Methodology research in clinical studies for Neglected Tropical Diseases (NTDs) of the skin: Addressing design, conduct and relevance for context and target population

Astrid C. Erber

The Global Health Network, Division of Tropical Medicine and Global Health

Nuffield Department of Medicine

Green Templeton College, University of Oxford

Submitted for the degree of

Doctor of Philosophy in Clinical Medicine

Trinity Term 2017

Methodology research in clinical studies for Neglected Tropical Diseases (NTDs) of the skin:

Addressing design, conduct and relevance for context and target population

Astrid C. Erber, Green Templeton College, submitted for DPhil Clinical Medicine, TT2017

Abstract

Buruli Ulcer (BU) and Cutaneous Leishmaniasis (CL) are two Neglected Tropical Diseases (NTDs) of the skin. Both are 'diseases of poverty' and can cause long-term disabilities. This work comprises two research methodology studies, suggesting novel approaches to design and conduct of clinical studies in order to make research more relevant for their context, and for patients.

The first is a study evaluating a diagnostic test for BU at a point-of-care (PoC) setting in rural Ghana, complemented by interviews and focus group discussions (FGDs) with relevant stakeholders. It reports results from the diagnostics trial and points out roles for a Bayesian statistical framework using existing data on the one hand, and qualitative research in diagnostics development on the other hand, addressing challenges such as those related to the context or small sample sizes. The second, qualitative, study sought to address methodological issues of clinical trials for CL interventions. Semi-structured interviews were conducted with 74 CL patients from seven endemic countries about their disease experiences, and analysed for their preferred outcomes. Patients reported a number of treatment outcomes as important; the majority had only been insufficiently or not included in trials so far. Recommendations made include consideration of a group of outcomes into future trials *via* suitable outcome measures, and suggest information and/or psychological support for patients to address their fears related to other, undesired ones.

The thesis shows a role for qualitative research to enrich clinical studies for skin NTDs. It emphasizes the importance of consulting appropriate stakeholders *via* suitable methods in order to make trials more relevant and suitable for the context and the intended target population. It also identifies areas that could be addressed with targeted capacity development or further operational research, and seeks to facilitate the conduct of pragmatic trials in low-resource areas.

Acknowledgements

This DPhil journey took me through a number of countries, meeting inspiring people on the way, some of which became my friends.

I would like to thank my supervisors Professors Trudie Lang, Kevin Marsh and Piero Olliaro, who led and instructed me, each in their own way. I feel very privileged to have been able to work with such accomplished researchers and personalities in their field. I would also like to express my gratitude to the past and present team members of The Global Health Network, and to John Minogue and Paul Hogben at the Division of Tropical Medicine and Global Health. I am also greatly indebted to all my collaborators for both projects, in particular Dr. Anthony Ablordey, all CL patients and stakeholders who were interviewed, as well as Dr. Emma Plugge for her ongoing support and advice on qualitative research, Dr. Ben Lambert for modelling, and Abdul Masoudi and Dr. Christine Halleux for help with the CL project.

Oxford can be a confusing place to be at times, but wonderful and strange things await! I would like to mention a very special group of people, in no particular order –Ally Siddharth, Ghia, Gina, Fahd, Elisa, Maria, Thanasis, Nicoletta, Marc, Moritz, Felix, Ninni, Vasiliki, Stefan, Hakan, Amalee, Amélie, Inga and Klemens. Sharing thoughts, meals, a lot of laughter, but never the same opinion (of course!); sharing moments such as the eerie beauty of Port Meadow in the moonlight, and actually looking forward to rowing at 6 freezing o'clock on a winter morning. Maggie's living room proved a great, but imperfect, refuge from approximately four thousand and five hundred ways of craziness. The combination of all that will be greatly missed.

A big thank you goes out to my family, and in particular to my parents, without which this endeavour would not have been possible and who welcomed me back home for writing up during a scorching hot summer, and my former boss Dr. Faduma Issa who guided me during a time where nothing seemed possible. Last, but not least, I would also like to thank Sebastian for appearing in my life at the most confusing time - and staying.

If I could give one piece of advice to a dreamy and bespectacled seven-year old girl fascinated by her microscope, then I would perhaps tell her one thing: To never feel obliged to justify why she loves worms.

Table of contents

1. Introduction: Trials in low-resource settings

This research work aims at contributing to improved clinical trials design for neglected tropical diseases (NTDs) that take into account public health objectives as well as the appropriate operational considerations for trials to be conducted in resource-constrained settings in low- and middle-income countries (LMICs). It is expected that by standardising trial design and ensuring it is adapted both to the patients' and programme needs as well as the conditions in which it is conducted will generate better knowledge, better tools, evidence-informed guidelines, and ultimately better health outcomes.

Two case-scenarios have been studied: trials evaluating diagnostic tests and trials testing therapeutic interventions for tropical skin diseases: Buruli Ulcer (BU) and Cutaneous Leishmaniasis (CL), respectively. Their common aim is to suggest novel and more suitable approaches, and to identify problem areas that could be addressed with specific and targeted capacity development, or further operational research.

Despite recent wider recognition of and increased funding towards NTDs, it could be argued that the '10/90 gap' still exists. It was defined in the early 2000s and describes the phenomenon whereby 10% of global health research expenditure is allocated to diseases primarily affecting 90% of the world's population (Davey & Global Forum for Health Research (Organization), 2004). Awareness of NTDs has increased with a World Health Organization (WHO) 2020 roadmap (WHO/NTD, 2012) and the subsequent London Declaration (WHO, 2012), a collaborative programme launched in 2012 with the aim of controlling, and eliminating where possible, 10 NTDs by 2020. The London Declaration was endorsed by a number of donors, commercial partners, national developmental programmes, product development partnerships (PDPs) and non-governmental organizations (NGOs). Several product development partnerships, such as Drugs for Neglected Diseases *initiative* (DND*i*), the Foundation for Innovative New Diagnostics (FIND) or the Program for Appropriate Technology in Health (PATH) are specifically dedicated to research interventions and diagnostics for NTDs, independent of business considerations traditionally driving commercial development of drugs or diagnostic tests.

In addition to health research expenditure, the question of where the research is conducted, and who leads it and does it, is essential: Local empowerment and stewardship is more likely to address the context-specific challenges, more likely to influence local policy and practice, and is therefore more sustainable (Costello & Zumla, 2000; Franzen et al., 2013; Lang & Siribaddana, 2012; Lang et al., 2010; J. Page et al., 2003). In order for this to happen, local capacities must be strengthened. Despite the increasing number of trials being conducted in LMICs, overall clinical research capacity in LMICs remains insufficient (Aksoy, 2010; Lang et al., 2010). A study among clinical researchers in Ethiopia explored potential reasons, and found that researchers reported a great number of barriers on an individual, operational, organisational and system level, such as limited funding, learning opportunities, human and material capacity, as well as poor incentives. Additional barriers on a personal level include lack of awareness, confidence and motivation (Franzen, 2014; Franzen et al., 2013).

The central assumption and emphasis behind this work, connecting the two research strands, is that qualitative research methodology can be used to meaningfully complement quantitative methodology in research studies in NTDs. Qualitative research is suitable to address questions surrounding research hypotheses and instruments, challenges with trial design, recruitment, data analysis and reporting. It could also help to explain findings, and to understand how interventions can be implemented, to assess acceptability of interventions to service users and health care professionals (HCPs) (Rapport et al., 2013).

The work seeks to facilitate the conduct of pragmatic trials relevant for the setting in low-resource areas by consulting stakeholders and by including contextual information and already existing data. It takes two NTDs affecting the skin as examples. At the same time, the methodology developed in both research strands is not disease-specific, and could be used for a wide range of NTDs.

The aspects found to be critical and that I sought to address with this research are:

- The comprehensiveness of diagnostic research and its poor methodological standardization (Knottnerus, Buntinx, & van Weel, 2009; Knottnerus, van Weel, & Muris, 2002), in particular during prospective assessment of a diagnostic test in the designated patient

population. For this work, a study assessing the performance and suitability of a diagnostic test for BU for use at the point of care (PoC) was chosen, with first, a number of relevant contextual factors and challenges (such as access issues due to remote settings, a lengthy and poorly defined diagnostic pathway, availability of suitable equipment, training of HCPs, and environmental factors, for example elevated temperature or poor electricity supply), and second, a great amount of data and local informal expert and stakeholder estimates and opinions available, that could meaningfully contribute to the trial.

- The lack of patient consultation and poorly standardized methodologies of clinical trials for NTD interventions. For this work, CL was chosen as an example for a skin NTD, addressing issues related to trials identified in a number of previous publications, (González et al., 2009; González, Pinart, Reveiz, & Alvar, 2008; González et al., 2010; Monge-Maillo & López-Vélez, 2013; Olliaro et al., 2013; Reveiz, Maia-Elkhoury, Nicholls, Sierra Romero, & Yadon, 2013), such as the lack of patients' opinions on outcomes, preferred modes of administration and other critical issues, as well as the lack of consensus among HCPs and researchers on outcomes, outcome measures and eligibility criteria.

1.1. Aims and objectives

Aims: To incorporate qualitative methodology research into clinical studies for interventions and diagnostic tests for skin NTDs conducted in low-resource settings, in order to identify key factors that could improve research design, conduct and relevance to their target population.

Two case studies:

- to explore the design, conduct and context of an evaluation trial of a diagnostic test for BU in a low-resource setting.

- to determine patient-preferred outcomes for clinical trials for CL.

Objectives:

- To explore the process and context of a diagnostic test evaluation trial for BU
    - to evaluate performance of a diagnostic test for BU under field conditions including historical data and expert estimates, and
    - to explore contextual factors, the process and challenges, in order to enhance the overall study design.
- To consult patients on relevant outcomes for CL clinical trials
    - to capture disease experiences of CL patients in endemic countries by conducting semi-structured interviews, and
    - to identify outcomes for intervention trials in CL that are important from a patient's perspective through analysis of the interviews.

1.2. Methodological approach

In the following, I will discuss qualitative and quantitative research approaches in general, alongside of their associated (post-)positivistic and constructivist paradigms, and related topics. The discussion aims at clarifying the terminology, and situating this work in the discourse in a pragmatic manner by making the philosophical worldview relevant for this research explicit.

In contrast to a quantitative research paradigm, qualitative research differs from the former in the nature of data used, the methods for sampling, data collection and analysis (methodology), the underlying ontology (asking about the 'nature of reality') and epistemology (the nature of the relationship between the inquirer and the known, or knowable) (Guba, 1990).

Traditionally, clinical studies are rooted in a (post-) positivistic tradition of quantitative research, historically understood as the scientific approach, or *normal science*, a term shaped by Kuhn (1970). Quantitative research, thus strongly associated with positivism, tests pre-defined (*a priori* defined)

hypotheses using empirical, numerical data, and subsequent decision making based on the results, in disciplines such as clinical epidemiology. The approach is seen as confirmatory or 'top-down'. Data being collected is thus quantitative, with measurement as precise as possible, using structured and validated instruments for data collection, with the researcher being as removed as possible. Analysis tries to find statistical relationships between variables. Sampling strategies are therefore aiming at statistical representativeness for the overall population from which the sample is drawn: Findings are aimed at being generalizable, providing representation of an objective outsider viewpoint. Ontologically, the reality, or truth, which is sought, is seen as objective, material, and agreed-upon. Human thoughts and behaviour are considered regular and predictable. Quantitative description, causal explanation and prediction are among the research objectives (Johnson & Christensen, 2010; Neuman, 2013). While a purely positivist metaphysical stance would argue that reality is absolutely accessible, postpositivism, developed by Thomas Kuhn (1970), Sir Karl Popper (1935) and others over the course of the twentieth century as criticism of logical positivism, argue that reality is often accessible only imperfectly and probabilistically. Verification of *a priori* defined hypotheses is not possible; they can only be subject to falsification. Postpositivism could be considered as the basis of contemporary scientific thinking.

Instead of focussing on numerical data and the confirmatory scientific method, qualitative research allows for exploring and explaining data. It is often used when little is known about a topic or phenomenon, to explore and learn more about it in order to gain a better understanding. Data being collected is qualitative in nature, such as words or images, and are collected *via* instruments such as in-depth interviews or open ended questions, participant observation, and field notes. The researcher is the primary data-collection instrument, and subjectivity embraced and made as explicit as possible. Analysis uses searches for patterns, themes, and holistic features; and appreciates difference/variation through contrast and comparison. Findings are context-specific and provide insider viewpoints. This results in a concept of ontology that is subjective, mental, personal, and constructed. Human thoughts and behaviour are consequently seen as situational, social, contextual, personal, and unpredictable. Research objectives of qualitative research are therefore subjective description, empathetic understanding, and exploration (Johnson & Christensen, 2010). Qualitative research typically uses nonprobability samples, not aiming at representativeness of the sample for

the overall population (Neuman, 2013).Constructivism is typically associated with a purist qualitative research approach, and its proponents feel that positivism and postpositivism are severely flawed and must be entirely replaced (Guba, 1990; Guba & Lincoln, 1994). They argue that reality is not absolute, but exists only in the context of a mental framework, a construct, for thinking about it. No unequivocal explanation for a given phenomenon is ever possible, as there can be many constructions to arrive at a given, observable reality, without a foundational way to choose among them. Hence, reality can only be seen through a window of theory, as well as a window of values. Another point of criticism is the interactive nature of the inquirer-inquired-into dyad, or the notion of objectivity of enquiry that does not exist, a notion *'devastating to both positivism and postpositivism. (…) Knowledge is seen as a human construction, never certifiable as ultimately true, but problematic and ever changing. (…) Realities are multiple, and they exist in people's minds.(…)'* (Guba, 1990, p. 26). Constructivism is therefore ontologically relativistic and epistemologically subjective. Proponents of constructivism see this paradigm as fully incommensurable to postpositivism. One of the paradigm worldviews considered suitable for studies mixing quantitative and qualitative methodology by Creswell and Plano Clark (2011) is pragmatism, addressing the strong dichotomy between postpositivism and constructivism. Pragmatism is focusing on 'what works' and is mainly considering the practical effects. Originating with philosophers of the late 19th century, such as Peirce, James, Mead and Dewey, it was formally linked to mixed methods research in the early 2000s, and is currently a paradigm taken up by many mixed methods researchers. For pragmatism, the research question is of primary importance, rather than the method or the philosophical worldview, and all approaches available to understand the problem are being used. In addition to the methodology, ontology and epistemology are flexible: Pragmatism posits an external world independent of the mind, as well as that lodged in the mind, and truth is not based on a duality between reality independent of the mind or within the mind (Cherryholmes, 1992; Creswell, 2013; Creswell & Plano Clark, 2011).

In mixed methods research, Creswell and Plano Clark suggest that more than one paradigm/worldview may be used (Creswell & Plano Clark, 2011), either according to or independent of the phases, and study design. In the latter approach, methods guide procedures, e.g. first using quantitative methods, such as surveys and experiments which are rather linked to a

postpositivist worldview, and framing the project within an *a priori* theory being tested. Subsequently, in an e.g. more exploratory phase using focus group discussions (FGDs), the worldview shifts to constructivism guiding the research.

I considered, for the purpose of this work and outside of it, that many of the points brought up by constructivists against (post-)positivism (Guba, 1990; Guba & Lincoln, 1994) have been addressed already, quite a number in the transition between positivism and postpositivism (Kuhn, 1970; Popper, 1935), and others are restricted to problems in the social sciences, being of very limited applicability in the medical sciences. A healthy scientific approach takes into account alternative explanations, and does provide tools to choose among them in order to arrive at recommendations. For the purpose of this work, and hoping to inform evidence-based medicine, the postpositivist mindset of the researcher can't be discarded, and is seen as central. For the application of both qualitative and quantitative methods in the context of this work, a pragmatist worldview is seen as appropriate.

2. Diagnostics evaluation studies in low-resource settings: Case study of a point-of-care (PoC) test for diagnosis of Buruli Ulcer (BU)

This chapter describes a study evaluating a diagnostic test for BU over a certain period of time, as well as an embedded research methodology study addressing the design, conduct and context of the test evaluation.

2.1. Background and literature review

2.1.1. Evaluation trials of diagnostic tests

Diagnostic tests of high quality and accuracy play an essential role in patient management and infectious disease control. Diagnostics are an indispensable element of case management and public health, and as such help identify who is affected by a certain condition, who should be treated or not, and what the treatment outcome was. Therefore, new diagnostic tests must be thoroughly evaluated before being introduced in clinical practice. However, the process whereby diagnostic tests are developed, approved and adopted is both cumbersome and inadequately standardised and regulated. As a consequence, often times health systems especially in LMICs have to select tests based on insufficient and/or inappropriate information provided by the manufacturer, or published data originating from diagnostic accuracy studies that may be poorly designed and reported, and inadequately powered (Lijmer, Leeflang, & Bossuyt, 2009; Peeling, Smith, & Bossuyt, 2008).

Importantly, crucial information about the test performance for the intended use case (Knottnerus et al., 2009, 2002) may not be available, including for instance whether the test is to be used for detecting or excluding a condition, contributing to the decision-making process as to which treatment should be administered, assessing prognosis, or monitoring the clinical course of a disorder and treatment outcome (test of cure). Test performance includes a range of attributes, ranging from analytical and clinical sensitivity and specificity, test characteristics and procedures, and their consequences in terms of ease or burden to the patient (invasive or non-invasive procedure) and health providers, and costs. There is a choice between qualitative (e.g. mapping the

structure of the decision-making process) and quantitative (e.g. assessment of test discrimination and the probability of the clinical outcome) approaches to address these issues, or a combination of both.

In a systematic review of 19 published schemes for phased evaluation of diagnostic tests conducted between 1978 and 2007, Lijmer et al. (2009) find, despite their variability, substantial commonalities between schemes. Each of the schemes consists of four to seven different elements. Two examples are given below:

In analogy to the four-to-five phase hierarchical model of clinical trials for interventions, Sackett and Haynes (2002) and Haynes and You (2009) distinguish 5 distinct phases of evaluation studies of diagnostic tests, according to the types of questions being asked during that phase. Phase I and II are looking at questions that are answered with the range of test results. This includes *Do test results in patients with the target disorder differ from those in normal people?* and *Are patients with certain test results more likely to have the target disorder than patients with other test results?*, that is, sensitivity and specificity of a test under investigation, as compared to the known disease status of patients. Both phases therefore use a preselected group of confirmed patients, or diagnosis under 'ideal circumstances'. Phases III – V focus on a broader perspective: Phase III is looking at the question '*Among patients in whom it is clinically sensible to suspect the target disorder, does the test results distinguish those with and without the target disorder?'*, adding diagnosis based on clinical criteria. Phase IV is looking at the patients' health outcomes, and Phase V at a combination of health outcomes and costs, adding a health technology assessment (HTA) component. Using a different approach originally proposed for cancer biomarker research, Pepe (2004) distinguishes also 5 phases: Phase 1 uses exploratory investigations to identify promising tests and settings for application. Phases 2 and 3 use retrospective validation and retrospective refinement, typically within a case-control study design, where the disease status of the patient has already been determined. In contrast, a Phase 4 study is a prospective cohort study, where the subjects' disease status is generally unknown: Subjects are enrolled into a cohort study and evaluated for diseases using both the reference standard test or *reference test (*ideally, the *gold standard*, that is, the best available diagnostic test under reasonable conditions), and the diagnostic test under investigation, or *index test*. The key objectives in this phase are determination

of the test's operating characteristics when the test is applied prospectively in a defined population for the first time. Studies then evaluating the overall test impact on the population are called Phase 5 studies.

The reason for this variability of proposals for the phased development of diagnostic tests may be again the absence of a unified regulatory framework, with a lack of international standards and little agreement on the evidence required with regards to diagnostic tests (Lijmer et al., 2009; Price & Christenson, 2008; Walley, 2008). Knottnerus and Buntinx (2009) have pointed out that, in contrast to the methodology for randomized controlled trials (RCTs) on effectiveness for therapeutic interventions, or for etiologic study designs, the methodology for diagnostic test evaluations is not yet as established, or 'crystallized'. A clear challenge is the number of elements that diagnostic research comprises (such as in terms of contextual factors, treatment options including 'watchful waiting' strategies, the embeddedness in a workflow of therapeutic management), and some authors argue that diagnostic research, for these reasons, is more complex than treatment research (Knottnerus et al., 2009).

Significant steps towards the methodological standardization of diagnostics studies have been made with the work on diagnostic studies architecture (Sackett & Haynes, 2002), the Standards for Reporting of Diagnostic Accuracy studies (STARD) (Bossuyt, Cohen, Gatsonis, Korevaar, & STARD group, 2016; Bossuyt et al., 2003) and QUADAS (Whiting et al., 2011, 2006), a tool aiming at the Quality Assessment of Diagnostic Accuracy Studies that are to be included in systematic reviews.

From the practical perspective of a researcher developing a new diagnostic tool, there is a lack of guidance and/or methodological standardization with regard to diagnostics evaluation trials in low-resource settings, as well as a lack of published examples in operational approaches in running these trials (Banoo et al., 2006; Peeling et al., 2008; P. G. Smith & Morrow, 1996). This is true in particular if they are performed for the first time prospectively in the designated patient population. The ASSURED criteria (Peeling, Holmes, Mabey, & Ronald, 2006) have been specifically developed by the World Health Organization Sexually Transmitted Diseases Diagnostics Initiative (SDI). They serve as a benchmark to decide if diagnostic tests address disease control needs, and

have been translated into a scorecard to assess and compare tests. ASSURED stands for Affordable, Sensitive, Specific, User-friendly (simple to perform in a few steps with minimal training), Rapid and robust (results available in less than 30 min), Equipment-free and Deliverable to end-users.

The performance of diagnostic tests in different settings can be influenced by many variables. These include population or pathogen characteristics (prevalence, genetic variation in the host or in the pathogen), as well as operating characteristics, and local diagnostic practice and skills, such as how the test is being used and embedded in a diagnostic work flow (Banoo et al., 2006). The assumption underlying this research is that exploring these variables, their interdependencies and correlations with the test's performance, and making them explicit, will help with a better understanding and reporting of the performance and operational characteristics of a diagnostic assay (Banoo et al., 2006; Bossuyt et al., 2003). These variables are rarely considered in the design of diagnostic evaluation studies.

To address and investigate the above-mentioned issues, I set out to conduct a research methodology study, or Study Within A Trial (SWAT) (Clarke, 2012; Clarke, Savage, Maguire, & McAneney, 2015; The Global Health Network Editorial Team, 2015; Van Loggerenberg, 2017), embedded in, and following, a planned evaluation study of a diagnostic test at the PoC (which would therefore correspond to a Phase III study or later, as defined above (Haynes & You, 2009; Sackett & Haynes, 2002), or a Phase 4 study using the definition of Pepe (2004)). Furthermore, and in line with the suggestions made by Knottnerus et al. (2009, 2002) I propose that combining quantitative with qualitative data, is a suitable and necessary approach in this context. It will allow for exploring rich data (in particular with regards to the diagnostic workflow, the context and challenges), and triangulating them with empirical, quantitative, data on test performance. In the broader sense, it could also be considered an implementation research study, as discussed in Peters et al. (2013), with the intent to understand what, why, and how an implementation could work in a real world setting, and to test approaches to improve it. Context (including for example the social, cultural, economic, political and physical environment, including institutional settings and their stakeholders and interactions) plays a central role in implementation research, with an important

implication being that various actors should be involved in the identification, design and conduct of research, instead of only being targets for dissemination of study results. Outcome variables for research could e.g. include acceptability, adoption, appropriateness, feasibility, implementation cost, coverage, and sustainability. A feature of implementation research is the wide range of qualitative and quantitative methods being used, with a focus on participatory research methods.

A SWAT(Clarke, 2012; Clarke et al., 2015; The Global Health Network Editorial Team, 2015; Van Loggerenberg, 2017) is commonly a smaller study being embedded in a larger trial looking at specific methodological aspects of trials. SWATs are increasingly used in RCTs testing treatments and complex interventions, either addressing specific questions essential to the trial, or looking at it from a broader viewpoint. Examples for SWATs include e.g. a study testing whether site visits in trials are beneficial for RCT recruitment (V. Smith, Clarke, Begley, & Devane, 2015; V. Smith et al., 2013), and, as a follow-up, whether it matters who conducts site visits (Nollett, Kelson, & Hood, 2016), or a research methodology study to map initiation and operation of an RCT for podoconiosis treatment in Northern Ethiopia (Lang et al., 2013).

This particular SWAT was embedded (as described in section 2.2.1 Study design) in a prospective evaluation trial of a DNA-detection based diagnostic assay for BU, performed under the leadership of Dr. Anthony Ablordey of the Noguchi Memorial Institute for Medical Research (NMIMR), University of Ghana, Legon, Ghana. The trial tested the diagnostic assay at the PoC in a health clinic in Pakro, a rural area of the Eastern Region of Ghana, with the reference test conducted at the laboratory facilities of NMIMR.

2.1.2. BU: Pathology, epidemiology and treatment

BU is the most common mycobacterial disease in humans after tuberculosis and leprosy. It is caused by *Mycobacterium ulcerans*, which has evolved from the aquatic dwelling *M. marinum* by acquisition of a virulence plasmid pMUM enabling it to produce a macrolide toxin termed mycolactone (ML). ML has cytotoxic and immunosuppressive activities (Junghanss, Johnson, & Pluschke, 2014; Röltgen & Pluschke, 2015). A genomic analysis of human *M. ulcerans* isolates has

indicated a differentiation into two principal lineages (Röltgen & Pluschke, 2015), with a designated *ancestral* lineage (closer related to *M. marinum*) containing strains of Asian and American origin, and a *classical* lineage, with strains from Africa, Australia and Papua New Guinea.

The mode of transmission is unclear, but distribution patterns of BU cases vary between lineages: Those associated with the *classical* lineage occur in highly endemic foci, whereas those in regions with *ancestral* lineage are distributed in a scattered and sporadic manner. Röltgen and Pluschke (2015) point out that the latter may be associated with an ecological niche where humans would only occasionally be infected, whereas the former could be present in an environmental reservoir associated with a higher risk for transmission to humans. Supported by an analysis of the risk factors involved, and evidence for the ability of *M. ulcerans* to persist in the environment as well as in a range of animal hosts, the authors indicate that multiple modes of transmission might be possible. Their conceptual model includes environmental as well as animal reservoirs, and infection *via* insect vectors (aquatic or airborne), or wounds (Jacobsen & Padgett, 2010; Röltgen & Pluschke, 2015). The prevalence of the disease is unknown, with the cumulative number of BU cases in the African region in 2008 estimated to be 60.000, with likely under-reporting. Approximately 75% of the cases are children below the age of 15 (Junghanss et al., 2014).

M. ulcerans mostly causes skin lesions, which contain clusters of extracellular bacilli. Pathogenesis is mediated by ML, destroying the host tissue by causing necrosis and apoptosis, and strong local immunosuppression being mediated by inflammatory cells being killed within the lesion before they reach its active centre, as well as downregulation of systemic innate and adaptive immune responses (Hong et al., 2008; Junghanss et al., 2014). Frequent seroconversion observed in endemic foci suggests self-healing of minor infections in early stages. BU manifests itself as papules, nodules, plaques, oedemas or ulcers that can exceed 15 cm diameter, with a variety of differential diagnoses depending on the disease stage, including parasitic, bacterial and mycotic infections (Junghanss et al., 2014). Large-scale ulceration typically extends beyond the infected areas, and late-stage complications are osteitis/osteomyelitis and disseminated disease.

Various classifications of BU have been suggested to serve different purposes. As a basis for treatment recommendations, WHO defines three categories of lesions (Junghanss et al., 2014;

World Health Organization. Global Buruli Ulcer Initiative, 2004), depending upon the diameter of the lesion and surrounding indurated areas as defined by palpation: category I is a single lesion ≤ 5 cm category II is a single lesion between 5 and 15 cm; category III is a single lesion of more than 15 cm in diameter, or lesions at critical sites (eye, breast and genitalia), as well as osteomyelitis. An earlier classification (Asiedu, Scherpbier, & Raviglione, 2000) divided lesions into non-ulcerative, ulcerative and post-ulcerative (scars with and without sequelae) disease, whereas a descriptive morphological classification along dermatologically well-defined categories distinguishes between papules, nodules, plaques, oedema and ulcers (Junghanss et al., 2014; WHO/NTD, 2011). While early skin manifestations (papules, nodules and plaques) are very unspecific, later stages (ulcers with characteristically undermined edges and induration extending into surrounding skin which may be oedematous, white lesion floor with a cotton-wool-like appearance) are comparably easy to diagnose based on clinical presentation. Diagnostic methods include direct smear examination with staining to detect acid-fast bacilli (AFB), culture, histopathology and polymerase chain reaction (PCR), as discussed in 2.1.3 Diagnosis of BU (Junghanss et al., 2014; Portaels, 2014; Sizaire, Nackers, Comte, & Portaels, 2006). For decades, surgery was used for clinical management. Currently, the WHO recommends a combination of oral rifampicin and injectable streptomycin for 8 weeks (WHO, 2004), and an oral combination of clarithromycin and rifampicin is currently under investigation in a Phase II clinical trial (van der Werf, 2017). Other antimicrobial agents have shown activity on *M. ulcerans in vitro* and are used in clinical practice, although very few RCTs have been performed to date. An overview of treatment options and their evidence base can be found in (Farrar et al., 2014; Sizaire et al., 2006).

2.1.3. Diagnosis of BU

An overview of the currently available laboratory-based tests to confirm the clinical diagnosis of BU is shown in Table 1 (Portaels, 2014).

Table 1: **Sensitivity ranges of diagnostic tests for BU** (Portaels, 2014)

Method	Sensitivity in %
Direct smear examination for AFB (Ziehl-Neelsen staining)	30-80
In vitro culture	20-60
PCR targeting genomic region *IS2404*	> 90
Histopathological examination	90

Sensitivity rates show high variation depending on the type of sample: Direct smear examination and culture have been found to be less frequently positive when samples derived from nodules are used, as compared to samples derived from oedematous forms. If cultured from bone, the positivity rate was reported to be only 20% (Portaels, 2014). In clinical practice, diagnosis is typically based on clinical aspects and confirmed by molecular methods only if available, and often after initiation of treatment (Junghanss et al., 2014)

Ziehl-Neelsen staining is the only method that is easily implemented in the field. PCR is the quickest and most sensitive tool currently available, but has high technical requirements (cold chain, electric supply and training of laboratory staff) and therefore is not suitable for field implementation (Sizaire et al., 2006). This could be partly overcome by the development of a dry-reagent based PCR (Siegmund et al., 2007, 2005), which, however, still requires thermal cycling, electrophoresis and gel imaging equipment.

The loop mediated isothermal amplification method (LAMP) (Yasuyoshi Mori & Notomi, 2009; Nagamine, Hase, & Notomi, 2002; Nagamine, Watanabe, Ohtsuka, Hase, & Notomi, 2001; Notomi et al., 2000) emerged in the past couple of decades as a sensitive and suitable technique for the diagnosis of a wide variety of diseases (Dhama et al., 2014). LAMP is a modification of the original PCR protocol using a DNA polymerase with strand displacement activity, therefore eliminating the need for thermal cycling. A set of 4-6 specifically designed primers recognize a total of six distinct sequences on the target DNA, conferring high specificity. Within one hour under isothermal reaction conditions (a heat block or a water bath), typically 10^9 copies of the template are generated. The reaction was found to be very robust towards the presence of non-target DNA and contaminants, and sensitive enough to detect less than 10 target copies of the template. The

amplification product can be detected by turbidity, *via* the presence of magnesium pyrophosphate precipitate in solution as a by-product of amplification (Y. Mori, Nagamine, Tomita, & Notomi, 2001), or by fluorescence, e.g. by addition of calcein (Tomita, Mori, Kanda, & Notomi, 2008) (whose fluorescence is quenched due to the binding of manganese ions by pyrophosphate), a fluorescing DNA intercalating dye like SYBR Green, or a colour change in function of the pH change of the reaction using appropriate dyes (Tanner, Zhang, & Evans, 2015). Optimally, endpoint detection would not require any instruments, but should allow for naked eye detection, such as a clear colour change under varying light conditions.

LAMP has potential for being used as a PoC diagnostic test (Njiru, 2012), and meeting the ASSURED criteria (Peeling et al., 2006). The LAMP protocol was recently adapted for the detection of *M. ulcerans* (Ablordey et al., 2012; Beissner et al., 2015; Njiru, Yeboah-Manu, Stinear, & Fyfe, 2012), and a systematic review and meta-analysis of studies evaluating LAMP for detection of *M. ulcerans* was conducted for this research, indicating a sensitivity of 84% and a specificity of 98% as compared to PCR.

2.1.4. Study context and local environment

Pakro Health Centre is run by Ghana Health Service (GHS), a government body under the Ministry of Health in Ghana. It was originally established in 1991 through renovation of an abandoned warehouse, focussing on maternal and child health/family planning and obstetric care (Senah et al., 1997). Pakro Health Centre covers a population of around 9000 inhabitants in Pakro and around 28 surrounding farming villages and hamlets. Many farmers also live in the town and work on farming land in the surrounding area during the day (Figure 1 d). Previously a prosperous cocoa-growing area, severe drought and bush fires in the early 1980s destroyed all cocoa farms (Senah et al., 1997), and most of the villagers now are subsistence farmers growing maize, tomatoes and cassava (for producing garri, fried cassava flakes) for sale on markets.

Pakro is located in the Akuapim South municipal district in the Eastern Region of Ghana, at around 15 km east from Nsawam, the district capital. The road connecting Pakro to Nsawam is

only partially paved and poor, in particular during the rainy season. Nsawam has a population of around 45 000 inhabitants and a hospital (the Nsawam Government Hospital), and is connected to the capital Accra with a major road, with regular public transport options available. The health centre (Figure 1 a-c) is staffed with a disease control officer, responsible for BU, and other technical staff in a downstairs practice, as well as several nurses, orderlies and a doctor in an upstairs practice. A separate room downstairs, separated from the practice room by wooden panels with a door (Figure 1 b), was intermittently used for experiments. It is now equipped with staining reagents and equipment and a microscope, and being used by a member of technical staff responsible for malaria diagnosis. The room will be described in more detail in 2.2.3.3 Detection of *M. ulcerans* DNA under PoC conditions . Pakro Health Centre is connected to the public electricity grid and without a generator, which means supply can be sporadic and unreliable.

Figure 1: **Pakro Health Centre and its surroundings.** (a) Pakro Health Centre. (b) Room used for experimental work, inside view, (c) View from the entrance of the health centre. (d) Village endemic for BU in the region.

The National BU Control Programme was established in 2000. Its aims are to minimize the morbidity and disability associated with BU disease, with the key interventions being early detection (in the absence of an effective means of preventing BU) and standardized case management, reporting and surveillance. Health systems strengthening and advocacy for BU in the context of NTDs, research interventions and technical support to institutions are also included. Training of CBS volunteers as well as health workers forms an integral part of activities. Due to a lack of financial means, in particular the training and advocacy activities are seen as restricted. In addition, a fire at the Central Medical Stores, Tema, in 2015 destroyed a great amount of drugs as well as information material and other supplies of the programme. Collaboration with NGOs, who provide some additional funds, is being seen as important, and more are being solicited, e.g. to support patients' transport to and from health facilities. Integrated control of BU, yaws and leprosy has now started in 15 districts initially, with upscaling planned later: Field workers are trained to recognize and refer all three skin diseases, with the facilitators moving from one district to another. Ideally, initial training and refresher training twice a year would take place, but is currently not implemented because of financial constraints. The integration is seen as an opportunity to be more visible on the local level, and to encourage case reporting in general, and in an integrated format using the electronic DHIMS (District Health Information Management System).

2.1.5. The diagnostic workflow for BU patients at the research site

This chapter describes the diagnosis and treatment pathway for patients at Pakro Health Centre, based on information provided by the HCPs, CBS volunteers and researchers during interviews and a FGD, as well as field notes.

2.1.5.1. Case finding

Most patients in this rural area are identified and brought by CBS volunteers who are living and working in the villages. For BU, they collaborate with the health centre's disease control officer, who sometimes goes to the villages as well.

> *(…) I can't go to the whole community, so when there is a case at a particular community, they refer the…either the community volunteer comes with the patient, or the patient comes and goes, or the volunteers comes to inform me, we go to the community and go and see that particular disease. (H1)*

Originally, CBS volunteers were recruited and trained for a polio programme, and received additional training to identify and refer BU patients. CBS volunteers are nominated by approaching the community, rather than an individual: The need for a CBS volunteer from a specific village is communicated to the village chief and the group of village elders, and the community identifies the CBS volunteer who subsequently presents at the health centre and receives appropriate training.

Their work consists of communication and sensitization of the population through

a. *Debes* (singular *adeba*), which are informative presentations in front of smaller group gatherings. Monthly or biweekly debes with smaller groups of people are announced *via* poster and held, often times when the community nurses visit. Debes will focus on one or several specific diseases or health topics, and often include other issues, and people are encouraged (and will) ask questions. In addition, school classes are gathered and skin screenings of the school children take place.

b. Announcements during church services, where the pastor is asked for permission for an announcement.

c. Visits of family homes: One CBS volunteer notes that he usually visits 5 families per day in his scattered farming community. He explains that the optimal visiting time is 6pm, as all family members are then expected to be at home from farming work.

One HCP additionally describes sensitization campaigns with the help of volunteers, together with former BU patients who got healed:

> *Yes, at times we are moving, at times we go, […] we show film shows on BU, and then we do sensitization, so we go with some volunteers who had this BU and were healed. After the film then we talk with them and then we let the volunteers also talk. The experience they had, and how they also had their wounds healed, so they see, they become happy. (H1)*

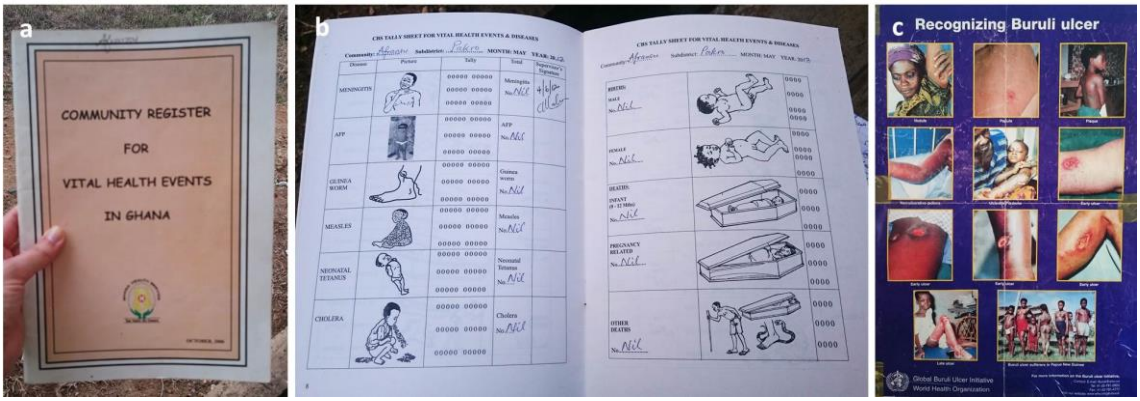


Figure 2: **Materials provided to CBS volunteers**. (a) and (b) *Community register for vital health events in Ghana* for reporting diseases and other health events on a monthly basis to the health centre. (c) Poster.

In addition to BU, CBS volunteers are responsible for identification and referral of patients to the health centre for a range of diseases. Materials that are provided to them are pamphlets and posters Figure 2 c) and the *Community register for vital health events in Ghana* (Figure 2 a, b), where they are required to note all occurrences for a range of diseases and other events (e.g. births and deaths). This information is collected on a monthly basis by the Pakro Health Centre, and recorded. CBS volunteers describe being trained for recognizing febrile illnesses including malaria, coughing (possibly indicating TB, which is being diagnosed in Nsawam district hospital) and diarrhoea, and treating or referring accordingly. For instance, they are equipped with malaria rapid diagnostic tests (RDTs) and medication, and if a child between 6 months and 5 years presents with fever and/or diarrhoea, it will be tested and, if necessary, immediately treated with antimalarials. Patients above that age are referred to the health centre.

> *If the disease is above our knowledge, we also refer. They tell us to just treat malaria, pneumonia, and diarrhoea. So if we see any other disease, we refer the patient to the clinic (C2).*

A family house visit is described as follows by one volunteer: He introduces himself, and asks whether someone has seen potential cases of a range of diseases. He then gives them information and general health advice, such as keeping their house and surroundings clean and free of weeds, covering open water, disposal of refuse and faeces, or using a water bucket to fetch water from the

local river instead of washing directly in the river, thus contaminating the river for the communities downstream. He notes that community members usually follow his advice.

CBS volunteers, together with the disease control officers, are also responsible for the local mass drug administration (MDA) campaigns, aiming at annual distribution of albendazole and ivermectin for chemoprophylaxis of infectious diseases like lymphatic filariasis (elephantiasis) and onchocerciasis (river blindness). The drugs get sent to Pakro Health Centre from the district level, and are then distributed to the CBS volunteers, who take them to the villages and administer them to the population:

> *Yes, we go to the communities when they are doing that mass drug distribution, we visit them to see how the drugs are being distributed. And then whether the people, [...] community members, are taking the drugs. So when you go they tell you they will not take it, then you will have to go and talk with them. (H1)*

One HCP describes skin rashes as common side effects of the drugs and CBS volunteers subsequently being accused by the village population '*that we are rather giving them sickness, the diseases*' (H1). Against the background that these MDA campaigns are aimed at chemoprophylaxis and not treatment after diagnosis, many are refusing the drugs.

The notion of BU as a spiritual, not medical, disease, amongst patients was mentioned frequently by the staff at the centre and the CBS volunteers: BU was described to be witchcraft, or a curse, and not a (medical) disease - thus requiring non-medical treatment:

> *…they would tell you it's [...] witchcraft, or somebody might have done something that is. But we tell them it's not witchcraft, it's real. So some of them, when they get it, they visit the herbalists, the traditionalist, to ask, or to perform certain ritual (..) (laughs)….before they come to the health centre to seek….and by that time the disease, especially BU, might have gone out of control. When it starts as a boil, instead of coming, they'll not come. They'll visit a whole lot of people before they come, then they get a big wound. (H1)*

(Closed lesions were typically described as 'boil'.) This notion of BU as non-medical condition is in line with findings from two qualitative studies, one from Benin (Aujoulat, Johnson, Zinsou, Guédénon, & Portaels, 2003), and one from Ghana (Stienstra, van der Graaf, Asamoa, & van der Werf, 2002). Both studies also pointed out that people were reluctant to visit health facilities, and usually presented late. The caveat is that at the time both studies were conducted, only surgery was available as therapy for BU and poorly accepted. A study conducted in the Ga West district in

Ghana found that 72% of patients sought advice from the herbalist first (Renzaho, Woods, Ackumey, Harvey, & Kotin, 2007). In addition, alternative treatment options were being perceived as being quicker than the 8 weeks of antibiotics given as standard treatment.

Alternative treatments were mentioned to be administered by community members described as herbalists, spiritualists or traditionalists (who are sometimes the same person), as well as by community members who claim to have medical knowledge despite having no medical training, and obtain inappropriate medication for the patients from drugstores. Spiritualists use prayers, and herbalists typically use particular herbs, unknown to the CBS volunteers, that are smeared on to the boil, thus leading to ulceration, or on the ulcer, often aggravating the lesions (described as '*whole arms become rotted*' (C1)). Some herbalists are described to have inherited the knowledge from their parents or grandparents; others use a trial and error method. In addition to monetary payment during treatment, they typically ask for animals and other items for sacrifice purposes, and fowl, as well as a specified sum upon cure.

CBS volunteers and the disease control officer describe specifically seeking BU patients at alternative healers:

> *Normally I visit them to see if they have got some cases. So we have to collaborate with them. [...] They are hiding in some peculiar places, they are not in town. This one where I went, there's this one is at a particular place - if nobody tells you there's somebody doing something there, you will never know. I went to his house, and I was told he has gone to his place where he does his spiritual. I went there, there were some people over there. So I told him: Today this place is busy, so maybe tomorrow or on Wednesday I have to go back and see. So that is how most patients behave here. (H1)*

Often, the healers will try and prevent them from taking 'their' patients to the clinic. If they can't be convinced, the disease control officer notes that he will allow the patient to continue the alternative treatment during the course of medical treatment.

One CBS volunteer describes the case of a lady and her son in Pakro town, both of which had BU in the form of small lesions on the back and the arm, respectively. During a home visit, the lady showed a small painless boil to the volunteer who advised her to go and seek help at the health centre. Both patients presented very late, with more complicated lesion: It turned out that in the meantime they had seen a herbalist for treatment. Both were healed with scars, and when the lady

thanked the CBS volunteer, she mentioned that if she got it a second time, she would go straight to the clinic.

Another CBS volunteer gave the example of a woman without any medical training who said she could help a patient, asking for money and fowl. The fowl's fat was used for treating the lesions, along with penicillin and flagyl from the chemist, dissolved in oil. The patient, a woman, was finally treated, after intervention of the disease control officer. He convinced the healer to 'release' the patient, under the condition that she would be brought back if the medical treatment was unsuccessful. She was treated and healed, but her very large lesion required her to be referred to Korlebu (Teaching Hospital in the capital, Accra).

A third CBS volunteer remembered 2 patients, children, whom he had found at the herbalist and brought to the clinic. The herbalist was very angry at him due to the loss of income, causing an ongoing conflict. The herbalist also developed BU, but refused to come to health centre, and his lesions subsequently became very severe. He left and never returned to the village again.

