible for the 2022 Public Health Emergency of International Concern (PHEIC).

Clade I mpox primarily circulates in Cameroon, the Central African Republic and Democratic Republic of the Congo (DRC), where it has recently experienced a substantial surge in case. [99] The DRC outbreak is notable for its expansion into new regions, spread into densely populated urban areas such as Kinshasa, and the first documented cases of sexual transmission of Clade I mpox. Both the World Health Organization (WHO) and European Centre for Disease Prevention and Control (ECDC) emphasised the importance of improving clinical awareness and diagnostic capacity as key response priorities. [99] [100]

Although Clade II mpox has a low fatality rate – 156 deaths among more than 91,000 cases reported outside Africa since May 2022 – the mortality rate for Clade I mpox is significantly higher, at 1–12%. [98] Since February 2023, over 12,500 cases and 581 deaths have been reported in the DRC alone. [99]

Diagnosing mpox remains a major challenge. Confirmation is typically done via PCR [101] [102], but data emerging from the DRC suggest that genetic mutations may be reducing the accuracy of the Clade I-specific PCR test recommended by the US CDC. [103] No rapid or point-of-care diagnostic tests are currently available, and limited laboratory capacity hampers testing – only 9% of suspected cases in the DRC have been PCR-tested to date. [100]

As a result, diagnosis in affected regions relies heavily on clinical assessment, which includes patient history, epidemiological context, and visual evaluation of skin lesions. This is complicated by the similarity of mpox lesions to those of other diseases, such as varicella, which co-circulate in the same regions.

While tecovirimat was licensed for the treatment of mpox under emergency measures during the 2022 Clade IIb outbreak, there is no clinical evidence supporting its efficacy for any clade of mpox. Several trials were initiated for Clade IIb mpox following the multi-country outbreak in 2022 to evaluate tecovirimat, none of which generated conclusive evidence. One clinical trial was however initiated for Clade I mpox, which recently published negative findings. [104] This does not appear to have impacted the existing licensing decisions of the use of tecovirimat for the treatment of mpox.

Chapter 1

Understanding the source of limitations in the development and definition of endpoints for outbreak-prone infectious diseases

---

1.1 Introduction

To optimise endpoint selection for outbreak-prone infections, it is crucial to first develop a detailed understanding of the core issues underlying the endpoint selection process. An important initial step is to identify a relevant outcome that the intervention is likely to positively impact. However, a significant barrier arises from the absence of standardised and comprehensive data describing key signs, symptoms, their progression, and overall patient outcomes. This challenge is particularly problematic when mortality is not a viable endpoint, and other meaningful clinical events must therefore be identified. This chapter will examine the role of clinical characterisation in shaping endpoint definitions, with a focus on two diseases—Lassa fever and plague—that illustrate these complexities.

As described in the Introduction, trials, while limited in number, have been conducted for both diseases but have not resulted in the generation of conclusive evidence to license an investigational product – despite one or more products having been incorporated in to case management guidelines for both diseases. (Table 1) [9] [17]

If future therapeutic trials were to be conducted, evaluating mortality alone would not be viable as an endpoint for either plague or Lassa fever, due to the expected numbers of mortality events being too small for a trial to detect significant differences between groups. Trials would therefore need to identify one or more alternative outcomes that have a clinically-meaningful impact on how a patient feels, functions or survives following administration of an intervention. To do this however robust clinical data describing the disease evolution are required.

To understand whether existing published literature can support the definition of a viable primary endpoint for both plague and Lassa fever, this chapter will investigate:

- How research for plague and Lassa fever has historically been conducted
- Whether a clear clinical picture of plague and Lassa fever – e.g. key clinical features at baseline, how the diseases evolve and important patient outcomes – can be built from the data generated by studies conducted to date
- The barriers, sources of bias and limitations that prevent the selection of clinically-meaningful endpoints

To answer these questions we conducted two systematic reviews – one for plague and one for Lassa fever. Systematic reviews have previously been used to establish a foundation for the identification of important disease outcomes, specifically in the development of Core Outcome Sets (a standardised set of outcomes that should ideally be measured in all trials for a specific disease). Two systematic reviews of cutaneous leishmaniasis interventions [105] [106] for example formed the foundation for a proposed standardised trial methodology, including its endpoints [107], and a review of tuberculosis meningitis research informed the proposed outcomes for future trials.[108]

As this approach has previously acted as a valuable foundation for outcome selection for other infectious diseases, it has been applied as the basis for this project. In the context

of outbreak-prone infectious diseases, where trial data may be sparse or inconsistent, this method is particularly important for ensuring robust data are available to support important choices for trial design, like endpoint selection. Systematically evaluating the signs, symptoms and outcomes reported in previous studies may expose the gaps in the overall clinical characterisation of plague and Lassa fever and the sources of bias and limitations that affect endpoint selection.

1.1.1 Clinical characterisation of Lassa Fever: A systematic review to inform clinical trial design

As there are a number of potential new therapeutic candidates progressing through the R&D pipeline for Lassa fever, it is important to establish a robust understanding of disease presentation and evolution to support endpoint selection for the evaluation of therapeutic efficacy. The data synthesised in this systematic review aims to develop a clinical picture of important features of Lassa fever from presentation at a health facility and the important events that arise over the course of the illness.

This systematic review is registered in PROSPERO under protocol number CRD42020220365.

Methods

Before initiating the systematic review, an initial search of Epistemonikos and Prospero was conducted, which returned no results, to understand if any high-quality systematic reviews had already been conducted that met any of the objectives of this review.

A search of biomedical literature databases and clinical trial registries was therefore conducted to identify studies describing Lassa fever that were published before 15 April 2021 (Appendix 2).

Studies that reported primary results of patients with a laboratory or clinically confirmed diagnosis of Lassa fever and described clinical features of the disease were eligible for inclusion in the systematic review. No language restrictions were applied.

All identified studies underwent two-stage screening for eligibility using Rayyan [109], which included, first, a review of titles and abstracts for potential inclusion, then a full text review of the articles to confirm eligibility. Inclusion or exclusion of each study required the

agreement of two independent reviewers, for whom screening decisions were blinded until screening was complete. Any conflicts between decisions of each reviewer were resolved through discussion.

Data were extracted from the included manuscripts on to a standardised electronic data capture form using REDCap [110]. Data extraction was completed by one reviewer and validated by a second reviewer. A record was created for each included article summarising the study design, demographic details of the study population, prevalence of signs, symptoms and important patient events at baseline and any timepoint post-baseline. A pre-defined list of 19 frequently reported signs and symptoms was created prior to data extraction to support standardised data capture about these events. Other signs and symptoms could be recorded using a free-text variable.

Analysis

A descriptive analysis summarising the demographic and clinical characteristics of patients with Lassa fever was performed using Microsoft Excel.

The analysis presents the number and percentage of articles according to article type (e.g. randomised-controlled trial, observational cohort study, case report). The number of articles published and number of patients reported in articles is presented per year as a combined bar and line graph.

Risk of bias assessments were completed using the Joanna Briggs Critical Appraisal Tools [111]. The results of the risk of bias assessments are presented per study type (e.g. randomised-controlled trial, observational studies, case reports) showing the outcome of each question per study as a stacked bar graph.

Demographic characteristics of the included study populations are summarised as the number of patients included per country and the percentage of patients included from the overall patient population in the systematic review. The number and percentage of male and female participants is also reported, along with the number of patients for whom sex could not be determined.

A narrative description of the endpoints used in the included studies is also provided, which includes the median follow-up duration with the inter-quartile range (IQR) and range.

Clinical characteristics are presented as the number and percentage of patients reporting each sign or symptom at baseline and post-baseline. Baseline is defined as the timepoint associated with the initial assessment and post-baseline is defined as any timepoint in the study after the initial assessment. To ensure accurate reporting on clinical features relating to Lassa fever, the analysis population includes only patients from articles in which 100% of the study population has laboratory confirmation of Lassa fever.

The prevalence of each sign or symptom among the analysis population is reported according to the number of patients for whom there was evidence of assessment of the sign or symptom in an article. The prevalence is therefore reported as the number of patients with a sign or symptom over the number of patients assessed for the sign or symptom as evidenced in the study's methods.

The signs and symptoms reported in the table in the main body of the text are those deemed clinically-significant by an independent clinician with experience in Lassa fever who reviewed a list of reported signs and symptoms before the analysis took place. Data on all signs and symptoms is reported in the appendices.

Recent articles have highlighted the importance of aspartate aminotransferase (AST) in Lassa fever outcomes. [94] [93] The analysis therefore presents the number and percentage of articles that report on AST, as well as the type of data recorded. AST is reported categorically according to the thresholds that were considered significant in these articles [94] [93] (< 150 IU/L and ≥ 150 IU/L)

Results

Search results

The literature search returned 4,794 publications. After removing duplicates, 2,704 titles and abstracts were screened and 195 full-text publications were assessed for eligibility, resulting in 147 studies that are included in the review (Figure 1.1).

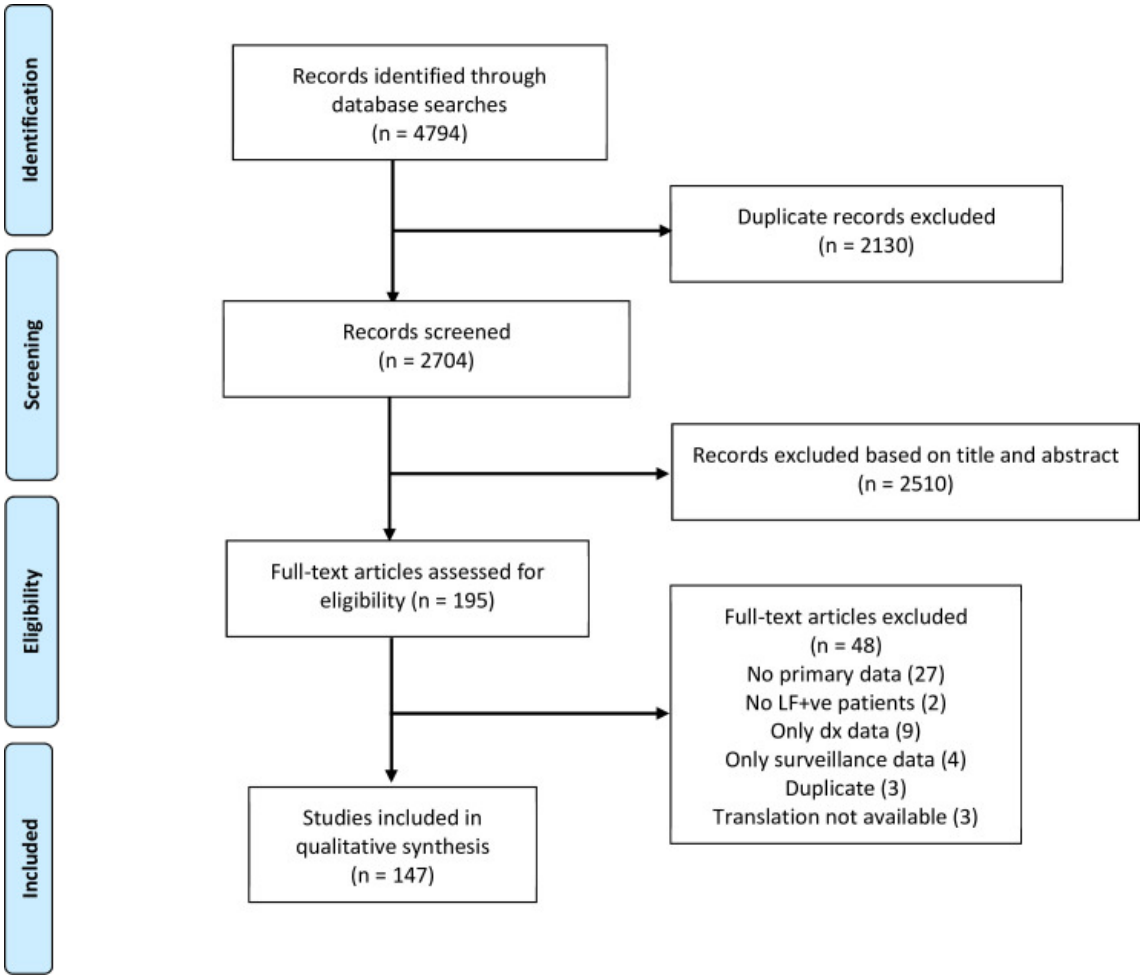


Figure 1.1: PRISMA (2009) flowchart summarising the article screening process

Included studies were published between 1970 and 2021 with increasingly higher numbers of studies being published after 2010 (Figure 1.2). Studies include 53 (36%) case reports, 41 (28%) case series, 30 (20%) cohort studies, 10 (7%) case-control studies, 11 (7%) cross sectional studies, and 2 (1%) quasi-randomised studies (defined as studies that uses randomisation for only part of its patient cohort or is not truly random) (Figure 1.2).

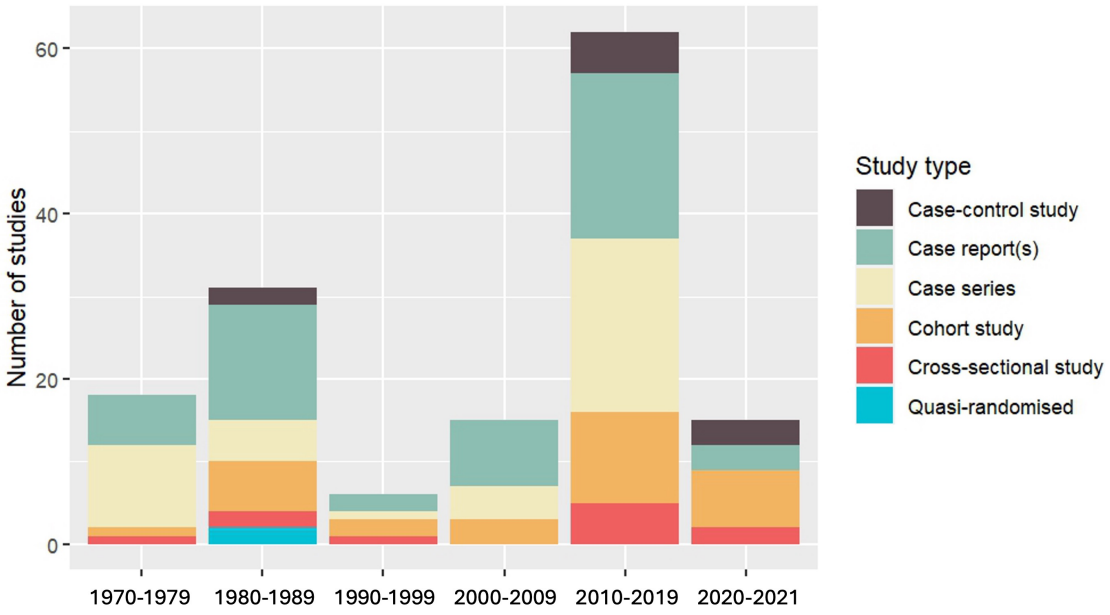


Figure 1.2: Summary of included studies by type and year of publication

Case reports and case series, which constitute the majority of the included studies, reported a median of one patient, while other protocolised studies enrolled a median of 58 participants.

Population characteristics

The overall study population included data on 8550 individuals, of whom 7862 were confirmed cases of Lassa fever. Most participants (91%) were enrolled in studies conducted in either Sierra Leone or Nigeria, where the highest numbers of Lassa fever cases are reported. The remainder of the study population was enrolled in studies conducted in other parts of West Africa, Europe, Asia and North America (Figure 1.3).

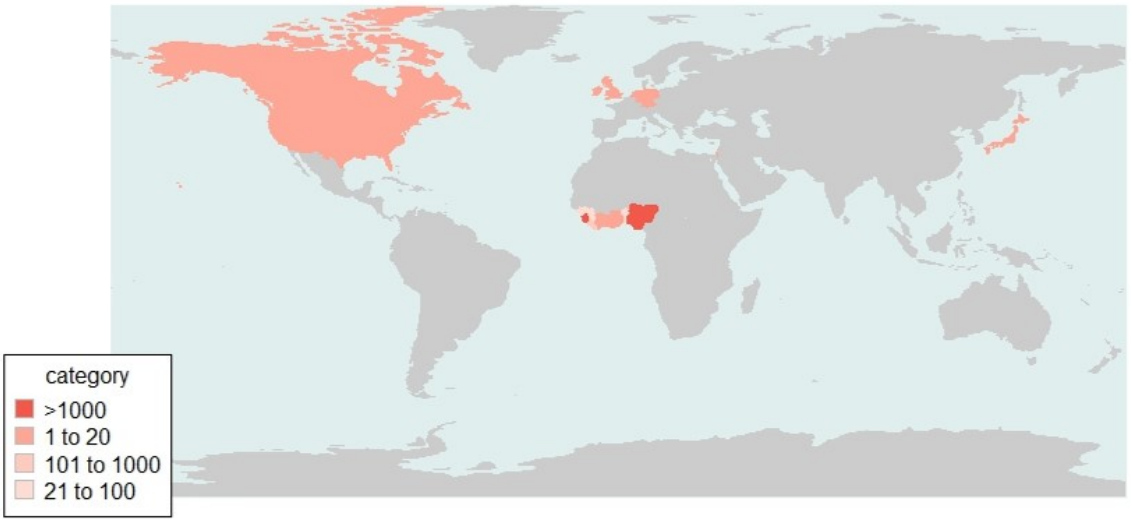


Figure 1.3: Number of patients included in the studies per country

The study population includes 3477 (41%) male participants and 3839 (45%) female participants, while sex was not reported for 1234 individuals (14%). Participants were aged between 0 and 90 years old, and 48 (37%) publications included a patient under the age of 16 years. At least one pregnant participant was included in 35 publications (24%).

Outcome measures

In the six (4%) publications that defined a primary endpoint, five (83%) publications used mortality and one (17%) used the diagnosis of acute kidney injury (AKI).

Three publications (2%) reported at least one secondary endpoint. Viral load was evaluated in two (66%) publications, and aspartate aminotransferase (AST), live birth, all-cause in-hospital fatality, frequency of acute kidney dysfunction, prognosis of AKI and AKF and time to hospital discharge were each listed as an outcome measure in one (33%) publication.

The median follow-up duration was 21 days (IQR 9–67 days).

Baseline clinical characteristics

At baseline, fever was the most prevalent symptom, identified in 88% of the individuals in whom it was assessed (1527/1730), followed by headache (809/1622, 50%), vomiting (806/1613, 49%), abdominal pain (660/1581, 42%) and cough (556/1581, 35%). (**Figure 1.4**)

A smaller number of patients presented with clinically severe or life-threatening signs and symptoms such as shock (12/187, 6%), breathing difficulty (21/310, 7%), and seizure (13/517, 3%). In pregnant women, labour complications were reported in 2/35 (6%) cases.

The median time from symptom onset to presentation was five days with a range of 0–32 days.

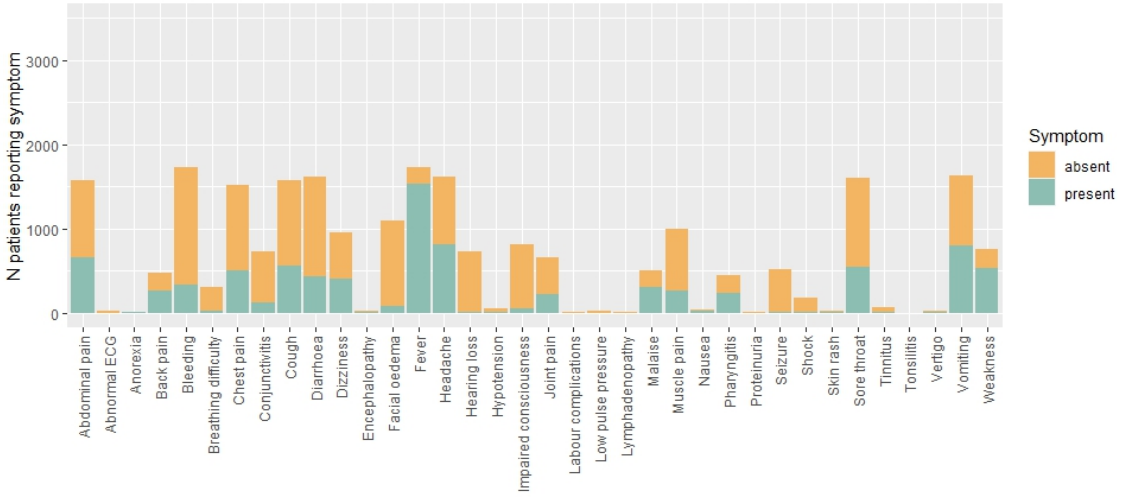


Figure 1.4: Reporting signs and symptoms at baseline

Post-baseline clinical characteristics

Fever was the most prevalent symptom during follow-up, reported for 3067/3300 individuals (93%) – representing a 5% increase in prevalence from baseline.

After fever, headache (2033/3200, 64%), vomiting (1695/3077, 55%) and abdominal pain (1594/3039, 52%) were again reported most frequently.

The prevalence of severe or life-threatening signs and symptoms increased post-baseline. Shock (87/175, 50%) had the greatest increase in prevalence, followed by breathing difficulty (265/1564, 17%) and seizure (37/548, 7%). Severe complications of Lassa fever were more frequently reported in case reports and case series than in other protocolised studies.

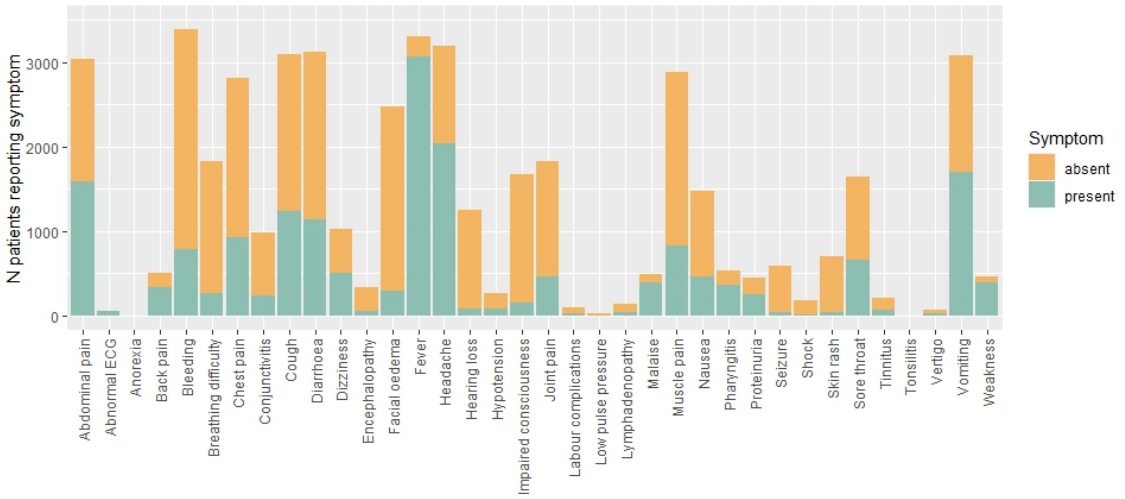


Figure 1.5: Reporting signs and symptoms at any point during follow-up

Mortality

Outcome data were available for 6373 participants, of which 1896 (30%) were reported to have died across 109 publications (74%). The median time from inclusion to death was 7.5 days (IQR 3–11).

Mortality was reported in a higher number of case reports and case series than other protocolised studies. Of the 94 case reports and series included in this review, 76 (81%) reported at least one death, while eight (15%) of the 53 included protocolised studies reported at least one death.

Aspartate aminotransferase (AST)

Aspartate aminotransferase (AST) levels were reported at baseline and post-baseline in 29 (20%) and 43 (29%) publications respectively, most of which reported a single AST value from a single patient. At baseline, 10 publications reported 19 individuals as having AST levels <150 IU/L and 15 publications reported 459 individuals as having AST levels $\geq$ 150 IU/L. During follow-up, the number of patients with AST <150 IU/L decreased to 18 and the the number of patients with AST $\geq$ 150 IU/L almost halved to 284.

Risk of bias assessments

Overall, there was a moderate risk of bias across all study types (Figure 1.6). Several criteria that were considered to be particularly important to this systematic review, including a clearly defined study population, reliable measurement of the exposure and reporting

of adverse events, were clearly reported in 60% or less of the evaluated publications. Outcomes were often reliably measured and reported in protocolised studies; however, in case reports and series, which constitute 64% of the included publications, approximately 50% of the studies did not clearly report follow-up information or outcomes.

Discussion

Despite fifty years of reporting and research, the clinical picture of Lassa fever remains unclear and inconsistent. A historic lack of investment in clinical research and drug development [112] has resulted in the limited anthology of published study reports presented in this review, which includes only one randomised controlled trial with a high risk of bias.

Although initiatives to support Lassa fever research and drug development have expanded in recent years, such as the launch of the World Health Organization (WHO) Research & Development Blueprint [113], the Coalition for Epidemic Preparedness Innovations (CEPI) "Enable" project [114] and U.S. FDA's Priority Review Voucher [115], future efforts to design interventional trials will face significant challenges. These challenges stem from the historic lack of harmonised approaches to clinical research and the resulting limited understanding of the clinical profile of Lassa fever.

Current research landscape for Lassa fever

The majority of published studies on Lassa fever to date consist of retrospective case reports and series involving single or small numbers of patients, thus limiting their generalisability, which is compounded by inconsistent reporting practices – for example, there is no standardised approach to documenting disease progression over a set follow-up period. Instead, case reports often focus on novel, unusual or severe patient events (such as mortality and life-threatening signs and symptoms), neglecting a more comprehensive summary of Lassa fever and often at the exclusion of patients whose clinical course involved mild symptoms and uneventful recovery.

Clinical characterisation of Lassa fever

While the signs and symptoms summarised in this systematic review demonstrate strong correlation with those already included in the case definitions of the World Health Organisation (WHO) [116] and Nigeria Centre for Disease Control (NCDC) [9], they also

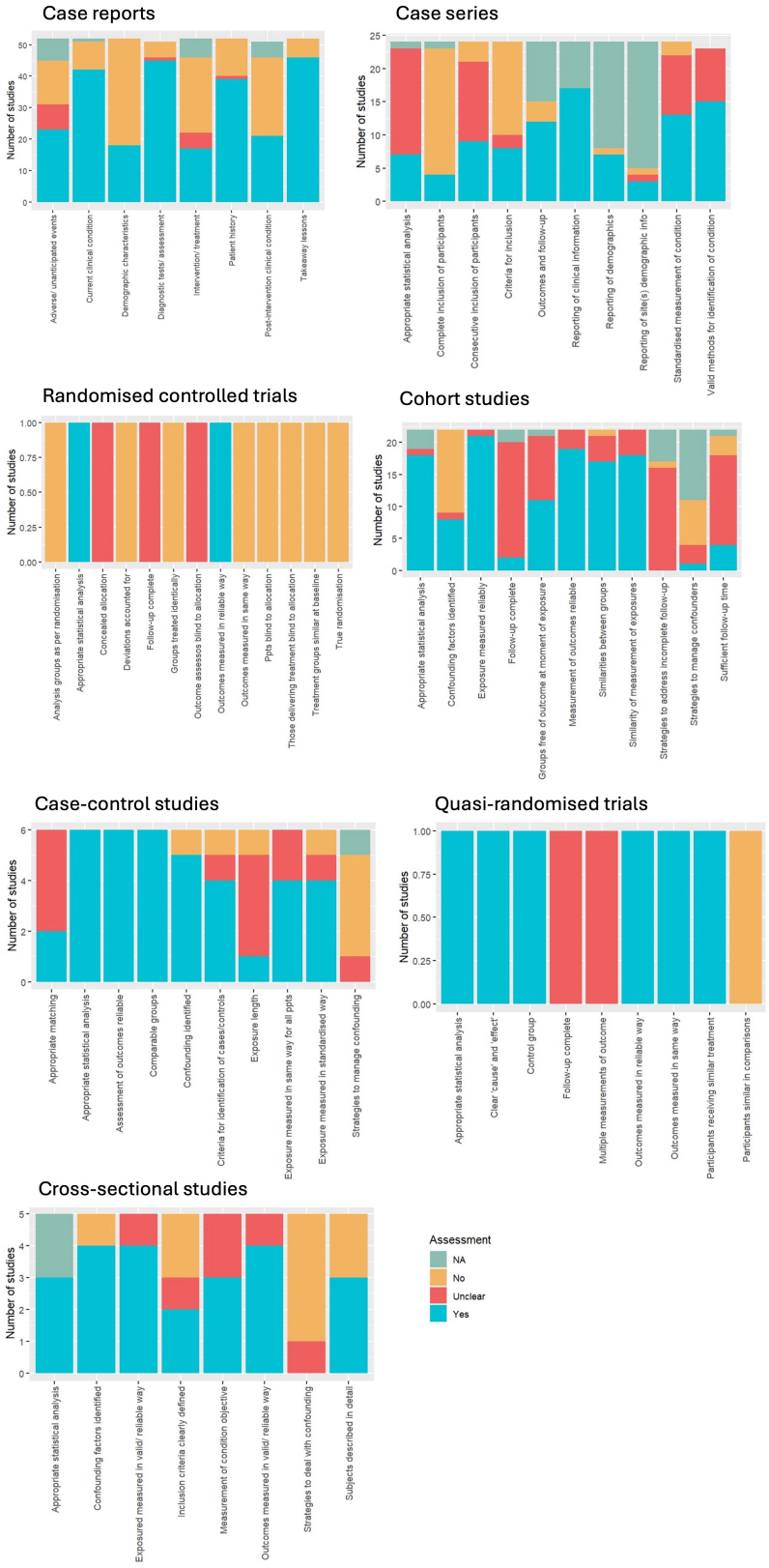


Figure 1.6: Risk of bias summary

reveal a much broader spectrum of symptoms that arise both at presentation and during the course of clinical illness.

Given the range of signs and symptoms presented in this review, generating a standardised characterisation of Lassa fever is challenging, particularly at presentation when the most common signs and symptoms (fever, headache, vomiting) are difficult to differentiate from other common febrile illnesses. The variability and type of signs and symptoms observed in this review also complicates the selection of meaningful outcomes for clinical trials, as those reported most frequently at baseline may not represent a clinically significant impact of an intervention on the disease.

Clinically severe or life-threatening signs and symptoms could provide a more suitable basis for identifying outcomes in a clinical trial. These indicators are more clinically meaningful than the common symptoms mentioned above, as they have the potential to significantly impact how a patient feels, functions or survives. The reliability around the estimations of their prevalence however remains uncertain, making it difficult to generate a reliable sample size. Furthermore, in this systematic review there is a clear discordance between the case fatality ratio (CFR), which is 30%, and the prevalence of severe and life-threatening signs and symptoms, which is much lower. This discrepancy raises questions about how often more severe symptoms actually appear and whether they can act as a reliable indicator of the severity or outcomes of Lassa fever.

Moreover, the high CFR presented in this review is likely due to the large volume of case reports and series that were included and may not be representative of the CFR that would occur in a clinical trial. Due to enhanced monitoring and availability of standardised supportive care in a trial setting, trial participant outcomes are generally expected to be better than those in the general patient population. [117] [118] A recent observational study of patients with Lassa fever demonstrated a CFR of 12%. [90]

Main barriers and sources of bias

The main sources of bias in the studies included in this review originate both from the study designs themselves and the quality of reporting on key characteristics essential to understanding the progression and outcomes of Lassa fever. Specifically, these biases are

linked to inconsistencies in the measurement of exposures, the documentation of critical patient outcomes, and the reporting of adverse events.

In particular, pregnancy is a known risk factor for severe outcomes in Lassa fever [88], including significantly higher case fatality rates. While several studies included in this systematic review included pregnant women, few stratified or adjusted their reporting for pregnancy. The inclusion of pregnant women without appropriate stratification may therefore have distorted the overall characterisation of disease severity, symptomatology, and prognosis within the general patient population. As it was challenging to extract data exclusively related to pregnant women, any potential biases present in the included studies have been inherited in the presentation of results in this systematic review.

Other significant factors influencing the risk of bias in the included studies may fall outside the scope of the standardised risk of bias assessment employed in this evaluation [111] While the evaluation incorporates elements such as the measurement and reporting of results, it does not address how outcomes were selected, the validity of their measurement, or their impact on other aspects of study design. For instance, in protocolised studies, the risk of bias assessment does not evaluate whether an appropriate effect size was chosen, the validity of assumptions underlying sample size calculations, or whether the selected outcome measures are clinically meaningful. Furthermore, the assessment does not consider whether study populations were representative of the broader patient population or identify potential sources of confounding that could influence the study results. These limitations may affect the reliability of the findings in this review, as unmeasured variables and methodological choices could introduce biases that remain undetected by standard tools. Thus, while the current evaluation offers insights into the quality of reporting and the potential for bias, it may not fully capture all factors that could influence the generation of a complete, accurate understanding of the clinical characterisation of Lassa fever.

Author contribution to "Clinical characterization of Lassa fever: A systematic review of clinical reports and research to inform clinical trial design"

Role description	Author contribution
Conceptualisation	-
Methodology	-
Software	-
Validation	X
Formal analysis	X
Investigation	-
Resources	-
Data curation	X
Writing (original draft)	X
Writing (review and editing)	X
Visualisation	X
Project administration	X
Funding acquisition	-

Publications

This chapter is based on the following publication:

- **Bourner J***, Merson L*, Jalloh S, Erber A, Salam AP, Flahault A, Olliaro PL. Clinical characterization of Lassa fever: A systematic review of clinical reports and research to inform clinical trial design. PLoS Negl Trop Dis. 2021 Sep 21;15(9):e0009788. doi: 10.1371/journal.pntd.0009788. PMID: 34547033; PMCID: PMC8486098.

*These two authors contributed equally to this manuscript and are listed as joint first authors

1.1.2 Clinical characterisation of bubonic plague: A systematic review to inform clinical trial design

As discussed in the Preface to this thesis, two clinical trials for plague have taken place and published their endpoints, and the perceived limitations of the composite primary endpoint of the IMASOY trial have inspired this work to take place.

Given the uncertainties around the role of the bubo (and other potential important patient characteristics and events) and lack of consensus on the optimal outcome measures for evaluating treatments for bubonic plague, this review seeks to provide a comprehensive summary of the clinical profile of patients with plague to guide the direction of changes to the IMASOY endpoint and endpoints that are used in future trials. Specifically, it examines the signs and symptoms (including the bubo) at presentation, their progression over time, and other important patient outcomes. This information will support the identification of clinically relevant and meaningful endpoints that reflect the disease's natural history and treatment effects for translation in to measurable, and patient-centred trial endpoints.

Methods

The literature search for peer-reviewed publications describing the clinical characteristics of patients with bubonic plague was conducted on bibliographic databases and clinical trial registries, including PubMED, Cochrane CENTRAL, clinicaltrials.gov, ISRCTN and the International Clinical Trials Registry Platform (ICTRP). A supplementary search was conducted on JSTOR to identify older reports (pre-1970) – although no formal search strategy was employed to obtain these publications due to the limited search and data extraction functionality of JSTOR.

All study designs were eligible for inclusion (from individual case reports to randomised controlled trials) providing the article contained individual patient data for adults or children of any age with suspected or confirmed bubonic plague, and described signs, symptoms and outcomes. There were no restrictions placed on language or publication year. Non-human, non-clinical, post-mortem and vaccine studies were excluded.

Two reviewers independently conducted screening in Rayyan [109]. Screening was completed in two stages: first, by reviewing titles and abstracts; then, by undertaking a full-text review of remaining articles. Data were extracted by the first reviewer using a standardised data capture form in Excel. Risk of bias was evaluated using the Joanna Briggs Institute Critical Appraisal tools. [111] The second reviewer verified the final list of included articles and performed a quality control review on 30% of the data. Disagreement about the data extraction was discussed between the two reviewers until a resolution was agreed.

Analysis

The analysis was completed using R v.4.2.2. [110] The screening and inclusion processes are summarised in the PRISMA flow diagram (**Figure 1.7**).

A comprehensive summary of all included studies is presented in a table, which details the study title, study type, country of conduct, number of bubonic plague cases, male-to-female ratio, and the median and range of ages in the study populations (in years).

Two bar charts visualise the data related to the types of articles included in the review: one showing the distribution of included studies per year by country, and the other illustrating the size of patient populations in the studies per year by country.

Demographic data extracted from each study include the overall male-to-female ratio and age (median and range) for the combined patient population across all included studies. Additionally, the total number and percentage of pregnant women and patients with comorbidities in the patient population are provided.

The clinical presentation of cases reported in the articles is summarised for patients with a confirmed clinical or laboratory diagnosis of bubonic plague. Signs and symptoms are reported at two time points: baseline (defined as the first recorded interaction with a healthcare provider) and post-baseline (all time points following the initial interaction). Signs and symptoms are presented as the number and percentage of patients reporting each symptom at least once. To ensure accurate denominators, only studies that reported a given sign or symptom for at least one patient were included in the calculation; studies that did not assess or report the evaluation of a given symptom were excluded.

Treatment data are summarised as the number and percentage of patients receiving treatment or no treatment. Treatments are categorised by antimicrobial class based on the classification from a recent systematic review by the U.S. Centers for Disease Control and Prevention. [119] Antimicrobials are divided into "high-efficacy" classes (aminoglycosides, tetracyclines, fluoroquinolones, sulfonamides, and amphenicols) and "other antibiotics". Mortality data, including the number and percentage of deaths, are reported for patients treated with high-efficacy antibiotics, other antibiotics, or no treatment.

Patient outcomes are summarised as the overall number and percentage of deaths in the entire patient population and stratified by studies conducted in high-income countries versus low- and middle-income countries. Additionally, the final status of the bubo (presence or absence) is reported as the number and percentage of patients with a known bubo outcome.

Results

In total, 2023 publications were identified in the search (**Figure 1.7**). After removing duplicates, 1984 publications were screened for inclusion. Following title, abstract and full-text screening, 1897 records were excluded, resulting in the inclusion of 87 publications for data synthesis (**section 4.6**).

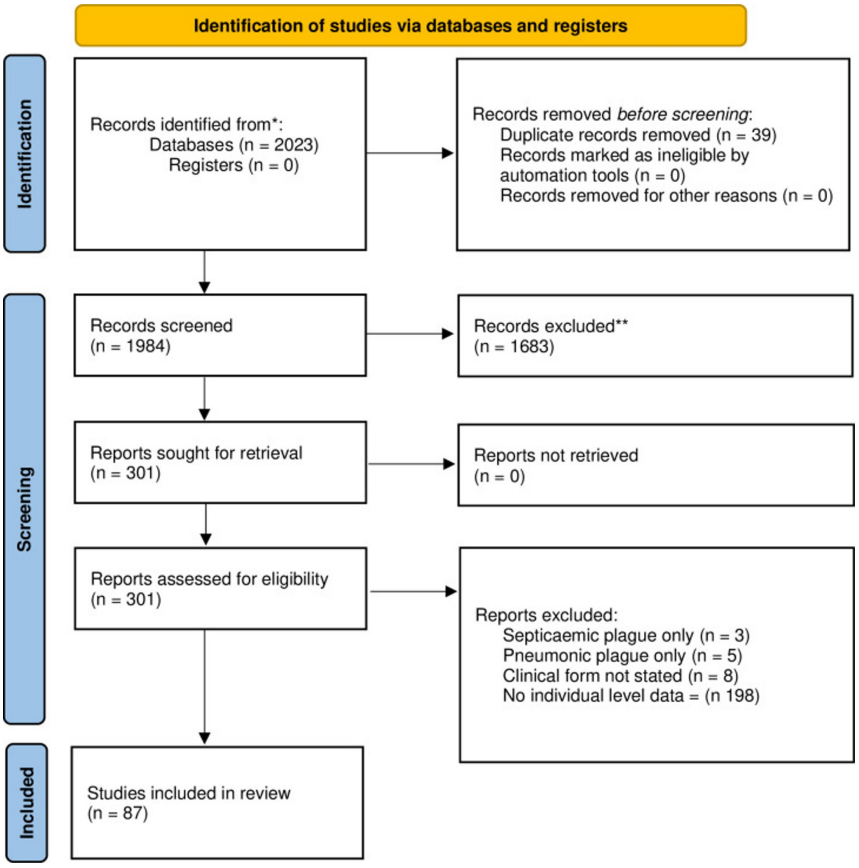


Figure 1.7: PRISMA (2009) flowchart summarising the article screening process

The majority of included studies were case reports (n=72, 83%). The remaining studies included 11 cohort studies (13%), four interventional studies (4%), of which one (1%) was randomised and three (3%) were non-randomised.

All studies were published between 1902 and 2021 (**Figure 1.8**). Most studies were case reports conducted in the USA (n=73, 84%) during the 1970s and 1980s. Aside from a small number of reports from the UK and Israel at the beginning of the 20th century, the remaining studies were conducted in Vietnam (n=3, 3%), Zambia (n=2, 2%), Libya (n=1, 1%), Madagascar (n=1, 1%), Mongolia (n=1, 1%), Tanzania (n=1, 1%), and Uganda (n=1, 1%).

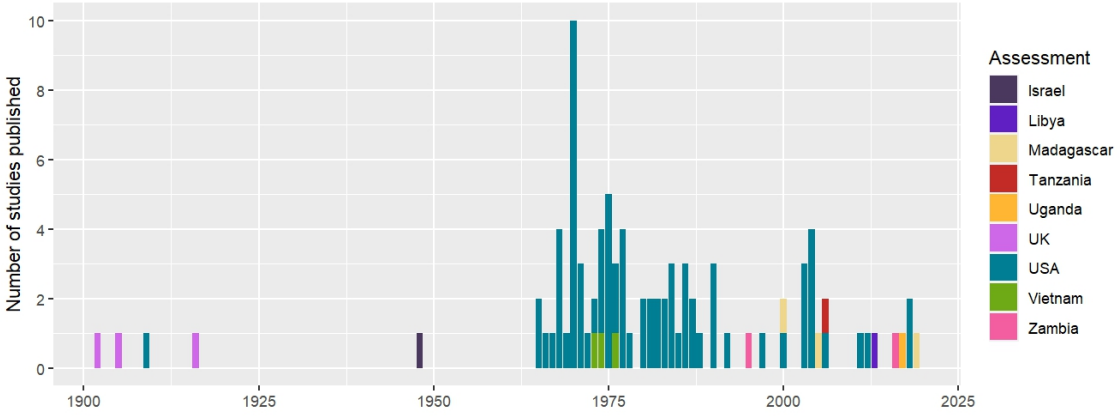


Figure 1.8: Summary of included studies by year and country

A total of 1343 probable and confirmed cases of bubonic plague (including patients with secondary pneumonic or septicemic plague) were reported across the included studies. While the majority of studies were conducted in the United States, the largest proportion of confirmed cases originated from a single study in Madagascar, which accounted for 870 cases (65%). The United States contributed 307 cases (23%) to the confirmed case cohort, followed Vietnam (n=76, 6%), Tanzania (n=65, 5%), Zambia (n=8, 1%), the United Kingdom (n=6,< 1%), Uganda (n=4, < 1%), Israel (n=3, < 1%), Mongolia (n=2, < 1%) and Libya (n=2, < 1%) (**Figure 1.9**).

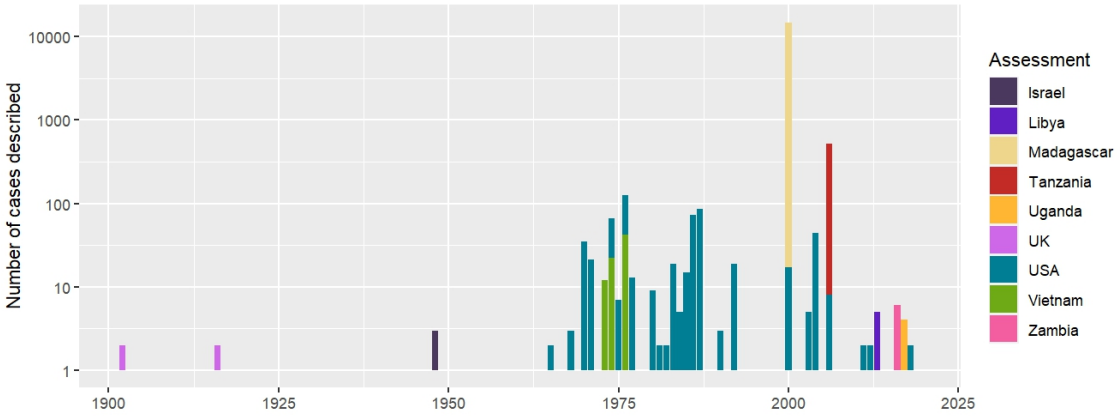


Figure 1.9: Summary of confirmed bubonic cases described in the included studies by year and country

The geographic distribution of patients included in this review changed over time. From the 1960s to the 1980s, most reports of plague originated from the United States,

accounting for 209/285 (73%) cases during this period. These were predominantly case reports of single patients or small case series, with the remaining cases reported from Vietnam. Since the 1990s, the focus has shifted toward sub-Saharan Africa, where 91% of cases included in this review were reported from Madagascar, Tanzania, Zambia, Uganda, and Libya. The remaining cases during this period were reported from the United States and Mongolia.

The study population consists of 769 (57%) males and 588 (43%) females with ages ranging from 0 to 82 years and a median age of 17 years. Five patients (<1%) were pregnant at the time of their plague diagnosis. Comorbidities were reported in 17 (< 1%) patients, including 13 (< 1%) who tested positive for malaria. Other reported comorbidities include chronic cough, chronic renal insufficiency, idiopathic thrombocytopenia, and heart murmur.

Baseline signs and symptoms

The median time from symptom onset to hospital admission was two days, with a range of 0 to 18 days.

Among confirmed cases, a wide variety of signs and symptoms were reported at baseline and post-baseline time points (**section 4.6**). The most frequently reported sign at baseline was the presence of a bubo recorded in 1216/1265 cases (96%). Patients had a median of one bubo, but 19 patients reported multiple buboes. The location of the bubo was documented in 1090 patients (90%), with the inguinal region being the most common site (590/1090, 54%), followed by the axillary (276/1090, 25%) and cervical regions (159/1090, 15%). Buboes were reported less frequently in other locations, such as the femoral and epitrochlear areas. Pain at the site of the bubo was reported in 230/320 patients (72%). Bubo size was infrequently documented, with only 19/88 studies (22%) recording one measurement for at least one patient. The median recorded bubo size was 30mm, ranging from 1mm to 150mm. The method of measurement was specified in three case reports, where ultrasound and CT scan were used. [120] [121] [122]

Following the presence of a bubo, fever was the most frequently reported systemic feature at baseline, present in 1004/1279 (78%) patients, and for which the median recorded temperature was 39.5°C (range: 36°C to 41.5°C) (**Table 1.2**). Headache was present in

127/244 (52%) patients, followed by altered mental status (33/84, 39%), chills (75/195, 38%) and malaise (22/68, 32%).

Table 1.2: Reported signs and symptoms at baseline and post-baseline, N (%) patients.

Sign/symptom	Baseline	Post-baseline
Fever	1004/1279 (78%)	52/120 (43%)
Headache	127/244 (52%)	3/69 (4%)
Altered mental status	33/84 (39%)	-
Chills	75/195 (38%)	-
Malaise	22/68 (32%)	1/1 (100%)
Fatigue	6/22 (27%)	1/3 (33%)
Myalgia	54/204 (26%)	-
Abdominal pain	48/194 (25%)	2/2 (100%)
Vomiting	63/261 (24%)	5/70 (7%)
Sore throat	17/73 (23%)	-
Cough	37/167 (22%)	12/115 (10%)
Nausea	41/191 (21%)	1/65 (2%)
Hypotension	15/87 (17%)	11/85 (13%)

Fewer patients exhibited clinical signs and symptoms indicative of severe illness at baseline, such as seizure (4/34, 12%), septic shock (5/59, 8%), and respiratory distress, arrest or failure (3/5, 60%).