One HCP told of two men who were diagnosed with diabetes and BU at Pakro, and referred to Nsawam hospital for treatment of their diabetes:

> *They can be treated (for diabetes). But both, they didn't go. They were complaining about financial problems, [...] and they both went to a herbalist. A month or so (later) we heard their condition deteriorated, and they were rushed to Nsawam and they couldn't survive, and they died. We were saying the money you went and used for the herbalist, you could have taken the money to Nsawam. If you go (to the herbalist) [...]they will charge you a whole lot of money. They take it in bits, they don't take it as a huge amount. If it is a huge amount, they will tell you: Come and put something down, so that they will start with the treatment. (H1)*

The HCPs agreed that case numbers have been going down over the recent years, and the lesions being seen at Pakro Health Centre from local villages are now less severe than they were in the past.

In summary, the collaboration between the disease control officer, the HCPs and the CBS volunteers plays an essential role in case finding and referral. CBS volunteers are based in the villages and trained for a number of tasks related to a range of diseases. The population is sensitized *via* debes, announcements during church services, and visits in school classes and the family homes, where cases can also be identified.

Even after diagnosis, the notion of BU as a curse or witchcraft, and the long treatment duration often leads patient to seek alternative (herbal, spiritual or alternative medical) treatments first. This is in particular against the background of a strong belief in witchcraft in the population, as experienced by the researcher, and as of yet unclear mode of transmission of BU (Merritt et al., 2010). Alternative treatments were identified as a major cause of delay of treatment onset, often leading to ulceration of previously closed lesions in the first place, or badly exacerbating existing lesions. The associated costs put an additional burden on impoverished patients. Often BU cases are sought specifically at alternative healers, and participants point out the importance of collaboration with them, as far as possible, in order to ensure treatment of patients. Sometimes, a conflict of interest seems unavoidable. Sensitization and education of the population about BU as a medical condition was considered as a successful strategy in this context.

2.1.5.2. Diagnosis

When patients present at the clinic, usually on Thursdays which are clinic days, HCPs ask about patients' disease history and take samples for BU diagnosis, either fine needle aspirates (FNAs) in case of closed lesions (edema), or swabs in case of open, ulcerated wounds.

Samples are being taken from Pakro to the tro-tro station Nsawam, usually on a motorbike or by shared taxis. Tro-tros are privately owned minibuses, usually leaving when they are filled to capacity. They are a very common and affordable form of public transport in Ghana, within and between towns. They are used to transport the samples onward to Madina in Accra, their terminal station, which is close to Legon, where NMIMR is. The total distance covered is around 55 km:

> *When we take the specimen, we have forms by the National BU Control Programme, that's the laboratory form. We fill 2 of them [...] and you add one to the specimen and keep the other one here. We send it, XXX takes the swabs and fills the forms, and sends them to the station, then gives them to the drivers. And we put them in a vaccine carrier with some little ice on it. Not proper ice, but it's chilled, chilled in water, so it keeps the temperature at least (…). It's very warm in here, hot, so just to keep the temperature a little down. And we call Noguchi, so they go to the station to meet the drivers and collect the samples from them. [...]That goes to Madina, very close to Legon, you know Madina. So they come to Madina, we give the numbers of the vehicles, the driver's name, the driver's number, and then they call them, meet them there, and you pay some token to them, the transport, so that's what they do. (H2)*

The form mentioned above corresponds to the *Request for laboratory confirmation of a Buruli Ulcer case form* (Portaels, 2014, p. 99) that is completed and submitted with the sample.

Since 2013, WHO recommends to national control programmes a strengthening of laboratory confirmation of cases in order to ensure that 70% of reported cases are laboratory-confirmed by positive PCR (Portaels, 2014). Usually, the sample would be sent to the district office, and *via* the regional office then onwards to one of the three reference laboratories in Ghana, of which Dr. Ablordey's lab at the NMIMR is one.

The time delay until it gets to the regional office was quoted as a week at least, but is possibly longer. In the case of Pakro, which is at around 55 km from NMIMR, GHS has accepted that samples are being sent directly to the laboratory in order to shorten the time to diagnosis; however, the costs have to be covered by the investigators or HCPs. This HCP sums up the situation with regards to sample logistics and transport options for patients, as they often end up paying for patients' transport back to their homes:

> *But in more remote places then the timeline would just be much much much longer. [...] And that is […] we have patients coming here, cause they hear of this place, and then we go and then they come. And we have a challenge of keeping these patients, cause sometimes they don't have the money [...]. After some time, they find it difficult to come, and getting them the transport too becomes a problem for us [...], it will come from our pockets. (H2)*

Once the samples arrive at the laboratory, they are analysed using PCR targeting *IS2404*, as described (Ablordey et al., 2012). The protocol uses an overnight incubation step for DNA extraction, following the modified Boom method specific for mycobacteria (Durnez et al., 2009).

> *Yes, we do microscopy, we do culture, but we lay emphasis on the PCR because it's the most sensitive. So PCR [...] is mostly done. [...] Culture takes about 6 – 8 weeks for growth, yes, so when it's good then we check whether it's PCR positive or negative. So the quickest at that level uses microscropy and then PCR. Then miscroscopy, culture, are not very sensitive as compared to PCR. [...] We can do microscopy, but we don't normally send the results unless….. PCR result is ready. (I1)*

The minimum time required for confirmation of diagnosis (from sending off the samples) is within around 48 hours, the typical time within a week, and the maximum time around 2 weeks. It was pointed out that BU diagnosis is incorporated in the group's other research activities, and depending on project resources. The diagnosis is confirmed to the HCP *via* e-mail or text message, and then the HCP informs the patient *via* the mobile phone number given (which could be a relative's or neighbour's number, or the number of a CBS volunteer). One HCP mentions that it

then takes on average 3 days to one week until patients present at the clinic for treatment, due to time and financial constraints.

In summary, upon arrival of patients, usually during clinic days, patients' history is being examined, and clinical diagnosis is made. Samples (either FNAs or swabs) are being taken and sent to NMIMR directly for analysis, together with the relevant forms requesting laboratory examination. Diagnosis is being done using PCR after overnight DNA extraction using a method specifically for mycobacteria. Results are communicated to Pakro *via* e-mail or text message, and patients are informed by calling them, relatives or neighbours, or the CBS volunteer, and asked to present at Pakro Health Centre for treatment.

2.1.5.3. Treatment and follow-up

Patients are treated with oral rifampicin and streptomycin injections for 8 weeks, usually requiring daily visits. During that time, the wounds are dressed, and often accordingly trained CBS volunteers help with wound dressing. Wounds of patients with a negative diagnosis are also managed appropriately.

BU wounds are dressed daily at the beginning, and if the lesions are getting better, on alternate days. During treatment, some patients have to be encouraged by the CBS volunteers to attend the health centre regularly during treatment, and they sometimes accompany them.

A recurring theme during interviews was the problematic transport situation for patients to and from the facility. One CBS volunteer gives an example: His community is about 3.5 miles from Pakro along the road. It's a scattered community of farmers, and produce is being sold, mostly by women, on the local markets in Nsawam and Adonso during specific market days. During these days, transport options for patients are available: Monday and Thursday are market days in Nsawam, and large public buses (Metro Mass Transit, or MMT) run along the main road in the morning and evening to transport passengers and goods to and from the market. Tuesday, there is a market day in Adonso, and (shared) taxis are usually available. On other days, no public transport options are available, sometimes motorbikes take people between villages: but '*It's just off/on, it's not always that you get it*' (C2). The standard shared taxi fare from the village to Pakro is 2 Ghana Cedis,

corresponding to the price of a meal with meat at a street kitchen, or around 35p. A volunteer describes one patient, his brother's wife, who was successfully cured. '*So now she is fine, she can go to the farming and can do farming work and other things*' whereas previously she could '*do nothing with the disease'* (C2). During her illness, she often had to cover the long distance for her daily visit to the clinic for treatment, as the buses weren't running during that time. Now, a CHPS compound (a Community-based Health and Planning Services clinic staffed with 2 community health nurses) is being established at the village.

> *If there's a health facility closer to the patients' home, the medication will be given to the patients directly, with a note attached to it, so they can be treated daily at the other facility, and will visit Pakro for check-up. After several weeks they come again for refilling them. (H2)*

If patients are immobile due to their old age, staff visits them for wound care and administration of antibiotics:

> *You go to their house and find out why they would not come. Couldn't come. [...] And the very old ones, sometimes they cannot walk, [...] because of the BU. So we'll go to their homes and then do the wound dressing and then give them their treatment, I will let the technical staff take their motorbike, go there, give the treatment, dress their wounds, and then they would come back. (H2)*

Wound healing takes a rather long time, and usually scars remain. One HCP describes caring for scars in order to avoid breakage of the skin.

> *(The time to cure is) depending on the size of the wound. At least around [...] 1 to 2 years. [...] There are scars. After the wounds healed, or the wound closes up, there are scars. So we [...] tell them, either they use some Vaseline or shea butter. [...] Massaging the place, so that it doesn't get breakages, when there should be a breakage, the wound will come again. During the Harmattan, when the whole place is very dry, you'll see people coming back… (H1)*

The Harmattan is a season between November and March where very dry, dusty and cold wind coming from the Sahara desert dominates Ghana's weather.

Patients are being taught exercises for prevention of disability. In case patients already present with disabilities, they improve only very little through exercises.

In summary, patients are treated with oral rifampicin and streptomycin injections for 8 weeks, accompanied by wound dressing, usually requiring daily visits at the beginning. The problematic

transport situation to and from the facility for patients was often mentioned, with possible solutions being treatment at a health facility closer to their home, or additional financial support.

2.1.6. Identified barriers for diagnosis and treatment

These barriers for diagnosis and treatment at Pakro Health Centre were identified by HCPs and researchers during interviews and a FGD.

1. Logistics and transport options: As reported in 2.1.4 Study context and local environment, transport options were limited. Despite BU treatment being free of charge, transport costs had to be covered by patients themselves. They were often unaffordable, preventing patients from presenting at the health centre, or from finishing their treatment. Reimbursement of transport costs at the health centre would enable them to come.

   > *(When patients hear about their diagnosis,) some of them don't feel happy, all they say, they have to take it, and they come. [...] Some, when you call, they refuse, they don't come. [...] Then we do a follow-up, we follow them. [...] We'll go to the communities, we'll get them, and then they'll promise they'll come, but they will not come. Some too, when they come, they don't finish the antibiotic treatment, they are not coming. When you call and you ask, they say financial problem. It's bad. (H1)*

   In case they live very far away from the nearest clinic, patients are being taught to care for their wounds, and are then asked to come weekly for review.

   > *Otherwise you'll need to have a form of reimbursing them. [...] If they know they'll get their money, that very day you'll see them here. But aside that, they have financial challenges, [...] we know where they are coming from. I mean look at this woman you saw. I mean, you can't even ask her to come every day, the lesion and the way she is. Ideally, you want to be treating her every day, but she cannot come. She definitely cannot come. So what we do here, we teach them, we tell them to look at what we are doing, how they should […] manage the lesions, so that on a weekly basis they come for a review. [...] But from the history, if you know they are very far away from the nearest clinic, we let them do it at home. (H2)*

   Despite no financial support available specifically for patients for transport to and from the health facility, possible solutions mentioned are:

   - Establishment of a CHPS compound in the village, so the patient can get treated locally
   - More public transport options

- The establishment of halfway homes, as observed during a visit in Benin:

Halfway home, it's like a hostel. [...] So if they are farmers, they can bring some food stuff, they come and stay close to the facility, so that every morning their wounds can be dressed. They come then they go to the hostel, they have their food, everything, they can live a little bit comfortable. Not necessary trekking home, commuting home, trekking in and out from some miles away, every day go and come, no, so they are around the facility. (They can stay for) at least for a month or whatever, at least so their wounds are better managed. And then they can go and come, even if they want to stay [...] they can. (H2)

- Specific support (such as by NGOs) to provide transport to and from the facility:

There are some patients, I mean, even money to go to the facility they don't have, so just identify facilities that have cases like that, and then they will support them with materials, and [...] for the patients. So that NGO is going directly to help the patients. We will not come in, but we will identify the patients, and they will come in. [...] Yes, they will support the patients to be able to come, some of them are so poor. (H2)

Due to a special arrangement with GHS, samples can be sent directly to NMIMR from Pakro, but costs for transport have to be covered by HCPs or investigators, indicating poorly defined sample logistics. In addition, costs for diagnosis of BU are not covered by the control programme.

Several volunteers mentioned mobility challenges, that is problems reaching families in scattered villages for home visits, in particular during the rainy season. It's only small footpaths connecting the hamlets, and visits usually happen after dark when all family members are back from their farming activities. Raincoats, umbrellas and wellington boots as well as torchlights would be needed. Some volunteers were given bicycles that have now broken down due to high humidity. They are in need of repair.

2. Traditional treatments: As discussed previously, BU is perceived as a non-medical condition, therefore alternative, herbal or spiritual treatment is sought first, delaying the onset of treatment. Sensitization and education are seen as crucial and have already led to a change in attitude.

I would say because of the sensitization, too, people come early. [...] as they perceive it as witchcraft, they don't come, they will not come, they visit a whole lot of herbalists before coming, that is why they are coming late with the bad ulcers. But now… (H1)

3. Time delay between taking of sample and treatment onset: Depending on delays with sample transport, analysis and the time patients take to present at the health centre after diagnosis, the time between sampling and onset of treatment can range from 5 days to several weeks.

4. Funding of the National BU Control Programme: The control programme does not receive regular funding and is soliciting for additional support from stakeholders such as NGOs. Funds would be needed in particular for additional personnel, and for training and advocacy activities.

5. Paradoxical reaction to treatment: An HCP emphasized the importance of explaining to patients that the lesions might enlarge in the initial phase of treatment, before getting better. In the past, patients have stopped treatment due to the paradoxical reaction, as they believed that the treatment didn't work. Informing patients about the possible occurrence of a paradoxical reaction at the beginning of treatment has helped.

6. Factors influencing diagnostic performance of PCR: One HCP describes false negatives in this context, which are cases where, despite a strong clinical suspicion, PCR is negative. However, patients were treated successfully based on this diagnosis on clinical grounds. In some cases, another sample being taken at a later stage then tests positive. The number of these cases is reported as low, but they potentially influence diagnostic performance.

> *But I must add sometimes [...] clinically [...] we are sure it's BU, and the test comes as negative. [...] Those cases we treat them, because we know the test is not 100% [...], so when we are sure that it's BU, and the test is negative, I say go ahead and treat. [...] Just a few, a handful of them. They have all the undermined edges, with the yellow slough and all those things, the smell and all those things, and the test comes as negative. [...] Oh they (react to treatment), the wounds heal. I mean, it heals. [...] When the clinical suspicion is so high, we go ahead and treat, even if the test is negative. [...] (When showing a picture of a typical BU lesion:) Everything clinically is BU, but sometimes you take the swab, it comes as negative. And then you take it again, depending on where you take it, it will come as positive. [...] We would continue to treat it, cause apart from BU nothing else would do this. (H2)*

A study conducted in a very similar patient population has shown that 92.9% of the cases with a BU diagnosis on clinical grounds were later confirmed by PCR (Kotey, 2013), confirming the high sensitivity of clinical diagnosis.

The HCP also noted that there are no differential diagnoses to BU that could be treated with the same combination of antibiotics, so the successful treatment clearly confirms the clinical

diagnosis. On the other hand there are also very severe cases that still test negative, and the HCP mentioned a patient in a village that remained untreated despite very strong clinical suspicion. He is now disabled and unable to walk. However, only cases confirmed as positive in the laboratory are reported as such.

This poor level of agreement between samples was also supported by one of the investigators, who mentioned diagnostically discordant samples taken from several lesions on one patient. Portaels (2014) recommended that sampling may need to be repeated if the results of PCR are negative despite a strong clinical diagnosis, and emphasizes training for HCPs taking samples. This is in line with the recommendations given during the interviews, where it was added that it's important to make sure that swabs are being taken from the undermined edges of the lesion.

On another occasion, feedback of the HCP to the laboratory over negative samples, despite a strong clinical suspicion, allowed for identification and correction of a technical error in the PCR machine.

A HCP noted that often patients don't accept a BU diagnosis on clinical grounds, but instead ask for the sample to be taken and submitted for diagnosis without clinical suspicion. He adds that this might negatively influence the sensitivity of PCR evaluated in the laboratory:

> *Yes, when they report, when the patients report, [...] we take the history, find out how long they've had it and what kind of treatment they have done, how it started. But most of them come in the ulcer stage. So when they come, looking at the lesion, we'll have a fair idea whether it's BU or not, using clinical judgement. But even when we think it's not BU we still take a swab, cause when they come and we don't do anything, we tell them it's not BU, but we dress for them, they find it very difficult to accept [...]. So we take the swab anyway, and then the results will come, we tell them the test is negative, but we still look after it. That's when they accept that they don't have it. [...] Now we virtually have to take everybody's, we swab everybody, just to allay their fear; we assure them they have not BU. [...] The problem now is we are sending those (….) we discussed, even those we think are not BU, we send it to them, so [...] their PCR sensitivity will come down. [...] There was a time we were indicating that this is non-BU, non-BU, but I think now we don't do it. (H2)*

2.2. Methodology

2.2.1. Study design

This study consists of a research methodology study (SWAT) embedded into an evaluation study of a diagnostic test for detection of BU infection. I conducted the literature review and meta-analysis, with support for the Bayesian hierarchical model, wrote the protocol for the SWAT, and, over a period of 6 months in Ghana, conducted and analyzed the experimental work (with training and some support from members of the research group) as well as the interviews and FGD.

Qualitative data from interviews and FGDs were collected within a study with a 'traditional' quantitative design, consisting of the experimental intervention collecting empirical data. The supplemental qualitative information was used to provide a rich picture of the context and the BU diagnostic and treatment pathway at the PoC the study was embedded in, and an overview of the research process. It also facilitated interpretation of quantitative findings and planning of further steps in the research process. Data collection was performed concurrently, that is, the interviews and FGD took place at several time points during the trial. Quantitative data (gained through experiments and analysis of diagnostic test evaluation results) and qualitative data (gained through observations, individual interviews and a FGD) were compared and contrasted. This allowed for topics or problems to be addressed as they arose, and thus for building an iterative process involving both components.

In line with participatory action research (PAR) (Wadsworth, 1998), I as the investigator conducting the SWAT, was working as part of the research team. In general, elements of PAR include

- a reflective, cyclical *research* process;
- *action* as a necessary 'consequence' of research, and research being aware of this change, or transformation, it brings about;
- a *participatory* element. Instead of operating within a traditional separation of researcher (those carrying out the research) – the researched (e.g. those interviewed) – the researched

for (the intended 'target group'), these conceptual categories are strongly overlapping in PAR., with active participation of all in the research process.

This study was not planned as PAR, and therefore might not comply to all criteria as set out by Wadsworth (1998), or later authors. However, the concept of PAR helps to understand, and reflect upon, the process and elements, as discussed further in 2.5 Conclusions.

The original protocol was based upon a single, validated protocol for the BU LAMP to be tested (Ablordey et al., 2012). However, during initial conversations with the PI, an iterative approach emerged. During interviews and discussions,

- (1) The diagnostic and treatment pathway and the research study context were explored.
- (2) Experiments were conducted based on discussions with the research team, with a particular focus on the requirements for an ideal, and technically feasible, diagnostic test for BU in this specific context. Research objectives were set accordingly.
- (3) Interview and FGD participants were given the opportunity to comment on the conduct and results of the experiments, providing feedback.

In the following, the ideas guiding the experimental work will be discussed, including the technological challenges for the BU LAMP and requirements for an ideal diagnostic test. They will be introduced by an overview of the different topic areas or elements of this diagnostic test evaluation study.

2.2.2. Interactions between qualitative and quantitative data

Figure 3 shows the different topic areas as the four main technological challenges which were explored during the interviews and FGDs of the SWAT, or the Bayesian hierarchical model of test accuracy built to address the samples size and imperfect gold standard challenge

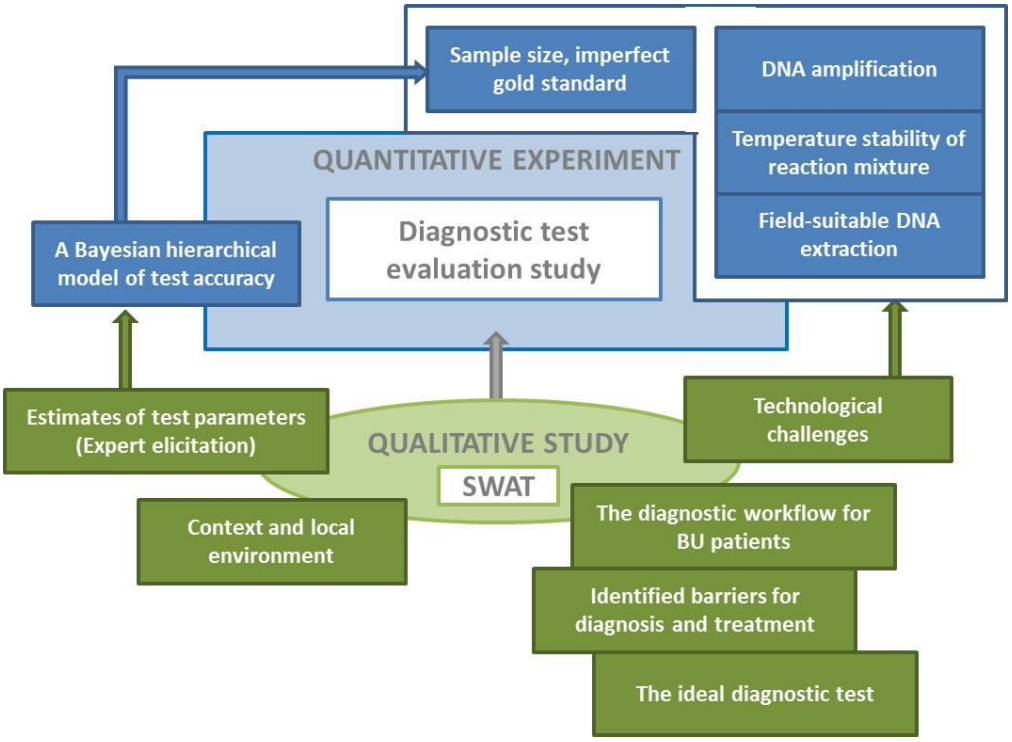


Figure 3: **Topic areas and their types of data included in the study.** Topic areas comprised of qualitative data, obtained during the SWAT, are marked in green. Topic areas comprised of quantitative data include a mathematical model of test accuracy and the experiments conducted during the diagnostic test evaluation study, and are marked in blue. Arrows indicate selected areas of interaction.

2.2.2.1. Technological challenges for the BU LAMP

Four main technological challenges in the current LAMP development were identified during the planning phase of this SWAT, during qualitative data collection (interviews and FGDs), as well as personal communication with the investigators. These challenges were at the same time guiding the quantitative and experimental part of the study, whose results will be presented in 2.3 Results.

a. Limited sample sizes and imperfect gold standard: During the planning phase and initial communication with the study PI, required sample sizes (that is, panels of clinical patient samples used in routine diagnosis for BU at NMIMR) to be used for the study were calculated (2.2.3.5 Sample size calculation). The uncertainty surrounding BU prevalence, as well as estimates, pointed out that the sample sizes used during development of the LAMP were

mostly too small in order to obtain statistically robust results using classical, frequentist statistics. In addition, PCR, the gold standard used for BU diagnosis, is known to be imperfect, that is, less sensitive than qPCR (sensitivity 84.04% (76.64–91.45) (Beissner et al., 2015). qPCR is not an option for routine diagnosis in Ghana due to its technical requirements. A Bayesian hierarchical model for calculation of LAMP sensitivity and specificity was built and will be reported in 2.3.1.1 Literature review and a Bayesian hierarchical model of the BU LAMP.

b. Field-suitable DNA extraction: One main challenge is to extract DNA in a setting without constant electricity supply (and without backup supply from a generator), as is the case in Pakro Health Centre, and which is likely to be representative for comparable settings. This excludes therefore a DNA extraction protocol using a conventional laboratory centrifuge.

> *I mean, if we had you know electricity, it would be no problem at all. So completely non-instrumented things, it's very difficult to achieve, especially DNA extraction. Thinking that you get the DNA from a mycobacteria, you know how difficult it is to obtain their DNA. (I2)*

> *So for now development of the […] point of care test is still in development. We have prototypes of various things that we are using for the diagnosis. Two major things: Since it's a […] nucleic-acid based detection assay, we need to have means of extracting DNA under field conditions. The second thing will be the amplification procedure itself. We need an incubation system […], we need a device that will be able to provide a constant temperature, I mean the optimal temperature, a constant point for one hour. So currently we have prototypes that we are field-testing to see if they would work under such field conditions. (I2)*

Mycobacterial DNA was pointed out as specifically challenging to extract, and experiments omitting DNA extraction (using crude, that is boiled or unboiled, samples) have yielded a lower sensitivity of the LAMP, but not PCR (Ablordey et al., 2012). Experiments using the syringe-based EasyNAT for DNA extraction have shown a lower sensitivity of the LAMP assay as well as PCR in detection of *M. ulcerans,* as compared to conventional methods, but above crude samples (Ablordey & de Souza, 2015; Ahortor, 2014). Another option mentioned was a handheld (hand-driven) centrifuge, which is available commercially. During the experiments, performance of EasyNAT was tested on a small range of samples, as reported in 2.3.1.2 Field-suitable DNA extraction.

c. Temperature stability of the reaction mixture: The current need for a cold chain for the LAMP reagents in its current state of development was mentioned. This is problematic in Pakro Health Centre, where refrigeration options are limited to 4° C storage and the capacity being

used for vaccines and drugs. Lyophilization, that is, freeze-drying of a master mix of reagents, was mentioned as a possible solution.

> *For now the issue would be the LAMP kit needs refrigeration, [...] so in settings like this where maybe there is no means of providing such [...] refrigeration conditions, then it becomes an issue. [...] There's a further step that we need to do: to produce or to make available a lyophilized or heat-stable form of the LAMP reagents, so that it can be kept at room temperature. (I2)*

A barrier encountered was that the polymerase used for amplification in the LAMP reaction was supplied in glycerol for storage, which would inhibit lyophilisation. Experiments were accordingly focusing on using a glycerol-free polymerase, and successful reconstitution of a lyophilized master mix, 2.3.1.3 Temperature stability of the reaction mixture.

d. <u>DNA amplification/incubation:</u> For amplification, LAMP needs incubation on 65° C for around 60 minutes, which has been pointed out as a challenge in the absence of a reliable electricity source for incubation. The NINA device and its high level of engineering and work in progress has been suggested as a good solution during discussions, and was used for experiments, as discussed in 2.3.1.4 DNA amplification/incubation:

> *NINA is fine, it has a lot of engineering […]. I don't know, they are constantly improving, they […] a lot of prototypes, so I'm quite hopeful in that with time […] they will get very good... (I2)*

2.2.2.2. The ideal diagnostic test

One broader topic area HCPs and researchers were asked to discuss was the ideal diagnostic test for this specific setting, the current challenges that will need to be dealt with and where the BU LAMP would possibly fit in. This was reflecting the experimental work, where it became clear that the current state of development of the BU LAMP would not yet allow implementation on the primary or peripheral health care level.

High sensitivity and specificity was reported as important, with the target being given at 100% for both during discussions. Specificity, the number of false positives, as a consequence of possible amplicon cross-contamination, was pointed out, and observed during experimental work. Training in prevention of contamination was seen as crucial. False positives due to other organisms that could be identified by the test were not seen as an issue, as indicated by an investigator.

The time to results should be as short as possible, with malaria diagnosis being mentioned as an example, where the RDTs are being conducted and results being available immediately. Estimates for conducting the LAMP are usually around 2.5 hrs, including DNA extraction. 2 hours duration to results is seen as likely if there is a pre-mix of reagents, which would be the case if reagents could be lyophilized. This time could potentially be shortened by 30 mins due to a shorter incubation time. In any case this range would allow for conducting the test while the patient is waiting for a diagnosis.

> *See, when you do it, and it's even negative, and you still suspect that this one, you can take another, and immediately you do it. But this one, you send it to the lab and then a week or two, and then you get it and it's negative. When you call the patient and tell the patient, and sometimes when they come, the results have not come [...] away. Anyway, they also, they don't have that much within their minds, so that's their problem. (H2)*

The potential lack of an intact cold chain at the peripheral health care level, as indicated by a non-continuous electricity supply, would ideally require the tests or components to be stable at temperatures exceeding 30°C over a longer period of time. However, refrigeration for storage of medical supplies is available in Pakro, and could potentially be used.

If the diagnostic test could be conducted in the clinic, sample transport would not be an issue. Otherwise, the costs for sample transport would need to be considered within the BU control programme. In general, low costs were seen as important, in particular in the context of other financial constraints for BU diagnosis and treatment.

Training of personnel was seen as crucial, in conduct of the test, in particular pipetting skills and prevention of potential cross-contamination. Ideally, the conduct of the test would require only minimal training. During the discussion participants mentioned that that technical personnel would be happy to receive comprehensive training.

2.2.3. Quantitative: The performance of the BU LAMP at the PoC

2.2.3.1. Determination of diagnostic test performance

For this study of the BU LAMP as index test, the reference test was PCR targeting genomic region *IS2404*, performed retrospectively in the laboratory facilities of the NMIMR (Guimaraes-Peres et al., 1999; Stinear et al., 1999). It is a nested PCR (using two primer sets in 2 consecutive steps) targeting *IS2404*, an extrachromosomal element repeated around 210 times in the genome. The research group's activities are also included in the national BU control programme of Ghana, identifying patients for referral, and to confirm their diagnosis using this test. The research group's laboratory forms part of a network confirming BU diagnosis (http://www.who.int/buruli/Global_network_laboratories_PCR_2014.pdf?ua=1), thus receiving routine diagnostic samples on a regular basis.

2.2.3.2. Sample collection and preparation for DNA extraction

Samples comprised FNAs and swabs. Swabs of the undermined edges of suspected BU lesion were obtained, kept in 2 ml PANTA transport medium (8.1 Appendix 1: Experimental reagents, primers and calculations) and gently vortexed or shaken for 5 secs. FNAs obtained by puncturing the nodules, plaques and oedematous lesions of suspected BU patients with a needle were transferred into sterile tubes containing 1ml phosphate buffered saline (1% PBS). Portions of the processed samples were used for DNA extraction.

2.2.3.3. Detection of *M. ulcerans* DNA under PoC conditions

Experiments mimicking field conditions were either conducted at the PoC in Pakro Health Centre, or an old laboratory area under uncontrolled conditions (not air-conditioned and with the windows left open, no use of electricity) at NMIMR for convenience when Pakro was difficult to access. In Pakro Health Centre, a work space was set up in a small 'laboratory area' within the treatment room of the centre, separated from the main room by wooden walls, and containing windows with glass shutters on two walls (Figure 1 b). A small hole to the main clinical area was used for passing through samples. The room had a tiled floor, benches, some drawers and a water tap with sink,

connected to an external tank, with water not always available. Due to its exposure and lack of ventilation, ambient temperatures within the room were higher than outside. The laboratory area was not fitted with an electric outlet. An outlet was installed in the treatment room, allowing for recharging of batteries, but electricity was only intermittently available. The temperature measured in the laboratory area during the period of experimental work was 29°C on average, ranging between 26 - 34.5°C, with temperatures being higher in the afternoon. Due to the high temperatures, in order to enable air to flow through, the window shutters could not remain closed during the day.

DNA was extracted using a syringe-based DNA extraction system EasyNAT (Ustar Biotechnologies, Hangzhou Ltd.) as follows, as described in (Ahortor, 2014) and adapted to conditions by using a commercial pocket warmer (Hokaron Haru-type, Lotte Health Products, Tokyo, Japan) as heat source for lysis: Before extraction, tubes were vigorously shaken. The BU samples (150 µl) were lysed by adding 600 µl of lysis buffer (1.6 M guanidinum HCl, 60 mM Tris (pH 7.4), 1% Triton X-100, 60 mM EDTA, 10% Tween20) and 50 µl Proteinase K (20 mg/ml). The mixture was heated to 60°-70°C for 30 min using a pocket warmer (Hatano et al., 2010) and allowed to cool to room temperature. Absolute ethanol (1400 µl) was added to the lysed sample, vortexed and sucked through the assembled silica syringe device (SSD) with the aid of the plunger. The lysed sample was then filtered back to the tube by pushing the plunger slowly (Figure 4a). DNA bound to the silica membrane at the base of the SSD was washed twice with 2 ml of 60% alcohol using air pressure generated from pushing the plunger of the syringe (Figure 4b). The 5 ml syringe barrel was detached from the SSD and replaced with a 1 ml syringe barrel for DNA elution. Nuclease free water (100 µl) was sucked through the nozzle of the syringe again by air pressure. With the aid of the plunger, nuclease free water was sucked over the silica membrane 3-4 times and finally eluted in a 1.5 ml eppendorf tube (Figure 4c). Care was taken not to contaminate the syringes during the extraction process.

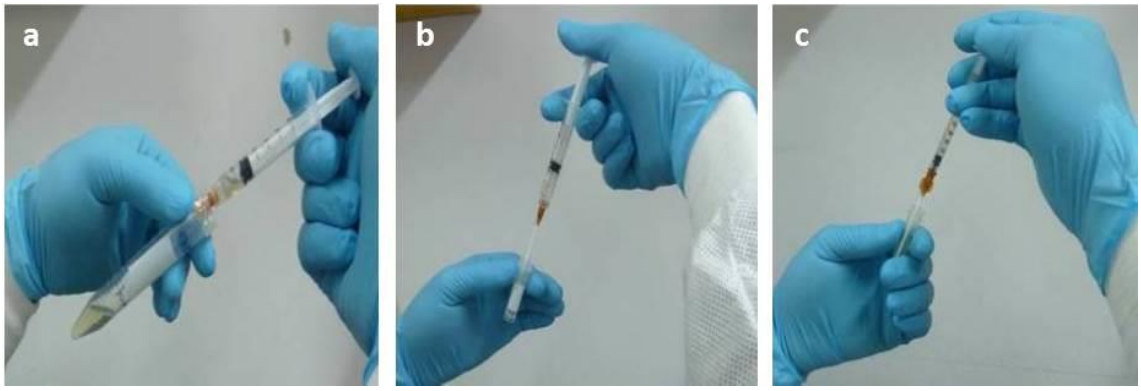


Figure 4: **DNA extraction procedure using disposable silica syringe device** . *(*a) Lysed sample sucked through the assembled device. (b) Washing of the DNA bound to the membrane in assembled device. (c) Elution of DNA from bound silica membrane with a 2 ml syringe. (Pictures reproduced with kind permission of E. Ahortor (Ahortor, 2014))

The LAMP assay for DNA amplification was performed as follows, using either a pre-mixed commercially available kit, or mixed from separate reagents in order to provide flexibility, e.g. for lyophilisation, and for cost reasons. In both cases, 4-6 different specifically designed primers (8.1 Appendix 1: Experimental reagents, primers and calculations), as described in Ahortor (2014), were used.

a. Commercial kit: Loopamp® DNA amplification kit RUO (Code No.: LMP205) and Loopamp® Fluorescence Detection Reagent (FDR) (Code No.: LMP221), both purchased from Eiken Chemical Co. Ltd. (Tokyo, Japan)). The LAMP reaction was performed in a 200 µl PCR tube in a total volume of 25 µl, using 1.6 mM each of primers FIP and BIP, 0.2 mM each of primers F3 and B3 and 0.8 mM each of primers LF and LB, 12.5 µl of 2X reaction mixture, 1 µl of Bst DNA polymerase, 1 µl of FDR, distilled water up to 20 µl, and 5 µl of extracted *M. ulcerans* DNA.

b. Mixed from separate reagents: The LAMP reaction was performed in a 200 µl PCR tube in a total volume of 25 µl. The master mix contained 2.5 µl of 10x reaction buffer, 4 mM $MgSO_4$, 0.4 mM dNTPs (NEB N0446S, dATP, dCTP, dGTP, dTTP: 0.4mM each), 1.6

mM each of primers FIP and BIP, 0.2 mM each of primers F3 and B3 and 0.8 mM each of primers LF and LB, 8 U Bst 2.0 WarmStart® Polymerase (kindly donated by NEB in a buffer either with glycerol, or glycerol-free) and Calcein/$MnCl_2$ solution (25 µM calcein, 0.5 mM $MnCl_2$ prepared as described in 8.1 Appendix 1: Experimental reagents, primers and calculations, double distilled H_2O (ddH_2O) up to 24 µl, and 1 µl of extracted *M. ulcerans* DNA.

A master mix was prepared and then aliquoted before addition of the DNA samples. As a positive control, DNA extracted from *M. ulcerans* isolates or from confirmed positive patient samples was added.

As a heat source for the amplification step, the Non-Instrumented Isothermal Nucleic Acid Amplification (NINA) heater (LaBarre et al., 2011) and DaqPRO 5300 Data logger device were kindly provided by Program for Appropriate Technology in Health (PATH, Seattle, WA). Experiments were conducted with two NINA prototypes kindly provided by PATH, H.V5.8 and H.V9, the latter having been optimized for ambient temperatures from 20- 36°C, as observed during experimental work with the prototype H.V5.8.

Before use, the NINA device was cooled to RT as well as possible by leaving it outside for sufficient time, or using running water.

After preparation, reaction tubes, controls (negative and positive) and one temperature sensing tube were arranged in the rack of the NINA device reaction (Figure 5). One heating pad (MgFe, provided by PATH) was put in the bottom of the flask and the exothermic reaction started by adding 5 ml 2% saline (provided by PATH) directly onto the heating pad, after which the NINA device was immediately assembled and covered with the lid, and temperature reading using the data logger was started. For experiments with NINA H.V9, reaction and control tubes were added after 10 mins waiting time from the start of the reaction according to manufacturer's instructions, once isothermal conditions were reached.

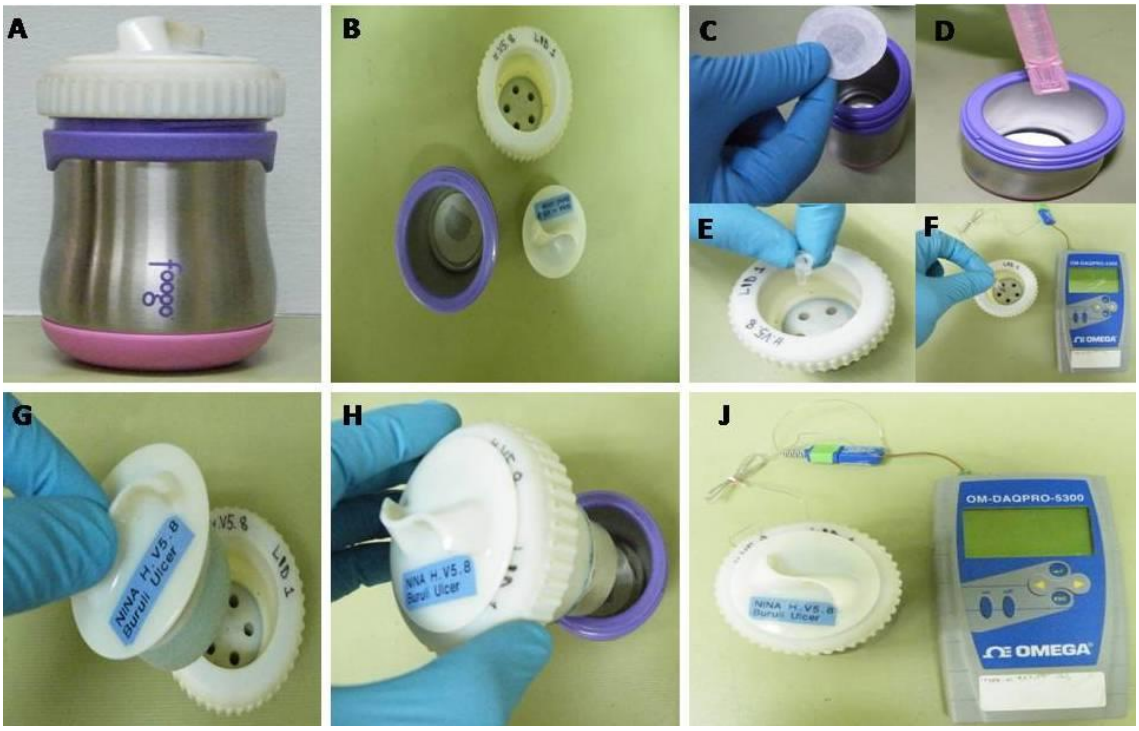


Figure 5: **Overview of NINA H.V5.8 device assembly and usage** (Ahortor, 2014). An assembled (A) and disassembled (B) NINA device. The thermochemical reaction is started by adding a MgFe heating pad (C) and saline (D). Thermal probes, connected to a temperature logger, were arranged in the rack (E and F). After closing the lid (G), the rack was inserted in the outer flask (H). Fully assembled NINA with temperature logger (J). (Pictures reproduced with kind permission of E. Ahortor (Ahortor, 2014))

2.2.3.4. Detection of *M. ulcerans* DNA under laboratory conditions

DNA was extracted from samples using the modified Boom method (8.1 Appendix 1: Experimental reagents, primers and calculations) or QIAamp DNA Mini kit (cat no 51304, Qiagen).

LAMP under laboratory conditions was conducted with the reaction mixture prepared as above, but instead of the NINA, a conventional thermal cycler providing a constant temperature of 65°C was used. In addition to visual inspection of the reaction tubes, results were visualized by inspection of tubes using a Kodak gel logic 100 molecular imaging system and/or viewed by electrophoresis

on 2% agarose gel prepared with 1X TAE buffer (40 mM Tris, 20 mM acetic acid, and 1mM EDTA) and stained with ethidium bromide. 5 µl of LAMP products were mixed with loading dye and added to the wells. The size of the amplicons was estimated by comparison with a 100 bp plus DNA ladder (Fermentas Life Sciences, EU) and visualized using a Kodak Gel Logic 100 molecular imaging system.

For lyophilisation of the LAMP reaction, a master mix was prepared as above without DNA template, adding trehalose (PHR1344, Sigma-Aldrich) ((Hayashida, Kajino, Hachaambwa, Namangala, & Sugimoto, 2015; Jain & Roy, 2010), N. Tanner, personal communication), 1 µl of a 2.5 M stock solution in ddH_2O for a 100 mM final concentration, with or without Calcein/$MnCl_2$ solution, filled up to 25 µl with ddH_2O and briefly centrifuged. The individual tubes with opened lids were carefully inserted into a 50 ml conical centrifuge tube (Falcon, 14-432-22) and frozen at -80°C overnight. Upon removing from the freezer, they were unscrewed and inserted into the glass vessels of a Labconco Vacuum Collector (Figure 6). This step was time critical in order to avoid thawing of reagents. Those centrifuge tubes containing PCR tubes with calcein solution were wrapped with aluminium foil for protection from UV radiation. Samples were dried for at least 4 hours at 0.22 mbar at a temperature of -46°C in the vacuum collector, and then stored, protected from light, either at 4°C for one week before reconstitution, or on RT for evaluation of stability.

For reconstitution of lyophilized reagents, 24 µl of ddH_2O were added to the tube and mixed by gently pipetting up and down. Once the lyophilized mass had dissolved completely, as determined by visual inspection, tubes were briefly centrifuged and 1 µl of DNA were added.

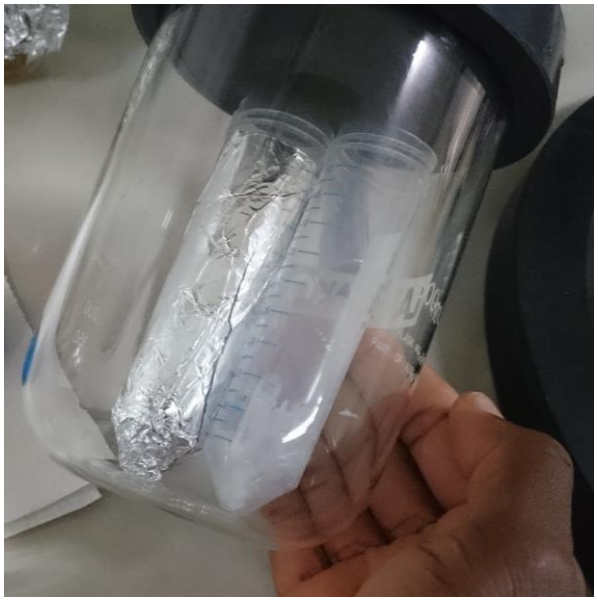


Figure 6: **Lyophilization of LAMP reaction mixtures.** PCR tubes containing LAMP reaction mixtures, inserted into centrifuge tube during lyophilisation. The centrifuge tubes were inserted upright in the glass vessel and attached to the vacuum collector, and care was taken to not thaw the contents in order to avoid spillage. Reaction mixtures containing calcein were protected from UV light by wrapping them in aluminium foil (left centrifuge tube).