Post-baseline signs and symptoms

The reporting of signs and symptoms declined following the baseline assessment. Patients were monitored for a median duration of seven days post-admission (IQR: 2–20 days). Reports of fever and headache persisted in 43% of patients, whereas other common symptoms observed at baseline, including malaise, fatigue, abdominal pain, vomiting,

cough, nausea, hypotension, and diarrhea, were infrequently reported thereafter. Although altered mental status and chills were initially reported in relatively high proportions of patients at baseline, no reports of these symptoms were documented post-baseline.

Only 19/88 (22%) studies provided data on the presence of buboes at any timepoint post-baseline. Among these patients, persistent buboes were reported in 61/95 (64%) patients. Bubo outcomes were reported for 93/1279 (7%) cases in which a bubo was documented at baseline. Two case reports included repeated measurements of buboes: one recorded an increase in inguinal lymphadenopathy at three days post-baseline, which subsequently reduced by 25mm to a size of 10mm at 26 days post-baseline. The other described an axillary bubo that decreased by 15mm to 5 m at 29 days post-baseline and fully resolved by 39 days post-baseline [123]. The bubo had completely resolved in 32/93 (34%) patients by the time of the last observation. In contrast, 61/93 (66%) patients still exhibited a bubo at the final reported observation, which occurred between 0 and 137 days post-admission.

Conversely, the frequency of certain signs and symptoms indicative of severe illness increased after baseline assessment. For example, respiratory distress, arrest, or failure was recorded in 8/65 (12%) patients, and disseminated intravascular coagulation was reported in 7/41 (17%) patients.

Outcome

Final clinical outcomes were recorded for 1300/1343 (97%) patients by the time of the last observation. Of these, 1090/1343 (81%) had fully recovered, while 208/1343 (15%) were reported to have died (**Table 1.3**). The case fatality ratio varied by national income level classification, ranging from 10% in high-income countries (HICs) to 17% in low- and middle-income countries (LMICs). The median time to death was one day in HICs and two days in LMICs, with a range of 0 to 16 days overall.

Table 1.3: Summary of patient outcomes.

Outcome	**All patients**	**HIC patients**	**LMIC patients**
Death, n/N (%)	208/1343 (15%)	32/316 (10%)	176/1027 (17%)
Time (days) from enrolment to death, median (range)	1 (0 to 16)	1 (0 to 16)	2 (1 to 5)
Fully recovered at last reported observation	1076/1343 (80%)	227/316 (72%)	849/1027 (83%)
Time (days) to defervescence, median (range)	3.5 (0 to 21)	4 (0 to 21)	2 (0 to 4)
Outcome unknown	45/1343 (3%)	57/316 (18%)	2/1027 (¡1%)

Treatment

Therapeutic intervention data, including cases where an absence of treatment was explicitly indicated, were available for 306/1343 (23%) patients. Of these, 271/306 (89%) received a high-efficacy antimicrobial, either alone or in combination with other antibiotics. A further 24/306 (8%) were treated with antibiotics not classified as highly efficacious, while 11/306 (4%) did not receive any antibiotic treatment (**Table 1.4**).

Streptomycin was the most commonly administered high-efficacy plague treatment, given to 141/306 (46%) patients either alone or in combination with other drugs. This was followed by gentamicin, tetracycline, and doxycycline, administered to 91/306 (30%), 85/306 (28%), and 42/306 (14%) patients, respectively. All other drugs were used in a small number of cases.

Among patients who received a high-efficacy antimicrobial at any time after presentation, 15/271 (6%) died. In contrast, 6/24 (25%) of those treated with only other antibiotics and 5/11 (45%) of those who received no antibiotic treatment died.

Risk of bias assessments

Table 1.4: Summary of treatments received by number of patients and number of deaths

	n (%) patients	**N deaths (CFR)**
High efficacy antibiotics	271 (89%)	15 (6%)
Streptomycin	141 (46%)	6 (4%)
Gentamicin	91 (30%)	11 (12%)
Levofloxacin	1 (<1%)	0 (0%)
Doxycycline	42 (14%)	2 (5%)
Ciprofloxacin	8 (3%)	0 (0%)
Tetracycline	85 (28%)	3 (4%)
Chloramphenicol	5 (2%)	5 (7%)
Amikacin	1 (<1%)	0 (0%)
Tobramycin	1 (<1%)	0 (0%)
Trimethoprim-Sulfamethoxazole	7 (2%)	0 (0%)
Other antibiotic treatment	24 (8%)	6 (25%)
No antibiotic treatment	11 (4%)	5 (45%)

Note: Other antibiotic treatments include antiobiotics that are not included in the U.S. CDC treatment guidelines for bubonic plague (2021) [76], such as: penicillin, cefotoxin, ceftriaxone, amoxicillin, cotrimoxazole, oxicillin, cefadroxil, cefalosporin, vancomycin, ticarcillin, methicillin, ampicillin, cephalexin, erythromycin, metronidazole, linezolid, tobramycin, cefoxatin, flucloxacillin, cefotaxime, cefuroxime, chloromyectin, azithromycin, clindamycin, piperacillin, nafcillin, cefalothin, sulphapyridine, sulfathiazole, sulphazidine, cephaloridine, sulfisoxazole, carbenicillin, dicloxacillin.

The risk of bias was moderate across all study types (**Figure 1.10**). Regarding the key information relevant to this systematic review – such as who is affected by plague, symptoms on first presentation, clinical progression, and outcomes – case reports often provided a well-documented summary of clinical characteristics and diagnostic testing. However, they contained limited demographic data, making it difficult to fully understand the populations affected. In contrast, this information was more challenging to extract from other study types. Interventional and cohort studies typically provided reliable descriptions and assessments of outcomes, although the completeness of follow-up data was frequently unclear.

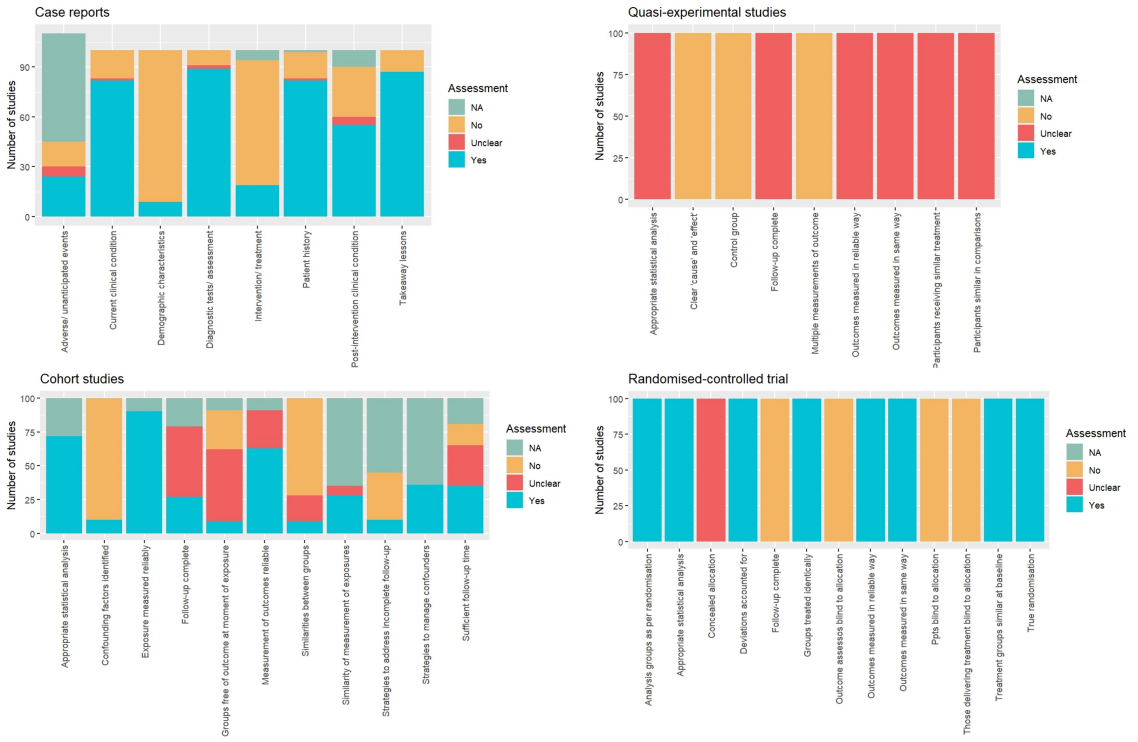


Figure 1.10: Summary of risk of bias assessments

Discussion

This systematic review aimed to summarise the clinical profile of patients with bubonic plague as described in academic literature, with the aim of identifying endpoints suitable as primary efficacy endpoints in late-stage therapeutic trials.

Limitations of this review

Although this systematic review compiled extensive data on the signs, symptoms, and outcomes of confirmed plague cases, it is limited to studies published in English. Consequently, important data from non-English sources, such as studies conducted in regions that have consistently reported cases of plague like China, Russia and Mongolia, are not included. For example, while plague cases have been well-documented in China [124], none of these studies were eligible for inclusion, and only one study describing two cases from Mongolia was reviewed.

Mortality as a primary efficacy endpoint

The overall case fatality ratio (CFR) across the reviewed patient population was 15%, suggesting that mortality alone may not be a feasible primary endpoint for clinical trials.

Based on crude calculations, a trial powered to detect a 50% reduction in mortality (which may be unrealistically high) would require an estimated sample size of 203 patients per arm (assuming a significance level of 5%, 80% power, and 10% lost to follow-up). This large sample size would likely be unattainable for a disease like plague, considering its wide geographic spread and relatively low case numbers, and would result in trials requiring several years to complete. Furthermore, smaller and more reasonable effect sizes would produce even larger sample sizes.

Furthermore, the CFR reported in this review may not reflect contemporary global trends in bubonic plague epidemiology. A significant proportion of the included studies originated from the United States between the 1960s and 1980s, while 80% of global plague cases in 2018 occurred in Madagascar and 18% in the Democratic Republic of Congo (DRC) [31]. Differences in healthcare accessibility and resources between these regions are likely to influence disease presentation and outcomes. Furthermore, Madagascar reported a CFR of 15% for bubonic plague cases between 1998 and 2016 [125], but this figure may be affected by limitations in diagnostic accuracy and regional sampling inconsistencies.

Alternative endpoints for plague trials

Given the limitations of using mortality as a primary endpoint, alternative or composite clinically relevant endpoints must be considered. The only randomised controlled trial included in this review used a composite endpoint – "cure or improvement in condition" – defined by resolution of fever, resolution of bubo pain and swelling, and recovery from pneumonia or other plague symptoms [58]. Similarly, the IMASOY trial (which was ongoing at the time of this review) used a composite endpoint encompassing survival, resolution of fever, a minimum 25% decrease in bubo size, no alternative plague treatment, and no extension of treatment beyond day 10 [1]. Both studies identified overall clinical recovery as their primary outcome of interest and included resolution of fever and bubo resolution or reduction in its definition.

Composite endpoints offer advantages by increasing the number of outcome events and improving the feasibility of detecting statistically significant treatment effects. However, they also introduce complexities in interpreting trial results. The clinical relevance of individual components of composite endpoints, such as fever resolution or bubo size reduction,

remains unclear. For example, fever resolution was documented for only 14% of patients in this review, and while the resolution rate (87%) was high, its relationship with overall clinical status remains insufficiently studied.

Additionally, the weighting of vastly different clinical events – such as fever resolution versus pneumonia recovery – within composite endpoints complicates interpretation. For instance, if Patient A recovers from fever alone while Patient B recovers from fever, pneumonia, and bubo-related symptoms, the composite endpoint may not adequately reflect differences in disease severity or treatment efficacy.

Need for improved disease characterisation

A clearer understanding of the relationship between individual signs, symptoms, and overall clinical status is critical for defining robust trial endpoints. This includes both physiological indicators of disease progression and laboratory measures, such as *Y. pestis* detection in bubo samples. Unfortunately, the limited longitudinal data available for plague patients currently prevent the identification of a single or composite endpoint that is both clinically relevant and achievable within reasonable sample sizes.

Challenges and broader implications

This review highlights how limited disease characterisation restricts the design of clinical trials for rare infectious diseases such as plague. The lack of comprehensive data on the clinical course and outcomes of plague not only complicates endpoint selection but also hampers the interpretation of trial results. These challenges persist despite previous interventional trials for plague having taken place, leaving questions about optimal treatment unresolved.

Although innovative trial designs could help address these limitations, the broader challenge lies in incentivising research and development (R&D) for diseases like plague, where pharmaceutical profitability is limited. Addressing this challenge will require coordinated efforts to prioritise neglected diseases, streamline trial designs, and foster collaborations to generate critical evidence for improving treatment outcomes.

Author contribution to "A systematic review of the clinical profile of patients with bubonic plague and the outcome measures used in research settings"

Role description	Author contribution
Conceptualisation	X
Methodology	X
Software	-
Validation	X
Formal analysis	X
Investigation	X
Resources	-
Data curation	X
Writing (original draft)	X
Writing (review and editing)	X
Visualisation	X
Project administration	X
Funding acquisition	-

Publications

This chapter is based on the following publication:

- **Bourner J**, Andriamarohasina L, Salam A, Kayem ND, Randremanana R, Olliaro P. A systematic review of the clinical profile of patients with bubonic plague and the outcome measures used in research settings. PLoS Negl Trop Dis. 2023 Nov 9;17(11):e0011509. doi: 10.1371/journal.pntd.0011509. PMID: 37943880; PMCID: PMC10662759.

1.1.3 Contribution of the chapter

Both bubonic plague and Lassa fever illustrate the complexities of conducting clinical trials for rare and neglected diseases. The shared challenges of variability in presentation, limited longitudinal data, and contextual differences underscore the need for innovative, pragmatic trial designs that can account for uncertainties in the frequencies of important patient events. Addressing these challenges is essential to accelerate the development of effective therapies for these diseases. Below is a summary of the themes that have emerged in the systematic reviews for both diseases.

Heterogeneous clinical presentation

Patients with confirmed plague and Lassa fever present with a highly variable list of signs and symptoms, ranging from mild, non-specific features to severe, life-threatening complications. However, the clinical relationship between these milder symptoms, disease severity, and eventual patient outcomes remains poorly understood. Severe manifestations are observed in a minority of patients included in these reviews, and the underlying factors contributing to their progression are not well-defined.

This variability in clinical presentation and disease progression poses significant challenges for the design of clinical trials. It complicates the definition of consistent endpoints to assess therapeutic efficacy and makes defining inclusion criteria and identifying the need for subgroup analyses particularly difficult, which may be crucial for determining treatment modifications tailored to specific patient populations.

Limited longitudinal data

Both reviews highlight the lack of robust longitudinal studies that track the evolution of signs and symptoms over time. In particular for plague, the morphology and persistence of the bubo was rarely reported beyond baseline. For Lassa fever, the contributing factors to a relatively high CFR remain poorly documented.

The absence of detailed longitudinal data complicate the selection of clinically relevant and measurable endpoints for trials, particularly those assessing recovery. For both plague and Lassa fever it is difficult to determine which clinical outcomes could act as a meaningful and reliable indicator of recovery, and when they are most likely to occur.

Case Fatality Ratios

These systematic reviews have demonstrated that mortality is an infrequent event in both diseases(15% for plague and 30% for Lassa fever), which is likely to be lower in a clinical trial context with enhanced patient monitoring, making it a difficult endpoint for trials due to the large sample sizes required to detect statistically significant differences.

Role of context and infrastructure

The substantial geographic and contextual differences in clinical outcomes for plague between HICs and LMICs highlights the influence of context-specific factors such as healthcare access, diagnostic capacity, and cultural practices on patient outcomes, necessitating tailored trial designs. As the majority of cases of both plague and Lassa fever occur in LMICs, trials must address contextual challenges, such as limited healthcare and diagnostic infrastructure, while ensuring that findings are applicable to real-world settings.

What do these findings mean for IMASOY?

Based on the findings of the systematic review for plague, it is challenging to determine whether and how the IMASOY endpoint should be amended. However, it is clear that the relationship between bubo presence and size, and the overall clinical status of the patient requires further investigation. Subsequent chapters will explore this relationship, in addition to the extent of potential measurement bias that could be introduced by using a digital caliper to measure the bubo.

What do these findings mean for Lassa fever?

As several therapeutics for Lassa fever may become available for testing in the next one to two years, it may not be feasible to wait for large, multi-country studies to generate the necessary data on the frequencies of signs and symptoms required to design reliable trial endpoints. In the absence of this data, other methods to identify key patient outcomes and define trial endpoints need to be considered. Subsequent chapters will summarise the use of a consortium approach based on clinician and researcher expertise to develop a trial design, including its endpoints, to address this gap.

Chapter 2

Challenges in Implementing Endpoints in Clinical Trials: Assessing Measurement Instrument Selection for Bubonic Plague

2.1 Introduction

Identifying a clinically meaningful outcome however is just the first step; the next challenge pertains to the selection of an appropriate measurement instrument. The chosen instrument must be valid – accurately measuring what it is intended to measure – and reliable – producing consistent results – within the target population.

Checklists and frameworks, such as the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) [126], are available to guide researchers in selecting valid and reliable measurement instruments. However, these methodologies depend on an extensive body of published research on the validity and reliability of potential measurement instruments. According to the COSMIN framework, researchers should first

consider conceptual factors, such as the target population and the setting in which the outcome will be measured. This is followed by systematic reviews and literature searches to identify existing tools that may be suitable. Finally, a quality assessment of validity and reliability studies should be conducted to ensure their appropriateness.

COSMIN has been successfully applied to identify measurement tools for diseases such as allergic rhinitis [127], prostate cancer [128], and pancreatitis [129], for which numerous measurement instruments have been evaluated in individual clinical studies and synthesised through systematic reviews.

For outbreak-prone infectious diseases however such research is scarce or non-existent. Specific disease outcomes that have been used in trials (regardless of their clinical meaningfulness), such as changes in bubo size [1] or mpox lesion status [130], lack consensus on the methods of assessment that should be used. For mpox, there is no clinical agreement on the precise physiological features of lesions at each stage of their evolution, and lesion evaluation may be highly subjective. [131] For plague, bubo measurement is infrequently performed in a clinical setting – in Madagascar descriptive comparisons are made to roughly indicate size (e.g. "petit pois", "lytchi", etc.), although a small number of trials and case reports conducted in other settings have described the use of a digital calliper, ultrasound, galium scan or CT scan to do so. [1] [120] [121] [122]; no studies however have been conducted to evaluate either the validity or reliability of these methods. The lack of available validated tools for clinical research is particularly critical during outbreaks, when there is no time to conduct validation studies, underscoring the importance of using inter-outbreak periods to identify and validate measurement instruments in preparation for future trials.

A further complication is that many outbreak-prone infections affect those living in communities with limited access to biomedical health care. [132] Even when valid and reliable instruments exist for measuring key physiological functions, they may not be accessible in resource-limited settings, due to cost, availability or lack of staff training to use specialist equipment. Ultrasound and CT scan facilities are examples of this challenge. While both of these tools may be able to produce accurate measurements of a plague bubo,

for example, this equipment is not available in the rural health facilities in Madagascar at which plague cases arise.

Nevertheless, clinical trials for outbreak-prone infectious diseases have taken place in the absence of validated measurement tools. As previously discussed, the IMASOY trial uses a digital caliper to measure bubo size [1] and several clinical trials for mpox have been initiated using "time to lesion resolution" – a clinical assessment of a lesion's status – as a primary endpoint. [133], [134] [130] This chapter evaluates the impact and risks of using such non-validated tools and assessments in clinical trials.

This chapter focuses on the validity and reliability of using a digital caliper to measure bubo size. In the first section, the results are presented of a large-scale study using artificial buboes to evaluate the accuracy and inter- and intra-rater measurement reliability of the caliper. While the artificial buboes facilitated the collection of a large dataset and provided sufficient insight in to inter- and intra-rater reliability, they were an imperfect surrogate for real buboes, which can be challenging to detect through palpation, located in areas of the body that present measurement challenges and many patients are small children (who may not remain still for the duration of the examination). A subsequent study, presented in the second section of this chapter, was therefore conducted to evaluate the validity of a digital caliper in a clinical setting with real buboes. However, unlike the study in which artificial buboes were created, it was not possible to know the "real" size of the buboes in the clinical study, making the assessment of measurement accuracy challenging. As ultrasound is commonly used in other areas of patient care to measure for example foetal anatomy, caliper measurements were compared to those taken by ultrasound, which was treated as the "gold standard". This study also presented a unique opportunity to explore the clinical meaningfulness of bubo size through its relationship with the overall clinical status of the patient.

By addressing these issues, the studies presented in this chapter highlight the complexities of selecting appropriate measurement instruments for outbreak-prone infectious diseases, particularly those that are conducted in challenging field conditions. By evaluating the validity and reliability of a digital caliper for measuring bubo size in plague, this chapter underscores the critical need for rigorous validation of measurement tools

during inter-outbreak periods. Establishing standardised and reliable outcome measures is essential to improving the quality of clinical trials and subsequently ensuring that research findings can meaningfully inform patient care and public health responses in future outbreaks.

2.2 Bubonic plague: can the size of buboes be accurately and consistently measured with a digital calliper?

Although the relationship between bubo size and overall clinical status has not yet been established, bubo size has been used to evaluate treatment response in two clinical trials. [1] [2] [58] Changes in bubo size and morphology have also been reported in clinical settings, although such reports are infrequent and inconsistent. [135]

However, only three studies in the existing literature have documented the measurement instruments used to evaluate bubo size. One case study conducted in the United States used ultrasound and CT imaging [120] and two others used a galium scan [121] [122], while the IMASOY trial [1] [2], which was conducted in remote rural villages in Madagascar, used a digital caliper.

While ultrasound and CT imaging likely provide more precise measurements of bubo size compared to digital calipers, their implementation at scale in a resource-limited health system, such as that of Madagascar, presents significant challenges for a clinical trial. The cost of the equipment, the need for specialised training, and infrastructural limitations necessitated the use of more affordable tools that are easier to deploy in a clinical trial. There are however no data informing the validity or reliability of the digital caliper approach, and it is possible to assume such tools – that rely on manual manipulation and subjective visual assessment of the bubo circumference – are prone to measurement error, which could prevent the accurate assessment of changes in bubo size over time. The scale of the potential measurement error has not however been quantified.

The aim of this study was therefore to assess the validity and reliability of a digital caliper to measure bubo size. However, as a large dataset of repeated measurements were required for this study, it was considered both unfeasible and unethical to measure real

buboes, due to the wide geographic dispersion of cases and the inconvenience and discomfort the measurement process may cause patients. The study was therefore conducted using artificial buboes that were created in different sizes, shapes and densities to mimic the various permutations of real buboes.

The results of this study were expected to inform decisions about amending the primary endpoint of the IMASOY trial and provide valuable insights for other trialists and clinicians who may record repeated bubo measurements over time.

2.2.1 Methods

The results of this study are reported according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) framework. [136]

Objectives

The primary objectives of this study were to (i) evaluate the accuracy of individual measurements against the true size of the artificial bubo when using a digital caliper and (ii) determine whether the characteristics of the artificial bubo influenced measurement accuracy.

In clinical and research settings, multiple measurements may be taken by either a single rater or multiple raters to track how a bubo changes in size over time. Understanding the reproducibility of these measurements is essential for ensuring reliability. The variation observed in repeated measurements of the same bubo could provide insights into the potential measurement error inherent in using a digital caliper, especially in the absence of more advanced technologies, such as ultrasound.

Preparation of artificial buboes

The study used 14 artificial silicone buboes designed to simulate clinical conditions. Bubo sizes were selected based on a review of data reported to the Institut Pasteur de Madagascar (IPM) by clinicians experienced in treating plague, which indicated that approximately three-quarters of buboes have a long axis within the 35-55mm range. To replicate the challenging conditions often encountered in clinical practice, whereby buboes are poorly delimited and embedded within surrounding oedema and inflammation, the artificial buboes were created in seven different sizes, with long axes ranging from 29.8 mm

to 81.8 mm. Variations included using hard or soft silicone and the presence or absence of an additional silicone layer representing oedema (Table 2.1). The same set of artificial buboes was used across the two training sessions in which study technicians were asked to complete the measurement exercise.

Table 2.1: Dimensions and characteristics of artificial buboes.

Artificial bubo ID	Total size	Density	Oedema
1	121.6	Hard	No
2	121.6	Soft	Yes
3	120.3	Hard	No
4	120.3	Soft	Yes
5	120.3	Soft	No
6	94.7	Hard	Yes
7	94.7	Hard	No
8	77.8	Hard	No
9	77.8	Soft	Yes
10	77.8	Soft	No
11	64.9	Hard	No
12	64.9	Soft	No
13	52.7	Hard	No
14	52.7	Soft	No

The artificial buboes were created by first preparing a set of standardised clay buboes with the axes as described in **Table 2.1**. A negative volume mould was then prepared using Crystacal R Casting Plaster and the soft and hard buboes were cast using Ecoflex™ 00–10 and Ecoflex™ 00–20 (Smooth-On, Inc. Macungie, PA). A second negative volume mould was created by spreading a thin layer of Dragon Skin FX- Pro™ (Smooth-On, Inc. Macungie, PA) over a skin-textured vinyl sheet. The resulting skin-textured latex sheet

was overlaid on a clay form and this was used to make the outer bubo Crystacal R Casting Plaster mould. The outer bubo mould was coated in Dragon Skin and, once the initial layer has cured, a bubo was placed in the cavity and overlaid with more Dragon Skin. The oedema effect was achieved by adding a layer of Soma Foama™ 15 around the bubo and allowing it to cure before overlaying with Dragon Skin (**Figure 2.1**).

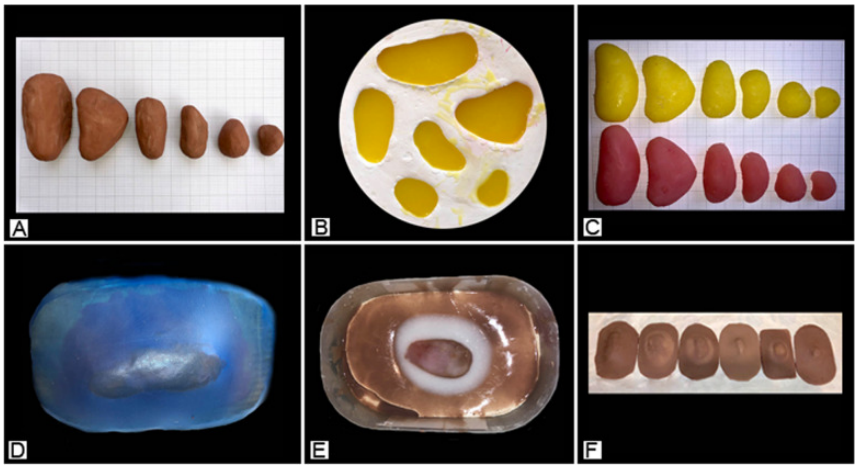


Figure 2.1: Artificial bubo preparation. A Standardised clay buboes. B Casting the buboes. C Set of soft (yellow) and hard (pink) buboes. D The outer bubo mould. E Casting a bubo in the outer mould with a layer of Soma Foama™ 15 to create the oedema effect. F A competed set of artificial buboes

Measurement process

Two training sessions were conducted in August 2020 and August 2021, during which raters measured each artificial bubo once using a digital caliper. They recorded the long- and short-axis measurements (in mm) on a standardised data collection form, which was subsequently transferred into a Microsoft Access database. All measurements were recorded individually, and raters remained blinded to each other's measurements.

The procedure for measuring artificial buboes followed the IMASOY trial protocol and the relevant standard operating procedure (**Figure 2.2**):

- Each bubo was palpated to assess its shape and locate the long and short axes
- The long axis was identified first, and the short axis was defined as the longest measurement perpendicular to the long axis

- To measure the long axis, an eyeliner pencil was used to mark a line bisecting the bubo, starting a couple of centimeters above and below the identified axis and stopping at the margins of the swelling
- This process was repeated for the short axis
- A digital caliper was then used to measure the distance between the pencil marks on each axis, and the measurements were recorded in millimeters.

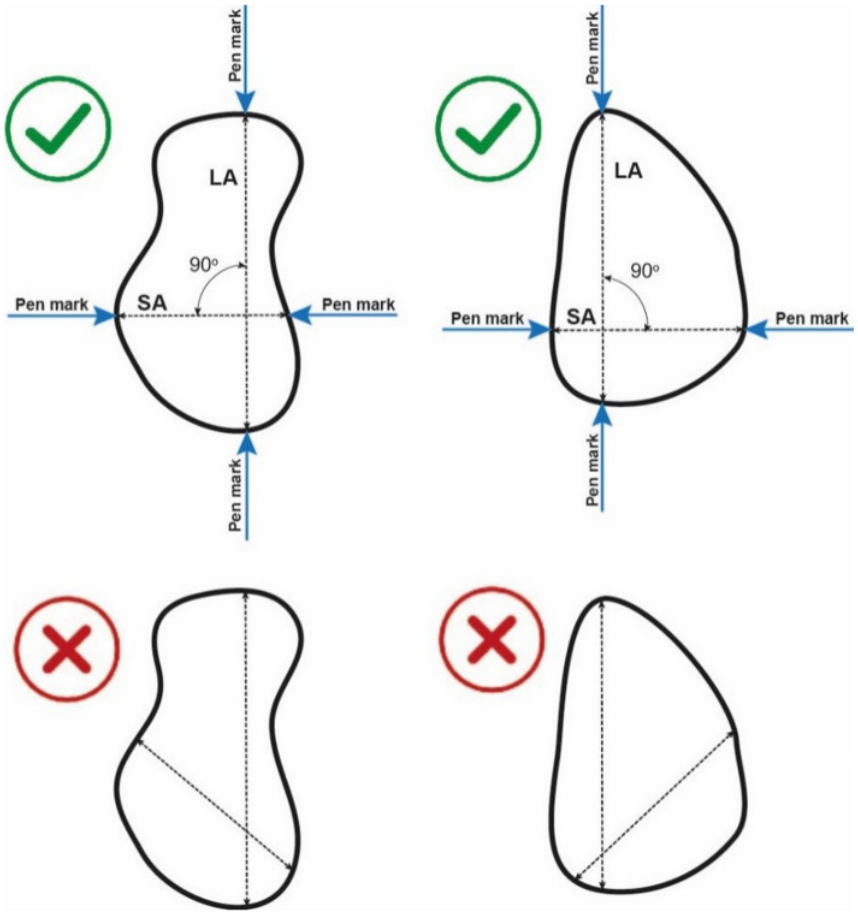


Figure 2.2: Bubo measurement method

Participants and sampling methods

The raters in this study were clinical study technicians (TECs) , who hold at least a nursing qualification, employed by the IMASOY trial to support study activities at trial sites. As part of their responsibilities in the IMASOY trial, TECs are tasked with measuring the buboes of patients enrolled in the study. To ensure consistency and accuracy, they

receive annual trial-specific training, including a dedicated session on bubo measurement techniques using artificial buboes. Prior to working on the IMASOY trial, most TECs had limited or no experience measuring buboes in a clinical setting using a digital caliper.

As this study was integrated into the annual training sessions of the ongoing IMASOY clinical trial, a convenience sampling approach was used. All TECs attending the training were asked to participate in the study. In August 2020, 28 TECs participated in the training, and in August 2021, 29 TECs participated. Of these, 20 TECs attended both training sessions (**Figure 2.3**). Data collected from both sessions were included in this analysis.

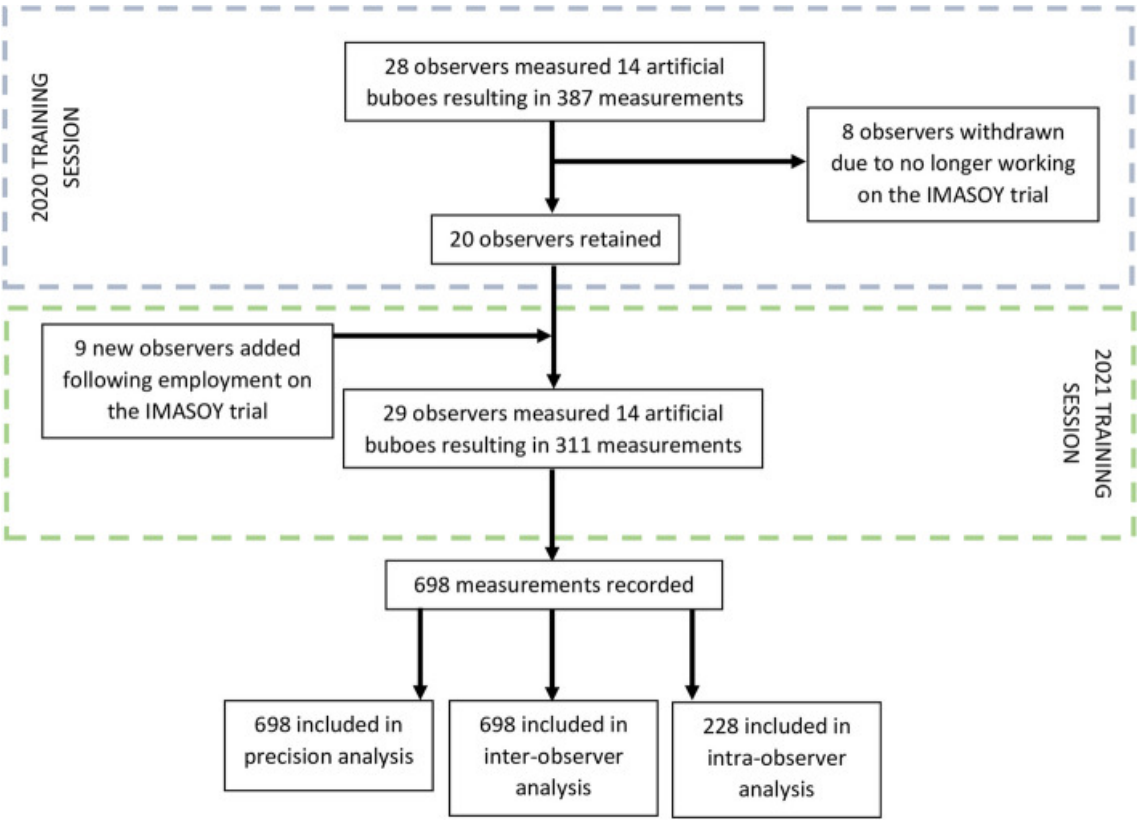


Figure 2.3: Flowchart of rater participation and recorded measurements per training session

Analysis

The following analyses were carried out using RStudio v. 1.3.1093. Where p-values are reported, we have used a significance level of 0.05.

Three types of analyses were performed to assess measurement accuracy (validity) and inter- and intra-rater agreement (reliability). Validity was defined as the degree to which an individual rater's measurement aligned with the true size of the artificial bubo. This was evaluated using the percentage difference and absolute difference (in mm) between the measured and true sizes. Intra-rater agreement assessed the consistency of repeated measurements by the same rater, while inter-rater agreement examined the consistency of measurements across multiple raters.

Accuracy: An analysis was conducted to evaluate the accuracy of the measured artificial bubo sizes compared to their true sizes (**Table 2.1**). For each recorded measurement, the absolute difference (in mm) and percentage difference from the true size were calculated.

Measurement accuracy was assessed by calculating the median percentage difference of all measurements and by summarising the number of measurements falling within intervals increasing in 5% increments from the true size. To facilitate interpretation, measurements within 5% and 25% intervals are highlighted.

To assess the impact of bubo characteristics –density, size, and presence of oedema – on the percentage difference between measurements and the true size, a linear mixed-effects regression model was fitted. The model included a random effect for rater ID and fixed effects for density (hard or soft), presence of oedema (present or absent), year of measurement (2020 or 2021), and size (modelled with five dummy variables for the six bubo sizes). Regression coefficients with 95% confidence intervals (CIs) were presented to illustrate the mean change in measurement difference associated with each characteristic.

Scatterplots were also generated to visualise the percentage difference between each measurement and the true size of the artificial bubo according to size, presence of oedema, and density, with a regression line and 95% CI.

Intra-rater agreement: Intra-rater agreement was evaluated across all artificial bubo measurements and by characteristic (density, size, and presence of oedema). To determine the consistency of measurements taken by the same rater one year apart, the percentage of second measurements falling within 5% intervals of the first measurement of the same artificial bubo was calculated.

The effect of bubo characteristics on the percentage difference between first and second measurements was analysed using a linear mixed-effects regression model. The results are presented in scatterplots. Additionally, the mean absolute difference and standard deviation between first and second measurements are reported.

Bland–Altman plots are used to visualise intra-rater agreement. These plots display each rater's average measurement (measurement 1 + measurement 2) against the absolute difference between the two measurements, with lines indicating the mean difference and limits of agreement (mean difference ± 1.96 SD).

As each artificial bubo was measured only once by each rater during each training session, the intra-rater analysis included only data from the 20 raters who participated in both training sessions. Data from raters who attended only one session were excluded.

Inter-rater agreement Inter-rater agreement, defined as the extent to which different raters provided similar measurements of the same artificial bubo, was assessed by calculating the percentage of measurements falling within 5% intervals of other raters' measurements. This analysis was de-emphasised as the same rater was required to record measurements of the same bubo during each patient visit in the IMASOY trial.

2.2.2 Results

There were 698 measurements in total recorded across the two training sessions by 57 raters, of whom 20 recorded measurements in both training sessions (**Figure 2.3**). The median number of artificial buboes measured per rater per year was 12 (range: 5 - 14) and, of the 57 raters, 29 (51%) measured all 14 artificial buboes.

Accuracy

The median percentage difference of all artificial bubo measurements compared to the reference bubo size was 13%, with a range from 0 to 171% (**Table 2.2**). Of the total measurements, 179 (26%) measurements fell within 5% of the reference size, and 508 (73%) were within 25% of the reference size.

Table 2.2: Summary of measurements

Artificial bubo characteristic	**N recorded measurements**	**Median absolute difference (range), mm**	**Median absolute difference (range), %**
Overall	698	12.0 (0.01 to 89.92)	13% (0 to 171%)
Hard	349	12.3 (0.01 to 89.92)	13% (0 to 171%)
Soft	349	11.9 (0.01 to 76.41)	13% (0 to 145%)
Oedema	213	10.3 (0.07 to 72.95)	10% (0 to 94%)
No oedema	485	12.6 (0.01 to 89.92)	15% (0 to 171%)
Size: 121.6mm	108	10.7 (0.32 to 73.04)	9% (0 to 60%)
Size: 120.3mm	160	9.7 (0.14 to 56.19)	8% (0 to 47%)
Size: 94.7mm	88	9.7 (0.07 to 68.19)	10% (0 to 72%)
Size: 77.8mm	150	12.1 (0.26 to 82.51)	16% (0 to 106%)
Size: 64.9mm	86	11.8 (0.01 to 46.84)	18% (0 to 72%)
Size: 52.7mm	106	17.6 (0.01 to 89.92)	28% (0 to 171%)
Year: 2020	387	10.69 (0.01 to 89.92)	10% (0 to 171%)
Year: 2021	311	15.75 (0.01 to 82.51)	17% (0 to 139%)

The majority of raters overestimated the size of the artificial buboes compared to their true size. Of the recorded measurements, 507 (74%) were larger than the true size, while 180 (26%) were smaller. Among individual raters, 47 (82%) more frequently overestimated bubo sizes than underestimated them.

Both the size of the artificial buboes and the year of training significantly influenced measurement accuracy. Measurement error decreased as the size of the artificial buboes increased (**Figure 2.4**). For the three largest artificial buboes (with true sizes of 121.6 mm, 120.3 mm, and 94.7 mm), the median measurement error ranged from 8% to 10% (**Table 2.2**). In contrast, the smallest artificial bubo exhibited the highest measurement error, with a median error of 28%.

The year of measurement also had a significant effect on accuracy. Measurements conducted in 2020 were more accurate than those in 2021 (**Table 2.2**), with statistically significant evidence of overestimation in 2021.

Neither the presence of edema nor the density of the artificial buboes significantly influenced measurement accuracy.

Intra-rater agreement During the two training sessions, 20 raters measured the artificial buboes twice, generating a total of 239 pairs of measurements. The median percentage difference between the first and second measurements of the same artificial bubo was 11%, ranging from 0% to 129% (**Figure 2.5**). The median absolute difference between the first and second measurements was 11.84mm, with a range of 0.03mm to 76.87mm.

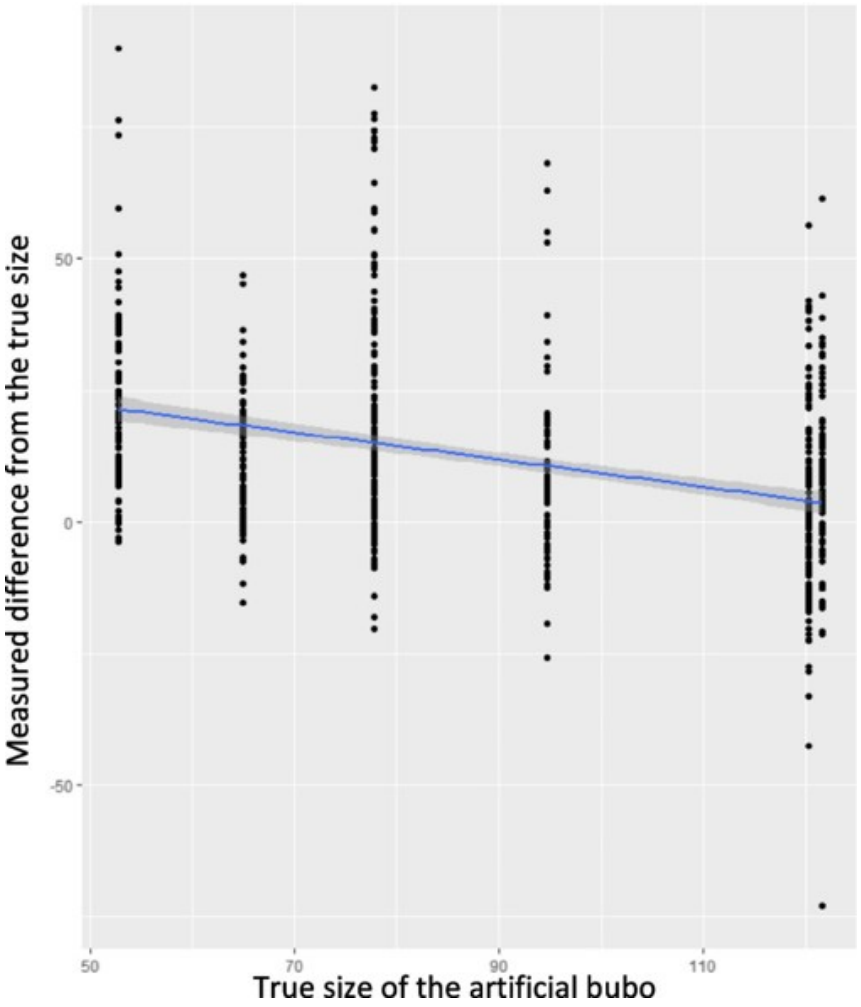


Figure 2.4: Scatterplot of percentage difference of each measurement from the true size plotted against the true size for per characteristic (size and presence of oedema)

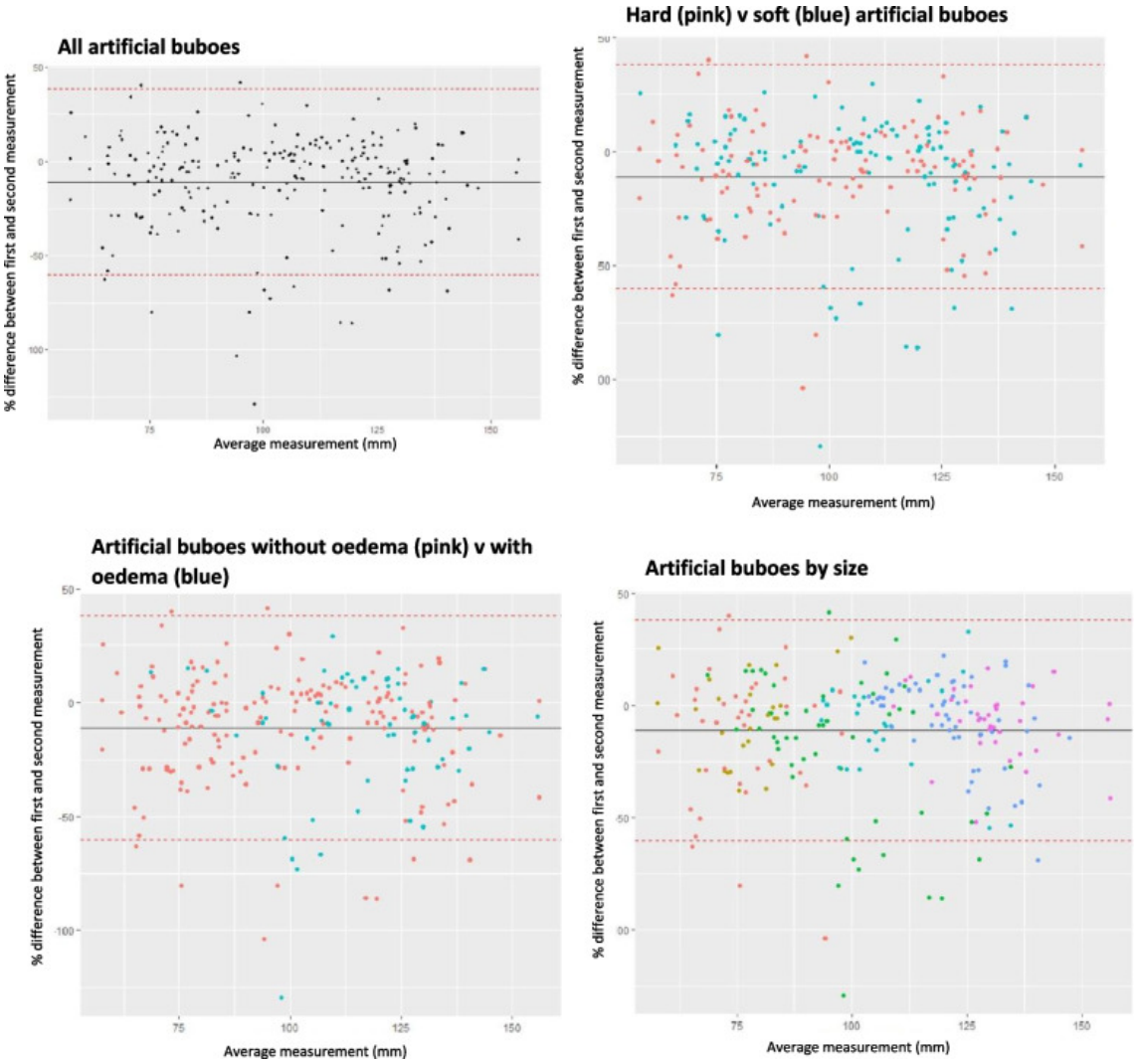


Figure 2.5: Bland–Altman plots showing the intra-rater agreement of the average measurement and percentage difference between measurements by artificial bubo characteristic

On average, each rater measured 12 of the 14 artificial buboes (86%) twice. Of these second measurements, 3 (25%) were within 5% of the corresponding first measurement, while 9 (75%) were within 25%.