Nested PCR over two rounds targeting the *IS2404* sequence was performed on all DNA extracts as described (Ablordey et al., 2012; Ahortor, 2014; Guimaraes-Peres et al., 1999) in 200 µl PCR tubes and a thermal cycler, using primers pGp1, pGp2, pGp3 and pGp4 (8.1 Appendix 1: Experimental reagents, primers and calculations). The first round was performed in a 30 µl reaction volume containing 25 pmol/µl of each primer (pGp1 and pGp2), 0.2 mM dNTPs, 3 µl of 10X PCR buffer containing 1.5 mM $MgCl_2$, 6 µl of Q-solution, 1U of HotStar Taq polymerase (Cat No./ID: 203203, Qiagen) and 3 µl of DNA extract. The thermal cycling conditions were as follows: 1x 95°C for 15', 40x (94° for 30", 64° for 1' and 72° for 1'30"), 1x 72°C for 10'. The second round was performed by adding 1 µl of the first run products to a 24 µl reaction volume containing 25 pmol/µl of each primer (pGp3 and pGp4), 5 µl of Q-solution, 2.5 µl of 10X PCR buffer, 0.2 mM dNTPs and 1U of HotStar Taq polymerase. The same thermal cycling profile was used, except for

the reduced number of cycles: 1x 95°C for 15', 40x (94° for 30", 64° for 1' and 72° for 1'30"). PCR products were viewed by electrophoresis on 2% agarose gel prepared with 1X TAE buffer (40mM Tris, 20mM acetic acid, and 1mM EDTA) and stained with ethidium bromide. 5 µl of second round PCR products were mixed with loading dye and added to the wells. The size of the amplicons was estimated by comparison with a 100 bp plus DNA ladder (Fermentas Life Sciences, EU) and visualized using a Kodak Gel Logic 100 molecular imaging system.

2.2.3.5. Sample size calculation

Sample size calculations for evaluation of a diagnostic test were performed at study onset following (Arkin & Wachtel, 1990; Buderer, 1996; S. R. Jones, Carley, & Harrison, 2003). Sample size requirements with clinically acceptable precision were calculated based on estimates and data from the literature for sensitivity and specificity, and the estimated disease prevalence in the target population: BU is a very rare disease, even in endemic areas, and limited epidemiological prevalence data are available. The disease has a distinctly focal transmission pattern, though the mode of transmission is not yet clear. An important assumption for this work is that in this specific context, prevalence refers to *prevalence in the target population of the diagnostic test* (that is, the group of subjects for whom the diagnostic test is designed, - i.e., with clinically-suspected BU - and who are actually being tested). For the purpose of this study, it is therefore defined as the positivity rate, that is, the fraction of patients with confirmed BU (disease positive using *IS2404* PCR), out of those with suspected BU based on clinical symptoms whose samples were submitted. This definition of prevalence strongly depends on the experience of the person performing the clinical screenings (A. Ablordey, personal communication), either in a referral procedure or in active case finding.

The number of subjects required to yield a 10% (clinically acceptable) width of a 2-sided 95% confidence interval (CI) for the sensitivity and specificity estimates ($\alpha = 0.05$) is 39 – 90 samples, depending on the prevalence of BU in the target population and the assumed sensitivity.

The range given above results from the extremes of the assumptions for prevalence and sensitivity compared to the reference test *IS2404* PCR: Prevalence was assumed to be 66-90%, according to an estimate for clinical routine samples during study planning (A. Ablordey, personal

communication), and a review and meta-analysis of the available data (8.2 Appendix 2: A systematic review and Bayesian meta-analysis of studies assessing the accuracy of BU LAMP). Sensitivity was set at ≥90% (Ablordey et al., 2012; Portaels, 2014; Sizaire et al., 2006; WHO/NTD, 2014); specificity was set at 100% (Guimaraes-Peres et al., 1999).

a) Assuming 66% prevalence of BU in the target population, and a sensitivity of 90% compared to the reference test, 90 individual samples would be necessary.

b) Assuming 90% prevalence in the target population, and a sensitivity equal to the reference test (90%) results in 39 samples required for the study.

Even the lower estimate of 39 samples was judged to exceed the typical sample sizes used for evaluating the BU LAMP during development due to technical and external constraints such as costs, and availability of patient samples (A. Ablordey, personal communication).

2.2.3.6. Quantitative data analysis

Analysis of experimental data and sample size calculations were conducted using the epi.tests package in R (3.3.0).

2.2.3.7. Literature review

A literature review of published and unpublished studies was conducted in order to extract data on the performance characteristics of LAMP for the diagnosis of BU in human samples as compared to PCR as reference standard (8.2.1 Systematic review protocol). A detailed list of databases searched and search terms can be found in 8.2.2 Databases and search terms. Studies identified through the literature search and from other were assessed according to eligibility criteria (8.2.3 PRISMA flow diagram): Those suitable for inclusion in the review and analysis had to be original studies assessing performance of LAMP as compared to PCR, or modifications of the PCR protocol (e.g. qPCR), containing paired data for comparison at the per subject level.. Of the included studies, data was extracted on the target sequence and primer sequence for LAMP primers, sample preparation and DNA extraction method, the heat source for LAMP amplification, the number of true/false positives/negatives, and sensitivity and specificity as compared to the

reference. Missing data could be calculated if not reported, after contacting the study authors if necessary using the epi.tests package in R (3.3.0). . An overview of the extracted data can be found in 8.2.4 Data extraction.

2.2.3.8. Bayesian meta-analysis

In addition, for quantification of test performance, an alternative Bayesian statistical approach was evaluated for this study. Frequentist models (those based on Neyman-Pearson inference as well as on Fisher's likelihood-based inference, which is regarded as a mixed form by some authors) are based on fixed, but unknown parameters (e.g. the mean of a normal distribution) and randomized, hypothetically repeated datasets. In Bayesian inference, unknown parameters are considered random, and the dataset is instead considered as fixed: The data collected at a given point in time are exactly known. A probability distribution (or probability density function (pdf)) is used to express the 'degree of belief' in a dataset representing *a priori* knowledge (which can be historical data, or expert estimates), taking the form of a uniform distribution if no prior data is available. An update of this *prior* probability distribution with experimental data (a pdf called likelihood) leads to a *posterior* probability distribution of parameters. Any inference is thus not only based on the information in the data, but also on information from previous studies or other sources, often allowing for smaller sample sizes (Basáñez, Marshall, Carabin, Gyorkos, & Joseph, 2004). Bayesian models are increasingly used in drug trials and diagnostic test evaluations (Berry, 2006; A.-L. Page et al., 2012).

This was in line with the following expected or observed challenges and uncertainties within this study that a Bayesian approach to statistical inference is able to handle well, in contrast to a classical or frequentist approach:

- *Limited sample size:* During study preparation, the sample size typically used for testing of the LAMP during development was expected to be insufficient in comparison to the required calculated sample size of 39-90 samples. Existing or historical data, e.g. from previously conducted and published studies of BU LAMP diagnostic accuracy, reports or

diagnostic records, can be incorporated for calculation of a valid prior. In their absence, a uniform, non-informative distribution can be used.

- *Expert elicitation:* Bayesian models are also able to incorporate expert estimates of relevant parameters, e.g. sensitivity, specificity and prevalence, in the construction of valid priors through expert elicitation techniques (Albert et al., 2012; Choy, O'Leary, & Mengersen, 2009; O'Hagan, 1998; Spiegelhalter, Abrams, & Myles, 2004). This was considered as particularly important with regards to the wide uncertainty surrounding prevalence, which could be addressed during discussions with researchers and HCPs.

- *Imperfect gold standard: IS2404* PCR, the reference standard used for this study in line with WHO recommendations, is an imperfect gold standard, with a reported sensitivity of around 84% as compared to qPCR (Beissner et al., 2015), which is however technically not feasible as reference standard in Ghana.

- *Missing data:* In the course of the study, missing data, e.g. incomplete paired datasets, were expected.

The analysis was based on data extracted during systematic review of published and unpublished studies assessing the accuracy of BU LAMP against PCR, as described in 2.2.3.7 Literature review. Data was incorporated into the model as follows: One *experiment* was defined as a panel of patient samples tested by both LAMP and PCR (paired data) within one study, resulting in 31 experiments (8.2.4 Data extraction, column *Experiment number and name*). Those experiment(s) from each study whose experimental protocol were as close as possible to the one under investigation were included in the analysis. In this case, this corresponded to a primary level field setting at the PoC, therefore those experiments that do not require electricity for DNA extraction and amplification were preferably selected.

This resulted in one experiment reported each for four studies (Ablordey et al., 2012; Ablordey & de Souza, 2015; Ahortor, 2014; de Souza, Quaye, Mosi, Addo, & Boakye, 2012), and two experiments reported for two studies (Beissner et al., 2015; Njiru et al., 2012) which contained two sample panels each.

A Bayesian prior pdf for sensitivity and specificity was kindly modelled by B. Lambert using R (3.3.0) and Stan® (2.13.0) (8.2.5 Meta-analysis of data). A point estimate and Bayesian Credible Interval (BCI) for sensitivity and specificity were calculated for each of the eight experiments separately, as well as for the entire dataset.

The assumptions for the model were as follows:

- Binomial distributions for the individual theta_sn and theta_sp parameters that represent random draws, independent of each other, from a population-level distribution.
- The positivity rate of *IS2404* PCR was used as proxy for prevalence in the target population, as discussed in 2.2.3.5 Sample size calculation.

The model is based on a study investigating diagnostic accuracy of a diagnostic test for the helminth strongyloides in a population of refugees using a Bayesian hierarchical model (Berkvens, Speybroeck, Praet, Adel, & Lesaffre, 2006; Broemeling, 2007; Dendukuri, Rahme, Bélisle, & Joseph, 2004; L. Joseph, Gyorkos, & Coupal, 1995; Lawrence Joseph, 2017). If available, additional estimates on local disease prevalence and test performance were gained from the individual interviews using expert elicitation (Spiegelhalter et al., 2004).

Differences in the results between a Bayesian and frequentist analysis were expected, as well as differences between results within the Bayesian approach, depending on the prior information is being used. However, only limited data is available addressing this (Austin, Naylor, & Tu, 2001; Shafer, 1982). This study aimed at making the assumptions underlying the established prior pdf explicit, and open for discussion with the research team, and this is considered as one of the strengths of a Bayesian approach and a possible point of triangulation between quantitative and qualitative data.

2.2.4. Qualitative: A Study Within A Trial (SWAT)

2.2.4.1. Observational: The setting

Field notes and photographs were taken by the researcher to document the setting.

2.2.4.2. Sampling

Participants for interviews and FGDs were selected based on a) their role in the diagnosis and treatment of BU at Pakro Health Centre and b) their role as researchers in the evaluation study after discussions with initially the study PI, and subsequently researchers and HCPs. Therefore, a convenience sampling approach with 'snowball' sampling was used.

2.2.4.3. Semi-structured individual interviews and focus group discussions (FGDs)

Interviews and the FGD were conducted in English, with the help of a translator if necessary. Before study onset, the objectives of this SWAT were thoroughly explained to each participant of individual interviews and the FGD. The consent form for interview and focus group participants (8.5 Appendix 5: Patient information sheet and informed consent declaration) was given to them, with the opportunity to ask questions. They were then asked to sign the attached volunteer agreement.

In-depth semi-structured interviews were performed with two researchers, two health care workers involved in diagnosis and treatment of BU at Pakro Health Centre, and five Community Based Surveillance (CBS) volunteers. The approximate duration of an interview was one hour, and refreshments were provided. The reason for conducting individual, semi-structured in-depth interviews was that this allows persons to speak free from social dynamics in a group context. Answers can be clarified and the interviewer is able to obtain more in-depth and perhaps sensitive information than in a social context (Bowling, 2009; Britten, 1995; Catherine Pope, Ziebland, & Mays, 2000).

I conducted the interviews, took notes, and recorded the interview for transcription and coding.

One FGD with a duration of around 1 hour was performed while at the PoC. Participants included two researchers and two HCPs treating patients in the health centre. A FGD is an unstructured interview technique with a small group of people (typically 6-12 participants), interacting with each other and the group leader and aiming to explore specific issues. This method was chosen because it allows participants to generate and explore their own and others' attitudes and ideas. It allows examination what participants think, and how and why they think that way, their understandings and priorities. More complex and rather technical issues can be discussed, and answers can be clarified (Bowling, 2009; Dawson, Manderson, & Tallo, 1993; Kitzinger, 1995, 2006). Importantly, FGDs are a popular technique in PAR, with participants as an active part of the research process (Kitzinger, 2006). A time point was chosen where I felt that participants were sufficiently familiar with me, and each other, and refreshments were provided, to enable an open and relaxed discussion atmosphere.

I facilitated the FGD, took observational notes, and audio recorded the discussion for transcription and analysis.

2.2.4.4. Qualitative data analysis

Audio recordings of the individual interviews and FGDs were transcribed into a word document, fully verbatim, and including hesitations, pauses, incomplete sentences and interruptions. The transcription was proof-read against the audio file to check for accuracy, and any missing or additional information was added.

The transcript conventions follow guidance given in Green and Thorogood (2014): I – interviewer, R – respondent, [] enclose material added by the author, […] indicate material omitted by the author. Irrespective of the gender of the respondent, the male grammatical form is used in order to protect his or her anonymity. Thematic content analysis of the transcripts, commonly used in exploratory healthcare research, was used for this study. During analysis, looking at data from multiple sources collectively, common ideas or elements were identified in an inductive manner. By grouping common ideas together, codes/themes were formed ('coding') using NVivo® 11 (QSR International), using thematic analysis (Braun & Clarke, 2006). Themes are concepts or

propositions that describe and help to interpret and explain aspects of the data, and they correspond to the final output of the analysis of the whole dataset. They are articulated and developed by comparison between and within interviews (Gale, Heath, Cameron, Rashid, & Redwood, 2013).

Transcription and preliminary thematic analysis for initial interviews were performed in order to inform subsequent interviews, incorporating emerging themes.

2.2.5. Ethical clearance

The SWAT protocol was approved by the Noguchi Memorial Institute's Institutional Review Board (IRB), as an amendment to the protocol containing the diagnostic test evaluation study, *Scaling up early detection and treatment to reduce Buruli ulcer morbidity in the Asante Akim North District of Ghana* (Certified Protocol Number CPN 026/10-11 amend 2015), and by the Oxford Tropical Research Ethics Committee (OxTREC) at Oxford University (Reference number 568-15).

2.3. Results

2.3.1. Introduction

The results of the experiments study will be presented in the following.

Experiments represent on-going technical development, using small samples sizes (or panels of samples). This study thus could be regarded as a longitudinal case study, embedded in and looking at a specific phase of the development of a diagnostic test. With this study, I would like to provide an example how qualitative methodology can be used to enrich and complement diagnostic evaluation studies in low-resource settings at rather early stages of the development of diagnostic test.

2.3.1.1. Literature review and a Bayesian hierarchical model of the BU LAMP test accuracy

A systematic review of available studies (published and unpublished) was conducted to construct a Bayesian hierarchical model of historical data on the test's sensitivity and specificity as described in 2.2.3.7 Literature review . Of the 442 identified studies from a search of relevant databases and two additional studies through other resources, 438 had to be excluded. 6 studies were included in the analysis, with a detailed overview of their experimental protocols and the reported diagnostic performance of the BU LAMP reported in 8.2.4 Data extraction. A total of 31 experiments testing a BU LAMP protocol against a PCR protocol could be identified, comprising 568 patient samples in total. In general, great heterogeneity in terms of methods (Primers used, sample preparation, heat source) was found.

Of these 31 experiments, eight (numbers 2, 5, 9, 10, 14, 16, 19 and 25 marked in dark orange in 8.2.4 Data extraction) were selected for meta-analysis, as detailed in 2.2.3.8 Bayesian meta-analysis and re-indexed to 1-8. Point estimates and BCIs were calculated for sensitivities and specificities (Figure 7) and an overall, or cumulative, pdf for sensitivity and specificity of the BU LAMP as compared to PCR (Figure 8) was constructed.

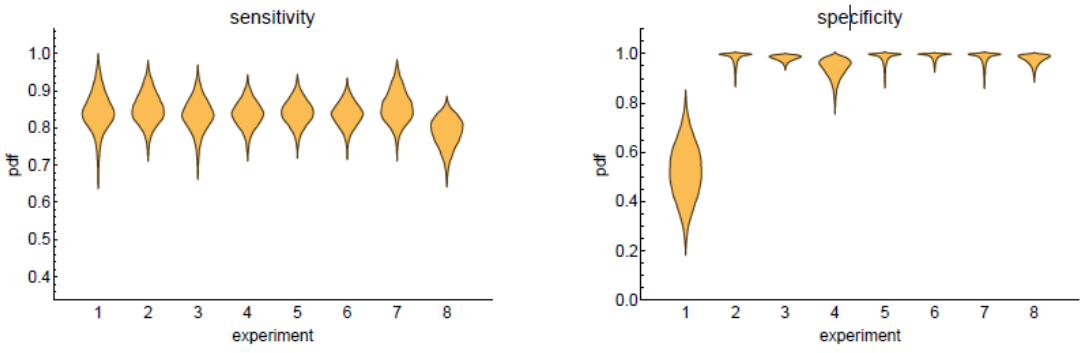


Figure 7: **Individual sensitivities and specificities of the BU LAMP experiments** included in the systematic review and meta-analysis of studies evaluating LAMP for detection of *M. ulcerans* in patient samples.

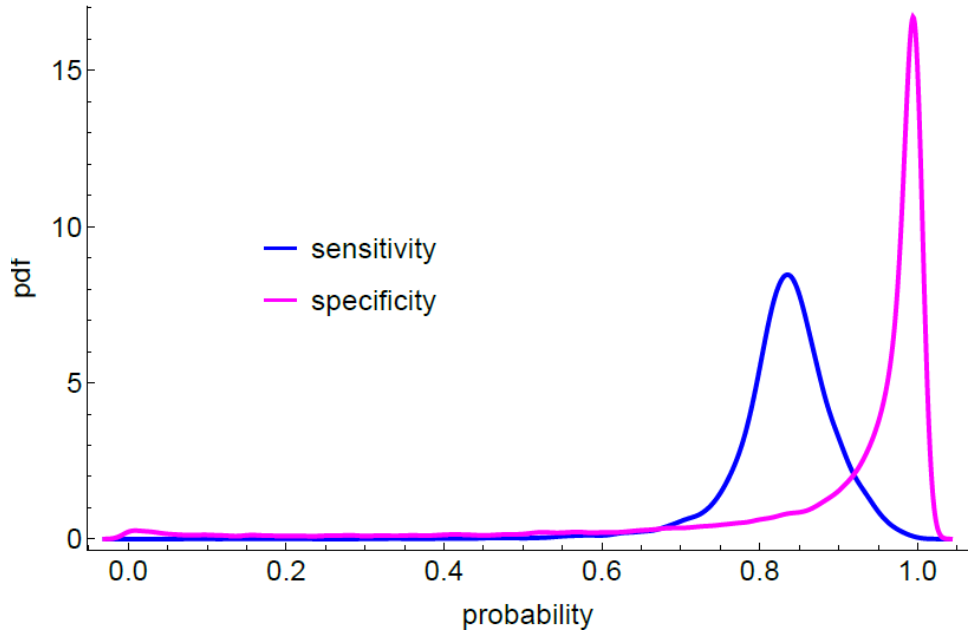


Figure 8**: Overall (cumulative) sensitivity and specificity of the BU LAMP** as compared to PCR**.**

These probability distributions for sensitivity and specificity were used as valid prior probability distributions (or priors) for update with experimental data to posterior probability distributions, or posteriors. Including the 568 samples reported in the literature, cumulative sensitivity of the BU LAMP was found to be 0.84 (0.80,0.87), with a specificity of 0.98 (0.9,1).

2.3.1.2. Field-suitable DNA extraction

Syringe-based extraction using the Easy-NAT kit followed by amplification using the NINA or amplification of previously isolated DNA from patient samples was performed at the PoC in Pakro during 14 experiments on n=16 patient samples, conducted in duplicate or triplicate per sample. After exclusion of samples in 3 contaminated experiments, 8 samples were found to be concordant with PCR and 3 to be discordant (Table 2). A sensitivity of 33.3% and a specificity of 83.3% was found for those extracted in the field (n=9 samples); a higher sensitivity of 66.6% and a specificity of 71.4% was found if samples that had not been extracted in the field were included also (n=13 samples).

Discordant samples were an issue in general; when testing a panel of 32 clinical samples which had been extracted with both QIAamp DNA Mini kit and the modified Boom protocol, followed by PCR, 16% of the samples were discordant.

Table 2: **Overview of sample concordancy**

Concordancy of samples IS2404 LAMP conducted at the PoC vs PCR	
Concordant	8
Discordant	5
Inconclusive	3
Total	16

Results from previous experiments conducted under laboratory condition suggest a lower sensitivity if DNA was extracted using EasyNAT: In one study, IS2404 PCR using syringe-based extraction was shown to have a sensitivity of 0.91 (0.80-0.97), LAMP after DNA extraction using syringe-based extraction 0.85 (0.73-0.93) and LAMP after DNA extraction using modified Boom 0.96 (0.87-1.00) (Ahortor, 2014). The specificity was 1 in all cases. Similar results were obtained in a different study, where LAMP after syringe-based DNA extraction had a sensitivity of 0.74 (0.64-0.83) as compared to PCR using the modified Boom method for extraction. (Ablordey & de Souza, 2015). An overview can be found in 8.2.4 Data extraction.

In summary, the obtained datasets could not be incorporated into the Bayesian model in order to model a posterior pdf and are only reported verbally. The high level of contamination indicated the difficulties, and it is unclear whether this method of DNA extraction is technologically feasible at the PoC at this stage.

2.3.1.3. Temperature stability of the reaction mixture

In the absence of a reliable cold chain at the PoC, lyophilisation of a master mix (that is, all reaction components without the sample in a dehydrated state) was explored as a possible solution. As

Glycerol, normally contained in the storage medium of the Bst 2.0 WarmStart® Polymerase, would inhibit lyophilisation, a glycerol-free polymerase was customized by NEB and tested first.

Analytical sensitivity of the Bst 2.0 WarmStart® Polymerase: LAMP using the Bst 2.0 WarmStart® Polymerase in glycerol-free buffer was found to have an analytical sensitivity, measured as limit of detection (LOD), of 10 fg of bacterial DNA using a thermocycler at 65°C (Observed range: 10 – 100 fg, depending on the template). This corresponds to 330 copies of *IS2404* or 1.6 bacteria. In comparison, previous studies (Ablordey et al., 2012; Ahortor, 2014) using commercial LAMP kits for amplification and detection, as described above, had found a detection limit of 30 copies for the *IS2404* LAMP. When using the NINA under uncontrolled conditions at an average temperature of 58.3°C, the LOD was found to be 100 fg.

Lyophilization of the reaction mixture: Reaction tubes containing Calcein/$MnCl_2$ instead of forming a compact mass at the bottom of the tubes (a *lyocake*), formed droplets, indicative of residual liquid, and often spilled out into the surrounding holding tube (Figure 9).

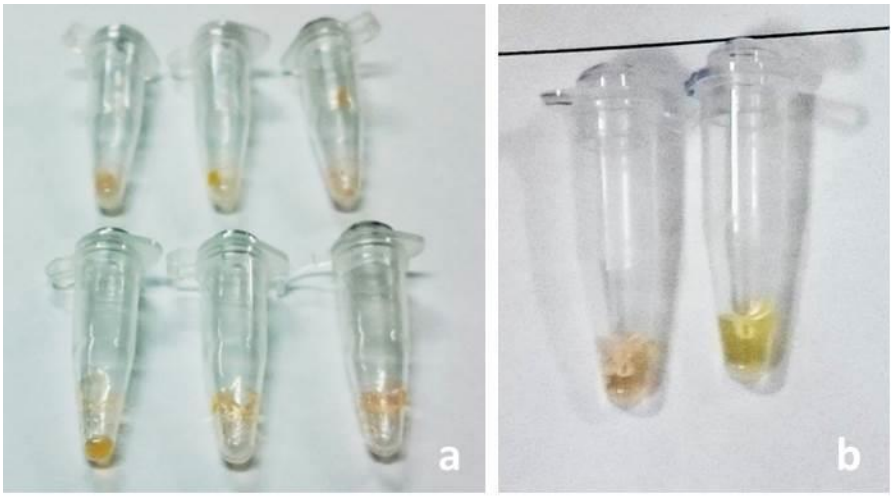


Figure 9: **Lyophilization in the presence of Calcein/$MnCl_2$ solution**. (a): After lyophilisation, reaction tubes containing Calcein/$MnCl_2$ solution do not show a compact lyocake, with droplets indicating residual liquids, and some spillage. (b) After reconstitution, the previously lyophilized reaction mixture (right tube) shows a greenish-orange colour as compared to non-lyophilized tube containing the Calcein/$MnCl_2$ solution (left tube).

Upon reconstitution, the solution assumed an orange-greenish colour upon lyophilisation, and showing strong autofluorescence under UV. No visual colour change and no change in fluorescence under UV were observed in case of amplification. Equivalent results were obtained using the commercial FDR. It was found that addition of trehalose (100 mM) was required for successful lyophilisation, and reconstitution of the reaction mixture. No amplification was observed without trehalose (Figure 10).

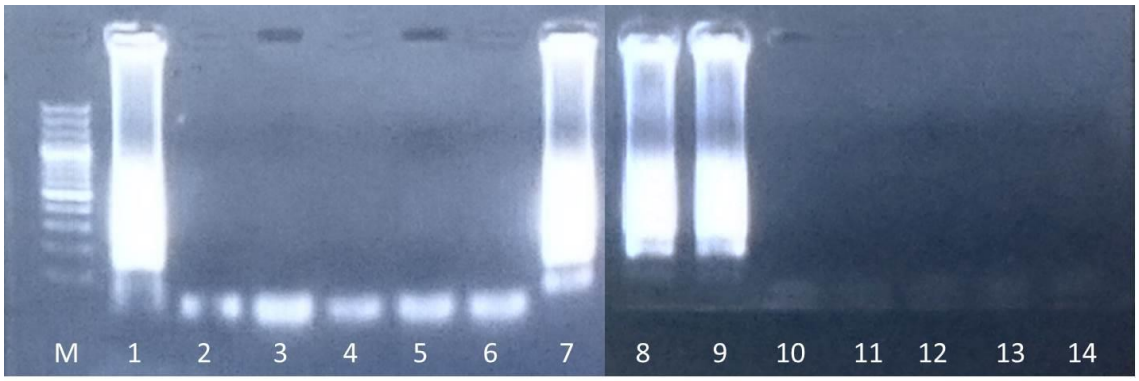


Figure 10: **Lyophilized LAMP requires trehalose for successful reconstitution ,** and detection of 1ng or 10ng *M. ulcerans* DNA. Bands indicate presence of amplification product, image merged from two horizontal lanes. Amplification could be observed only if trehalose was had been added before lyophilisation (lanes 7-9), and in a control sample which had not been lyophilized (lane 1), in the presence of DNA. Lanes:,1+2: -L, 1: +T, +FDR, DNA; 2: -T, -FDR, -DNA. Lanes 3-6: +L, +FDR, -T, 3+5: -DNA, 4+6: +DNA, 7-10: +L, +T, -FDR, 7: 1ng DNA, 8+9: 10ng DNA, 10: -DNA. 11-14: +L, -T, -FDR: 11: 1ng DNA, 12+13: 10ng DNA, 14: -DNA. M = marker, +L: lyophilized and reconstituted, -L: non-lyophilized control reaction, +/-FDR: includes/excludes Calcein/$MnCl_2$, .+/-DNA: Presence/absence of template.

In order to determine the sensitivity, varying concentrations of template DNA (in the range between 10 pg and 10 fg) were added to the reaction after 7 days of storage, and reconstitution (Figure 11). No loss of sensitivity as compared to a freshly prepared reaction mixture was found, but

a loss of reliability as one reaction of the duplicates each at concentrations of 100 fg and 10 fg failed to amplify.

In summary, it was shown that the commercial FDR could be replaced by a solution of Calcein/$MgCl_2$ at a fraction of the cost (low costs were among the elements of an ideal diagnostic test mentioned), but possibly at a loss of analytical sensitivity of one order of magnitude. Lyophilisation should be considered an option in order to confer stability to the reaction mixture at elevated temperatures, and to make the diagnostic workflow independent of a cold chain. However, the Calcein/$MnCl_2$ solution used for visual endpoint detection of amplification was found to be not stable during lyophilisation.

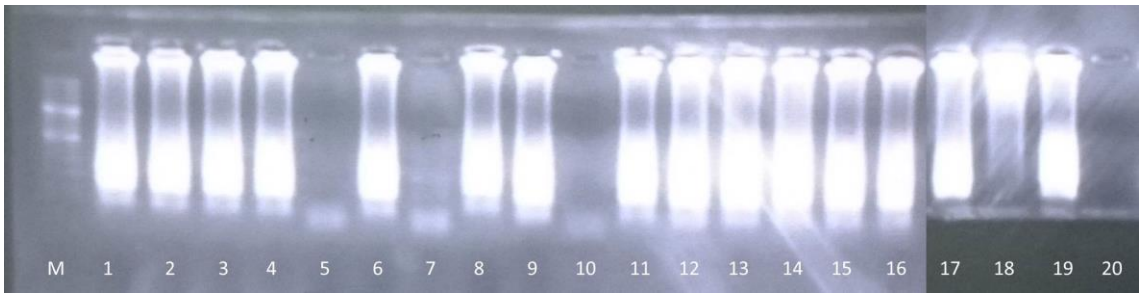


Figure 11: **Lyophilized LAMP with trehalose was able to detect concentrations as low as 10 fg of *M. ulcerans* DNA after storage of 7 days and reconstitution**, but not reliably so as compared to controls. Bands indicate presence of amplification product, image merged from two horizontal lanes. As compared to non-lyophilized controls, which were able to reliably detect concentrations as low as 10 fg in duplicates (lanes 17 and 18), the lyophilized reactions were less reliable and failed to amplify one reaction at 100 fg and 10 fg template DNA each (lanes 5 and 7). 10 fg template DNA corresponds to the LOD determined previously. Lanes 1-10: Lyophilized samples, reconstituted. 1+2: 10 pg DNA, 3+4: 1 pg DNA, 5+6: 100 fg DNA, 7+8: 10 fg DNA, 9: 1ng (positive control), 10: no DNA (negative control). Lanes 11-20: Non-lyophilized control reactions: Lanes 11+12: 10 pg, 13+14: 1 pg, 15+16: 100 fg, 17+18: 10 fg, 19: 1 ng (positive control), 20: no DNA (negative control).

2.3.1.4. DNA amplification/incubation

The NINA H.V5.8 device was capable of maintaining a reaction temperature between 60 – 70°C for at least one hour at high ambient temperature (29°C on average, ranging between 26 - 34.5°C, with some variability during the day) (Figure 13). An average temperature difference of 37°C was achieved.

As reaction temperatures above 70°C for the Bst 2.0 WarmStart® Polymerase are not recommended according to manufacturer's instructions, the NINA H.V9 prototype was specifically manufactured corresponding to the high ambient temperatures, and had a temperature range of 20-36°C according to specifications.

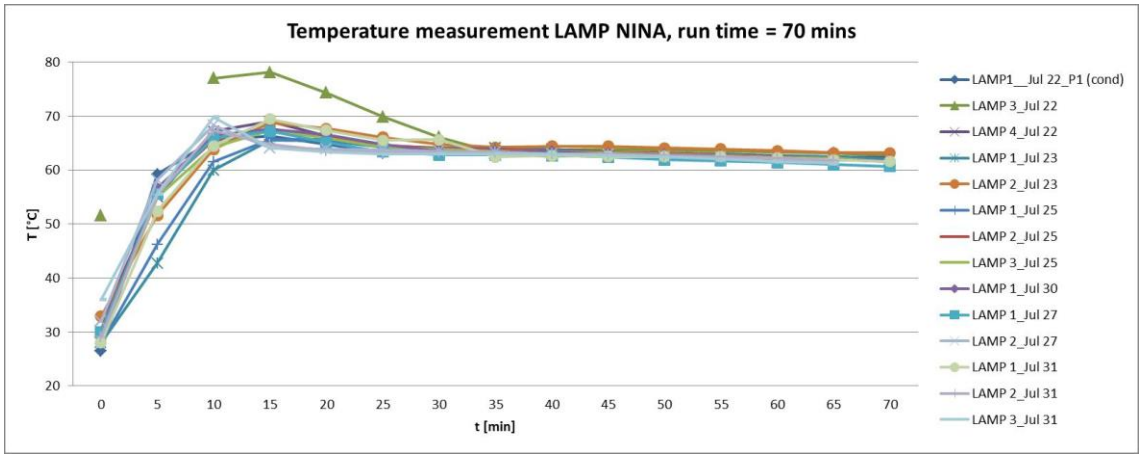


Figure 12: **Temperature curves of LAMP H.V5.8 experiments.** Measurements were taken for a run time of 70 minutes, at an average ambient temperature of 29°C. All experiments with starting temperatures between 26 – 36 °C (corresponding to equilibration at ambient temperature) were able to keep the reaction temperatures between 60 – 70 °C for one hour.

When tested with ambient temperatures, the NINA was able to keep the reaction temperatures between 52.6 –61.6°C for one hour after an initial peak in temperature, after which samples were added according to manufacturer's instructions (Figure 13). Specifications for the Bst 2.0

WarmStart® Polymerase state a temperature range between 60-70°C, with the activity optimum at 65°C.

Across four independent series of experiments comparing LAMP performed in a heat block at 65°C to LAMP performed using the NINA H.V5.8 at an average ambient temperature of 26°C (range from 24 - 29°C), reaction temperatures were found to be between 52.6 –61.6°C (Figure 13). Several non-systematic weakly positive (of an unusual pale orange green) reactions in tubes either containing negative control or very low DNA concentrations were observed (not shown). This occurred only when using the NINA and not a conventional thermal cycler, where control reactions were run in parallel. This could be due to handling errors, or unspecific reactions of the polymerase at temperatures below 60°C when using the NINA.

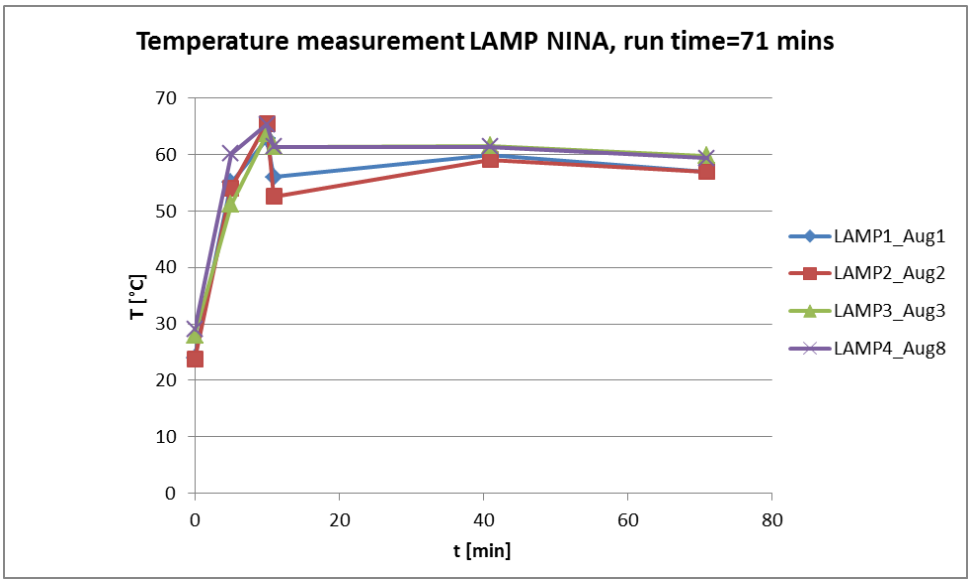


Figure 13: **Temperature curves of NINA H.V9 LAMP experiments.** Measurements were taken for a run time of 70 minutes, at an average ambient temperature and a starting temperature (corresponding to equilibration at ambient temperature) of 26°C (range 24 - 29°C). All experiments were able to keep the reaction temperatures between 52.6 –61.6°C for one hour.

In summary, the NINA was found to be a very versatile and robust instrument for non-instrumented DNA amplification, able to keep the T in a range corresponding to optimal polymerase activity. Some modifications depending on the ambient temperature might be required corresponding to the temperature profile of the polymerase used for the LAMP reaction.

2.3.2. Summary

This study is a study consisting of a SWAT conducted within, and reflecting upon an evaluation study of the BU LAMP, a newly developed diagnostic test for BU. The study assesses the potential of the LAMP for implementation at the primary or peripheral health care level (PoC) in a rural area in Ghana, Africa. The study site was Pakro Health Centre as well as NMIMR for laboratory facilities.

The pathway used for diagnosis and treatment of BU at Pakro Health Centre was discussed using individual interviews and FGDs: Collaboration with CBS volunteers, who are based in the villages and trained for a number of health-related tasks, plays an essential role in case finding and referral. The population is sensitized *via debes*, public talks, as well as in schools, and volunteers visit family homes for identifying cases. BU is often understood as curse or witchcraft. Against this background of a strong belief in witchcraft in the population, the as yet unclear mode of transmission, and the long treatment duration with antibiotics often leads patients to seek alternative (herbal, spiritual or alternative medical) treatments first. Alternative treatments were identified as a major cause of delay of treatment onset, and the costs put an additional burden on patients. Collaboration with traditional healers in order to ensure treatment of patients as well as education of the population about BU as a medical condition that can be treated appropriately is seen as important. Upon arrival of patients, usually during clinic days, patients' history is being examined, and clinical diagnosis is made. Samples (either FNAs or swabs) are being taken and sent to NMIMR directly for analysis, where PCR is being performed for diagnosis. Patients are informed, and asked to present at Pakro Health Centre for treatment. The current standard of treatment is oral rifampicin and streptomycin injections for 8 weeks, accompanied by wound dressing, usually requiring daily visits

initially. Interview participants reiterated the problematic transport situation to and from the facility for patients, with possible solutions being treatment at a health facility closer to their home, or additional financial support.

A number of context-specific barriers for diagnosis and treatment emerged from interviews and discussions. Some were related to logistics and transport options, that is, availability and affordability of local transport options for patients, poorly defined sample logistics, and the mobility of CBS volunteers (in particular when reaching smaller hamlets in scattered villages after nightfall or during the rainy season). Other barriers included patients' preference for traditional treatments, the time delay between sample taking and onset of treatment and the paradoxical reaction to treatment were also pointed out. From a technical/clinical point of view, factors influencing the diagnostic performance of PCR mentioned in this context were false negatives, that is, patient samples testing negative despite a high clinical suspicion (and successful treatment), and sample submission without clinical suspicion.

Four main current challenges for the BU LAMP as a PoC test specifically were brought up in the study planning phase and during the SWAT. These were addressed in the quantitative phase of the study, through a mathematical model and experiments:

- A Bayesian hierarchical model was constructed in order to address the problem of small sample sizes of clinical panels, and of PCR being an imperfect gold standard for diagnosis of BU. Prior distributions for sensitivity and specificity were constructed using data extracted from previously conducted studies. Unfortunately, the achieved sample size (mostly due to contamination issues) during experiments at the PoC did not produce a full dataset, so the model could not be updated using experimental data. A limitation of the model is the challenging definition of prevalence as the positivity rate of *IS2404* PCR, as this is heavily depending on the HCP conducting the clinical diagnosis of BU.
- Lyophilisation should be considered an option in order to confer long-term stability to the reaction mixture at elevated temperatures at the PoC, and to make the diagnostic workflow independent of a cold chain. Preliminary experiments have shown no loss of analytical sensitivity after 7 days of storage and reconstitution. However, the Calcein/$MnCl_2$ solution

used for visual endpoint detection of amplification was not found to be stable during lyophilisation.

- The NINA (LaBarre et al., 2011), a device used for incubation of the LAMP reaction which does not require an electricity supply, was found to be a very versatile and robust instrument for non-instrumented DNA amplification. Two prototypes were tested for this research, and some modifications might be required corresponding to the specific ambient temperature and the temperature profile of the polymerase used for the LAMP reaction.

Experiments conducted at the PoC pointed at a currently burdensome workflow and a high risk of technical failures and contamination. This work conducted during the exploratory study indicated that LAMP in its current state of technological development should not be considered yet for implementation on the primary, or peripheral, health care level for diagnosis of BU.

In summary, this study shows how easy-to-implement qualitative methodology is able to complement traditionally quantitative experimental work within a diagnostic evaluation study, providing a much richer picture on the context, and is able to reflect upon experimental efforts. This is in line with recommendations to take the diagnostic pathway and context much more into consideration during tests of a new diagnostic (Banoo et al., 2006; Knottnerus et al., 2009, 2002).

2.4. Discussion

2.4.1. Introduction

This discussion section will focus on four main interrelated points. First, it will look at the (technical) performance indicators of the BU LAMP and their clinical and practical significance in the context of the study. Second, it will discuss issues related to study design of diagnostic evaluation trials. Third, it will discuss limitations of the study and fourth, will suggest further research in the light of the study results and limitations.

2.4.2. Performance indicators of the BU LAMP

In the following, the performance indicators sensitivity and specificity will be revisited with regards to their clinical significance, and in connection with prevalence.

Modelled sensitivity for BU LAMP, based on previous studies, as compared to PCR is 0.84 (see section 2.3.1.1 Literature review and a Bayesian hierarchical model of the BU LAMP). qPCR was found to be more sensitive than PCR (Beissner et al., 2015), but requires equipment that is not available for routine diagnosis at the reference laboratories for BU in Ghana, and is therefore not taken into consideration. Specificity of BU LAMP was found to be 0.98 (see section 2.3.1.1 Literature review and a Bayesian hierarchical model of the BU LAMP). In general, specificity as the number of false positives, as a consequence of possible amplicon cross-contamination (that is, contamination of a reaction with products from previous reactions) at the current stage of LAMP's development was pointed out during discussions, and observed during experimental work. It is well documented for LAMP reactions due to its high sensitivity (Karthik et al., 2014). It could be mitigated by avoiding opening of the tube after the reaction, appropriate disposal and other field-suitable measures to counteract nucleic acid cross-contamination (Aslanzadeh, 2004), and it was emphasized that this should form an important part of the training. False positives due to other organisms that could be identified by the test were not seen as an issue during interviews. False positive test results would result in treating individuals with inappropriate medication, in this case an 8 week combination of rifampicin and streptomycin. Although the risk of side effects was deemed low during studies, streptomycin has a documented cochleovestibular or renal toxicity (Chauty et al., 2007).

Disease prevalence is an important parameter for a diagnostic test, as endemicity dictates the requirements to a suitable diagnostic test for disease control purposes (Bergquist, Johansen, & Utzinger, 2009). As discussed previously in section 2.2.3.5 Sample size calculation, prevalence in the context of study refers to prevalence in the target population of the diagnostic test, that is, those who are actually being tested. It is therefore defined as the positivity rate, that is, the fraction of patients with confirmed BU (that is, disease positive using IS2404 PCR), out of those with suspected BU based on clinical symptoms whose samples were submitted. A crucial underlying

assumption was that samples are only submitted in case of a clinical suspicion. However, it was found during interviews that this might not always be the case. In addition, as mentioned, this definition of prevalence strongly depends on the skills and experience of the person conducting the screening. In this study, prevalence therefore has a great amount of intrinsic uncertainty, thus influencing a) the Bayesian hierarchical model of sensitivity and specificity, as detailed in 2.3.1.1 Literature review and a Bayesian hierarchical model of the BU LAMP and b) the sample size calculations within a Frequentist paradigm, as discussed in 2.2.3.5 Sample size calculation.

In clinical practice, diagnosis is typically based on clinical aspects and confirmed by molecular methods only if available, and often after initiation of treatment (Junghanss et al., 2014).

2.4.3. Study design of diagnostic evaluation trials

A key question was whether the BU diagnostic evaluation trial could be considered a 'trial'. Clinical trial definitions are typically quite similar. WHO f.i. defines clinical trial as follows:

> *For the purposes of registration, a clinical trial is any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes. Interventions include but are not restricted to drugs, cells and other biological products, surgical procedures, radiological procedures, devices, behavioural treatments, process-of-care changes, preventive care, etc. (WHO, 2017)*

For the International Committee of Medical Journal Editors (ICMJE), a trial is

> *any research project that prospectively assigns people or a group of people to an intervention, with or without concurrent comparison or control groups, to study the cause-and-effect relationship between a health-related intervention and a health outcome. Health-related interventions are those used to modify a biomedical or health-related outcome; examples include drugs, surgical procedures, devices, behavioural treatments, educational programs, dietary interventions, quality improvement interventions, and process-of-care changes. Health outcomes are any biomedical or health-related measures obtained in patients or participants, including pharmacokinetic measures and adverse events (ICMJE, 2017).*

The findings in this study in general, but in particular from the qualitative stream using observations and interviews, reflecting upon the study design itself, indicate an iterative, complex research process strongly embedded in a particular context, rather than a trial according to definitions. At this stage, the study points out an on-going technical development, using small samples sizes (or panels of samples). The intervention, in this case the index test BU LAMP with reference test

IS2404 PCR, was modified repeatedly. Many different configurations or experimental protocols of the technology under investigation were not reflected in the study protocol, which was accordingly loosely defined. Mostly, routine diagnostic samples were used. This should not be seen as criticism, but seemed adequate and provided the flexibility needed in order to incorporate on-going modifications in this setting. Lijmer at al. (2009) have pointed out that existing evaluation frameworks for diagnostic evaluation trials, as discussed in section 2.1.1 Evaluation trials of diagnostic tests, can be useful to distinguish between study types, but they cannot be seen as a necessary sequence of evaluations. They argue that the evaluation of tests is most likely not a linear but a cyclic and repetitive process. This seems to be unlike the paradigm of the four-to-five-phased model for drug development, which is mostly presented as linear. However, Houn et al. (2000) have argued that overall, even the drug development process might actually be more cyclical and repetitive, starting with problem recognition through experimentation, assessment and adoption, and repeated if the technology is improved, e.g. using chemical modifications to improve the half-life or drug safety, or modified (repurposed) for new uses. Hunink and Krestin (2002), focussing on diagnostic imaging, suggest a circular approach as more appropriate, with concurrent development, assessment, and implementation of technology, and following up on that, Lijmer et al. (2009) extend this idea to the evaluation of tests in general, as seeing evaluation of tests as a linear series of phases *'will ultimately prove to be too restrictive, and may fail to do justice to the myriad of tests and the wide range of testing purposes'* (Lijmer et al., 2009, p. 8).