For hard artificial buboes, 95% of second measurements fell within 55% of the first measurement. In comparison, 95% of second measurements for soft artificial buboes were within 70% of the first measurement. Similarly, 95% of second measurements for artificial buboes with oedema were within 60% of the first measurement, while those without oedema were within 55%.

A linear regression analysis revealed that the size of the artificial bubo had a statistically significant effect on intra-rater agreement. Agreement between first and second measurements increased as the size of the artificial bubo increased (**Figure 2.6**). Neither density nor the presence of oedema had a statistically significant impact on intra-rater agreement.

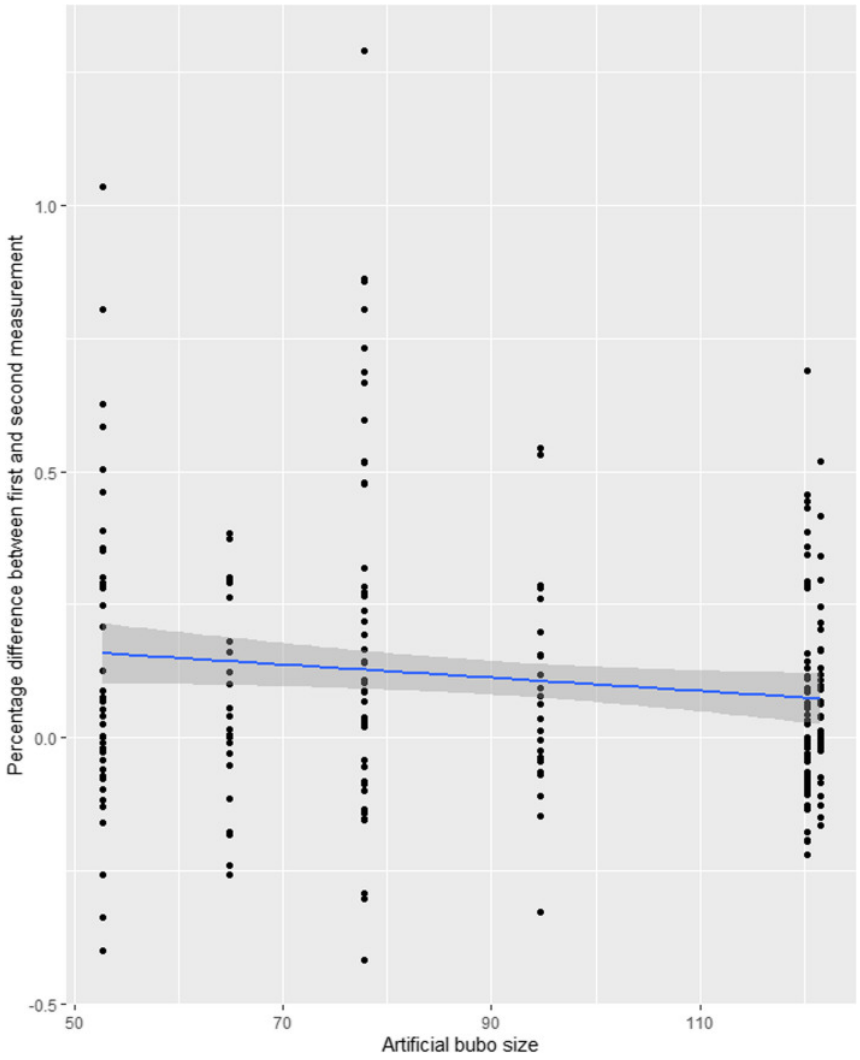


Figure 2.6: Scatterplot of the percentage difference between the first and second measurements of the same artificial bubo plotted against the true size of the artificial bubo

Inter-rater agreement The median standard deviation between measurements of the same artificial bubo by different raters was 17.1mm, ranging from 10.3mm to 23.0mm. Seventy-five percent (75%) of raters recorded measurements within 25% of another rater

who measured the same artificial bubo, while approximately 20% of raters recorded measurements within 5% of each other.

None of the bubo characteristics—density, size, or presence of edema—had a statistically significant effect on the raters' ability to measure the artificial buboes within 25% of each other's measurements.

2.2.3 Discussion

This study aimed to assess the validity and reliability of a digital caliper to measure artificial buboes, with the goal of evaluating its potential for use in clinical research settings. Additionally, the study explored the implications of employing this tool to assess changes in bubo size as part of evaluating treatment response.

Overall, nearly three-quarters of raters were able to measure the artificial buboes within 25% of their true size, with a median measurement difference of 13%. However, this also means that approximately one-quarter of the measurements deviated from the true size by more than 25%, with some instances of substantial error (up to 170%).

Measurement accuracy was influenced by both the year in which the measurement was made and the size of the artificial bubo. Larger buboes were measured with greater accuracy, while smaller buboes tended to be measured with greater error. The tendency to over-estimate the size of smaller buboes poses a particular challenge when using bubo size reduction as an indicator of treatment response. As buboes are expected to decrease in size over time and smaller buboes are more likely to be over-measured, changes in the size of buboes from baseline to end of treatment are more likely to be underestimated, potentially leading to inaccurate conclusions about treatment efficacy.

Although using the same rater to repeat measurements over the course of treatment may mitigate this risk, a significant proportion of the intra-rater comparisons – around one-third – showed discrepancies of more than 25%, either larger or smaller than the original measurement. Intra-rater agreement was also influenced by bubo size – again with smaller buboes yielding greater disagreement. This reinforces the potential risk of inaccurately estimating treatment effects, particularly when bubo size is used as an indicator of recovery.

A limitation of this study is that measurements were taken from artificial buboes, which do not fully replicate all the variables that may be encountered in clinical cases of bubonic plague. Additionally, the artificial buboes may not accurately reflect the full range of clinical scenarios, particularly given the challenges of measuring painful buboes or buboes in difficult anatomical locations. Thus, while the data provide useful insights, caution is needed when extrapolating these findings to clinical cases.

This study was not designed to assess the prevalence of incorrect determination of pre-specified reductions in bubo size, which would be critical for evaluating treatment response. However, the results do provide valuable parameter data that could inform the design of simulation studies. These studies could estimate (i) how often a specified percentage reduction in bubo size would be correctly classified under various measurement error assumptions and (ii) the potential impact of measurement error on treatment effect estimation in randomised controlled trials.

The findings of this study highlight that, despite most measurements being reasonably accurate, a substantial proportion exhibited significant error, particularly for smaller buboes, underscoring the potential for measurement inaccuracies using a digital caliper to distort assessments of treatment response. These findings emphasise the need for caution when using digital calipers to evaluate bubo size in both research and clinical practice and highlight the importance of identifying and validating alternative measurement techniques in real-world settings.

Author contribution to "Bubonic plague: can the size of buboes be accurately and consistently measured with a digital calliper?"

Role description	Author contribution
Conceptualisation	X
Methodology	X
Software	-
Validation	X
Formal analysis	X

Investigation	-
Resources	-
Data curation	X
Writing (original draft)	X
Writing (review and editing)	X
Visualisation	X
Project administration	X
Funding acquisition	-

Publications

This chapter is based on the following publication:

- **Bourner J**, Randriamparany R, Rasoanaivo TF, Denis E, Randremanana RV, Vaillant M, Salam AP, Gonçalves BP, Olliaro P. Bubonic plague: can the size of buboes be accurately and consistently measured with a digital calliper? Trials. 2023 Dec 19;24(1):815. doi: 10.1186/s13063-023-07835-7. PMID: 38115024; PMCID: PMC10729355.

2.3 A prospective cohort study to describe the morphology of buboes in patients with bubonic plague using ultrasound imaging

The findings of the previous section highlight some of the challenges that may be encountered when using a digital caliper to measure artificial buboes. However, given that artificial buboes do not accurately mimic conditions encountered in clinical cases of plague, the true scale of the potential measurement error generated by a digital caliper remains unknown. Determining the "true" size of a bubo in a clinical case of plague however is a challenge due to the absence of a validated measurement tool to accurately measure its dimensions. While no validated tool exists specifically to measure buboes, imaging tools commonly used to measure physiological structures in other clinical settings do exist. Imaging tools such as MRI, CT and ultrasound can capture accurate and reliable anatomical measurements and therefore could be used to generate a true size to which measurements recorded using a digital caliper could be compared. In this next section, which describes a prospective cohort to characterise bubo size and morphology, point-of-care ultrasound (POCUS) was used as the reference tool to which caliper measurements were compared. POCUS was selected in this setting due to its portability, which would facilitate the evaluation of patients at multiple sites in rural endemic regions of Madagascar.

As bubo size has been used as an indicator of treatment response, a separate and significant challenge identified in chapter 1 lies in exploring the relationships between the size and morphology of a plague bubo and the overall clinical status of a patient. Understanding this relationship could enhance both plague management and the definition of clinically meaningful clinical trial endpoints.

To address these questions, the primary and secondary aims of this study were to:

- assess the validity of using a digital caliper to measure plague buboes by comparing caliper and POCUS measurements

- describe the evolution and morphology of plague buboes by using POCUS to capture sonographic characteristics

As part of this study, local clinicians and research staff with limited ultrasound experience were trained to use POCUS to conduct lymph node assessments, thus increasing local radiological capacity. A tertiary objective of this study was therefore to evaluate the agreement between ultrasound assessments conducted by clinicians and research staff, and assessments conducted by radiologists.

2.3.1 Methods

This study was a multi-site prospective cohort study of patients with enlarged lymph nodes. Study assessments were conducted at inclusion (D1) and at least one follow-up visit on D11. An additional optional visit was conducted on D4, subject to local capacity.

The study was conducted at health centres in three districts participating in the IMASOY trial [1] [2]: Ambositra, Manandriana and Ambohimahasoa.

Participants in the study were individuals of any age presenting to participating health centers with suspected bubonic plague and regional lymphadenopathy who were willing and able to provide informed consent. Exclusion criteria included cases where the enlarged lymph nodes exhibited signs of suppuration or where high pain scores made ultrasound imaging intolerable.

Following enrolment, demographic information and biological samples were collected from all participants as part of routine care for plague and tested within the national plague surveillance programme. Based on the results, participants were retrospectively classified as confirmed cases or non-cases using the WHO plague case definition.[70]

Data were collected using a standardised paper Case Report Form (CRF) , which was subsequently transcribed into REDCap [110] [137].

Clinical and ultrasound assessments were conducted on D1, D4 and D11. The clinical assessments included a physical examination, measurement of the largest enlarged lymph node (referred to as the index node) detected in each anatomical zone using a digital caliper, and an overall evaluation of the patient's clinical status.

Ultrasound assessments were performed on the index node using a Butterfly iQ+ probe. Still images and cine clips of the index node were captured on D1, D4 and D11 for subsequent QA review by experienced radiologists. Node measurements were obtained using the Butterfly iQ+ measurement tool.

Before the study was initiated, local clinicians and research staff (henceforth referred to as "study sonographers") underwent a targeted ultrasound training programme, covering the use of ultrasound for lymph node evaluation. Two experienced radiologists provided the training and were responsible for the oversight of the study scanning. The training program involved two days of classroom-based training followed by seven days of supervised scanning, during which all images collected were subject to remote quality assurance (QA)
.

All images were independently reviewed by the two radiologists, who subsequently performed a joint review of their assessments at the study's conclusion to resolve any discrepancies in their evaluations. During the QA process, the radiologists completed a short CRF to document whether their assessments differed from those of the study sonographers. In instances where discrepancies were identified, the radiologists' evaluations were used for the final analysis.

Analysis

The analysis was conducted using R v3.6.0.

Participants were retrospectively categorised as "plague cases" or "non-cases" according to the WHO plague case definition following receipt of laboratory results from samples collected as part of routine care under the national plague surveillance programme. [70] In this study, plague cases are those who received either a confirmed or probable diagnosis. Enlarged lymph nodes of plague cases are referred to as "buboes". For non-plague cases, they are referred to as "enlarged lymph nodes".

Clinical and sonographic characteristics of interest were pre-defined on the CRF. The number and proportion of participants with each characteristic at baseline (D1) and D11 are presented in a stacked pyramid chart. An summary of the sonographic structures evaluated in the study can be found in the box below.

Normal and abnormal lymph node structures

An example of a plague bubo with normal structures is shown in Figure 2.7. While enlarged, this lymph node exhibits relatively normal structures including a well-defined, oval-shaped capsule with a distinct echogenic hilum and normal surrounding perinodal tissue.

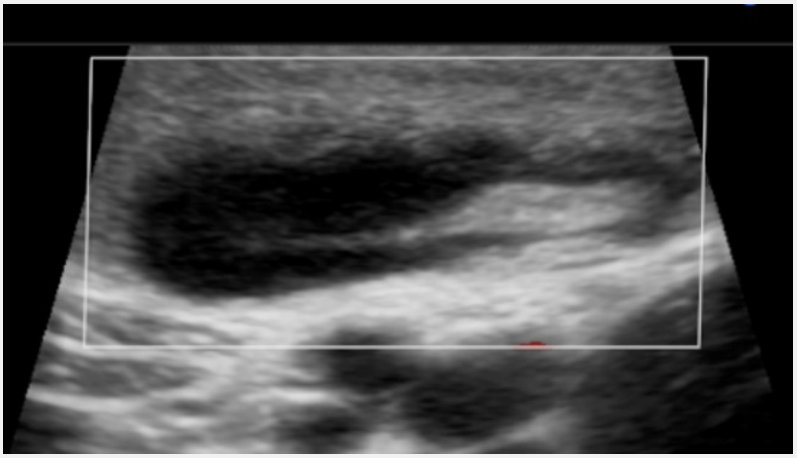


Figure 2.7: An example of a plague bubo with abnormal lymph node structures: oval-shaped; distinct capsule and hilum; normal perinodal tissue

Anomalous findings on ultrasound indicative radiological progression would include blurring to the edges of the capsule (which may be indicative of oedema), a circular-shaped capsule, a decrease in the size of the hilum or a total absence of the hilum (**Figure 2.8**).

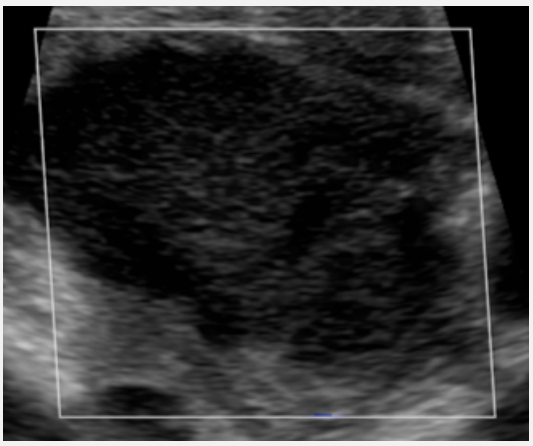


Figure 2.8: An example of a plague bubo with abnormal lymph node structures: round-shaped; absence of hilum; blurred capsule

A short report comparing the clinical and sonographic evolution of two clinical cases is described to highlight findings in the relationship between clinical status and radiological progression. Both cases presented are confirmed cases of plague.

Measurement validity is evaluated by comparing measurements taken by a digital caliper and those taken by ultrasound, which is used as the reference size. Bubo size is the sum of the long axis and short axis. The absolute and percentage difference between each caliper and ultrasound measurement was calculated in order to generate a mean difference overall and per zone (axillary, cervical and inguinal).

To assess agreement in ultrasound evaluations between study sonographers and radiologists, we report the absolute and percentage differences between measurements. The proportion of radiological features correctly identified by the study sonographers in comparison to the radiologists is represented in a heat map.

2.3.2 Results

Between January and March 2024, 16 participants were enrolled in the study, of which 12 were confirmed cases of plague and four were non-cases (**Table 2.4**). An equal number of male and female participants were enrolled in both the plague and non-case groups. The median age of plague cases was 14.5 years (range: 6 - 50 years) and 10.5 years (range: 3 - 23 years) for non-plague cases. Time between symptom onset and enrolment in both groups was one day (range: 1 - 2 days).

Table 2.4: Demographic characteristics

	All	**Plague cases**	**Non-cases**
Total number of cases enrolled	16	12	4
Male : female	8 : 8	6 : 6	2 : 2
Age (years), median [range]	13.5 [3 – 50]	14.5 [6 – 50]	10.5 [3 – 23]
Duration of symptoms before enrolment (days), median [IQR]	1 [1 – 2]	1 [1 – 2]	1 [1 – 2]

Alternative diagnoses for non-plague cases are not known due to limited laboratory testing capacity at the sites involved in the study.

Clinical and sonographic characteristics

At baseline, all participants presented with a single enlarged lymph node. Among confirmed plague cases, seven inguinal and three axillary buboes were observed, compared to two enlarged cervical lymph nodes and two enlarged inguinal lymph nodes in non-cases. Most were visible (83%) and hard (58%), with few buboes exhibiting erythema (17%) or oedema (33%) (**Figure 2.4**). The median pain score of plague buboes at baseline was 6/10.

All participants presented with mild plague symptoms, except one participant who presented with severe symptoms.

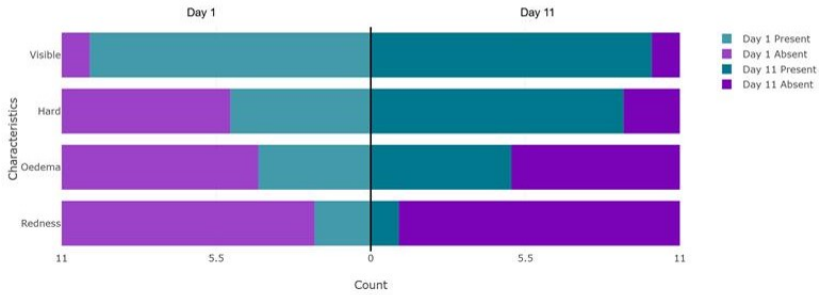


Figure 2.9: Clinical characteristics of plague buboes at D1 and D11

The majority of plague buboes exhibited normal structural characteristics (Figure 2.7) on ultrasound. More than half of the buboes were oval-shaped (67%), had a distinct capsule (56%) and homogeneous cortex (56%), and just under half had a distinct hilum (44%). Few buboes exhibited severe structural anomalies such as necrosis (11%), hypervascularity (22%) and coalescence (11%) (**Figure 2.10**).

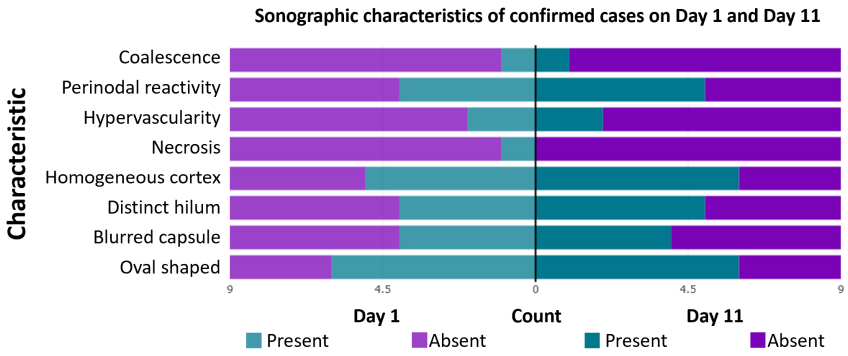


Figure 2.10: Sonographic characteristics of plague buboes at D1 and D11

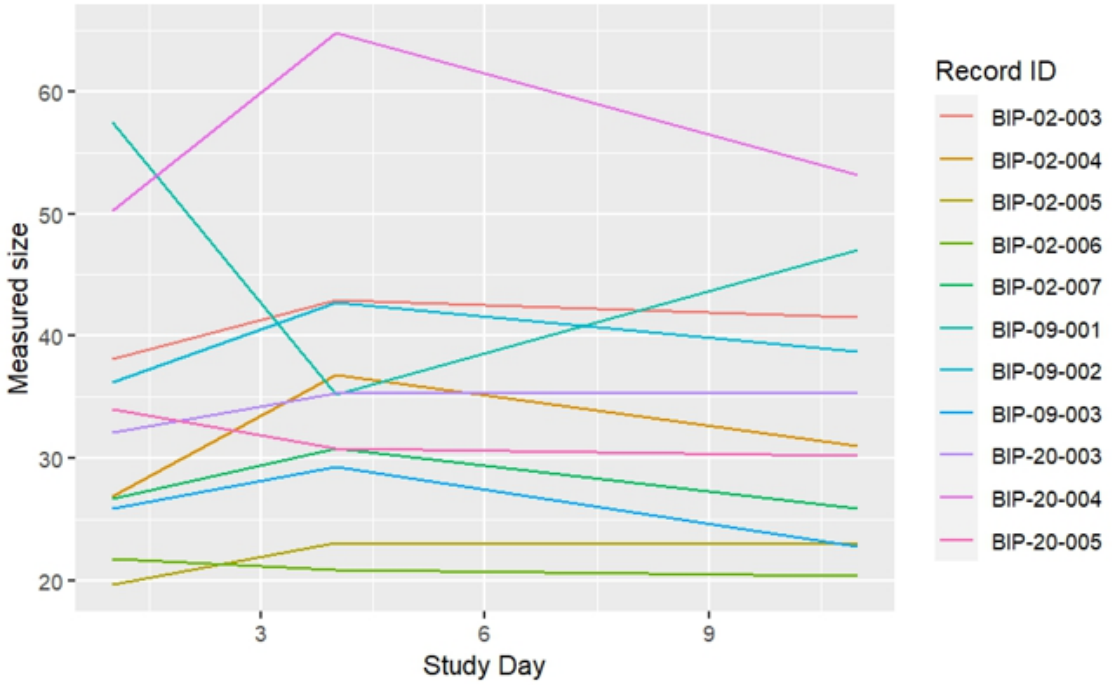


Figure 2.11: Scatterplot of bubo sizes per participant on D1, D4 and D11

By D11, all participants had completed treatment and showed signs of clinical improvement with mild persistent symptoms of plague, including the participant who initially presented in a severe condition. All buboes remained detectable by ultrasound in all plague cases, with two additional buboes detected in a single case. Few changes in clinical and sonographic characteristics were observed (**Figure 2.9** & **Figure 2.10**). The median pain score decreased from 6 to 3/10 (range: 0-6).

There were no detectable enlarged lymph nodes in non-cases on D11.

Bubo size did not substantially change over time (**Figure 2.11**). The median size measured by ultrasound was 30.8mm at baseline and 29.4mm at D11. The median absolute difference between D1 and D11 size was 3.3mm (range: 0.8 – 10.5), representing a median percentage difference of 10%.

Case reports

The clinical and sonographic progression of two cases included in the study is presented below to illustrate the lack of observed relationship between clinical status and sonographic features of the bubo.

Case #1 was a 21 year old male participant, who was evaluated as being in a clinically severe state at the point of admission. On palpation, a left axillary bubo was detected which was hard with surrounding oedema and painful (pain score: 8/10).

On D1, the bubo demonstrated relatively normal structures with a distinct capsule and visible hilum, and measured 28.8mm (**Figure 2.12**). No necrosis or hypervascularity was detected.

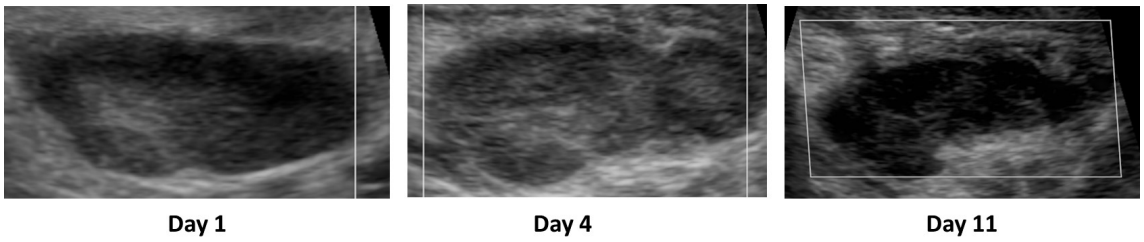


Figure 2.12: Ultrasound images taken on Day 1, Day 4 and Day 11 for case report #1

On D4, no changes to the clinical characteristics of the bubo were observed – although the patient reported a lower pain score of 6/10. While the overall sonographic characteristics remained similar to that observed on D1, the capsule became lobulated and some visibility of the hilum was lost. Interruption to the capsule was also observed indicating perinodal reaction or increasing oedema. The size of the bubo increased by 7.5mm to 35.3mm.

By D11, the patient was reported to have improved clinically and only mild plague symptoms remained. While the pain score further decreased to 3/10 and no changes to the clinical characteristics of the bubo were observed, further radiological progression was present. The capsule became more lobulated, the hilum was no longer visible and capsular interruption was still present. The size of the bubo remained stable from the previous measurement taken on D4 at 35.4mm.

Case #2 was a 38 year old male participant, who presented with mild symptoms of plague at admission and a right axillary bubo, which on palpation was hard without oedema and a pain score of 8/10.

The bubo was oval-shaped with a relatively distinct capsule – although some capsular interruption is seen in the blurred margins between the lymph node and perinodal tissue –

a moderately visible hilum and homogeneous cortex (**Figure 2.13**). The bubo measured 50.3mm.

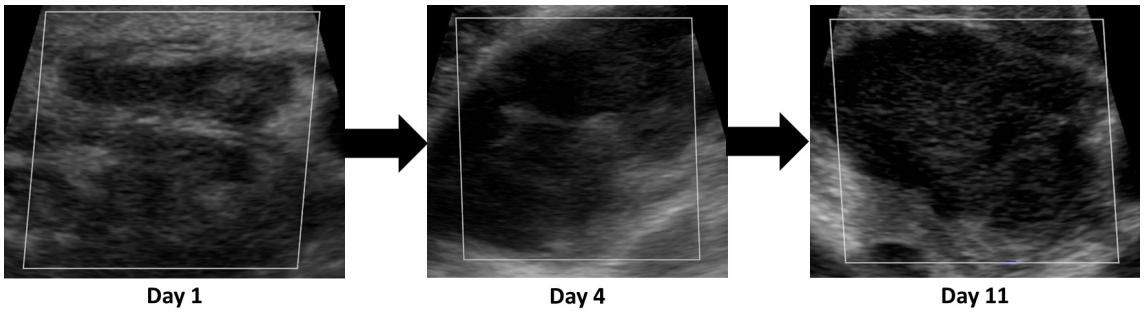


Figure 2.13: Ultrasound images taken on Day 1, Day 4 and Day 11 for case report #2

By D4, while the bubo remained hard and the pain score decreased to 6/10, oedema was noted on palpation. On ultrasound examination the bubo demonstrated a more rounded shape with a less visible hilum and blurred capsule. The bubo increased in size from D1 by 14.5mm to 64.8mm.

By D11, the patient's clinical status had improved overall but a few mild persistent symptoms remained. The size of the bubo decreased from D4 to 53.2mm, but remained larger than the measured size on D1.

Measurement validity

Sixteen digital caliper measurements were compared to corresponding ultrasound measurements of the same enlarged lymph nodes. Substantial discrepancies were observed across all anatomical zones when comparing the two methods.

Location	**N nodes**	**Absolute difference, Median (IQR), mm**	**% difference, Median (IQR)**
Cervical	1	2.67 (2.67 – 2.67)	157.06 (157.06-157.06)
Inguinal	9	3.15 (1.14-4.93)	112.76 (44.02-159.90)
Axillary	6	1.71 (0.67-2.78)	44.53 (15.72-55.66)

Axillary lymph nodes showed the smallest differences, with a mean absolute difference of 1.71mm (range: 0.67–2.78 mm), corresponding to a mean percentage difference of 44.53% (range: 15.72%–55.66%). Inguinal lymph nodes exhibited greater variability, with a mean absolute difference of 3.15mm (range: 1.14–4.93mm) and a mean percentage difference of 112.76% (range: 44.02%–159.9%). For cervical lymph nodes, only a single

bubo was assessed using both methods, resulting in a mean absolute difference of 2.67mm and a corresponding percentage difference of 157%.

Rater agreement

Minimal discrepancies were identified between ultrasound measurements of lymph nodes recorded by radiologists and those recorded by study technicians (**Table 2.5**). In the cervical zone, complete agreement was observed, with radiologists concurring with all three ultrasound measurements recorded by the study sonographers. In the inguinal zone, the mean absolute difference between measurements was 0.23mm (range: 0–2.58mm), corresponding to a mean percentage difference of 5% (range: 0–45%). In the axillary zone, the mean absolute difference was slightly higher at 0.375mm (range: 0–2.25mm), with a mean percentage difference of 7% (range: 0–45%).

Table 2.5: Median (range) absolute and percentage difference between expert and non-expert ultrasound measurements

Location	**N nodes**	**Expert v non-expert mean difference, (range), mm**	**Expert v non-expert mean % difference, (range)**
Cervical	3	0 (0-0)	0 (0-0)
Inguinal	13	0.23 (0-2.58)	4.58 (0-44.87)
Axillary	6	0.375 (0-2.25)	7.46 (0-44.73)

Identification of lymph node characteristics on ultrasound however was more challenging and there was less observed agreement between the experts and non-experts (**Figure 2.14** & **Figure 2.15**). There was overall good agreement on the distribution of the lymph nodes, the presence of necrosis, hypervascularity and coalescence. Least agreement was observed relating to the homogeneity of the cortex. Agreement however varied by zone. Higher levels of agreement were observed in the cervical and inguinal zone, and lower levels were observed in the axillary zone.

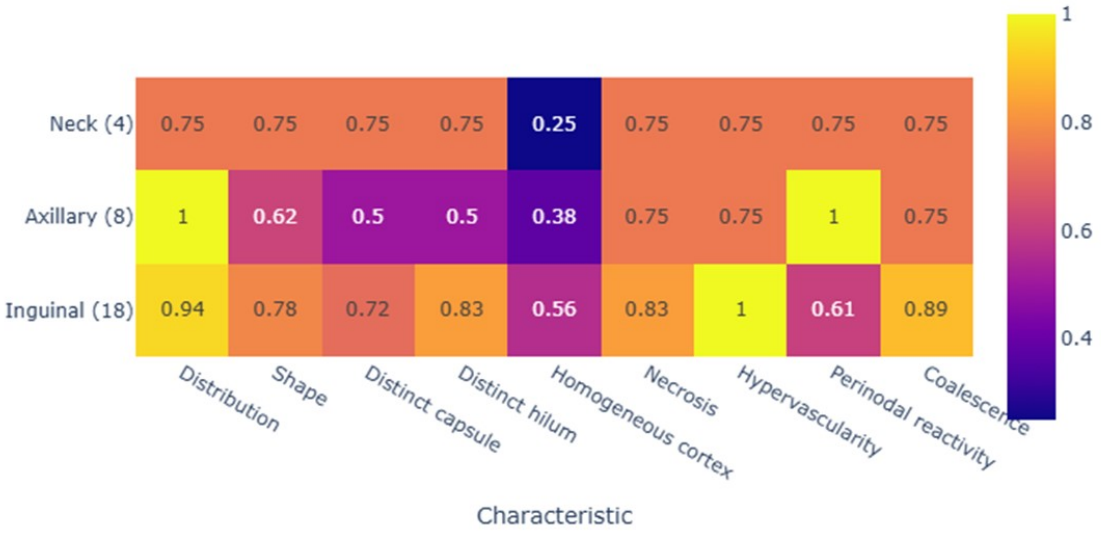


Figure 2.14: Agreement between experts and non-experts by characteristic and zone

Lower agreement was also observed on overall on the D11 scans (**Figure 2.15**) with substantially lower agreement occurring in the characterisation of the capsule and presence of necrosis.

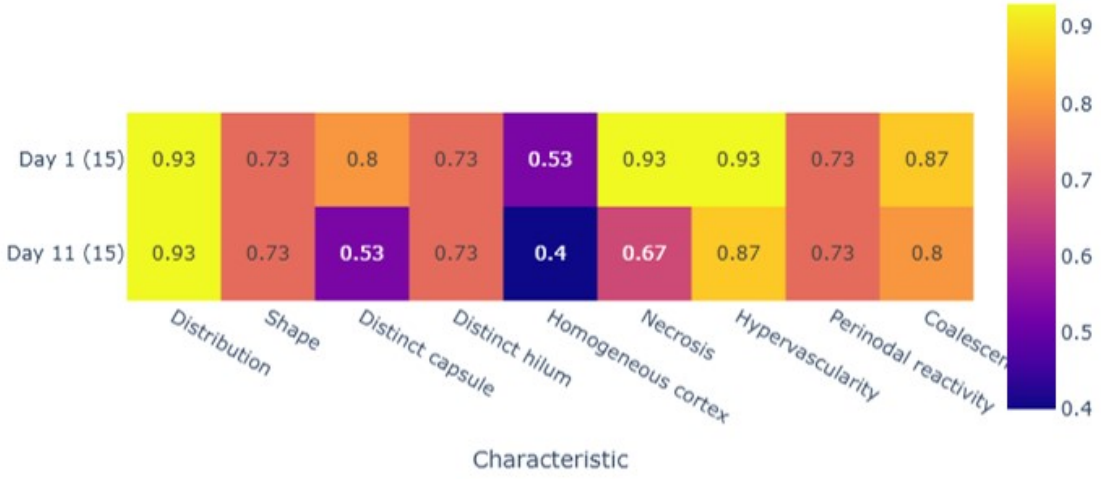


Figure 2.15: Agreement between experts and non-experts by characteristic and study day

2.3.3 Discussion

This study presents findings from a small cohort of patients with suspected bubonic plague, including 12 confirmed cases and four non-cases, and is the first study of plague to systematically characterise the progression of enlarged lymph nodes from admission to the end of treatment (D11).

All confirmed plague cases presented with a detectable bubo at admission, most of which exhibited typical clinical and sonographic features of enlarged lymph nodes. Only a few cases demonstrated severe irregularities, such as hypervascularity and necrosis.

Our findings highlight the absence of an observed association between bubo size, morphology, and clinical status, which is particularly evident in the case reports. While a larger dataset would be necessary to confirm this finding, our observations suggest that bubo size may not be a reliable indicator of treatment response. On D1, all (but one) patient presented with mild plague symptoms and all patients showed overall clinical improvement by the end of treatment on D11 with only mild symptoms remaining. However, bubo size remained relatively stable across the cohort and did not decrease with clinical improvement, contrary to the hypothesis of the IMASOY endpoint, which expected a reduction in size of at least 25%. The other clinical trial conducted for plague in Uganda assumed a total resolution of the bubo in its endpoint. [58] Furthermore, in the two case reports presented, ultrasound imaging revealed substantial progression, with worsening radiological findings over time despite clinical recovery. These findings suggest that bubo size may not accurately reflect treatment efficacy and could potentially confound the interpretation of trial outcomes. Given that patients must meet all IMASOY endpoint criteria to be classified as having responded to treatment, it is unlikely – based on these data, where the median percentage change in bubo size from D1 to D11 was approximately 10% – that any patient will achieve the required 25% reduction in bubo size, even if they are considered to be clinically well. A larger cohort is however needed to validate these conclusions.

Our study further highlights the limitations and risks associated with using a digital caliper to measure plague buboes. Consistent with the findings in section 2.2, digital caliper measurements differed substantially from those measured by ultrasound, which was

used as the reference standard. Measurement error was considerable across all anatomical regions assessed, with the greatest discrepancies observed in the inguinal region, where most buboes were located.

This study not only suggests a lack of association between bubo size and clinical status but also highlights the significant measurement error that may be introduced by using a digital caliper for bubo assessment. These findings indicate that incorporating bubo measurement as a component of the primary endpoint in the IMASOY trial may in turn lead to incorrect classifications of patients as either having or (more likely) not having had a treatment response as a result of mismeasurement of their bubo.

The study was conducted by clinicians and nurses with limited or no prior experience in ultrasound imaging. Training clinical staff with minimal prior sonographic experience to perform ultrasound scans for various diseases may offer several benefits for patient care and public health. When combined with the use of portable ultrasound devices, this approach could increase access to sonographic assessment for patients in underserved areas. Additionally, it may prove valuable in managing high case burdens by facilitating more effective patient triage. While a comparison of the lymph node measurements revealed good agreement between the radiologists and the study sonographers, there was limited agreement in the identification of lymph node characteristics. This may be due to the perceived subjectivity of some of the assessments, such as the distinctness of the capsule and homogeneity of the cortex. These results suggest that while objective measures can be performed well by newly trained clinical staff, characterisation of clinical features of lymph nodes may require a longer training and supervision period with regular expert reviews.

Author contribution to "A prospective cohort study to describe the morphology of buboes in patients with bubonic plague using ultrasound imaging"

Role description	Author contribution
Conceptualisation	X

Methodology	X
Software	-
Validation	X
Formal analysis	X
Investigation	-
Resources	-
Data curation	X
Writing (original draft)	X
Writing (review and editing)	X
Visualisation	X
Project administration	X
Funding acquisition	X

Publications

An abstract reporting the findings of this study has been accepted for presentation at the International Pandemic Sciences Conference 2025.

The article associated with this chapter is being drafted for submission for publication.

2.4 Contribution of the chapter

This chapter explores the impact of the lack of validated, accepted measurement tools on the assessment of clinical outcomes of patients with bubonic plague enrolled in clinical trials. Before initiating the projects in this chapter, several concerns were raised about incorporating "a 25% reduction in bubo size" as an indicator of treatment response as part of the composite primary endpoint in the IMASOY trial.

The findings presented in this chapter confirm the significant suspected limitations using changes in bubo size as a measure of treatment response. They also demonstrate

the risks involved in selecting endpoints that are based on anecdotal evidence of changes to physiological features that have not been systematically evaluated or documented.

First, no direct relationship could be observed in the imaging study between bubo size and the patients' overall clinical status – both of which remained relatively stable from admission to the end of treatment (D11). Second, the imaging study also demonstrated that the average reduction in bubo size between D1 and D11 was substantially lower than expected at approximately 10%. None of the buboes evaluated in the study met the 25% reduction criteria of the IMASOY trial. Finally, the findings of both the artificial bubo measurement study and imaging study demonstrate the likelihood of substantial measurement error when using a digital caliper to measure buboes. Challenges were also observed relating to inter- and intra-rater reliability and accuracy, which are complicated in a clinical setting by the inherent variability in bubo presentation between patients.

Combined, these issues confirm concerns about the robustness of using bubo size reduction as a component of the composite primary endpoint in the IMASOY trial. Following presentation of these results to the IMASOY Trial Steering Committee and its Data Safety Monitoring Board, the primary endpoint was amended to remove "a 25% reduction in bubo size" (**Table 2.7**). This amendment was made before enrolment to the trial stopped and any analysis took place.

Table 2.7: Original and revised primary endpoint in the IMASOY trial

Original primary endpoint	Amended primary endpoint
• Death • Persistent fever • No reduction or less than 25% reduction in bubo size • Development of secondary pneumonic plague • Administration of alternative or additional treatment for plague	• Death • Persistent fever • ~~No reduction or less than 25% reduction in bubo size~~ • Development of secondary pneumonic plague • Administration of alternative or additional treatment for plague

The original primary endpoint however was retained as a secondary endpoint to ensure transparency of reporting and demonstrate the effect of bubo size in a larger patient population as an indicator of therapeutic response.

Although a composite endpoint was also used for one of the other two plague trials [58], selecting an endpoint for a disease with relatively low mortality, heterogeneous clinical manifestations, and limited published descriptive data is not unique to plague; it is a common challenge in trials of rare diseases in general. In such cases, clinical trials often rely on surrogate or composite endpoints to enhance the sensitivity of the outcome measure by increasing the number of detectable events. [138] Composite primary endpoints can be particularly useful for diseases with heterogeneous manifestations, such as plague, as they allow for the evaluation of multiple important outcomes rather than focusing on a single outcome at the expense of others. [139] While it was clearly necessary to remove bubo size as a component of the primary endpoint due to its potential to introduce substantial measurement bias, a question could still be raised around the use and interpretation of the amended composite endpoint.

McCoy (2018) states that three questions should be asked before proceeding with a composite primary endponint [60]:

- Are all of the components of similar importance?
- Is each component expected to occur with a similar frequency?
- Does each component share similar relative risk reductions?

If the response to any of these questions is "No" then McCoy states that the composite endpoint should be dropped and a single endpoint should be selected instead.

If these questions are asked in relation to the IMASOY endpoint, the response to each question would be either "No" or "Unknown". To take the first question as an example, clearly death and persistence of fever are not of equal clinical importance. A particular challenge with including persistence of fever in the composite endpoint is that, while it is intended to reflect fever due to plague, the fever could be attributable to other causes, such as co-infection with diseases like malaria, which is also prevalent in plague-endemic regions of Madagascar. [140] The final component of the endpoint – decision

to continue treatment – is also highly subjective and the antibiotics used to treat plague (gentamicin and ciprofloxacin) are also commonly used to treat other infections. An individual clinician's decision to continue to treatment may be out of an abundance of caution if the patient is displaying any mild persistent symptoms, which may or may not be due to plague or any other suspected co-infection.

As patients must meet all of the events in the endpoint in order to be considered as having a response to treatment, a general problem with composite endpoints is that there is a risk that one component, over the others, drives the outcome and therefore it is difficult to determine between groups which component is having an effect on patient outcomes. [138] A higher rate of treatment failures in the IMASOY investigational arm may lead to the interpretation that there are more deaths in this group, when in fact there may be more patients with persistent fever or clinicians deciding to continue treatment. Understanding the treatment effect therefore becomes unclear, which is a considerable concern as the results of the trial are intended to support medical decision-making.

While this chapter intended to focus just on one component of the IMASOY primary endpoint, there are clear challenges relating other aspects that may introduce bias and uncertainty in to the trial results. Using a composite endpoint approach to manage low and uncertain event frequencies clearly has some considerable limitations that may be just as problematic as selecting a single endpoint. As biomarkers are often used as an alternative strategy in the definition of endpoints to manage this challenge, chapter 3 will explore the assessment of biomarkers and surrogate endpoints for another outbreak-prone infection, mpox. (Note: A different disease has been selected for study in chapter 3 due to the availability of data). The results of the IMASOY trial will then be presented in chapter 4 to review the impact of the change to the endpoint on the interpretation of the trial results.

Chapter 3

Challenges in Implementing Endpoints in Clinical Trials: Assessing Measurement Instrument Selection for Clade 1 Mpox

3.1 Introduction

The challenges associated with selecting appropriate outcome measures and measurement instruments for plague share similarities with those encountered in clinical research on mpox. These challenges derive from a limited understanding of the frequency, progression, and clinical meaningfulness of signs and symptoms, as well as important patient outcomes. Before the multi-country outbreak of Clade IIb mpox that began in 2022, few clinical studies had been published, most of which described individual cases or small cohorts of patients with Clade I or Clade IIa mpox. [141] [142] [143] [144] [145] These studies not only involved small sample sizes but also exhibited substantial variability in the reporting of clinical features, sample collection methods, and patient outcomes. Re-

porting inconsistencies were evident in both the types of data collected and the time points at which observations were recorded. Consequently, when trials were urgently needed in 2022 to evaluate therapeutic options for a new variation of mpox, there was a significant lack of clinical data to inform the selection and measurement of appropriate endpoints.

Despite these uncertainties, clinical studies for both Clade I and Clade IIb mpox have used "time to lesion resolution" as a primary endpoint. [133] [134] [130] [146] Unlike the trials for plague that used composite endpoints, lesion resolution represents a single, observable outcome. Given that a high lesion burden is a defining characteristic of Clade I mpox and is a highly observable outcome, it was likely considered a practical and viable endpoint. Moreover, lesion assessment has been used as an outcome measure in clinical trials for other infectious diseases, such as herpes simplex virus [147] [148] and cutaneous leishmaniasis, [149] establishing a precedent with regulatory bodies and ethics committees overseeing the scientific integrity of these studies.

However, at the time these studies were initiated, no evidence supported a clear association between lesion resolution and clinical improvement. Furthermore, there was no standardised approach for distinguishing between "resolved" and "active" lesions in mpox, raising concerns about the validity and reliability of lesion resolution as an endpoint.

There had also been limited exploration of alternative endpoints beyond lesion resolution, such as severe clinical manifestations that may develop over time or virological outcomes. The lack of detailed clinical data on disease progression and patient outcomes in Clade I disease has been compounded by the absence of longitudinal sampling data, which has limited the selection of virological endpoints.

To address these issues, the section 3.2 of this chapter examines the "time to lesion resolution" endpoint widely used in mpox trials. It specifically investigates the extent of consensus among clinicians when assessing lesions at different stages of evolution and evaluates whether "time to lesion resolution" is a measurable and reliable outcome. Although most mpox trials have focused on Clade IIb, this project – and the investigations described in subsequent sections – focuses on Clade I mpox, for which only one trial was initiated during the course of these studies [146], but others are needed to evaluate therapeutics in this clade.

Building on this analysis, section 3.3 and section 3.4 explore alternative endpoints, focusing on whether virological outcomes – such as changes in viral load and time to PCR negativity (which have been used in other clinical trials for viral diseases like HIV and COVID-19) [150] [151] [152] – could serve as viable outcome measures, and identifying other clinically meaningful signs and symptoms that may warrant further investigation.

A key challenge with virological endpoints in mpox however is the variability in MPXV DNA detection across sample types. [153] For Clade IIb, viral DNA has been most reliably detected in genital and peri-anal swabs, but this is unlikely to apply to Clade I due to differences in transmission routes and clinical presentation. Lesion samples have shown wide variability in MPXV detection, with positivity rates ranging from 56% in some studies [153] to 94% in others. [154] Moreover, MPXV DNA has been detected for up to four weeks in some cases, [153] [154] limiting the correlation between viral load and clinical outcomes. Most of this research, however, has focused on Clade IIb, and longitudinal sampling data for Clade I remains scarce. Therefore section 3.3 explores the feasibility of using different sample types and PCR primers to detect MPXV DNA in Clade I.

As section 3.2 and section 3.3 rely primarily on non-clinical or retrospective data to identify potential methodological challenges with lesion and virological evaluation, section 3.4 of this chapter draws on data from a prospective cohort study of Clade I mpox patients in the Central African Republic. This analysis aims to validate the earlier findings and determine whether additional clinical outcome measures could be considered for future clinical studies of Clade I mpox.

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

As stated in the introduction, there are inherent difficulties associated with visual lesion assessment, particularly those conducted by multiple observers over time. "Time to lesion resolution" has however been recently used as the primary endpoint for both Clade I and Clade IIb studies. [133] [146] [130] [134]

Many areas of clinical practice and trials currently use imaging with or without Artificial Intelligence (AI) enhancements to support the assessment of dermal lesions, such as for acne [155] and skin cancer [156]. This however is not the case for mpox, which in all clinical trials to date has relied on visual inspection of lesions to classify disease stage. Issues with the subjective clinical assessment of dermal lesions for other diseases have however been noted, such as psoriasis where the detection of changes in plaques over time is challenging and there is a lack of validation of existing scoring systems. [157] The lack of exploration of methods to classify dermal lesions appears to be a common problem across many disease areas, including mpox for which limitations in the clinical identification of lesion stages have been previously noted.[131] Before the launch of several mpox trials in 2022, no studies had been conducted to investigate whether clinicians can (i) reliably distinguish mpox lesions from other viral lesions, which is particularly important in settings where laboratory testing is limited or (ii) accurately differentiate between stages of Clade I and Clade IIb mpox lesions. The ability to identify and reliably classify Clade I mpox lesions based on clinical assessment has significant implications for low-resource settings in endemic regions, particularly where laboratory testing for mpox is limited, and for preparedness in historically non-endemic regions, where there may be low clinical suspicion of mpox. Clinical suspicion of mpox likely relies on the identification of characteristic lesions and accurate lesion classification is critical for informing treatment decisions, guiding public health interventions, and ensuring the reliability of clinical research outputs.

This study aims to evaluate inter-clinician agreement in determining a differential diagnosis between Clade I mpox and varicella, and reliability classifying Clade I lesion stages. We chose to evaluate whether clinicians can accurately distinguish between mpox and varicella as both diseases exhibit similar lesion presentation and co-circulate in the same geographic region. Given that diagnosis in such settings is frequently based on clinical assessment alone, it is critical to determine the accuracy of clinicians' visual and clinical differentiation, particularly when treatment decisions depend on this initial evaluation.

A parallel study by these authors on the reliability of Clade IIb lesion assessment also revealed only moderate agreement among assessors and highlighted potential challenges in consistent lesion classification.[158] Repeating this study for Clade I mpox is important for

several reasons, including differences in lesion presentation between clades – with Clade I lesions being more extensive and patients having a higher burden – and as a number of other diseases with a similar clinical presentation commonly circulate in the same geographical areas as Clade I mpox, such as varicella zoster virus (VZV), making difficult to clinical distinguish between diseases.

3.2.1 Methods

The objective of this study was to assess inter-rater reliability and agreement among clinicians with experience evaluating lesions in patients with Clade I mpox. Two key areas of interest were examined:

- differential diagnosis between Clade I mpox and varicella based on lesion presentation,
- classification of lesion stages.

The study's design, methodology, and results are reported according to the Guidelines for Reporting Reliability and Agreement Studies (GRAAS).[136]

Ethical approval was obtained from the University of Oxford Medical Science Interdivisional Research Ethics Committee (R84355/RE001).

Participants

Participants included clinicians with experience diagnosing and managing patients with Clade I mpox. Given the potentially limited pool of suitably experienced clinicians, no pre-specified target sample size was established. As a result, the study was designed to be descriptive.

Data Collection

Using the RedCap survey tool [110, 137], participants were presented with 17 sets of anonymised images of patients with lesions confirmed by PCR to be either Clade I mpox or varicella. The images, sourced from the Institut Pasteur de Bangui (IPB) clinical image library, depicted lesions at various stages and on different areas of the body.

Participants independently reviewed each image and indicated their most likely diagnosis – mpox or varicella – and classified the lesions as either active, scabbed, or resolved.

An option to indicate an inability to classify the lesion was also provided. Before beginning the survey, participants were shown the World Health Organization's working definitions for each lesion stage.[159]

The survey collected limited demographic information about the clinicians and their experience, including each clinician's country of practice, the average number of Clade I mpox patients they managed annually, and their self-assessed confidence in evaluating mpox lesions. Confidence was rated on a 10-point ordinal scale, with 1 indicating no confidence and 10 indicating complete confidence. The questionnaire was provided in English and French.

Data Analysis

All analyses were performed using the *irr* package in R Statistical Software (v4.3.2)[34] and validated with the *statsmodels* package in Python (v3.11.5) by two independent analysts.

Accuracy in identifying mpox or varicella was measured as the percentage of raters selecting the correct diagnosis for each image. Spearman's rank correlation coefficient (ρ) was used to assess the relationship between experience – measured by self-reported expertise and the number of patients managed annually – and their diagnostic accuracy. Spearman's method was selected for its appropriateness with small sample sizes and ordinal data, though the results should be interpreted cautiously given the limited statistical power and susceptibility to noise and outliers in small datasets.

Inter-rater reliability was assessed using Fleiss' kappa coefficient (κ), while inter-rater agreement was evaluated through the proportion of partial agreement (ρ_0) and proportion of exact agreement (ρ_{0e}). Partial agreement was defined as alignment on any single selection in a multiple-choice response, whereas exact agreement referred to complete alignment across all selections in a multiple-choice response.

Lesion classification analyses focused exclusively on images from patients with confirmed mpox diagnoses, on which two separate analyses were performed:

- Responses to images showing only a single or homogeneous lesion type, for which κ and ρ_{0e} were calculated,

- Responses to images with multiple lesion types (combinations of active, scabbed, and resolved lesions), for which κ, ρ_0, and ρ_{0e} were reported.

3.2.2 Results

The questionnaire was circulated to 38 clinicians, of whom 17 accessed the survey and completed the eligibility screening. One respondent was deemed ineligible as they were not directly involved in the clinical management of patients with Clade I mpox.

All 16 eligible participants who began the questionnaire completed it in its entirety. Half of the participants were from the Democratic Republic of Congo (n=8, 50%) (**Table 3.1**). There were two (13%) respondents each from the Central African Republic, France and the UK, and there was one (6%) respondent each from Nigeria and Belgium. Half of all respondents had managed more than 50 cases of mpox. The respondents had a median confidence score of eight (range: 7.5-8).

Table 3.1: Summary of study participants

Country of work, n (%):	
Democratic Republic of Congo	8 (50%)
Central African Republic	2 (13%)
France	2 (13%)
United Kingdom	2 (13%)
Nigeria	1 (6%)
Belgium	1(6%)
Number of mpox patients the respondents has personally managed, n (%):	
<5	0
5-10	2 (13%)
10-20	1 (6%)
20-50	5 (31%)
>50	8 (50%)
Confidence assessing mpox lesions, median (IQR):	
Confidence score	8 (7.5 - 8)

Diagnosis

The accuracy of selecting the most likely diagnosis between mpox and varicella based on lesion presentation varied across the images, with correct identification rates ranging from 25% to 100% and a median accuracy of 75% (Q1: 62.5%, Q3: 93.75%). No significant correlation was observed between diagnostic accuracy and either self-rated con-

fidence (Spearman's $\rho = 0.3$) or clinical experience (Spearman's $\rho = 0.17$) in assessing lesions (**Table 3.2**).

Table 3.2: Correlation between raters' diagnostic accuracy, self-rated confidence and experience using Spearman's correlation (ρ)

	Image	**Self-rated confidence**	**Experience**
Image	-	0.3	0.168
Self-rated confidence	0.3	-	-0.042
Experience	0.168	-0.042	-

When asked to select the most likely diagnosis between varicella and mpox based on lesion presentation, inter-rater reliability was found to be poor ($\kappa = 0.223$; z = 10.1), while agreement was moderate ($\rho_{0e} = 68\%$).

Additionally, there appeared to be improved diagnostic accuracy when a single lesion type was present in the image, as opposed to images displaying lesions in multiple concurrent stages of evolution (**Table 3.3**). No relationship could however be observed between diagnosis and accuracy, as the small sample size limited the ability to draw any generalisable conclusions.

Table 3.3: Summary of question characteristics and accuracy score

Question	**Accuracy (%)**	**Diagnosis**	**Lesion type(s)**
Question 4	25	Mpox	Multiple
Question 19	43.75	Mpox	Multiple
Question 5	50	Varicella	Single
Question 16	56.25	Varicella	Multiple
Question 11	62.5	Mpox	Multiple
Question 9	68.75	Varicella	Multiple
Question 13	68.75	Mpox	Single
Question 18	68.75	Mpox	Single
Question 20	75	Mpox	Single
Question 6	81.25	Mpox	Single
Question 8	81.25	Mpox	Multiple
Question 10	81.25	Mpox	Multiple
Question 14	93.75	Mpox	Multiple
Question 17	93.75	Mpox	Single
Question 7	100	Mpox	Single
Question 12	100	Mpox	Single
Question 15	100	Mpox	Single

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 ($\rho_0 = 78\%$) (**Table 3.4** & **Figure 3.1**).

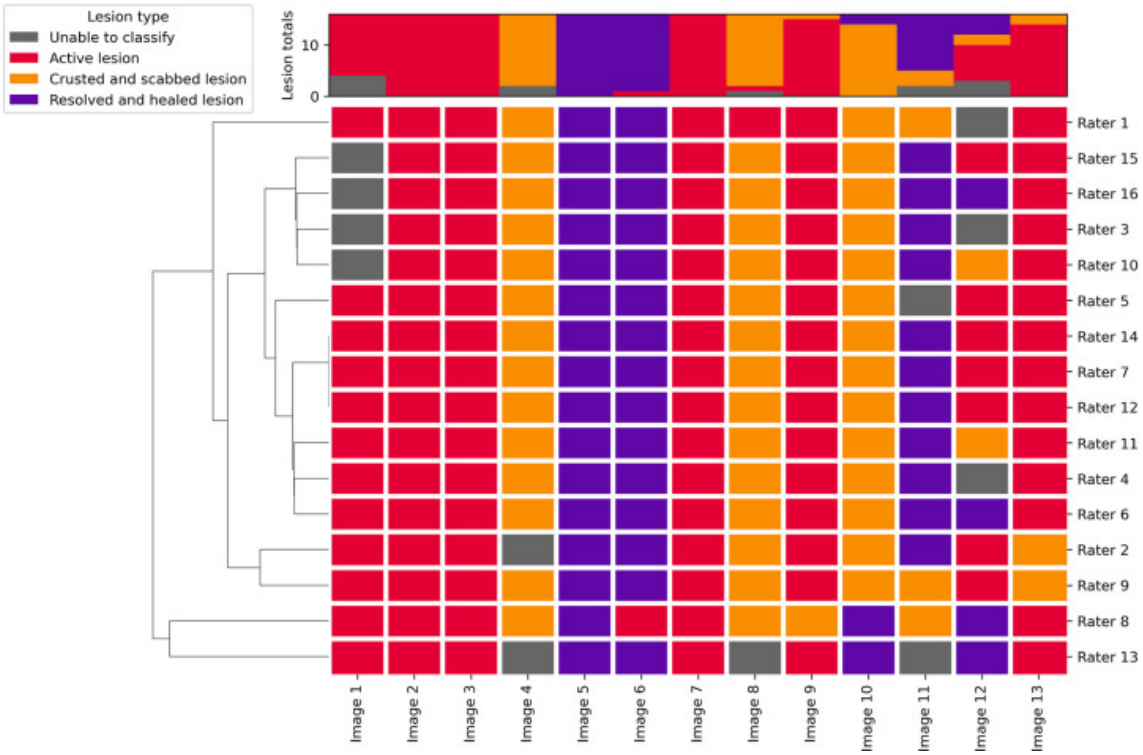


Figure 3.1: Dendrogram of the respondents' lesion assessments for which a single lesion type was present in the image

However, when multiple lesion types were shown in an image, both inter-rater reliability ($\kappa = 0.153$, z = 10.5) and agreement ($\rho_0 = 29\%$) substantially decreased (**Table 3.4**).

For partial agreement, inter-rater reliability was moderate ($\kappa = 0.447$, z = 25.9), and agreement was good ($\rho_0 = 73\%$) (**Table 3.4** & **Figure 3.2**).

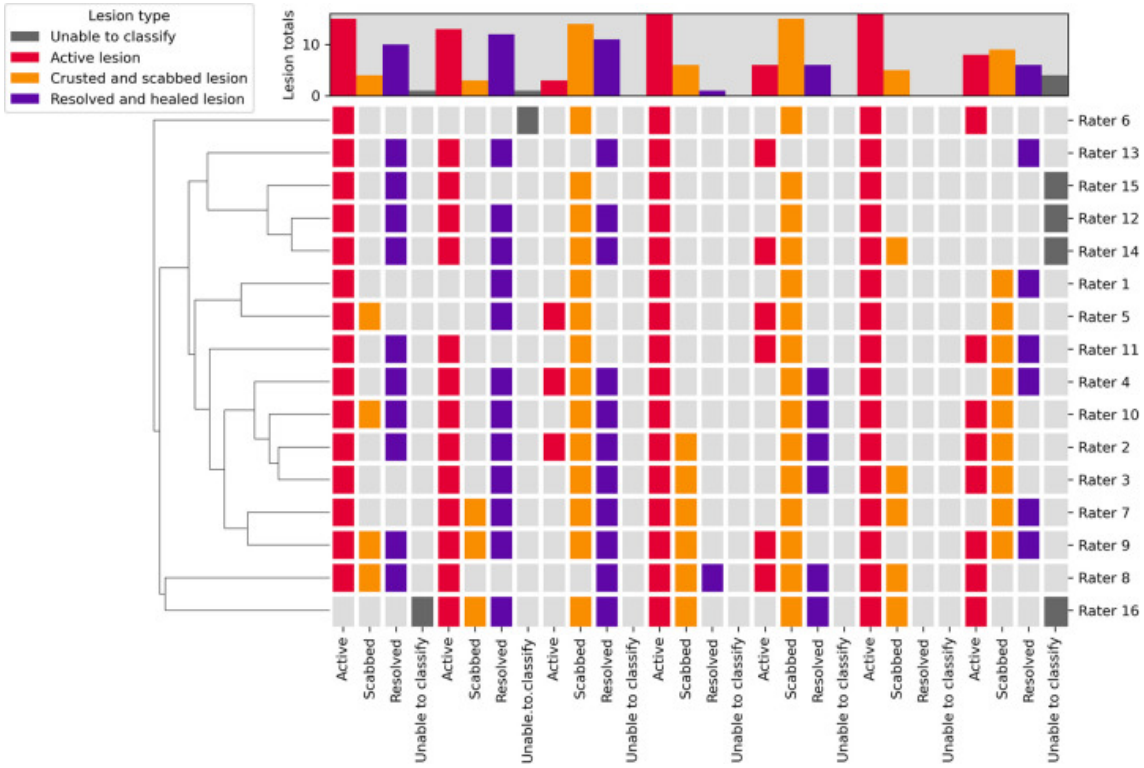


Figure 3.2: Dendrogram of the respondents' lesion assessments for which multiple lesion types were present in the image

3.2.3 Discussion

This study highlights the current limitations in using clinical assessment for diagnosing Clade I mpox and evaluating lesion stages and treatment outcomes.

Our findings reveal moderate diagnostic accuracy, poor reliability, and moderate inter-rater agreement among clinicians distinguishing between mpox and varicella based on lesion presentation. In settings where laboratory testing for mpox is limited or unavailable, there is a risk of misdiagnosis, which may result in mismanagement, such as prescribing incorrect or unnecessary antiviral treatments. Misclassification of cases can have significant public health implications, including inaccurate disease surveillance, inappropriate vaccine distribution, erroneous drug procurement, and failure to meet obligations for reporting notifiable diseases. These misclassifications also impact the evaluation of treatment efficacy, both in clinical practice and research, by introducing inconsistencies in outcome classification.

Table 3.4: Summary of inter-rater reliability and agreement results

Diagnosis	**Response selection formats**	**Lesion types present in image**	**Fleiss' kappa**	**Z score**	**% agreement type**	**% agreement result**
Mpox & VZV	Single choice	Single and multiple lesion types	0.223	10.1	Percentage of exact agreement	67.745
Mpox	Single choice	Single lesion types	0.671	40.6	Percentage of exact agreement	78.14
Mpox	Multiple choice	Multiple lesion types	0.447	25.9	Percentage of partial agreement	73.18
Mpox	Multiple choice	Multiple types	0.153	10.5	Percentage of exact agreement	29.04

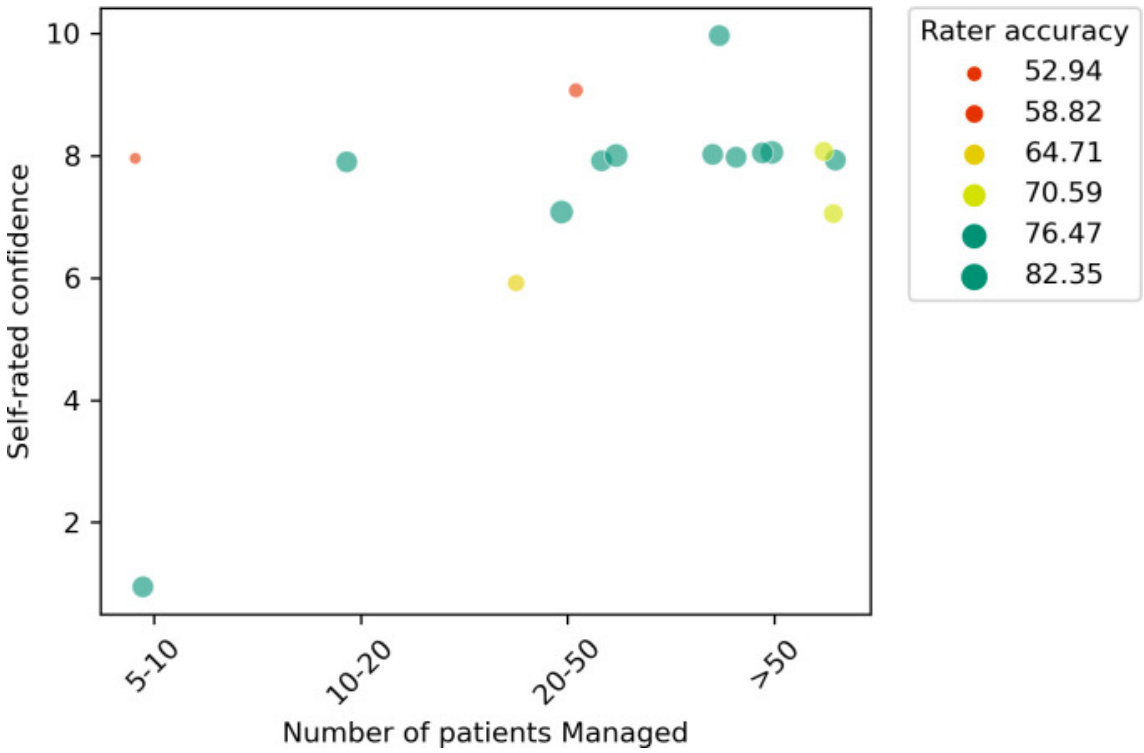


Figure 3.3: Scatterplot of respondents' self-rated confidence plotted against experience (measured according to the number of patients managed)

For mpox lesion classification, we observed moderate reliability and good agreement when a single lesion type was present, although these results should be interpreted with caution, as they do not necessarily indicate strong agreement beyond chance [160] [161]. However, achieving consistent assessments became notably more challenging when multiple lesion types were present. Since lesions likely evolve through different stages over time, the poor reliability and agreement observed when multiple lesion types coexisted poses a substantial challenge to accurate clinical assessments. This issue is especially critical in clinical trials, where a single assessment of a lesion is often used as a surrogate for overall disease progression and response to treatment.

These findings emerged despite the participants' high self-reported confidence (IQR: 7.5-8) in diagnosing mpox lesions. No significant correlation was found between confidence, clinical experience and accuracy (**Figure 3.3**). Given the sporadic and widespread nature of Clade I mpox reports across large geographic areas, it is possible that confidence and agreement among real-world clinicians – who may only occasionally treat mpox or varicella – could be lower than what was observed in this relatively expert group.

A parallel study examining Clade IIb mpox lesions found higher agreement and reliability among clinicians. [158] This may be attributed to differences in lesion presentation, as Clade IIb lesions tend to be more localised and fewer in number than those of Clade I. The surveys were not combined due to the distinct clinical presentations between the clades, and the clinical experience of clinicians spanning both clades remains limited to a very small group of highly specialised experts.

There is an ongoing need for resources, such as the WHO lesion assessment atlas, and exploration of the use of artificial intelligence (AI) and machine learning (ML) to assist clinicians and clinical trialists to improve lesion assessments, particularly in low-resource settings where laboratory facilities are limited. These technologies could prove crucial before rapid diagnostic tests become widely available. While AI and ML have shown efficacy in diagnosing Clade IIb mpox, evidence on their effectiveness with Clade I remains limited. [162] [163] By leveraging these technologies, we can bridge diagnostic gaps, ensuring timely and accurate detection of mpox, even in challenging environments, ultimately improving diagnostic accuracy, patient outcome assessments in trials and disease control.

This study has several limitations. The sample size was relatively small due to the limited number of clinicians worldwide with experience managing Clade I mpox patients. The response rate was further impacted by challenges in identifying and reaching clinicians working in remote areas with limited internet connectivity. Additionally, visual inspections of images may not fully capture all the information typically considered in clinical assessments, such as epidemiological context, other clinical signs (e.g., fever, adenopathy), patients' descriptions of lesion evolution, and associated symptoms like pain. We chose to assess reliability and agreement rather than accuracy because there is no established "gold standard" for classifying mpox lesions. Furthermore, images featuring mpox and varicella co-infection were not included in the study.

Overall, clinicians face significant challenges in distinguishing Clade I mpox from varicella and in reliably assessing disease stage in Clade I mpox patients. More robust and accessible diagnostic tools and indicators are needed to inform clinical decisions, public health strategies, and research priorities – especially in resource-limited settings where mpox is prevalent.

Author contribution to "Challenges in Clinical Diagnosis of Clade I Mpox: Highlighting the Need for Enhanced Diagnostic Approaches"

Role description	Author contribution
Conceptualisation	X
Methodology	X
Software	-
Validation	X
Formal analysis	X
Investigation	X
Resources	-
Data curation	X
Writing (original draft)	X
Writing (review and editing)	X
Visualisation	X
Project administration	X
Funding acquisition	-

Publications

This chapter is based on the following publications:

- **Bourner J**, Garcia-Gallo E, Mbrenga F, Boum Y 2nd, Nakouné E, Paterson A, Jones B, Olliaro P, Rojek A. Challenges in clinical diagnosis of Clade I Mpox: Highlighting the need for enhanced diagnostic approaches. PLoS Negl Trop Dis. 2024 Jun 24;18(6):e0012087. doi: 10.1371/journal.pntd.0012087. PMID: 38913721; PMCID: PMC11226010.

- Jones B, Paterson A, AlKhoury N, **Bourner J**, Dunning J, Olliaro P, Rojek A. Variability in clinical assessment of clade IIb mpox lesions. Int J Infect Dis. 2023

Dec;137:60-62. doi: 10.1016/j.ijid.2023.10.004. Epub 2023 Oct 15. PMID: 37848125; PMCID: PMC10914632.

3.3 Laboratory Diagnosis of Mpox, Central African Republic, 2016–2022

While lesion resolution presents challenges for evaluation in clinical trials, alternative strategies – such as viral clearance or PCR negativity – have been proposed and utilised as secondary endpoints in some studies. [134] [164] [133] However, at the time of initiating these studies for Clade IIb, limited data were available on the utility of this measure and its correlation with clinical response to treatment. Early studies conducted during the Clade IIb outbreak also revealed variations in viral load and the time to viral clearance across different sample types.[165] Notably, these findings were specific to Clade IIb, and no studies had characterised the viral load kinetics over time in Clade I mpox cases. Extrapolating findings between clades is challenging, as differences in clinical presentation and disease severity may influence viral kinetics and response to treatment. For example, Clade IIb mpox cases typically present with a lower and more localised lesion burden compared to Clade I, which may result in different viral load dynamics. In addition, case fatality ratios vary between clades, with Clade I historically associated with higher CFRs (6–25%) [166] [167], suggesting a more aggressive disease course and potentially longer or higher-level viraemia. In contrast, Clade IIb has shown a substantially lower CFR (estimated at 0.19% [168]), which may reflect more rapid viral clearance.

Different PCR testing methods can be employed primarily for the diagnosis of mpox, utilising either generic pan-mpox primers (G2RG) or clade-specific primers (C3L for Clade I, for example). However, these methods have not been systematically compared for their effectiveness in detecting mpox or assessing changes in viral load over time.

If viral clearance is to be used as an endpoint in clinical trials, it is essential to identify the optimal specimen types for sampling and select the most appropriate primers. However, low numbers of geographically dispersed cases detected in Clade I-endemic regions has historically made conducting a prospective cohort study infeasible. To address this limitation, we conducted a retrospective descriptive study using biobanked samples from IPB to evaluate viral load across different specimen types and primer sets. To inform the design of future Clade I clinical trials, this study also aimed to characterise the geographic

distribution of cases across the Central African Republic (CAR) and assess the proportion of suspected mpox cases diagnosed with VZV or identified as co-infected.

3.3.1 Methods

The Ministry of Health and Population in CAR set up a passive surveillance programme for mpox in 2010. Under this programme, specimens are collected from all suspected cases of mpox meeting the standardised case definition, which is disseminated to all health professionals in CAR through regular training sessions and posters displayed in health facilities. Specimens are sent for biologic confirmation by PCR to the national reference laboratory at IPB. Whenever possible, contact tracing is conducted after identification of confirmed cases.

Since 2016, each specimen received at IPB is tested for MPXV by real-time PCR. After specimen processing, 200 μL of each sample are extracted. The reactions are performed in 25 μL volume containing 12.5 μL of TaqMan Universal PCR Master Mix, 4.5 μL of nuclease-free water, 1 μL of each 10 μmol/L primer, using the generic primer (G2RG) and clade I–specific (C3L) primers and 5 μL of extracted DNA. [169] On the basis of these same concentrations, VZV primers are also used. [170]

By using results from all specimens collected from patients with suspected mpox under the national mpox surveillance programme during 2016–2022, this study aims to describe the mpox landscape in CAR and evaluate the agreement of mpox test results (including cycle threshold (Ct) values) generated using the G2RG and C3L primers and different specimen types (blood, active lesion, and scab).

3.3.2 Results

During 2016–2022, a total of 494 specimens (278 blood, 99 active lesion, 95 scab, and 22 oropharyngeal) from 302 patients were tested for suspected mpox at IPB. Of the total 302 suspected cases, 105 (35%) were positive for MPXV on $\geq$1 specimen (varying 19%–64% annually) (**Figure 3.4**). Of the 105 MPXV-positive patients, three (3%) were also positive for VZV. Of the 197 MPVX-negative patients, 82 (42%) were positive for VZV and 108

(55%) were negative for both MPXV and VZV. The remaining seven patients were not tested for VZV.

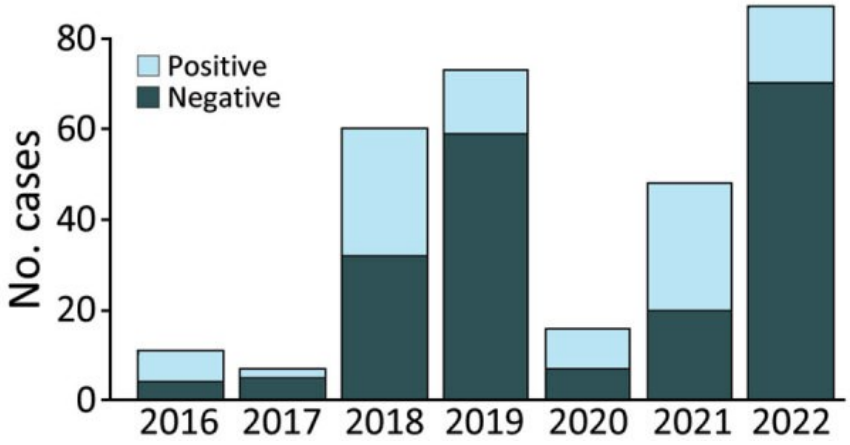


Figure 3.4: Laboratory test results for suspected mpox cases by year in CAR, 2016–2022. Of 302 suspected cases during the study period, 105 (35%) had positive results for mpox virus on ≥1 specimen.

The highest percentage of MPXV-positive specimens derived from the Lobaye and Mbomou prefectures, which together contributed 58% of mpox cases overall. MPXV detection rates varied by prefecture: Sangha Mbaere, 24/40 specimens (60%); Lobaye, 35/106 specimens (33%); Mbomou, 25/74 specimens (34%); and Bangui 2/41 specimens (5%).

Significantly more female patients were among MPXV-positive than VZV-positive patients ($p = 0.03$) but not among patients who were negative for both viruses. The median age across all suspected patients was 14 years; we observed no statistically significant difference between the median ages of confirmed patients with mpox (17 years) and VZV (20 years) infections. The median age of patients who tested negative on both tests was significantly lower (9 years).

Blood specimens were positive for MPXV on G2RG in 77/278 (28%) cases, active lesions in 45/102 (44%), scabs in 36/98 (37%), and oropharyngeal specimens in 3/22 (14%) (**Table 3.6**).

Of specimens returning a positive result on G2RG, the median Ct was 32.11 (IQR 29.12–35.45) for blood specimens, 18.92 (IQR 17.42–23.43) for active lesions, 18.07 (16.19–19.82) for scabs, and 30.15 (28.04–32.56) for oropharyngeal specimens (**Table 3.7**). Similar values

Table 3.6: Test results by specimen type and test type for MPXV and VZV, n/N (%)

Specimen type	MPXV (G2RG)		MPXV (C3L)		VZV	
	Positive	Negative	Positive	Negative	Positive	Negative
Blood	77/278 (28)	201/278 (72)	73/278 (26)	205/278 (74)	62/260 (24)	198/260 (76)
Active lesion	45/102 (44)	57/102 (56)	45/102 (44)	57/102 (56)	42/108 (39)	66/108 (61)
Scab	36/98 (37)	62/98 (63)	37/98 (38)	61/98 (62)	38/100 (38)	62/100 (62)
Oropharyngeal	3/22 (14)	19/22 (86)	2/22 (9)	20/22 (91)	6/22 (27)	16/22 (73)

were returned by C3L. For paired specimens, we observed either substantial (κ 0.61–0.80) or almost perfect (κ 0.81–1.00) agreement of a positive or negative result on pairwise comparisons of tests conducted on different specimens types on either G2RG or C3L.

Table 3.7: Cycle threshold values obtained using G2RG and C3L PCR primers on different specimen types. Data are median cycle threshold value (interquartile range). C3L, clade I–specific primer; G2RG, generic primer; MPXV, monkeypox virus; VZV, varicella–zoster virus.

Specimen type	MPXV (G2RG)	MPXV (C3L)	VZV
Blood	32.11 (29.12–35.45)	32.93 (30.25–35.94)	34.41 (31.38–36.01)
Active lesion	18.92 (17.42–23.43)	19.61 (18.05–23.57)	19.23 (17.69–20.82)
Scab	18.07 (16.19–19.82)	18.13 (16.46–21.46)	15.78 (13.63–18.42)
Oropharyngeal	30.15 (28.04–32.56)	28.19 (26.79–29.59)	34.31 (32.95–35.67)

The Ct values of G2RG and C3L on blood were significantly higher than in active lesion and scabs, whereas we observed no difference between active lesion and scab specimens. We observed no statistically significant difference between the Ct values generated on G2RG and C3L on the same specimens (**Figure 3.5**).

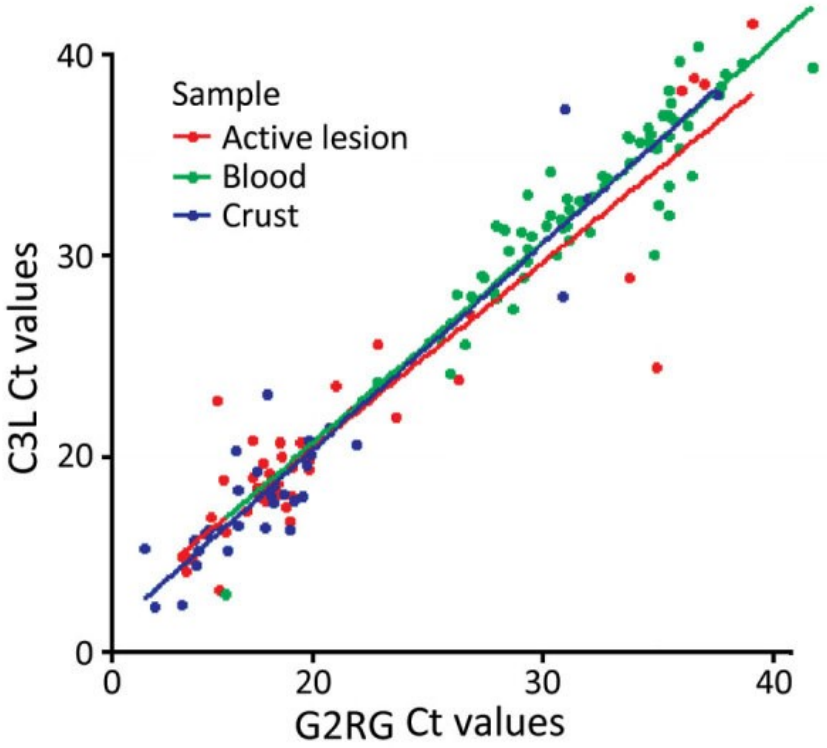


Figure 3.5: Distribution of Ct values obtained using G2RG and C3L primers of monkeypox virus–positive active lesion, blood, and scab specimens in study assessing laboratory diagnosis of mpox, Central African Republic, 2016–2022. C3L, clade I–specific primer; Ct, cycle threshold; G2RG, generic primer

3.3.3 Discussion

Approximately one third of suspected mpox cases in CAR are confirmed MPXV infections; an additional 2/5 are VZV infections, leaving $\approx$ 3/5 cases of papulovesicular cutaneous eruptions undiagnosed. Most mpox and VZV infections were diagnosed in teenagers and young adults, with an even younger population remaining undiagnosed.

Although cases of mpox are generally detected across the heavily forested, southern parts of CAR, mpox detection rates vary across prefectures. Some prefectures, such as Sangha Mbaere, have a high detection rate of MPXV (60%) over VZV (5%), whereas in others, such as Bangui, detection is much lower (MPXV 5%, VZV 46%). The varying detection rates between prefectures could be linked to local lifestyles and practices, as well as social instability. In the southwest region, local communities primarily subsist through hunting and gathering, spending long periods in mpox-endemic forest, which may increase

the risk for exposure to the virus; however, in the southeast, mpox-endemic bushlands are used for farming and as a place of passage or temporary habitation for communities that have been displaced by social instability. This data is useful to inform the location of future trial sites – which would be most efficient if conducted in the southeastern region – taking in to account potential logistical and security challenges.

Our results demonstrate very high agreement in PCR results between primers. The results also highlight the need to prioritise active lesion and scab specimens over blood and oropharyngeal specimens, given that their relatively higher viral loads for MPXV and VZV enable better detection. This correlates with findings in other studies highlighting higher viral load and persistent viral DNA detection in skin lesions over other sample types. [165]

However, these data are not longitudinal and are limited to diagnostic samples only. The duration of symptoms at the time of sampling is unknown, as is the extent to which Ct values may change throughout treatment and follow-up. Consequently, determining the optimal timing for evaluating endpoints related to viral clearance has not been resolved in this study. Therefore, before virological endpoints are used in trials more research is needed to understand changes in viral load between initial presentation and end of treatment, and whether viral load has any relationship with the overall clinical status of a patient.

Author contribution to "Laboratory Diagnosis of Mpox, Central African Republic, 2016–2022"

Role description	Author contribution
Conceptualisation	X
Methodology	X
Software	-
Validation	X
Formal analysis	X
Investigation	-

Resources	-
Data curation	X
Writing (original draft)	X
Writing (review and editing)	X
Visualisation	X
Project administration	X
Funding acquisition	-

3.3.4 Publications

This chapter is based on the following publications:

- Garba-Ouangole S, **Bourner J**, Mbrenga F, Gonofio E, Selekon B, Manirakiza A, Kalthan E, Malaka C, Boum Y 2nd, Olliaro P, Nakouné E. Laboratory Diagnosis of Mpox, Central African Republic, 2016-2022. Emerg Infect Dis. 2023 Sep;29(9):1846-1849. doi: 10.3201/eid2909.230514. Epub 2023 Jul 12. PMID: 37437563; PMCID: PMC10461668.

3.4 Tecovirimat for Monkeypox in Central African Republic under Expanded Access

There are clearly a number of complications affecting endpoint selection for mpox. In the absence of a standardised definition for lesion stages, clinical assessment is prone to substantial measurement bias, potentially affecting the reliability of trial data. Prospective studies investigating viral clearance through serial lesion sampling from diagnosis through post-treatment follow-up are crucial to determining whether viral load can serve as a meaningful measure of patient outcomes. Additionally, these virological data would need to be evaluated alongside clinical data on signs and symptoms to establish whether viral clearance correlates with how a patient feels, functions and survives.

We therefore conducted an Expanded Access Programme (EAP) in CAR, in which clinical and virological data were prospectively collected from a cohort of patients with suspected mpox. The EAP provided patients with suspected mpox with access to tecovirimat, which otherwise would have been unavailable. EAPs – sometimes referred to as "compassionate use" – act as a regulatory framework within which patients who have a serious illness can access treatment with an Investigational Medicinal Product (IMP) where no appropriate treatment option exists. [171]

Tecovirimat is an antiviral medication aimed at the treatment of orthopoxvirus infections, including smallpox and mpox. While it received regulatory approval in some countries for mpox (mainly under exceptional circumstances in response to the Clade IIb mpox outbreak), it had not at the time of this study received approval in CAR. Tecovirimat works by inhibiting the viral envelope protein VP37, limiting its ability to reproduce and cause disease progression. [14] At the time of conducting this study, no clinical trials of tecovirimat had taken place in humans with mpox infection. Only studies in healthy volunteers had been conducted demonstrating that tecovirimat is generally well-tolerated, with a favourable safety profile.

While the primary objective of the programme was to provide patients with access to treatment with tecovirimat, secondary objectives were to describe the clinical and virological characteristics of the cohort and describe their overall outcomes following treatment.

This prospective data collection aimed to highlight the clinical and virological characteristics of a patient cohort in a Clade-I endemic region, including presenting signs and symptoms, clinical outcomes, and changes in Ct values over time. The findings are intended to support the selection of relevant endpoints in future clinical trials.

3.4.1 Methods

The methods and results of this study are reported according to the CONSORT checklist, which has been adapted for a non-randomised research programme.

The EAP was Sponsored by the University of Oxford and received ethical approval from the Oxford Tropical Research Ethics Committee (OxTREC) (ref: 1-20) and the Comité Ethique et Scientifique de la Faculté des Sciences de la Santé de l'Université de Bangui (ref: 10/UB/FACSS/IPB/CES/023). The study is registered on ISRCTN (ISRCTN43307947).

A partnership agreement between the University of Oxford, Institut Pasteur de Bangui and Ministry of Health in CAR supported the operationalisation of the programme.

Participants

The EAP provided patients with clinically-suspected mpox with access to tecovirimat under a standardised treatment protocol. Treatment was provided to all eligible patients at two health centres, Mbaiki district hospital and Bangui hospital. Patients with clinically suspected mpox were identified through the national mpox surveillance system and active case finding in the Bangui, Lobaye, Sangha Mbaéré and Manbéré-Kadéï prefectures. Suspected cases were notified to the EAP Principal Investigator (PI) in CAR, who was deployed to the patient's village or health centre to evaluate the patient for eligibility and invite the patient to join the EAP. All eligible patients were provided with an information sheet about the EAP and had the opportunity to ask questions and discuss participation with their family or caregiver and the PI.

Patients who were eligible to join the EAP and access treatment with tecovirimat had to meet the following criteria:

- $\geq$13kg in weight,
- clinically suspected to have mpox and awaiting laboratory confirmation or had laboratory confirmation by PCR of mpox,

- willing and able to provide informed consent to participate in the EAP.

Patients who were enrolled on the basis of a clinical suspicion of mpox prior to laboratory confirmation were removed from the study immediately upon receipt of a negative PCR result for mpox.

Patients were excluded if they met the following criteria:

- taking repaglinide or midazolam,
- intolerant to galactose,
- lactase deficiency,
- malabsorption of glucose-galactose.

Before tecovirimat could be administered and data collected on the CRF, all participants had to sign a consent form. Participants under 16 years of age were asked to assent to participate and a guardian was asked to sign the consent form. Participants for whom informed consent was obtained were transferred to Mbaiki district hospital or Bangui hospital for treatment with tecovirimat and observed until clinically well enough to be discharged.

Intervention

All patients who consented to participate in the EAP received tecovirimat (TPOXX, SIGA Inc.) [172] in the form of 200mg immediate-release capsules for 14 days according to the dosing schedule described in Table 3.9. Tecovirimat was administered by the PI who provided the patient with a meal containing moderate or high fat content 30 minutes before each treatment dose.

Table 3.9: Dosing schedule for adults and children

Weight	**Dosage**
Adults above 40kg	
40kg to 120kg	600mg twice daily
≥120kg	600mg three times daily
Adults below 40kg and children	
13kg to 24kg	200mg twice daily
25kg to 39kg	400mg twice daily
≥40kg	600mg twice daily

Study procedures

After 14 days patients were assessed for discharge. If patients remained clinically unwell or returned a positive PCR result for mpox on D14, they would remain in hospital until well enough for discharge and they received a negative PCR result for mpox.

Blood, lesion and throat samples were collected on D1, D4, D8, D14, D21 (for patients who tested positive for mpox at D14 only) and D28. All samples were tested by PCR for mpox at each timepoint and the overall results with the Ct value were captured on the CRF. Clinical data, including signs and symptoms, vital signs and adverse events, were collected daily during hospitalisation and at D28. Voluntary malaria RDT, HIV and pregnancy testing were offered to patients, as appropriate according to standard practice, following consent.

Outcomes

The clinical, virological and safety outcomes evaluated in this EAP are described in Table 3.10. Pregnant women were followed-up until the outcome of their pregnancy was known, which was subsequently captured on the CRF.

Sample size

As EAP primarily aimed to provide treatment with tecovirimat to patients, no formal sample size was calculated. However, ethical approval was obtained for inclusion of 100 participants.

Analysis

The primary analysis population includes only patients who received a positive PCR result for mpox at baseline and treatment with tecovirimat. There is no comparator group.

The data collected in this EAP have been used to summarise the prevalence of signs and symptoms among the patient cohort presented as the number and percentage patients who presented with the signs and symptoms at baseline, end of treatment (D14) and end of follow-up (D28).

Time to lesion resolution is presented as the median number of days for all lesions to have resorbed, crusted or desquamated in the patient cohort.

Clinical outcomes are presented as the number and percentage of patients in each category of the ordinal scale described in Table 3.10.

Table 3.10: Outcomes of interest

Outcome	**Outcome measure**
Clinical	Time to lesion resolution from treatment start until D14. Lesion resolution is defined as the first day on which all lesions are resorbed, crusted or desquamated and mucosal ulcers have healed, in the absence of any serious complication.
	Overall clinical status of the patient from treatment start until D14 and D28 assessed using an ordinal scale: • absence of active lesions and no serious complications • active lesions and absence of serious complications • serious complications of mpox • death
Virological	Virological status is summarised according to the presence of viral DNA in lesion, blood and throat samples at D4, D8, D14 and D28.
Safety	Number and type of Serious Adverse Events (SAEs) reported from consent until D28.

Virological outcomes were assessed by reporting the number and percentage of patients with a PCR-positive result at D4, 14, 21, and 28 for each sample type. Additionally, median Ct values with interquartile ranges (IQR) are summarised for blood, lesion, and throat samples. The agreement between Ct values obtained using G2RG and C3L primers was evaluated through the intra-class correlation coefficient. Differences in Ct values across sample types were analysed for statistical significance using the Kruskal-Wallis test.

SAEs have been reported according to the Medical Dictionary for Regulatory Activities (MedDRA) [173] Preferred Term (PT) and System Organ Class (SOC) and summarised as the number and percentage of patients for whom each event was reported. Causality between the event and tecovirimat was performed independently by the PI and a Medical Monitor. In the event the causality assessment differed between parties, the assessment with the strongest relationship to tecovirimat was selected. Causality was categorised as:

- unrelated to tecovirimat,
- unlikely to be related,
- possibly related,
- probably related,
- definitely related.

Events considered as either unrelated or unlikely to be related to tecovirimat received an overall classification of "unrelated" and events considered as possible, probably or definitely related to tecovirimat received an overall classification of "related".

Descriptive analyses were performed in R version 4.2.1. [34] Figure plots were generated using the "ggplot" package.

3.4.2 Results

This study took place between December 2021 and December 2023 and enrolled a total of 31 participants with suspected mpox, of whom 26 (84%) received a positive PCR result and 24 (77%) tested positive for Clade I mpox using the C3L clade-specific primer. The five (16%) participants who tested negative for mpox on samples collected at baseline

were removed from the study and treatment was terminated immediately upon receipt of laboratory results. All 26 participants who tested positive for mpox completed treatment according to the protocol and are included in the final analysis.

Demographic characteristics

The primary analysis population includes 11 male participants and 15 female participants with a median age of 23 years (**Table 3.11**). One female participant was pregnant at the point of enrolment. Of the 23 participants who were tested for malaria at baseline, 21 (91%) were positive. One (4%) participant was positive for HIV. The median time from symptom onset to treatment start was 12 days.

The D28 visit was completed by 23 (88%) participants, of whom 14 completed a delayed visit beyond D28 due to logistical reasons, for which the median visit day was D49 (range: D35-D77). The D28 visit was not completed by two participants for logistical reasons and one participant died before the visit could be completed.

Table 3.11: Demographic characteristics

Demographic characteristics	
Male : female	11:15
Age (years): median (range)	23 (4-42)
Pregnant women	1/15 (7%)
Comorbidities	
Malaria: n/N (%) patients	21/23 (91%)
HIV: n/N (%) patients tested	1/26 (4%)
General characteristics	
Time (days) from symptom onset to treatment start: median (range)	12 (5-45)

Clinical signs and symptoms at baseline and end of treatment

Clinical signs and symptoms at baseline were varied, with the most common symptoms including lymphadenopathy which affected all participants, headache affecting 25/26 (96%) participants and muscle and joint pain affecting 21/26 (81%) participants (**Table 3.12**). All other signs and symptoms were recorded in fewer than 20 participants. No serious or life-threatening signs or symptoms were reported at baseline.

By the end of treatment and end of follow-up, the prevalence of all signs and symptoms had substantially reduced to affect between zero and four participants, except for

lymphadenopathy which was still present in 25/26 participants at D14 and 14/23 participants on D28 (**Table 3.12**).

Table 3.12: Number and percentage (n/N (%)) of patients for which each sign or symptom was recorded on D1, D14 and D28

Signs and symptoms	**D1**	**D14**	**D28**
Active lesions	20 / 26 (77%)	10 / 26 (38%)	6 / 23 (26%)
Scabbed lesions	14 / 26 (54%)	10 / 26 (38%)	5 / 23 (22%)
Lymphadenopathy	26 / 26 (100%)	25 / 26 (96%)	14 / 23 (61%)
Headache	25 / 26 (96%)	3 / 26 (12%)	1 / 23 (4.3%)
Muscle pain	25 / 26 (96%)	4 / 26 (15%)	2 / 23 (8.7%)
Fever (temperature >38·0 °C)	21 / 26 (81%)	1 / 26 (3.8%)	1 / 23 (4.3%)
Joint pain	21 / 26 (81%)	3 / 26 (12%)	2 / 23 (8.7%)
Upper respiratory symptoms	16 / 26 (62%)	3 / 26 (12%)	0 / 23 (0%)
Cough	11 / 26 (42%)	2 / 26 (7.7%)	0 / 23 (0%)
Lower respiratory symptoms	9 / 26 (35%)	2 / 26 (7.7%)	0 / 23 (0%)
Eye complications	3 / 12 (25%)	1 / 12 (8.3%)	0 / 10 (0%)
Keratitis	5 / 26 (19%)	4 / 26 (15%)	4 / 23 (17%)
Pharyngitis	2 / 12 (17%)	0 / 12 (0%)	0 / 10 (0%)
Diarrhoea	3 / 26 (12%)	0 / 26 (0%)	0 / 23 (0%)
Encephalitis	1 / 12 (8.3%)	1 / 12 (8.3%)	0 / 10 (0%)
Deep tissue abscess	1 / 12 (8.3%)	1 / 12 (8.3%)	1 / 10 (10%)
Vomiting	1 / 26 (3.8%)	0 / 26 (0%)	0 / 23 (0%)

Active lesions were present in 20/26 (77%) participants at baseline and scabbed lesions were reported in 14/26 (54%) participants. The number of patients with active and scabbed lesions reduced by end of treatment and end of follow-up. Active lesions were detected in 10 (38%) participants at D14 and by six (26%) participants at D28. Scabbed lesions were detected in lower numbers of participants at both timepoints. The median time to lesion resolution was eight days, with a range of one to 29 days (**Figure 3.6**).