However, some authors have pointed out a poorly defined methodology of diagnostic evaluation trials as yet (Knottnerus et al., 2009, 2002). A better defined methodology would also provide guidance to researchers, as discussed in section 2.1.1 Evaluation trials of diagnostic tests. I argue that methodology studies, such as SWATs, have the potential to address specific methodology issues within trials, and are therefore a good instrument to improve and develop the methodology of diagnostic evaluation trials.

A finding that I consider important is that both parts of this study – the quantitative trial and the qualitative SWAT - could not be conducted independently, as a mere observer role would not be accepted and would make a complete understanding of the context difficult. Considering PAR, -

this study follows the rather broad definition by Wadsworth (1998) among the many available – allowed for collaboration, active participation and immersion of the investigator in the research process, which was conducted in a cyclical, rather than linear manner, while being fully aware of its inevitable intervention in the social situations within which it operated.

There are examples of SWATs within diagnostic evaluation trials (despite not being called that), such as a study testing a LAMP for pulmonary tuberculosis in microscopy centres of Peru, Bangladesh and Tanzania (Boehme et al., 2007). The study, assessing clinical performance, also included a methodology component to assess the operational feasibility. This component addressed issues such as training needs, hands-on time, interrater reliability, perceived utility and feasibility of LAMP during routine use, and priorities for simplification of the workflow. The study combined empirical data with individual questionnaires, containing quantitative and qualitative items, assessing the current version of the LAMP and setting research priorities for further development(Boehme et al., 2007).

2.4.4. Limitations of the study

One clear limitation – at the same time, one of its strengths – is that it is a case study focusing on a very specific setting and context. Intrinsic to qualitative research is that it does not aim for generalizability or statistical representativeness, as previously discussed in 1.2 Methodological approach. The contrast between the qualitative and the quantitative research strands is perhaps best visible in the different sampling strategies used (as described in 2.2.3.5 Sample size calculation for the quantitative strand, aiming at statistical robustness of the sample, and 2.2.4.2 Sampling for the qualitative strand, using a convenience sampling approach). However, the overall study aims at transferability of findings in two ways: First, transferability to other diagnostics evaluation studies in similar settings. This linkage to the context, in line with the qualitative research paradigm, moves away from generalizability, as an all-important principle, to transferability to other settings or contexts. A prerequisite for transferability is that enough information is provided alongside of the results: The findings include an accurate description of the study site, including personnel involved

and resources available, elicited through observations, interviews and the FGD. The aim is to allow researchers conducting similar studies to assess transferability of this study's findings to their own trials. Second, transferability to routine use of the diagnostic assay: This topic area was covered during interviews and the FGD, focussing e.g. on perceived training needs of the intended users, possible opportunities for training, and detailed discussion of the diagnostic pathway with suggestions for improvement. Transferability was a point brought up and addressed during the IRB review of the protocol for this study.

The clearly exploratory nature of the study could be considered another limitation. Experiments reported here could only be conducted in a low number of repeats under at times challenging experimental conditions. This study in its entirety should therefore be considered as exploratory only, with elements of an implementation research study. This specific type of study, as defined by Peters et al. (2013), aims at evaluating acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, coverage, and sustainability of an intervention. These aspects in relation to a diagnostic test for BU could be further investigated in a larger study, using similar methodology to this work.

Interoperator and interrater reliability were not quantified for this study. As *per* the protocol, the calculation of interoperator reliability (as a measure of robustness of results if the test was conducted by different persons) was planned, but could not be conducted due to a lack of time and resources. In this particular context, I would consider this parameter relevant for the testing process, as well as the procedure of taking patient samples (FNAs or swabs): During the interviews it was pointed out that the positivity rate was depending on the skillset and experience of the person performing the screening. Both interoperator and interrater reliability, that is, robustness of results when test readouts are judged by different readers as either negative or positive, could be seen as indicative for training requirements, another important element brought up during interviews.

The conflicting data available on prevalence of BU at the onset of the study, as discussed in 2.2.3.5 Sample size calculation and 2.2.3.6 Quantitative data analysis poses a clear limitation. Unfortunately values for prevalence could not be established with more reliability in the course of the study,

neither during interviews and the FGD, nor from diagnostic records. This is in particular unfortunate given the central role of this parameter for the performance of diagnostic tests in general (e.g. the dependence of a test's positive predictive value (PPV) on prevalence, if sensitivity and specificity are constant (Parikh, Mathai, Parikh, Chandra Sekhar, & Thomas, 2008)). For this study, prevalence is a central parameter for both sample size calculation and modelling sensitivity and specificity using a Bayesian approach, where having a more robust estimate, or data, would have helped to establish a more reliable prior.

Another limitation is that I was not able to approach and consult BU patients directly, due to limited resources and the language barrier. I think their narratives would have very much enriched the study. The CBS volunteers that were consulted, living in the communities with patients, are seen as their placeholders – for some aspects, this admittedly might be inadequate.

2.4.5. Further research

Target product profiles (TPPs) are seen as a useful guidance for development of interventions or diagnostic tests. Described by the Food and Drug Administration (FDA) as '*format for a summary of a drug development program described in terms of labelling concepts*' (Food and Drug Administration (FDA), 2007, p. 5), it is also increasingly used as an instrument for development of diagnostic tests, providing details on performance and operational characteristics (minimum/optimum) of tests in development. As an example, FIND publishes TPPs for their priority diagnostic tests on their website (Foundation for Innovative New Diagnostics (FIND), 2017), they are used as consensus instrument between stakeholders such as developers, users, researchers, and donors to inform and focus development efforts. Substantial information that could inform a TPP can be found in the interviews, during discussion of the ideal diagnostic test (see 2.2.2.2), such as sensitivity, specificity, technical skills (connected to training needs), and the nature of samples (incl. sampling procedures).

As the possibility of adding an appropriate stain after the reaction has finished provides a great contamination risk, alternative dye systems for visual endpoint detection should be considered, e.g. pH-based dyes that have shown promising results when used with LAMP (Tanner et al., 2015). The

dyes yielded a very clear colour change in case of amplification as shown during experiments (tested as a prototype, data not shown), and remained stable during lyophilisation (N. Tanner, personal communication). Unfortunately, appropriate experiments could not be conducted due to a lack of time.

As discussed previously in 2.4.4 Limitations of the study, a quantitative measure of interoperator and interrater reliability would be important for further work. Methods used could be either Fleiss' Kappa or AC1/AC2 (Blood & Spratt, 2007; Gwet, 2010; Viera & Garrett, 2005; Wagai, Senga, Fegan, & English, 2009; Wongpakaran, Wongpakaran, Wedding, & Gwet, 2013). In a technical context, exploring and quantifying the interrater reliability is of relevance for dichotomization of the LAMP results, that is, the robustness of the threshold in determining the cut off between a negative and a positive sample. The LAMP method allows in principle for a fluorometric or colorimetric readout with naked eye detection, or using a simple hand-held device measuring UV or turbidity (Njiru, 2012). Analysis of test reliability is expected to quantify inter-individual differences between operators and raters, and also provide indicators for standardization (e.g. creating a standard operating procedure (SOP)). Ideally, this would use the difference in performance between operators with different skill levels as an indicator for transferability, following (A.-L. Page et al., 2012)).

Another possibly useful area of research is a more formal approach of expert consultation for obtaining estimates of relevant parameters for a useful Bayesian hierarchical model, as described by authors for other fields of research (Albert et al., 2012; Choy et al., 2009; O'Hagan, 1998; Spiegelhalter et al., 2004). In my opinion, in particular prevalence, and the uncertainty around this central parameter due to a lack of empirical data and the great variability due to sampling methods, could gain from reliable estimates and make sample size calculations and the model more robust.

2.5. Conclusions

Reflecting back on the original aims and objectives of this study, it was shown that first, using a combination of qualitative and quantitative data sources or types, and second, working with a range

of stakeholders, was able to provide a rich overall account of a diagnostic evaluation trial in a low-resource setting. Moreover, it provides an insight into an iterative process of research, using points of triangulation between data types: Technical challenges to BU diagnosis could be identified with stakeholders, and partly addressed with experiments; importantly, context and identified barriers for diagnosis and treatment of BU provide guidance two-fold: To help with planning future research and conducting further evaluation studies of the BU LAMP in this setting on one hand, but also to improve access to diagnosis and treatment for patients on the other hand. Continuous work with stakeholders (researchers, HCPs and CBS volunteers based in the villages) was a crucial factor for making this possible. The use of stakeholder estimates for eliciting crucial parameters such as prevalence was not successful, but merits further research. To me this is a fascinating point of triangulation between qualitative data (expert opinions) and quantitative data (using this parameter to quantify diagnostic test accuracy in a Bayesian statistical paradigm).

Results of the experimental part of this study so far indicate that LAMP in its current state of development is not suitable for implementation in a primary, or peripheral, health care setting. There is on-going research towards a PoC screening test for symptomatic patients at the district hospital level, with BU LAMP among the candidates (Foundation for Innovative New Diagnostics (FIND), 2017 and A. Ablordey, personal communication). This is in line with successful usage of LAMP for detection of other pathogens, e.g. *Plasmodium* or *Trypanosoma brucei* (Hayashida et al., 2015; Sema et al., 2015).

As noted in 2.2.1 Study design, I as the investigator conducting the SWAT was also working as part of the research team for the diagnostic test evaluation study: Predominantly with the study PI for initial planning of the study, and at later stages consulting the other researchers and HCPs participating in the trial. As a trained molecular biologist, I was working alongside the researchers involved with the diagnostic evaluation trial, and actively conducting setup and experiments. I would argue this study fulfils the criteria for PAR following the definition of Wadsworth (1998), as laid out in 2.2.1 Study design. The iterative research process outlined above is in line with PAR: Research directions were suggested taking stakeholders' experience and knowledge and the context into consideration; results of experiments were then reflected upon with the research group, and

thus informed (and will inform) future steps. By contrast, a mere observer role of the investigator would not have been accepted and, from my perspective, would have made understanding of the research process difficult.

3. Using patient consultations to better define outcome measures for Cutaneous Leishmaniasis (CL) clinical trials

This part describes a study addressing the need to consult patients about their preferred treatment effects (outcomes) for clinical trials for CL. Outcomes are crucial elements for clinical study design, and more generally providing better and more adapted health interventions to patients, which have insufficiently been taken into consideration so far in particular for NTDs.

3.1. Background and literature review

3.1.1. Introduction and epidemiology of CL

Cutaneous Leishmaniasis (CL) is the most common of three forms of leishmaniasis caused by protozoan parasites belonging to the genus *Leishmania*, which are transmitted by the bite of infected sand flies. CL causes lesions on exposed parts of the body, which will potentially leave life-long scars. Approximately 95% of CL cases occur in the Americas, the Mediterranean basin, and the Middle East and Central Asia, and over two-thirds of new CL cases occur in six countries: Afghanistan, Algeria, Brazil, Colombia, the Islamic Republic of Iran and the Syrian Arab Republic (Alvar et al., 2012). An estimated 0.7 million to 1.3 million new cases occur worldwide annually (World Health Organization, 2014), and altogether, the leishmaniases account for more than 3.3 million DALYs (all ages DALYs (Murray et al., 2012)).

Like many other NTDs, leishmaniasis often occurs in remote locations and is focally distributed, rendering extrapolation from official data sources difficult (Alvar et al., 2012; Bern, Maguire, & Alvar, 2008). Alvar et al. (2012) found up to 10-fold underreporting of the annual CL incidence when consulting national and international experts about their opinion on probable degrees of underreporting for specific countries. The estimates were subsequently also used for countries judged as being similar in their degree of underreporting.

The epidemiology of leishmaniasis is dynamic, with reports of (re-)emergence and spread of the disease, with three major risk factors challenging the control: human-made/environmental changes,

immune status (also discussed in section 3.1.2 Pathology, aetiology and transmission of CL) and treatment failure and drug resistance (Dujardin & Decuypere, 2013).

3.1.2. Pathology, aetiology and transmission of CL

Leishmaniasis is a complex of mammalian diseases that are transmitted by female sand flies. Human leishmaniasis is caused by ca. 20 different species of *Leishmania*. The parasite is obligate intracellular dwelling in cells of the monocytic-phagocytic system of vertebrates. Pathogenesis is caused by cell destruction, the rupture of the 'nests' of the parasite's intracellular stage in the macrophages, connected to an intense inflammatory reaction (A. Ponte-Sucre, 2013).

The most common form, localized CL, is mostly caused by *L. major* and *L. tropica* in the Old World (OWCL), and by *L. (Viannia) braziliensis*, *L. mexicana* and related species in the New World (NWCL). While some forms may be self-limiting, the majority of lesions persist and progress if untreated. Active lesions and scars can be disfiguring and disabling, are a cause of stigma, and can adversely affect peoples' well-being and social interactions. A mucosal involvement (mucosal leishmaniasis (ML) or mucocutaneous leishmaniasis (MCL) may occur in less than 5% of the infections caused by *L .(Viannia) braziliensis*. It causes extensive destruction and distortion of oronasal and pharyngeal cavities incl. the nasal septum and palate, leading to severe facial mutilations. Disseminated cutaneous leishmaniasis (DCL) (a diffuse nodular non-ulcerating form) and leishmaniasis recidivans (localized non-healing lesions) can occur in particular in *L. aethiopica* infections, are difficult to treat and can be severe. Visceral leishmaniasis (VL), caused by *L. donovani* and *L. infantum (*syn. *L. chagasi)*, is characterized by splenomegaly and hepatomegaly, progressive fever and weight loss, hypergammaglobulinemia and pancytopenia and is, if untreated, fatal in the majority of cases (Bern et al., 2008; Boelaert & Sundar, 2014; Guerrant, Walker, & Weller, 2011; A. Ponte-Sucre, 2013).

Defective cell-mediated immune responses caused by concomitant diseases or malnutrition exacerbate the disease and cure. Co-infection with human immunodeficiency virus (HIV) correlates with a rapid increase in the incidence of leishmaniasis and recrudescence after treatment (A. Ponte-

Sucre, 2013). Recent reports have also pointed out an increased incidence of CL in refugee populations in the Middle East, the extent of which is particularly hard to assess due to the catastrophic living conditions of refugees, and the lack of even basic provisions and health care (Du, Hotez, Al-Salem, & Acosta-Serrano, 2016).

The sand flies involved in transmission are representatives of the genus *Lutzomyia* in the New World and *Phlebotomus* in the Old World, both belonging to the family *Psychdidae*. Their habitat ranges from tropical rain forests to very dry regions, and they tolerate altitudes from sea level to 1500 m. Females need blood meals for the maturation of their offspring. After fertilization, between 40 and 70 eggs are deposited in damp dark areas, and the imago develops *via* larval and pupal stages. In anthroponotic foci, transmission occurs between humans *via* sand flies, and in zoonotic foci, the parasite is transmitted between local animal hosts and from these to humans *via* sand flies (A. Ponte-Sucre, 2013).

The *Leishmania* life cycle alternates between a stage in vertebrates as an intracellular form (amastigotes in macrophages) and a stage in the insect vector as an extracellular form (promastigotes in the gut). The insect regurgitates promastigotes when puncturing the skin of the vertebrate host while taking a blood meal. The parasites are recognized by surface receptors of macrophages and dendritic cells of the human immune system and ingested, where they migrate into the phagolysosome and differentiate into amastigotes upon encountering a higher temperature and acidic pH. Amastigotes multiply by binary division, and eventually rupture densely infected macrophages, causing the release and subsequent uptake by macrophages, thus causing the infection to spread within the host. Upon uptake by insect vectors, the amastigotes are transformed into promastigotes in the digestive tract, differentiate into metacyclic promastigotes (the infective parasites) and migrate to the insect's proboscis *via* the heart valve, thus being released into the new host upon taking a blood meal (A. Ponte-Sucre, 2013).

Differential gene expression between the life cycle stages seems to be regulated by two different cues in the environment: temperature and pH, with temperature also driving tissue tropism. Those species causing CL and MCL/ML prefer temperatures below 35° C and multiply in exposed areas of the skin. Species causing VL require a temperature of 37° C for differentiation to amastigotes.

They migrate to the bone marrow, spleen and liver (Aebischer & Mrva, 2016; Bates & Rogers, 2004; A. Ponte-Sucre, 2013; Ridley, 1979).

All *Leishmania* species are morphologically similar, but were shown to have great genomic plasticity, with differences even between individual organisms in the same isolate. They have a unique ability to alter the copy number of individual genes, groups of genes, chromosomes and even the entire genome, allowing for great flexibility with expression of gene products. Therefore, *Leishmania* can be considered aneuploid, rather than diploid, when sampled from the natural population, and this has been suggested as a possible source for drug resistance (Aebischer & Mrva, 2016; A. Ponte-Sucre, 2013; Ubeda et al., 2008). In addition to clonal reproduction as described above, there is evidence for frequent sexual reproduction from natural hybrids, mosaic genotypes and gene flow between populations, and cross-species genetic exchange between visceral and cutaneous strains has been experimentally demonstrated with the extracellular promastigote stages in the sand fly vector (Aebischer & Mrva, 2016; Ponte-Sucre, 2013; Romano et al., 2014; Schönian, Cupolillo, & Mauricio, 2013).

3.1.3. Treatment options for CL

An overview of the treatment options for CL, listed separately for OWCL and NWCL and pharmaceutical and physical therapies, is provided in Table 3 (Monge-Maillo & López-Vélez, 2013). There is great variability depending on the *Leishmania* species; therefore treatments recommendations are often restricted to a limited region. For OWCL, and against the background of a high spontaneous cure rate, local therapies such as thermotherapy, PMM ointments or intralesional antimonials are considered first line treatments in most cases. For NWCL, systemic treatment is preferred, with its efficacy depending on the species and geographical area. Pentavalent antimonials are used mostly, although they often cause side effects. Sb^V are recommended for infections with *L. amazonensis* and *L. (Viannia) peruviana*. Oral treatment with ketoconazole or miltefosine is recommended for *L. mexicana*, and for *L. (Viannia) guyanensis* and *L. (Viannia)*

panamensis, pentamidine and miltefosine could be recommended. For *L. (Viannia) braziliensis*, amphotericin B and LAB are good alternatives.

Table 3: **Overview of CL treatment options used for OWCL and NWCL**, and their strength of recommendations (Monge-Maillo & López-Vélez, 2013). PMM: paromomycin, SSG: sodium stibogluconate, MA: meglumine antimoniate, Sb^v: Pentavalent antimonials LAB: Liposomal amphotericin B. Species abbreviations: L.a.: *L. aethiopica*, L.b.: *L. (Viannia) braziliensis*, L.d.: *L. donovani*, L.g: *L. (Viannia) guyanensis*, L. m.: *L. major*, L.p: *L. panamensis*, L.pe: *L. peruviana,* L.t.: *L. tropica.* SOR/QOE: Strength of recommendation and quality of evidence, following the Infectious Diseases Society of America (IDSA) grade classification.

	Treatment		SOR/QOE	Country and species
Pharmaceutica				
OWCL				
	Topical PMM regimens		AI	Israel (L.m.), Tunisia (L.m.)
			DI	Turkey (L.t.), Iran, Sudan, Tunisia (L.m.,L.t.)
	Intralesional antimonials (SSG or MA)		AI	Saudi Arabia (L.m.), Sri Lanka (L.d.)
			BII	Syria, Iraq, Iran (L.m. or L.t.), mediterranean
			CIII	Other areas (No data for L.a.)
	Topical imidazoles (Miconazole or clotrimazole)		DI	Saudi Arabia (L.m.)
	Azole regimens	Oral fluconazole	AI	Saudi Arabia and Iran (L.m.)
		Oral ketoconazole	AI	Iran and Kuwait (L.m.)
			CI	Turkey (L.t.)
		Oral itraconazole	AI	Iran (L.m.)
			BI	India (L.t.)
	Oral miltefosine		BI	Iran (L.m.)
	Sb^V			
		SSG or MA intramuscular or intravenous	BIII	Iran and Afghanistan (L.m.)
			CIII	Kenya and Sudan (L.t.)
		SSG or MA intramuscular or intravenous and allopurinol	BI	Iran (L.m.)
		SSG and pentoxifylline	BI	Iran (L.m.)
NWCL				
	Topical PMM regimens		BI	Ecuador and Guatemala (L.pa. or L.b. or L.m.)
			BII	Colombia (L.pa.)
	Azole regimens	Oral ketoconazole	BI	Guatemala (L.m.) and Panama (L.pa.)
		Oral itraconazole	DI	Colombia
	Oral miltefosine		BI	Colombia (L.pa.) and Brazil (L.g. or L.b.)
	Azithromycine		DI	Argentina (L.b.)
	Allopurinol		DI	Colombia or Ecuador (L.pa. or L.b.)
	Sb^V			
		SSG or MA	AI	Panama, Colombia and Ecuador (L.pa.), Guatemala (L.b.)
			BII	Brazil (L.g.)
			BI	Peru (L.b.), Panama (L.pa.), Guatemala (L.b.)
		LAB regimens	BII	Bolivia and Brazil (L.b.)
		Pentamidine iseothionate	BI	Colombia (L.pa.)
Physical				
OWCL				
	Thermotherapy		AI	Afghanistan, Iraq, Iran, Kuwait (L.m. or L.t.)
			CI	Afghanistan (L.t.), Iraq, Kuwait (L.m.)
	CO2 laser thermotherapy		AI	Iran (L.m. or L.t.)
	Photodynamic therapy		AI	Iran (L.m. or L.t.)
	Cryotherapy	Liquid nitrogen	BI	Iran (L.m. or L.t.)
			BII	UAE, Jordan (L.m.), Turkey (L.t.)
			BIII	Greece, Israel, Jordan (L.m.)
		Liquid nitrogen in combination with SSG or MA	AI	Iran (L.m. or L.t.)
NWCL				
	Thermotherapy		BI	Guatemala (L.m., L.b.)
			BI	Colombia (L.pa. or L.b. or L.g.)
	Topical nitric oxide		BI	Colombia (L.pa.)
			BII	Brazil (L.b.)
			BIII	French Guyana and Surinam (L.g.)
			DI	Peru (L.b.)
	Intramuscular PMM		BI	Brazil (L.b.)
			DI	Belize and Colombia (L.pa. or L.b.)

Further candidates are being developed by DND*i*: An oxaborole compound (DNDI-6148, with back-ups DNDI-5421 and DNDI-5610 at the lead optimization stage) and a nitroimidazole (DNDI-0690), as well as CpG-D35 as an immunomodulatory adjunct to drug therapy, currently in preclinical development (DNDi, 2016). Therapeutic switching, or repurposing, of drugs originally developed for different indications, can be seen as an important factor driving the antileishmanial drug development pipeline (e.g. PMM, AmpB and MIL) (Shakya, Bajpai, & Gupta, 2011).

The target for all drugs is the amastigote residing in host macrophages. Great differences in drug susceptibility according to species is complemented by host factors, e.g. the immune status, nutrition, age and gender (Maes, Luz, Cos, & Yardley, 2013).

Overall, systematic reviews show that in addition to the limited number of available options, recommendations for CL treatment rely on a weak evidence base with regards to their efficacy and safety. On the other hand, interventions discarded as ineffective in inconclusive studies could prove to have therapeutic benefit if evaluated in an adequate study. A general poor quality, as well as a lack of methodological harmonization in the conduct and analysis of clinical trials of CL interventions was repeatedly pointed out, such as for outcomes and outcome measures and eligibility criteria (González et al., 2009, 2008, 2010; Monge-Maillo & López-Vélez, 2013; Olliaro et al., 2013; Reveiz et al., 2013)In order for both current and newer treatments to be adequately assessed, standardized methodologies are needed which can be applied generally, while at the same time allowing the flexibility required to cover the very diverse disease manifestations of CL. This would improve the quality of clinical research and the usefulness of results for meta-analysis, and would provide clinical investigators with guidance for the design, conduct, analysis and report of future clinical trials of treatments for CL - including the definition of measurable, reproducible and clinically meaningful outcomes. Similar to other NTDs, as only limited resources are available for CL research, it would be beneficial if resources were concentrated in clinical studies of excellence that meet international quality standards.

This issue is addressed by two guidance documents, one on the methodology of clinical trials assessing CL interventions (Olliaro et al., 2013), and a second one in order to harmonize clinical trial methodologies within an extensive network of clinical researchers (Olliaro et al., 2018). Central

to this initiative is an innovative approach that engages with a broader stakeholder community to seek agreement on key methodological questions regarding whom to treat (eligibility criteria) and how to assess and measure treatment effects(outcomes measures), so that results can later be shared or combined for meta-analysis and provide consistent and reliable information for policy decisions. The work presented here focusses on consulting and including patients as stakeholders in defining outcomes for CL intervention trials.

3.1.4. Core Outcome Sets (COS) and Core Eligibility Criteria (CEC)

An outcome in the context of a clinical trial refers to what is being measured, that is, any identified result arising from exposure to a causal factor or a health intervention (Boers et al., 2014; Prinsen et al., 2014). A Core Outcome Set (COS), consisting of Core Outcomes or Core Outcome Measures (COMs), is an agreed minimum set of outcomes that should be measured and reported in clinical trials of a specific disease or trial population (Boers et al., 2014; Clarke, 2007; Williamson, Altman, Blazeby, Clarke, & Gargon, 2012b; Williamson et al., 2012a). A COS ideally also includes the instruments used to measure the COM, if ambiguity exists.

Adopting a COS does not mean that measurements should be restricted to these outcomes, but rather that outcomes contained in the COS will always be collected and reported, complemented by additional ones deemed important by the PI. This makes it all more important to achieve consensus on both the definition and the measurement of the COM. A consensus on a COS helps to make sure that outcomes deemed core by the scientific community are measured and reported during trials.

The COMET (Core Outcomes Measures in Effectiveness Trials) initiative (http://www.comet-initiative.org/) is dedicated to the development of COS. As part of the OMERACT (Outcome Measures in Rheumatology) initiative Boers et al. (2014) suggest a framework for developing a COS for clinical trials, *via* a definition of concepts and core areas. The first concept, *Impact of health conditions*, comprises the core areas *Death*, *Life impact* and *Resource use/Economical impact* and the second concept is *Pathophysiological manifestations*, which is also a core area. The authors suggest

selecting, and obtaining consensus on, outcomes belonging to domains within each of the different core areas, as well as how to measure them (definition of measurement instruments). As part of this framework, the OMERACT Filter 2.0 suggests to validate outcome measurement instruments by using the key properties of *validity*, *reliability and responsiveness*, and *usability*.

Involving patients in the definition of treatment outcomes, though not frequent practice yet, is aligned to recent recommendations (Boers et al., 2014; Sinha, Smyth, & Williamson, 2011; Tugwell et al., 2007) and already proved effective in identifying outcomes not previously identified by other stakeholders (Williamson et al., 2012b, 2012a)

In analogy to a COS, but not previously mentioned in the literature, Core Eligibility Criteria (CECs) are defined in this context as a core set of inclusion and exclusion criteria. They explicitly define the patient spectrum to be included in/excluded from an intervention trial. Adequate definition and reporting are crucial: Overly rigorous eligibility criteria compromise the external validity of a trial, that is, the extent to which its results are generalizable to the overall target patient population. Poorly or too loosely defined eligibility criteria compromise patient safety and make comparison of results between different studies very difficult (Rothwell, 2006).

3.2. Methodology

3.2.1. Study design

This qualitative study was initiated, supported and sponsored by the Special Programme for Research & Training in Tropical Diseases (TDR) in collaboration with DND*i*, and coordinated by the investigator: Study set up and establishment of the international research collaboration will be described in the following chapter. I was responsible for drafting the protocol and supporting documents, incorporating feedback, general project and workshop coordination, and analysis and write-up.

The study workflow is depicted in Figure 14. Individual, semi-structured interviews in regional languages were conducted at eight sites in seven countries endemic for OWCL (Burkina Faso, Iran,

Morocco, and Tunisia) and NWCL (Brazil, two regions in Colombia, and Peru). All sites used equivalent protocols, collaboratively developed during and after the study workshops. Minor adaptations were made depending on the specific ethics committees' requirements or feedback. Interviews elicited the patients' disease history, treatment experiences and attitudes, following an interview topic guide (Appendix 8: Interview topic guide, modified from Erber et al. (2018)). The interviews were conducted, transcribed and translated by the group of PIs based in the endemic regions (local PIs), who are therefore also referred to as 'interviewers' here. Pooled translated transcripts were analysed by me for outcomes for CL treatment that are considered important from a patient's perspective for the purpose of this work.

Additional analyses of findings that have emerged from the interviews and which are relevant to a specific country or regional context are on-going. These are conducted by the local PIs and investigating themes such as preferred modes of administration and accepted trade-offs, traditional treatments used in certain contexts, direct and indirect costs related to diagnosis and treatment, and psychological impact of the disease.

Individual, semi-structured in depth interviews were chosen to allow patients to speak freely about their disease experiences in order to elicit spontaneous, non-biased reports, also of sensitive topics such as stigmatization.

Questions in the interview topic guide are open ended, allowing for further questioning (probing) based on the patient's responses, thereby enabling the interviewer to clarify complex problems associated with a condition, and their inter-relationships. The interviewer can tailor the order, language, phrases and syntax to the cultural and personal context as much as possible, while the topic guide ensured that the same topics were covered across all sites for consistency, thus providing guidance during the interview. (Bowling, 2009; Brédart, Marrel, Abetz-Webb, Lasch, & Acquadro, 2014; Britten, 1995).

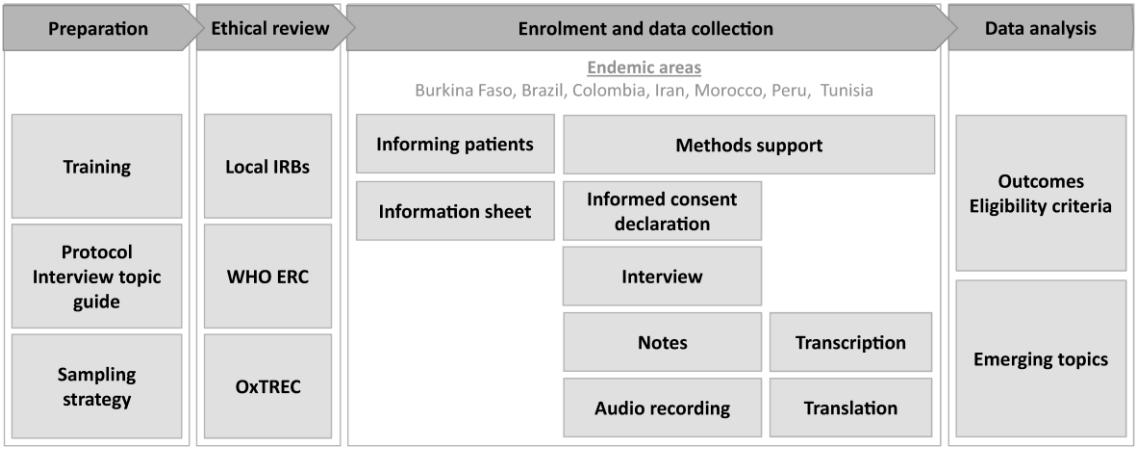


Figure 14: **Study workflow of CL patient interviews**. The master protocol including the interview topic guide and the sampling strategy were developed in collaboration during a workshop (Preparation). Individual sites were responsible for submission of the master protocol and adapted site-specific protocols to the responsible bodies for ethics review (Review). After ethical clearance, the patient interviews and their transcription and translation were conducted locally, with methodological support from an experienced qualitative researcher (Conduct of patient interviews). They were analysed centrally using a collaboratively developed coding framework for outcomes, as well as locally for emerging topics relevant to a specific country or regional context (Analysis).

3.2.2. Establishment of the collaboration and protocol development

Participating experts (the local PIs) were identified through an open competitive call for application published widely on relevant websites (WHO/TDR, 2015), and selected by a review panel. 56 applications in total were received, with 22 applications from NWCL-affected areas, and 34 from OWCL-affected areas. The aim was to cover as much as possible the different forms of CL across the world (OWCL and NWCL). Selection criteria for candidates were the geographical origin (corresponding to specific *Leishmania* species), access to current and prior CL patients, the number and type of CL cases seen per year, the patient spectrum at the site, and available resources and collaborators.

TDR originally supported eight participants, although the Ethiopian PI had to withdraw participation due to a change of institution. A ninth participant (a second site in Colombia) was

supported by DND*i*. Of the selected participants, four are male and four female. All of them (3.2.6 Ethical clearance) are CL researchers and/or HCPs with experience with CL patients.

TDR convened two 5-day training workshops for the PIs to train investigators, and to provide opportunities for collaboration. The first workshop, conducted at the beginning for protocol development, contained training on planning and conducting qualitative health research with a specific focus on interviewing patients. During this workshop, the interview topic guide (Appendix 8: Interview topic guide, modified from Erber et al. (2018)) was also developed and tested. The second workshop contained training on analysis of qualitative data using NVivo® 11 as well as development of a coding framework for data analysis based on three initial interviews. Both trainings were held by the Health Experiences Research Group (HERG) at the Nuffield Department of Primary Care Health Sciences at the University of Oxford.

3.2.3. Sampling

A two-step sampling procedure was based on the assumption that the local PIs, as HCPs and researchers, were able to purposively choose suitable patients most informative to the study objectives, and knew how to approach them and conduct interviews, all appropriate for the specific context.

Aligned with the qualitative nature of the study, a nonprobability sampling approach, that is, a sampling approach not guided by the idea of random selection and statistical representativeness of the sample for the overall population, was chosen, and no sample size calculation was conducted. However, representativeness of the findings for a patient spectrum that is as broad as possible was considered important for this study, despite being rather unusual in the context of qualitative research *per se* (Gobo, 2004; Mays & Pope, 1995; Neuman, 2013). This aspect was covered by using a maximum variation sampling approach, as a variation of purposive sampling (Bowling, 2009).

In a first step, we defined a specific patient spectrum along the characteristics detailed in Figure 15, and sought to obtain maximum variation within this spectrum. A central assumption is that using

this strategy allowed us to obtain diverse input in patient experiences across all sites, in line with the exploratory purpose of the interviews.

a) Gender
b) Age
c) Clinical presentation: Anatomical localization and number of lesions
d) Treatment status:
    a. Patients currently seeking treatment
    b. Patients under treatment
    c. Patients who had been treated previously
e) Patients
    a. Treated as outpatients
    b. Treated as inpatients
f) Leishmania species: *L. major, L. tropica,, L. (Viannia) braziliensis, L. (Viannia) panamensis* and *L. (Viannia) peruviana*
g) Treatment modality (Treatment, route of administration, and dosing regimen)

Figure 15: **Patient spectrum composition.** For patient recruitment to the multicentre study, maximum variation along these characteristics or parameters was sought, thus guiding recruitment at individual sites.

In a second step, this desired patient spectrum composition for the entire study (across all sites) was then translated collaboratively into a sampling approach that is meaningful for each specific region, taking into consideration the typical patient populations at each site (see 8.7

Appendix 7: Background parameters informing the sampling strategy). For instance, for the entire study, we considered an overall gender balance important, but at some sites, predominantly men (e.g. soldiers) were seen as patients, and therefore interviewed.

In each of the eight sites, 9 interviews on average were conducted. The sample size was determined by the available resources. Data saturation, describing a point in qualitative data collection when no new or no more relevant information emerges with respect to the theory guiding the sampling (Given & Saumure, 2008), was deemed as unlikely to be achieved, and not a consideration for this study. This is discussed further in 3.4.4 Limitations of the study.

Only patients above the age of consent in the specific region were interviewed, that is, 18 years of age in all countries except Burkina Faso, where the legal age of consent is 20 years.

3.2.4. Enrolment and data collection

An overview of the interview process can be found in Figure 15.. Participants were identified and invited to participate by the local PIs amongst their patients with a confirmed CL diagnosis (newly diagnosed, on treatment and after treatment). The PIs, as HCPs and/or researchers, are part of the medical team. An overview of recruited patients can be found in Table 4. Overall, 74 patients were recruited, with 50% infected by OWCL and 50% with NWCL. Leishmania species considered were *L. (Viannia) braziliensis* 24%, *L. (Viannia) braziliensis* or *L. (Viannia) peruviana* 9%, *L. (Viannia) panamensis* 16% (NWCL) and *L. major* 14%, *L. major* or *L. tropica* 26%, and *L. tropica* 11 % (OWCL). The overall gender ratio is 38% women and 62% men, corresponding to the expected ratio of 40% women and 60% men (Appendix 7: Background parameters informing the sampling strategy). The average number of lesions across all sites is 3.4, and considerable variation across sites (no data is available from Tunisia and Peru). 5% of patients have had CL diagnosed, 47% were under treatment, 46% had completed their treatment and 1% had remained untreated.

After having been explained the study, and being given the patient information sheet (8.5 Appendix 5: Patient information sheet and informed consent declaration), a separate appointment for the

interview was made. Patients who did not consent to being audio recorded were not interviewed. It was made clear that participation in the interviews was voluntary, and participation or refusal to participate is and completely independent of any current or future treatment, or participation in clinical trials.

Table 4: **Characteristics of patients recruited for the study**

	Brazil	Burkina-Faso	Colombia	Colombia	Iran	Morocco	Peru	Tunisia	Overall
Institution/Study site	Centro de Pesquisa René Rachou (CPqRR), Fundação Oswaldo Cruz (FIOCRUZ) Minas Gerais	Centre MURAZ, Bobo-Dioulasso, Burkina Faso	Centro Internacional de Entrenamiento de Investigaciones Médicas (CIDEIM), Cali	Programa de Estudio y Control de Enfermedades Tropicales (PECET), Medellín	Molecular Dermatology Research Center, Shiraz University of Medical Sciences, Shiraz	National School of Public Health, Rabat, Morocco	Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia, Lima	Institut Pasteur de Tunis, Tunis	
Area where the study is conducted	Belo Horizonte	Bobo-Dioulasso and Ouagadougou	Cali and Tumaco	Leishmaniasis Recovery Center, Boyacá	Shiraz and vicinity	Sefrou, Moulay Yacoub	Lima, San Martin, Cusco	Sidi-Bouzid and Gafsa	
***Leishmania* species**	*L. (Viannia) brazilliensis (L.b.)*	*L. Major (L.m.)*	*L. (Viannia) panamensis (L.pa.)*	*L. (Viannia) braziliensis, L. (Viannia) panamensis*	*L. major, L. tropica (L.t.)*	*L.tropica*	*L. (Viannia) braziliensis and L. (Viannia) peruviana (L.pe.)*	*L. major*	L. b. 24%, L. b. or L. pe. 9%, L. m. 14%, L. m. or L. t. 26%, L. pa. 16%, L. t. 11%
Number of patients	10	10	10	10	10	8	7	9	74
Age range of patients (mean, median) in years	19-71 (42, 41)	24-73 (52, 56)	18-52 (34, 34)	20-32 (25, 25)	21-63 (38, 33)	23-57 (42, 50)	25-68 (42, 37)	27-65 (40, 37)	18-73 (39, 35)
Gender ratio (F:M)	50:50=F:M	40:60=F:M	20:80=F:M	0:100=F:M	50:50=F:M	63:37=F:M	29:71=F:M	56:44=F:M	38:62=F:M
Average number of lesions (mean, median)	1-1 (1, 1)	1-56 (11, 6)	1-4 (1.7, 1)	1-5 (2.4, 1)	1-5 (1.8, 1)	1-6 (2.5, 2)	N/A	N/A	1-56 (3.4, 1)
Treatment status in % (diagnosed:under treatment:completed :untreated)	0:40:60:0	0:0:90:10	20:40:40:0	20:30:50:0	0:70:30:0	0:25:75:0	0:100:0:0	0:89:11:0	5:47:46:1

At the specified appointment, interviews were conducted after patients have had the opportunity to ask any remaining questions, and have given their consent by signature (or appropriate alternative) to the interviewer or a designated person using the Informed consent declaration (8.5 Appendix 5: Patient information sheet and informed consent declaration). Consent was sought corresponding to

local IRB requirements by signature, or a suitable alternative, and in the presence of one or several witnesses if necessary. To obtain consent from participants who were unable to sign, at least one literate witness was chosen from outside the research team. After the participant had given oral consent, the witness(es) signed on his/her behalf. Whenever possible, participants chose the witness(es) themselves. Documents were presented in languages required by the local IRBs, and explained to patients if necessary. Participation in the interviews was voluntary, and patients did not receive any payment for their participation. If applicable, they received compensation for their time, and reimbursement of travel costs.

If possible, the interviews took place in the health care facility, optimally in a separate room to where treatment took place, or in patients' homes, and during an appointment reserved for the interview only.

Interviews were conducted in the patients' mother tongue, or a language that he or she was sufficiently comfortable with. The interviews roughly followed the topic guide (Appendix 8: Interview topic guide, modified from Erber et al. (2018)) which provided guidance for the interviewer to ensure consistency. The interviews with a duration of about one hour were audio recorded, and the interviewers took notes before, during and after the interview. The patients were free to stop at any time during the interview, and to refuse to answer questions they felt uncomfortable with.

Interviews were transcribed verbatim by the interviewers, or designated collaborators, into the original language, and were subsequently translated into English. Translated transcripts, as well as notes, were also shared for analysis. Patients remained anonymous; their names were replaced by an ID during the transcription and their identity was not disclosed for the report and any publications. All investigators, and any collaborators involved in the study, signed a Statement of compliance and confidentiality.

3.2.5. Data analysis

To ensure consistency across all sites, the pooled translated interview transcripts from all sites were analysed by me, with additional information provided through the notes taken by the local PIs and through discussion. I was supervised and supported by an experienced qualitative researcher. NVivo® 11, the OSOP (One Sheet of Paper) approach (Ziebland & McPherson, 2006) and the method described by Green and Thorogood (2014) were used for qualitative data management and thematic analysis. In a first deductive analysis approach, codes were grouped into a pre-defined and collaboratively developed thematic coding frame. The analysis is described in more detail in 3.3.2 Coding framework and thematic analysis in the Results section. This is in line with approaches in qualitative research, where the data guide the analysis.

In addition, further inductive analyses are currently ongoing on a local level by the local PIs, allowing categories to be generated from the data and addressing findings that have emerged on a local level (individual countries, or regions). This will be reported in the section 3.4.5 Relevance to clinical trial design and further research, providing a rich insight into CL patients' experiences in a LMIC context.

3.2.6. Ethical clearance

Ethical clearance of the protocol used a master protocol as well as site-specific protocols over several submission rounds.

In a first step, a master protocol including supporting documents (Patient information sheet, Informed consent declaration and Interview topic guide (Appendix 5: Patient information sheet and informed consent declaration and Appendix 8: Interview topic guide), developed in collaboration, was submitted for ethical clearance to, and received approval, from:

- World Health Organization Research Ethics Review Committee (WHO ERC), Geneva, Switzerland,

as well as the IRBs and ethics committees (ECs) responsible for the respective sites as follows, with the local PIs named in brackets:

- Comité d'éthique de la Faculté de Médecine et du CHU Hassan II Fes, Fez, Morocco (Dr. Issam Bennis)

- Comité d'éthique Biomédicale de l'Institut Pasteur de Tunis, Institut Pasteur de Tunis, Tunis, Tunisia (Prof. Afif Ben Salah, Aicha Boukthir)

- Comité Institucional de Ética de Investigación en Humanos (CIEIH), Centro Internacional de Entrenamiento e Investigaciones Médicas (CIDEIM), Cali, Colombia (Dr. Maria del Mar Castro Noriega)

- Comité d'éthique institutionnel du Centre MURAZ, Bobo-Dioulasso, Burkina Faso (Dr. Mamoudou Cissé)

- Centro de Pesquisa René Rachou (CPqRR), Fundação Oswaldo Cruz (FIOCRUZ), Minas Gerais, Brazil, and Comissão Nacional de Ética em Pesquisa – CONEP, Brazilia, Brazil (Dr. Glaucia Fernandes Cota)

- IRB of Mekelle University, College of Health Sciences, Mekelle, Ethiopia (Dr. Mairie Kebede)

- Shiraz University of Medical Sciences (SUMS) Ethics in Research Committee, Shiraz, Iran (Prof. Farhad Handjani)

- Sede de Investigación Universitaria (CBEIH-SIU) of the University of Antioquia, Medellin, Colombia (Dr. Liliana López Carvajal)

- Comité Institucional de Etica para Humanos, Universidad Peruana Cayetano Heredia (CIEH – UPCH), Lima, Perú. (Prof. Dalila Martinéz Medina)

- Oxford Tropical Research Committee (OxTREC), University of Oxford, Oxford, UK.

In a subsequent step, site-specific protocols based on the approved master protocol and containing modifications and translations required by the individual IRBs were submitted for review and

received approval. The study was classified as minimal risk non-interventional study as it did not include any diagnostic or therapeutic intervention, and due to the type of data being collected.

Data are jointly owned by the local PIs and TDR, subject to a WHO standard Technical Services Agreement (TSA).

3.3. Results

3.3.1. Introduction

This chapter is going to present the results of the analysis for outcomes for CL intervention trials obtained from interviews with 76 patients, 50% of which were male and 50% female, in Burkina Faso, Brazil, two regions in Colombia, Iran, Morocco, Peru, and Tunisia, and compared with those included in trials in the literature.

Interviews were analysed using an iterative thematic analysis approach, *via* a coding framework built up from initial interviews and its subsequent update and adaptation, and the analysis results. For exploring outcomes, data (translated transcripts) were coded to a selected subset of nodes. Emerging overall themes were then derived, supported by quotes, and categorized into domains corresponding to their meaning for patients.