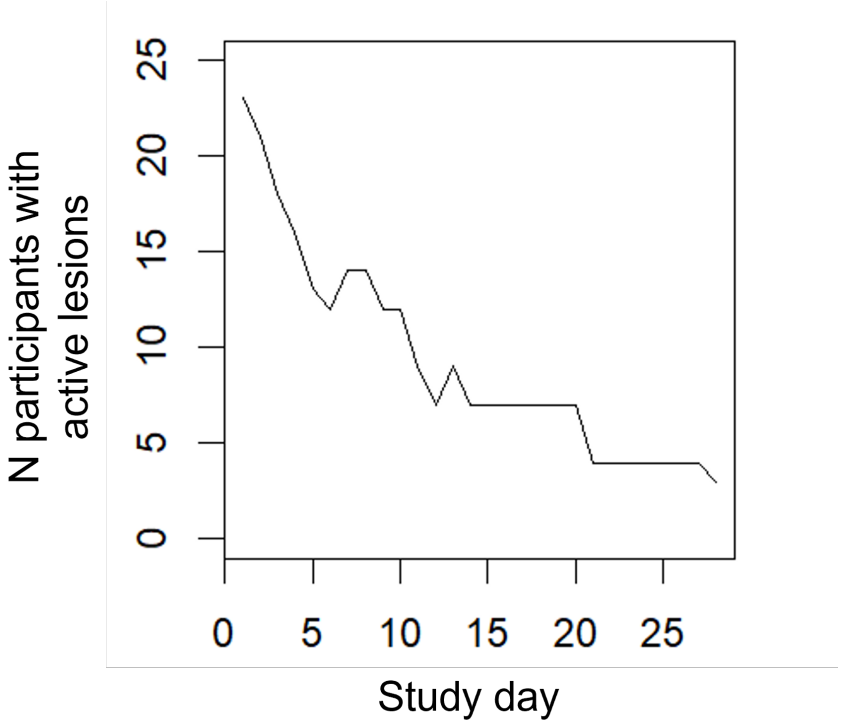


Figure 3.6: Time to lesion resolution

At the last patient visit, all lesions had completely resolved with no serious complications for 11 (45%) participants, active lesions had resolved with no serious complications for nine (39%) participants and active lesions remained without serious complications for one (4%) participant. Two (9%) participants remained hospitalised with serious complications at D28 and there was one (4%) death at D17.

Virological results

At baseline blood specimens were positive on G2RG in 19/26 (73%) participants, active lesions in 14/16 (88%) participants, and scabbed lesions in 13/14 participants. By the end of treatment (D14), nine participants remained positive on blood (3/17 (18%) participants), active lesion (7/8 (88%) participants), or scabbed lesion (8/10 (80%)) samples. By the end of follow-up, six participants remained positive on either blood (3/17 (18%)), active lesion (5/5 (100%)) or scabbed lesion (4/5 (80%)) samples (**Figure 3.7**).

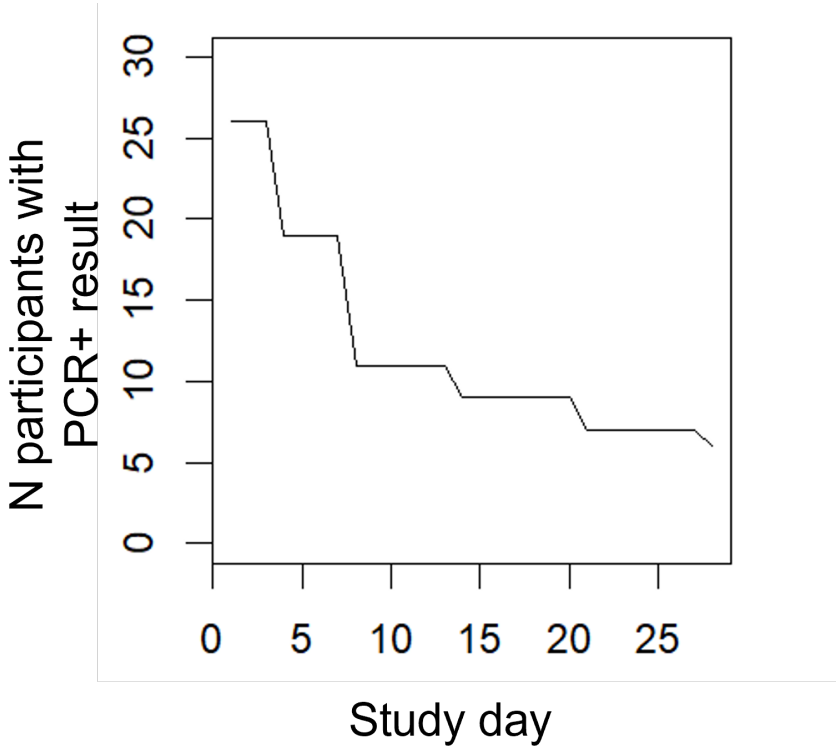


Figure 3.7: Number of participants with a PCR positive result on each study day, D1 to D28

Of the specimens positive on G2RG at baseline, the median Ct was 32.2 in blood, 24.6 in active lesion and 21.4 in scabbed lesions (**Table 3.13**).

Ct values were significantly higher in blood samples from D1 until D28 than in active or scabbed lesion samples. There was excellent reliability between the results generated on G2RG and by C3L primers (ICC: 0.925 [95% CI: 0.886-0.952]).

Safety

Between the point of consent and the D28 visit, six SAEs were recorded in five (19%) participants (**Table 3.14**). Two events – peritonitis and vaginal infection – were observed in one participant. Three events – tuberculosis, vaginal infection and trismus – were present before enrolment in the EAP and worsened during the course of treatment.

Four events were life-threatening, one event resulted in prolonged hospitalisation and one event resulted in death, which sadly appears to be a result of homicide. All remaining patients either saw an improvement in their condition or made a full recovery.

All events were considered to be unrelated to tecovirimat.

Table 3.13: Number and percentage (n/N (%)) participants positive on blood, scab and active lesion samples from D1 to D28 with median (sd) [min-max]) Ct values

	Blood		**Crust**		**Active lesion**	
	PCR +	CT-value	PCR +	CT-value	PCR +	CT-value
D1	19 / 26 (73%)	32.2 (4.9) [20.6-41.8]	13 / 14 (93%)	21.4 (5.5) [15.0-32.0]	14 / 16 (88%)	24.6 (7.4) [17.2-39.1]
D4	9 / 23 (39%)	35.7 (3.4) [29.4-39.9]	10 / 13 (77%)	20.8 (4.9) [14.2-30.4]	11 / 15 (73%)	20.7 (4.0) [16.0-28.4]
D8	5 / 19 (26%)	31.0 (8.7) [15.7-36.6]	9 / 11 (82%)	24.5 (5.9) [14.9-32.7]	7 / 11 (64%)	24.4 (7.8) [14.8-37.0]
D14	3 / 17 (18%)	32.9 (1.6) [31.4-34.6]	8 / 10 (80%)	18.7 (4.3) [13.2-26.0]	7 / 8 (88%)	22.8 (5.0) [15.9-28.2]
D21	1 / 8 (12%)	39.5 (NA) [39.5-39.5]	6 / 7 (86%)	23.4 (8.2) [13.2-33.4]	5 / 6 (83%)	22.7 (6.8) [16.5-32.7]
D28	3 / 17 (18%)	35.0 (5.7) [28.8-39.9]	4 / 5 (80%)	19.9 (3.1) [16.2-23.2]	5 / 5 (100%)	22.1 (6.6) [15.6-31.3]

Table 3.14: Summary of Serious Adverse Events

Event	**Grade**	**Seriousness**	**Study day on which the event started**	**Study day on which the event ended**	**Outcome**	**Causality with tecovirimat**
Anaemia	4	Life-threatening	D9	D14	Recovery without sequelae	Not related
Unexplained death[1]	5	Death	D48	D48	Death	Unlikely to be related
Tuberculosis (recurrent)	2	New or prolonged hospitalisation	D6	D15	Improvement in condition	Not related
Vaginitis [2]	4	Life-threatening	D2	D17	Recovery without sequelae	Not related
Peritonitis [2]	4	Life-threatening	D12	D26	Recovery without sequelae	Unlikely to be related
Trismus	4	Life-threatening	D4	D4	Recovery without sequelae	Not related

3.4.3 Discussion

This study represents the first published data on a prospective cohort of patients with Clade I mpox.

Viable endpoints for mpox remain unclear. In particular, "time to lesion resolution" may not accurately reflect changes in patients' clinical status over time, especially by the end of treatment. At baseline, 22 patients (85%) had active lesions, which decreased to six patients (23%) at the end of treatment and three patients (12%) by the final follow-up visit, which was beyond D28 for several patients (Figure 3.6). A key challenge with using lesion resolution as an endpoint is that some patients may not present with active lesions at baseline – which was the case for 23% of patients included in this study – making it impossible to calculate time to lesion resolution for this subgroup. However, excluding these patients from trials – if the presence of active lesions is used as an inclusion criterion for endpoint calculation – would be inappropriate, as it would not be possible to determine the clinical impact of an intervention on mpox for approximately one quarter of patients.

Second, throughout the study period the number of patients with active lesions fluctuates, which is unexpected as active lesions are not anticipated to reappear once resolved. The fluctuation also does not correlate with the presence of signs and symptoms as by the end of treatment approximately 10 patients still had active lesions but few patients had any other signs and symptoms – most patients were well enough to be discharged. This fluctuation is however likely a result of the challenges accurately counting and categorising a high volume of lesions in multiple patients on a daily basis. Patients with clade I mpox typically have a high lesion burden and, in this particular programme, up to 8000 lesions were counted in some cases. [174]

The final important challenge is the definition of an "active", "scabbed" and "resolved" lesion and the meaning of each of these stages in the course of illness. [131] Recent studies have shown wide variability between the categorisation of mpox lesions by experienced clinicians highlighting the risk of serious measurement error that could be associated with outcomes that rely on evaluating lesion stages. [175] [158] Baseline data show the presence of both active and scabbed lesions in some patients, whereas by D14, at least nine patients have neither active nor scabbed lesions. However, challenges in consistently clas-

sifying lesions demonstrated in the above-cited studies generate some uncertainty about the accuracy of these findings.

Time to PCR negativity was also assessed as a potential endpoint. Blood samples were taken at baseline, then on D4, 8, 14, 21 and 28. At D14 (the end of treatment) approximately half of the patients were still positive on PCR for mpox (Figure 3.7), which considering most patients reported a resolution in their clinical symptoms could mean that PCR negativity is not the most informative indicator of treatment response. There is also little correlation between the presence of mpox signs and symptoms in Table 3.12, the presence of active lesions and PCR negativity, which complicates understanding of the significance of each clinical feature in the course of mpox illness. The limited understanding about the suitability of this potential endpoint may however be a result of the sample collection method using blood rather than lesion or saliva samples for a large part of the cohort, which have a much higher viral load detectable for a longer period of time. [176]

While a number of the signs and symptoms in Table 3.12, such as headache, muscle and joint pain, appear frequent at baseline and sharply reduce by D14 and D28, it is not clear what relationship these signs and symptoms have with clinical recovery from mpox infection, nor are they particularly clinically meaningful. They, along with other common symptoms like fever, are also associated with malaria, which was highly prevalent among the study cohort. Although there was one death and three SAEs recorded, none could be attributed to disease progression or treatment with tecovirimat. Outcome measures relating to the frequency of signs and symptoms of Clade I mpox are therefore unsuitable candidates for use as part of an endpoint.

While surrogate endpoints such as time to lesion resolution or PCR negativity could be useful for a disease like mpox, which has no viable clinical outcome measures, their measurability generates substantial risks for the interpretation of the trial results and importantly their significance in the course of mpox illness remains unclear. Ultimately they would not meet regulatory requirements for surrogate endpoints in that these outcome measures would lack the substantial clinical evidence to demonstrate the surrogate outcome correlates with clinical benefit and their representation of clinical changes is not well understood. [48]

Several challenges with data collection in this study may affect the validity of the results. First, sampling strategies were inconsistent across the cohort. For the first 14 cases, only blood samples were collected; thereafter, attempts were made to collect multiple samples from multiple anatomical sites at each visit. However, the number of oropharyngeal and scabbed lesion samples collected remained low. Second, the final visit, scheduled for D28, occurred much later for 14 participants – up to D77. As a result, the clinical and virological status at D28 is unclear for approximately half of the cohort. Finally, challenging field conditions made it difficult to validate and quality-assure some of the data. These limitations should be considered when interpreting the results.

Author contribution to "An Expanded Access Programme of tecovirimat for Clade I mpox in the Central African Republic"

Role description	Author contribution
Conceptualisation	-
Methodology	X
Software	-
Validation	X
Formal analysis	X
Investigation	-
Resources	-
Data curation	X
Writing (original draft)	X
Writing (review and editing)	X
Visualisation	X
Project administration	X
Funding acquisition	-

Publications

This chapter is based on the following publication:

- Mbrenga F, Nakouné E, Malaka C, **Bourner J**, Dunning J, Vernet G, Horby P, Olliaro P. Tecovirimat for Monkeypox in Central African Republic under Expanded Access. N Engl J Med. 2022 Dec 15;387(24):2294-2295. doi: 10.1056/NEJMc2210015. Epub 2022 Nov 30. PMID: 36449745; PMCID: PMC10117058.

Note: A subsequent manuscript describing the remainder of the study cohort is pending publication.

3.5 Contribution of the chapter

This chapter explores potential endpoints for use in Clade I mpox trials. Although several trials have used "time to lesion resolution" to evaluate the efficacy of study interventions, evidence from section 3.2 suggests that substantial differences in clinician assessment of lesion stages may lead to misclassification of "active" versus "resolved" lesions. Furthermore, findings from section 3.4 indicate that lesion stages may not correlate with clinical response to treatment.

A similar issue arises with virological outcomes as indicators of treatment response. Evidence from section 3.4 suggests that there is unlikely to be a meaningful relationship between virological outcomes – such as MPXV positivity or viral load – and overall clinical improvement. Most patients enrolled in the EAP were sufficiently well to be discharged by the end of treatment, despite remaining MPXV-positive. Virological outcomes may, however, be more useful for evaluating public health measures during an outbreak, such as assessing the risk of onward transmission in community settings following treatment. Since the EAP was embedded in the national surveillance system in CAR and the study team conducted active case finding without detecting further cases linked to individuals who remained positive on D14 or D28, this suggests a limited risk to the community. Therefore, virological outcomes may have limited value as endpoints in clinical trials. Nonetheless, patients enrolled in the EAP appeared to display an overall clinical improvement, even

if this did not align with virological changes – although larger studies will be needed to confirm the impact of tecovirimat on clinical outcomes in patients with Clade I mpox.

To date, only one study has been initiated primarily to evaluate tecovirimat for Clade I mpox [146]. Initial results published in 2024 demonstrated that, while tecovirimat was safe, it had no significant effect on patient outcomes between the control and investigational arms [104]. This study used saytime to lesion resolution as the primary endpoint, alongside several secondary endpoints related to lesion and virological status. Additional secondary endpoints included mortality – which was low at 1.7% – and the presence of signs and symptoms, which are still under analysis [104].

Given the low mortality and mild morbidity in both acute and long-term outcomes associated with Clade I mpox from these studies and others [177] – regardless of whether treatment is received or not – and the uncertainty surrounding suitable virological or clinical endpoints for trials, it may be valuable to explore patient perspectives. Since Clade I mpox appears to present limited clinical, virological, or public health challenges, understanding patient-reported challenges and identifying meaningful health or social improvements from their perspective could provide valuable insights for future trial design. Studies on patient-reported outcomes have not yet been conducted for any mpox clade. However, such outcomes will likely need to be clade-specific due to differences in clinical presentation and social impact based on the geographical distribution of each clade.

Challenges in identifying mpox lesions and plague bubo structures on ultrasound highlight the limitations of subjective assessments. Studies in section 2.3 and section 3.2 reported significant disagreement among assessors in evaluating these features. For future trials involving outbreak-prone infections, AI and ML could be explored to improve the identification and tracking of physical disease manifestations. Early studies have demonstrated the potential for ML to support real-time mpox diagnosis using images taken with mobile applications [178]. This approach would be particularly valuable in settings with limited laboratory capacity, where delayed diagnostic confirmation may affect timely clinical decision-making. ML has also been applied to ultrasound-based identification of anatomical structures with high sensitivity in other areas of healthcare [179, 180]. While these systems are primarily used for clinical diagnosis, they could be adapted to clinical

trials to monitor changes over time in disease manifestations, providing more objective and consistent data for endpoint evaluation. However, while these applications may improve accurate detection and tracking of physical disease manifestations, it is unclear how these innovations could be applied to diseases like mpox. Nevertheless there is opportunity for exploration of these opportunities in future studies.

Mpox exemplifies an outbreak-prone infection for which the identification of suitable endpoints remains unresolved. Current endpoints in use face several challenges, and the reliance on suboptimal measures is not unique to mpox; similar issues have been observed in diseases such as plague in previous chapters. In the case of plague, it took a considerable amount of time for the research team to recognise the problems within the endpoint and make an amendment. For mpox, "time to lesion resolution" is a widely used endpoint despite being subjective and there being evidence published about the risk of substantial measurement error. This is likely due to the urgent need for therapeutic evidence and the lack of consensus on validated alternative endpoints. This underscores the need for the research community to develop and adopt innovative approaches to endpoint selection. In Chapter 4, we will explore alternative methods that may enhance the design of endpoints challenging infectious diseases, like mpox, but primarily using Lassa fever as an example.

Chapter 4

Improving the design of viable endpoints for outbreak-prone infectious diseases

---

4.1 Introduction

Building on the insights from the previous chapters, this chapter shifts focus toward potential solutions to endpoint-related challenges in clinical trials for outbreak-prone infections, specifically plague and Lassa fever. The previous chapters identified substantial limitations in endpoint selection and measurability, which complicate the interpretation of trial results and limit the understanding of the effects of tested interventions. These challenges have had a different impact on each of the diseases studied in this project.

As discussed in detail in the previous chapter, the way forward for Clade I mpox in particular remains uncertain. Selecting appropriate endpoints has presented several challenges due to the complex nature of the disease, variability in clinical presentation, and limitations in virological assessments.

For plague, previous chapters have highlighted the issues surrounding bubo size as a clinically-meaningful patient outcome. Identifying alternative, clinically-meaningful end-

points however is challenging. Past trials have evaluated the persistence of signs and symptoms, as seen in Mwengee et al. (2006), where improvement was defined as "the persistence of any sign or symptom in a milder degree" [58]. However, this approach presents challenges due to the heterogeneous clinical presentation of plague, making it difficult to determine which specific outcomes are being affected by the intervention. Furthermore, this definition lacks clarity on whether the intervention is effective in both mild and severe manifestations of the disease, further complicating comparability across patient groups. Furthermore, evaluating clinical recovery is difficult to standardise without a composite endpoint that explicitly defines the clinical criteria for "recovery". In the absence of such standardisation, this measure is likely to remain subjective and susceptible to variability in clinician interpretation, particularly when different risk-based approaches to patient management are applied. Overall, this represents another instance of a plague-related endpoint that calls into question the validity of trial results – results which, in this case, have been presented as positive evidence supporting the use of gentamicin in the treatment of plague.

This chapter will first highlight the impact of removing the problematic "$\geq$25% reduction in bubo size" component of the IMASOY endpoint. This analysis aims to illustrate the potential consequences of relying on outcome measures that introduce high levels of bias and subjectivity, ultimately affecting the validity and interpretability of trial results.

The next sections will focus on Lassa fever. In previous chapters, no outcomes were assessed for Lassa fever, as only one clinical trial has been conducted to date, using mortality as its primary endpoint. However, as discussed earlier in this thesis, mortality is not a viable endpoint for many outbreak-prone infections, including Lassa fever, primarily due to the low number of mortality events that limit the detection of statistically significant effects between treatment groups. Given the challenges in endpoint selection for Lassa fever and the lack of precedent for alternative outcomes, this chapter establishes a consortium-led approach aimed at facilitating the collaborative sharing of knowledge and experience to identify viable, clinically-meaningful endpoints. The resulting framework is then applied to the design of a clinical trial, providing a structured methodology for endpoint selection in future studies.

Given the challenges identified for each of these diseases, questions remain around how researchers can optimise endpoint selection for outbreak-prone infections. The research conducted during the course of this project is formed of systematic reviews, validity and reliability studies, and cohort studies, and has taken several years to complete. In the event of an outbreak, these activities and the time it has taken to complete them are likely not going to be available. Other systems for optimising endpoint selection therefore need to be identified so robust, clinically-meaningful evidence can be generated in urgent situations.

To address this broader concern, the chapter will conclude by proposing a structured framework for researchers conducting clinical trials on outbreak-prone infections. This framework will aid in selecting robust and clinically relevant endpoints while identifying potential biases and limitations, with the objective of improving the reliability, comparability, and regulatory acceptance of trial endpoints for these diseases and ultimately enhancing the evidence base for novel therapeutics in challenging infectious disease contexts.

4.2 A randomised controlled trial of a ciprofloxacin monotherapy vs an aminoglycoside + ciprofloxacin for the treatment of bubonic plague (IMASOY)

This subsection presents the main results of the IMASOY trial, focusing on the impact of removing "a 25% reduction in bubo size" from the original primary endpoint. The results of both endpoints – the original and amended endpoint – are reported to show how individual components of a composite endpoint can substantially alter the overall trial outcome. By comparing these two endpoints, the analysis aims to underscore the sensitivity of a trial's conclusions to good endpoint selection.

Furthermore, in light of the findings discussed in Section section 2.2, which highlighted the methodological difficulties associated with accurately measuring smaller buboes, this subsection also reports on the distribution of bubo sizes among patients enrolled in the

IMASOY trial. This allows for a comparative assessment between the artificial buboes and those occurring in clinical cases.

4.2.1 Methods

IMASOY (NCT04110340) is a multi-centre open-label, randomised controlled trial evaluating the non-inferiority of a ciprofloxacin monotherapy against an aminoglycoside + ciprofloxacin for the treatment of bubonic plague in Madagascar. The ciprofloxacin monotherapy is currently the third-line treatment regimen in Madagascar, while an aminoglycoside + ciprofloxacin is the current first-line treatment regimen.

The trial was approved by Oxford Tropical Research Ethics Committee (45-18), Comité d'Ethique et de Recherche Biomédicale à Madagascar (Authorisation N°116-MSANP/CERBM dated 10/09/2018)), and the London School of Hygiene and Tropical Medicine (17911).

Recruitment and participants

Between January 2020 and April 2024, a total of 82 primary care centres and hospitals in 12 rural plague-endemic districts in Madagascar were activated during the annual plague transmission season, which typically runs from August to April.

Patients with clinically suspected bubonic plague, of any age and sex, were eligible if they had a recent onset of fever, one or more buboes, and residence in or travel to a plague endemic area within 14 days before symptom onset. Patients were excluded from participation if they were pregnant, had a known allergy to aminoglycosides or fluoroquinolones, tendinitis, myasthenia gravis, theophylline or warfarin use, or had already received treatment for plague in the preceding three months.

Patients who received a confirmed or probable diagnosis of plague according to the WHO case definition [70] were considered "cases" and those who did not meet this definition were considered "non-cases" (**Table 4.1**).

Trial interventions

The investigational treatment regimen consisted of oral ciprofloxacin (500mg for adults and 15mg/kg for children) given twice daily for 10 days.

The control group received an aminoglycoside (streptomycin or gentamicin) for three days followed by oral ciprofloxacin (as above) for seven days. Intramuscular streptomycin

Table 4.1: Bubonic plague case definition

	Confirmed case	Probable case
Bubonic plague case defintion	• qPCR is positive on D1 sample OR • culture is positive on D1 sample OR • there has been a seroconversion between D1 and D21 samples or a four-fold increase in antibody titre on two separate serological samples between D1 and D21	• RDT (conducted at central laboratory) positive on D1 sample AND • qPCR negative on D1 sample AND • culture negative on D1 sample AND • no evidence of seroconversion, nor a four-fold increase in antibody titre

was given at 1g twice daily to adults and 15mg/kg twice daily to children. Intravenous gentamicin was given at 2.5mg/kg.

Outcomes

The composite primary endpoint used in the final analysis was treatment failure assessed on D11, defined as:

- Death
- Persistent fever
- Development of secondary pneumonic plague
- Administration of alternative or additional treatment for plague

A secondary endpoint extended the primary definition of treatment failure by including an additional criterion: less than a 25% reduction in bubo size by D11.

Additional secondary endpoints are reported in the results publication.

Sample size

The trial defined a target sample size of 190 confirmed and probable cases to demonstrate the non-inferiority of an oral ciprofloxacin monotherapy with 90% power, based on the assumption that 90% of individuals receiving an aminoglycoside plus ciprofloxacin

would meet the primary endpoint, with a 15% non-inferiority margin and a one-sided alpha of 2.5% and allowing for 10% loss to follow-up.

Randomisation and blinding

Following consent, patients were randomised by a member of the trial team with a 1:1 allocation ratio, using a computer-generated randomisation sequence with random block sizes generated from a master list by the trial statistician and stratified by health facility. Blinding of patients and the trial team to treatment allocation was not possible due to the different treatment administration routes of the treatment regimens.

Statistical methods

While the main results publication reports the trial results among several analysis populations, which are supplemented by sensitivity analyses to test the reliability of the results under different pre-specified conditions, this chapter will focus on exploring the impact of retaining or removing the "25% reduction in bubo size" component of the composite primary endpoint. The primary analysis population in this section therefore comprises confirmed and probable cases of bubonic plague who were randomised and received the trial drugs, referred to as the intention-to-treat infected (ITTI) population.

Baseline characteristics are summarised by treatment arm, using appropriate descriptive statistics such as number and percentage for categorical variables, and median and range for continuous variables.

The frequency of each component of the composite primary endpoint is reported as the number and percentage of patients in each arm. The overall risk of treatment failure per arm was accompanied by a corresponding 95% CI, as well as the risk difference in failure rates between arms. To establish non-inferiority of ciprofloxacin compared to the control treatment, the upper bound of the 95% CI for the risk difference was required to be less than 15%.

The primary analysis of the primary endpoint was adjusted for trial site. Unadjusted results are also presented.

The median and interquartile range (IQR) of bubo sizes (in mm) measured with a digital caliper are presented for all buboes included in the ITTI analysis, as well as stratified by body zone.

4.2.2 Results

In total, 933 suspected bubonic plague patients were screened for eligibility to participate in the study, of whom 450 were enrolled and randomised, with 222 receiving either a confirmed or probable diagnosis of bubonic plague (**Figure 4.1**).

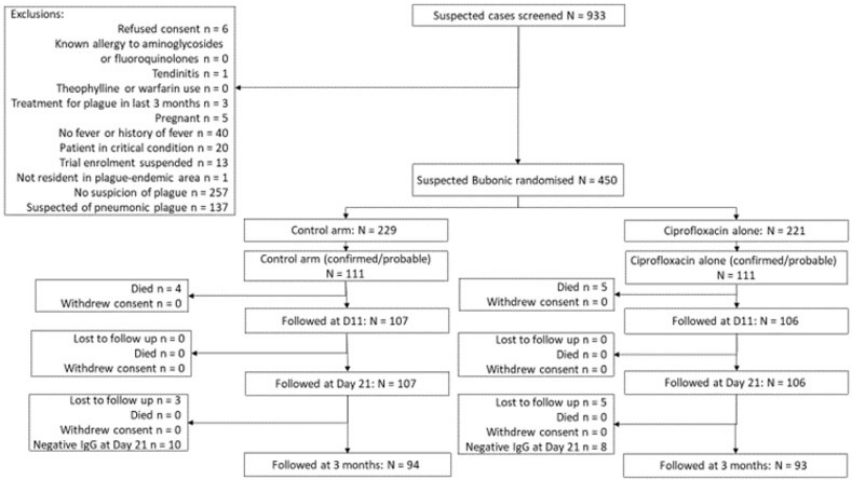


Figure 4.1: Participant flow in confirmed and probable bubonic plague patients

The median age of the study participants was 14 years (range: 2-72 years) and the male-to-female ratio was 1:1. All patients presented with at least one bubo (range: 1-5 buboes per patient), with the majority (71%) detected in the inguinal area (**Table 4.2**).

Table 4.2: Demographic characteristics

	Overall	**Control**	**Intervention**
N	222	111	111
Male (%)	118 (53.2)	63 (56.8)	55 (49.5)
Age (years)			
Median (range)	14.0 (2.0 - 72.0)	14.0 (2.0 - 72.0)	14.0 (2.0 - 64.0)
Median duration of fever (range) days	1.0 (0.0 - 7.0)	1.0 (0.0 - 7.0)	1.0 (0.0 - 7.0)
Number of buboes at baseline, median (range)	1 (1 - 5)	1 (1 - 5)	1 (1 - 3)
Bubo location			
Axillary, n (%)	41 (18.5)	18 (16.2)	23 (20.7)
Cervical, n (%)	21 (9.5)	9 (8.1)	12 (10.8)
Inguinal, n (%)	157 (70.7)	83 (74.8)	74 (66.7)

There were 9 (8.1%) treatment failures in the control group and 10 (9.0%) in the interventional group. In the primary efficacy analysis, which excluded bubo size from the composite endpoint, this generated an adjusted risk difference of 0.9 (95% CI: -6.0, 7.8), therefore meeting the criterion for non-inferiority (**Table 4.3**). However, by including bubo size as part of the composite endpoint, there were 40 (36.0%) treatment failures in the control group and 52 (46.8%) in the interventional group, with an adjusted risk difference of 10.8 (95% CI: 4.3, 17.3). Had bubo size remained part of the composite primary endpoint, the criterion for non-inferiority would not have been met.

Table 4.3: Summary of efficacy analyses with and without bubo size

	Control	Ciprofloxacin
Died by D11, n (&)	4 (3.6)	5 (4.5)
Fever at D11, n (&)	1 (0.9)	2 (1.8)
Secondary pneumonic plague, n (&)	3 (2.7)	3 (2.7)
Extra treatment before or at end of treatment, n (&)	2 (1.8)	2 (1.8)
Less than 25% reduction in bubo size, n (&)	35 (31.5)	46 (41.4)
Primary Total: treatment failure	9 (8.1, 3.8 - 14.8)	10 (9.0, 4.4 - 15.9)
Secondary Total: treatment failure	40 (36.0, 3.8 - 14.8)	52 (46.8, 4.4 - 15.9)
Primary Unadjusted risk difference (failure) % (one-sided 2.5% CI UB)	-	0.9 (-6.5, 8.3)
Secondary Unadjusted risk difference (failure) % (one-sided 2.5% CI UB)	-	10.8 (-2.1, 23.7)
Primary Adjusted risk difference (failure) % (one-sided 2.5% CI UB)	-	0.9 (-6.0, 7.8)
Secondary Adjusted risk difference (failure) % (one-sided 2.5% CI UB)	-	10.8 (4.3, 17.3)

At baseline, the median bubo size was 40.97 mm (Q1: 26.08; Q3: 60.40) (**Table 4.4**). Size appeared to vary with bubo location, with the largest buboes located in the inguinal zone (median: 43.90mm) and smaller buboes being located in the axillary (median: 33.76mm) and cervical (38.18mm) zones. By D11, the majority of buboes, particularly in

the axillary and cervical zones, were no longer measurable. Buboes located in the inguinal zone were still measurable in most cases, although the median bubo size reduced by 70%.

Table 4.4: Bubo sizes

Location	**Day 1**	**Day 11**
Overall (mm), median [Q1, Q3]	40.97 [26.08, 60.40]	0.00 [0.00, 58.53]
Axillary (mm), median [Q1, Q3]	33.76 [19.62, 48.06]	0.00 [0.00, 51.38]
Cervical (mm), median [Q1, Q3]	38.18 [25.37, 61.17]	0.00 [0.00, 37.00]
Inguinal (mm), median [Q1, Q3]	43.90 [28.03, 64.27]	13.38 [0.00, 61.39]

4.2.3 Discussion

IMASOY is the first RCT to deliver conclusive evidence on treatment for bubonic plague, confirming that a 10-day course of oral ciprofloxacin is non-inferior to an injectable aminoglycoside followed by seven days of oral ciprofloxacin.

The measurement of bubo size throughout this thesis has not been demonstrated to be clinically meaningful or objective. A key finding of the IMASOY trial, which supports the decision to exclude bubo size from the composite primary endpoint, is that the buboes measured during the trial were generally smaller than the artificial buboes described in section 2.2 and varied based on their anatomical location (**Table 4.4**). Notably, the median bubo sizes across all zones were smaller than the smallest artificial bubo (52.7 mm), meaning the risk of measurement error was likely greater in the IMASOY trial than in the artificial bubo study.

Using the original primary endpoint – which included an additional criterion of <25% reduction in bubo size – non-inferiority would not have been demonstrated. Instead, the ciprofloxacin monotherapy would have been deemed inferior to the current standard of care. Although the projects in section 2 demonstrated the lack of clinical meaningfulness of a reduction in bubo size as an indicator of treatment response, it is clear that the combination of an aminoglycoside with ciprofloxacin has a statistically significant impact on this outcome.

Had the primary endpoint not been modified, rejecting a ciprofloxacin monotherapy could have had several negative implications for public health interventions in Madagascar. First, as many of the health centres involved in the trial had limited bed capacity (with some having only a single available bed), an oral ciprofloxacin monotherapy would reduce the burden of plague on the healthcare system by allowing patients to be treated at home, thereby reserving health centre resources for patients with more urgent care needs. Second, regimens containing an aminoglycoside are significantly more expensive than an oral ciprofloxacin regimen. The cost of gentamicin is approximately $14 USD for intravenous administration and $6.50 USD for intramuscular injection, whereas ciprofloxacin capsules cost only $0.75 USD. Adopting a ciprofloxacin monotherapy could therefore lead to substantial cost savings for the public health system. Additionally, indirect costs borne by patients – such as transportation, accommodation and subsistence expenses – could be significantly reduced.

Finally, ciprofloxacin is generally considered to have a more favourable safety profile than streptomycin, particularly in terms of ease of administration and lower risk of severe organ toxicity. Adverse events associated with streptomycin are well-established, including nephrotoxicity and ototoxicity, which may be irreversible and particularly concerning with prolonged use or in vulnerable populations. [181] It is also designated as FDA pregnancy category D due to documented foetal risk. [182] While ciprofloxacin, on the other hand, avoids the organ-specific toxicities typically associated with aminoglycosides, it is not without significant safety concerns. Fluoroquinolones have been linked to tendon rupture, peripheral neuropathy, and aortic aneurysm. [183] Moreover, while ciprofloxacin is not formally contraindicated in pregnancy, caution is advised due to concerns about cartilage toxicity, particularly in developing foetuses and children. While ciprofloxacin carries its own risks, given its overall safety advantages and practical benefits, it may represent a safer and more feasible alternative to streptomycin for the treatment of plague in many clinical contexts.

Section 2 also examined other potential challenges associated with the composite endpoint, including discrepancies in severity between endpoint components and unequal event frequencies. While differences in the clinical importance of these events have already been

discussed earlier in this thesis, Table 4.3 indicates minimal variation in the observed frequencies of events in the trial data related to the amended primary endpoint. In the original endpoint, however, reduction in bubo size was the most frequently observed event, occurring in 35 patients in the control group and 46 in the investigational group. The other events included in the amended endpoint affected between one and five patients only. If the original endpoint had been retained, concerns could have arisen regarding its validity, as the least clinically meaningful component would have been the most frequently observed. Although aminoglycosides may significantly affect bubo size, there appears to be no meaningful difference in the frequency of other endpoint events, either overall or between treatment arms.

Nevertheless, a challenge remains regarding the final component of the endpoint – the decision to continue treatment beyond day 11 – which may be confounded by other potential infections and is highly subjective, evidenced by the fact that there were nine patients who either developed pneumonic plague or had persistent fever on D11 but only four continued treatment. Since there is no statistically significant difference between treatment arms and no notable discrepancy in the frequency of this event relative to others included in the endpoint, the inclusion of this component does not affect the overall interpretation of the results.

Overall, this study underscores the inherent complexity in selecting an appropriate composite endpoint for a clinical trial, especially when the individual components differ in both clinical severity and subjectivity. The amendment to the original composite endpoint was made to minimise the influence of a component deemed less clinically meaningful and prone to measurement bias. However, this decision, while justifiable, does not eliminate the broader challenges associated with interpreting the endpoint; specifically, each component contributes unequally to the overall effect and may either dilute or exaggerate perceived treatment benefit. In this case, both the original and amended composite endpoints carry limitations and introduce potential biases in different ways. While the final results demonstrate non-inferiority between the combination regimen and ciprofloxacin monotherapy, the clinical significance of this finding remains uncertain given the heterogeneity and subjectivity of the outcomes measured.

Author contribution to "A trial of ciprofloxacin vs aminoglycoside-ciprofloxacin for bubonic plague"

Role description	Author contribution
Conceptualisation	-
Methodology	X
Software	-
Validation	X
Formal analysis	-
Investigation	X
Resources	-
Data curation	X
Writing (original draft)	X
Writing (review and editing)	X
Visualisation	-
Project administration	X
Funding acquisition	-

Publications

The article associated with this chapter has been accepted by the New England Journal of Medicine and is pending publication.

4.3 A standardised Phase III clinical trial framework to assess therapeutic interventions for Lassa fever

The path forward for Lassa fever presents different challenges to plague. Given that over 40 years have passed since the last randomised controlled trial (RCT) was conducted, and at the time of conducting this project, no new trials were planned, there remain significant uncertainties in endpoint selection. The trial by McCormick et al. (1986) [92] highlighted methodological limitations that led to inconclusive evidence, including the use of historical controls and retrospectively pooled data. [93] This study relied on a mortality endpoint; however, subsequent reanalysis of the available data has revealed that the estimates reported by McCormick et al. may be inaccurate and misleading. [94] Notably, recent findings suggest that ribavirin could be harmful to certain subgroups of patients. These uncertainties underscore the urgent need for well-designed clinical trials to establish robust evidence for the treatment of Lassa fever.

Before a new era of Lassa fever clinical trials can begin, robust methodological frameworks must be established to ensure that trials are conducted consistently and generate reliable, comparable data. This project aims to enhance the comparability of Lassa fever studies and accelerate the evaluation of Lassa fever therapeutics by defining the foundational framework on which future trials can be conducted. To achieve this, we convened a consultation group comprising clinicians and researchers with prior experience in Lassa fever patient management or research – individuals most likely to be involved in future trials. The group was tasked with identifying the core components essential for a Phase III pivotal trial, including eligibility criteria, case definitions, outcome measures, and key data variables required to assess new and existing therapeutics (**Box 1**).

Box 1. Project outputs and definitions

The definitions of the four key outputs of the consultation are as follows:

Core Eligibility Criteria (CEC)–the characteristics of the study population

Core Case Definition (CCD)–how to identify a patient with confirmed Lassa fever

Core Outcome Set (COS)–what outcomes to measure in order to assess treatment efficacy

Core Data Variables (CDV)–list of recommended data variables for collection in all patients to standardise the characterisation of Lassa fever

While the development of COS is supported by the Core Outcome Measures in Effectiveness Trials (COMET) Initiative [184], the group also recognised the value of developing additional outputs (CEC, CCD and CDV) that would improve the comparability of Lassa fever clinical trials.

The full methodology and results for this project are described in Olayinka et al. (2022) [185]. This section focuses on the development of the COS.

4.3.1 Methodology

The methodology of this project is reported according to the COS-STAR (Core Outcome Set-STAndards for Reporting) Statement. [186]

The first step in development a COS is to establish the outcomes that have previously been used in clinical studies of the disease of interest. [184] However, as only one clinical trial for Lassa fever had been completed at the point of conducting this project, this approach was not feasible. Instead, as reported in subsection 1.1.1, a systematic review [63] was conducted to establish the important clinical features of the disease and common patient outcomes that may inform the subsequent stage of the project development. Second, a stakeholder consultation was established, using a modified-Delphi methodology, to obtain consensus on the key elements of a COS.

A stakeholder group, consisting of experts in the field of Lassa fever, was convened to participate in the COS development. Stakeholders were selected based on experience of Lassa fever patient management and/or who had been involved in clinical research for Lassa fever. In total 43 individuals were invited to participate in the consultation group, of whom four declined. The membership of the group covered 16 organisations across

six countries. Of the 39 individuals who agreed to participate, five were involved in the central coordination, development and analysis of the consultation surveys and therefore did not complete the questionnaire.

The consultation took place remotely over five stages, including:

- one initial kick-off meeting to introduce the project concept to the participating stakeholders
- two modified-Delphi consultations
- two post-Delphi consultation meetings, one of which was a final consultation meeting

Modified-Delphi process

The consultation followed a modified Delphi approach, maintaining the core structure of a traditional Delphi approach, in that an expert panel was assembled, and two rounds of questionnaires were administered, with the second questionnaire refined based on responses from the first. Following each round, a summary of the anonymised responses was shared with the wider group to facilitate consensus-building.

The consultation departed from the traditional structure of a Delphi methodology in two ways. First, a small pilot of the first questionnaire was conducted ahead of the first round to aid the development of relevant, clear questions and response options. Second, stakeholder meetings were also held after each survey round, in which members could discuss the complex themes that emerged from the survey. The second and final stakeholder meeting aimed at discussing the results of the second round and building consensus on the final COS framework.

Both rounds of the Delphi survey were developed using an Excel spreadsheet, allowing stakeholders to indicate their preferences for each proposed outcome within the framework using a Likert scale. Participants could also suggest additional outcomes for inclusion and provide explanatory comments. An Excel-based questionnaire was selected over online survey tools to offer greater flexibility, enabling participants to complete the survey over multiple sessions and facilitating participation from those in regions with weak or unstable internet connectivity.

Consensus thresholds in Delphi studies vary considerably depending on response types (e.g., yes/no, Likert scale) and the subject matter. [187] Given the heterogeneity of observational research in Lassa fever and the limited availability of conclusive evidence, stakeholder opinions and experiences were expected to vary widely. Therefore, the consensus threshold for the Delphi rounds was set at a relatively low level at 60%. Greater emphasis was placed on addressing divergences in opinion through group discussions during the consultation meetings.

The systematic review informed the selection of outcomes proposed in the first Delphi round, which included a total of 145 items, 13 of which related specifically to outcome measures. The second round incorporated all items that had not reached consensus in the first round, as well as any new items proposed by stakeholders either through the survey or during consultation meetings.

Following the analysis of each round, a study report was circulated to all stakeholders, providing an anonymised summary of the results, including the percentage of consensus achieved for each item, along with additional comments from the group.

Consultation meetings

The results of the Delphi rounds were presented to stakeholders during consultation meetings, where participants had the opportunity to ask questions and provide additional input to refine the framework. Following the second Delphi round, a lack of consensus remained regarding the optimal trial outcomes. To address this, the central coordination team developed workable proposals to navigate key challenges, which served as a basis for further discussion during the meetings.

After deliberation, these proposals were presented using Zoom's polling function, allowing stakeholders to cast anonymous votes. The poll could be modified in real time and relaunched if discussions necessitated adjustments. Once all votes were cast, the results – including the percentage of votes for each option – were shared with the group, followed by an opportunity for further discussion. The option receiving the highest number of votes, provided it exceeded the 60% consensus threshold, was incorporated into the framework.

Analysis

All analyses were performed using Excel. The results of each round were presented using descriptive statistics: N and % of each response per item.

All explanatory comments and proposals were presented to stakeholders verbatim for further consideration.

4.3.2 Results

Round 1

The questionnaire was distributed to 30 stakeholders, of whom 13 (43%) returned a completed response by the deadline.

Consensus was reached for 12 (86%) of the 14 proposed outcome measures, with more than 60% of respondents indicating that they "strongly agree" or "agree" to include the outcome measure in the framework (**Table 4.6**).

The highest level of consensus (100%) was observed for survival/mortality, followed by progression to severe disease (90%) and unfavourable outcome (85%).

However, there was wide disagreement regarding the most appropriate timepoint for assessing mortality/survival, between the options of 14, 21, and 28 days. Two participants suggested that 28 days may be the most suitable to capture late-stage mortality and prolonged hospitalisation due to complications.

Consensus was achieved for the inclusion of all five elements comprising the composite outcome measure "progression to severe disease":

- Death
- Acute Kidney Injury (AKI)
- Encephalopathy
- Shock
- Acute Respiratory Distress Syndrome (ARDS)

Several additional outcome measures were proposed by the stakeholders, including survival with complications, time to PCR negativity, and resolution of fever. Additional

Table 4.6: Summary of stakeholder responses in Round 1

Outcome measure	**Strongly agree**	**Agree**	**No preference**	**Disagree**	**Strongly disagree**
Survival/ mortality	75%	25%	0%	0%	0%
14 days	50%	38%	13%	0%	0%
21 days	20%	60%	20%	0%	0%
28 days	29%	43%	14%	14%	0%
Progression to severe disease	30%	60%	0%	0%	10%
Renal failure	67%	33%	0%	0%	0%
Encephalopathy	67%	33%	0%	0%	0%
Shock	42%	50%	8%	0%	0%
Respiratory failure	42%	42%	8%	8%	0%
Progression to severe disease based on clinical assessment of unspecified criteria	8%	23%	38%	31%	0%
"Unfavourable outcome" (composite outcome of mortality AND progression to severe disease)	31%	54%	1%5	0%	0%
Serious Adverse Events (SAEs)	31%	38%	23%	8%	0%
Grade 3 and 4 adverse events	0%	42%	50%	8%	0%

outcome measures were also proposed for pregnant women, including vertical transmission, time between pregnancy outcome and discharge, time between treatment start and pregnancy outcome. No additional outcome measures were specified for children.

Round 2

In Round 2, five additional stakeholders joined the group, bringing the total to 35. The questionnaire was subsequently circulated to all 35 stakeholders, with 24 (69%) submitting completed responses. Unlike Round 1, where outcome measures were assessed differently, Round 2 required stakeholders to specify whether each outcome measure should be included as a primary or secondary outcome measure or be excluded altogether.

Of the 21 outcome measures proposed in Round 2, four (19%) achieved consensus for consideration as a primary outcome measure (**Table 4.7**). Survival/mortality received the greatest support to be considered as a primary outcome measure (96%), followed by unfavourable outcome (78%), progression to severe disease (65%) and pregnancy outcome (for trials in pregnant women) (62%).

Table 4.7: Summary of stakeholder responses in Round 1

Outcome measure	**Primary**	**Secondary**	**Exclude**
Survival/ mortality	96%	4%	0%
Survival with complications	35%	57%	9%
Haemoglobin level and other haematologic parameters	22%	52%	26%
Renal function	39%	57%	4%
Resolution of fever	43%	26%	30%
Malignancies	0%	14%	86%
Progression to severe disease	65%	35%	0%
Unfavourable outcome	78%	22%	0%
Time to hospital discharge	35%	57%	9%
Time to PCR negativity	48%	48%	4%
Time to cessation of fever	39%	35%	26%
Time to cessation of symptoms	35%	39%	26%
Presence of severe anaemia	13%	70%	17%
Hearing loss	26%	65%	9%
AST	18%	59%	23%
ALT	18%	55%	27%
Outcome of subsequent pregnancies	0%	32%	68%
Vertical transmission	36%	50%	14%
Time between pregnancy outcome and hospital discharge/ recovery	18%	68%	14%
Time between treatment start and pregnancy outcome	18%	73%	9%
Pregnancy outcome	62%	38%	0%

The proposed outcome measures that gained over 60% agreement as a secondary outcome measure were time between treatment start and pregnancy outcome (73%), presence of severe anaemia (70%), time between pregnancy outcome and hospital discharge/ recovery (68%), and hearing loss (65%).