Findings were compared and contrasted with the literature, that is, outcomes used in clinical trials of CL interventions. Several themes corresponded to outcomes used in trials, but many provide a much more detailed pictures of aspects related to quality of life (QoL) that are relevant for patients. In addition, analysis of the interviews yielded some unexpected outcomes. Related findings which were explored in addition to outcomes are reported in 3.3.5 Related findings.

3.3.2. Coding framework and thematic analysis

A draft coding framework for data analysis was developed collaboratively with the local PIs during a workshop (described in 3.2.2 Establishment of the collaboration and protocol development), after

familiarization with the data. The framework was based on three initial interview transcripts, which were read and re-read, and sections organized under different, suitable headings. Care was taken to keep this process as inductive as possible, that means guided by and open towards the data and emerging themes, instead of informed by medical or scientific knowledge about CL. Thematic analysis and the definition of a theme followed Braun and Clarke (2006), and the six distinct steps laid out by the authors. Headings were grouped into 7 broader domains (nodes) *Costs, Symptoms, Impact on daily life, Relationship with health system, Treatment, Knowledge and understanding of disease*, and *Diagnosis*, as well as suitable sub-nodes (Figure 16). Consensus was achieved through discussion. This initial coding framework was then manually entered into NVivo® 11 for coding of all subsequent interviews: Relevant sections in all remaining interviews were grouped under suitable nodes/sub-nodes of the framework, using NVivo® 11.

During analysis, the coding framework was subject to continuous updating: For example, the sub-node *Scar treatment* was added after initial analysis of the Iranian interviews, where this theme emerged as being very important for patients, who also mentioned treatment options for scars. The sub-node *No treatment* was added to the node *Medical* (in the category *Treatment*) to be able to capture experiences of patients who have not received any medical treatment, such as in Burkina Faso.

For identifying outcomes, data in all interviews were coded to nodes *Outcomes*, *Expectations* and *Preferences* (in category *Treatment*), as well as *Scar* (in category *Symptoms*), *Medical* and *Self-treatment* as well as *Side effects*, to capture contextual information.

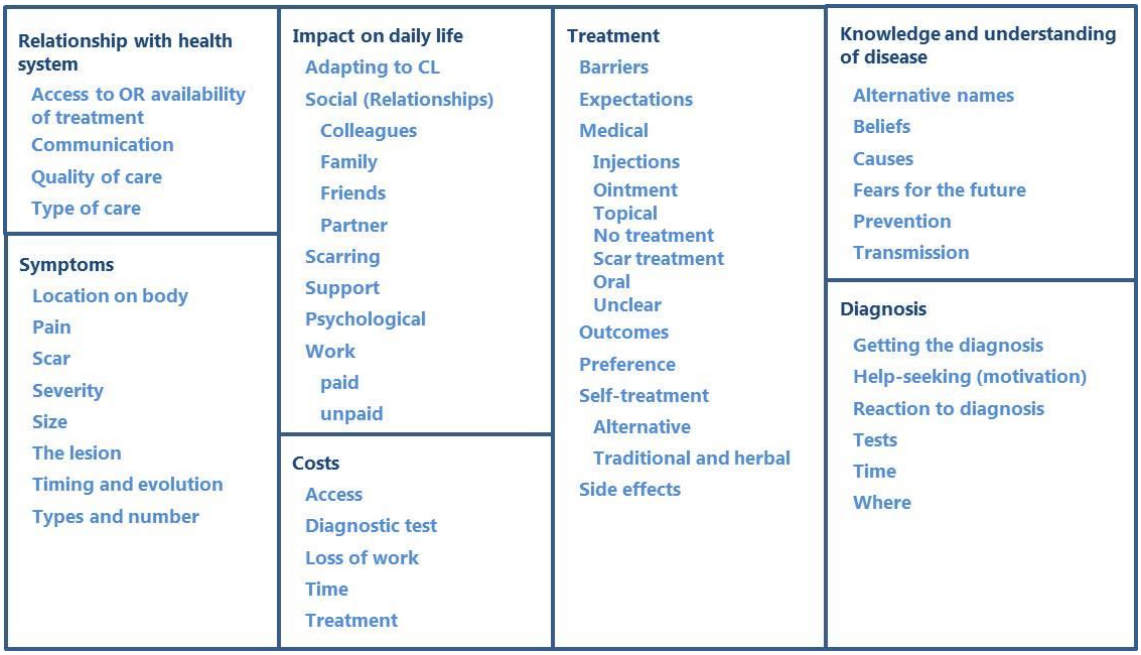


Figure 16: **Coding framework for interview transcripts.** This framework was used in NVivo® 11 for coding interviews, that is, sorting relevant sections. Nodes in dark blue, sub-nodes in blue.

After coding of all interviews using NVivo® 11, data was analysed using pen-and-paper methods as described in Green and Thorogood (2014) and Ziebland and McPherson (2006) and printouts, due to the high volume of data. Transcript sections that had been coded to relevant nodes in NVivo® 11, and could therefore be easily retrieved and printed, were manually arranged into key themes (outcome domains) and subthemes (outcomes) (Figure 17). In a subsequent step, domains and outcomes were assessed for inclusion into clinical trials based on whether they were clinically meaningful and measurable using appropriate instruments; otherwise, alternatives were discussed and suggested.

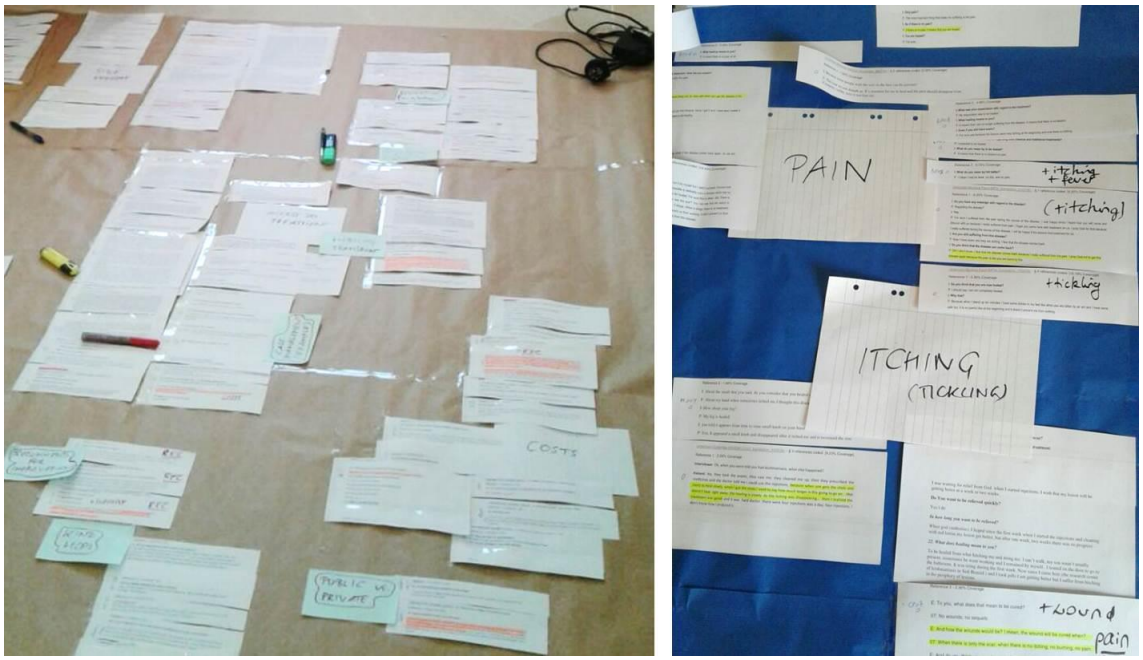


Figure 17: **Analysis of interview data using pen-and-paper** and manual arrangement of excerpts from interview transcripts. The example shown in detail (blue background) is the key theme 'Pain and unpleasant sensations', and development of its subthemes such as 'Itching, stinging or tickling' and 'Unspecified pain', as reported as follows in 3.3.3 Outcomes: The patients' perspectives.

Themes are concepts or propositions that describe, help to interpret and explain aspects of the data. They correspond to the final output of the analysis of the whole dataset (report), where the key themes, or domains, form headers of the individual report sections.: Themes are articulated and developed by comparison between and within interviews, and between countries, specifically looking for emerging patterns, similarities and differences (Gale et al., 2013). Examples for themes include social stigmatization, or fear of sequelae, as discussed in more detail as follows. In addition, for each key theme and subtheme, representative quotes were selected for the report. The notes of the interviewers taken during the interviews and reflections written up afterwards informed the overall process.

3.3.3. Outcomes: The patients' perspectives

Eight key themes and their subthemes, related to outcomes, that is, what patients would like to see or expect from treatment, and what they would not like to see, emerged from analysis of the interviews. They are depicted in Figure 18 and presented in this chapter, supported by quotes. They are reported in no particular order, thus avoiding prioritizing of certain outcomes over others. This was very rich data, and thus much related information and original quotes are deliberately included. Selected quotes are quite lengthy accounts of patients' experiences, but have been left as intact as possible, as they provide a very interesting and often touching insight on how their lives are affected by CL.

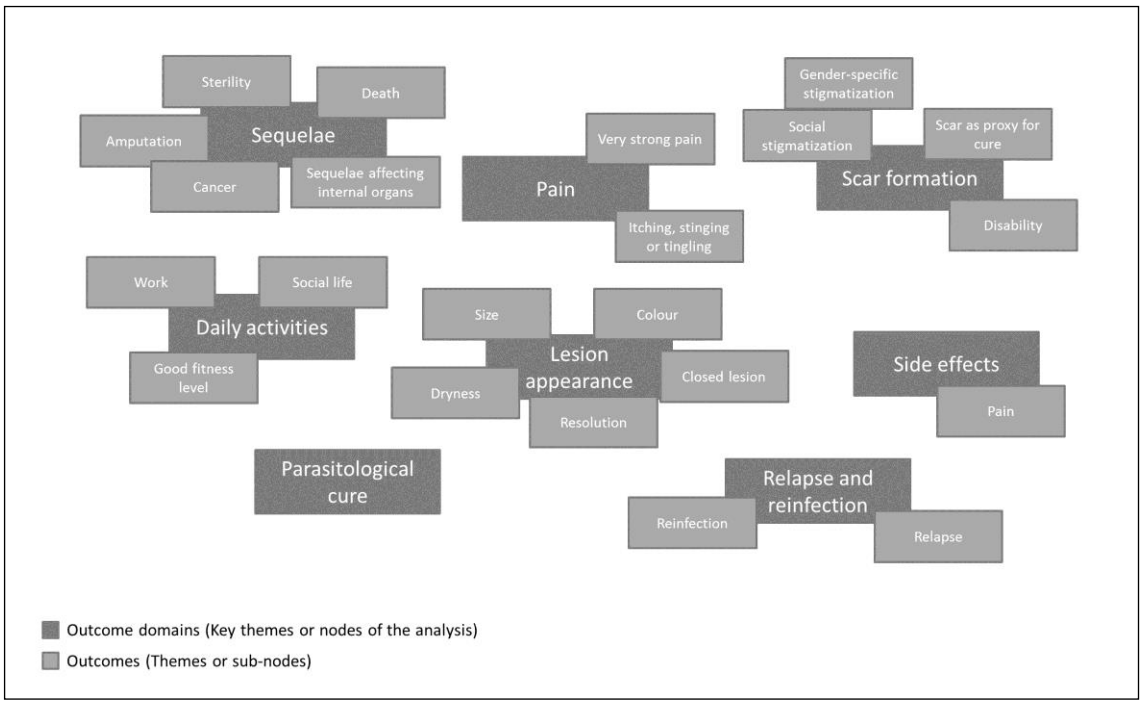


Figure 18: **Key themes and subthemes of outcomes relevant for patients.** Eight general themes were identified from the interviews and termed (high-level) outcome domains, containing individual outcomes: *Sequelae*, *Pain*, *Daily activities*, *Lesion appearance*, *Parasitological cure, Scar formation, Relapse and reinfection* and *Side effects*. Outcome domains and outcomes are discussed in more detail in the text; most of them were found to be highly interrelated. The outcome domain *Side effects* only contains *Pain* as a major outcome mentioned (the remaining ones are discussed in the text).

In the following, the key themes (outcome domains) are numbered, and subthemes (outcomes) are presented with letters. The transcript conventions follow guidance given in Green and Thorogood (2014) and as laid out in 2.2.4.4 Qualitative data analysis.

1. Sequelae: Patients saw successful treatment as avoiding sequelae that could be for example large lesions leading to amputation of limbs, disease progression to affect internal organs, cancer and even death.

    a) Amputation: If not successfully treated, several patients feared amputation of limbs as potential complication, such as this man from Brazil:

    […] For me it is a bad bug starving, it follows eating, eating and eating. […] The injury will go deeper and it can take a nerve, one may lose the movement and may need to have to amputate a limb! (BR05)

    For a diabetic young woman from Tunisia, her fear of amputation of a leg due to CL is so strong that she reported suicidal thoughts:

    A catastrophe [...] in my life named leishmanias I swear. I can't talk. I'm thinking how my leg could be amputated. How my life will look like after amputating my leg, I think about suicide at the same time I think about my daughter how she will calm down after my death. I can't live without my leg. (TU04)

    b) Sequelae affecting internal organs: Several patients in Brazil, Colombia and Morocco mentioned the possibility of the disease affecting internal organs. This could indicate a fear of the visceral form, VL, which is co-endemic in these countries, although typically in different populations, with VL e.g. mainly occurring in children and immunocompromised adults in Colombia (L. López, personal communication). The vagueness of description indicates that not much is known about VL. An atypical form of CL caused by *L. infantum* (syn. *L. chagasi*) and presenting with non-ulcerating lesion that potentially evolves into a visceral form in patients with a suppressed immune system has been described from

Nicaragua and Venezuela, mostly in children older than five years, and young adults (Belli et al., 1999; De Lima et al., 2009; Reveiz et al., 2013).

This man from Brazil was afraid of the disease affecting internal organs:

I think that cure would be the decontamination of the infection. The most important thing is the elimination of the risk of compromising the internal organs (BR09)

c) Cancer: While seeking treatment at the health centre, someone there who remained unspecified mentioned to this Moroccan patient that CL could develop into cancer if left untreated over several years:

I: What do you mean by it heals? R: I was afraid on the beginning when they said to me the disease stayed two months on me before appearing. He said to me if someone didn't treat it, it could change to a cancer. I: Is someone on the health centre who told you if you will not take this treatment it will maybe change to a cancer? R: Yes it will change to a cancer that what he said. But it will not change to a cancer on the same year. He said to me it possible to change to a cancer with years and he said I had a lucky chance to not get the disease inside me [visceral form]. This form of disease is the most dangerous one. (MO08)

d) Death: Patients feared death as potential consequence of CL, without specifying where this attitude might come from.

For example, a female patient from Larama in Burkina Faso was afraid that CL could kill her. This might reflect upon Larama being an emerging disease focus of *L. major* at the time of the 2013 outbreak, with CL being unknown to the population at that time.

e) Sterility: A number of male patients in Colombia being treated with Sb^V or pentamidine isethionate talked about their fears that CL treatment could leave them sterile, that is, unable to father children. This is not among the documented adverse effects for either drug. Patients referred to rumours they have heard, or someone telling them about it.

R: Many people say you cannot have children. So that could be it. I: Who says that? Who are those? R: The ones who have had the illness. Those are comments you hear. From that to reality, I'd say there is a 20% chance of that to happen. So that doesn't affect me at all. I: Do you know about a case of someone becoming sterile because of this? R: No. Only comments. I: But who make the comments? R: Friends. I: Or people to the medics? R: No. Friends always say that:

You have children already? No. Oh boy, you lost your chance now. [laughs] I don't pay attention to them though. It's all because the injection is so strong they think it can affect. (CP10)

[The patient is talking about someone he knows] And then all those shots, I tell you doctor, [...], I think those shots kind of make the person sterile, because the man has had women but he has not had any children. So they are saying that the problem of injections, as he took two wounds of that stuff, then they gave him about 200 injections. [...] He has been with several women and has not had any children. And they're saying that this problem was caused by the injections that they gave him. [...] Women who had been with him left him, they have looked for other men, other husbands, have children with them, but not with him. (CC03)

Five out of the six male Colombian patients mentioning fear of infertility were soldiers with the Colombian army. This interesting finding could indicate a cultural notion related to a certain profession, or gender. Fertility-related behaviours - including choice of sexual partners, use of contraception, and fathering a child - are common social domains for the performance of masculinity (Pierotti, 2013). As a significant life event, childbearing can be integral to the production of a masculine identity, and the inability to do so could threaten that identity. For example, a US-based research project found that one fifth of young male teenagers believe that impregnating a woman would make them feel 'more like a man' (Morrell, 2006). For Colombia, the local PI pointed out the importance of the childbearing ability in the local context. She indicated that soldiers often remained on duty for 8 months continuously before seeing their wives or girlfriends, and might not be able to father offspring in the limited time available. The PI noted in addition that, since many of the soldiers were very young and undisciplined, their commanders were struggling to keep them in the military base while on treatment; the commanders therefore told them having sexual relations while on treatment could confer sterility. Sterility as possible sequel encountered in the interviews could be one of several variations of this rumour (L. López, personal communication). A Colombian anthropologist, working at the same site for her research on armed conflict and CL in Colombia, confirmed this finding and mentioned that while working with the soldiers there she often encountered this fear of infertility herself (L. Pinto García, personal communication). In summary, the medically unfounded fears surrounding male infertility as possible sequel of CL

treatment could be due to interrelated rumours in a very male, and male dominated, culture of machismo in the Colombian army, and there are indications that these rumours might have been produced on purpose.

2. Pain and unpleasant sensations: Itching, stinging or tickling sensations or unspecified pain, compared to e.g. being burnt by fire, was mentioned repeatedly by patients as a symptom.

    a) Itching, stinging or tickling: Itching, stinging or tickling was described by several patients as a symptom of CL; disappearance of these sensations was associated with healing.

    About my hand when [it] sometimes itched me I thought this disease is not yet healed. (MO04)

    b) Very strong pain: Unspecified, very strong pain was mentioned by several patients, and most notably by the majority of patients in Burkina Faso. Patients described it differently: One female patient describes it as like 'being burnt by fire', rendering her unable to work, or a male patient compares it to chili pepper being put into the lesion.

    I went to the health centre and they told me that there is no treatment for this disease. Lesions became painful and I went to health centre again and they told me that there is no treatment for this disease. Lesions were painful and itching. […] I didn't work in my farm because of the pain that is like someone burnt by fire. I didn't get anything in that year from my farm. […] I was happy since I heard that you will come and discuss with us because I really suffered from pain. I hope you came here with treatment for us. (BF02)

    The lesions were very painful and the pain was not like you have been cut by a knife. If it was like that many people would have died. But it was like you have put chili pepper into the lesions. (BF04)

    When discussing this strong emphasis on pain among patients in Burkina Faso with the local PI, he noted that it occurred before traditional treatments were administered, and saw it as part of the natural evolution of the disease. Results should be interpreted against the background that patients were from an emerging

CL focus and have never experienced symptoms before (M. Cissé, personal communication). CL in new foci tends to be more aggressive (Olliaro et al., 2013).

In contrast, other patients emphasized that CL itself was painless; pain was experienced as treatment side effect by this patient in Colombia, where CL has been endemic for a long time:

From pain, because as I'd told you, the illness does not cause any pain, but the treatment itself does. (CC01)

CL lesions (nodules and ulcers) are typically not painful, unless secondary bacterial infection takes place. It is common, leading to pain and serious disability (González et al., 2008).

3. Daily activities: Patients mentioned that being cured for them means being able to perform work, to have a normal social life without their family being affected, and a good fitness level. These rather heterogeneous themes were grouped under *Daily activities*.

   a) Work: Being able to go back to work, either paid or supporting family was seen as important, as it was associated with being healthy again.

   This young Colombian soldier described the work he was doing with his father, and how important it was for him to be able to help him after he's cured again:

   Yes, I was helping my dad to bring lumber from the forest, and as we had built on the beach, we have cottages that we built there; […] I: How did the skin lesions affect the other aspects of your life? I mean, in any other way, how was your life besides work? With family and friends. R: Oh for sure, everything. I could not do anything, I could not go out in the cold evenings, I could not walk, I could not go to anywhere. I was totally useless. […] I: […] What do you expect from the treatment? What does it mean for you to be cured of Leishmaniasis? R: […] I need many things, mostly to help my dad and be there, we are building other cabins, and he has not started because of me, because I'm sick. (CC02)

   This male patient in Burkina Faso reported being unable to work while sick, thereby decreasing his income to an extent that he was unable to feed his family, and therefore noted:

   [Healing] means that you got a good recovery and your family is no longer suffering and your kids too. (BF09).

b) Social life: For several patients, to be able to go out and have a normal social life, free from stigmatization, was a very important component of being healed.

To experience social stigmatization, or the fear of stigmatization, was a common theme emerging.

A very clear and open account of social stigmatization is from a female patient from Larama in Burkina Faso. She had about 50 CL lesions disseminated on her body, all of which resulted in scars. She told her story as to how she was unable to work, and to go to the mosque or to the market. In addition, she describes being stigmatized by her family as well as by her village. This turned out to be mainly due to her husband, who, ignorant about the mode of transmission, blamed an ointment and his wife's ethnic group for it. This turned out to be unfounded when he contracted the disease himself. In contrast, she experienced great support from her friends in the village:

I have really suffered during the course of this disease because in that year I didn't work. I suffered from fever. I couldn't eat. People were laughing at me. They said this is my ethnic group's disease. This has really disappointed me. There are some people who said that this is "BF03's disease" [using the patient's real name]. I suffered from that mockery. They said that I am the one who has brought the disease in the village. I stayed praying God. Then after [this], the disease has touched several persons in the village even in my own family but no one died. Today no one is suffering from this disease but I am afraid that it comes back again. I couldn't approach people because of the bad smell of the lesions even my own family. For example I couldn't eat with my family. I stayed alone. […] My son was not very happy with his father because he is the one who stigmatized my disease at home and in the village. In addition my husband said that my disease worsened because I used medical treatment. I was very sad because of my husband's bad thoughts and mockery. I slept with one of my kids and he didn't get the disease and then I think that this disease is not contagious. […] They [my friends] were very compassionate with me. They used to pay me visit, they talked with me and encouraged. They advised me not to forgive going to the mosque because of this disease. I was finally embarrassed by their visits but they told me that it is mandatory for them as friends to help. (BF03)

This indicates that the population of Larama, an emerging disease focus, was not sufficiently familiar with CL to know about the mode of transmission, and treatment options, at the time of the CL outbreak in 2013. The bad smell of the lesions that BF03 mentioned – which is not usually symptomatic for CL - could be due to her using local tree bark (*Khaya senegalensis)* and leaves as traditional

treatment in the evenings, causing the lesions to be suppurated the following morning.

c) Good fitness level: This was an important outcome in particular for several soldiers being interviewed in Colombia. This could be due to the physical activity soldiers have to perform, thus being associated with good health.

4. Lesion appearance: Different components of the lesion appearance were considered as important outcomes of treatment, mostly whether the lesion was closed, its dryness, its size, and its colour, and often a combination of these elements.

a) Closed lesion: This corresponds to re-epithelialization in medical terms, that is, covered with an intact epithelium.

[Healing means] let the wound close, heal. [...] covered, that's what I imagine. (PE05)

b) Dryness: Several patients mentioned a dry wound indicating cure, often in combination with its size.

c) Size: A reduction in lesion size was seen as an indicator of cure.

[Healing means] It's a god's blessing, when my sore shrunk and dry (TU06)

d) Colour: The skin colour, or absence of bruises, was mentioned by some patients as an expectation after successful treatment.

[After treatment I expect the lesions] at least to return to the same skin colour, right? Because I see it all bruised (PE03)

e) Resolution: Often, patients mentioned 'normal skin' as an outcome, but did not specify it further.

I: When you started injections, what did you expect as a result? R: I guess it will disappear totally. I will not see any knob on it and my skin will return normal. (MO13)

5. Parasitological cure: Tests for the absence of the pathogen, and, importantly, confirmation by a HCP were seen as a relevant outcome by many patients. The pathogen was in some cases seen as a virus, parasite, or bacterium.

   Cured, well, to know that the parasite is eliminated from the body a hundred percent, basically. Regardless of the scars left. That is the least of my worries. That will be something one can handle as time goes by. [So my main concern is] that there aren't any creatures in my body. (CC01)

   [...] I would get tested to know if the virus is still there or not. [...] Well the essential is that there is no longer the virus. (PE04)

6. Relapse and reinfection: Fear of relapse and reinfection was an issue for many patients, and in most cases patients were clearly able to distinguish between relapse, with the lesions occurring in the same skin area, and reinfection by a sand fly, in particular if they went back to the endemic areas.

   New world species *Leishmania* parasites have been shown to persist in humans after clinical cure: As DNA in peripheral blood (*L.* (*Viannia*) *braziliensis*) (Delgado, Guevara, Silva, Belfort, & Ramirez, 1996) and scars (*L.* (*Viannia*)) (Schubach et al., 1998) of patients. Several studies provide evidence for the viability and infectivity of the parasite in some specimens isolated after clinical cure: Old world species (*L. infantum*) (syn. *L. chagasi*) (le Fichoux et al., 1999; Morales et al., 2002)), as well as new world species (*L. (Viannia)* (Mendonça et al., 2004)), strongly suggesting that persistence of parasites in both old world and new world species might be the rule rather than the exception.

   Clinically, several forms of leishmaniasis can occur after healing: Leishmaniasis recidivans is a chronic form of CL due to *L. tropica* or *L. (Viannia) braziliensis* occurring 1-2 years after healing of the initial lesion, usually in the face, causing microsatellite and confluent lesions that finally ulcerate at the border of the initial scars. Mucocutaneous leishmaniasis (MCL), caused by *L. (Viannia) braziliensis* and other new world species, evolves after cure of CL and a variable period of latency, and affects nasal and oral mucosa, potentially leading to disfiguring mutilations due to widespread tissue necrosis. In OWCL, involvement of the mucosa is very rare. Not directly a relapse form of CL, but worth mentioning, is post kala-azar dermal leishmaniasis (PKDL), which is a form of diffuse CL and a sequel of VL,

potentially appearing after up to 20 years in individuals that have been untreated, partially or even adequately treated (Boelaert & Sundar, 2014; González et al., 2008).

Leishmaniasis is also a common opportunistic infection of HIV, in addition rendering the distinction between viscerotropic species (causing VL) and dermatotropic species (causing CL) less valid (Boelaert & Sundar, 2014).

This patient from Larama, an emerging focus of CL in Burkina Faso, talks about his concerns on an epidemiological (village) level as well as on an individual level, and he is afraid that his scars could turn into active lesions again:

I: Do you think that this disease can come back again? P: Oh! I pray God so that this disease never comes [back again]. [...] This disease has appeared suddenly and has disappeared now. It could come back again. [...] I want God to eradicate definitely this disease here [in this village] so that people can be in good health and then this disease can't come back again. [...] I can say that it is not still present. My lesions have been healed but I am afraid that the scars become lesions. (BF04).

a) Relapse: Patients felt very concerned, and uncertainty at the same time about the possibility of a relapse, and some specifically mentioned that they did not feel cured as long as there was this possibility.

   I consider it [late complications and relapse] and it makes me very worried. My biggest concern about the disease is it. According to the search I performed, future lesions can appear and mucosal lesions as well. My biggest concern is that today: the disease relapse. It is very important. [...] and, about the risk of future lesions, we can't even know. It is a concern that I have, the greatest concern, but I have no idea whether I'm or not cured. (BR03)

   I'm not sure actually but people say it may start [relapse] again. Because the last time I went for cryotherapy, they told me: come to us if you see any signs again. [...] Yeah I'm always afraid of that. (IR07)

   Some people say leishmaniasis "bursts" after 5 years again. It's like a myth, right? Something the patients say. You see how it bursts again in many cases. [...] It appears in other part, or it appears again in the same place, or simultaneously in several parts. Then you have to go for a second treatment. [...] Because I've seen men who after 4 or 5 years after the treatment, the have the disease again. So I think the parasite remains somewhere. (CP02)

b) Reinfection: Often, patients were aware that they could be reinfected after having successfully completed treatment, and were afraid of it happening.

   In particular the soldiers interviewed in Colombia were well informed about the danger of reinfection if they went back to the endemic regions, and often mention

having seen colleagues in the army getting infected multiple times. The local PI pointed out that, in this context, CL is seen as an occupational disease, that is, a disease related to the army profession (L. López, personal communication).

I had it once. Then a second time. I mean, the treatment cures you permanently, but if you keep going to the endemic zone, then you catch it again. (CP08)

One patient talked about rumours that traditional treatments, often used in areas where patients were unable to obtain treatment in health centres, would also prevent reinfection.

I've been told that there are people who can cure that, they know their remedies, their tricks and that whoever they cure never get sick again. That is what I have heard. And they say that there are people out there in the Caquetá, back in Los Llanos, Putumayo; that there are people who have been affected by that and that they cannot get to a health center, and that they attend prayer leader, shamans, as they are called; and that they get healed and never get anything back. Well that's what I've heard, I do not know how truth that is. (CC09)

Similarly, an elderly male patient from Iran discusses possible immunity to reinfection after having healed by itself, after around one year. This is also reflected in the traditional name *Salak* (in Persian *Sal* means *year*) given to CL:

I: Okay, what do you think will happen to your disease finally? R: Well according to experiences, the reason that they named it Salak that it lasts one year and then it will get healed. [...] It will get healed and then one's body is immune to CL. I wanted to ask you this question that if it naturally heals itself…[...] I: Yes if it naturally …R: So you hope that it will get healed after one year? R: Because it's called Salak. [...]As I said, I heard this from people. (IR03)

Patients are concerned about reinfection, and talk about rumours that treatments could prevent reinfection.

[...] I always have this concern that it won't get infected again... everywhere I go I have this concern that it can get infected, like the water (he means the pool) I said before. I always take care of it. I should use tetracycline to prevent infection. So the first thing that matters is that first it gets well soon. (IR04)

7. <u>Scar formation:</u>

In general, scars at skin areas not visible were not seen as great problems by most patients. If in visible areas, many patients expressed their fear of social stigmatization, and several persons emphasized gender-specific aspects, that is, the problems a visible scar, e.g. in the

face, could pose for unmarried women. A second notion of the scar was that its absence was seen as a proxy of being healed.

Some patients, predominantly men, did not see their scars as problematic regardless of the localization. For example, two Colombian soldiers had other scars from a machete and gunpowder, respectively, already, and seemed to even be proud of them, showing them to the interviewer.

[The remaining scars are] normal. I have other scars from other accidents. For example, this was a deep cut with a machete. (CP09)

This Tunisian man with a scar on his shoulder, talks about being so happy to be healed, and does not care much about the scar at all:

No they are completely healed but the scar remained. The most important thing that it is over, I don't care about the scar. I'll marry another woman! I'll undress my shoulders, I don't care. (TU02)

In some regions where it's available, patients freely talk about the availability of scar treatments, for example in Iran and Tunisia, where cosmetic surgery seems to be commonly available. However, the costs are not covered by insurance, and pose a financial problem to many patients.

An elderly male patient at Sidi Bouzid, Tunisia, talks about his scar, which does not seem to bother him at all, even in the face. He discusses availability of plastic surgery, and the associated costs.

The scar is normal for me even if it's here [in the face]. [If my daughter had a lesion or two in her face] If she can she should do a plastic surgery. I know a woman who had a lesion [a beauty spot] in her face it looks at an olive pill, she removed it with a plastic surgery. In my point of view they can removed the scar, it's a usual act. If I'll have a scar I'll do a plastic surgery. If they have money they could do the plastic surgery, poor people they have [to wait] because plastic surgery needs money. (TU02)

This male patient from Iran talks about the prohibitively high costs of scar treatment. He also mentions an interesting meaning of a scar as spiritual damage.

The laser's expenses are also so high that I have given up, the scar [pause] is also a spiritual damage, also a so-called… loss. [...]The scar [thinking]... well in my opinion, as I said there is laser therapy which is expensive [pause] but maybe there may be an organization, person, or ministry that accepts to give credit to CL patients. (IR01)

a. Disability: As with his wound, this young male soldier from Colombia suffered a lot from its location very close to his eye, and is afraid that the complications and pain would continue with a scar resulting from that wound:

And now, that I got here it began to expand and it's more painful, my hand has swollen up, my eye; my eye hurts a lot, that's more than anything. Most than anything, I am very concerned about the eye, concern about losing it, or something like that.(CC02)

b. Social stigmatization: Many patients report being affected by social stigmatization with a visible scar. This ranges from embarrassment about others asking about the origin of the scar, to being ridiculed; whereas others report fears of not finding a job with a visible scar.

This young man from Morocco freely talks about the role of the social environment with respect to scars, which could provide stigmatization as well as support to affected patients:

With complete sincerity, when I see someone with a scar on his face I avoided asking him questions to not hurt him, even if he has a large scar I told him comparing to other persons [it] is nothing. The society is unforgiving, you can hear from your beloved words that hurt you psychologically but on the other side there are words that can reassure and support you. If you're not able to help someone at least tell him good words. For example, saying there is a new medication and new surgery; or perhaps a new therapy exists abroad and Morocco is the first in the African and Arabic countries that evolved in medicine. [...] As I explained before, our society is unforgiving, undereducated and close-minded. (MO02)

This elderly man from Burkina Faso explains

I: I can see a scar on your left arm, what do you think about that? P: What can I do to remove it? You know, it is not possible to remove this scar from my arm unless you have a treatment for that or God removes it one day. I: Does this scar cause you any trouble? R: For sure no. I: Is it embarrassing for you if someone sees the scar on your arm? R: He can't see it since it is protected. I: How would you react if it was located on your face? R: For sure it would be embarrassing. I: How? R: People would say what happens to this man and you are obliged to tell your story. This is troublesome but if it is hidden no one will react. (BF01)

A female Brazilian patient discusses her scars, and that she would cover them in order to not having to explain to others where the scars came from, rather than being stigmatized:

I'm not going to wear long clothes ... or rather, if I go to a party using long clothing [it] is to avoid to be explaining to everyone. The other day I went to a wedding, I used a long dress. For this reason, it is not because I was worried about people avoiding me or hating me. I did this to avoid to be explaining all the time... [...]It's all the same: explaining, explaining, explaining. It's not that I'm ignorant, but it's kind of weird, huh? If you have time, you explain and everyone understands, but one comes and make a question, another question comes ... it's not that you don't want to talk ...but…. (BR06)

By contrast, this Peruvian patient explains neither being ashamed, nor being afraid to be stigmatized:

I: And, do you mind that scar? R: No doctor, normal, in my work. I: And, what about the one that will be on your face? R: The same. I: And is there no problem? R: No problem. What can I do doctor? I: Don't you mind? Are you not ashamed? R: No, no. I: Some people feel ashamed; they do not want to go out. R: But, since I was a boy, and above all I'm on my land, all my countrymen know me, and we know each other, what can I do? It will be the same. (PE07)

A woman from Tunisia who is currently unemployed worries about finding a job if the sores were located in the face:

Thanks god my sores are in a hidden location. If it is in my face it will be a big problem it could prevent me to find a job. (TU07)

c. <u>Gender-specific stigmatization:</u> Several patients point out how a scar affects girls or women more than men, specifically if the scars are located in their face or other visible areas. All except one of these concerns are from the old world, where it's often pointed out that if the girl is still unmarried, it could pose problems in particular.

The only exception, as an example from the new world, is this Peruvian patient talking about his daughter, with visible scars on her cheeks, and his fear of her being bullied. He explains to consider surgery for her:

[...] We were just talking about that, we would expect a year to decide for surgery, because as she's a girl would perhaps disfigure her face. [...] At school, right? So her friends, classmates maybe, they can put she a nickname. Or perhaps they can bully her, perhaps mistreat her psychologically. I: Until now have they done her something? Has she been bullied or discriminated? P: So far not yet [she has not been bullied or discriminated], but in the future it could be. (PE04)

A Tunisian man points out how a friend's daughter undergoes laser treatment for her scars which are located in the face:

Actually I didn't think about the scar because I am a man and my sores are in my leg it doesn't matter for me, but when beautiful women get a scar in her face it's a big deal. The daughter of my friend get a leishmaniasis sore which left a scar in her face now she is treating the scar by laser, a one session costs 150TD, not every patient could afford that. [...] He told me that she is doing better but she still need more laser session to reduce the scar. There are patients who are completely disfigured. (TU06)

A young man from Morocco talks about the problem that scars pose for young women. He points out psychological problems due to deformations, and compares the CL scar to deformed teeth:

Maybe the [scar] problem for her is more serious than mine, because women in general do not like to be affected by a scar, especially in their face. She will have a serious problem which can lead to psychological complexes in her life. [...] She can become psychologically sick because of the scar, I know people who have a problem by The Will of God. For example, a girl with deformed teeth and defected look, it makes her psychically sick, although for the teeth there is a solution because the medical science became now more developed. (MO02)

He continues explaining the impact he thinks his scar will have on his relationships with women:

I: Concerning your relationship with people? For example, a girl who meets you, will [she] be afraid? R: I can convince her by giving details that it is a disease and I [that] followed the treatment prescribed by the doctor, to not be scared any more. I: If this disease is not located on your face, you will not have any problem? R: I will not have any problem, especially when the doctor reassured me that I will have a small scar like a black spot on my forehead while unlucky others have this thing on their faces. I: if the scar will remain, you may have any problem in your work? R: I don't understand, do you mean I must keep my basketball cap whole working day to hide the scar? No, I can't do it all the day. This kind of idea does not exist in my surroundings. (MO02)

An elderly Moroccan woman with a scar on her cheek contrasts her situation with that of a young, unmarried girl, and society's potential reaction. Towards the end she teases the (male) interviewer when being asked about the means women use to hide their scars:

I: Did you apply anything to treat the dark scar on your cheek? [...] R: No, I didn't apply anything. People with white skin and black scar could worry, but me why would I be worried. My skin colour and the scar are similar (laughs). Thanks to God (laughs). I: If we hypothesize that a young girl was affected like you. R: If she is still young and unmarried [she] should ablate it, but I'm an old woman, there was no need to remove it. [...] I: How was your relative's behaviour towards you? R: Normal, thanks God. If I was young and unmarried it might perhaps be a different behaviour, a girl is not like an old woman. [...] I: What can women do to hide the scar? R: You know what women do, I have nothing to explain to you (laughs). I: I would not ask you if I knew it. R: I don't know, she will do all things that she is able to do (laughs). (MO03)

A woman from Iran, with three scars on her face and hands, tells her story, and how the scar from CL she had when she was 5 or 6 years old has severely affected her social life and two marriages. She describes getting a divorce after getting married at the age of 17, and her husband insulting her scars during an argument. She remarried, and despite her husband not judging her, she perceives herself as inferior. Despite having been married for around 5 years , she has not shown her face to her husband without make up on. .She has suffered from stigmatization due to her CL scars, and sees in laser therapy an opportunity to eradicate her scar, after previous attempts had been unsuccessful, having resulted in even worse scars. She discusses that the CL scar affects her specifically, and she would not mind a scar looking different (e.g. one left by a knife cut across her face).

I was 5 years old when I got CL and didn't know what it was. We lived in Lamerd [a city 383 km from Shiraz] at that time. [...] All the other children had it on their hands or feet. Me and my sister had it on our face... [...] But there was not a hospital at that time to treat this. They used different drugs on our face. Well it was finished and it got better and emm... still I... [...] I got married at 17. Before I was 17, my cousins and all the children mocked me. When we were sleeping I told my sister that I wish it would disappear in the morning. [...] Yeah my sister also had it. [...] I didn't know that it's something that has happened to so many people. We didn't have much money at that time, we lived in a crowded family. I wasn't supported financially so I could not do surgery or something. [...] Well I got married at 17 and I got divorced because of this issue. He told me something about my face [she had a scar on the face at that time] and I finally got separated when we were engaged.[...] I got married again, and my husband's skin is white without any spots. I had to use creams since I was a kid. [...] Yeah, I married twice and I didn't have a wedding ceremony both times due to my CL.[...] Because they took films at weddings and the camera takes a close shot from your face and I didn't want that... my sister also doesn't want to have a wedding ceremony. She has been engaged for 3 years now, her fiancé says that we should have a wedding but she says that the camera takes a close shot of my face, I'll be embarrassed in front of people....[...] We'll be mocked by people. The boys in our family can't understand... their skin is smooth... they were all sent abroad when they were kids... so they didn't know what it was and they all mocked us.[...] Even now when I go outside, many people who don't know CL, look at me in another way... it's like I'm weird. Sometimes that my clothes are not suitable, I think they are looking at my clothes but actually they are looking at my face to see what it is. As much as I remember, I put on makeup all the time..[...] I married when I was 17 and got divorced at 18 and I immediately performed the surgery after that. Because I felt too much pressure, I felt insulted.[...] he mocked my face when we had a fight.[...] Well thank God my current husband has no problem with it. He even told me to forget about it... well people are different.[...] My current husband is from Shiraz and although he is superior to me at all levels... his skin is fine, he is tall, he has a good family.... but he has no problem with it. But I've got tired. I have a kid..... Wherever I go, people first look at my face. [...] It's been 4.5-5 years that I married my current husband... emmm... I think yesterday was the first time that he saw me without makeup [pause] I feel ashamed to look in his face... I looked down. I sleep at night with my day makeup, I don't want that... [...] I think that if I wash my face and look ugly he might leave me too... this will be my second divorce....[...] It doesn't look good. People may think that my first divorce was also my fault, not my husband's. [...] I hate makeup. I wore makeup and used creams and

beauty supplies since I was a kid so much that I hate them. I want to go outside but without makeup. [...] Now that I performed this surgery, I'm really pleased.[...] I don't care at all [about the scar that remains]. I told you the other day.... you can pick a knife and draw it on my face from top to bottom, I don't care. [...] It's important that the CL won't be evident. [...] I don't want CL on my face. [...] Yeah the scar should go away. Maybe the line on my face is important to you, because your own face is smooth or you may want to do laser therapy. But I don't care about this line and I won't do laser therapy. (IR08)

d. Scar as proxy for cure: Several patients saw the absence of scars as a sign that the disease is cured, either by reminding them of CL, or they felt that the skin area was particularly vulnerable to becoming a wound again.

This Brazilian woman mentions, and reiterates, that a visible scar might indicate that they are not yet fully cured, so the absence of a scar is seen as a proxy of cure.

I think that if this scar was disappeared, then I would have the notion that it is healed. The scar being here makes me remember of this disease. [...] I think it's important to have the lesion healed, in my case I also wish that this scar disappear. I guess that's it, it would give me the security of being cured. But I do not have much sense. I think when you see the wound, when you see the scar, you're remembering. Sometimes, this question arises... Will it be gone? It is insecure. (BR03)

This male patient from Burkina Faso is afraid of a relapse, that the scars could become an open wound again:

Because the lesions are healed but the scars are vulnerable. The scars are different from the other parts of my body. If something injures me the scars may become wounds. [...] For sure it [the disease] may come back again. As I told you the scars are vulnerable. Then if something injures me the scars may become wounds. (BF09)

8. Side effects: For this specific theme, analysis focusses on fears and conceptions of patients, as well as 'discrepant voices', that is, opinions not generally represented in other settings or interviews. In general, people felt informed about side effects related to the treatment. The most common side effects was local pain, e.g. at the injection or treatment site, which is reported below. Other common adverse effects were also headaches, local swelling, pain (e.g. back or stomach ache, pain in the joints, muscles, kidneys, legs or the testicles), loss of appetite, fever, tiredness and dizziness. Patients also complained about tachycardia, shivers, weakness, weight loss and vomiting to a lesser extent. These side effects are documented in the literature (González et al., 2009, 2008).

Patients had gotten information about possible side effects from different sources, e.g. by reading, or *via* rumours.

I: and have you searched for information about this disease after diagnosis? R: Yes. And I was very concerned to the duration of the process and how much this drug is toxic, huh? [...] I: what is your main fear? Both for you and for your child, in relation to the treatment? What are you afraid to happen? R: I'm concerned that it (the treatment) leaves any sequel (irreversible side effect). I: Understood. Is it your fear after reading more about the disease? R: Yes, that's what I'm afraid. (BR04)

I've heard it's to check the organs to see if they are in good shape; because the Glucantime it's so heavy that effect the heart, kidneys liver...It affects pretty much every organ, so you must have them in good shape. The blood test it's pretty much to check that out. If something is not good then the medicine can screw it up. [...] The doctor tells you: the treatment is going to affect this and that [the kidneys and the other organs]. [...] Also [he] told me I couldn't drink black beverages, not even coffee. No fat because of the liver. It's serious stuff because there are people who get affected very much. [...]I've seen partners who have two and three treatments already and it's not the same, doctor. I've seen them having troubles with their kidneys and how that decays their bodies. I've seen people that with the Glucantime only they lose a lot of weight. (CP02)

a) Pain: As an example, pain was described as a common side effect of treatment, sometimes perceived as unbearable.

   Yes, it was very hurtful. When you have already received 5 from the 10 injections, the medicine doesn't go through because the buttocks are so swollen that they are forced to suspend the treatment. You cannot endure the pain. I was one of those. I went to therapy but it wasn't enough. I was all the time lying in bed. I couldn't even walk. Very painful. [...] The therapy is a series of massages to help you alleviate the pain. They massage you with a machine. They put you hot and cold things in your buttocks in order to lower the swelling. That helps a lot. It's helped me. It also helped the fact that I was suspended by 5 days because the medicine wouldn't go through. (CP04)

   Oh my sister, what I felt? Say that I lost my feeling. I felt the pain, the stronger pain I felt in my life, you couldn't stand it, a pain until you can't feel your leg any more it was like that it didn't belongs to your body anymore. I couldn't move my leg any more. The doctor tried to move it and help me to stand up. He asked me not to walk after that I felt that he anesthetized it. After walking a little bit I felt the double or the triple of the previous pain. (TU03)

b) Sexual behaviour: Related to fears of sterility discussed previously, CL treatment affecting sexual activity was mentioned in the same context, again by patients in Colombia. One patient complained about the loss of sexual activity during treatment in combination with pain in his testicles, but talks about the possibility of fathering a child, so otherwise this bears no resemblance to the fears of CL drugs causing infertility described previously.

A Colombian soldier talks about advice being given to him by colleagues to refrain from sexual activity during treatment, complementing advice from the HCP to not consume alcohol, nicotine, to get a sufficient amount of sleep and to pay attention to cholesterol intake:

R: Well, you know what other people say. You know about the injections and to take care because of the treatment… I: What recommendations do they give you? R: Not drinking, not smoking, not staying up late, avoid colesterol and not having sex. They always say the same thing. They say that affects the treatment. […] The hardest part was to be on treatment without sex. I don't like the booze or the cigarette. I don't stay up late; I don't like to eat food high in cholesterol but I do like to have sex, so that was the hardest thing about it. (CP10)

In summary, findings showed a variety of treatment outcomes that patients consider important, reflecting upon their personal health and wellbeing as well as their social and socioeconomic context.

3.3.4. Outcomes used in clinical trials for CL interventions

Clinical trials conducted to test CL interventions show great heterogeneity with regards to outcomes as well as a focus on the lesion as unit of analysis, insufficiently taking patients into consideration. Of 106 studies of CL interventions included in a review by López et al. (2018), 14.2% specifically defined initial cure, and 90.6% final cure. In 100% and 86.35% of the studies which included initial and definitive cure, complete re-epithelisation of the ulcer was the definition of 'cure', respectively. Additional criteria that were variably present included: absence of active lesion, no relapse, negative parasitological test, clinical improvement, no appearance of new lesions, and/or reversible hypopigmentation. In 11.6% of the studies definitive cure was defined as re-epithelisation, ranging from >60% to >90%. For all the studies including nodular and papular lesions, cure was defined as resolution and flattening of the lesion.

With regards to outcomes concerning patients rather than lesions, clearly fewer are included in trials: Of the studies included in the Cochrane collaboration reviews for OWCL and NWCL (González et al., 2008, 2010), the degree of functional and aesthetic impairment was not measured by any study. Prevention of scarring was measured by 8 out of 49 studies in OWCL, but was not

considered by any study included in the review of NWCL. QoL was included in 2 out of 49 studies in OWCL, but by none in NWCL. Adverse effects were considered by all except three in OWCL (46 out of 49) and by all except 6 in NWCL (32 out of 38).

3.3.5. Related findings

In the following, findings related to outcomes from the interview analysis will be discussed, providing a rich picture of the context.

Despite not being the main focus of the analysis, mental health and the severe psychological impact of CL was a common and central theme throughout the interviews. Epidemiological studies have shown that around 50% of CL lesions are located in the face, and the visibility of lesion is considered an important risk factor for depression in dermatological conditions (Bailey et al., 2017; Karimkhani et al., 2016). Evidence comparing patients with active and inactive lesions with a healthy control group shows significant association of CL with anxiety, depression and decreased QoL in patients. This concerned lesions on exposed body parts such as face and hands, and active CL for more than 1 year, permanent scar formation and social stigmatization (Yanik, Gurel, Simsek, & Kati, 2004). This is supported by findings from many of the interviews, perhaps best illustrated by the Tunisian woman with diabetes (TU04), who reported suicidal thoughts should her leg be amputated because of CL.

Self-healing is a common phenomenon in cases of CL caused by *L. major*. Several Iranian patients explain the traditional name for CL in Persian, *Salak*, which can be translated as 'one year' and refers to the time it takes for the disease to heal naturally. Similarly, patients in Tunisia talk about CL as healing by itself after a period of time, with one patient noting that aged people named leishmaniasis *the six months sore*, after which it would heal. A very clear picture of self-healing cases emerged from Burkina Faso, where all except one patient (who received Glucantime) interviewed in Burkina Faso remained untreated, apart from traditional treatments, antibiotics, wound care (e.g. antiseptics or soap), anti-inflammatory drugs, or tetanus vaccines. They report self-healing of the lesions typically after several months.

An interesting contrast in patient perceptions was seen between regions who have been endemic for CL for a long time, such as Shiraz in Iran, and new, emerging foci, for example Larama in Burkina Faso, where people were affected during an outbreak in 2013. In Larama, patients interviewed reported very little or no knowledge about transmission, etiology and consequences of the disease, leading to consequences such as social stigmatization. A woman with a large number of CL lesions on her body explains how others in the village associated her disease with her ethnic group, or her person, who had brought the disease into the village. She is also unsure about the severity of the disease, believing it could potentially lead to death:

When I got the disease I was very sad. I thought that it would kill me. (BF03)

In comparison, patients in endemic regions for instance in Colombia or Iran, knew about transmission and treatment of disease. Colombian soldiers knew the disease very well and saw it as a professional risk. The majority knew about transmission and prevention, and often received the correct diagnosis from colleagues who had been infected previously.

Within Iran, there was an interesting contrast between endemic areas, e.g. one patient (IR03) received advice from workers from Afghanistan who knew the disease well. To IR07's family in Tabriz (the Aserbaijan province of Iran), the disease was unknown, and explained that she did not tell her mother out of fear that she would get upset, whereas to her husband's family in Shiraz, an endemic area, the disease was familiar.

3.3.6. Summary

Patients reported outcomes important to them from different domains (Figure 18): Absence of pain was seen as essential, with symptomatic and strong pain described mainly by patients in Burkina Faso; others specified it as itching, stinging or tingling sensation. Possible sequelae of CL that patients were afraid of were amputation of extremities; others that constitute unfounded fears were sequelae affecting internal organs, cancer and death. The fear to become sterile was also mentioned; however, this was restricted to patients from Colombia, and mostly soldiers with the Colombian army. Among the likely sequelae, scar formation was seen as an important outcome. Patients feared

disability and social stigmatization – very often, a special emphasis was put onto gender-specific stigmatization of women due to visible scars. Scars are also seen as proxy for cure, as a reminder of a potentially still active disease. The lesion appearance – size, colour, dryness, being closed and 'normal skin' - was also seen as an indicator of cure, as well as parasitological cure, that is, absence of parasite, and often confirmation by a HCP. Patients feared relapses of CL, and reinfection. Among the side effects being reported were headaches, local swelling, pain (e.g. back or stomach ache, pain in the joints, muscles, kidneys, legs or the testicles), loss of appetite, fever, tiredness and dizziness. Patients also complained about tachycardia, shivers, weakness, weight loss and vomiting to a lesser extent. Additionally, effects on sexual behaviour were mentioned, specifically loss of libido, and related to that, the advice to abstain from sexual activity during treatment as an advice given by colleagues in the Colombian army. The impact on daily activities concerned themes like the ability to work, to participate in a social life, and a generally good fitness level.

Contrasted with the outcomes included in studies so far, it clearly shows that the outcomes considered as important by patients are insufficiently considered in trials: Outcomes related to the degree of functional and aesthetic impairment (which we suggest could be mapped to the outcomes contained in the domain *Sequelae* or *Disability* associated with scars) was not measured by any study. Prevention of scarring, which would correspond to the outcomes domain *Scar formation*, was measured by 16% of studies in OWCL, and not in a single study in NWCL. QoL, which could be associated with outcomes contained in the domains *Daily activities* or *Pain*, was included in 4% of studies in OWCL, but by none in NWCL. Adverse effects were considered more often (González et al., 2009, 2008). Most of the outcomes brought up in the interviews were found to be highly interrelated, making it difficult to exactly map them onto those considered among the secondary and tertiary outcomes by both Cochrane Collaboration reviews. I would argue that this mapping might not be necessary.

3.4. Discussion

3.4.1. Introduction

The broader aim of this study part was to determine optimal outcomes for clinical trials for CL. These data are important for two reasons: Firstly, for CL patients, who were consulted about their opinion, therefore giving them a voice in clinical study design, and secondly, for other research groups who would like to consult patients on issues related to clinical trials. Both of these aspects also contribute to understanding better trial design and why outcome measure selection is an important element that needs to be considered.

This discussion section will focus on four broad themes: First, to discuss the study itself, and the challenges encountered during planning, initiation, conduct and analysis (section 3.4.2). Second, to make recommendations based on the treatment outcomes considered important by patients (section 3.4.3). Third, to point out limitations of the study (section 3.4.4), and fourth, to consider how this might guide future study design, and to suggest further research (section 3.4.5).

3.4.2. Study planning, initiation, conduct and analysis

The investigators encountered a number of challenges specific to the conduct of this multicentre study, a) related to the investigators' dual role as interviewers and HCPs and/or researchers, b) the number of languages (Arabic, Dioula, Farsi, Mòoré, Portuguese and Spanish) patients were interviewed in, and the partly strong dialects being used and c) the division of labour, mostly between conduct, translation and transcription of interviews being done by one researcher, and analysis being conducted by someone else.

Dual role of investigators: For this study, HCPs assumed the role of interviewers, and in most cases patients being interviewed were aware of the interviewers' professional background. This could potentially have influenced the conversation by introducing an undesired relationship, where HCPs would be expected to answer questions and provide medical advice, or even a hierarchy. If necessary, interviewers clarified their role towards patients before or during the interview: that they, as HCPs, were wearing a 'different hat' as interviewers. Specifically in a situation where HCPs

interview patients, Britten (1995) recommends to ensure that patients are able to speak freely, and not be corrected if they make wrong statements about treatment for example. In cases where patients were found to ask questions concerning medical issues during interviews, interviewers offered to clarify these after the interview, and without being recorded.

Related to this dual role of the investigators in this study, ERCs emphasized the importance of avoiding coercion of patients. In order to not make patients believe that their medical treatment depended on their willingness to participate in the study, or their answers given during the interview, appropriate recruitment procedures were followed: If the interviewer was treating the patient as HCP, that is, interviewing his/her 'own patient', the informed consent was taken by someone else not involved in the patient's treatment. If the interviewer was not treating the patient as HCP, the patient was invited to participate by their treating physician/nurse, or someone from the medical team, and the informed consent was then taken by the investigator or someone from the medical team.

Interview languages: Interviews were being conducted in patients' mother tongues, and transcripts were then translated into English for analysis. However, the original languages the interviews were conducted in were still preserved in many elements in the transcripts: Very often, patients used colloquial expressions, also when relating to medical terminology. In addition, PIs noted the partly strong dialects that patients used when being interviewed to describe their experience, e.g. in Tumaco, the Pacific region of Colombia (M. Castro Noriega, personal communication). This made the conduct, transcription and translation of interviews difficult. It was found that consulting colleagues or collaborators who were familiar with the local dialects helped.

Division of labour: In relation to that, this study setup, where local investigators conducted, transcribed and translated interviews which were then sent to another investigator for analysis across all regions, was unusual for qualitative research, where there is usually very little division of labour (Ziebland & McPherson, 2006).

In order to resolve the challenges surrounding the languages and the division of labour, thorough note-taking by the interviewers before, during and after the interviews was encouraged, and

recognized as essential for the analysis. Interviewers were asked to take notes e.g. on the setting, atmosphere and context of the interview, the main points made by the respondent during this interview, whether they encountered anything surprising during the interview, or any problems with the topic guide (e.g. wording, order of topics, missing topics). The notes greatly helped to interpret and understand the information contained in the transcripts. In addition, iterative discussions and written feedback rounds helped to clarify any remaining issues.

3.4.3. Recommendations based on patient-preferred outcomes

Information obtained on preferred outcomes was very rich, and general key themes could be identified as well as context-specific themes (Figure 18). In summary, outcomes identified from the literature as included and reported in clinical trials are much more 'clinical', that is, focussing on cure of lesions after a certain time, insufficiently considering patients (3.3.4 Outcomes used in clinical trials for CL interventions).

Based on the outcomes and outcome domains identified Figure 18, I have grouped them over several steps in order to be able to make recommendations (Figure 19): Those that were found to be already incorporated and reported in clinical trials, as discussed in 3.3.4 Outcomes used in clinical trials for CL interventions, are depicted in green. Among the remaining ones, a number of outcomes surrounding *Sequelae* were based on unfounded fears in relation to CL, these are depicted in red. The remaining outcomes were either found to be not yet sufficiently taken into consideration in trials. For a number, depicted in blue, their inclusion, using appropriate measurement instruments, could be suggested if they are found to be clinically meaningful and measurable. Those that comprised of outcomes with a longer-term impact such as *Amputation* and *Disability* or gender-specific and cultural connotations surrounding the psychological and social impact of scar formation (as discussed in 3.3.3 Outcomes: The patients' perspectives and 3.3.5 Related findings) were grouped separately and depicted in brown and gold.

The groups will be discussed in more detail in the following.

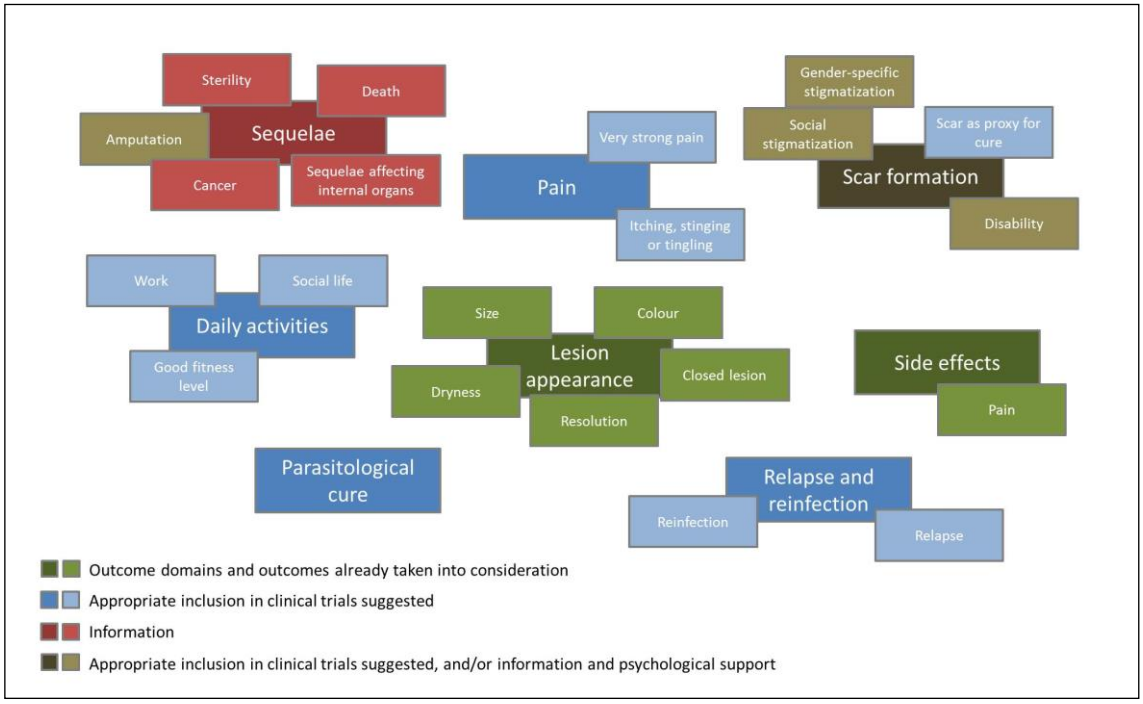


Figure 19: **Reported outcomes grouped by recommendations :** First, outcomes that are already taken into consideration in trials (dark and light green). Second, those that could be addressed by clinical trials in clinical trials using appropriate instruments (dark and light blue) and third, those that could be included in trials in combination with information and psychological support, taken their often complex sociocultural background into consideration if necessary (brown and gold). Fourth, those that could be addressed by information to patients (dark and light red).

The first group contains outcome domains and outcomes depicted in dark green and green, respectively, and mostly concerning *Lesion appearance*, and also the most common *Side effects*. I would argue that these are concepts corresponding to the clinical definition of cure with regards to lesion appearance as already reported in 3.3.4 Outcomes used in clinical trials for CL interventions, that is, that all inflammatory signs have disappeared, and epithelialization has occurred in ulcerative lesions.

The second group, depicted in blue, contains the domains *Daily activities, Pain, Parasitological cure* and *Relapse and reinfection* as well as the outcome *Scar as proxy for cure.* I would argue that *Daily activities, Pain* and *Scar as proxy for cure* (as well as its immediate impact on daily life) could be summarized under a broader category *Quality of Life*, and would suggest using an appropriate measurement

instrument for patients' QoL, such as the Dermatology Quality of Life Index (DLQI) (Finlay & Khan, 1994), which is standardized, used widely and has been translated into a great number of languages, to assess quality-of-life related aspects in clinical trials. *Relapse and reinfection* are addressed in trials, but to an insufficient extent: Only around 40% of studies in NWCL and 25% of studies in OWCL address recurrence after 6 months of follow-up (González et al., 2009, 2008), so it is recommended to include these outcomes more widely by using longer follow-up times. *Parasitological cure*, that is, the absence of parasites in the body, is an outcome patients would like to see; but there is evidence of parasite persistence in scars even after clinical cure, as discussed in more detail 3.3.3. This suggests that parasitological cure might not be associated with clinical cure in CL. Given these limitations, the inclusion of *Parasitological cure* as a potentially meaningful outcome could be explored *via* a consensus survey, such as suggested in 3.4.5 Relevance to clinical trial design and further research.

The third group, in brown and gold, contains outcomes broadly contained in the outcome domain *Scar formation*, such as *Social stigmatization* (and in particular *Gender-specific stigmatization* towards women) and *Disability,* as well as *Amputation* in the outcome domain *Sequelae.* Amputation is a rare, but possible, sequel in patients with comorbidities such as diabetes, or those who develop MCL. These are particularly sensitive issues, and I would recommend informing patients, for example about treatment options for scars, and if possible, in addition provide special attention and psychological support and/or sensitization. This is considered important in particular against the background of an association of CL with anxiety and depression, as discussed in 3.3.5 Related findings. In addition, their inclusion in trials *via* a QoL questionnaire such as the DLQI, would need to take the cultural background into consideration.

The fourth group, in dark and light red, contains outcomes patients are afraid of, but which are not based on a medical or clinical foundation. A number are grouped in the domain *Sequelae* consist of fears around *Sterility, Death, Cancer* and *Sequelae affecting internal organs*. These could be considered as misconceptions, as discussed in more detail in section 3.3.3 Outcomes: The patients' perspectives. I would suggest addressing outcomes in this group carefully by informing patients to avoid these

misconceptions and alleviate their fears, preferably through personal communication. This could help them to e.g. distinguish between CL and VL in co-endemic areas in relation to *Sequelae*.

3.4.4. Limitations of the study

One on the limitations of this study was the fact that, as a qualitative study, it can't talk about representativeness of the results for a given study population. Despite aiming at covering a patient population as large and diverse as possible, the study was by no means able to cover outcomes important for other patients, and the findings are thus restricted to the number of patients that were actually consulted.

Second, and in relation to this, data saturation was deemed unlikely to be achieved, due to the maximum variation sampling approach and the small sample size per site. Conclusions related to single countries or regions, and comparisons and contrasting between countries and regions were possible to only a very limited extent, as an interesting finding reported from one region might have just been missed at another site.

Third, by design, there was a distinct focus on clinical trials, and it was therefore avoided to include patients that would not normally be enrolled, that is, patients with severe comorbidities or more severe disease manifestations. Depending on the region, CL patients very often suffer from comorbidities potentially exacerbating their symptoms, such as HIV-positive patients which were pointed out as a special patient population in Burkina Faso, developing a much greater number of lesions than immunocompetent patients (M. Cissé, personal communication). Diabetes as comorbidity was common, typically interfering with the healing process of CL lesions and potentially leading to amputation of affected limbs. This was a central theme for one Tunisian patient, who even mentioned suicide, should her leg be amputated, as discussed in section 3.3.3 Outcomes: The patients' perspectives.

Fourth, it was decided during one of the training workshops that children had to be excluded, despite forming a significant part of the patient population in some countries such as Iran.

Interviewing children, as was pointed out during the planning phase of the study, requires a skillset and precautions that we did not find attainable within the scope of the study.

3.4.5. Relevance to clinical trial design and further research

As a follow-up, already considered in the design for this patient interview study, I consider it important to conduct a study to consult researchers and HCPs working in the field of CL about outcomes for CL interventions, as stakeholder groups that are planning and conducting trials, and treating patients, respectively. Figure 20 shows the workflow of a proposed Delphi-type survey for HCPs treating CL patients and researchers working with CL. Outcomes reported in the literature will be combined with outcomes reported during this study and recommended for inclusion in clinical trials (Figure 19, in blue) in a questionnaire. During the first round (R1) participants will be asked to rate these outcomes according to their importance, and to suggest further outcomes to include. The two following rounds (R2 and R3) only aim at quantitative scoring according to perceived importance among HCPs and researchers. The aim of this study is the development of a COS as described in section 3.1.4, consisting of a set of proposed outcomes and their measurement instruments to include in future clinical trials, therefore working towards harmonization of clinical trial methodology (Olliaro et al., 2018). The survey also provides the opportunity of seeking consensus with regards to eligibility criteria. The study is outlined in more detail in 8.9 Appendix 9: Delphi consensus survey.

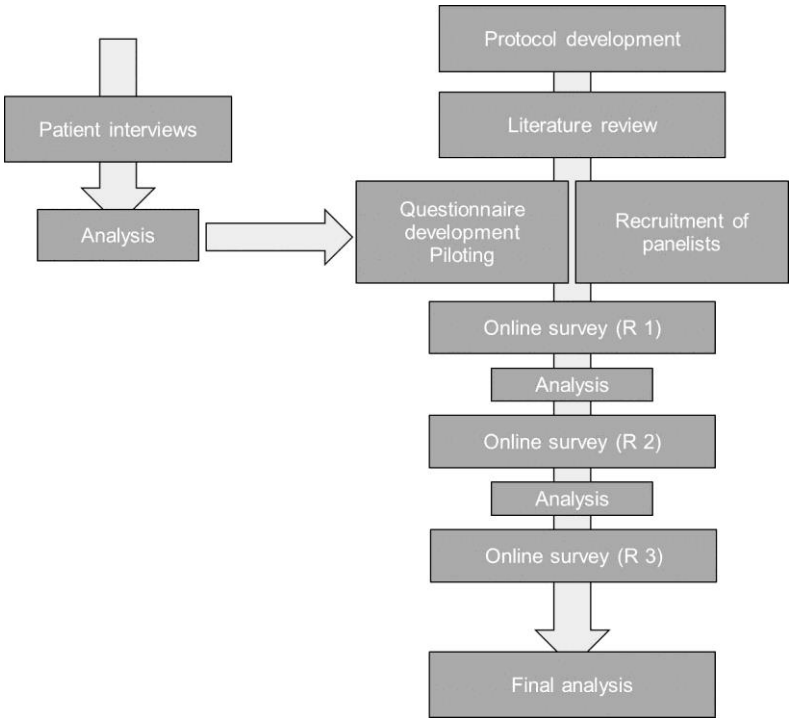


Figure 20: **Schematic overview of the Delphi process**. The figure depicts the proposed Delphi consensus survey over three rounds. The survey questionnaire is constructed after a review and analysis of the literature, and from the findings of the patient interviews. After piloting, the survey is run over three rounds in order to obtain as much consensus as possible. Modified from Prinsen et al. (2014)

An important question to address is whether the outcomes perceived as important by patients will be accepted and given equal importance by HCPs and researchers. However, and even if there is no convergence between the patients' and the experts' attitudes and perceptions, it is important to state that the data obtained from this qualitative study exploring patients' preferences has its intrinsic validity.

Further data analysis and follow-up studies are being planned/conducted on other themes, such as the socio economic and psychological burden of disease (lesion and scar) and treatment in everyday life, traditional treatments of CL, direct and indirect costs associated with CL and preferred modes of administration or dosage, which could be useful to guide product development, such as establishment of a TPP.

It is also seen as important that consulting children as patients form part of further research, e.g. on a more regional level, after consultation of local experts and IRBs, in order to develop a protocol that sufficiently takes appropriate methods and regulations, the context as well as cultural norms into consideration.

3.5. Conclusions

This study intended to capture disease experiences of CL patients in endemic countries by conducting semi-structured interviews, and to identify treatment outcomes for intervention trials in CL important from a patient perspective through analysis of the interviews.

It clearly showed that such an approach is feasible: The analysis resulting in a rich account of outcomes of a very diverse group of 74 patients in countries endemic for CL (Brazil, Burkina Faso, Colombia, Iran, Morocco, Peru and Tunisia). While some, such as *Lesion appearance* and *Side effects*, were already considered in existing trials, a large group of outcomes focussed on issues related to QoL (those grouped mostly under *Daily activities* and *Pain*) has been insufficiently taken into consideration so far in clinical trials. A group emerged as relating to the fear of possible sequelae (the domain *Sequelae*), most of which are unfounded and are based on misconceptions; these could be alleviated by providing appropriate information and psychological support, if necessary, to patients. *Scar formation* emerged as an important domain, and outcomes, as related to QoL, could be included to a certain extent in clinical trials. In addition the psychosocial impact of scar formation, often related to gender-specific stigmatization, should be addressed by psychological support and sensitization or information. I was able to make clear recommendations regarding treatment outcomes.

These results indicate the importance and feasibility of consulting stakeholders, and in particular patients, on elements of clinical trial design. This is with the aim of designing better and more appropriate studies.

4. Study summary and conclusions

This part will provide a short summary of the two research projects separately, and then will contrast and compare results of both. This discussion will focus on four, very interrelated elements: First, study design, second, the importance of outcomes relevant for patients, third, stakeholder consultations and fourth, using qualitative methodology.

4.1. Summary

This work consists of two studies addressing clinical-trial specific issues for two skin NTDs. Despite being different in their approaches, and the subjects being approached, the studies' overall results indicate a great role for methodology research in order to improve study design. The first study (chapter 2: Diagnostics evaluation studies in low-resource settings: Case study of a point-of-care (PoC) test for diagnosis of Buruli Ulcer (BU)) used a study design combining qualitative and quantitative data to investigate an evaluation study of a diagnostic to test for *M. ulcerans* infection in humans conducted in a rural area in Ghana. Results indicate an iterative research process, and several points of triangulation between both data types: Together with stakeholders, technical challenges were identified and addressed with experiments. A rich account of the context and identified barriers for diagnosis and treatment of BU will hopefully provide useful directions for future developments of a PoC test in this specific setting, even if the LAMP as candidate technology turns out to be not feasible. A Bayesian statistical approach for determination of test accuracy was seen as advantageous in order to address challenges such as small sample sizes typical during the development of a diagnostic test and an imperfect gold standard; however, the use of expert estimates for eliciting relevant parameters such as prevalence for more appropriate modelling of test accuracy merits further research.

The second study (chapter 3 Using patient consultations to better define outcome measures for Cutaneous Leishmaniasis (CL) clinical trials) resulted in a set of outcomes of CL treatments that a diverse international patient population reported as important to them. The majority of these outcomes have been insufficiently taken into consideration in trials so far, these mostly belong to

the outcome domains of *Relapse and reinfection*, *Parasitological cure*, *Scar formation*, *Pain*, *Daily activities* and *Sequelae*. I suggest including a number of them in trials using standardized instruments such as the DLQI and addressing others, some of which are unfounded fears, *via* information to patients and psychological support if needed. The chosen methodology of conducting semi-structured interviews with patients was also found suitable to explore a broad range of themes in addition, such as socioeconomic or psychological impact on patients' lives, and thus has the potential improve case-management in routine practice in often difficult settings.

4.2. Discussion

First, the thesis reflects upon adequate study designs for testing interventions and diagnostic tests in resource-limited settings. It reported a diagnostic evaluation trial at its early stage as a rather cyclical process of trial-and-error, on-going development and improving. More methodology research is needed in order to provide a robust body of knowledge, and, importantly, guidance for researchers who would like to conduct diagnostic test evaluation trials, in particular in resource-limited settings. In contrast to this, researchers wanting to conduct trials for interventions have better guidance and regulatory oversight available to them; however, methodological deficiencies need to be addressed. For CL interventions, trials were found to use poorly harmonized outcome measures, and in addition, only a limited range of outcomes being taken into consideration and measured. This study showed that this could be addressed by consulting patients as important stakeholders, and making suggestions for outcome measures to be included in trials - thus making trials, and the resulting interventions, more suitable and relevant for patients.

Second, looking at results of both studies in combination, the central role of outcomes suitable for patients in studies becomes clear: For trials of interventions such as for CL, I can clearly recommend to consider an additional set of outcomes considered as important by patients. This is in line with the literature, and research conducted by the US-based Patient Centered Outcomes Research Institute (PCORI) or by the NHS in an initiative called Patient Centred Outcome Measures (PCOMs) (Brédart et al., 2014; L. Forsythe, Heckert, Margolis, Schrandt, & Frank, 2017;

L. P. Forsythe et al., 2016; Frank et al., 2015; NHS England, 2015, 2017). However, this type of research is rarely implemented for infectious diseases, and in particular NTDs. In parallel to this, the importance of outcomes for patients – and in particular outcomes important from a patients' perspective - when evaluating diagnostic tests is something that is hardly considered. As suggested by Lijmer et al. (2009), setting measurement of outcomes for patients is a very important aim for diagnostics evaluation studies. Test accuracy is one component of test evaluation (Ferrante di Ruffano, Hyde, McCaffery, Bossuyt, & Deeks, 2012), but establishing a benefit to patient health must be the priority for diagnostic evaluations.

Third, and closely related to this, both studies also point out the importance of identifying and consulting suitable stakeholders during design and operation of clinical studies. I would argue that a central role should be given to patients, as the 'end users' of interventions. As particularly emphasized when consulting CL patients about outcomes that they prefer, they are linking treatment outcomes to their specific settings. As such, outcomes get almost 'dragged' out from a medical domain into so many aspects of their lives, as discussed in 3.3.3 Outcomes: The patients' perspectives and 3.4.3 Recommendations based on patient-preferred outcomes. These are aspects such as patients' social lives, their mental health and psychological state, but also economic domains - if they simply would like to heal in order to be able to feed their families *via* subsistence farming. I would argue that this additional perspective does not invalidate the more technical, medical domains trials conventionally use as outcomes. Both domains will need to be taken into consideration in order to ensure that the resulting interventions are appropriate for the setting they are meant to be implemented in. HCPs in a broader sense are another group of stakeholders I was working with for the BU project. For this study, I deliberately included village-based CBS volunteers among HCPs, due to their essential role in identifying cases, treatment and follow-up. At the same time, they were trained individuals appointed by, and living in, the communities. They were able to provide rich accounts about patients' attitudes where we could not consult patients directly.

It was found for both studies that qualitative data could be used in order to understand and to be able to overcome barriers to research studies in NTDs. This thesis addressed this in the two following ways:

- By adding an exploratory element, which yields additional findings or perspectives that are not normally known, or addressed. This includes CL patients' views on the treatment effects they would like to see. Most of these have not been sufficiently taken into consideration in trials far, whereas others are misconceptions that could be clarified (such as sterility fears related to treatment common among Colombian soldiers). This also includes a description of the context where a new intervention is to be implemented by consulting various stakeholder groups such as HCPs or CBS volunteers. These stakeholder groups are able to link an intervention to 'real life', that is, the setting and context where it is to be implemented – a notion at the core of implementation research (Peters et al., 2013).

- By introducing an additional explanatory component, which is not necessarily taken into consideration in classical, quantitative study designs and is potentially able to answer the question '*Why?*'. This allows for explaining observations: Patients presented late for diagnosis of BU because they perceived BU as caused by witchcraft rather than being a medical disease, and therefore sought treatment with traditional healers (spiritualists or herbalists) first. Of particular importance here is that the qualitative paradigm also provides space for many interesting narratives in between, such as collaborations of CBS volunteers with traditional healers in order to ensure that patients get treated adequately.

It needs to be emphasized that the methodology developed in both research strands is not disease-specific, and could be used for a wide range of NTDs.

4.3. Conclusions and recommendations

This research work consists of two different studies, both addressing deliberately diverse issues around clinical studies for NTDs of the skin. Despite being challenging to work with, this provides perspectives from very different angles, and points at several common issues trials in low-resource settings have. This work aimed to incorporate methodology research into clinical studies for interventions and diagnostic tests for skin NTDs conducted in low-resource settings, in order to identify key factors that could improve research design, conduct and relevance for patients – against the background of a lack of research and good quality evidence in this field.

The aim of the first study was to *explore the design and conduct of an evaluation trial of a diagnostic test for BU in a low-resource setting*. The specific objectives were two-fold, *to explore the process and context of a diagnostic test evaluation trial for BU* and *to evaluate performance of a diagnostic test for BU under field conditions including historical data and expert estimates*. These objectives were partly met: The performance of the BU LAMP was evaluated in a series of experiment in the laboratory facilities of the NMIMR or at the field site, a health clinic in Pakro, a rural area in Ghana. The context and diagnostic workflow as well as the process of the evaluation study was explored using FGDs and individual interviews with researchers, HCPs and CBS volunteers, also identifying barriers to BU diagnosis and treatment as well as research directions, resulting in an overall iterative research process.

The objectives of the CL study were *to capture disease experiences of CL patients in endemic countries by conducting semi-structured interviews*, and *to identify outcomes for intervention trials in CL that are important from a patient's perspective through analysis of the interviews*. These objectives were met: In-depth interviews of about 1 hour duration with 74 patients in Brazil, Burkina Faso, Colombia, Iran, Morocco, Peru and Tunisia were conducted and analysed. A set of outcomes relevant from a patients' perspective were reported, and further recommendations formulated: First, for their consideration in clinical trials, advice and information to patients in order to clarify misconceptions, and psychological support if needed.

Following up on the previous chapter, and taking into consideration the themes discussed there (study design, patient-relevant outcomes, stakeholder consultations and using qualitative

methodology for clinical studies), I would like to make recommendations in line with the results of this thesis.

Issues around study design brought up in this thesis differed, depending on whether an intervention or a diagnostic test is being evaluated:

For trials testing interventions, methodological deficiencies such as poorly defined, non-standardized outcomes and eligibility criteria should be addressed appropriately. Patients could be consulted about preferred outcomes *via* individual, semi-structured interviews using a maximum variation sampling approach. In line with the results of this study, I would recommend keeping the scope of the interview sufficiently broad, such as interviewing patients about their overall disease experience. This has the potential to also improve treatment of patients by exploring their treatment preferences, socioeconomic or psychological burden, or access to treatment. The thesis assessed two ways of consulting stakeholders issues related to clinical trials, *via* semi-structured interviews and FGDs. This study was able to show that patients consider outcomes as relevant that have not been considered so far in trials, so patient-relevant outcomes should be taken into consideration when designing research studies, regardless of the nature of the intervention (treatment or diagnostic test). Consultation of HCPs and researchers, another important group of stakeholders, could not be realized due to resource limitations. This is planned *via* a consensus survey aiming at harmonization of CL clinical trial design, as laid out in more detail in 3.4.5 Relevance to clinical trial design and further research: In order to encourage incorporating patient-preferred outcomes into future trials, I would therefore recommend, as a clear subsequent step, a Delphi survey or similar methodology. The survey would obtain feedback from researchers and HCPs on a broad list of outcomes, consisting of those included previously in trials and reported *via* a systematic review, and those reported by patients as relevant. In addition the survey could address other methodological elements of trials, such as eligibility criteria, which were found to be poorly harmonized too across CL clinical trials. The aim of the Delphi would be to obtain a COS for inclusion in future trials that is considered meaningful and relevant for the target population, that is, patients, researchers and HCPs, as well as CEC as guidance for future trials.

For diagnostic test evaluation studies at that stage of development, results pointed to an iterative process consisting of separate experiments with continuous improvement, instead of testing a defined intervention. The lack of guidance and methodological standardization is in contrast to the complexity of the context a diagnostic test might need to operate in. This is particularly true in settings where f.i. a diagnostic pathway might be strongly influenced by infrastructural constraints and great economic burden on patients (even if diagnosis and treatment are provided free of charge). I would recommend incorporating interviews and, preferably, FGDs with a mixed group of researchers, HCPs and patients, or patient representatives, at several stages of a diagnostic test evaluation. The discussion should contain both exploratory and explanatory elements, and could help to ensure that the context and barriers are taken into account, even at early stages when planning research activities. In line with this iterative study design, a Bayesian statistical approach, modelling a prior distribution from existing data and expert estimates, would be better suited to smaller sample sizes, an imperfect gold standard and missing data expected during the research study.

Once a diagnostic test has been developed and validated, qualitative methodology consulting a variety of stakeholder such as researchers, HCPs and patients could be used during implementation trials of the test (Peters et al., 2013) in order to better adapt it to the local context, e.g. incorporation into an existing diagnostic workflow. It could also be used for a process or gap analysis, or could complement a cost-effectiveness analysis for implementation of a new diagnostic test in a specific setting.

Whereas working with a variety of stakeholders is beneficial, it is important to emphasize that there are several methods of stakeholder consultation, and not all might be appropriate in a specific setting, and for all groups.

Overall, this thesis sought to explore key elements of clinical study design of NTDs of the skin that could inform future study designs, taking two diverse approaches into consideration. This could be addressed, and in summary it can be concluded that methodology research considering qualitative as well as quantitative data can be able to address specific challenges to better help understand and overcome barriers for research studies in NTDs.

5. List of abbreviations

AFB	Acid-Fast Bacilli
BCI	Bayesian Credible Interval
BU	Buruli Ulcer
CBS	Community Based Surveillance
CEC	Core Eligibility Criteria
CHPS	Community-based Health and Planning Services
CI	(Frequentist) confidence interval
CL	Cutaneous Leishmaniasis
COM	Core Outcome Measure
COS	Core Outcome Set
ddH_2O	double distilled H_2O
DHIMS	District Health Information Management System
DRB	Dry Reagent Based
FDA	Food and Drug Administration
FDR	Fluorescence detection reagent
fg	femtogram (10^{-15}g)
FGD	Focus Group Discussion
FIND	Foundation for Innovative New Diagnostics
FIT	Foreign-Initiated trial
FNA	Fine Needle Aspirate
GHS	Ghana Health Service
GSK	GlaxoSmithKline
HERG	Health Experiences Research Group
HIV	Human Immunodeficiency Virus
HTA	Health Technology Assessment
IIT	Investigator-initiated trial
IM	Intramuscular
IRB	Institutional Review Board
LAB	Liposomal Amphotericin B
LAMP	Loop-mediated isothermal reaction
LMIC	Low and middle income countries
LOD	Limit Of Detection

LRS	Low Resource Settings
MA	Meglumine Antimoniate
MCL	Mucocutaneous Leishmaniasis
MDA	Mass Drug Administration
ML	Mycolactone
MMV	Medicines for Malaria Ventures
NAAT	Nucleic-Acid Amplification Test
NGO	Non-Governmental Organization
NHS	National Health System
NINA	Non-Instrumented Nucleic Acid Amplification
NMIMR	Noguchi Memorial Institute for Medical Research
NWCL	New-World CL
OMERACT	Outcome Measures in Rheumatology
OWCL	Old-World CL
OxTREC	Oxford Tropical Research Ethics Committee
PAR	Participatory Action Research
PATH	Program for Appropriate Technology in Health
PBS	Phosphate Buffered Saline
PCOMs	Patient Centred Outcome Measures
PCOR	Patient-Centered Outcomes Research
PCORI	Patient-Centered Outcomes Research Institute
PCR	Polymerase Chain Reaction
pdf	probability density function
PDP	Product Development Partnership
pg	picogram (10^{-12}g)
PI	Principal Investigator
PKDL	Post Kala-azar Dermal Leishmaniasis
PMM	Paromomycin
PoC	Point of Care
PPV	Positive Predictive Value
QoL	Quality of Life
RCT	Randomized Controlled Trial
RDT	Rapid Diagnostic Test

Sb^V	Pentavalent antimonial
SDI	(World Health Organization) Sexually Transmitted Diseases Diagnostics Initiative
SOP	Standard Operating Procedure
SSD	Silica Syringe Device
SSG	Sodium-Stibogluconate
STARD	Standards for Reporting of Diagnostic Accuracy Studies
SWAT	Study Within A Trial
TDR	Special Programme for Research and Training in Tropical Diseases
TPP	Target Product Profile
TSA	Technical Services Agreement
VL	Visceral Leishmaniasis
WHO	World Health Organization
WRAIR	Walter Reed Army Institute of Research

6. Table of figures

7. References

Ablordey, A., Amissah, D. A., Aboagye, I. F., Hatano, B., Yamazaki, T., Sata, T., … Katano, H. (2012). Detection of Mycobacterium ulcerans by the loop mediated isothermal amplification method. *PLoS Neglected Tropical Diseases*, *6*(4), e1590. https://doi.org/10.1371/journal.pntd.0001590

Ablordey, A., & de Souza, D. K. (2015). *BU LAMP Project Report to FIND (unpublished)*.

Aebischer, A., & Mrva, M. (2016). Leishmania. In J. Walochnik & M. Duchêne (Eds.), *Molecular Parasitology* (pp. 195–216). Springer Vienna. https://doi.org/10.1007/978-3-7091-1416-2_7

Ahortor, E. (2014). *Evaluating IS2404 loop mediated isothermal amplification (LAMP) method for field diagnosis of Buruli Ulcer in endemic communities in Ghana* (MPhil thesis). MPhil thesis, University of Ghana, Legon.

Akins, R. B., Tolson, H., & Cole, B. R. (2005). Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. *BMC Medical Research Methodology*, *5*, 37. https://doi.org/10.1186/1471-2288-5-37

Aksoy, S. (2010). Solutions to Neglected Tropical Diseases Require Vibrant Local Scientific Communities. *PLOS Neglected Tropical Diseases*, *4*(3), e662. https://doi.org/10.1371/journal.pntd.0000662

Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K., & Rousseau, J. (2012). Combining Expert Opinions in Prior Elicitation. *Bayesian Analysis*, *7*(3), 503–532. https://doi.org/10.1214/12-BA717

Alvar, J., Vélez, I. D., Bern, C., Herrero, M., Desjeux, P., Cano, J., … WHO Leishmaniasis Control Team. (2012). Leishmaniasis worldwide and global estimates of its incidence. *PloS One*, *7*(5), e35671. https://doi.org/10.1371/journal.pone.0035671

Arkin, C. F., & Wachtel, M. S. (1990). How many patients are necessary to assess test performance? *JAMA*, *263*(2), 275–278.

Asiedu, K., Scherpbier, R., & Raviglione, M. (2000). *Buruli ulcer. Mycobacterium ulcerans infection.* Geneva: WHO/CDS/CPE/GBUI/1 WHO. Retrieved from http://whqlibdoc.who.int/hq/2000/WHO_CDS_CPE_GBUI_2000.1.pdf?q=buruli-ulcer

Aslanzadeh, J. (2004). Preventing PCR Amplification Carryover Contamination in a Clinical Laboratory. *Annals of Clinical & Laboratory Science*, *34*(4), 389–396.

Aujoulat, I., Johnson, C., Zinsou, C., Guédénon, A., & Portaels, F. (2003). Psychosocial aspects of health seeking behaviours of patients with Buruli ulcer in southern Benin. *Tropical Medicine & International Health*, *8*(8), 750–759. https://doi.org/10.1046/j.1365-3156.2003.01089.x

Austin, P. C., Naylor, C. D., & Tu, J. V. (2001). A comparison of a Bayesian vs. a frequentist method for profiling hospital performance. *Journal of Evaluation in Clinical Practice*, *7*(1), 35–45.

Bailey, F., Mondragon-Shem, K., Hotez, P., Ruiz-Postigo, J. A., Al-Salem, W., Acosta-Serrano, Á., & Molyneux, D. H. (2017). A new perspective on cutaneous leishmaniasis—Implications for global prevalence and burden of disease estimates. *PLOS Neglected Tropical Diseases*, *11*(8), e0005739. https://doi.org/10.1371/journal.pntd.0005739

Banoo, S., Bell, D., Bossuyt, P., Herring, A., Mabey, D., Poole, F., … TDR Diagnostics Evaluation Expert Panel. (2006). Evaluation of diagnostic tests for infectious diseases: general principles. *Nature Reviews. Microbiology*, *4*(12 Suppl), S20-32. https://doi.org/10.1038/nrmicro1570

Basáñez, M.-G., Marshall, C., Carabin, H., Gyorkos, T., & Joseph, L. (2004). Bayesian statistics for parasitologists. *Trends in Parasitology*, *20*(2), 85–91.

Bates, P. A., & Rogers, M. E. (2004). New insights into the developmental biology and transmission mechanisms of Leishmania. *Current Molecular Medicine*, *4*(6), 601–609.

Beissner, M., Phillips, R. O., Battke, F., Bauer, M., Badziklou, K., Sarfo, F. S., … Bretzel, G. (2015). Loop-Mediated Isothermal Amplification for Laboratory Confirmation of Buruli Ulcer Disease—Towards a Point-of-Care Test. *PLoS Negl Trop Dis*, *9*(11), e0004219. https://doi.org/10.1371/journal.pntd.0004219

Belli, A., García, D., Palacios, X., Rodriguez, B., Valle, S., Videa, E., … Harris, E. (1999). Widespread atypical cutaneous Leishmaniasis caused by Leishmania (L.) Chagasi in Nicaragua. *The American Journal of Tropical Medicine and Hygiene*, *61*(3), 380–385.