Stakeholders agreed that malignancies (86%) and outcome of subsequent pregnancies (68%) should be excluded as core outcome measures in future trials.

Stakeholders were also asked to select a preferred definition of the clinical syndromes proposed as a component of "unfavourable outcome", as follows:

- **Acute Kidney Injury**: Proposals gaining over 60% consensus were:

  - fall in urine output to less than 0.5ml/kg/h for more than 6 hours in adults and more than 8 hours in children and young people gained the highest acceptance among the delegates
  - rise in serum creatinine of 26mmol/L or greater within 48 hours
  - 50% or greater rise in serum creatinine known or presumed to have occurred in the past 7 days

- **Encephalopathy**: No predefined definitions of encephalopathy were provided for stakeholder selection. Instead, stakeholders were asked to propose their own definitions. The most commonly suggested themes included:
  - altered mental status or sensorium that cannot be otherwise explained
  - presence of confusion, disorientation, behavioural changes, seizures
  - declining or lack of consciousness (potentially using the Glasgow Come Score)

- **Shock**: 86% of stakeholders agreed that shock should be defined based on MAP <65mmHg and lactate >2mmol/L. However, only 20% of the stakeholders stated it was possible to measure lactate at their site. Several stakeholders suggested that, if this definition is to be used, equipment to measure lactate should be provided for use in a clinical trial.

- **Acute respiratory distress syndrome**: 86% of stakeholders considered that respiratory failure should be defined as SpO2 <90% and 77% of stakeholders considered SpO2 90% with decision to start O2 therapy to be an appropriate definition.

Final consultation meeting

Of the 35 stakeholders invited to attend the meeting, 24 participated representing an attendance rate of 69%.

During the presentation of results following Round 2, a concern was raised by one of the stakeholders regarding the sample size required to facilitate a clinical trial of Lassa fever using a survival/mortality primary endpoint. During the final consultation meeting sample size estimations were shown to the stakeholders to guide discussions about the feasibility using a survival/mortality primary endpoint. It was widely agreed that the sample

sizes presented would not be achievable for Lassa fever and, upon voting, the majority of stakeholders (83%) agreed that "unfavourable outcome" would be more feasible.

During the meeting measurement instruments for the four pathologies characterising "unfavourable outcome" were proposed for evaluation.

- **Acute Kidney Injury**: 64% of the group agreed that AKI should be assessed on the basis of either the Sequential Organ Failure Assessment (SOFA) score [188] or urine output alone (where creatinine testing is not possible).
- **Acute Respiratory Distress Syndrome**: 68% of the stakeholders agreed that Acute Respiratory Distress Syndrome (ARDS) should be assessed by SOFA score using one of two methods (partial pressure of oxygen (PaO_2)/fraction of inspired oxygen (FiO_2) or oxygen saturation (SpO_2)/FiO_2)
- **Shock**: The majority of stakeholders (76%) agreed that shock can be assessed by SOFA score (with the option to not record inotropes if unavailable)
- **Encephalopathy**: Following discussions and multiple rounds of voting, stakeholders agreed that encephalopathy should be assessed on the basis of 'ACVPU + seizures'

Table 4.8 presents the final outcome framework endorsed by the stakeholders. The group also agreed that an adapted assessment system for paediatric patients was required.

Table 4.8: Final outcome framework

Body system	**Renal**	**Respiratory**	**Cardiovascular**	**Nervous**
Pathology	AKI (Acute kidney injury)	ARDS (Acute respiratory distress syndrome)	Shock	Encephalopathy
Assessment method	Creatinine; urine output	Arterial blood gas (ABG) analysis: • PaO_2 • Pulse oximetry: SpO_2	Blood Pressure: Mean Arterial Pressure (MAP)	ACVPU [189]
Acceptable definitions	SOFA [188] 0-4 (creatinine test preferred; where creatinine testing is not available, urine output is acceptable provided it is measured accurately)	SOFA [188] 0-4 • If ABG available: PaO_2/FiO_2 • if ABG not available: SpO_2/FiO_2	SOFA [188] 0-4	ACVPU [189] + seizure

4.3.3 Discussion

This project represents the consensus position of a wide range of stakeholders in the development of Lassa fever therapeutics on the core primary outcome measure that should be measured in future trials. The core primary outcome measure is a composite that has been identified as a result of widespread agreement that a survival/mortality endpoint would not be feasible. The data collected within the scope of the framework also indicate a number of other outcomes that stakeholders would be interested to measure as secondary outcomes measures.

A multi-stakeholder approach was used to lay the foundation for the collaborative and coordinated advancement of Lassa fever trials while securing the engagement of potential investigators and treatment centres. Establishing these partnerships is vital to prevent a fragmented approach to trial implementation for a disease that necessitates a well-integrated research strategy. Moreover, prioritising an efficient clinical trial pathway will

enhance resource allocation, ensuring the optimal use of the limited resources available for studies of this nature.

The involvement of clinicians and researchers from potential research sites has been instrumental in selecting outcomes that are measurable within the specific settings where trials will be conducted. Given that many of these clinical research centres operate with limited resources, and resource availability may vary between sites, the primary endpoint has been designed to accommodate these constraints. Consequently, alternative detection methods have been incorporated for both acute kidney injury (AKI), using urine output, and acute respiratory distress syndrome (ARDS), using the SpO_2/FiO_2 ratio, to ensure feasibility across diverse trial environments.

Limitations

The definition of this primary endpoint framework however is not without its challenges and limitations.

The scope of this project was limited to gathering the perspectives of clinicians and researchers, and as a result a number of integral voices are not represented. To enhance the utility and specificity of the primary outcome measure outlined above, future work must incorporate input from survivors and communities affected by Lassa fever, regulatory agencies and ethics boards. In particular, as traditional hard endpoints, such as mortality, which are commonly used in pivotal trials, are not feasible for Lassa fever, it will be essential to initiate discussions with regulatory agencies on the acceptability of composite and surrogate endpoints as viable alternatives to ensure trial efforts are not wasted.

A modified Delphi approach was employed to develop the framework. The findings of this project highlight the limitations of using the Delphi technique to achieve consensus on complex topics where limited evidence is available to guide decision-making. Notably, without the proactive contributions of certain individuals – who voluntarily estimated sample sizes and raised concerns about recruitment feasibility – a survival/mortality primary outcome would have been selected based on consensus alone. Furthermore, undetected issues related to the final primary outcome measure may require further evaluation before its implementation in a future clinical trial. This study underscores that a Delphi survey alone is insufficient for determining trial outcomes. Additional checks and balances are

necessary before finalising group recommendations, which should be transparently documented and published alongside a core outcome set. Section 4.5 describes a framework that could be implemented when designing trial endpoints to ensure they are clinically-meaningful, reliable and feasible.

Author contribution to "A standardised Phase III clinical trial framework to assess therapeutic interventions for Lassa fever"

Role description	Author contribution
Conceptualisation	X
Methodology	X
Software	-
Validation	X
Formal analysis	X
Investigation	X
Resources	-
Data curation	X
Writing (original draft)	X
Writing (review and editing)	X
Visualisation	X
Project administration	X
Funding acquisition	-

Publications

This chapter is based on the following publication:

- **Bourner J***, Olayinka AT*, Akpede GO, Okoeguale J, Abejegah C, Ajayi NA, Akude C, Ayodeji O, Bausch DG, de Clerck H, Dan-Nwafor C, Dunning J, Erameh C, Eze JN, Formenty P, Gillesen A, Jalloh S, Jaspard M, Jegede T, Maikere J,

Malvy D, Ogbaini-Emovon E, Ojo OE, Okogbenin S, O'Neill K, Orji ML, Owhin SO, Ramharter M, Samuels RJ, Shehu N, Merson L, Salam AP, Kayem ND, Horby P, Ihekweazu C, Olliaro P. A standardised Phase III clinical trial framework to assess therapeutic interventions for Lassa fever. PLoS Negl Trop Dis. 2022 Jan 6;16(1):e0010089. doi: 10.1371/journal.pntd.0010089. PMID: 34990453; PMCID: PMC8769305.

*These two authors contributed equally to this manuscript and are listed as joint first authors

4.4 An adaptive design for a Phase II/III platform trial of Lassa fever therapeutics

As outlined in earlier sections, there is an urgent need to both re-evaluate ribavirin and investigate other potential therapeutic options for Lassa fever, despite the existing methodological and operational challenges, including:

- the limited number of hospitalisations each year and, as a result, potential inclusions in a clinical trial [190],
- the likely heterogeneity in treatment effect introduced by geographically widespread sites operating in different health systems with varying resource capacity, supportive care practices and experience managing Lassa fever cases,
- the limited available data on the frequencies of the outcomes described in the composite endpoint developed in the previous section, which is critical to estimate a target sample size that would detect a significant treatment effect.

These complexities underscore the critical need for future Lassa fever clinical trials to adopt standardised methodologies that account for the aforementioned challenges and to maximise resource efficiency while avoiding duplicative efforts. Future trials therefore must be collaborative – spanning multiple sites and countries. Achieving this will require consensus among a range of stakeholders on key aspects of trial design and implementation. Without a co-developed and harmonised approach, future trials risk being fragmented, underpowered, biased, and difficult to interpret across the varying settings in which they are conducted.

The West Africa Lassa Fever Consortium (WALC) was therefore established in 2021 to advance therapeutic development for Lassa fever, from clinical research to ensuring access to effective treatments. The consortium brought together over 100 stakeholders – including public health experts, academic researchers, industry representatives, and regulatory authorities – to develop a comprehensive clinical development plan for new and existing therapeutics. This plan encompasses a Target Product Profile for Lassa fever

therapeutics, a research capacity development strategy, a clinical trial proposal, and a value proposition, which are published in their entirety on the ISARIC website. [46]

This section focuses on the clinical trial proposal developed by the WALC, focusing on the key design considerations essential for ensuring that future trials generate reliable and clinically meaningful results. The pre-positioned protocol associated with this proposal has been published separately. [191]

4.4.1 Methods

A consultation group, consisting of members with broad expertise in Lassa fever, was established to develop a pre-positioned protocol for an adaptive Phase II/III randomised controlled platform trial to evaluate multiple Lassa fever therapeutics.

The consultation group was formed of 56 stakeholders representing the follow groups described in **Table 4.11**.

Table 4.10: Summary of stakeholders involved in the consortium

Member role	**Eligibility criteria**
Clinicians	Experience managing patients with Lassa fever
Clinical researchers	Involvement in the development of randomised controlled trials for emerging infections
Drug developers	Involvement in the development of novel therapeutics for Lassa fever
Ethicists	Member of an ethics committee in a Lassa fever-endemic country
NGOs	Involvement in the clinical or public health management of Lassa fever in an endemic country
Public health officials	Involvement in the public health management of Lassa fever in an endemic country
Regulators	Responsible for the oversight of clinical trials in West Africa
Social scientists	No specific criteria
Statisticians	Experience designing multi-arm, multi-stage randomised controlled trials

The majority of stakeholders represented organisations based in West Africa (57%), the rest represented organisations based in Europe and the United States (36%), pan-African organisations (4%) and international organisations (4%).

Table 4.11: Summary of stakeholders involved in the consortium

	Total N	**West Africa, N (%)**	**Pan-Africa, N (%)**	**Outside Africa, N (%)**	**International, N (%)**
Clinical researchers	15	1 (7%)	0 (0%)	14 (93%)	0 (0%)
Clinician	7	7 (100%)	0 (0%)	0 (0%)	0 (0%)
Drug developer	3	0 (0%)	0 (0%)	3 (100%)	0 (0%)
Ethics committee	2	2 (100%)	0 (0%)	0 (0%)	0 (0%)
NGO	9	8 (89%)	0 (0%)	0 (0%)	1 (11%)
Public health agency	11	9 (82%)	1 (9%)	0 (0%)	1 (9%)
Regulatory body	1	0 (0%)	1 (100%)	0 (0%)	0 (0%)
Social scientists	5	5 (100%)	0 (0%)	0 (0%)	0 (0%)
Statisticians	3	0 (0%)	0 (0%)	3 (100%)	0 (0%)
Total	56	32 (57%)	2 (4%)	20 (36%)	2 (4%)

The trial design and protocol were developed through group discussion. Consultation group members were grouped in to themes to develop specific areas of the protocol, such as the overall trial design, eligibility criteria, patient pathway, among others. Meetings were scheduled on a bi-weekly basis on average, where each sub-group reported back on progress and requested group feedback. One multi-day face-to-face meeting was held in Abidjan in February 2022 to generate consensus on key design issues.

Discussions were informed by published data, unpublished data available from Lassa fever treatment centres, and via clinician representatives who were members of the consultation group.

To supplement the available data, two further research studies were conducted:

- a systematic review of supportive care guidelines (SCGs) used to inform the supportive care requirements described in the protocol,
- a survey of clinicians who work in Lassa fever treatment centres to understand, across sites, the variation in Lassa fever treatment practices and the acceptability of placebo-controlled trials.

Note: both projects are referred to in the Results section below, but full details are published elsewhere. [192]

4.4.2 Results

We designed a multi-arm Phase II/III adaptive randomised controlled platform trial with an equal allocation ratio and using a superiority framework. To optimise research effort, the trial uses a portfolio approach, which allows for the simultaneous evaluation of multiple drug candidates within a single platform. The below results summarise the elements of the study design related to the primary endpoint, but information about the study design in its entirety can be found in a separate publication. [193]

Primary endpoint Based on the results described in section 4.3 [185] and a review of data presented in published scientific literature [86] [90], the consultation group defined a composite primary endpoint of "Unfavourable outcome", consisting of death or new onset of AKI, ARF or shock (**Table 4.12**).

Table 4.12: Summary of stakeholders involved in the consortium

Parameter (any of the following)	**Measurement definition**	**Assessment time point**
Death	1) Yes, 2) No	D28
New onset of acute kidney injury	KDIGO 3	Hospital discharge or D28, whichever is earliest
New onset of acute respiratory failure	Arterial oxygen partial pressure/fractional inspired oxygen ¡315 [194] based on 2 consecutive measurements taken >4 h apart meeting the above criteria	Hospital discharge or D28, whichever is earliest
New onset of shock	Mean arterial pressure <65 mm Hg [195] based on 2 consecutive measurements taken >4 h apart meeting the above criteria	Hospital discharge or D28, whichever is earliest

The composite primary endpoint assesses the first instance of any one of the events any time after the initiation of treatment. If a patient develops two of the events during the course of treatment, only the first event will be included in the final analysis. Any events meeting the criteria described in Table 4.12 that are present at the point of enrolment (and before the initiation of treatment) would not be included in the final analysis; only

the new onset of an event that develops following the initiation of treatment would be included.

Sample size and interim analysis

Due to a lack of available published data on several events included in the composite endpoint (e.g. number of patients who experience two consecutive recordings of MAP $<$ 65mmHg or two consecutive recordings of SpO2/FiO2 $\leq$ 315), the sample size calculation will initially assume that the frequency of composite endpoint is the same as the mortality endpoint. The sample size will therefore be calculated based on a mortality rate of 15% in the control arm [90], an effect size of 33% relative risk reduction (a 10% absolute risk reduction for clinical importance), 90% power and 10% loss to follow-up, generating a target sample size of 1010 in each arm, which may take over six years to achieve for a two-arm trial.

After enrolling 300 patients, an interim analysis evaluating the overall event rate – blind to treatment allocation and including any variation by site – will be conducted by an independent Data Safety and Monitoring Board (DSMB) to assess the feasibility of achieving a sample size informed by data relating to the outcomes included in the composite endpoint (**Figure 4.2**). To account for any variation in outcomes observed between health centres, subsequent randomisation will be stratified by site.

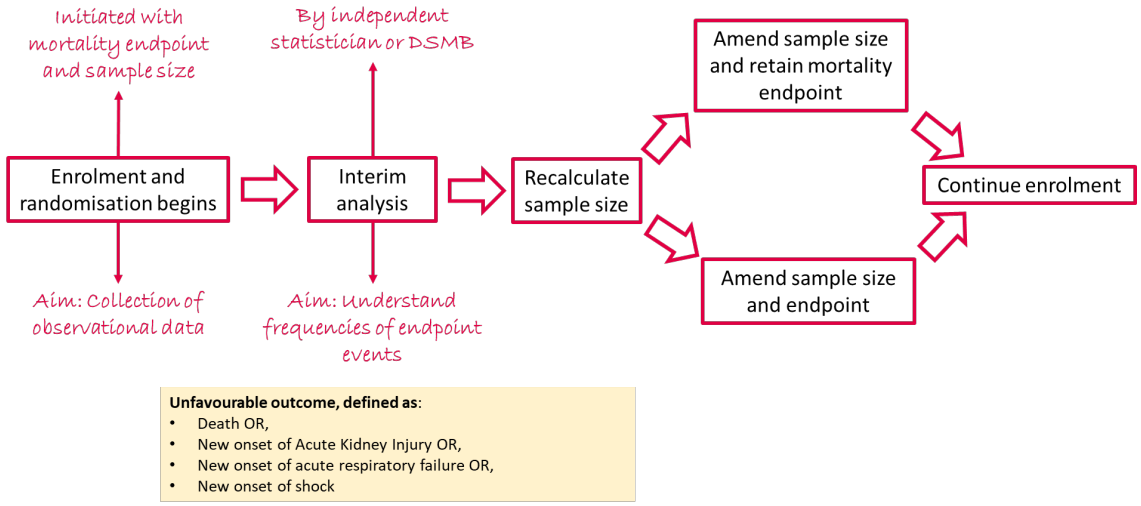


Figure 4.2: Proposed adaptive trial design

The composite endpoint will be deemed feasible if the re-estimated sample size substantially lowers the target sample size and is attainable within a shorter time frame. For example, scenarios for projected recruitment times under a new target sample size could be presented to the data monitoring committee alongside revised expectations for timing of delivery of further results and treatments for patients. In this case, the sample size will be adjusted via an amendment, and the trial will continue with the composite endpoint described in Table 4.12.

If the frequency of the events in the composite endpoint is not considerably higher than the frequency of mortality in the study population, the trial team will consider the feasibility of continuing the study with either a single endpoint of mortality or the composite endpoint. For example, there is no clear indication that a composite endpoint will reduce the sample size such that clinically relevant results will be available sooner.

To ensure trial integrity [196], details on the sample size re-estimation and decision process will not be shared with investigators to avoid indirect inferences on the interim data.

4.4.3 Discussion

This trial has been designed to efficiently generate sufficient, reliable data for multiple Lassa fever therapeutics taking in to consideration the limited available evidence on which to base design decisions and the practical, contextual challenges that may limit the implementation of the trial (e.g. differences in supportive care practices) and interpretation of results (e.g. use of ribavirin as the control arm regimen).

The design aims to prevent the need for multiple trials to competitively generate evidence across multiple different comparisons, while enabling data to be collected consistently and in a comparable manner. The design also avoids focus and investment being dedicated to one lead product without overlooking other potential candidates in the R&D pipeline and reduces time to reach a clinical decision through the existence of an established research infrastructure.

The composite primary endpoint shows some differences to the endpoint designed in the COS exercise [63]. While it has retained most of its key components – all of which are

similarly clinically meaningful – encephalopathy has been removed. Encephalopathy was widely considered to be clinically meaningful, particularly as its correlation with mortality in Lassa fever has been previously demonstrated, [86] but identifying a suitable reliable measurement instrument was challenging and there were wide variations in diagnostic practices across sites.

However, while using the composite endpoint confers many benefits for the trial's design, it is not without its challenges. In particular, the working group were mindful about selecting a composite outcome measure that is both clinically meaningful and reliably measured. The clinical meaningfulness of a single blood pressure recording of MAP $<$ 65mmHg that resolves either spontaneously or with minimal fluid therapy is unclear, particularly when no association between hypotension and mortality has been reported in large observational cohorts. [90] [86] Similarly, the clinical relevance of a single SpO_2/FiO_2 recording ≤ 315 is also unclear, as oxygen saturations naturally fluctuate. For this reason, to meet the definition of acute respiratory failure and shock, respectively, the evaluation of SpO_2/FiO_2 and MAP requires two consecutive measurements, taken at least four hours apart, that meet the threshold criteria.

Another challenge of the composite endpoint is ensuring that patients are correctly classified as being event-free at baseline. A patient with a SpO2/FiO2 = 320 on admission who subsequently develops SpO2/FiO2 ≤ 315 soon after enrolment would be classified as having an unfavourable outcome, but the difference between these two values may just be the result of natural fluctuations in vital signs or progression of a pathophysiologic pathway that was already well underway prior to the initiation of an effective antiviral therapy. This issue should, however, be resolved by randomisation.

Finally, another important challenge is determining the appropriate action that should be taken in the event that the interim analysis shows no reduction in sample size when the sample size calculation is informed by more robust data on the composite endpoint. In that scenario, based on pre-defined parameters established before the start of the trial, the DSMB would need to decide which endpoint would generate the most clinically meaningful data for the evaluation of the therapeutics included in the trial.

There are undoubtedly other important aspects that are not addressed here but can be addressed when setting up a trial. For instance, the trial could, as a secondary objective, gather information other potential outcomes, such as post-acute phase sequelae through long-term follow-up, and/or outcomes measures on which there was insufficient agreement or evidence – e.g., encephalopathy [197] and haematological alterations [198]. Collectively, these data will help improve the currently limited body of knowledge about clinical manifestations and potential outcome measures of Lassa fever.

With this trial proposal – developed in collaboration with a broad range of stakeholders in the Lassa fever research landscape – WALC aimed to catalyse research progress for a disease that has for decades remained dormant. The INTEGRATE trial has recently been launched using elements of the trial design developed by the WALC, including the composite primary endpoint. [59]

The published protocol is freely available and can be adapted by any research team who obtains funding to initiate a trial. [191]

Author contribution to "An adaptive design for a Phase II/III platform trial of Lassa fever therapeutics"

Role description	Author contribution
Conceptualisation	X
Methodology	X
Software	-
Validation	X
Formal analysis	X
Investigation	X
Resources	-
Data curation	X
Writing (original draft)	X
Writing (review and editing)	X
Visualisation	X

Project administration	X
Funding acquisition	-

Publications

This chapter is based on the following publications:

- **Bourner J**, Salam AP, Jaspard M, Olayinka A, Fritzell C, Goncalves B, Vaillant M, Edwards T, Erameh C, Ajayi N, Ramharter M, Olliaro P; WALC Work Package 2 Working Group. The West Africa Lassa fever Consortium pre-positioned protocol for a Phase II/III adaptive, randomised, controlled, platform trial to evaluate multiple Lassa fever therapeutics. Wellcome Open Res. 2023 Jun 2;8:122. doi: 10.12688/wellcomeopenres.19041.2. PMID: 39211525; PMCID: PMC11358687.

- **Bourner J**, Vaillant M, Abdel Salam A, Jaspard M, Fritzell C, Jacob ST, et al. Adaptive Design for Phase II/III Platform Trial of Lassa Fever Therapeutics. Emerg Infect Dis. 2025;31(2):9-16. https://doi.org/10.3201/eid3102.240251

4.5 The ISARIC framework for selection of endpoints in treatment trials for emerging infections

As explored in previous chapters, many emerging infectious diseases prove difficult to study for a number of reasons, particularly when survival/mortality cannot be used as the primary endpoint and identifying alternative endpoints is often challenged by the limited availability of data describing the range, frequency and heterogeneity of other important, clinically-meaningful patient outcomes (and the validated tools to measure them) that could inform endpoint selection. [63] [135] [131]. These are often the diseases for which there are limited therapeutic options available to patients and for which advances are desperately needed. To facilitate clinical research and accelerate product development, it is therefore imperative that clinical trials for such diseases are robustly designed, given the difficulty in recruiting and monitoring patients and the few opportunities there are to conduct research.

Mpox has been a prime example of this challenge, where "lesion resolution" has been used as an indicator of treatment response in several trials. [199] [134] [164] [133] [146] However as described in Chapter 3, its measurability poses some considerable challenges for which there is no clear path forward. Unlike trials for other diseases, such as solid tumours which have guidelines to support outcome measurement [19], there are limited frameworks to support trial endpoint selection for emerging infections, such as mpox. Core Outcome Sets (COS) may provide an indication of events that the research community considers to be important to a disease or syndrome, [184] but COS are developed agnostic to any specific trial design and therefore often stop short of classifying outcomes as "primary", "secondary", "tertiary" or "exploratory". Specifically, COS do not take into account important design considerations in clinical trials that assess treatment efficacy, such as outcome frequency, reliability of measurement, timing of appearance during the clinical course, amongst others.

Trials for emerging infections therefore often take place without well-defined frameworks or validated measurement instruments in place and as many outbreaks are short-lived, it would not be feasible to wait for complete and accurate disease characterisations

or validated tools to be made available. The selection of suboptimal outcome measures and tools can however be problematic; for example, they may risk measurement error, generate results with limited clinical relevance or lead to inaccurate interpretation of trial data. [158] [200] To support the generation of robust, clinically-meaningful and reliable data in trials for emerging infectious diseases, the aim of this project was to utilise an international emerging infections consortium to develop a framework that supports the selection of primary endpoints in Phase II/III therapeutic clinical trials for emerging infectious diseases.

4.5.1 Methods

Using a consensus-based approach, we conducted a multi-step structured methodology that elicited the expert opinions of key stakeholders on the design, conduct, analysis and regulatory review of clinical treatment trials. This approach followed established consensus methodologies that have previously been applied to make decisions in the field of clinical research. [164] [134] This project focused on Clinician-Reported Outcome measures (ClinROs) [201] – the assessment of outcomes that involve a clinical evaluation of the observable signs and symptoms related to a disease or condition.

The approach was conducted in four phases.

Phase 1: In the first phase, a small working group was convened to act as the steering committee to develop an initial list of endpoint characteristics and the criteria that endpoints should meet before being selected for use in a trial.

Phase 2: A two-round Delphi survey was then initiated which included the initial list of endpoint characteristics and criteria, as well as other exploratory questions to understand important endpoint criteria to different types of stakeholders. Stakeholders included senior experts in the design, analysis and regulation of clinical trials, identified through the International Severe Acute Respiratory and Emerging Infections Consortium (ISARIC) and knowledge of recently published clinical trials in the field. Of the identified stakeholders, 31 (43%) worked in organisations based in low- and middle-income countries and 41 (57%) worked in organisations based in high income countries. To delineate the scope of the questionnaire, in this and subsequent phases, the diseases of interest were referred to

as "difficult-to-study infectious diseases", the definition of which was communicated to respondents before completing the questionnaire. In each Delphi round, respondents were asked if they would include the proposed items in the framework, modify them or remove them. Consensus among the group was determined if 75% of respondents either agreed or disagreed with the inclusion of an item in the framework.

After each round we presented a bar chart showing the proportion of respondents who opted to include, modify or remove each item. In order to ensure the framework was relevant and applicable in different contexts, it was important to understand whether there was disagreement between respondents from HICs and LMICs. To allow the steering committee to highlight strong disagreement in subsequent rounds and phases for further consideration, each bar on the bar chart was therefore split in to the proportion of responses from HIC respondents and LMIC respondents. A draft framework was produced following the completion of each round for review by the Delphi participants to demonstrate the included items.

Phase 3: Following the Delphi survey, a consensus meeting was held in-person with a smaller group of stakeholders who were either involved in the small working group or the Delphi consultation to discuss the Delphi findings and finalise the framework. All elements of the draft framework were discussed during the face-to-face meeting, as well as any remaining free-text comments and considerations generated by the questionnaires. Stakeholders were selected to participate in this phase based on expertise and active contribution to either of the previous phases, ensuring diversity of trials expertise among the group. Consensus on the items included in the framework was achieved through group discussion. During the consensus meeting, a draft framework was produced and ratified by the group.

Phase 4: The draft framework was subsequently posted on the website of the ISARIC, which is openly accessible, and circulated to its 119 member organisations and partners for public consultation, which was open for one month and comments on the framework were captured on a short online form, developed using JISC Online Surveys. All comments received were reviewed and discussed by the small working group to determine their integration into the framework.

The framework contains a four-level structure:

- **Domains**: refer to the top-level descriptive category that summarises the characteristics it contains,
- **Critical characteristics**: are the essential characteristics that all endpoints need to possess to be viable for selection,
- **Ideal characteristics**: are other characteristics that endpoints should possess to generate meaningful evidence for trials (but acknowledging they may not be reasonably achieved in all scenarios for these diseases),
- **Additional considerations**: are other important considerations that shouldn't necessarily affect endpoint selection, but contain practical advice.

4.5.2 Results

In total, 38 individuals provided input to the development of the framework across the four project phases (**Figure 4.3**). Participants in the Phase 1 working group consisted of six clinical researchers (four senior investigators and two junior researchers) in emerging infectious diseases, two biostatisticians and one expert in community engagement. Participants in the modified-Delphi survey (Phase 2) were 22 clinical researchers with experience conducting studies for infectious diseases (of whom 16 were senior investigators in prominent research institutions), four senior biostatisticians, and three senior regulators from two multinational regulatory bodies. Attendees at the small working group (Phase 3) were nine clinical researchers (including seven senior investigators from prominent research institutions) in emerging infectious diseases and one senior regulator from a regional regulatory body. During the public consultation, one feedback submission was received but the proposal was considered not to be relevant to the framework as it did not align with the scope of the project. The 38 participants in the three project phases represented 20 countries, of which 26 (68%) participants were from HICs and 12 (32%) were from LMICs.

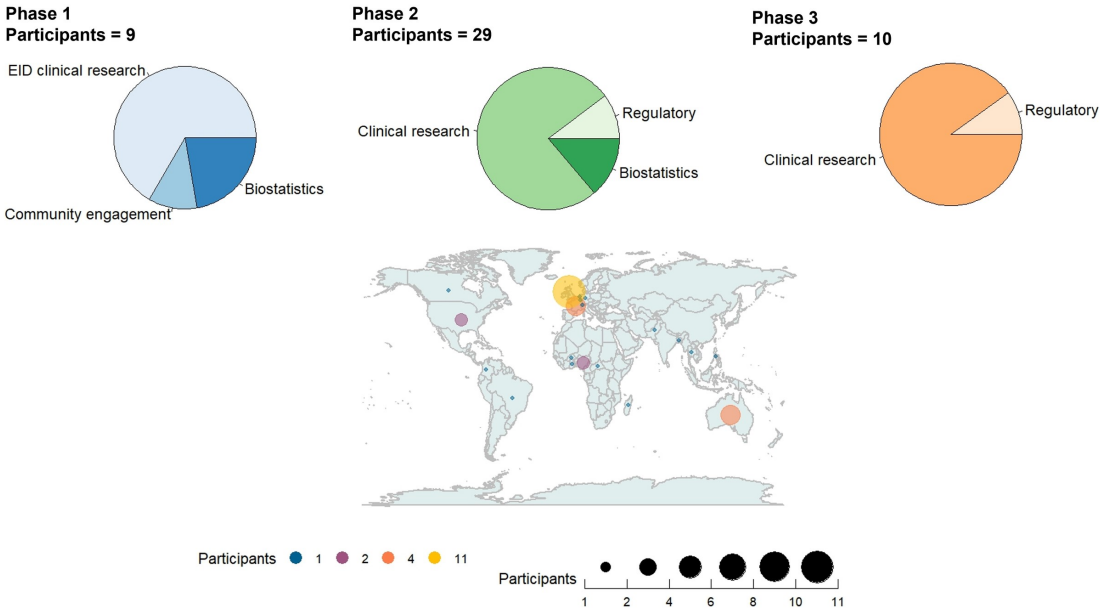


Figure 4.3: Summary of participant expertise and country across Phases 1, 2 and 3

Following the completion of Phase 1, which generated the first of two modified-Delphi questionnaires, the first questionnaire was circulated to 64 potential participants, of whom 23 (36%) returned a response. In this round, consensus was achieved to include all the proposed domains (**Figure 4.4**) and 10 of the proposed sub-domains in the framework. The Phase I working group then triaged the responses and free-text comments to develop the second questionnaire, which was sent to the same list of 64 potential participants, of whom 17 (27%) returned a response. Following this round, it was determined that the framework should contain seven of the 10 domains, alongside 17 sub-domains.

An example of the evolution of the framework is presented in Figure 4.4.

Figure 4.4: An example of domains and sub-domains included across modified-Delphi Round 1 and 2

All items that did not meet the target threshold for inclusion were reviewed and discussed by the working group, who categorised items in to "critical" and "ideal" characteristics of endpoints. These components of the framework were added to ensure that the framework was non-restrictive and could be flexibly implemented in a wide variety of trial contexts. The working group also identified the need for a new domain to represent the "public health meaningfulness" of the endpoint, which contains important considerations for trials of emerging infections where, for example, risk of infection transmission could be measured. All items under "public health meaningfulness" were categorised as "ideal" characteristics, as the working group agreed that clinically meaningful endpoints should be prioritised.

While no substantive changes were made in response to the public consultation, minor additional modifications were made to the framework for clarity by the steering committee following the closure of the consultation.

The final framework **Table 4.14** has been developed to support the selection of primary endpoints in Phase II/III trials for emerging infectious diseases that aim to identify a new

treatment where no other treatment exists or where the current standard of care is deemed sub-optimal or ineffective.

It encompasses seven key domains that balance scientific validity, clinical relevance, and operational feasibility. The domains are as follows:

- First, endpoints should demonstrate clinical meaningfulness, reflecting outcomes that are both frequent in the study population and serious in nature (e.g., mortality, life-threatening complications, or long-term disability). Where appropriate, validated biological markers may be considered, although non-validated surrogates should remain exploratory.

- Second, ideally the public health meaningfulness should be evaluated (where appropriate). Endpoints that reflect transmission potential, disproportionately affect vulnerable populations, or place significant demand on healthcare infrastructure can enhance the broader relevance of trial findings. This however may not be possible or relevant for all trials.

- Third, the validity of the measurement tool is critical for all trials. Endpoints should be based on instruments with demonstrated accuracy, construct and content validity, and minimal measurement error across relevant populations.

- Fourth, reliability is also essential. Outcome measurements should be consistent both within and between raters, and across patient subgroups, to allow for robust and reproducible findings.

- Fifth, the feasibility of measuring the endpoint should be considered in terms of standardisation, availability of tools, acceptability to participants, and capacity at trial sites. Cost and resource implications must align with the local context and funding scope.

- Sixth, safety must be prioritised. Measuring the outcome should not introduce undue risk to patients or healthcare workers, and any additional risks should be clearly mitigated by safety protocols.

- Finally, statistical and design considerations should guide endpoint selection. Assumptions used for sample size calculations must be grounded in reliable evidence where available. For novel conditions with limited prior data, endpoint assumptions should be revisited and revised as new information becomes available during the course of the trial.

Table 4.14: A framework to support the selection of primary endpoints in clinical trials for emerging infections.

Domain	**Critical characteristics of the primary endpoint**	**Ideal characteristics of the primary endpoint**	**Additional considerations**
[1] Clinical meaningfulness of the outcome measure	[a] The outcome is frequent in the potential patient pool allowing a sufficient number of events to be detected in the study population. [b] The outcome is serious and severe, defined as: outcomes that without a safe and efficacious intervention may result in one of the following: i. death; ii. a life-threatening condition; iii. new or prolonged hospitalisation; iv. disability or incapacity; v. in a congenital anomaly; vi. a life-shortening condition; vii. a long-term, life-altering condition.	-	When a biological marker has been validated as a surrogate for clinical efficacy, the biological meaningfulness of the outcome should be considered. Non-validated surrogate endpoints can be included as exploratory endpoints.

Table 4.14: A framework to support the selection of primary endpoints in clinical trials for emerging infections.

Domain	**Critical characteristics of the primary endpoint**	**Ideal characteristics of the primary endpoint**	**Additional considerations**
[2] Public health meaningfulness	-	[a] The endpoint is associated with infection/transmission risk (where uncertainty exists); [b] The outcome can be measured in vulnerable groups who are often excluded from clinical trials (e.g. pregnant women); [c] The outcome impacts the health and social care system, defined as outcomes that put an increased demand on healthcare infrastructure (e.g., outcomes that result in patients requiring prolonged hospitalisation and/or more complex or specialised care)	See also comment on biological meaningfulness above.

Table 4.14: A framework to support the selection of primary endpoints in clinical trials for emerging infections.

Domain	**Critical characteristics of the primary endpoint**	**Ideal characteristics of the primary endpoint**	**Additional considerations**
[3] Validity (accuracy) of the measurement tool	There should be data available from robust experimentation/testing that can be used as evidence of (as applicable): i. the ability of the measurement instrument to detect change; ii. content validity; iii. construct validity; iv. accuracy; v. risk of measurement error; vi. the validity of the measurement instrument in different sub-populations included in the trial	-	-

Table 4.14: A framework to support the selection of primary endpoints in clinical trials for emerging infections.

Domain	**Critical characteristics of the primary endpoint**	**Ideal characteristics of the primary endpoint**	**Additional considerations**
[4] Reliability (consistent measurement) of the measurement instrument	There should be data available from robust experimentation/testing that can be used evidence of (as applicable): i. subjectivity measuring and interpreting the outcome; ii. inter- and intra-rater consistency measuring and recording the outcome; iii. measurement consistency across different sub-populations	-	-

Table 4.14: A framework to support the selection of primary endpoints in clinical trials for emerging infections.

Domain	**Critical characteristics of the primary endpoint**	**Ideal characteristics of the primary endpoint**	**Additional considerations**
[5] Feasibility measuring the outcome	[a] The measurement of the outcome should be easily standardised between studies; [b] The required frequency and duration of the outcome measurements is acceptable to the study population and the study team at the research site have sufficient resources to conduct the measurements according to the protocol.	[c] The measurement instrument is readily available locally; [d] The cost of the measurement instrument and measuring the outcome is acceptable within the context of the disease, setting, available resources and scope of the trial funding.	The level of experience of trial staff using the measurement instrument and measuring the outcome should be taken into consideration to determine the feasibility of its use.
[6] Safety measuring the outcome	[a] The risks associated with measuring the outcome do not outweigh the benefit, and are mitigated by appropriate safety measures for both patients and healthcare staff.	[b] There is no more than minimal additional risk conferred to both patients and healthcare staff when measuring the outcome compared to routine care.	-

Table 4.14: A framework to support the selection of primary endpoints in clinical trials for emerging infections.

Domain	**Critical characteristics of the primary endpoint**	**Ideal characteristics of the primary endpoint**	**Additional considerations**
[7] Statistical and design considerations related to endpoint selection	[a] For known conditions: there must be reliable evidence supporting the assumptions for the sample size calculation; [b] The outcome should, for an adequate statistical power, require a target sample size that can be recruited within a feasible and relevant timeframe, defined as: i. approximately 6 to 18 months during an outbreak; ii. 24 months outside an outbreak (Although those using this framework should consider their own specific needs and context)	[c] The data used to inform the endpoint selection and related aspects of the trial design (e.g., effect size, expected frequency of the event in the control arm) should derive from either, in order of preference: i. robust data from multiple large-scale clinical trials, where available; ii. robust data from one large-scale clinical trials or observational study; iii. multiple small trials and observational studies	For unknown conditions: if no or limited data are available to inform trial design and endpoint selection, the trial assumptions should be updated regularly as new data become available (e.g., through interim meetings of the DSMB).

This framework is not suited for the scenarios described in Table 4.15.

Table 4.15: Additional framework guidance

Scenario 1	The framework should not be used when mortality/ survival has been identified as a more appropriate primary endpoint based on the following criteria: • Mortality/survival is considered the single most clinically-meaningful outcome • The intervention is believed to affect the patient's chance of survival • The frequency of mortality in the potential patient pool would allow the trial to be adequately powered and a significant treatment effect to be detected between arms ideally within a period of approximately 6 to 18 months during an outbreak or 24 months outside an outbreak (although the time frame required to gather evidence may vary between outbreaks, diseases and trials)
Scenario 2	If a Core Outcome Set (COS) has been developed for the disease and population of interest, the study team should aim to evaluate the outcomes identified in the COS as either primary or secondary endpoints, where feasible to do so. The framework should be used to complement the COS, assess the operationalisation of an endpoint in the trial protocol and determine any potential risks or limitations.
Scenario 3	The framework focuses on ClinROs. Where it is necessary to evaluate other types of outcomes (such as Patient-Reported Outcome measures (PROs) or a surrogate primary endpoint), it is recommended that trial teams discuss plans with their national or regional regulatory body before committing the endpoints to trial protocols.

4.5.3 Discussion

This ISARIC-adopted framework has two primary functions: firstly, to guide researchers through the essential and ideal characteristics that primary endpoints in Phase II/III trials for emerging infectious diseases should possess in order to generate clinically meaningful, robust and reliable data; secondly, to act as a tool to give an indication of the strength of a potential primary endpoint and to identify areas of potential bias, weakness or operational challenge that researchers may encounter when selecting an endpoint to evaluate their trial data. This tool will help to design and conduct more clinically-meaningful and robust

clinical trials, particularly in trials for diseases that have historically been difficult to study.

While the framework has been developed with a specific focus on primary endpoints for emerging infectious diseases, it could be broadly applied to other diseases and would be particularly helpful for diseases that are poorly characterised. As some emerging infectious diseases arise more frequently in settings with limited resources, the framework has been developed with a focus on clinical trials that take place in this context, where access to diagnostics, laboratories, and imaging may be delayed or impossible. However, researchers in any setting are encouraged to adapt the framework to their specific needs and context.

Discussions around the development of this framework were limited to ClinROs and therefore the framework currently omits considerations around Patient-Reported Outcome measures (PROs), Observer-Reported Outcome measures (ObsROs) and Performance Outcome measures (PerfOs). [201] Components of the framework, such as items [3], [4] and [5], could be appropriately applied to these types of outcome measures, but further investigation is needed to determine how they can be adequately addressed by the framework and how it could integrate with other quality assessment tools, such as the U.S. Food and Drug Administration's Clinical Outcome Assessment (COA) Qualification Program [202], which would evaluate only the PRO tool itself (items [3], [4] and [6] on the framework) and do not consider the broader context of the trial, its context or population.

It should also be acknowledged that the framework was largely developed with input from experts in HICs (68%). Notwithstanding the lower numbers of stakeholders from LMICs, it is important to highlight that the responses from HIC and LMIC participants did not substantially differ in subsequent rounds and phases of the framework development. While researchers, including those in LMICs, are encouraged to adapt the framework to their own contexts, the research community needs to find better ways of engaging with stakeholders in LMICs during the development phase to ensure equitable access and relevance of new research tools.

Overall, this framework aims to support the generation of conclusive clinical evidence in treatment trials by guiding the selection of clinically meaningful, reliable and actionable endpoints, particularly for diseases that have limited treatment options and have

historically faced challenges generating robust pivotal data that are required for regulatory approval of new products. We anticipate that the framework will be used primarily to support protocol development, but it could also be applied to the evaluation of the quality of published trial outputs.

Author contribution to "A standardised Phase III clinical trial framework to assess therapeutic interventions for Lassa fever"

Role description	Author contribution
Conceptualisation	X
Methodology	X
Software	-
Validation	X
Formal analysis	X
Investigation	X
Resources	-
Data curation	X
Writing (original draft)	X
Writing (review and editing)	X
Visualisation	-
Project administration	X
Funding acquisition	-

Publications

This article associated with this chapter has been drafted and is pending submission for publication.

4.6 Contribution of the chapter

This chapter examines the impact of different approaches to endpoint selection and modification for trials of plague and Lassa fever. In the case of plague, the IMASOY trial had already established and implemented an endpoint. However, concerns regarding the interpretation of one component (a 25% reduction in bubo size) prompted an evaluation of its measurability in assessing the efficacy of the trial drugs. In contrast, the approach to endpoint selection for Lassa fever followed a different pathway; a consortium was convened prior to the initiation of a trial to refine both the endpoint and overall study design. Although both endpoints have been successfully incorporated into clinical trials, challenges remain in ensuring their reliability for accurately assessing the impact of the interventions on disease outcomes. These limitations and biases are made clearer when applying the endpoint selection framework developed at the end of this chapter.

For plague, while one problematic component of the composite endpoint was identified and removed, the endpoint framework highlights potential limitations and biases associated with the remaining four elements, each of which may influence the interpretation of the trial results (**Table 4.17**). Death, on the one hand, is a clinically meaningful and objectively measurable outcome with minimal additional safety risks beyond those inherent to routine clinical practice. Fever is also measurable and presents limited safety concerns; however, its frequency is not well established in the academic literature making it difficult to determine the event rate necessary for appropriate sample size calculations, and in isolation, it is not a particularly severe or serious condition.

Similarly, challenges arise with the inclusion of the administration of additional therapeutics and the continuation of therapy beyond D11. While both reflect a clinical assessment that the patient requires further treatment, such judgments are inherently subjective and vary according to the risk tolerance of the individual clinician. In resource-limited settings such as Madagascar, these decisions are further influenced by external factors, including the availability of additional therapeutics and hospital capacity.

This assessment does not imply that the endpoint is unable of accurately capture the effect of the intervention on patient outcomes. However, it does highlight potential sources of bias that could influence the interpretation of the results. In particular, the

factors identified in the framework introduce biases that may limit the confidence with which the trial findings can be generalised to different populations, healthcare settings, and regions affected by plague.

Table 4.17: Application of the endpoint framework to the IMASOY primary endpoint

Domain	**Death**	**Presence of fever**	**Administration of additional plague therapeutics**	**Continuation of treatment beyond D11**
Clinical meaningfulness	Meets criteria	Unknown frequency and does not meet the seriousness or severity criteria	Unknown frequency and does not meet the seriousness or severity criteria	Unknown frequency but may result in prolonged hospitalisation
Public health meaningfulness	Does not meet public health criteria	Does not meet public health criteria	Does not meet public health criteria	May result in greater burden on the health system if patients require prolonged hospitalisation
Validity	Can be accurately measured	Can be accurately measured	Data on administration of additional therapeutics can be collected accurately from patient records	Data on continuation of therapy can be collected accurately from patient records

Table 4.17: Application of the endpoint framework to the IMASOY primary endpoint

Domain	**Death**	**Presence of fever**	**Administration of additional plague therapeutics**	**Continuation of treatment beyond D11**
Reliability	Can be consistently measured	Can be consistently measured	A subjective measure influenced by the clinician's individual risk tolerance	A subjective measure influenced by the clinician's individual risk tolerance
Feasibility	Can be feasibly measured	Can be feasibly measured	A subjective measure influenced by the clinician's individual risk tolerance and availability of additional therapeutics	A subjective measure influenced by the clinician's individual risk tolerance
Safety	Risks can be mitigated by IPC procedures	Risks can be mitigated by IPC procedures	No additional measurement risk	No additional measurement risk

Table 4.17: Application of the endpoint framework to the IMASOY primary endpoint

Domain	**Death**	**Presence of fever**	**Administration of additional plague therapeutics**	**Continuation of treatment beyond D11**
Statistical and design considerations	Variable CFR reported in academic literature (15%); studies in Madagascar report CFR of 5% to 20%	Limited available data on defervescence	Limited data on the administration of additional therapeutics	Limited available data on continuation of therapy beyond D11

For Lassa fever, the endpoint framework identifies relatively few problematic components that could significantly influence the interpretation of trial results (**Table 4.18**). Each component is clinically meaningful and objectively measurable, provided that at least two consecutive readings of the clinical syndromes are recorded. This approach helps mitigate the risk of a single transient measurement leading to misclassification of a patient's condition.

The assumptions regarding event frequencies for sample size calculations could largely be based on data from a robust, large observational cohort study. [90] However, this study was conducted at a single treatment centre in Nigeria, raising concerns about its generalisability to other Lassa fever treatment centres across the region. Given the considerable variability in data from other studies, it remains uncertain whether this dataset accurately reflects the broader epidemiological landscape of Lassa fever and whether the data alone are sufficient to base assumptions for the sample size calculation.

Table 4.18: Application of the endpoint framework to the primary endpoint developed for Lassa fever

Domain	**Death**	**New onset of AKI**	**New onset of ARF**	**New onset of shock**
Clinical meaningfulness	Meets criteria	Meets criteria for seriousness and severity, but inconsistent data available on frequency	Meets criteria for seriousness and severity, but inconsistent data available on frequency	Meets criteria for seriousness and severity, but inconsistent data available on frequency
Public health meaningfulness	No direct public health meaningfulness	No direct public health meaningfulness	No direct public health meaningfulness	No direct public health meaningfulness
Validity	Can be accurately measured	Can be accurately measured with at least two consecutive readings several hours apart	Can be accurately measured with at least two consecutive readings several hours apart	Can be accurately measured with at least two consecutive readings several hours apart
Reliability	Can be consistently measured	Can be consistently measured	Can be consistently measured	Can be consistently measured
Feasibility	Can be feasibly measured	Can be feasibly measured	Can be feasibly measured	Can be feasibly measured
Safety	Risks can be mitigated by IPC procedures	Risks can be mitigated by IPC procedures	Risks can be mitigated by IPC procedures	Risks can be mitigated by IPC procedures

Table 4.18: Application of the endpoint framework to the primary endpoint developed for Lassa fever

Domain	**Death**	**New onset of AKI**	**New onset of ARF**	**New onset of shock**
Statistical and design considerations	Data on CFR available from large observational cohort study, but conducted at a single site and variation observed between other studies	Data on CFR available from large observational cohort study, but conducted at a single site and variation observed between other studies	Data on CFR available from large observational cohort study, but conducted at a single site and variation observed between other studies	Data on CFR available from large observational cohort study, but conducted at a single site and variation observed between other studies

While the framework identifies several challenges for both plague and Lassa fever, its primary aim is not to discourage the use of the endpoint but rather to highlight potential sources of bias that may affect the interpretation of trial outcomes. By applying the framework prospectively, researchers can anticipate and address these biases, allowing for modifications to the endpoint that enhance its validity and reliability in assessing treatment effects. By applying the framework retrospectively researchers can better interpret the trial results and determine whether certain factors may have influenced the findings.

Overall, this chapter has examined the role of endpoint selection and modification in clinical trials for plague and Lassa fever, highlighting key challenges and biases that may influence the interpretation and generalisability of trial results. The application of the endpoint framework has provided valuable insights into the strengths and limitations of the endpoints used in both diseases, emphasising the importance of ensuring clinically meaningfulness, reliability, and validity, and that the assumptions underpinning trial design decisions are based on robust data.

The findings underscore the importance of applying an endpoint framework prospectively to refine endpoint selection and trial design before study initiation. The framework is also usefully applied retrospectively to critically assess potential biases in completed trials and inform the design of future trials.

Concluding remarks

This project has underscored the complex challenges involved in designing clinical trials for outbreak-prone infections. A primary limitation in trial design for these diseases is the scarcity of robust, standardised, and consistent data in the academic literature. This lack of data hinders the identification of clinically-meaningful patient outcomes for clinical trials and complicates sample size calculations due to unclear event frequencies. As a result, trials risk being underpowered, leading to inconclusive findings. Consequently, treatments may be prematurely integrated into clinical care based on weak evidence or erroneously dismissed, depriving patients of potentially life-saving therapeutics.

Despite these challenges, this project has also identified practical solutions for improving clinical trials in this field. Specifically it has made two important direct contributions to clinical trials for plague and Lassa fever:

- **Lassa fever**: The primary endpoint developed for Lassa fever trials in section 4.3 and section 4.4 has been successfully implemented in the INTEGRATE trial, alongside key trial design components such as the adaptive platform framework [59].
- **Bubonic plague**: By investigating the meaningfulness and potential measurement error associated with measuring buboes in the IMASOY trial, a crucial amendment to the primary endpoint was made to remove the measurement of buboes from the composite endpoint, thus eliminating a major source of bias. Without this adjustment, the trial would have produced negative results, limiting its potential to influence treatment practices.

Beyond these specific contributions, the endpoint framework developed in this project offers a structured approach to endpoint selection for future trials of outbreak-prone in-

fections and other diseases where endpoint selection is particularly challenging. While no framework can guarantee the identification of a "perfect" endpoint, it can help researchers anticipate limitations early in the trial design process rather than addressing them retrospectively. This, in turn, facilitates more transparent reporting of biases and methodological constraints. The framework may be especially valuable for researchers developing Core Outcome Sets (COS), as illustrated in section 4.3. Without a guiding structure, consensus-driven approaches can result in the selection of endpoints that are impractical for trial implementation – as was the case for Lassa fever, where mortality was initially chosen as the preferred primary endpoint despite infeasibility due to large sample size requirements. Integrating a framework into these discussions can help ensure that endpoint selection is both evidence-based and practical, ultimately strengthening the design and impact of clinical trials for outbreak-prone infections.

This project has also had indirect yet significant impacts on other research in the field of outbreak-prone infections. Notably, findings from the prospective cohort study using ultrasound to characterise the morphology of buboes have informed the design of a trial investigating the use of ultrasound to detect complications of dengue. This dengue study adopts a similar approach, training clinicians with limited ultrasound experience. By recognising the potential measurement errors introduced by the subjective assessment of different physiological structures in the bubonic plague study, the dengue study has been able to refine its training protocols and standard operating procedures (SOPs) to mitigate assessment bias.

There are also several streams of work that need to be undertaken in the future to further develop some of the concepts explored in this project. First, for mpox, further investigation is needed to determine optimal endpoints that accurately and meaningfully capture treatment outcomes. This thesis has identified multiple challenges associated with clinical and laboratory outcome measures. As an alternative approach, examining the patient experience of mpox and identifying patient-reported outcomes that could be applied to a trial may result in the identification of workable endpoints – provided regulators are also consulted on the likelihood of these endpoints supporting licensing decisions. Notably, the scientific literature lacks qualitative research on this topic for Clade I mpox and such

research would only be considered useful if conducted across the endemic region to ensure cultural and contextual relevance.

Furthermore, patient perspectives should be explicitly incorporated into the endpoint framework to ensure that selected endpoints are not only clinically relevant but also meaningful to patients. To support this aim, conducting focus groups with individuals who have experienced the types of "difficult-to-study" diseases the framework is aimed at could help elucidate what constitutes a meaningful outcome to patients.

Finally, collaboration with initiatives such as the COMET Initiative would be valuable to enhance the structure of consensus-building activities. Integrating the endpoint framework into the COS development process would help ensure that COS are both practically implementable and contextually appropriate for the clinical trials in which they are applied.

In summary, this project addresses a critical gap in the design of clinical trials for outbreak-prone infections, where data limitations and methodological challenges often undermine the reliability and interpretability of trial outcomes. Through direct contributions to trials for Lassa fever and bubonic plague, as well as the development of a structured framework for endpoint selection, this work has generated both immediate and long-term impact to researchers working in research for outbreak-prone infections. By promoting more systematic, evidence-based approaches to endpoint design and highlighting the importance of anticipating bias and feasibility issues from the outset, this project also lays the groundwork for more robust, interpretable, and ultimately impactful clinical trials in the future.

Appendices

Appendix 1: Summary of published papers associated with this project

Bourner J, Vaillant M, Abdel Salam A, Jaspard M, Fritzell C, Jacob ST, et al. Adaptive Design for Phase II/III Platform Trial of Lassa Fever Therapeutics. Emerg Infect Dis. 2025;31(2):9-16. https://doi.org/10.3201/eid3102.240251

Bourner J, Garcia-Gallo E, Mbrenga F, Boum Y 2nd, Nakouné E, Paterson A, Jones B, Olliaro P, Rojek A. Challenges in clinical diagnosis of Clade I Mpox: Highlighting the need for enhanced diagnostic approaches. PLoS Negl Trop Dis. 2024 Jun 24;18(6):e0012087. doi: 10.1371/journal.pntd.0012087. PMID: 38913721; PMCID: PMC11226010.

Bourner J, Randriamparany R, Rasoanaivo TF, Denis E, Randremanana RV, Vaillant M, Salam AP, Gonçalves BP, Olliaro P. Bubonic plague: can the size of buboes be accurately and consistently measured with a digital calliper? Trials. 2023 Dec 19;24(1):815. doi: 10.1186/s13063-023-07835-7. PMID: 38115024; PMCID: PMC10729355.

Bourner J, Andriamarohasina L, Salam A, Kayem ND, Randremanana R, Olliaro P. A systematic review of the clinical profile of patients with bubonic plague and the outcome measures used in research settings. PLoS Negl Trop Dis. 2023 Nov 9;17(11):e0011509. doi: 10.1371/journal.pntd.0011509. PMID: 37943880; PMCID: PMC10662759.

Jones B, Paterson A, AlKhoury N, **Bourner J**, Dunning J, Olliaro P, Rojek A. Variability in clinical assessment of clade IIb mpox lesions. Int J Infect Dis. 2023 Dec;137:60-62. doi: 10.1016/j.ijid.2023.10.004. Epub 2023 Oct 15. PMID: 37848125; PMCID: PMC10914632.

Bourner J, Salam AP, Jaspard M, Olayinka A, Fritzell C, Goncalves B, Vaillant M, Edwards T, Erameh C, Ajayi N, Ramharter M, Olliaro P; WALC Work Package 2 Working Group. The West Africa Lassa fever Consortium pre-positioned protocol for a Phase II/III adaptive, randomised, controlled, platform trial to evaluate multiple Lassa fever therapeutics. Wellcome Open Res. 2023 Jun 2;8:122. doi: 10.12688/wellcomeopenres.19041.2. PMID: 39211525; PMCID: PMC11358687.

Mbrenga F, Nakouné E, Malaka C, **Bourner J**, Dunning J, Vernet G, Horby P, Olliaro P. Tecovirimat for Monkeypox in Central African Republic under Expanded Access. N Engl J Med. 2022 Dec 15;387(24):2294-2295. doi: 10.1056/NEJMc2210015. Epub 2022 Nov 30. PMID: 36449745; PMCID: PMC10117058.

Bourner J*, Olayinka AT*, Akpede GO, Okoeguale J, Abejegah C, Ajayi NA, Akude C, Ayodeji O, Bausch DG, de Clerck H, Dan-Nwafor C, Dunning J, Erameh C, Eze JN, Formenty P, Gillesen A, Jalloh S, Jaspard M, Jegede T, Maikere J, Malvy D, Ogbaini-Emovon E, Ojo OE, Okogbenin S, O'Neill K, Orji ML, Owhin SO, Ramharter M, Samuels RJ, Shehu N, Merson L, Salam AP, Kayem ND, Horby P, Ihekweazu C, Olliaro P. A standardised Phase III clinical trial framework to assess therapeutic interventions for Lassa fever. PLoS Negl Trop Dis. 2022 Jan 6;16(1):e0010089. doi: 10.1371/journal.pntd.0010089. PMID: 34990453; PMCID: PMC8769305.

Bourner J*, Merson L*, Jalloh S, Erber A, Salam AP, Flahault A, Olliaro PL. Clinical characterization of Lassa fever: A systematic review of clinical reports and research to inform clinical trial design. PLoS Negl Trop Dis. 2021 Sep 21;15(9):e0009788. doi: 10.1371/journal.pntd.0009788. PMID: 34547033; PMCID: PMC8486098.

*These two authors contributed equally to this manuscript and are listed as joint first authors

Appendix 2: Search strategy for "Clinical characterisation of Lassa Fever: A systematic review to inform clinical trial design"

African Journal Online (https://www.ajol.info/)

Search date: 02/12/2018

Search strategy: lassa*

Cochrane Central Register of Controlled Trials (CENTRAL)

(https://www.cochranelibrary.com/central/about-central)

Date: 07/01/2020 (Issue 12 of 12, December 2019)

Search strategy:

#1 MeSH descriptor: [Lassa Fever] explode all trees 2

#2 MeSH descriptor: [Lassa virus] explode all trees 0

#3 lassa* 153

#4 #1 or #2 or #3 153

Embase (https://www.embase.com/login)

Data: 1974 to present

Search strategy: #1 Lassa fever/ (904)

#2 Lassa virus/ (1000)

#3 lassa*.ti,ab. (1554)

#4 #1 or #2 or #3 (1983)

#5 4 (1983)

#6 limit 5 to yr="1883 - 2019" (1978)

Global Health (Ovid)

(http://www.ovid.com/site/catalog/databases/30.jsp)

Date: <1973 to 2019 Week 51>

Search strategy:

#1 exp Lassa virus/or exp Lassa fever/(789)

#2 lassa*.ti,ab. (812)

#3 1 or 2 (851)

#4 3 (851)

#5 limit 4 to yr="1931 - 2019" (851)

Global Index Medicus

(http://www.globalhealthlibrary.net/php/level.php?lang=en&component=17&item=107)

Database: Global Index Medicus, regional indices only: African Index Medicus (AIM) and Index Medicus of the Eastern Mediterranean Region (IMEMR)

Search strategy:

Lassa*

PubMed/MEDLINE

(https://www.nlm.nih.gov/bsd/pmresources.html):

Database: Medline (Ovid MEDLINE® Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily and Ovid MEDLINE®) 1946 to present

Search strategy:

#1 Lassa Fever/ (617)

#2 Lassa virus/ (625)

#3 lassa*.ti,ab. (1376)

#4 1 or 2 or 3 (1512)

#5 4 (1512)

#6 limit 5 to yr="1860 - 2019" (1508)

Clinicaltrials.gov

Date: 07/01/2020

Condition or disease: Lassa – 10

ISRCTN Registry

(https://www.isrctn.com/editAdvancedSearch)

Date: 07/01/2020

Condition: lassa – 0

Pan African Clinical Trials Registry

(http://www.pactr.org/)

Date: 07/01/2020

Search Terms: Lassa – 0

WHO International Clinical Trials Registry

(http://www.who.int/ictrp/en/)

Date: 21/9/2018

Advanced search: ALL trials

Condition: lassa – 69

Appendix 3: Summary of studies included in "Clinical characterisation of bubonic plague: A systematic review to inform clinical trial design"

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Two Cases Of Bubonic Plague Occurring On Board Ship *Barnett H. N., 1902* [203]	Case report(s)	United Kingdom	2	2: 0	NK
A case of bubonic plague on a vessel arriving in the Mersey *NK, 1905* [204]	Case report(s)	United Kingdom	1	1: 0	NK
Subacute plague in man due to ground squirrel infection *McCoy G. W., 1909* [205]	Case report(s)	United States	1	1: 0	13
Three Cases of Bubonic Plague Arising In England *Rendle-Short A., 1916* [206]	Case report(s)	United Kingdom	3	3: 0	16 (10 to 23)
Streptomycin in Bubonic Plague *Haddad C. H., 1948* [207]	Interventional, non-randomised	Israel	3	3: 0	20 (8 to 35)

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Plague – New Mexico *United States Centres for Disease Control and Prevention, 1965* [208]	Case report(s)	United States	1	1: 0	14
Plague: Shasta County, California *United States Centres for Disease Control and Prevention, 1965* [209]	Case report(s)	United States	1	1: 0	5
Suspected Case of Imported Bubonic Plague – Texas *United States Centres for Disease Control and Prevention, 1966* [210]	Case report(s)	United States	1	1: 0	21
Plague – Arizona *United States Centres for Disease Control and Prevention, 1967* [211]	Case report(s)	United States	1	1: 0	4
Plague in San Diego *Connor J. D., 1968* [212]	Case report(s)	United States	1	1: 0	3

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Presumptive Bubonic Plague — Denver, Colorado *United States Centres for Disease Control and Prevention, 1968* [213]	Case report(s)	United States	1	0: 1	6
Bubonic Plague Death – Lemhi County, Idaho *United States Centres for Disease Control and Prevention, 1968* [214]	Case report(s)	United States	1	1: 0	32
Plague Case – Navajo Reservation – Kayenta, Arizona *United States Centres for Disease Control and Prevention, 1968* [215]	Case report(s)	United States	1	0: 1	8
Plague – New Mexico *United States Centres for Disease Control and Prevention, 1969* [216]	Case report(s)	United States	1	1: 0	3

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Bubonic plague in the Southwestern United States *Reed W. P., 1970* [217]	Case report(s)	United States	25	15: 10	14 (2 to 74)
Bubonic Plague – California *United States Centres for Disease Control and Prevention, 1970* [218]	Case report(s)	United States	1	1: 0	8
Bubonic Plague – California *United States Centres for Disease Control and Prevention, 1970*	Case report(s)	United States	1	1: 0	45
Human Bubonic Plague – Cochiti, New Mexico *United States Centres for Disease Control and Prevention, 1970* [219]	Case report(s)	United States	1	1: 0	39
Bubonic Plague – Santa Fe, New Mexico *United States Centres for Disease Control and Prevention, 1970* [220]	Case report(s)	United States	1	0: 1	20

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Plague – New Mexico *United States Centres for Disease Control and Prevention, 1970* [221]	Case report(s)	United States	1	1: 0	13
Plague – New Mexico *United States Centres for Disease Control and Prevention, 1970* [222]	Case report(s)	United States	2	1: 1	24.5 (15 to 34)
Plague – California *United States Centres for Disease Control and Prevention, 1970* [223]	Case report(s)	United States	1	0: 1	10
Plague - Rio en Medio, New Mexico *United States Centres for Disease Control and Prevention, 1970* [224]	Case report(s)	United States	1	0: 1	9
Plague – New Mexico *United States Centres for Disease Control and Prevention, 1970* [225]	Case report(s)	United States	1	1: 1	11.5 (7 to 16)

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Human Bubonic Plague – Oregon *United States Centres for Disease Control and Prevention, 1971* [226]	Case report(s)	United States	1	1: 0	10
Human Bubonic Plague – New Mexico *United States Centres for Disease Control and Prevention, 1971* [227]	Case report(s)	United States	1	0: 1	26
Clinical Features of Plague in the United States: the 1969-1970 Epidemic *Palmer D. L., 1971* [228]	Cohort	United States	19	12: 7	25.5 (2 to 49)
Human Bubonic Plague - Coconino County Colorado *United States Centres for Disease Control and Prevention, 1972* [229]	Case report(s)	United States	1	1: 0	19

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Human Bubonic Plague – Arizona *United States Centres for Disease Control and Prevention, 1973* [230]	Case report(s)	United States	1	0: 1	9
Co-trimoxazole in Bubonic Plague *Nguyen-Van-Ai, 1973* [231]	Interventional, non-randomised	Vietnam	12	4: 8	38.5 (6 to 63)
Human Bubonic Plague – New Mexico *United States Centres for Disease Control and Prevention, 1974* [232]	Case report(s)	United States	1	0: 1	12
Human Plague – New Mexico *United States Centres for Disease Control and Prevention, 1974* [233]	Case report(s)	United States	1	1: 0	19
Human Plague – New Mexico, Utah *United States Centres for Disease Control and Prevention, 1974* [234]	Case report(s)	United States	1	1: 0	5

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Yersinia pestis Infection in Vietnam. I. Clinical and Hematologic Aspects *Butler T., 1974* [235]	Cohort	Vietnam	22	17: 5	16 (4 to 67)
Plague and the gallium scan *Stahly T. L., 1975* [122]	Case report(s)	United States	1	1: 0	11
Fatal Bubonic Plague – California *United States Centres for Disease Control and Prevention, 1975* [236]	Case report(s)	United States	1	0: 1	1
Bubonic Plague – Arizona *United States Centres for Disease Control and Prevention, 1975* [237]	Case report(s)	United States	2	0: 2	17 (3 to 31)
Plague in Humans – New Mexico *United States Centres for Disease Control and Prevention, 1975* [238]	Case report(s)	United States	6	0: 6	10.5 (3 to 28)

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Human Plague Case – Bernalillo County, New Mexico *United States Centres for Disease Control and Prevention, 1975* [239]	Case report(s)	United States	1	1: 0	11
Bubonic Plague from Exposure to a Rabbit: A Documented Case, and a Review of Rabbit-Associated Plague Cases in The United States *Von Reyn C. F., 1976* [240]	Case report(s)	United States	1	0: 1	62
Human Plague – Arizona, California, New Mexico *United States Centres for Disease Control and Prevention, 1976* [241]	Case report(s)	United States	3	2: 1	45 (15 to 63)

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Yersinia pestis Infection in Vietnam. II; Quantitative Blood Cultures and Detection of Endotoxin in the Cerebrospinal Fluid of Patients with Meningitis *Butler T, 1976* [242]	Cohort	Vietnam	42	20: 22	15 (3 to 70)
Plague and pregnancy. A case report *Mann J, 1977* [243]	Case report(s)	United States	1	0: 1	28
Epidemiological and clinical features of an outbreak of bubonic plague in New Mexico *Von Reyn C. F., 1977* [244]	Cohort	United States	7	3: 4	19 (5 to 62)
Plague – Arizona, Colorado, New Mexico *United States Centres for Disease Control and Prevention, 1977* [245]	Case report(s)	United States	4	3: 1	30.5 (3 to 43)

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Plague – United States *United States Centres for Disease Control and Prevention, 1977* [246]	Case report(s)	United States	1	1: 0	6
Plague – Arizona, California, New Mexico *United States Centres for Disease Control and Prevention, 1978* [247]	Case report(s)	United States	1	1: 0	14
Plague in the United States: the "black death" is still alive *Hoffman S. L., 1980* [248]	Case report(s)	United States	1	1: 0	55
Plague – United States *United States Centres for Disease Control and Prevention, 1980* [249]	Case report(s)	United States	9	6: 3	8 (4 to 55)
Human Plague – Texas, New Mexico *United States Centres for Disease Control and Prevention, 1981* [250]	Case report(s)	United States	1	1: 0	25

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Human plague associated with domestic cats–California, Colorado *United States Centres for Disease Control and Prevention, 1981* [251]	Case report(s)	United States	1	1: 0	49
Peripatetic Plague *Mann J., 1982* [252]	Case report(s)	United States	1	0: 1	16
Febrile lymphadenitis in the American West *Mann J., 1982* [253]	Case report(s)	United States	1	1: 0	5
Plague - South Carolina *United States Centres for Disease Control and Prevention, 1983* [254]	Case report(s)	United States	1	0: 1	13
Plague in the United States 1982 *Barnes A. M., 1983* [255]	Cohort	United States	19	11: 8	20 (4 to 78)
Plague Pneumonia – California *United States Centres for Disease Control and Prevention, 1984* [256]	Case report(s)	United States	1	1: 0	35

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Winter Plague – Colorado, Washington, Texas, 1983-1984 *United States Centres for Disease Control and Prevention, 1984* [257]	Case report(s)	United States	3	3: 0	38.5 (29 to 48)
Human Bubonic Plague Transmitted by a Domestic Cat Scratch *Weniger B. G., 1984* [123]	Case report(s)	United States	1	0: 1	10
Nineteen cases of plague in Arizona. A spectrum including ecthyma gangrenosum due to plague and plague in pregnancy *Welty T. K., 1985* [258]	Cohort	United States	19	6: 13	25 (2 to 78)
Multiple lung cavities in a 12-year-old girl with bubonic plague, sepsis, and secondary pneumonia *Florman A. L., 1986* [259]	Case report(s)	United States	1	0: 1	12

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Plague masquerading as gastrointestinal illness *Hull H. F., 1986* [260]	Cohort	United States	47	NK: NK	61 (13 to 71)
Plague in a pregnant patient *Wong T. W., 1986* [261]	Case report(s)	United States	1	0: 1	25
Plague meningitis–a retrospective analysis of cases reported in the United States, 1970-1979 *Becker T. M., 1987* [262]	Case report(s)	United States	2	1: 1	10.5 (10 to 11)
Imaging in plague *Moreno A. J., 1987* [121]	Case report(s)	United States	1	1: 0	8
Human Plague – United States, 1988 *United States Centres for Disease Control and Prevention, 1988* [263]	Case report(s)	United States	3	3: 0	41 (19 to 82)

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Imported bubonic plague – District of Columbia *United States Centres for Disease Control and Prevention, 1990* [264]	Case report(s)	United States	1	0: 1	47
Bubonic plague in a child presenting with fever and altered mental status *Migden D., 1990* [265]	Case report(s)	United States	1	1: 0	8
Plague in New Mexico *Owens C., 1990* [266]	Case report(s)	United States	1	1: 0	47
Plague - A clinical review of 27 cases *Crook L. D., 1992* [267]	Cohort	United States	27	15: 12	41 (2 to 80)
An Outbreak of Plague in Northwestern Province, Zambia *McClean K. L., 1995* [268]	Case report(s)	Zambia	1	1: 0	23

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Fatal human plague–Arizona and Colorado, 1996 *United States Centres for Disease Control and Prevention, 1997* [269]	Case report(s)	United States	2	1: 1	17 (16 to 18)
Current epidemiology of human plague in Madagascar *Chanteau S., 2000* [270]	Cohort	Madagascar	917	515: 402	15[b]
Cases of cat-associated human plague in the Western US, 1977-1998 *Gage K. L., 2000* [271]	Case report(s)	United States	17	11: 6	28 (6 to 58)
Imported plague–New York City, 2002 *United States Centres for Disease Control and Prevention, 2003* [272]	Case report(s)	United States	2	1: 1	50 (47 to 53)
Painful lymphadenopathy and fulminant sepsis in a previously healthy 16-year-old girl *Chmura K., 2003* [273]	Case report(s)	United States	2	0: 1	16

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Gentamicin and Tetracyclines for the Treatment of Human Plague: Review of 75 cases in New Mexico, 1985-1999 *Boulanger L. L., 2004* [274]	Cohort	United States	50	31 :19	NK
Human plague–four states, 2006 *United States Centres for Disease Control and Prevention, 2006* [275]	Case report(s)	United States	4	2: 2	34 (28 to 43)
Treatment of Plague with Gentamicin or Doxycycline in a Randomized Clinical Trial in Tanzania *Mwengee W., 2006* [58]	Interventional, randomised controlled trial	Tanzania	65	41: 24	13 (0 to 65)
Notes from the field: two cases of human plague–Oregon, 2010 *United States Centres for Disease Control and Prevention, 2011* [276]	Case report(s)	United States	2	NK: NK	29.5 (17 to 42)

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Misidentification of Yersinia pestis by Automated Systems, Resulting in Delayed Diagnoses of Human Plague Infections—Oregon and New Mexico, 2010–2011 *Tourdjman M., 2012* [277]	Case report(s)	United States	2	1: 1	37.5 (17 to 58)
Plague Outbreak in Libya, 2009, Unrelated to Plague in Algeria *Cabanel N., 2013* [278]	Cohort	Libya	2	1: 1	19 (14 to 24)
Outbreak of Plague in a High Malaria Endemic Region — Nyimba District, Zambia, March–May 2015 *Sinyange N., 2016* [279]	Cohort	Zambia	7	NK: NK	8 (3 to 18)
Successful Treatment of Human Plague with Oral Ciprofloxacin *Apangu T., 2017* [77]	Interventional, non-randomised	Uganda	4	1: 3	31 (10 to 52)

Study title ***First author, Year***	**Study type**	**Country**	**N cases**	**N male: N female**	**Age (years), Median[a] (range)**
Case report *Lazet K., 2018* [120]	Case report(s)	United States	1	0: 1	33
Human case of bubonic plague resulting from the bite of a wild Gunnison's prairie dog during translocation from a plague endemic area *Melman S. D., 2018* [280]	Case report(s)	United States	1	1: 0	66
Two fatal cases of plague after consumption of raw marmot organs *Kehrmann J., 2020* [281]	Case report(s)	Mongolia	2	1: 1	37.5 (37 to 38)
Delays in Identification and Treatment of a Case of Septicemic Plague — Navajo County, Arizona, 2020 *Dale A. P., 2021* [282]	Case report(s)	United States	1	1: 0	67

Appendix 4: Reported signs and symptoms in confirmed cases of bubonic plague at baseline and post-baseline (n/N (%) patients)

Sign or symptom	Baseline	Post-baseline
Abdominal distension	-	4/107 (4%)
Acute renal failure	3/5 (60%)	1/2 (50%)
Altered mental status	33/84 (39%)	-
Anaemia	-	1/1 (100%)
Anorexia	8/52 (15%)	1/1 (100%)
Anterior hemiblock	1/22 (5%)	-
Anuria	1/1 (100%)	-
Anxiety	2/28 (7%)	1/1 (100%)
Arthralgia	2/10 (20%)	-
Atelectasis	-	1/1 (100%)
Bilateral pulmonary infiltrates	1/4 (25%)	1/4 (25%)
Bloodshot eyes	1/3 (33%)	-
Bloody sputum	1/3 (33%)	5/7 (71%)
Blurred vision	2/2 (100%)	-
Bradycardia	1/22 (5%)	-
Breathing difficulty	1/2 (50%)	3/17 (18%)
Bubo suppuration	-	2/65 (3%)
Bundle branch block	1/22 (5%)	-
Cardiac arrest	-	1/2 (50%)

Sign or symptom	Baseline	Post-baseline
Cellulitis	-	3/18 (17%)
Cerebral edema	-	1/1 (100%)
Chest heaviness	-	1/1 (100%)
Chest pain	1/17 (6%)	3/27 (11%)
Cholelthiasis	-	1/47 (2%)
Coma	2/25 (8%)	-
Confusion	4/46 (9%)	-
Congested	1/3 (33%)	-
Congestive heart failure	-	1/25 (4%)
Conjunctivitis	1/19 (5%)	1/25 (4%)
Conjunctival suffusion	1/2 (50%)	1/2 (50%)
Contractions	1/1 (100%)	-
Costovertebral angle tenderness	1/47 (2%)	-
Cutaneous ecchymoses	-	1/4 (25%)
Cutaneous edema	1/15 (7%)	-
Cyanosis	1/2 (50%)	1/1 (100%)
Decreased breath sounds	1/1 (100%)	-
Decreased platelet count	-	1/15 (7%)
Decreased white cell count	-	1/25 (4%)
Decreased hematocrit	1/1 (100%)	-
Dehydration	4/4 (100%)	-
Delirium	14/220 (6%)	1/3 (33%)

Sign or symptom	**Baseline**	**Post-baseline**
Diaphoresis	1/2 (50%)	-
Diarrhoea	22/147 (15%)	3/67 (4%)
Difficulty walking	1/9 (11%)	-
Diffuse aching	1/25 (4%)	-
Disorientation	2/2 (100%)	-
Disseminated intravascular coagulation	1/1 (100%)	7/41 (17%)
Diverticulosis	-	1/2 (50%)
Dizziness	6/93 (6%)	1/65 (2%)
Dry, coated tongue	2/3 (67%)	-
Dyspnea	2/10 (20%)	2/6 (33%)
Ecthyma gangrenosum	1/15 (6%)	-
Edema	-	3/29 (10%)
Elevated blood pressure	1/1 (100%)	-
Elevated liver function tests	1/1 (100%)	-
Elevated transaminases	1/1 (100%)	-
Elevated white cell count	2/2 (100%)	1/25 (4%)
Emaciation	-	1/1 (100%)
Endophthalmitis	-	1/1 (100%)
Endotoxemia	-	3/42 (7%)
Enlarged liver	1/2 (50%)	-
Erythema	7/16 (44%)	4/4 (100%)
Exophthalmos	-	1/1 (100%)

Sign or symptom	Baseline	Post-baseline
Fetal distress	-	1/1 (100%)
Fetal tachycardia	1/1 (100%)	1/1 (100%)
Fibrin thrombosis	-	1/1 (100%)
Fluctuant nodes	-	1/1 (100%)
Fluid filled lesion on thumb	1/3 (33%)	-
Flushing	1/1 (100%)	1/1 (100%)
Fremitus	1/1 (100%)	-
Gangrene	1/1 (100%)	-
Gangrenous lymph node	1/1 (100%)	-
Gastrointestinal bleeding	-	1/17 (6%)
Granular casts	1/1 (100%)	-
Hallucinations	2/27 (7%)	2/16 (13%)
Haemhorrhage - bilateral adrenal	-	1/1 (100%)
Haemorrhage - ear	-	1/1 (100%)
Haemorrhage - eye	-	1/1 (100%)
Haemorrhage - lymph node	-	1/25 (4%)
Hematuria	1/1 (100%)	-
Hepatosplenomegaly	1/1 (100%)	1/1 (100%)
Herpes labialis	-	1/65 (2%)
High proportion of immature neutrophils	1/17 (6%)	-
Hyperventilation	-	1/15 (67%)

Sign or symptom	Baseline	Post-baseline
Hypoactive blowel sounds	1/47 (2%)	-
Hypotension	15/87 (17%)	11/85 (13%)
Hypoxemia with bilateral pulmonary edema	-	1/1 (100%)
Icterus	-	1/1 (100%)
Impaired consciousness	2/2 (100%)	1/3 (33%)
Incontinence	-	1/3 (33%)
Increased partial thromboplastin times	-	2/19 (11%)
Increased prothrombin time	-	1/4 (25%)
Infectious syndrome	1/2 (50%)	-
Inflammation of breast	1/25 (4%)	-
Inflammation of shoulder	1/25 (4%)	-
Injected conjunctivae	1/3 (33%)	-
Inflamed eardrum	1/1 (100%)	-
Inflamed throat	1/1 (100%)	-
Irritability	1/1 (100%)	-
Ischaemia	-	1/2 (50%)
Jaundice	1/3 (33%)	-
Lactic acid acidosis	-	1/17 (6%)
Lesion haemorrhage on toe	1/1 (100%)	-
Lethargy	3/67 (4%)	2/2 (100%)
Leukocytosis	1/17 (6%)	-

Sign or symptom	Baseline	Post-baseline
Loss of appetite	1/1 (100%)	-
Loss of consciousness	1/1 (100%)	-
Lymph node necrosis	1/4 (25%)	1/25 (4%)
Lymphadenitis	-	2/2 (100%)
Lymphadenopathy - generalised	-	1/1 (100%)
Malaise	22/68 (32%)	1/1 (100%)
Membrane rupture	1/1 (100%)	-
Meningeal irritation	-	1/1 (100%)
Meningitis	2/8 (25%)	6/62 (10%)
Meningism	2/28 (7%)	-
Metastatic anterior chamber endophthalmitis	1/1 (100%)	-
Microhaematuria	1/1 (100%)	-
Moribund appearance	-	1/1 (100%)
Muscle pain	2/26 (8%)	-
Myalgia	54/204 (26%)	-
Myocarditis	-	1/25 (4%)
Necrosis of the spleen	-	1/1 (100%)
Non-purposeful movement of all extremities	1/1 (100%)	-
Nuchal rigidity	4/20 (20%)	1/1 (100%)
Numbness and tingling sensation in arms	1/1 (100%)	-

Sign or symptom	Baseline	Post-baseline
Obtunded	2/77 (3%)	1/15 (7%)
Osteomyelitis	-	1/3 (33%)
Pain non-specific	2/2 (100%)	-
Pain in arm	3/17 (18%)	-
Pain in back	2/18 (11%)	-
Pain in elbow	1/1 (100%)	-
Pain in extremities	1/1 (100%)	-
Pain in flank	1/47 (2%)	-
Pain in groin	1/2 (50%)	1/1 (100%)
Pain at injection site	-	1/1 (100%)
Pain in knee/leg	1/1 (100%)	-
Pain in neck	5/6 (83%)	-
Perianal abrasions	1/1 (100%)	-
Pericarditis	-	1/25 (4%)
Perivenous haemhorrage	-	1/1 (100%)
Petechiae	1/15 (7%)	1/2 (50%)
Pharyngeal erythema	3/8 (38%)	-
Photophobia	1/25 (4%)	-
Pleocytosis	-	1/1 (100%)
Pleural effusion	-	2/5 (40%)
Pneumonia	2/22 (10%)	1/42 (2%)
Pneumonitis	2/2 (100%)	1/3 (33%)
Prostration	3/21 (14%)	1/1 (100%)

Sign or symptom	Baseline	Post-baseline
Proteinuria	2/2 (100%)	-
Pulmonary edema	1/1 (100%)	1/1 (100%)
Pustular eruption	-	1/1 (100%)
Pyuria	2/2 (100%)	-
Rapid and shallow breathing	2/4 (50%)	1/1 (100%)
Rapid pulse	-	1/1 (100%)
Rales	1/1 (100%)	-
Rash	2/4 (50%)	1/1 (100%)
Rectal bleeding	-	1/1 (100%)
Red blood cell casts	1/1 (100%)	-
Red throat (with exudate on tonsils)	2/2 (100%)	-
Reduced urine output	-	1/1 (100%)
Refractory shock	-	1/2 (50%)
Respiratory arrest/ failure	1/1 (100%)	3/19 (16%)
Respiratory distress	2/4 (50%)	5/46 (11%)
Restlessness	1/3 (33%)	-
Rigors	5/6 (83%)	-
Seizure	4/34 (12%)	2/66 (3%)
Septic shock	5/59 (8%)	2/4 (50%)
Skin lesions	1/1 (100%)	-
Sinus tachycardia	3/22 (14%)	-
Splenomegaly	1/1 (100%)	-

Sign or symptom	**Baseline**	**Post-baseline**
ST-segment depressions	1/22 (5%)	-
Staggering gait	1/1 (100%)	-
Stiff back	1/25 (4%)	-
Stiff neck	2/26 (8%)	1/slash3 (33%)
Stupor	1/1 (100%)	-
Subpleural pulmonary haemorrhages	-	1/1 (100%)
Sweating	2/26 (8%)	-
Tachycardia	4/28 (14%)	1/25 (4%)
Tachypnoea	1/1 (100%)	2/2 (100%)
Thrombocytopenia	2/2 (100%)	4/25 (16%)
Thrombosis	-	1/1 (100%)
Tonsillitis	2/2 (100%)	-
Tracheal displacement	1/(25%)	-
Tremor	1/1 (100%)	-
Tubercule-like bodies in iris	-	1/1 (100%)
Unspecified gastrointestinal complaints	40/42 (95%)	-
Upper respiratory infection	-	1/65 (2%)
Vertigo	1/1 (100%)	-

Bibliography

[1] Rindra Vatosoa Randremanana et al. "An open-label, randomized, non-inferiority trial of the efficacy and safety of ciprofloxacin versus streptomycin + ciprofloxacin in the treatment of bubonic plague (IMASOY): study protocol for a randomized control trial". en. In: *Trials* 21.1 (Aug. 2020), p. 722.

[2] Rindra Vatosoa Randremanana et al. "An open-label, randomized, non-inferiority trial of the efficacy and safety of ciprofloxacin versus an aminoglycoside + ciprofloxacin in the treatment of bubonic plague (IMASOY): study protocol for a randomized control trial-an update to the published protocol". en. In: *Trials* 25.1 (July 2024), p. 457.

[3] European Medicines Agency. *ICH guideline E8 (R1) on general considerations for clinical studies*. en. Accessed: 2024-7-23. Apr. 2022. URL: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ich-guideline-e8-r1-general-considerations-clinical-studies_en.pdf.

[4] United States Food and Drug Administration. *Multiple Endpoints in Clinical Trials*. URL: https://www.fda.gov/files/drugs/published/Multiple-Endpoints-in-Clinical-Trials-Guidance-for-Industry.pdf. (accessed: 02.08.2024).

[5] Sabue Mulangu et al. "A randomized, controlled trial of Ebola virus disease therapeutics". en. In: *N. Engl. J. Med.* 381.24 (Dec. 2019), pp. 2293–2303.

[6] United States Food and Drug Administration. *Animal Rule Information*. URL: https://www.fda.gov/emergency-preparedness-and-response/mcm-regulatory-science/animal-rule-information. (accessed: 10.04.2024).

[7] United States Food and Drug Administration. *Product Development Under the Animal Rule: Guidance for Industry*. URL: https://www.fda.gov/emergency-preparedness-and-response/preparedness-research/animal-rule-information. (accessed: 10/07/2025).

[8] National Institute of Allergy and Infectious Diseases. *Biodefense & Emerging Infectious Diseases*. en. Accessed: 2025-01-26. URL: https://www.niaid.nih.gov/research/niaid-biodefense-pathogens.

[9] Nigeria Centre for Disease Control. *National Guideline for Lassa Fever Case Management*. URL: https://ncdc.gov.ng/themes/common/docs/protocols/92_1547068532.pdf. (accessed: 12.12.2020).

[10] World Health Organisation. *Therapeutics for Ebola virus disease*. URL: https://iris.who.int/bitstream/handle/10665/361697/9789240055742-eng.pdf?sequence=1. (accessed: 10.04.2024).

[11] United States Food and Drug Administration. *FDA approves treatment for ebola virus*. URL: https://www.fda.gov/drugs/news-events-human-drugs/fda-approves-treatment-ebola-virus. (accessed: 17.07.2024).

[12] United States Food and Drug Administration. *FDA Approves First Treatment for Ebola Virus*. URL: https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-ebola-virus. (accessed: 17.07.2024).

[13] World Health Organisation. *Clinical Management and infection prevention and control for monkeypox*. URL: https://iris.who.int/bitstream/handle/10665/355798/WHO-MPX-Clinical_and_IPC-2022.1-eng.pdf?sequence=1. (accessed: 26.07.2024).

[14] European Medicines Agency. *Tecovirimat SIGA*. en. Accessed: 2024-7-23. 2023. URL: https://www.ema.europa.eu/en/medicines/human/EPAR/tecovirimat-siga.

[15] Centers for Disease Control and Prevention (CDC). *CDC Guidelines for the Prevention and Treatment of Anthrax, 2023*. 2023. URL: https://www.cdc.gov/mmwr/volumes/72/rr/rr7206a1.htm. (accessed: 26.07.2024).

[16] United States Food and Drug Administration. *Products Approved for Anthrax*. URL: https://www.fda.gov/drugs/bioterrorism-and-drug-preparedness/products-approved-anthrax. (accessed: 26.07.2024).

[17] Centers for Disease Control and Prevention (CDC). *Recommended antibiotic treatment for plague*. 2015. URL: https://www.cdc.gov/plague/resources/Recommended-antibiotics-for-plague_revision-Aug-2015_Final-(00000002).pdf#:~:text=Not%20FDA%20approved%20but%20considered%20an%20effective%20alternative,or%20dual%20therapy%20for%20patients%20with%20bubonic%20disease. (accessed: 01.05.2024).

[18] United States Food and Drug Administration. *Products Approved for Other Bioterrorism Emergencies*. URL: https://www.fda.gov/drugs/bioterrorism-and-drug-preparedness/products-approved-other-bioterrorism-emergencies. (accessed: 01.05.2024).

[19] E A Eisenhauer et al. "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)". en. In: *Eur. J. Cancer* 45.2 (Jan. 2009), pp. 228–247.

[20] Filip Janku et al. "Outcomes of phase II clinical trials with single-agent therapies in advanced/metastatic non-small cell lung cancer published between 2000 and 2009". en. In: *Clin. Cancer Res.* 18.22 (Nov. 2012), pp. 6356–6363.

[21] Lesley Seymour et al. "iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics". en. In: *Lancet Oncol.* 18.3 (Mar. 2017), e143–e152.

[22] Riccardo Lencioni and Josep M Llovet. "Modified RECIST (mRECIST) assessment for hepatocellular carcinoma". en. In: *Semin. Liver Dis.* 30.1 (Feb. 2010), pp. 52–60.

[23] United States Food and Drug Administration. *Diabetes Mellitus: Efficacy Endpoints for Clinical Trials Investigating Antidiabetic Drugs and Biological Products*. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/diabetes-mellitus-efficacy-endpoints-clinical-trials-investigating-antidiabetic-drugs-and-biological. (accessed: 16.09.2024).

[24] European Medicines Agency. *Clinical development of medicinal products for treatment of HIV infection - Scientific guideline*. en. 2017. URL: https://www.ema.europa.eu/en/clinical-development-medicinal-products-treatment-hiv-infection-scientific-guideline. (accessed: 16.09.2024).

[25] World Health Organisation. *Guidelines on clinical evaluation of vaccines: regulatory expectations*. URL: https://www.who.int/publications/m/item/WHO-TRS-1004-web-annex-9. (accessed: 16.09.2024).

[26] UK Health Security Agency. *High consequence infectious diseases (HCID)*. URL: https://www.gov.uk/guidance/high-consequence-infectious-diseases-hcid. (accessed: 31.07/2024).

[27] United States Centres for Disease Control and Prevention. *National Center for Emerging and Zoonotic Infectious Diseases (NCEZID)*. URL: https://www.cdc.gov/ncezid/topics-programs/index.html. (accessed: 31.07.2024).

[28] World Health Organisation. *Prioritizing diseases for research and development in emergency contexts*. URL: https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts. (accessed: 31.07.2024).

[29] M Roser, H Ritchie, and F Spooner. *Burden of disease*. URL: https://ourworldindata.org/burden-of-disease.

[30] Hyuna Sung et al. "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". en. In: *CA Cancer J. Clin.* 71.3 (May 2021), pp. 209–249.

[31] Eric Betherat. *Plague around the world 2019 - La peste dans le monde 2019*. URL: https://iris.who.int/bitstream/handle/10665/325482/WER9425-289-292-en-fr.pdf. (accessed: 10.04.2024).

[32] Vinod Soman Pillai, Gayathri Krishna, and Mohanan Valiya Veettil. "Nipah virus: Past outbreaks and future containment". en. In: *Viruses* 12.4 (Apr. 2020), p. 465.

[33] Md Zakiul Hassan et al. "Nipah virus disease: what can we do to improve patient care?" en. In: *Lancet Infect. Dis.* 24.7 (July 2024), e463–e471.

[34] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2023. URL: https://www.R-project.org/.

[35] S Champely et al. *pwr*. URL: https://CRAN.R-project.org/package=pwr.

[36] Moreno Ursino and Nigel Stallard. "Bayesian approaches for confirmatory trials in rare diseases: Opportunities and challenges". en. In: *Int. J. Environ. Res. Public Health* 18.3 (Jan. 2021), p. 1022.

[37] Lori E Dodd et al. "Design of a randomized controlled trial for Ebola virus disease medical countermeasures: PREVAIL II, the Ebola MCM study". en. In: *J. Infect. Dis.* 213.12 (June 2016), pp. 1906–1913.

[38] Nigeria Centre for Disease Control. *Lassa Fever Situation ReportvEpi Week 52: 25th – 31st December 2023*. 2023. URL: https://ncdc.gov.ng/diseases/sitreps/?cat=5&name=An%20update%20of%20Lassa%20fever%20outbreak%20in%20Nigeria. (accessed: 31.07.2024).

[39] Mahmood M Dalhat et al. "Epidemiological trends of Lassa fever in Nigeria, 2018-2021". en. In: *PLoS One* 17.12 (Dec. 2022), e0279467.

[40] David M Kent et al. "Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal". en. In: *Trials* 11.1 (Aug. 2010), p. 85.

[41] Samuel R P Franzen, Clare Chandler, and Trudie Lang. "Health research capacity development in low and middle income countries: reality or rhetoric? A systematic meta-narrative review of the qualitative literature". en. In: *BMJ Open* 7.1 (Jan. 2017), e012332.