Bergquist, R., Johansen, M. V., & Utzinger, J. (2009). Diagnostic dilemmas in helminthology: what tools to use and when? *Trends in Parasitology*, *25*(4), 151–156. https://doi.org/10.1016/j.pt.2009.01.004

Berkvens, D., Speybroeck, N., Praet, N., Adel, A., & Lesaffre, E. (2006). Estimating disease prevalence in a Bayesian framework using probabilistic constraints. *Epidemiology (Cambridge, Mass.)*, *17*(2), 145–153. https://doi.org/10.1097/01.ede.0000198422.64801.8d

Bern, C., Maguire, J. H., & Alvar, J. (2008). Complexities of assessing the disease burden attributable to leishmaniasis. *PLoS Neglected Tropical Diseases*, *2*(10), e313. https://doi.org/10.1371/journal.pntd.0000313

Berry, D. A. (2006). Bayesian clinical trials. *Nature Reviews Drug Discovery*, *5*(1), 27–36. https://doi.org/10.1038/nrd1927

Blood, E., & Spratt, J. F. (2007). Disagreement on Agreement: Two Alternative Agreement Coefficients. In *SAS Global Forum 2007 Proceedings*. Retrieved from www2.sas.com/proceedings/forum2007/186-2007.pdf

Boehme, C. C., Nabeta, P., Henostroza, G., Raqib, R., Rahim, Z., Gerhardt, M., … Perkins, M. D. (2007). Operational Feasibility of Using Loop-Mediated Isothermal Amplification for Diagnosis of Pulmonary Tuberculosis in Microscopy Centers of Developing Countries. *Journal of Clinical Microbiology*, *45*(6), 1936–1940. https://doi.org/10.1128/JCM.02352-06

Boelaert, M., & Sundar, S. (2014). Leishmaniasis. In J. Farrar, P. Hotez, T. Junghanss, G. Kang, D. Lalloo, & W. Nicholas J. (Eds.), *Manson's Tropical Diseases Expert Consult.* (pp. 631–655). Saunders Ltd.

Boers, M., Kirwan, J. R., Wells, G., Beaton, D., Gossec, L., d'Agostino, M.-A., … Tugwell, P. (2014). Developing core outcome measurement sets for clinical trials: OMERACT filter

2.0. *Journal of Clinical Epidemiology*, *67*(7), 745–753. https://doi.org/10.1016/j.jclinepi.2013.11.013

Bossuyt, P. M., Cohen, J. F., Gatsonis, C. A., Korevaar, D. A., & STARD group. (2016). STARD 2015: updated reporting guidelines for all diagnostic accuracy studies. *Annals of Translational Medicine*, *4*(4), 85. https://doi.org/10.3978/j.issn.2305-5839.2016.02.06

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., … Standards for Reporting of Diagnostic Accuracy. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. *Clinical Chemistry*, *49*(1), 1–6.

Bowling, A. (2009). *Research Methods In Health: Investigating Health and Health Services*. McGraw-Hill International.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brédart, A., Marrel, A., Abetz-Webb, L., Lasch, K., & Acquadro, C. (2014). Interviewing to develop Patient-Reported Outcome (PRO) measures for clinical research: eliciting patients' experience. *Health and Quality of Life Outcomes*, *12*, 15. https://doi.org/10.1186/1477-7525-12-15

Britten, N. (1995). Qualitative interviews in medical research. *BMJ (Clinical Research Ed.)*, *311*(6999), 251–253.

Broemeling, L. (2007). *Bayesian Biostatistics and Diagnostic Medicine*. Chapman and Hall/CRC. Retrieved from http://www.crcpress.com/product/isbn/9781584887676

Buderer, N. M. (1996). Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*, *3*(9), 895–900.

Chauty, A., Ardant, M.-F., Adeye, A., Euverte, H., Guédénon, A., Johnson, C., … Grosset, J. (2007). Promising Clinical Efficacy of Streptomycin-Rifampin Combination for Treatment

of Buruli Ulcer (Mycobacterium ulcerans Disease). *Antimicrobial Agents and Chemotherapy*, *51*(11), 4029–4035. https://doi.org/10.1128/AAC.00175-07

Cherryholmes, C. H. (1992). Notes on Pragmatism and Scientific Realism. *Educational Researcher*, *21*(6), 13–17. https://doi.org/10.3102/0013189X021006013

Choy, S. L., O'Leary, R., & Mengersen, K. (2009). Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology*, *90*(1), 265–277.

Clarke, M. (2007). Standardising outcomes for clinical trials and systematic reviews. *Trials*, *8*(1), 39. https://doi.org/10.1186/1745-6215-8-39

Clarke, M. (2012). Education section - Studies Within A Trial (SWAT). *Journal of Evidence-Based Medicine*, *5*(1), 44–45. https://doi.org/10.1111/j.1756-5391.2012.01169.x

Clarke, M., Savage, G., Maguire, L., & McAneney, H. (2015). The SWAT (study within a trial) programme; embedding trials to improve the methodological design and conduct of future research. *Trials*, *16*(2), P209. https://doi.org/10.1186/1745-6215-16-S2-P209

Costello, A., & Zumla, A. (2000). Moving to research partnerships in developing countries. *BMJ (Clinical Research Ed.)*, *321*(7264), 827–829.

Creswell, J. W. (2011). *Designing and conducting mixed methods research*. Los Angeles; London: SAGE.

Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.). SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). SAGE Publications.

Davey, S., & Global Forum for Health Research (Organization). (2004). *The 10/90 report on health research, 2003-2004*. Geneva, Switzerland: Global Forum for Health Research.

Dawson, S., Manderson, L., & Tallo, V. L. (1993). WHO | A manual for the use of focus groups. International Nutrition Foundation for Developing Countries (INFDC). Retrieved from http://archive.unu.edu/unupress/food2/UIN03E/UIN03E00.HTM

De Lima, H., Rodríguez, N., Feliciangeli, M. D., Barrios, M. A., Sosa, A., Agrela, I., … Lopez, O. (2009). Cutaneous leishmaniasis due to Leishmania chagasi/Le. infantum in an endemic

area of Guarico State, Venezuela. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *103*(7), 721–726. https://doi.org/10.1016/j.trstmh.2008.11.019

de Souza, D. K., Quaye, C., Mosi, L., Addo, P., & Boakye, D. A. (2012). A quick and cost effective method for the diagnosis of Mycobacterium ulcerans infection. *BMC Infectious Diseases*, *12*, 8. https://doi.org/10.1186/1471-2334-12-8

Delgado, O., Guevara, P., Silva, S., Belfort, E., & Ramirez, J. L. (1996). Follow-up of a human accidental infection by Leishmania (Viannia) braziliensis using conventional immunologic techniques and polymerase chain reaction. *The American Journal of Tropical Medicine and Hygiene*, *55*(3), 267–272.

Dendukuri, N., Rahme, E., Bélisle, P., & Joseph, L. (2004). Bayesian sample size determination for prevalence and diagnostic test studies in the absence of a gold standard test. *Biometrics*, *60*(2), 388–397. https://doi.org/10.1111/j.0006-341X.2004.00183.x

Dhama, K., Karthik, K., Chakraborty, S., Tiwari, R., Kapoor, S., Kumar, A., & Thomas, P. (2014). Loop-mediated isothermal amplification of DNA (LAMP): a new diagnostic tool lights the world of diagnosis of animal and human pathogens: a review. *Pakistan Journal of Biological Sciences: PJBS*, *17*(2), 151–166.

DNDi. (2016). DNDi - Portfolio. Retrieved January 18, 2017, from http://www.dndi.org/diseases-projects/portfolio/

Du, R., Hotez, P. J., Al-Salem, W. S., & Acosta-Serrano, A. (2016). Old World Cutaneous Leishmaniasis and Refugee Crises in the Middle East and North Africa. *PLOS Neglected Tropical Diseases*, *10*(5), e0004545. https://doi.org/10.1371/journal.pntd.0004545

Dujardin, J.-C., & Decuypere, S. (2013). Epidemiology of Leishmaniasis in the Time of Drug Resistance. In Alicia Ponte-Sucre, E. Diaz, & M. Padrón-Nieves (Eds.), *Drug Resistance in Leishmania Parasites* (pp. 65–83). Springer Vienna. https://doi.org/10.1007/978-3-7091-1125-3_1

Durnez, L., Stragier, P., Roebben, K., Ablordey, A., Leirs, H., & Portaels, F. (2009). A comparison of DNA extraction procedures for the detection of Mycobacterium ulcerans, the causative

agent of Buruli ulcer, in clinical and environmental specimens. *Journal of Microbiological Methods*, *76*(2), 152–158. https://doi.org/10.1016/j.mimet.2008.10.002

Erber, A. C., Arana, B., Bennis, I., Salah, A. B., Boukthir, A., Noriega, M. del M. C., … Olliaro, P. (2018). An international qualitative study exploring patients' experiences of cutaneous leishmaniasis: study set-up and protocol. *BMJ Open*, *8*(6), e021372. https://doi.org/10.1136/bmjopen-2017-021372

Farrar, J., Hotez, P., Junghanss, T., Kang, G., Lalloo, D., & Nicholas J., W. (2014). *Manson's Tropical Diseases Expert Consult.* Saunders Ltd.

Ferrante di Ruffano, L., Hyde, C. J., McCaffery, K. J., Bossuyt, P. M. M., & Deeks, J. J. (2012). Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ (Clinical Research Ed.)*, *344*, e686.

Finlay, A. Y., & Khan, G. K. (1994). Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. *Clinical and Experimental Dermatology*, *19*(3), 210–216. https://doi.org/10.1111/j.1365-2230.1994.tb01167.x

Food and Drug Administration (FDA). (2007). Guidance for Industry and Review Staff Target Product Profile — A Strategic Development Process Tool. Retrieved from http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm080593.pdf

Forsythe, L., Heckert, A., Margolis, M. K., Schrandt, S., & Frank, L. (2017). Methods and impact of engagement in research, from theory to practice and back again: early findings from the Patient-Centered Outcomes Research Institute. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*. https://doi.org/10.1007/s11136-017-1581-x

Forsythe, L. P., Ellis, L. E., Edmundson, L., Sabharwal, R., Rein, A., Konopka, K., & Frank, L. (2016). Patient and Stakeholder Engagement in the PCORI Pilot Projects: Description and Lessons Learned. *Journal of General Internal Medicine*, *31*(1), 13–21. https://doi.org/10.1007/s11606-015-3450-z

Foundation for Innovative New Diagnostics (FIND). (2017). Target product profiles (TPPs) for diagnostic tests. Retrieved September 12, 2017, from https://www.finddx.org/target-product-profiles/

Frank, L., Forsythe, L., Ellis, L., Schrandt, S., Sheridan, S., Gerson, J., … Daugherty, S. (2015). Conceptual and practical foundations of patient engagement in research at the patient-centered outcomes research institute. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *24*(5), 1033–1041. https://doi.org/10.1007/s11136-014-0893-3

Franzen, S. R. P. (2014). *Barriers and enablers to locally-led clinical trial conduct in Low and Middle Income Countries; Strategies for developing locally sustainable health research capacity* (DPhil thesis). University of Oxford, Oxford, UK.

Franzen, S. R. P., Chandler, C., Enquselassie, F., Siribaddana, S., Atashili, J., Angus, B., & Lang, T. (2013). Understanding the investigators: a qualitative study investigating the barriers and enablers to the implementation of local investigator-initiated clinical trials in Ethiopia. *BMJ Open*, *3*(11), e003616. https://doi.org/10.1136/bmjopen-2013-003616

Fyfe, J. A. M., Lavender, C. J., Johnson, P. D. R., Globan, M., Sievers, A., Azuolas, J., & Stinear, T. P. (2007). Development and Application of Two Multiplex Real-Time PCR Assays for the Detection of Mycobacterium ulcerans in Clinical and Environmental Samples. *Applied and Environmental Microbiology*, *73*(15), 4733–4740. https://doi.org/10.1128/AEM.02971-06

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*, *13*, 117. https://doi.org/10.1186/1471-2288-13-117

Given, L., & Saumure, K. (2008). Data Saturation. In L. Given, *The SAGE Encyclopedia of Qualitative Research Methods*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. https://doi.org/10.4135/9781412963909.n99

Gobo, G. (2004). Sampling, Representativeness And Generalizability. In C. Seale, G. Gobo, J. F. Gubrium, & D. Silverman (Eds.), *Qualitative Research Practice* (pp. 405–427). SAGE Publications.

González, U., Pinart, M., Rengifo-Pardo, M., Macaya, A., Alvar, J., & Tweed, J. A. (2009). Interventions for American cutaneous and mucocutaneous leishmaniasis. In *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD004834.pub2/abstract

González, U., Pinart, M., Reveiz, L., & Alvar, J. (2008). Interventions for Old World cutaneous leishmaniasis. In *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD005067.pub3/abstract

González, U., Pinart, M., Reveiz, L., Rengifo-Pardo, M., Tweed, J., Macaya, A., & Alvar, J. (2010). Designing and reporting clinical trials on treatments for cutaneous leishmaniasis. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *51*(4), 409–419. https://doi.org/10.1086/655134

Green, J., & Thorogood, N. (2014). *Qualitative methods for health research* (3rd ed.). SAGE Publications.

Guba, E. G. (1990). *The Paradigm Dialog*. SAGE Publications.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). SAGE.

Guerrant, R. L., Walker, D. H., & Weller, P. F. (2011). *Tropical infectious diseases: principles, pathogens and practice*.

Guimaraes-Peres, A., Portaels, F., de Rijk, P., Fissette, K., Pattyn, S. R., van Vooren, J., & Fonteyne, P. (1999). Comparison of two PCRs for detection of Mycobacterium ulcerans. *Journal of Clinical Microbiology*, *37*(1), 206–208.

Gwet, K. L. (2010). *Handbook of Inter-Rater Reliability (Second Edition)*. Advanced Analytics Press.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008–1015.

Hatano, B., Maki, T., Obara, T., Fukumoto, H., Hagisawa, K., Matsushita, Y., … Katano, H. (2010). LAMP using a disposable pocket warmer for anthrax detection, a highly mobile and reliable method for anti-bioterrorism. *Japanese Journal of Infectious Diseases*, *63*(1), 36–40.

Hayashida, K., Kajino, K., Hachaambwa, L., Namangala, B., & Sugimoto, C. (2015). Direct blood dry LAMP: a rapid, stable, and easy diagnostic tool for Human African Trypanosomiasis. *PLoS Neglected Tropical Diseases*, *9*(3), e0003578. https://doi.org/10.1371/journal.pntd.0003578

Haynes, R. B., & You, J. J. (2009). The architecture of diagnostic research. In J. A. Knottnerus & F. Buntinx (Eds.), *The evidence base of clinical diagnosis* (2nd ed., pp. 20–41). Oxford; Hoboken, NJ: Wiley-Blackwell Pub./BMJ Books.

Hong, H., Coutanceau, E., Leclerc, M., Caleechurn, L., Leadlay, P. F., & Demangel, C. (2008). Mycolactone diffuses from Mycobacterium ulcerans-infected tissues and targets mononuclear cells in peripheral blood and lymphoid organs. *PLoS Neglected Tropical Diseases*, *2*(10), e325. https://doi.org/10.1371/journal.pntd.0000325

Houn, F., Bright, R. A., Bushar, H. F., Croft, B. Y., Finder, C. A., Gohagan, J. K., … Wagner, R. F. (2000). Study design in the evaluation of breast cancer imaging technologies. *Academic Radiology*, *7*(9), 684–692. https://doi.org/10.1016/S1076-6332(00)80524-3

Hunink, M. G. M., & Krestin, G. P. (2002). Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology. *Radiology*, *222*(3), 604–614. https://doi.org/10.1148/radiol.2223010335

ICMJE. (2017). ICMJE | Recommendations | Clinical Trial Registration. Retrieved September 5, 2017, from http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html

Jacobsen, K. H., & Padgett, J. J. (2010). Risk factors for Mycobacterium ulcerans infection. *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases*, *14*(8), e677-681. https://doi.org/10.1016/j.ijid.2009.11.013

Jain, N. K., & Roy, I. (2010). Trehalose and protein stability. *Current Protocols in Protein Science*, *Chapter 4*, Unit 4.9. https://doi.org/10.1002/0471140864.ps0409s59

Johnson, B., & Christensen, L. (2010). *Educational Research: Quantitative, Qualitative, and Mixed Approaches* (4th ed.). SAGE.

Jones, J., & Hunter, D. (1995). Consensus methods for medical and health services research. *BMJ : British Medical Journal*, *311*(7001), 376–380.

Jones, S. R., Carley, S., & Harrison, M. (2003). An introduction to power and sample size estimation. *Emergency Medicine Journal*, *20*(5), 453–458. https://doi.org/10.1136/emj.20.5.453

Joseph, L., Gyorkos, T. W., & Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology*, *141*(3), 263–272.

Joseph, Lawrence. (2017). Lawrence Joseph. Bayesian statistics. Retrieved from http://www.medicine.mcgill.ca/epidemiology/Joseph/

Junghanss, T., Johnson, R. C., & Pluschke, G. (2014). Mycobacterium ulcerans Disease. In J. Farrar, P. Hotez, T. Junghanss, G. Kang, D. Lalloo, & W. Nicholas J. (Eds.), *Manson's Tropical Diseases Expert Consult.* (pp. 631–655). Saunders Ltd.

Karimkhani, C., Wanga, V., Coffeng, L. E., Naghavi, P., Dellavalle, R. P., & Naghavi, M. (2016). Global burden of cutaneous leishmaniasis: a cross-sectional analysis from the Global Burden of Disease Study 2013. *The Lancet Infectious Diseases*, *16*(5), 584–591. https://doi.org/10.1016/S1473-3099(16)00003-7

Karthik, K., Rathore, R., Thomas, P., Arun, T. R., Viswas, K. N., Dhama, K., & Agarwal, R. K. (2014). New closed tube loop mediated isothermal amplification assay for prevention of product cross-contamination. *MethodsX*, *1*, 137–143. https://doi.org/10.1016/j.mex.2014.08.009

Kitzinger, J. (1995). Qualitative research. Introducing focus groups. *BMJ (Clinical Research Ed.)*, *311*(7000), 299–302.

Kitzinger, J. (2006). Focus groups. In C Pope & N. Mays (Eds.), *Qualitative research in health care* (3rd ed., pp. 21–31). Massachusetts, USA: Blackwell Publishing Ltd.

Knottnerus, J. A., & Buntinx, F. (2009). *The evidence base of clinical diagnosis: theory and methods of diagnostic research* (2nd ed.). Oxford; Hoboken, NJ: Wiley-Blackwell Pub./BMJ Books.

Knottnerus, J. A., Buntinx, F., & van Weel, C. (2009). General introduction: evaluation of diagnostic procedures. In J. A. Knottnerus & F. Buntinx (Eds.), *The evidence base of clinical diagnosis* (2nd ed., pp. 1–19). Oxford; Hoboken, NJ: Wiley-Blackwell Pub./BMJ Books.

Knottnerus, J. A., van Weel, C., & Muris, J. W. M. (2002). Evaluation of diagnostic procedures. *BMJ*, *324*(7335), 477–480. https://doi.org/10.1136/bmj.324.7335.477

Kotey, N. K. (2013). *Antibiotic Treatment Outcomes of Buruli Ulcer in Akwapem South and Suhum-Kraboa-Coaltar Districts* (Thesis). University of Ghana. Retrieved from http://ugspace.ug.edu.gh/handle/123456789/5472

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed., Vol. 2). Chicago: University of Chicago Press.

LaBarre, P., Hawkins, K. R., Gerlach, J., Wilmoth, J., Beddoe, A., Singleton, J., … Weigl, B. (2011). A simple, inexpensive device for nucleic acid amplification without electricity-toward instrument-free molecular diagnostics in low-resource settings. *PloS One*, *6*(5), e19738. https://doi.org/10.1371/journal.pone.0019738

Lang, T., Clarke, M., Gashe, F. E., Negussie, H., Molla, M., McKay, A., … Davey, G. (2013). *A research methodology study to map the process of initiating and operating a randomized trial of podoconiosis treatment in Northern Ethiopia* (Protocol (unpublished)).

Lang, T., & Siribaddana, S. (2012). Clinical Trials Have Gone Global: Is This a Good Thing? *PLOS Medicine*, *9*(6), e1001228. https://doi.org/10.1371/journal.pmed.1001228

Lang, T., White, N. J., Tran, H. T., Farrar, J. J., Day, N. P. J., Fitzpatrick, R., … Marsh, K. (2010). Clinical research in resource-limited settings: enhancing research capacity and working together to make trials less complicated. *PLoS Neglected Tropical Diseases*, *4*(6), e619. https://doi.org/10.1371/journal.pntd.0000619

le Fichoux, Y., Quaranta, J. F., Aufeuvre, J. P., Lelievre, A., Marty, P., Suffia, I., … Kubar, J. (1999). Occurrence of Leishmania infantum parasitemia in asymptomatic blood donors living in an area of endemicity in southern France. *Journal of Clinical Microbiology*, *37*(6), 1953–1957.

Lijmer, J. G., Leeflang, M., & Bossuyt, P. M. (2009). Proposals for a Phased Evaluation of Medical Tests. In *Medical Tests—White Paper Series*. Rockville (MD): Agency for Healthcare Research and Quality (US). Retrieved from http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=350

López-Carvajal, L., Vélez, I., Arbeláez, M. P., & Olliaro, P. (2018). Eligibility criteria and outcome measures adopted in clinical trials of treatments of cutaneous leishmaniasis: systematic literature review covering the period 1991-2015. *Tropical Medicine & International Health: TM & IH*. https://doi.org/10.1111/tmi.13048

Maes, L., Luz, R. A. I. da, Cos, P., & Yardley, V. (2013). Classical Versus Novel Treatment Regimens. In Alicia Ponte-Sucre, E. Diaz, & M. Padrón-Nieves (Eds.), *Drug Resistance in Leishmania Parasites* (pp. 301–319). Springer Vienna. https://doi.org/10.1007/978-3-7091-1125-3_15

Mandimika, N., Barnes, K., Chandler, C., Pace, C., Mehta, U., & Allen, E. (2013). Development of harmonized approaches for detecting and recording participant-reported antimalarial drug safety data. Protocol for part 3: The Delphi Process. Study protocol, unpublished.

Mays, N., & Pope, C. (1995). Rigour and qualitative research. *BMJ (Clinical Research Ed.)*, *311*(6997), 109–112.

McKenna, H. P. (1994). The Delphi technique: a worthwhile research approach for nursing? *Journal of Advanced Nursing*, *19*(6), 1221–1225. https://doi.org/10.1111/j.1365-2648.1994.tb01207.x

Mendonça, M. G., de Brito, M. E. F., Rodrigues, E. H. G., Bandeira, V., Jardim, M. L., & Abath, F. G. C. (2004). Persistence of leishmania parasites in scars after clinical cure of American

cutaneous leishmaniasis: is there a sterile cure? *The Journal of Infectious Diseases*, *189*(6), 1018–1023. https://doi.org/10.1086/382135

Merritt, R. W., Walker, E. D., Small, P. L. C., Wallace, J. R., Johnson, P. D. R., Benbow, M. E., & Boakye, D. A. (2010). Ecology and transmission of Buruli ulcer disease: a systematic review. *PLoS Neglected Tropical Diseases*, *4*(12), e911. https://doi.org/10.1371/journal.pntd.0000911

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Monge-Maillo, B., & López-Vélez, R. (2013). Therapeutic options for old world cutaneous leishmaniasis and new world cutaneous and mucocutaneous leishmaniasis. *Drugs*, *73*(17), 1889–1920. https://doi.org/10.1007/s40265-013-0132-1

Morales, M. A., Cruz, I., Rubio, J. M., Chicharro, C., Cañavate, C., Laguna, F., & Alvar, J. (2002). Relapses versus reinfections in patients coinfected with Leishmania infantum and human immunodeficiency virus type 1. *The Journal of Infectious Diseases*, *185*(10), 1533–1537. https://doi.org/10.1086/340219

Mori, Y., Nagamine, K., Tomita, N., & Notomi, T. (2001). Detection of loop-mediated isothermal amplification reaction by turbidity derived from magnesium pyrophosphate formation. *Biochemical and Biophysical Research Communications*, *289*(1), 150–154. https://doi.org/10.1006/bbrc.2001.5921

Mori, Yasuyoshi, & Notomi, T. (2009). Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases. *Journal of Infection and Chemotherapy: Official Journal of the Japan Society of Chemotherapy*, *15*(2), 62–69. https://doi.org/10.1007/s10156-009-0669-9

Morrell, R. (2006). Fathers, fatherhood and masculinity in South Africa. In L. Richter & R. Morrell (Eds.), *BABA: men and fatherhood in South Africa* (pp. 13–25). South Africa: Human Sciences Research Council. Retrieved from

http://www.academia.edu/2296327/Fathers_fatherhood_and_masculinity_in_South_Africa

Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F., Askham, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment (Winchester, England)*, *2*(3), i–iv, 1–88.

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., … Memish, Z. A. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, *380*(9859), 2197–2223. https://doi.org/10.1016/S0140-6736(12)61689-4

Nagamine, K., Hase, T., & Notomi, T. (2002). Accelerated reaction by loop-mediated isothermal amplification using loop primers. *Molecular and Cellular Probes*, *16*(3), 223–229.

Nagamine, K., Watanabe, K., Ohtsuka, K., Hase, T., & Notomi, T. (2001). Loop-mediated isothermal amplification reaction using a nondenatured template. *Clinical Chemistry*, *47*(9), 1742–1743.

Neuman, W. L. (2013). *Social Research Methods: Qualitative and Quantitative Approaches* (7th ed.). Pearson.

NHS England. (2015). NHS England » Children and young people to get more say in their health care. Retrieved October 3, 2017, from https://www.england.nhs.uk/2015/02/pcoms-cyp/

NHS England. (2017). NHS England » Patient Centred Outcome Measures. Retrieved October 3, 2017, from https://www.england.nhs.uk/ourwork/pe/pcoms/

Njiru, Z. K. (2012). Loop-Mediated Isothermal Amplification Technology: Towards Point of Care Diagnostics. *PLoS Negl Trop Dis*, *6*(6), e1572. https://doi.org/10.1371/journal.pntd.0001572

Njiru, Z. K., Yeboah-Manu, D., Stinear, T. P., & Fyfe, J. A. M. (2012). Rapid and sensitive detection of Mycobacterium ulcerans by use of a loop-mediated isothermal amplification test. *Journal of Clinical Microbiology*, *50*(5), 1737–1741. https://doi.org/10.1128/JCM.06460-11

Nollett, C., Kelson, M., & Hood, K. (2016). Site visits to initiate recruitment in a clinical trial: does it matter who conducts the visit? Protocol for implementation in trials. *Journal of Evidence-Based Medicine*, *9*(4), 225–227. https://doi.org/10.1111/jebm.12212

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., & Hase, T. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Research*, *28*(12), E63.

O'Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(1), 21–35. https://doi.org/10.1111/1467-9884.00114

Olliaro, P., Grogl, M., Boni, M., Carvalho, E. M., Chebli, H., Cisse, M., … Arana, B. (2018). Harmonized clinical trial methodologies for localized cutaneous leishmaniasis and potential for extensive network with capacities for clinical evaluation. *PLOS Neglected Tropical Diseases*, *12*(1), e0006141. https://doi.org/10.1371/journal.pntd.0006141

Olliaro, P., Vaillant, M., Arana, B., Grogl, M., Modabber, F., Magill, A., … Alvar, J. (2013). Methodology of clinical trials aimed at assessing interventions for cutaneous leishmaniasis. *PLoS Neglected Tropical Diseases*, *7*(3), e2130. https://doi.org/10.1371/journal.pntd.0002130

Page, A.-L., Alberti, K. P., Mondonge, V., Rauzier, J., Quilici, M.-L., & Guerin, P. J. (2012). Evaluation of a rapid test for the diagnosis of cholera in the absence of a gold standard. *PloS One*, *7*(5), e37360. https://doi.org/10.1371/journal.pone.0037360

Page, J., Heller, R. F., Kinlay, S., Lim, L. L.-Y., Qian, W., Suping, Z., … Macharia, W. (2003). Attitudes of developing world physicians to where medical research is performed and reported. *BMC Public Health*, *3*, 6.

Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, *56*(1), 45–50.

Peeling, R. W., Holmes, K. K., Mabey, D., & Ronald, A. (2006). Rapid tests for sexually transmitted infections (STIs): the way forward. *Sexually Transmitted Infections*, *82 Suppl 5*, v1-6. https://doi.org/10.1136/sti.2006.024265

Peeling, R. W., Smith, P. G., & Bossuyt, P. M. M. (2008). A guide for diagnostic evaluations. *Nature Reviews Microbiology*, *8*, S2–S6. https://doi.org/10.1038/nrmicro1522

Pepe, M. S. (2004). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. OUP Oxford.

Peters, D. H., Adam, T., Alonge, O., Agyepong, I. A., & Tran, N. (2013). Implementation research: what it is and how to do it. *BMJ*, *347*, f6753. https://doi.org/10.1136/bmj.f6753

Pierotti, R. S. (2013). Masculinity: An Overlooked Cultural Influence on Fertility. Presented at the Annual Meeting of the Population Association of America, New Orleans, LA. Retrieved from http://paa2013.princeton.edu/papers/131653

Ponte-Sucre, A. (2013). Introduction: Leishmaniasis – The Biology of a Parasite. In Alicia Ponte-Sucre, E. Diaz, & M. Padrón-Nieves (Eds.), *Drug Resistance in Leishmania Parasites* (pp. 1–12). Springer Vienna. https://doi.org/10.1007/978-3-7091-1125-3_1

Pope, Catherine, Ziebland, S., & Mays, N. (2000). Analysing qualitative data. *BMJ*, *320*(7227), 114–116. https://doi.org/10.1136/bmj.320.7227.114

Popper, K. R. (1935). *Logik der Forschung; zur Erkenntnistheorie der modernen Naturwissenschaft*. Wien: J. Springer.

Portaels, F. (2014). *Laboratory diagnosis of buruli ulcer*. Geneva: WHO.

Price, C. P., & Christenson, R. H. (2008). Evaluating New Diagnostic Technologies: Perspectives in the UK and US. *Clinical Chemistry*, *54*(9), 1421–1423. https://doi.org/10.1373/clinchem.2008.108217

Prinsen, C. A. C., Vohra, S., Rose, M. R., King-Jones, S., Ishaque, S., Bhaloo, Z., … Terwee, C. B. (2014). Core Outcome Measures in Effectiveness Trials (COMET) initiative: protocol for an international Delphi study to achieve consensus on how to select outcome measurement

instruments for outcomes included in a 'core outcome set.' *Trials*, *15*(1), 247. https://doi.org/10.1186/1745-6215-15-247

Rapport, F., Storey, M., Porter, A., Snooks, H., Jones, K., Peconi, J., … Russell, I. (2013). Qualitative research within trials: developing a standard operating procedure for a clinical trials unit. *Trials*, *14*, 54. https://doi.org/10.1186/1745-6215-14-54

Renzaho, A. M. N., Woods, P. V., Ackumey, M. M., Harvey, S. K., & Kotin, J. (2007). Community-based study on knowledge, attitude and practice on the mode of transmission, prevention and treatment of the Buruli ulcer in Ga West District, Ghana. *Tropical Medicine & International Health: TM & IH*, *12*(3), 445–458. https://doi.org/10.1111/j.1365-3156.2006.01795.x

Reveiz, L., Maia-Elkhoury, A. N. S., Nicholls, R. S., Sierra Romero, G. A., & Yadon, Z. E. (2013). Interventions for American Cutaneous and Mucocutaneous Leishmaniasis: A Systematic Review Update. *PLoS ONE*, *8*(4), e61843. https://doi.org/10.1371/journal.pone.0061843

Ridley, D. S. (1979). The pathogenesis of cutaneous leishmaniasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *73*(2), 150–160.

Röltgen, K., & Pluschke, G. (2015). Epidemiology and disease burden of Buruli ulcer: a review. *Research and Reports in Tropical Medicine*, *2015*(6), 59–73. https://doi.org/10.2147/RRTM10.2147/RRTM.S62026

Romano, A., Inbar, E., Debrabant, A., Charmoy, M., Lawyer, P., Ribeiro-Gomes, F., … Sacks, D. L. (2014). Cross-species genetic exchange between visceral and cutaneous strains of Leishmania in the sand fly vector. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(47), 16808–16813. https://doi.org/10.1073/pnas.1415109111

Rothwell, P. M. (2006). Factors That Can Affect the External Validity of Randomised Controlled Trials. *PLoS Clinical Trials*, *1*(1). https://doi.org/10.1371/journal.pctr.0010009

Sackett, D. L., & Haynes, R. B. (2002). The architecture of diagnostic research. *BMJ : British Medical Journal*, *324*(7336), 539–541.

Schönian, G., Cupolillo, E., & Mauricio, I. (2013). Molecular Evolution and Phylogeny of Leishmania. In Alicia Ponte-Sucre, E. Diaz, & M. Padrón-Nieves (Eds.), *Drug Resistance in Leishmania Parasites* (pp. 15–44). Springer Vienna. https://doi.org/10.1007/978-3-7091-1125-3_2

Schubach, A., Haddad, F., Oliveira-Neto, M. P., Degrave, W., Pirmez, C., Grimaldi, G., & Fernandes, O. (1998). Detection of Leishmania DNA by polymerase chain reaction in scars of treated human patients. *The Journal of Infectious Diseases*, *178*(3), 911–914.

Sema, M., Alemu, A., Bayih, A. G., Getie, S., Getnet, G., Guelig, D., … Pillai, D. R. (2015). Evaluation of non-instrumented nucleic acid amplification by loop-mediated isothermal amplification (NINA-LAMP) for the diagnosis of malaria in Northwest Ethiopia. *Malaria Journal*, *14*, 44. https://doi.org/10.1186/s12936-015-0559-9

Senah, K. A., Richardson, D., Kwofie, G., Wilson, J. B., Collison, A. H., & Tinkorang, E. K. (1997). From abandoned warehouse to life-saving facility, Pakro, Ghana. *International Journal of Gynecology & Obstetrics*, *59*, S91–S97. https://doi.org/10.1016/S0020-7292(97)00152-5

Shafer, G. (1982). Lindley's Paradox. *Journal of the American Statistical Association*, *77*(378), 325–334. https://doi.org/10.2307/2287244

Shakya, N., Bajpai, P., & Gupta, S. (2011). Therapeutic switching in leishmania chemotherapy: a distinct approach towards unsatisfied treatment needs. *Journal of Parasitic Diseases: Official Organ of the Indian Society for Parasitology*, *35*(2), 104–112. https://doi.org/10.1007/s12639-011-0040-9

Siegmund, V., Adjei, O., Nitschke, J., Thompson, W., Klutse, E., Herbinger, K. H., … Bretzel, G. (2007). Dry Reagent–Based Polymerase Chain Reaction Compared with Other Laboratory Methods Available for the Diagnosis of Buruli Ulcer Disease. *Clinical Infectious Diseases*, *45*(1), 68–75. https://doi.org/10.1086/518604

Siegmund, V., Adjei, O., Racz, P., Berberich, C., Klutse, E., van Vloten, F., … Bretzel, G. (2005). Dry-reagent-based PCR as a novel tool for laboratory confirmation of clinically diagnosed Mycobacterium ulcerans-associated disease in areas in the tropics where M. ulcerans is

endemic. *Journal of Clinical Microbiology*, *43*(1), 271–276. https://doi.org/10.1128/JCM.43.1.271-276.2005

Sinha, I. P., Smyth, R. L., & Williamson, P. R. (2011). Using the Delphi Technique to Determine Which Outcomes to Measure in Clinical Trials: Recommendations for the Future Based on a Systematic Review of Existing Studies. *PLoS Med*, *8*(1), e1000393. https://doi.org/10.1371/journal.pmed.1000393

Sizaire, V., Nackers, F., Comte, E., & Portaels, F. (2006). Mycobacterium ulcerans infection: control, diagnosis, and treatment. *The Lancet. Infectious Diseases*, *6*(5), 288–296. https://doi.org/10.1016/S1473-3099(06)70464-9

Smith, P. G., & Morrow, R. (1996). *Field Trials of Health Interventions in Developing Countries: A Toolbox* (Second edition). London: Macmillan Education.

Smith, V., Clarke, M., Begley, C., & Devane, D. (2015). SWAT-1: The effectiveness of a "site visit" intervention on recruitment rates in a multi-centre randomised trial. *Trials*, *16*, 211. https://doi.org/10.1186/s13063-015-0732-z

Smith, V., Clarke, M., Devane, D., Begley, C., Shorter, G., & Maguire, L. (2013). SWAT 1: what effects do site visits by the principal investigator have on recruitment in a multicentre randomized trial? *Journal of Evidence-Based Medicine*, *6*(3), 136–137. https://doi.org/10.1111/jebm.12049

Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons.

Stienstra, Y., van der Graaf, W. T. A., Asamoa, K., & van der Werf, T. S. (2002). Beliefs and attitudes toward Buruli ulcer in Ghana. *The American Journal of Tropical Medicine and Hygiene*, *67*(2), 207–213.

Stinear, T., Ross, B. C., Davies, J. K., Marino, L., Robins-Browne, R. M., Oppedisano, F., … Johnson, P. D. (1999). Identification and characterization of IS2404 and IS2606: two distinct repeated sequences for detection of Mycobacterium ulcerans by PCR. *Journal of Clinical Microbiology*, *37*(4), 1018–1023.

Sumsion, T. (1998). The Delphi Technique: An Adaptive Research Tool. *The British Journal of Occupational Therapy*, *61*(4), 153–156.

Tanner, N. A., Zhang, Y., & Evans, T. C. (2015). Visual detection of isothermal nucleic acid amplification using pH-sensitive dyes. *BioTechniques*, *58*(2), 59–68. https://doi.org/10.2144/000114253

The Global Health Network Editorial Team. (2015). Article: Studies Within a Trial or Review. Retrieved September 28, 2017, from https://globalresearchmethods.tghn.org/articles/studies-within-trial-or-review/

Tomita, N., Mori, Y., Kanda, H., & Notomi, T. (2008). Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. *Nature Protocols*, *3*(5), 877–882. https://doi.org/10.1038/nprot.2008.57

Tugwell, P., Boers, M., Brooks, P., Simon, L., Strand, V., & Idzerda, L. (2007). OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials*, *8*, 38. https://doi.org/10.1186/1745-6215-8-38

Ubeda, J.-M., Légaré, D., Raymond, F., Ouameur, A. A., Boisvert, S., Rigault, P., … Ouellette, M. (2008). Modulation of gene expression in drug resistant Leishmania is associated with gene amplification, gene deletion and chromosome aneuploidy. *Genome Biology*, *9*(7), R115. https://doi.org/10.1186/gb-2008-9-7-r115

van der Werf, T. S. (2017). WHO Drug Study for Buruli Ulcer - Comparison of SR8 and CR8 - Full Text View - ClinicalTrials.gov,. Identifier NCT01659437. Retrieved August 28, 2017, from https://clinicaltrials.gov/ct2/show/NCT01659437

Van Loggerenberg, F. (2017). Blog post : Global Health Methodology Research and SWATs (Studies Within A Trial) - Global Health Methodology Research. Retrieved September 28, 2017, from https://globalresearchmethods.tghn.org/community/blogs/post/3269/2015/07/global-health-methodology-research-and-swats-s/

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Family Medicine*, *37*(5), 360–363.

Wadsworth, Y. (1998). What is Participatory Action Research? *Action Research International*. Retrieved from http://www.aral.com.au/ari/p-ywadsworth98.html

Wagai, J., Senga, J., Fegan, G., & English, M. (2009). Examining agreement between clinicians when assessing sick children. *PloS One*, *4*(2), e4626. https://doi.org/10.1371/journal.pone.0004626

Walley, T. (2008). Evaluating laboratory diagnostic tests. *BMJ (Clinical Research Ed.)*, *336*(7644), 569–570. https://doi.org/10.1136/bmj.39513.576701.80

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., … QUADAS-2 Group. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*(8), 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

Whiting, P. F., Weswood, M. E., Rutjes, A. W. S., Reitsma, J. B., Bossuyt, P. N. M., & Kleijnen, J. (2006). Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Medical Research Methodology*, *6*, 9. https://doi.org/10.1186/1471-2288-6-9

WHO. (2004). Buruli ulcer disease : mycobacterium ulcerans infection : an overview of reported cases globally. Retrieved from http://apps.who.int//iris/handle/10665/72521

WHO. (2012). London Declaration on Neglected Tropical Diseases. Uniting to combat neglected tropical diseases. Retrieved from http://www.who.int/neglected_diseases/London_Declaration_NTDs.pdf

WHO. (2017). WHO | Clinical trials. Retrieved September 6, 2017, from http://www.who.int/topics/clinical_trials/en/

WHO/NTD. (2011). *Buruli Ulcer. Recognize and act now ! A guide for field health workers*. Retrieved from http://whqlibdoc.who.int/publications/2011/9789241501224_eng.pdf?ua=1

WHO/NTD. (2012). Accelerating work to overcome the global impact of neglected tropical diseases. A roadmap for implementation. Retrieved from http://www.who.int/neglected_diseases/resources/WHO_HTM_NTD_2012.1/en/

WHO/NTD. (2014). Report of a WHO–FIND consultative meeting on diagnostics for Buruli ulcer. (Dr K. Asiedu/Buruli Ulcer, Ed.). Retrieved from http://apps.who.int/iris/bitstream/10665/112669/1/WHO_HTM_NTD_IDM_2014.2_eng.pdf

WHO/TDR. (2015). Call for Expression of Interest. Retrieved July 18, 2017, from http://www.who.int/tdr/grants/CL_Call_for_Expression_of_Interest.pdf?ua=1

Williamson, P., Altman, D., Blazeby, J., Clarke, M., & Gargon, E. (2012b). Driving up the quality and relevance of research through the use of agreed core outcomes. *Journal of Health Services Research & Policy*, *17*(1), 1–2. https://doi.org/10.1258/jhsrp.2011.011131

Williamson, P., Altman, D. G., Blazeby, J. M., Clarke, M., Devane, D., Gargon, E., & Tugwell, P. (2012a). Developing core outcome sets for clinical trials: issues to consider. *Trials*, *13*(1), 132. https://doi.org/10.1186/1745-6215-13-132

Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Medical Research Methodology*, *13*, 61. https://doi.org/10.1186/1471-2288-13-61

World Health Organization. (2014, April 22). Leishmaniasis fact sheet, http://www.who.int/mediacentre/factsheets/fs375/en/, accessed 2014-04-22. Retrieved April 22, 2014, from

World Health Organization. Global Buruli Ulcer Initiative. (2004). Provisional guidance on the role of specific antibiotics in the management of Mycobacterium ulcerans disease (Buruli ulcer). Retrieved from http://www.who.int/buruli/information/antibiotics/en

Yanik, M., Gurel, M. S., Simsek, Z., & Kati, M. (2004). The psychological impact of cutaneous leishmaniasis. *Clinical and Experimental Dermatology*, *29*(5), 464–467. https://doi.org/10.1111/j.1365-2230.2004.01605.x

Ziebland, S., & McPherson, A. (2006). Making sense of qualitative data analysis: an introduction with illustrations from DIPEx (personal experiences of health and illness). *Medical Education*, *40*(5), 405–414. https://doi.org/10.1111/j.1365-2929.2006.02467.x

8. Appendices

8.1. Appendix 1: Experimental reagents, primers and calculations

Preparation of PANTA Transport Medium

Difco Middlebrook 7H9 broth (4.7g) and 2 ml glycerol was dissolved in 900 ml of distilled water and autoclaved for 10 minutes at 121°C. With the aid of a syringe and under aseptic, a vial of BBL MGIT PANTA was dissolved in sterilised 7H9 Middlebrook/glycerol solution. The medium was then dispensed in 2 ml cryo vials for sample collection

Modified Boom DNA extraction method (Ahortor, 2014; Durnez et al., 2009)

250µl of the processed sample were lysed with 250µl of lysis buffer (1.6 M GuHCl, 60mM Tris (pH 7.4), 1% Triton X-100, 60 mM EDTA, 10% Tween-20), 50µl Proteinase K (20mg/ml) and 500 µl glass beads in a sterile 1.5 ml eppendorf tube. Tubes were incubated horizontally in an orbital shaker (200 rpm) at 60°C for 12 hours (overnight). 40µl of diatomaceous earth solution (diatomaceous earth (10g) dissolved in 50 ml of distilled water containing 500µl of 37% (w/v) HCl) was added to the lysed sample and incubated at 37°C for 1 hour to capture DNA. The mixture was pelleted by centrifuging at 14000 rpm for 1 min at RT, and the supernatant discarded. DNA bound to diatomaceous earth was washed twice with 900 µl of ice cold 70% alcohol and 900µl of absolute acetone, pulsed vortexed and centrifuged at 14000 rpm for 60 seconds at room temperature. Supernatant was discarded and DNA bound to diatomaceous earth was dried at 50°C for 20 min using a heat block before resuspension in 80µl of nuclease free water and incubation at 68°C for 20 min to unbind the DNA. The DNA extract was centrifuged at 14000 rpm for 2 min and about 60µl of the resulting supernatant transferred to a sterile tube and stored at -20°C. Negative and positive control extractions were also performed in each batch of DNA extraction. The negative control extractions included all the reagents for DNA extraction except the inclusion of a clinical specimen whiles the positive control extraction included 250 µl of *M. ulcerans* isolate suspension.

Primer sequences for *IS2404* nested PCR (Guimaraes-Peres et al., 1999)

PCR Primer name	Sequence (5'-3')	Expected Band size (bp)
pGp1	AGGGCAGCGCGGTGATACGG	546
pGp2	CAGTGGATTGGTGCCGATCGAG	
pGp3	GGCGCAGATCAACTTCGCGGT	218
pGp4	CTGCGTGGTGCTTTACGCGC	

Primer sequences for LAMP assay (Ablordey et al., 2012; Ahortor, 2014)

LAMP primers were designed using online primer software (PrimerExplorer, version 4, http://primerexplorer.jp/e/v4_manual/index.html).