[42] Marjorie J Good et al. "Measuring clinical trial-associated workload in a community clinical oncology program". en. In: *J. Oncol. Pract.* 9.4 (July 2013), pp. 211–215.

[43] Delanyo Dovlo. "Wastage in the health workforce: some perspectives from African countries". en. In: *Hum. Resour. Health* 3.1 (Aug. 2005), p. 6.

[44] Davies Adeloye et al. "Health workforce and governance: the crisis in Nigeria". en. In: *Hum. Resour. Health* 15.1 (May 2017), p. 32.

[45] Syed Masud Ahmed et al. “The health workforce crisis in Bangladesh: shortage, inappropriate skill-mix and inequitable distribution”. en. In: *Hum. Resour. Health* 9.1 (Jan. 2011), p. 3.

[46] ISARIC. *Clinical development strategy to deliver new Lassa fever therapeutics*. URL: https://isaric.org/research/lassa-fever-resources/walc/. (accessed: 26.04.2024).

[47] Jae W Song and Kevin C Chung. “Observational studies: cohort and case-control studies”. en. In: *Plast. Reconstr. Surg.* 126.6 (Dec. 2010), pp. 2234–2242.

[48] United States Food and Drug Administration. *Surrogate Endpoint Resources for Drug and Biologic Development*. URL: https://www.fda.gov/drugs/development-resources/surrogate-endpoint-resources-drug-and-biologic-development. (accessed: 06.08.2024).

[49] Nick Freemantle et al. “Composite outcomes in randomized trials: greater precision but with greater uncertainty?” en. In: *JAMA* 289.19 (May 2003), pp. 2554–2559.

[50] Robert Goldberg et al. “Individual and composite study endpoints: separating the wheat from the chaff”. en. In: *Am. J. Med.* 127.5 (May 2014), pp. 379–384.

[51] PREVAIL II Writing Group et al. “A randomized, controlled trial of ZMapp for Ebola virus infection”. en. In: *N. Engl. J. Med.* 375.15 (Oct. 2016), pp. 1448–1456.

[52] Jake Dunning et al. “Experimental treatment of Ebola virus disease with brincidofovir”. en. In: *PLoS One* 11.9 (Sept. 2016), e0162199.

[53] Jake Dunning et al. “Experimental treatment of Ebola virus disease with TKM-130803: A single-arm phase 2 clinical trial”. en. In: *PLoS Med.* 13.4 (Apr. 2016), e1001997.

[54] Mandy Kader Konde et al. “Interferon β-1a for the treatment of Ebola virus disease: A historically controlled, single-arm proof-of-concept trial”. en. In: *PLoS One* 12.2 (Feb. 2017), e0169255.

[55] Daouda Sissoko et al. “Experimental treatment with favipiravir for Ebola virus disease (the JIKI trial): A historically controlled, single-arm proof-of-concept trial in guinea”. en. In: *PLoS Med.* 13.3 (Mar. 2016), e1001967.

[56] Johan van Griensven et al. “Evaluation of convalescent plasma for Ebola virus disease in Guinea”. en. In: *N. Engl. J. Med.* 374.1 (Jan. 2016), pp. 33–42.

[57] H T Chong et al. “Treatment of acute Nipah encephalitis with ribavirin”. en. In: *Ann. Neurol.* 49.6 (June 2001), pp. 810–813.

[58] William Mwengee et al. “Treatment of plague with gentamicin or doxycycline in a randomized clinical trial in Tanzania”. en. In: *Clin. Infect. Dis.* 42.5 (Mar. 2006), pp. 614–621.

[59] *ISTH/ANRS 0409s INTEGRATE Lassa Fever Study*. en. Accessed: 2024-07-17. 2024. URL: https://clinicaltrials.gov/study/NCT06212336?cond=lassa%20fever&rank=1#study-plan.

[60] C Eric McCoy. “Understanding the use of composite endpoints in clinical trials”. en. In: *West. J. Emerg. Med.* 19.4 (July 2018), pp. 631–634.

[61] Sue Ross. “Composite outcomes in randomized clinical trials: arguments for and against”. en. In: *Am. J. Obstet. Gynecol.* 196.2 (Feb. 2007), 119.e1–6.

[62] European Medicines Agency. *Guideline on multiplicity issues in clinical trials*. en. Accessed: 2024-9-25. 2017. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-multiplicity-issues-clinical-trials_en.pdf.

[63] Laura Merson et al. “Clinical characterization of Lassa fever: A systematic review of clinical reports and research to inform clinical trial design”. en. In: 15.9 (Sept. 2021), e0009788.

[64] Andrea M McCollum et al. “Epidemiology of human mpox - worldwide, 2018-2021”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 72.3 (Jan. 2023), pp. 68–72.

[65] Voahangy Andrianaivoarimanana et al. “Understanding the persistence of plague foci in Madagascar”. en. In: *PLoS Negl. Trop. Dis.* 7.11 (Nov. 2013), e2382.

[66] World Health Organisation. *Lassa fever*. URL: https://www.who.int/health-topics/lassa-fever. (accessed: 12.12.2020).

[67] Camille Besombes et al. “National Monkeypox surveillance, Central African Republic, 2001-2021”. en. In: *Emerg. Infect. Dis.* 28.12 (Dec. 2022), pp. 2435–2445.

[68] World Health Organisation. *Plague*. URL: https://www.who.int/news-room/fact-sheets/detail/plague. (accessed: 30.11.2021).

[69] R Barbieri et al. “Yersinia pestis: the Natural History of Plague”. en. In: *Clin. Microbiol. Rev.* 34.1 (Dec. 2020).

[70] World Health Organisation. *Plague outbreak toolbox*. en. URL: %5Curl%7Bhttps://www.who.int/emergencies/outbreak-toolkit/disease-outbreak-toolboxes/plague-outbreak-toolbox%7D. Accessed: 2024-4-30.

[71] Sophie Jullien, Harsha A Dissanayake, and Marty Chaplin. “Rapid diagnostic tests for plague”. en. In: *Cochrane Database Syst. Rev.* 6 (June 2020), p. CD013459.

[72] United States Centres for Disease Control and Prevention. *Plague: Symptoms*. URL: https://www.cdc.gov/plague/symptoms/index.html. (accessed: 30.11.2021).

[73] Rindra Randremanana et al. “Epidemiological characteristics of an urban plague epidemic in Madagascar, August-November, 2017: an outbreak report”. en. In: *Lancet Infect. Dis.* 19.5 (May 2019), pp. 537–545.

[74] United States Food and Drug Administration. *CDER Drug and Biologic Animal Rule Approvals*. URL: https://www.fda.gov/media/150191/download?attachment. (accessed: 01.05.2024).

[75] United States Food and Drug Administration. *Products Approved for Other Bioterrorism Emergencies*. URL: https://www.fda.gov/drugs/bioterrorism-and-drug-preparedness/products-approved-other-bioterrorism-emergencies. (accessed: 30.11.2021).

[76] Christina A Nelson et al. “Antimicrobial treatment and prophylaxis of plague: Recommendations for naturally acquired infections and bioterrorism response”. en. In: *MMWR Recomm. Rep.* 70.3 (July 2021), pp. 1–27.

[77] Titus Apangu et al. “Successful treatment of human plague with oral ciprofloxacin”. en. In: *Emerg. Infect. Dis.* 23.3 (Mar. 2017).

[78] Nigeria Centre for Disease Control. *An update of Lassa fever outbreak in Nigeria.* URL: https://ncdc.gov.ng/diseases/sitreps/?cat=5&name=An%20update%20of%20Lassa%20fever%20outbreak%20in%20Nigeria. (accessed: 09.04.2024).

[79] Jeffrey G Shaffer et al. "Space-time trends in Lassa fever in Sierra Leone by ELISA serostatus, 2012-2019". en. In: *Microorganisms* 9.3 (Mar. 2021), p. 586.

[80] Emmanuel Dwalu et al. "Trend of Lassa fever cases and factors associated with mortality in Liberia, 2016 - 2021: a secondary data analysis". en. In: *Pan Afr. Med. J.* 47 (2024).

[81] Adrian Q N Mylne et al. "Mapping the zoonotic niche of Lassa fever in Africa". en. In: *Trans. R. Soc. Trop. Med. Hyg.* 109.8 (Aug. 2015), pp. 483–492.

[82] Africa Centres for Disease Control and Prevention. *Lassa Fever*. en. Accessed: 2024-4-25. Sept. 2019. URL: %5Curl%7Bhttps://africacdc.org/disease/lassa-fever/%7D.

[83] Stephan Günther and Oliver Lenz. "Lassa virus". en. In: *Crit. Rev. Clin. Lab. Sci.* 41.4 (2004), pp. 339–390.

[84] Nnennaya A Ajayi et al. "Containing a Lassa fever epidemic in a resource-limited setting: outbreak description and lessons learned from Abakaliki, Nigeria (January-March 2012)". en. In: *Int. J. Infect. Dis.* 17.11 (Nov. 2013), e1011–6.

[85] A Mba et al. "Lassa fever infection among health care workers in Nigeria, 2019". In: *International Journal of Infectious Diseases* (2020).

[86] Peter Okokhere et al. "Clinical and laboratory predictors of Lassa fever outcome in a dedicated treatment facility in Nigeria: a retrospective, observational cohort study". en. In: *Lancet Infect. Dis.* 18.6 (June 2018), pp. 684–695.

[87] George O Akpede et al. "Caseload and case fatality of Lassa fever in Nigeria, 2001-2018: A specialist center's experience and its implications". en. In: *Front. Public Health* 7 (June 2019), p. 170.

[88] Nzelle D Kayem et al. "Lassa fever in pregnancy: a systematic review and meta-analysis". en. In: *Trans. R. Soc. Trop. Med. Hyg.* 114.5 (May 2020), pp. 385–396.

[89] Robert J Samuels et al. “Lassa fever among children in eastern province, Sierra Leone: A 7-year retrospective analysis (2012-2018)”. en. In: *Am. J. Trop. Med. Hyg.* 104.2 (Nov. 2020), pp. 585–592.

[90] Alexandre Duvignaud et al. “Lassa fever outcomes and prognostic factors in Nigeria (LASCOPE): a prospective cohort study”. en. In: *Lancet Glob. Health* 9.4 (Apr. 2021), e469–e478.

[91] J D Frame et al. “Lassa fever, a new virus disease of man from West Africa. I. Clinical description and pathological findings”. en. In: *Am. J. Trop. Med. Hyg.* 19.4 (July 1970), pp. 670–676.

[92] J B McCormick et al. “Lassa fever. Effective therapy with ribavirin”. en. In: *N. Engl. J. Med.* 314.1 (Jan. 1986), pp. 20–26.

[93] Kirsten Alexandra Eberhardt et al. “Ribavirin for the treatment of Lassa fever: A systematic review and meta-analysis”. en. In: *Int. J. Infect. Dis.* 87 (Oct. 2019), pp. 15–20.

[94] Alex Paddy Salam et al. “Time to reconsider the role of ribavirin in Lassa fever”. en. In: *PLoS Negl. Trop. Dis.* 15.7 (July 2021), e0009522.

[95] *A Study to Assess the Safety, Tolerability, and Pharmacokinetics of Oral ARN-75039 in Healthy Adult Subjects*. en. Accessed: 2021-04-25. 2023. URL: https://classic.clinicaltrials.gov/ct2/show/NCT05735249?term=treat&cond=lassa+fever&draw=2&rank=5.

[96] *Multiple Ascending Oral Dose 14-Day Trial of LHF-535 in Healthy Participants*. en. Accessed: 2021-02-18. 2019. URL: https://clinicaltrials.gov/ct2/show/NCT03993704?cond=Lassa+Fever&draw=3&rank=12.

[97] *NCT04055454 A Trial to Evaluate the Optimal Dose of MV-LASV*. en. Accessed: 2021-02-18. 2019. URL: https://classic.clinicaltrials.gov/ct2/show/NCT04055454?term=treat&cond=lassa+fever&draw=2&rank=4.

[98] Oriol Mitjà et al. “Monkeypox”. en. In: *Lancet* 401.10370 (Jan. 2023), pp. 60–74.

[99] European Centre for Disease Prevention and Control. *Implications for the EU/EEA of the outbreak of mpox caused by Monkeypox virus clade I in the Democratic Republic of the Congo*. en. Accessed: 2024-4-25. Dec. 2023. URL: %5Curl%7Bhttps://www.ecdc.europa.eu/en/publications-data/implications-eueea-outbreak-mpox-caused-monkeypox-virus-clade-i-democratic%7D.

[100] World Health Organisation. *Mpox (monkeypox) - Democratic Republic of the Congo*. en. Accessed: 2024-4-25. URL: %5Curl%7Bhttps://www.who.int/emergencies/disease-outbreak-news/item/2023-DON493%7D.

[101] Eskild Petersen et al. "Human Monkeypox: Epidemiologic and clinical characteristics, diagnosis, and prevention". en. In: *Infect. Dis. Clin. North Am.* 33.4 (Dec. 2019), pp. 1027–1043.

[102] Yunfan Zhou and Zixin Chen. "Mpox: a review of laboratory detection techniques". en. In: *Arch. Virol.* 168.8 (Aug. 2023), p. 221.

[103] Leandre Murhula Masirika et al. "Ongoing mpox outbreak in Kamituga, South Kivu province, associated with monkeypox virus of a novel Clade I sub-lineage, Democratic Republic of the Congo, 2024". en. In: *Euro Surveill.* 29.11 (Mar. 2024).

[104] *The antiviral tecovirimat is safe but did not improve clade I mpox resolution in Democratic Republic of the Congo*. en. Accessed: 2025-3-14. Aug. 2024.

[105] Urbà González et al. "Interventions for Old World cutaneous leishmaniasis". en. In: *Cochrane Database Syst. Rev.* 4 (Oct. 2008), p. CD005067.

[106] Urbà González et al. "Interventions for American cutaneous and mucocutaneous leishmaniasis". en. In: *Cochrane Database Syst. Rev.* 2 (Apr. 2009), p. CD004834.

[107] Piero Olliaro et al. "Methodology of clinical trials aimed at assessing interventions for cutaneous leishmaniasis". en. In: *PLoS Negl. Trop. Dis.* 7.3 (Mar. 2013), e2130.

[108] Ben J Marais et al. "Standardized methods for enhanced quality and comparability of tuberculous meningitis studies". en. In: *Clin. Infect. Dis.* 64.4 (Feb. 2017), pp. 501–509.

[109] Mourad Ouzzani et al. "Rayyan-a web and mobile app for systematic reviews". en. In: *Syst. Rev.* 5.1 (Dec. 2016), p. 210.

[110] Paul A Harris et al. "Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support". en. In: *J. Biomed. Inform.* 42.2 (Apr. 2009), pp. 377–381.

[111] Joanna Briggs Institute. *Critical Appraisal Tools*. URL: https://jbi.global/critical-appraisal-tools.

[112] Nick Chapman et al. *Landscape of emerging infectious disease research and development: Preventing the next pandemic*. URL: https://s3-ap-southeast-2.amazonaws.com/policy-cures-website-assets/app/uploads/2020/09/16021708/EID20_Report_Web_1_FINAL.pdf. (accessed: 10.04.2024).

[113] World Health Organisation. *An R&D Blueprint for Action to Prevent Epidemics*. URL: https://www.who.int/publications/m/item/an-r-d-blueprint-for-action-to-prevent-epidemics. (accessed: 10.04.2024).

[114] CEPI. *Largest ever Lassa fever study expands to more countries in West Africa*. URL: https://cepi.net/largest-ever-lassa-fever-study-expands-more-countries-west-africa. (accessed: 10.04.2024).

[115] United States Food and Drug Administration. *Tropical Disease Priority Review Voucher Program*. URL: https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/tropical-disease-priority-review-voucher-program. (accessed: 14.08.2021).

[116] World Health Organisation. *Lassa fever technical guidance*. URL: https://www.who.int/health-topics/lassa-fever/technical-guidance. (accessed: 12.12.2020).

[117] S K Nijjar et al. "Participation in clinical trials improves outcomes in women's health: a systematic review and meta-analysis". en. In: *BJOG* 124.6 (May 2017), pp. 863–871.

[118] Joseph M Unger et al. "Comparison of survival outcomes among cancer patients treated in and out of clinical trials". en. In: *J. Natl. Cancer Inst.* 106.3 (Mar. 2014), dju002.

[119] Kiersten J Kugeler et al. “Antimicrobial treatment patterns and illness outcome among United States patients with plague, 1942-2018”. en. In: *Clin. Infect. Dis.* 70.70 Suppl 1 (May 2020), S20–S26.

[120] Katherine Lazet and Stephanie Rutterbush. “6-day history of fever · groin pain and swelling · recent hiking trip in Colorado · Dx?” en. In: *J. Fam. Pract.* 67.12 (Dec. 2018), pp. 780–782.

[121] A J Moreno et al. “Imaging in plague”. en. In: *Eur. J. Nucl. Med.* 13.6 (1987), pp. 315–317.

[122] T L Stahly and J D Shoop. “Plague and the gallium scan: Case report”. en. In: *J. Nucl. Med.* 16.11 (Nov. 1975), pp. 1031–1032.

[123] B G Weniger et al. “Human bubonic plague transmitted by a domestic cat scratch”. en. In: *JAMA* 251.7 (Feb. 1984), pp. 927–928.

[124] Lei Xu et al. “Nonlinear effect of climate on plague during the third pandemic in China”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 108.25 (June 2011), pp. 10214–10219.

[125] Voahangy Andrianaivoarimanana et al. “Trends of human plague, Madagascar, 1998-2016”. en. In: *Emerg. Infect. Dis.* 25.2 (Feb. 2019), pp. 220–228.

[126] Cecilia A C Prinsen et al. “How to select outcome measurement instruments for outcomes included in a “Core Outcome Set” - a practical guideline”. en. In: *Trials* 17.1 (Sept. 2016), p. 449.

[127] Rafael José Vieira et al. “Control of Allergic Rhinitis and Asthma Test: A systematic review of measurement properties and COSMIN analysis”. en. In: *Clin. Transl. Allergy* 12.9 (Sept. 2022), e12194.

[128] Maria Monica Ratti et al. “A systematic review to evaluate patient-reported outcome measures (PROMs) for metastatic prostate cancer according to the COnsensus-based standard for the selection of health Measurement INstruments (COSMIN) methodology”. en. In: *Cancers (Basel)* 14.20 (Oct. 2022), p. 5120.

[129] STARSurg Collaborative. "Quality of life instruments in acute and chronic pancreatitis: a consensus-based standards for the selection of health measurement instruments (COSMIN) approach". en. In: *HPB (Oxford)* 26.7 (July 2024), pp. 859–872.

[130] *Study of Tecovirimat for Human Monkeypox Virus (STOMP)*. en. Accessed: 2021-04-25. 2022. URL: https://clinicaltrials.gov/study/NCT05534984?cond=mpox&rank=6.

[131] Amanda Rojek, Jake Dunning, and Piero Olliaro. "Monkeypox: how will we know if the treatments work?" en. In: *Lancet Infect. Dis.* 22.9 (Sept. 2022), pp. 1269–1270.

[132] Nick Feasey et al. "Neglected tropical diseases". en. In: *Br. Med. Bull.* 93.1 (2010), pp. 179–200.

[133] *Antiviral treatment with tecovirimat for patients managed at home with mpox*. en. Accessed: 2021-04-25. 2022. URL: https://www.isrctn.com/ISRCTN17461766?q=tecovirimat&filters=&sort=&offset=1&totalResults=3&page=1&pageSize=10.

[134] *Assessment of the Efficacy and Safety of Tecovirimat in Patients With Monkeypox Virus Disease (UNITY)*. en. Accessed: 2024-07-17. 2023. URL: https://clinicaltrials.gov/study/NCT05597735?intr=tecovirimat&rank=3#study-plan.

[135] Josephine Bourner et al. "A systematic review of the clinical profile of patients with bubonic plague and the outcome measures used in research settings". en. In: *PLoS Negl. Trop. Dis.* 17.11 (Nov. 2023), e0011509.

[136] Jan Kottner et al. "Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed". en. In: *J. Clin. Epidemiol.* 64.1 (Jan. 2011), pp. 96–106.

[137] Paul A Harris et al. "The REDCap consortium: Building an international community of software platform partners". en. In: *J. Biomed. Inform.* 95.103208 (July 2019), p. 103208.

[138] Gerald F Cox. "The art and science of choosing efficacy endpoints for rare disease clinical trials". en. In: *Am. J. Med. Genet. A* 176.4 (Apr. 2018), pp. 759–772.

[139] Johan Verbeeck et al. “Composite endpoints, including patient reported outcomes, in rare diseases”. en. In: *Orphanet J. Rare Dis.* 18.1 (Sept. 2023), p. 262.

[140] Michele Nguyen et al. “Mapping malaria seasonality in Madagascar using health facility data”. en. In: *BMC Med.* 18.1 (Feb. 2020), p. 26.

[141] Yvan J F Hutin et al. “Outbreak of human Monkeypox, democratic republic of Congo, 1996 to 1997”. In: *Emerg. Infect. Dis.* 7.3 (June 2001), pp. 434–438.

[142] Mary G Reynolds et al. “Detection of human monkeypox in the Republic of the Congo following intensive community education”. en. In: *Am. J. Trop. Med. Hyg.* 88.5 (May 2013), pp. 982–985.

[143] L Khodakevich et al. “Monkey pox virus infection in humans in the Central African Republic”. fr. In: *Bull. Soc. Pathol. Exot. Filiales* 78.3 (1985), pp. 311–320.

[144] Anne W Rimoin et al. “Endemic human monkeypox, Democratic Republic of Congo, 2001-2004”. en. In: *Emerg. Infect. Dis.* 13.6 (June 2007), pp. 934–937.

[145] Adesola Yinka-Ogunleye et al. “Outbreak of human monkeypox in Nigeria in 2017-18: a clinical and epidemiological report”. en. In: *Lancet Infect. Dis.* 19.8 (Aug. 2019), pp. 872–879.

[146] *Tecovirimat for Treatment of Monkeypox Virus*. en. Accessed: 2021-04-25. 2022. URL: https://clinicaltrials.gov/study/NCT05559099?term=palm%20007&rank=1.

[147] *Phase III Study of ASP2151 in Herpes Zoster Patients*. en. Accessed: 2025-03-07. 2018. URL: https://clinicaltrials.gov/study/NCT01959841?term=lesion%20resolution&cond=herpes&rank=2#study-plan.

[148] *Monoclonal Antibody Therapy Against Chronic Herpes Simplex Virus 1 Infection (MATCH-1)*. en. Accessed: 2025-03-07. 2023. URL: https://clinicaltrials.gov/study/NCT04539483?term=lesion%20resolution&cond=herpes&rank=1#study-plan.

[149] *Combination, Miltefosine Monotherapy for Cutaneous Leishmaniasis in New World*. en. Accessed: 2025-03-07. 2024. URL: https://clinicaltrials.gov/study/NCT04515186?cond=cutaneous%20leishmaniasis&rank=5#study-plan.

[150] *Once Daily Dosing of Lonafarnib Co-administered With Ritonavir for Treatment of Chronic Hepatitis D Virus Infection (LOWR6)*. en. Accessed: 2025-03-11. 2023. URL: https://clinicaltrials.gov/study/NCT05229991?outc=viral%20load&rank=2#study-plan.

[151] *Study to Evaluate the Efficacy of MONotherapy of TiviCAY® Versus a Triple Therapy in HIV-1-infected Patients (MONCAY)*. en. Accessed: 2025-03-11. 2023. URL: https://clinicaltrials.gov/study/NCT02596334?outc=viral%20load&rank=10#study-plan.

[152] *A Study of LAM-002A for the Prevention of Progression of COVID-19*. en. Accessed: 2025-03-11. 2023. URL: https://clinicaltrials.gov/study/NCT04446377?outc=viral%20load&rank=8#study-plan.

[153] Darrell H S Tan et al. "Longitudinal analysis of mpox virus DNA detectability from multiple specimen types during acute illness: A cohort study". en. In: *Open Forum Infect. Dis.* 11.2 (Feb. 2024), ofae073.

[154] Antonio Piralla et al. "Dynamics of viral DNA shedding and culture viral DNA positivity in different clinical samples collected during the 2022 mpox outbreak in Lombardy, Italy". en. In: *Travel Med. Infect. Dis.* 59.102698 (May 2024), p. 102698.

[155] Diane Thiboutot et al. "Parametric Acne Severity (PAS) score and lesion counts from multi-modality facial image analysis correlates strongly with investigator assessment". en. In: *J. Drugs Dermatol.* 20.6 (June 2021), pp. 642–647.

[156] Yao Zhang et al. "Toward automated assessment of mole similarity on dermoscopic images". en. In: *J. Med. Imaging (Bellingham)* 8.1 (Jan. 2021), p. 014506.

[157] S R Feldman and G G Krueger. "Psoriasis assessment tools in clinical trials". en. In: *Ann. Rheum. Dis.* 64 Suppl 2.suppl_2 (Mar. 2005), ii65–8, discussion ii69–73.

[158] Benjamin Jones et al. "Variability in clinical assessment of clade IIb mpox lesions". en. In: *Int. J. Infect. Dis.* 137 (Oct. 2023), pp. 60–62.

[159] World Health Organisation. *Atlas of mpox lesions: a tool for clinical researchers, 28 April 2023, version 1.0*. en. Accessed: 2024-4-25. URL: %5Curl%7Bhttps://

www.who.int/publications-detail-redirect/WHO-MPX-Clinical-Lesions-2023.1%7D.

[160] Douglas G Altman. *Practical statistics for medical research*. Chapman and Hall/CRC, Nov. 1990.

[161] J R Landis and G G Koch. "The measurement of observer agreement for categorical data". en. In: *Biometrics* 33.1 (Mar. 1977), pp. 159–174.

[162] Tushar Nayak et al. "Deep learning based detection of monkeypox virus using skin lesion images". en. In: *Med. Nov. Technol. Devices* 18 (June 2023), p. 100243.

[163] Krishnaraj Chadaga et al. "Application of artificial intelligence techniques for Monkeypox: A systematic review". en. In: *Diagnostics (Basel)* 13.5 (Feb. 2023).

[164] *European Trial Into Mpox Infection (EPOXI)*. en. Accessed: 2021-04-25. 2023. URL: https://clinicaltrials.gov/study/NCT06156566?cond=mpox&rank=8.

[165] Clara Suñer et al. "Viral dynamics in patients with monkeypox infection: a prospective cohort study in Spain". en. In: *Lancet Infect. Dis.* 23.4 (Apr. 2023), pp. 445–453.

[166] Auwal I Kabuga and Mohamed E El Zowalaty. "A review of the monkeypox virus and a recent outbreak of skin rash disease in Nigeria". en. In: *J. Med. Virol.* 91.4 (Apr. 2019), pp. 533–540.

[167] Hermann Meyer et al. "Outbreaks of disease suspected of being due to human monkeypox virus infection in the Democratic Republic of Congo in 2001". en. In: *J. Clin. Microbiol.* 40.8 (Aug. 2002), pp. 2919–2921.

[168] Candida Diaz Brochero et al. "Decoding mpox: a systematic review and meta-analysis of the transmission and severity parameters of the 2022-2023 global outbreak". en. In: *BMJ Glob. Health* 10.1 (Jan. 2025).

[169] Yu Li et al. "Real-time PCR assays for the specific detection of monkeypox virus West African and Congo Basin strain DNA". en. In: *J. Virol. Methods* 169.1 (Oct. 2010), pp. 223–227.

[170] R J Cohrs et al. “Analysis of individual human trigeminal ganglia for latent herpes simplex virus type 1 and varicella-zoster virus nucleic acids using real-time PCR”. en. In: *J. Virol.* 74.24 (Dec. 2000), pp. 11464–11471.

[171] United States Food and Drug Administration. *Expanded Access*. URL: https://www.fda.gov/news-events/public-health-focus/expanded-access. (accessed: 27/09/2024).

[172] SIGA Technologies. *TPOXX (tecovirimat) Investigator's Brochure*. 2021. URL: https://www.siga.com/about-siga/#products.

[173] Elliot G Brown. “Methods and pitfalls in searching drug safety databases utilising the Medical Dictionary for Regulatory Activities (MedDRA)”. en. In: *Drug Saf.* 26.3 (2003), pp. 145–158.

[174] Festus Mbrenga et al. “Tecovirimat for Monkeypox in Central African Republic under expanded access”. en. In: *N. Engl. J. Med.* 387.24 (Dec. 2022), pp. 2294–2295.

[175] Josephine Bourner et al. “Challenges in clinical diagnosis of Clade I Mpox: Highlighting the need for enhanced diagnostic approaches”. en. In: *PLoS Negl. Trop. Dis.* 18.6 (June 2024), e0012087.

[176] Hakyoung Kim et al. “Viral load dynamics and shedding kinetics of mpox infection: a systematic review and meta-analysis”. en. In: *J. Travel Med.* 30.5 (Sept. 2023), taad111.

[177] Christoph Zeyen et al. “Clinical spectrum and long-term outcomes of mpox: a cohort study spanning from acute infection to six-month follow-up”. en. In: *BMC Infect. Dis.* 24.1 (Mar. 2024), p. 317.

[178] Ameera S Jaradat et al. “Automated Monkeypox skin lesion detection using deep learning and transfer learning techniques”. en. In: *Int. J. Environ. Res. Public Health* 20.5 (Mar. 2023).

[179] Olivier X Miguel et al. “Deep learning prediction of renal anomalies for prenatal ultrasound diagnosis”. en. In: *Sci. Rep.* 14.1 (Apr. 2024), p. 9013.

[180] Carlos A Meza Ramirez et al. “Advancing cervical cancer diagnosis and screening with spectroscopy and machine learning”. en. In: *Expert Rev. Mol. Diagn.* 23.5 (May 2023), pp. 375–390.

[181] Mitchell Waters and Prasanna Tadi. “Streptomycin”. en. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, Jan. 2025.

[182] United States Food and Drug Administration. *Sterptomycin for Injection, USP*. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2012/064210s009lbl.pdf. (accessed: 10/07/2025).

[183] Diva Baggio and Michelle R Ananda-Rajah. “Fluoroquinolone antibiotics and adverse events”. en. In: *Aust. Prescr.* 44.5 (Oct. 2021), pp. 161–164.

[184] Paula R Williamson et al. “The COMET handbook: Version 1.0”. en. In: *Trials* 18.S3 (June 2017).

[185] Adebola Tolulope Olayinka et al. “A standardised Phase III clinical trial framework to assess therapeutic interventions for Lassa fever”. en. In: *PLoS Negl. Trop. Dis.* 16.1 (Jan. 2022), e0010089.

[186] Jamie J Kirkham et al. “Core Outcome Set-STAndards for reporting: The COS-STAR Statement”. en. In: *PLoS Med.* 13.10 (Oct. 2016), e1002148.

[187] Heiko A. von der Gracht. “Consensus measurement in Delphi studies: Review and implications for future quality assurance”. In: *Technological Forecasting and Social Change* 79.8 (2012), pp. 1525–1536. ISSN: 0040-1625. DOI: https://doi.org/10.1016/j.techfore.2012.04.013. URL: https://www.sciencedirect.com/science/article/pii/S0040162512001023.

[188] J L Vincent et al. “The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine”. en. In: *Intensive Care Med.* 22.7 (July 1996), pp. 707–710.

[189] 19 December. *National Early Warning Score (NEWS) 2*. https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2. Accessed: 2024-4-25. Dec. 2017.

[190] John-Ugwuanya A Grace, Ifunanya J Egoh, and Nnenna Udensi. "Epidemiological trends of Lassa fever in Nigeria from 2015-2021: A review". en. In: *Ther. Adv. Infect. Dis.* 8 (Jan. 2021), p. 20499361211058252.

[191] Josephine Bourner et al. "The West Africa Lassa fever Consortium pre-positioned protocol for a Phase II/III adaptive, randomised, controlled, platform trial to evaluate multiple Lassa fever therapeutics". en. In: *Wellcome Open Res.* 8 (June 2023), p. 122.

[192] *Lassa fever therapeutics Developing protocols for Phase II-III clinical trials*. en. Accessed: 2025-02-24. 2022. URL: https://isaric.org/research/lassa-fever-resources/walc/.

[193] Josephine Bourner et al. "Adaptive design for phase II/III platform trial of Lassa fever therapeutics". en. In: *Emerg. Infect. Dis.* 31.2 (Feb. 2025), pp. 9–16.

[194] Elisabeth D Riviello et al. "Hospital incidence and outcomes of the acute respiratory distress syndrome using the Kigali modification of the Berlin definition". en. In: *Am. J. Respir. Crit. Care Med.* 193.1 (Jan. 2016), pp. 52–59.

[195] Mervyn Singer et al. "The third international consensus definitions for sepsis and septic shock (sepsis-3)". en. In: *JAMA* 315.8 (Feb. 2016), pp. 801–810.

[196] Christy Chuang-Stein et al. "Sample size reestimation: A review and recommendations". en. In: *Drug Inf. J.* 40.4 (Oct. 2006), pp. 475–484.

[197] John J Chiosi et al. "Evaluation of three clinical prediction tools to predict mortality in hospitalized patients with Lassa fever". en. In: *Am. J. Trop. Med. Hyg.* 107.4 (Oct. 2022), pp. 856–862.

[198] S Fisher-Hoch et al. "Hematologic dysfunction in Lassa fever". en. In: *J. Med. Virol.* 26.2 (Oct. 1988), pp. 127–135.

[199] *Tecovirimat in Non-hospitalized Patients With Monkeypox (PLATINUM-CAN)*. en. Accessed: 2024-07-17. 2024. URL: https://clinicaltrials.gov/study/NCT05534165?intr=tecovirimat&rank=2#study-plan.

[200] Josephine Bourner et al. “Bubonic plague: can the size of buboes be accurately and consistently measured with a digital calliper?” en. In: *Trials* 24.1 (Dec. 2023), p. 815.

[201] United States Food and Drug Administration. *Clinical Outcome Assessment (COA): Frequently Asked Questions*. URL: https://www.fda.gov/about-fda/clinical-outcome-assessment-coa-frequently-asked-questions. (accessed: 17.07.2024).

[202] United States Food and Drug Administration. *Clinical Outcome Assessment (COA) Qualification Program*. URL: https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/clinical-outcome-assessment-coa-qualification-program. (accessed: 17.07.2024).

[203] T L Anderson. “Two cases of bubonic plague on board ship”. en. In: *BMJ* 2.2175 (Sept. 1902), pp. 730–730.

[204] “Plague The. A Case Of Bubonic Plague On A Vessel Arriving In The Mersey”. In: *The British Medical Journal* 1 (1905), pp. 383–384.

[205] G W McCoy and W B Wherry. “Subacute plague in man due to ground squirrel infection”. In: *J. Infect. Dis.* 6.5 (Nov. 1909), pp. 670–675.

[206] A R Short. “Three cases of bubonic plague arising in England”. en. In: *Br. Med. J.* 2.2905 (Sept. 1916), p. 327.

[207] Ch Haddad and A Valero. “Streptomycin in bubonic plague”. en. In: *BMJ* 1.4560 (May 1948), pp. 1026–1027.

[208] “PLAGUE - New Mexico”. In: *Morb. Mortal.* 14.33 (1965), pp. 285–286.

[209] “PLAGUE: Shasta County, California”. In: *Morb. Mortal.* 14.38 (1965), pp. 325–332.

[210] “SUSPECTED CASE OF IMPORTED BUBONIC PLAGUE - Texas”. In: *Morb. Mortal.* 15.44 (1966), pp. 377–378.

[211] “PLAGUE — Arizona”. In: *Morb. Mortal.* 16.27 (1967), pp. 222–222.

[212] J D Connor et al. “Plague in San Diego”. en. In: *West. J. Med.* 129.5 (Nov. 1978), pp. 394–406.

[213] “PRESUMPTIVE BUBONIC PLAGUE — Denver, Colorado”. In: *Morb. Mortal.* 17.27 (1968), pp. 253–260.

[214] “BUBONIC PLAGUE DEATH – Lemhi county, Idaho”. In: *Morb. Mortal.* 17.44 (1968), pp. 405–406.

[215] “PLAGUE CASE - Navajo Reservation - Kayenta, Arizona”. In: *Morb. Mortal.* 17.29 (1968), pp. 269–270.

[216] “PLAGUE – New Mexico”. In: *Morb. Mortal.* 18.42 (1969), pp. 365–366.

[217] W P Reed et al. “Bubonic plague in the Southwestern United States. A review of recent experience”. en. In: *Medicine (Baltimore)* 49.6 (Nov. 1970), pp. 465–486.

[218] “BUBONIC PLAGUE – California”. In: *Morb. Mortal.* 19.27 (1970), pp. 261–261.

[219] “HUMAN BUBONIC PLAGUE - Cochiti, New Mexico”. In: *Morb. Mortal.* 19.20 (1970), pp. 197–198.

[220] “BUBONIC PLAGUE – Santa Fe, New Mexico”. In: *Morb. Mortal.* 19.32 (1970), pp. 313–314.

[221] “PLAGUE - New Mexico”. In: *Morb. Mortal.* 19.25 (1970), pp. 241–242.

[222] “Front Matter”. In: *Morb. Mortal.* 19.32 (1970), p. 370.

[223] “PLAGUE — California”. In: *Morb. Mortal.* 19.45 (1970), pp. 437–438.

[224] “PLAGUE — Rio en Medio, New Mexico”. In: *Morb. Mortal.* 19.33 (1970), pp. 322–322.

[225] “A FATAL CASE OF DYSENTERY DUE TO SHIGELLA DYSENTERIAE 1 IN AN AMERICAN RESIDENT - Florida”. In: *Morb. Mortal.* 19.28 (1970), pp. 269–270.

[226] “HUMAN BUBONIC PLAGUE – Oregon”. In: *Morb. Mortal.* 20.34 (1971), pp. 303–304.

[227] “BUBONIC PLAGUE – New Mexico”. In: *Morb. Mortal.* 20.32 (1971), pp. 283–284.

[228] D L Palmer et al. “Clinical features of plague in the United States: the 1969-1970 epidemic”. en. In: *J. Infect. Dis.* 124.4 (Oct. 1971), pp. 367–371.

[229] "HUMAN BUBONIC PLAGUE – Coconino County, Arizona". In: *Morb. Mortal.* 21.10 (1972), pp. 81–82.

[230] "HUMAN BUBONIC PLAGUE – Arizona". In: *Morb. Mortal.* 22.30 (1973), pp. 249–250.

[231] Nguyen-Van-Ai et al. "Letter: Co-trimoxazole in bubonic plague". en. In: *Br. Med. J.* 4.5884 (Oct. 1973), pp. 108–109.

[232] "HUMAN BUBONIC PLAGUE - New Mexico". In: *Morb. Mortal.* 23.26 (1974), pp. 231–232.

[233] "HUMAN PLAGUE - New Mexico". In: *Morb. Mortal.* 23.41 (1974), pp. 350–350.

[234] "HUMAN PLAGUE — New Mexico, Utah". In: *Morb. Mortal.* 23.37 (1974), pp. 317–318.

[235] T Butler et al. "Yersinia pestis infection in Vietnam. I. Clinical and hematologic aspects". en. In: *J. Infect. Dis.* 129.Supplement 1 (May 1974), Suppl:S78–84.

[236] "FATAL BUBONIC PLAGUE – California". In: *Morb. Mortal.* 24.24 (1975), p. 211.

[237] "BUBONIC PLAGUE – Arizona". In: *Morb. Mortal.* 24.22 (1975), p. 190.

[238] "PLAGUE IN HUMANS - New Mexico". In: *Morb. Mortal.* 24.40 (1975), pp. 341–342.

[239] "HUMAN PLAGUE CASE - Bernalillo County, New Mexico". In: *Morb. Mortal.* 24.10 (1975), pp. 90, 95.

[240] C F von Reyn et al. "Bubonic plague from exposure to a rabbit: a documented case, and a review of rabbit-associated plague cases in the United States". en. In: *Am. J. Epidemiol.* 104.1 (July 1976), pp. 81–87.

[241] "Human Plague — Arizona, California, New Mexico". In: *Morb. Mortal.* 25.19 (1976), p. 155.

[242] T Butler et al. "Yersinia pestis infection in Vietnam. II. Quantiative blood cultures and detection of endotoxin in the cerebrospinal fluid of patients with meningitis". en. In: *J. Infect. Dis.* 133.5 (May 1976), pp. 493–499.

[243] J M Mann and R Moskowitz. “Plague and pregnancy. A case report”. en. In: *JAMA* 237.17 (Apr. 1977), pp. 1854–1855.

[244] C F von Reyn et al. “Epidemiologic and clinical features of an outbreak of bubonic plague in New Mexico”. en. In: *J. Infect. Dis.* 136.4 (Oct. 1977), pp. 489–494.

[245] “Plague — Arizona, Colorado, New Mexico”. In: *MMWR Morb. Mortal. Wkly. Rep.* 26.26 (1977), pp. 215–216.

[246] “Plague – United States”. In: *MMWR Morb. Mortal. Wkly. Rep.* 26.41 (1977), p. 337.

[247] “Plague – Arizona, California, New Mexico”. In: *MMWR Morb. Mortal. Wkly. Rep.* 27.30 (1978), pp. 259–260.

[248] S L Hoffman. “Plague in the United States: the “black death” is still alive”. en. In: *Ann. Emerg. Med.* 9.6 (June 1980), pp. 319–322.

[249] “Plague — United States, 1980”. In: *MMWR Morb. Mortal. Wkly. Rep.* 29.31 (1980), pp. 371–377.

[250] Centers for Disease Control (CDC). “Human plague – Texas, New Mexico”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 30.12 (Apr. 1981), pp. 137–138.

[251] Centers for Disease Control (CDC). “Human plague associated with domestic cats–California, Colorado”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 30.22 (June 1981), pp. 265–266.

[252] J M Mann et al. “Peripatetic plague”. en. In: *JAMA* 247.1 (Jan. 1982), pp. 47–48.

[253] J M Mann and R F Shaw. “Febrile lymphadenitis in the American west”. en. In: *Hosp. Pract. (Off Ed)* 18.1 (Jan. 1983), pp. 51, 54, 56.

[254] Centers for Disease Control (CDC). “Plague–South Carolina”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 32.32 (Aug. 1983), pp. 417–418.

[255] A M Barnes and J D Poland. “Plague in the United States, 1982”. en. In: *MMWR CDC Surveill. Summ.* 32.3 (Aug. 1983), 19SS–24SS.

[256] Centers for Disease Control (CDC). “Plague pneumonia–California”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 33.34 (Aug. 1984), pp. 481–483.

[257] Centers for Disease Control (CDC). “Winter plague–Colorado, Washington, Texas, 1983-1984”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 33.11 (Mar. 1984), pp. 145–148.

[258] T K Welty et al. “Nineteen cases of plague in Arizona. A spectrum including ecthyma gangrenosum due to plague and plague in pregnancy”. en. In: *West. J. Med.* 142.5 (May 1985), pp. 641–646.

[259] A L Florman, R R Spencer, and S Sheward. “Multiple lung cavities in a 12-year-old girl with bubonic plague, sepsis, and secondary pneumonia”. en. In: *Am. J. Med.* 80.6 (June 1986), pp. 1191–1193.

[260] H F Hull, J M Montes, and J M Mann. “Plague masquerading as gastrointestinal illness”. en. In: *West. J. Med.* 145.4 (Oct. 1986), pp. 485–487.

[261] T W Wong. “Plague in a pregnant patient”. en. In: *Trop. Doct.* 16.4 (Oct. 1986), pp. 187–189.

[262] T M Becker et al. “Plague meningitis–a retrospective analysis of cases reported in the United States, 1970-1979”. en. In: *West. J. Med.* 147.5 (Nov. 1987), pp. 554–557.

[263] Centers for Disease Control (CDC). “Human plague–United States, 1988”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 37.42 (Oct. 1988), pp. 653–656.

[264] Centers for Disease Control (CDC). “Imported bubonic plague–District of Columbia”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 39.49 (Dec. 1990), pp. 895, 901.

[265] D Migden. “Bubonic plague in a child presenting with fever and altered mental status”. en. In: *Ann. Emerg. Med.* 19.2 (Feb. 1990), pp. 207–209.

[266] C Owens. “Plague in New Mexico”. en. In: *J. Community Health Nurs.* 7.3 (1990), pp. 153–158.

[267] L D Crook and B Tempest. “Plague. A clinical review of 27 cases”. en. In: *Arch. Intern. Med.* 152.6 (June 1992), pp. 1253–1256.

[268] K L McClean. “An outbreak of plague in northwestern province, Zambia”. en. In: *Clin. Infect. Dis.* 21.3 (Sept. 1995), pp. 650–652.

[269] Centers for Disease Control and Prevention (CDC). “Fatal human plague–Arizona and Colorado, 1996”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 46.27 (July 1997), pp. 617–620.

[270] S Chanteau et al. “Current epidemiology of human plague in Madagascar”. en. In: *Microbes Infect.* 2.1 (Jan. 2000), pp. 25–31.

[271] K L Gage et al. “Cases of cat-associated human plague in the Western US, 1977-1998”. en. In: *Clin. Infect. Dis.* 30.6 (June 2000), pp. 893–900.

[272] Centers for Disease Control and Prevention (CDC). “Imported plague–New York City, 2002”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 52.31 (Aug. 2003), pp. 725–728.

[273] Kathryn Chmura et al. “Painful lymphadenopathy and fulminant sepsis in a previously healthy 16-year-old girl”. en. In: *Chest* 124.1 (July 2003), pp. 379–382.

[274] L Lucy Boulanger et al. “Gentamicin and tetracyclines for the treatment of human plague: review of 75 cases in new Mexico, 1985-1999”. en. In: *Clin. Infect. Dis.* 38.5 (Mar. 2004), pp. 663–669.

[275] Centers for Disease Control and Prevention (CDC). “Human plague–four states, 2006”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 55.34 (Sept. 2006), pp. 940–943.

[276] Centers for Disease Control and Prevention (CDC). “Notes from the field: two cases of human plague–Oregon, 2010”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 60.7 (Feb. 2011), p. 214.

[277] Mathieu Tourdjman et al. “Misidentification of Yersinia pestis by automated systems, resulting in delayed diagnoses of human plague infections–Oregon and New Mexico, 2010-2011”. en. In: *Clin. Infect. Dis.* 55.7 (Oct. 2012), e58–60.

[278] Nicolas Cabanel et al. “Plague outbreak in Libya, 2009, unrelated to plague in Algeria”. en. In: *Emerg. Infect. Dis.* 19.2 (Feb. 2013), pp. 230–236.

[279] Nyambe Sinyange et al. “Outbreak of plague in a high malaria endemic region - Nyimba District, Zambia, March-May 2015”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 65.31 (Aug. 2016), pp. 807–811.

[280] S D Melman et al. “Human case of bubonic plague resulting from the bite of a wild Gunnison’s prairie dog during translocation from a plague-endemic area”. en. In: *Zoonoses Public Health* 65.1 (Feb. 2018), e254–e258.

[281] Jan Kehrmann et al. “Two fatal cases of plague after consumption of raw marmot organs”. en. In: *Emerg. Microbes Infect.* 9.1 (Dec. 2020), pp. 1878–1880.

[282] Ariella P Dale et al. “Notes from the field: Delays in identification and treatment of a case of septicemic plague - Navajo county, Arizona, 2020”. en. In: *MMWR Morb. Mortal. Wkly. Rep.* 70.31 (Aug. 2021), pp. 1063–1064.