LAMP Primer name	Sequence (5'-3')
Buruli- F3	CGAGAACAGCCTGCACTG
Buruli- B3	CGGTTGGCGGTCAAAGC
Buruli-FIP	GTGCGCCGTGTCCGGTATGGATACGCGATGTCACCTTC
Buruli- BIP	AGGTCCTAGCAACGCTACGCAAATCCGGCAGGCTTCGG
Buruli-LF	GCCTTTGACGGTCTTCGTC
Buruli-LB	CACCGCGATCAATCTGCAC

Calcein/$MnCl_2$ solution for LAMP product detection (Tomita et al., 2008)

Calcein (Sigma Aldrich C0875-5G) 625µM was added to a 12.5mM solution of $MnCl_2$ (Sigma Aldrich 63535-50G-F) in ddH_2O. Calcein has very poor water solubility, and requires rigorous shaking at RT for at least 2 hrs, protected from light and optimally in a large volume (e.g. 50ml conical centrifuge tube (Falcon, 14-432-22)). The resulting mixture is bright yellow and can be stored at -20°C in aliquots in order to avoid repeated freeze-thaw cycles. Upon addition to the reaction tubes, it assumes the same pale orange colour, and in case of a positive reaction, the same bright greenish-yellow colour as observed with the commercially available FDR (Eiken Chemical Co. Ltd. (Tokyo, Japan)).

Calculation of gene copy number for determination of analytical sensitivity (Ahortor, 2014)

The *M. ulcerans* genome has a chromosome size of 5632kb, and 174kb in extrachromosomal plasmid DNA, giving a total genome size of 5806kb. Its mass was calculated using

$$\textit{M. ulcerans}\ \text{genome mass} = \frac{(n)\ (1.096 \times 10^{-21} g)}{\text{bp}}$$

where n = genome size of *M. ulcerans*. The mass per genome calculated was 6363 fg/genome.

The mass of a single *IS2404* copy, of which *M. ulcerans* has an average number of 207 copies per genome, was calculated using

$$1\ \textit{IS2404} = \frac{6363\ \text{fg/genome}}{207\ \text{copies/genome}}$$

resulting in 0.030741 fg/copy of *IS2404*.

8.2. Appendix 2: A systematic review and Bayesian meta-analysis of studies assessing the accuracy of BU LAMP

8.2.1. Systematic review protocol

based on PROSPERO's handbook (http://www.crd.york.ac.uk/PROSPERO/) using the PICO (Patient-Intervention-Comparison-Outcome) guidelines

Citation

Review question(s)

What are the performance characteristics of assays for the detection of *Mycobacterium ulcerans* in human samples based on loop-mediated isothermal amplification (LAMP) protocols, as compared to the polymerase chain reaction (PCR) reference standard?

A special emphasis will be put on the inclusion of LAMP protocols suitable as point-of-care (PoC) tests in endemic countries.

Searches

We aim to include published studies in English, published in the period up to 31 December 2016, as well as unpublished studies available in the period up to 31 December 2016.

For literature search, we will be using Medline/PubMed, Scopus, Biosis/Web of Science, EMBASE, Global Health Database, and African Journals Online.

Types of study to be included

Studies assessing performance of LAMP as compared to PCR, or modifications of the PCR protocol (e.g. qPCR) will be included. They will need to contain paired data for comparison at the per subject level.

Exclusion criteria are as follows:

A) Lack of a comparator assay, or incomplete or poor quality of comparator diagnosis.

B) Duplication of publications, or review of previously published studies

C) Reporting errors or conflicting data.

Condition or domain being studied

Buruli Ulcer (BU), caused by *M. ulcerans*, is the most common mycobacterial disease in humans after tuberculosis and leprosy. *M. ulcerans* mostly cases necrotizing skin lesions containing clusters of extracellular bacilli. Pathogenesis is mediated by mycolactone, destroying the host tissue by necrosis and apoptosis, and strong local immunosuppression. Large-scale ulceration typically extends beyond the infected areas, and late-stage complications are osteitis/osteomyelitis and disseminated disease. In the later stages there is good consensus on case detection and diagnosis based on clinical presentation. By contrast, diagnosis is difficult in the earlier stages. Methods include direct smear examination with staining to detect acid-fast bacteria, culture, histopathology, which are lacking accuracy and/or require a long time, and PCR, requiring appropriate technical equipment. A reliable test that could be used at the PoC with minimal technical requirements does not exist. LAMP is a promising candidate technology for a PoC test for BU.

The mode of transmission has not yet been clarified, but it is generally assumed that infection takes place through a skin trauma or insect bites; endemic areas are typically associated with slow-flowing or stagnant water bodies, and transmission of individual haplotypes is very focal. Disease prevalence is unknown, with the cumulative number of BU cases in the African region in 2008 estimated to be 60000, and underreporting being suspected. Around 75% of the cases are children below the age of 15.

Participants/ population

All types of diagnostic samples of humans suspected of BU will be included. Representativeness of the sample to the general population is not a necessary criterion.

Intervention(s), exposure(s)

Diagnostic assays based on a LAMP protocol on samples obtained from humans (Index test).

Comparator(s)/ control

Reference standard: PCR, or a modification of the PCR protocol (e.g. qPCR)

Outcome(s)

Primary Outcome:

Performance characteristics of LAMP-based diagnostics in individual BU patient samples

Secondary Outcome:

Performance of LAMP-based diagnostics stratified by sample type

Data extraction, (selection and coding)

One reviewer will extract, compile and sort the identified studies according to inclusion and exclusion criteria. She will extract the following data for each LAMP method included in a study:

- Number of samples included
- Target sequence and primer sequence for LAMP primers
- Sample preparation and DNA extraction method
- Heat source for LAMP amplification
- Number of true/false positives/negatives (calculated if not reported)
- Sensitivity as compared to reference standard with a 95% confidence interval (calculated if not reported)
- Specificity as compared to reference standard with a 95% confidence interval (calculated if not reported)

In addition, information on the sample type (swabs or fine needle aspirates (FNAs), if available, will be included, and used for subgroup analysis.

Authors will be contacted for missing data.

Risk of bias (quality) assessment

Anticipatory steps will be taken to correct the effects of biases (publication bias and selection bias) by including unpublished studies, and attempting to discuss sample selection with the publications' authors.

Strategy for data synthesis

The aim of data synthesis is to calculate a meaningful Bayesian prior probability distribution (or prior) for sensitivity and specificity, so that further experimental data on LAMP accuracy during assay development and optimization can be included.

A set of patient samples tested by both LAMP and PCR (paired data) within a study will be denoted one experiment. If studies contain more than one index and/or comparator test, any suitable pairs of LAMP/PCR will be denoted an experiment. One study can contain more than one experiment.

The analysis aim is to include those experiment(s) from each study that are as close as possible to the method/experimental configuration we intend to use for the field setting, that is, electricity not required for DNA extraction (Sample preparation (and DNA extraction method) and amplification (Heat source for amplification). Selection of suitable experiments for inclusion into the model will be discussed between the authors to reach consensus. The assumption behind this selection is that the values for sensitivity and specificity will form the most meaningful prior probabilities.

LAMP sensitivity and specificity (and their confidence intervals) for experiment will be reported (or calculated if necessary). A Bayesian credible interval (BCI) will be calculated. Sensitivity and specificity values across all experiments will be included for calculation of an overall sensitivity and specificity and their BCI, and reported.

Analysis of subgroups or subsets

If adequately reported, an analysis stratified by sample types (FNA or swabs) will be conducted

Dissemination plans

Contact details for further information

Review team

Astrid Erber, Ben Lambert

Details of any existing review of the same topic by the same authors

Anticipated or actual start date

June 2016

Anticipated completion date

January 2017

Funding sources/sponsors

University of Oxford

Conflicts of interest

None known

Language

English

Country

UK

Subject index terms (MeSH)

Humans; Polymerase Chain Reaction; Nontuberculous Mycobacteria; Buruli Ulcer; Mycobacterium Ulcerans; Sensitivity and Specificity; Diagnostic Techniques and Procedures

8.2.2. Databases and search terms

Databases searched and search terms using wild cards, covering the following terms:

- buruli
- ulcerans
- diagnosis and/or diagnos*
- detection and/or detect*
- lamp
- loop-mediated
- isothermal
- amplification

Publications up to December 2016 are being considered:

Medline (http://www.ncbi.nlm.nih.gov/pubmed):(((buruli OR ulcerans) AND (diagnosis OR detection OR diagnos* OR detect*) AND (lamp OR loop-mediated OR isothermal OR amplification))) in *all fields* – 25 results *

Scopus (http://www.scopus.com/home.url): (TITLE-ABS-KEY(buruli OR ulcerans) AND TITLE-ABS-KEY(diagnos* OR detect*) AND TITLE-ABS-KEY(lamp OR loop-mediated OR isothermal OR amplification)) in *title*, *abstract* and *keyword* fields – 41 results

Biosis (http://apps.webofknowledge.com/): **TOPIC:** (buruli OR ulcerans) AND **TOPIC:** (diagnos* OR detect*) *AND* **TOPIC:** (lamp OR loop-mediated OR isothermal OR amplification) in *topic* fields– 31 results

EMBASE (http://www.embase.com/login): ((buruli or ulcerans) and (diagnos* or detect*) and (lamp or loop-mediated or isothermal or amplification)).af. – 53 results

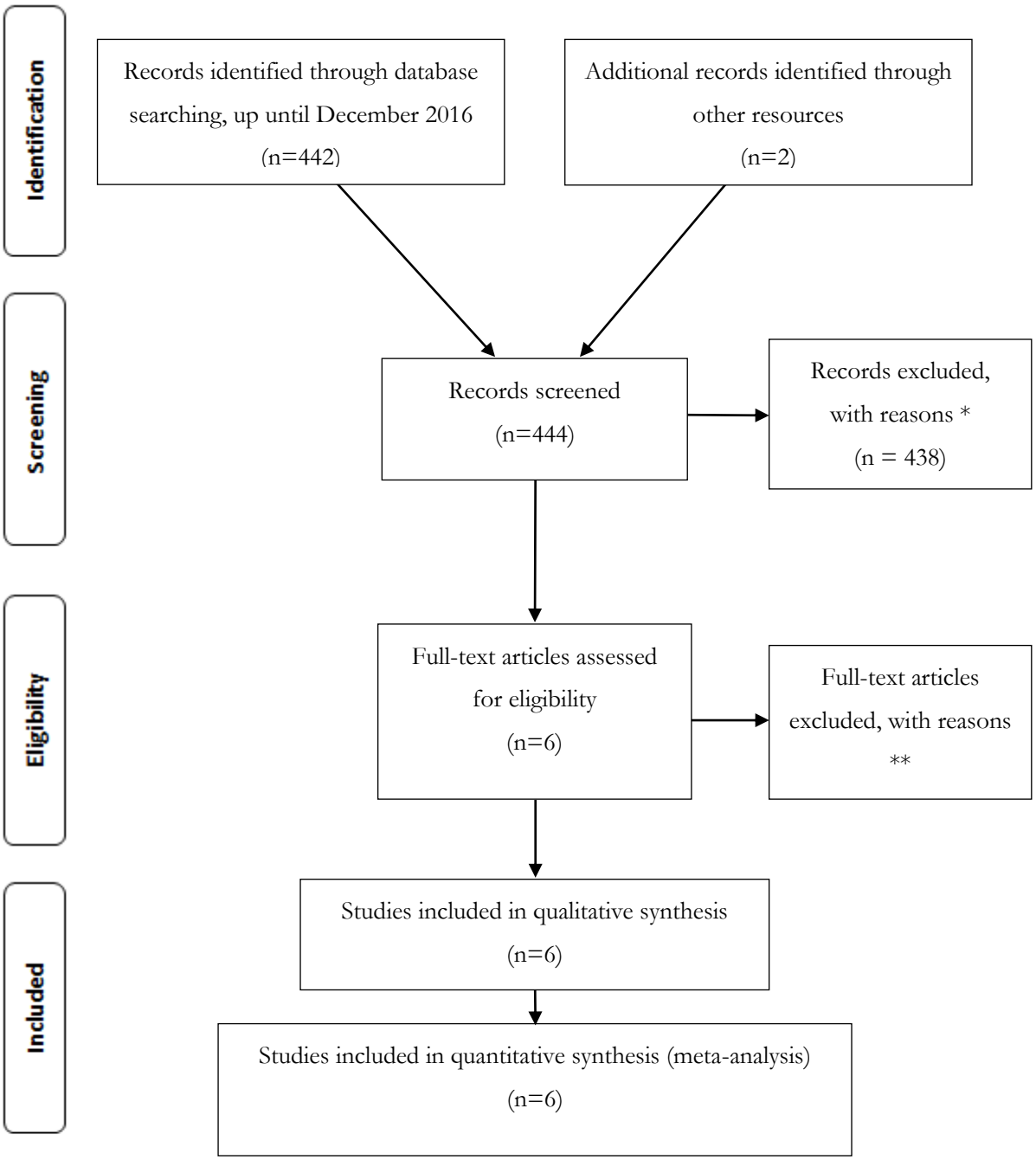
Global Health Database (Ovid)
(http://www.ovid.com/site/catalog/databases/30.jsp?cid=ovid-web-OvidHomepage): ((buruli or ulcerans) and (diagnos* or detect*) and (lamp or loop-mediated or isothermal or amplification)).af. – 292 results

African Journals Online (http://www.ajol.info/index.php/index/search/search): (buruli OR ulcerans) AND (diagnos* OR detect*) – no results

***** Although technically use of the wildcard expressions *diagnos** and *detect** should also include the expressions *diagnosis* and *detection*, this was not found to be true for Medline; therefore, both the wildcard terms and the full terms were included in the search strategy to obtain the maximum number of publications.

8.2.3. PRISMA flow diagram

(Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009)



*** Exclusion criteria:** Not original studies assessing performance of LAMP as compared to PCR (or modifications of the PCR protocol), or not containing paired data for comparison at the per subject level.

**** Exclusion criteria:** Lack of a comparator assay, or incomplete or poor quality of comparator diagnosis, or duplication of publications, or review of previously published studies, or reporting errors or conflicting data.

8.2.4. Data extraction

	Reference short	Experiment number and name	Sample preparation	Heat source	n (+)	n (+) conc	n (-)	n (-) conc	pos rate	sn (comp to ref std)	95% CI for sn[a]	sp (comp to ref std)	n (samples)	Spec
1	de Souza *et al.* (2012)	*pMUM001* LAMP	boiled[a]	thermal cycler, water bath[b]	6	2	16	15	0.27	0.67	0.09-0.99	0.79	22	1
2		*pMUM001* LAMP	purified (Qiagen)	thermal cycler, water bath[b]	13	3	9	9	0.59	1.00	0.19-1.00	0.47		2
		IS2404 PCR	purified (Qiagen)	thermal cycler	3		19		0.14					3
3	Ablordey *et al.* (2012)	*IS2404* LAMP	unboiled extract [c]	pocket warmer	12	12	18	9	0.40	0.57	0.34-0.78	1.00	30	4
4		*IS2404* LAMP	boiled extract [d]	pocket warmer	9	9	21	9	0.30	0.43	0.22-0.66	1.00		
5		*IS2404* LAMP	purified (modified Boom [e])	pocker warmer	19	19	11	9	0.63	0.90	0.70-0.99	1.00		
6		*IS2404* LAMP	unboiled extract [c]	heat block	12	12	18	9	0.40	0.57	0.34-0.78	1.00		
7		*IS2404* LAMP	boiled extract [d]	heat block	9	9	21	9	0.30	0.43	0.22-0.66	1.00		
8		*IS2404* LAMP	purified (modified Boom [e])	heat block	21	21	9	9	0.70	1.00	0.77-1.00	1.00		
		IS2404 PCR	all of above [c,d,e] - same results	thermal cycler	21		9		0.70				5	
9	Njiru *et al.* (2012)	*IS2404* LAMP	purified (Qiagen) [f]	thermal cycler	17	15	119	116	0.13	0.83	0.59-0.96	0.98	136	6
		IS2404 TaqMan PCR	purified (Qiagen) [f]	thermal cycler	18		118		0.13					7
10		*IS2404* LAMP	purified (Qiagen) [f]	thermal cycler	44	42	35	27	0.56	0.84	0.71-0.93	0.93	79	8
		IS2404 PCR	purified (Qiagen) [f]	thermal cycler	50		29		0.63				9	
11	Ahortor (2014)	*IS2404* LAMP	purified (modified Boom [e])	heat block	52	52	12	10	0.81	0.96	0.87-1.00	1.00	64	10
12		*IS2404* LAMP	syringe-based	heat block	46	46	18	10	0.72	0.85	0.73-0.93	1.00		
13		*IS2404* LAMP	syringe-based	pocket warmer	46	46	18	10	0.72	0.85	0.73-0.93	1.00		
14		*IS2404* LAMP	syringe-based	NINA	46	46	18	10	0.72	0.85	0.73-0.93	1.00		
15		IS2404 PCR	syringe-based	thermal cycler	49	49	15	10	0.77	0.91	0.80-0.97	1.00		11
		IS2404 PCR	purified (modified Boom [e])	thermal cycler	54		10		0.84					12
16	Beissner *et al.* (2015)	*IS2404* LAMP	purified (Qiagen Gentra Puregene, modified)	heat block	57	57	34	23	0.63	0.84	0.75-0.93	1.00	91	13
17		*IS2404* (c/DRB) PCR	purified (Qiagen Gentra Puregene, modified)	thermal cycler	59	59	32	23	0.65	0.87	0.79-0.95	1.00		14
		IS2404 qPCR	purified (Qiagen Gentra Puregene, modified)	thermal cycler	68		23		0.75					15
18		*IS2404* LAMP	purified (Qiagen Gentra Puregene, modified)	heat block	23	23	9	8	0.72	0.96	0.88-1.00	1.00	32	
19		*IS2404* DRB LAMP	purified (Qiagen Gentra Puregene, modified)	heat block	22	22	10	8	0.69	0.92	0.81-1.00	1.00		
20		*IS2404* (c/DRB) PCR	purified (Qiagen Gentra Puregene, modified)	thermal cycler	24	24	8	8	0.75	1.00	0.80-1.00	1.00		16
		IS2404 qPCR	purified (Qiagen Gentra Puregene, modified)	thermal cycler	24		8		0.75					17
	Ablordey, de Souza (2015)													
21		*IS2404* LAMP	syringe-based	thermal cycler	70	70	66	42	0.51	0.74	0.64-0.83	1.00	136	18
22		*IS2404* LAMP	syringe-based	thermal cycler	60	60	76	42	0.44	0.64	0.53-0.73	1.00		19
23		*IS2606* LAMP	syringe-based	thermal cycler	39	39	97	42	0.29	0.41	0.31-0.52	1.00		20
24		*IS2404* PCR	syringe-based	thermal cycler	80	80	56	42	0.59	0.85	0.76-0.92	1.00		21
25		*IS2404* LAMP	syringe-based (using lysis buffer of modified Boom [e,h])	thermal cycler	71	70	65	41	0.52	0.74	0.64-0.83	0.98		22
26		*IS2404* LAMP	syringe-based (using lysis buffer of modified Boom [e,h])	thermal cycler	63	62	73	41	0.46	0.66	0.64-0.83	0.98		23
27		*IS2606* LAMP	syringe-based (using lysis buffer of modified Boom [e,h])	thermal cycler	28	27	108	41	0.21	0.29	0.20-0.39	0.98		24
28		IS2404 PCR	syringe-based (using lysis buffer of modified Boom [e,h])	thermal cycler	71	71	65	42	0.52	0.76	0.66-0.84	1.00		25
29		*IS2404* LAMP	purified (modified Boom [e])	thermal cycler	87	86	49	41	0.64	0.91	0.84-0.96	0.98		26
30		*IS2404* LAMP	purified (modified Boom [e])	thermal cycler	86	81	50	37	0.63	0.86	0.78-0.92	0.88		27
31		*IS2606* LAMP	purified (modified Boom [e])	thermal cycler	54	54	82	42	0.40	0.57	0.88-1.00	1.00		28
		IS2404 PCR	purified (modified Boom [e])	thermal cycler	94		42		0.69					29
													568	

References (Columns *Sample preparation* and *Heat source*)

a boiled: ground in mortar, incubation at 90°C for 20 mins

b paper reports 'similar results' for both

c 250µl suspension of unboiled specimen

d 250µl suspension of specimen boiled for 10 mins and centrifuged for 5 mins at 14000 rpm, supernatant used

e (Durnez et al., 2009)

f Swabs and homogenized fresh tissue biopsies: Roche AMPLICOR respiratory specimen preparation kit (Roche Diagnostics Co., Indianapolis, IN), followed by a cleanup to remove potential PCR inhibitors by using the QIAamp DNA Mini kit (QIAGEN Inc., Valencia, CA). Formalin-fixed, paraffin-embedded tissue sections: Following a 24-h proteinase K digestion of dewaxed samples, using the QIAamp DNA Mini kit. (Fyfe et al., 2007)

g Calculated using epi.tests in R, with the exception of Beissner (2015), where values were reported in the publication

h Without glass beads

References (Column: *Spec*):

1 Samples: 5 swabs in no medium, 10 swabs in medium, 4 FNAs, 3 punch biopsies, LAMP target reported as mycolactone encoding plasmid sequence (*pMUM001*), is *IS2606*

2 LAMP target reported as mycolactone encoding plasmid sequence (*pMUM001*), is *IS2606*

3 PCR targeting *IS2404*, single round as reference standard.

4 Samples: 20 swabs in no medium, 10 FNAs

5 PCR targeting *IS2404*, two rounds, with 1µl of product carried over

6 Origin: Australia, Samples: Swabs, tissue biopsy, fixed and paraffin embedded tissue

7 PCR = TaqMan RT-PCR, Kappa = 0.87 (reported; calculated: 0.84)

8 Origin: Ghana, Samples: Swabs, tissue biopsy, fixed and paraffin embedded tissue

9 PCR targeting *IS2404* one rounds, Kappa = 0.74 (reported and calculated)

10 Unclear which samples were used for LAMP evaluation. All positives concordant (Ahortor, pers. communication)

11 PCR targeting *IS2404* two rounds, with 1µl of product carried over

12 PCR targeting *IS2404* two rounds, with 1µl of product carried over

13 Samples from 91 suspected cases: 66 FNAs, 32 swabs, 42 punch biopsies; 95% CI reported in publication using McNemar chi-square for matched pairs of samples with categorical test results. Results reported by cases, in order to be in line with remaining publications.

14 Conventional PCR and dry reagent based (DRB) PCR showed identical performance, both were targeting *IS2404*

15 No false positives as compared to qPCR observed

16 Conventional PCR and DRB PCR showed identical performance, both were targeting *IS2404*

17 No false positives as compared to qPCR observed

18 'Ablordey primers'

19 'Njiru primers'

20 'Dziedzom/de Souza primers'

21 PCR targeting *IS2404* two rounds, with 1µl of product carried over

22 'Ablordey primers'

23 'Njiru primers'

24 ’Dziedzom/de Souza primers'

25 PCR targeting *IS2404* two rounds, with 1µl of product carried over

26 ‘Ablordey primers'

27 ‘Njiru primers'

28 ‘Dziedzom/de Souza primers'

29 PCR targeting *IS2404* two rounds, with 1µl of product carried over

8.2.5. Meta-analysis of data

8.2.5.1. Assumptions for the model are discussed in 2.2.3.6 Quantitative data analysis

Analysis of experimental data and sample size calculations were conducted using the epi.tests package in R (3.3.0).

.

s_LAMPMetaAnalysis_update190816.R

```
rm(list = ls())

setwd("")

## EpiR for performance calculations

## http://www.inside-r.org/packages/cran/epiR/docs/epi.tests

library(epiR)

desouza1<-matrix(c(2,1,4,15),2,2)

desouza2<-matrix(c(3,0,10,9),2,2)

ablordey3<-matrix(c(12,9,0,9),2,2)

ablordey4<-matrix(c(9,12,0,9),2,2)

ablordey5<-matrix(c(19,2,0,9),2,2)

ablordey6<-matrix(c(12,9,0,9),2,2)

ablordey7<-matrix(c(9,12,0,9),2,2)

ablordey8<-matrix(c(21,0,0,9),2,2)

njiru9<-matrix(c(15,3,2,116),2,2)

njiru10<-matrix(c(42,8,2,27),2,2)
```

```
ahortor11<-matrix(c(52,2,0,10),2,2)

ahortor12<-matrix(c(46,8,0,10),2,2)

ahortor13<-matrix(c(46,8,0,10),2,2)

ahortor14<-matrix(c(46,8,0,10),2,2)

ahortor15<-matrix(c(49,5,0,10),2,2)

beissner16<-matrix(c(57,11,0,23),2,2)

beissner17<-matrix(c(59,9,0,23),2,2)

beissner18<-matrix(c(23,1,0,8),2,2)

beissner19<-matrix(c(22,2,0,8),2,2)

beissner20<-matrix(c(24,0,0,8),2,2)

ablordey21<-matrix(c(70,24,0,42),2,2)

ablordey22<-matrix(c(60,34,0,42),2,2)

ablordey23<-matrix(c(39,55,0,42),2,2)

ablordey24<-matrix(c(80,14,0,42),2,2)

ablordey25<-matrix(c(70,24,1,41),2,2)

ablordey26<-matrix(c(62,32,1,41),2,2)

ablordey27<-matrix(c(27,67,1,41),2,2)

ablordey28<-matrix(c(71,23,0,42),2,2)

ablordey29<-matrix(c(86,8,1,41),2,2)

ablordey30<-matrix(c(81,13,5,37),2,2)
```

```
ablordey31<-matrix(c(54,40,0,42),2,2)

dat<-ablordey31

colnames(dat) <- c("Ref+","Ref-")

rownames(dat) <- c("LAMP+","LAMP-")

rval <- epi.tests(dat, conf.level = 0.95)

print(rval); summary(rval)

# Input data from experiments --------

## Sample sizes for sensitivity and specificity respectively (PCR detected)

## Have one repeated-entries for different tests within one paper

## Updated for v.5 (220716) on 190816

NPos <- c(3,21,18,50,54,68,24,94)

NNeg <- c(19,9,118,29,10,23,8,42)

## Numbers correctly identified by LAMP for each test

## Updated for v.5 (220716) on 190816

YPos <- c(3,19,15,42,46,57,22,70)

YNeg <- c(9,9,116,27,10,23,8,41)

lenPos <- length(YPos)

lenNeg <- length(YNeg)

lPos <- YPos/NPos

lNeg <- YNeg/NNeg
```

```
# Fit model using Stan ------

library(rstan)

# library(shinystan)

options(mc.cores=4)

fit <-
stan('metaAnalysis.stan',data=list(lenPos=lenPos,lenNeg=lenNeg,NPos=NPos,NNeg=NNeg,YPos
=YPos,YNeg=YNeg),

        iter=20,chains=1)

fit <-
sampling(get_stanmodel(fit),data=list(lenPos=lenPos,lenNeg=lenNeg,NPos=NPos,NNeg=NNeg,
YPos=YPos,YNeg=YNeg),

        iter=8000,chains=4)

print(fit)

aFit <- as.shinystan(fit)

launch_shinystan(aFit)

## Extract posterior samples from overall sensitivity and specificity

lSn <- extract(fit,'theta_sn_average')[[1]]

lSp <- extract(fit,'theta_sp_average')[[1]]

hist(lSn)

hist(lSp)

quantile(lSp,0.1)

write.csv(file='LAMPSensitivity.csv',lSn)
```

```
write.csv(file='LAMPSpecificity.csv',lSp)

## Extract out the individual-experiment sensitivities and specificities

lSn <- extract(fit,'theta_sn')[[1]]

lSp <- extract(fit,'theta_sp')[[1]]

write.csv(file='LAMPIndividualSensitivity.csv',lSn)

write.csv(file='LAMPIndividualSpecificity.csv',lSp)
```

metaAnalysis_nonCentered.stan

```
data{
  int lenPos;
  int lenNeg;
  int YPos[lenPos];
  int YNeg[lenNeg];
  int NPos[lenPos];
  int NNeg[lenNeg];
}
parameters{
  real mu_sn_raw[lenPos];
  real mu_sp_raw[lenNeg];
  real mu_sn_overall;
```

```
  real mu_sp_overall;
  real<lower=0> sigma_sn;
  real<lower=0> sigma_sp;
  }
transformed parameters{
  real<lower=0,upper=1> theta_sn[lenPos];
  real<lower=0,upper=1> theta_sp[lenNeg];
  real mu_sn[lenPos];
  real mu_sp[lenNeg];
    for(i in 1:lenPos){
    mu_sn[i] = mu_sn_overall + mu_sn_raw[i]*sigma_sn;
    theta_sn[i] = inv_logit(mu_sn[i]);
  }
  for(i in 1:lenNeg){
    mu_sp[i] = mu_sp_overall + mu_sp_raw[i]*sigma_sp;
    theta_sp[i] = inv_logit(mu_sp[i]);
  }
}
model{
  ## Likelihood for sensitivity
```

```
  for(i in 1:lenPos){
    YPos[i] ~ binomial_logit(NPos[i],mu_sn[i]);
  }
  ## Likelihood for specificity
  for(i in 1:lenNeg){
    YNeg[i] ~ binomial_logit(NNeg[i],mu_sp[i]);
  }
    ## Priors
  mu_sn_raw ~ normal(0,1);
  mu_sp_raw ~ normal(0,1);
    # Hyper-priors
  mu_sn_overall ~ normal(0,5);
  mu_sp_overall ~ normal(0,5);
  sigma_sn ~ cauchy(0,2);
  sigma_sp ~ cauchy(0,2);
}
generated quantities{
  real mu_sn_average;
  real mu_sp_average;
  real<lower=0,upper=1> theta_sn_average;
```

```
  real<lower=0,upper=1> theta_sp_average;

    mu_sn_average = normal_rng(mu_sn_overall,sigma_sn);

  mu_sp_average = normal_rng(mu_sp_overall,sigma_sp);

    theta_sn_average = inv_logit(mu_sn_average);

  theta_sp_average = inv_logit(mu_sp_average);

}
```

8.3. Appendix 3: Consent form for interview and focus group participants and volunteer agreement

CONSENT FORM

FOR INTERVIEW AND FOCUS GROUP PARTICIPANTS

Title: A study within a trial looking at methodological approaches to evaluating a diagnostic test for Buruli Ulcer in a field setting

Principal Investigator: Astrid Christine Erber (MSc.)

Address and Contact Information: Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine Research Building (NDMRB), University of Oxford, Old Road Campus, Roosevelt Drive, Headington, OX3 7FZ, U.K..

E-Mail: astrid.erber@ndm.ox.ac.uk, telephone number: +44 18656 12963 (UK), +233 574294390 (GHA).

General Information about Research

This study is a so-called study within a trial, or SWAT. It aims to help understand the influence of individual, process and context factors on the performance of a diagnostic test under field conditions. It is embedded in a trial evaluating a newly developed diagnostic test for detection of Buruli Ulcer, performed under the leadership of Dr. Anthony Ablordey of the Noguchi Memorial Institute.

This study uses individual interviews and a focus group discussion with researchers and research staff, as well as health care professionals involved with the trial. It also uses results from the main trial on how the diagnostic test performs under different conditions.

Because of your role in the trial, we would greatly appreciate your collaboration!

- You will be asked for two interviews, with one hour duration each, about your experiences with the evaluation trial. These are open interviews and will not use questionnaires. There will be 2-3 weeks between the two interviews.

- The focus group's aim is to discuss the trial from a process point of view. It involves as many people from the group as possible, and will take about 2 hours. It will take place during the evaluation trial, around a week after the first interview.

Possible Risks and Discomforts

We do not envisage any harm from this study. In order to minimize discomfort, refreshments will be provided for you during the interviews and the focus group discussion.

Possible Benefits

Participation in this study is voluntary and not remunerated. There will be no direct benefits from participating. However, this will be an opportunity for you to share your knowledge and experience.

Confidentiality

We will protect information about you to the best of our ability. Your personal data will be kept confidential, and you will not be named in any reports or publications. Audio recordings, transcripts and obtained data will be kept on the person of the researcher or in a locked vehicle or room during the study. After study closure, data will be stored electronically and in paper form in a locked cabinet in the project office in Oxford, UK. Astrid Erber, the researcher, may sometimes look at the records.

Compensation

Participation is voluntary, and no compensation will be given.

Voluntary Participation and Right to Leave the Research

All participation in research is voluntary. You are free to decide if you want to take part or not. If you do agree to take part now, you can still change your mind by notifying Astrid Erber at any time, without consequences.

Contacts for Additional Information

If you have problems or questions relating to this project, please call Dr. Anthony Ablordey on 0275652022 or Astrid Erber (Principal Investigator), astrid.erber@ndm.ox.ac.uk, telephone number: +44 18656 12963 (UK), +233 574294390 (GHA), Prof. Trudie Lang (PhD supervisor), trudie.lang@ndm.ox.ac.uk, telephone number: +44 1865612955 (UK).

Your Rights as a Participant

This research has been reviewed and approved by the Institutional Review Board of Noguchi Memorial Institute for Medical Research (NMIMR-IRB) (Reference number 026/10-11). If you have any questions about your rights as a research participant you can contact the IRB Office between the hours of 8am-5pm through the landline 0302916438 or email addresses: nirb@noguchi.mimcom.org

This research has been reviewed and approved by the Oxford Tropical Research Ethics Committee (OxTREC), University of Oxford, Oxford, UK (Reference number 568-15).

VOLUNTEER AGREEMENT

The above document describing the benefits, risks and procedures for the research title A study within a trial looking at methodological approaches to evaluating a diagnostic test for Buruli Ulcer in a field setting has been read and explained to me. I have been given an opportunity to have any questions about the research answered to my satisfaction. I agree to participate as a volunteer.

_____ _____

Date Name and signature or mark of volunteer

If volunteers cannot read the form themselves, a witness must sign here:

I was present while the benefits, risks and procedures were read to the volunteer. All questions were answered and the volunteer has agreed to take part in the research.

_____ _____

Date Name and signature of witness

I certify that the nature and purpose, the potential benefits, and possible risks associated with participating in this research have been explained to the above individual.

_____ _____

Date Name and Signature of Person Who Obtained Consent

8.4. Appendix 4: Ethical approvals

NOGUCHI MEMORIAL INSTITUTE FOR MEDICAL RESEARCH
Established 1979
A Constituent of the College of Health Sciences
University of Ghana

INSTITUTIONAL REVIEW BOARD

Phone: +233-302-916438 (Direct)
+233-289-522574
Fax: +233-302-502182/513202
E-mail: nirb@noguchi.mimcom.org
Telex No: 2556 UGL GH

Post Office Box LG 581
Legon, Accra
Ghana

My Ref. No: DF.22
Your Ref. No:

6th May, 2015

ETHICAL CLEARANCE

FEDERALWIDE ASSURANCE FWA 00001824 **IRB 00001276**

NMIMR-IRB CPN 026/10-11 *amend. 2015* **0000908 IORG**

On 6th May 2015, the Noguchi Memorial Institute for Medical Research (NMIMR) Institutional Review Board (IRB) at a full board meeting conducted continuing review and and amended your protocol titled:

TITLE OF PROTOCOL : **Scaling up early detection and treatment to reduce Buruli Ulcer morbidity in the Asante Akim North District of Ghana (A study within a trial looking at methodolical approaches to evaluating a diagnostic test for Buruli Ulcer in a field setting)**

PRINCIPAL INVESTIGATOR : **Anthony Ablordey, PhD**

CO – INVESTIGATORS : **Dr. William Amponsah & Astrid Christine Erber**

Please note that a final review report must be submitted to the Board at the completion of the study. Your research records may be audited at any time during or after the implementation.

Any modification of this research project must be submitted to the IRB for review and approval prior to implementation.

Please report all serious adverse events related to this study to NMIMR-IRB within seven days verbally and fourteen days in writing.

This certificate is valid till 5th May, 2016. You are to submit annual reports for continuing review.

Signature of Chair: .....
Mrs. Chris Dadzie
(NMIMR – IRB, Chair)

cc: Professor Kwadwo Koram
Director, Noguchi Memorial Institute
for Medical Research, University of Ghana, Legon

Oxford Tropical Research Ethics Committee

University of Oxford
Research Services, University Offices
Wellington Square, Oxford OX1 2JD
Tel. +44 (0)1865 (2)82106
E-mail: oxtrec@admin.ox.ac.uk

Ms Astrid Erber
Green Templeton College
43 Woodstock Road
Oxford
OX2 6HG

06 August 2015

Dear Ms Erber

Full Title of Study: A study within a trial looking at methodological approaches to evaluating a diagnostic test for Buruli Ulcer in a field setting

OXTREC Reference: 568-15

Thank you for your email of the 3 August 2015, and for your minimal risk application form. Other documents reviewed were:

Documents:	Version:	Date:
Protocol (and included documents)	1.2	17/7/15

The OxTREC executive team reviewed the above application on the 6 August 2015 and gave approval for this study.

Approval is given for the first five years and is subject to receiving the local ethical approval (if this approval has not yet been received).

Any subsequent changes to the application must be submitted to the Committee as an Amendment. This should include a letter to give the reasons for the proposed modifications and all revised documents with changes tracked.

Please ensure that you submit a completed Annual Report form on every anniversary of this approval and a final End of Study Report. The relevant forms can be found on the OxTREC website: http://www.tropicalmedicine.ox.ac.uk/oxtrec.

Yours sincerely

Mary Warrell

Dr Mary Warrell
OxTREC Chairman

Direct Line Tel: +44 (0)1865 (2)82106
Email: oxtrec@admin.ox.ac.uk
Web: www.admin.ox.ac.uk/rso/

8.5. Appendix 5: Patient information sheet and informed consent declaration

(Master protocol versions)

Patient Information Sheet

Disease Experiences of Cutaneous Leishmaniasis (CL) Patients

[please insert your institutional affiliation here]

and

Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford

and

UNICEF/UNDP/World Bank/WHO Special Programme for Research & Training in Tropical Diseases (TDR), Geneva, Switzerland

You are being asked to take part in a study. Before you decide whether or not to take part, please take time to read the following information carefully. It is important for you to understand why the work is being done and to take time to decide whether or not you wish to take part.

What is the purpose of the study?

This is a study which looks at the experiences of cutaneous leishmaniasis (CL) patients. Interviews are conducted to learn about issues that are important from a patient's perspective with regard to treatment, and inclusion in clinical trials (research studies that test how well new medical drugs work in humans).

Why have I been chosen?

You have been chosen because you have been diagnosed with CL, either recently or in the past.

Do I have to take part?

No. It is entirely voluntary. You are able to ask the researchers questions before you decide. If you do decide to take part you can still withdraw at any time without giving reasons. Your current treatment for CL, or treatment in the future, current or actual medical attention, or participation in clinical trials will not be affected in any way.

What will happen to me if I take part?

You will be interviewed by one researcher, who is the local principal investigator (PI). You will discuss issues with regards to your disease experience. Some questions will touch upon sensitive issues, like disability, or social interactions. Examples are: *Is there anything you can't do or no longer do because of your disease?*, or *How did people respond to your disease?* There are no right or wrong answers - we are interested in your disease experience. Providing you agree, the discussion will be audio recorded, and the researcher will take notes. The interview will last around 60 minutes.

Will my part in this study be kept confidential?

All information collected about you during the study will be kept confidential. Details which might identify you will be either removed or changed so that you cannot be recognized.

What will happen to the results of the study?

The interviews will be transcribed (written down), and translated into English. Names will be replaced with a code ID. Electronic recordings (audio files) of the interviews will be kept only by the researcher. Transcripts and their English translations and notes will be stored in a locked cabinet, or electronically, and can only be accessed by this researcher (or designated persons) and researchers at the University of Oxford for data analysis. Data will be kept for 5 years.

Interviews will be analyzed to identify elements that are important from a CL patient's perspective. Any publications or presentations arising from this research will not identify any person. Findings will be presented in scientific meetings, and published. The researcher will inform you about the results and explain them to you.

Will I benefit from taking part in this study?

You may not benefit directly from taking part. We hope that, within this study, we will find out about elements that are important from a patient's perspective with regards to CL treatment, and enrollment in clinical trials for new CL treatments - and will therefore benefit future patients. There are no risks involved in taking part. Participation in the interviews is completely independent of any treatment or medical attention you are currently receiving, or which you might need in the future. It is also independent of your participation in clinical trials that might take place in the future.

You may need to absent from work to attend the interview, and other costs (such as for transportation) may incur. In this case, you will receive compensation for your time, and reimbursement of travel costs (if applicable). Please understand that this does not represent payment for your participation in the interview.

What if there is a problem?

If you have a concern about any part of this study, please speak to the researcher who will do his or her best to answer your query. If you remain unhappy and wish to make a formal complaint, please

contact the Chair of the [please insert name and contact details for the contact person at your local ethics committee here].

Who has reviewed the study?

This study has been reviewed and cleared by the following ethics committees:

- [please insert your ethics committee here]

- the World Health Organization Ethics Research Committee (WHO ERC)

- Oxford Tropical Research Ethics Committee (OxTREC)

Contact for further information

[please insert the name and contact data of the person consenting here, if this is done by someone else, or your name]

Informed Consent Declaration

Disease Experiences of Cutaneous Leishmaniasis (CL) Patients

[please insert your institutional affiliation here]

and

Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford

and

UNICEF/UNDP/World Bank/WHO Special Programme for Research & Training in Tropical Diseases (TDR), Geneva, Switzerland

Researcher:

[please insert your name and contact data here]

1. I confirm that I have read and understood the information sheet for the above study and have had the opportunity to ask questions. I have received satisfactory answers to these questions.

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason and without any legal rights being affected.

3. I understand that the study has been reviewed by the following ethics committees: [please insert your ethics committee here], the World Health Organization Ethics Research Committee (WHO ERC), Oxford Tropical Research Ethics Committee (OxTREC), and has received clearance.

4. I understand that only the research team will have access to the data for analysis. The data will be stored safely for five years.

5. If I am interviewed, anything I say during the course of this may be recorded electronically and used as anonymous quotes in any presentation of the research.

6. If I receive compensation for my time and reimbursement of travel costs (if applicable), I also understand that this does not represent payment for my participation in the interview.

7. I agree to take part in the above study.

8. I understand how to raise a concern and make a complaint.

[please modify adapt this signature page according to your local IRB requirements, e.g. fingerprints, number of witnesses – below a generic example]

_____	_____	_____
Name of participant	date	signature or fingerprint

_____ _____ _____

Name of researcher date signature

_____ _____ _____

Name of witness date signature

8.6. Appendix 6: Ethical approvals

World Health Organization

Research Ethics Review Committee (WHO ERC)

20, AVENUE APPIA – CH-1211 GENEVA 27 – SWITZERLAND – HTTP://INTRANET.WHO.INT/HOMES/RPC/ERC – HTTP://WWW.WHO.INT/RPC/RESEARCH_ETHICS

WHO ERC
Review Summary

Protocol ID: ERC.0002635
Country: N/A
Protocol Title: Disease Experiences of Cutaneous Leishmaniasis (CL) Patients. Patient Interviews – Master Protocol
Version: 1.0 **Dated**: 15/07/2015
WHO Responsible Staff Member: Piero Luigi Olliaro
Responsible Unit: WHO/HQ/HTM/TDR
Meeting Date:

Dear Dr. Piero Luigi Olliaro,
Please find the review summary of the Protocol "Disease Experiences of Cutaneous Leishmaniasis (CL) Patients. Patient Interviews – Master Protocol", which was submitted to the Secretariat on 17/07/2015. This proposal underwent expedited review.

The outcome of the review is provided below. When responding, please submit the following:

1. A cover memorandum that addresses your responses, POINT BY POINT, to each of the queries in sections A and B.
   Section C contains Suggestions to improve the proposal but there is no obligation to follow them.

2. An Amended protocol including the responses in bold, highlighted or in track changes. The protocol should include all relevant documentation (ICF, study instruments, peer review, etc.) even if already submitted.

Please note that comments in the introductory paragraph are meant for the WHO Responsible Staff Member, though you may decide to share them with the PI.

PLEASE RESPOND TO THIS REVIEW SUMMARY WITHIN A 3 MONTH PERIOD, OR PROVIDE THE ERC SECRETARIAT A VALID JUSTIFICATION FOR THE DELAY.

A. Amendments (Response and change required)
This section includes queries and comments on your protocol, study instruments or the informed consent form for which the ERC requires your response and where relevant, appropriate amendments to the protocol, study instruments or the informed consent.

1. Protocol
1.1. Please provide an amended proposal specifying the version number and/or date on each page.
1.2. In relation to the responses provided to the comments made by the ERC Secretariat:
1.2.1. Please include them in the corresponding sections of the protocol.
1.2.2. Please expand and clearly indicate who will owe the data and how it will be used.
1.3. Researchers have an obligation to report study results back to participants and their communities. Please describe the dissemination plan that will be followed to meet this obligation. Please include this information also in the consent documents.
1.4. As per the protocol, an "appropriate alternative" will be offered to obtain consent from those who cannot give their consent by signature. Please confirm that an independent witness will be chosen by the participant and describe the process.

2. Informed Consent Forms
2.1. Some of the terms provided in the information sheet could be further explained or substituted (e.g., clinical trial) to ensure comprehension by potential participants.
2.2. Please indicate not only how long the interview will last but also the anticipated duration of the study.

2.3. Some of the questions included in the semi-structured interview can be embarrassing to participants. Please inform participants about the nature of the questions and include a sample of such questions in the information sheet so that they are fully aware of what will be expected of them as research participants.

B. Clarifications (Response required but change may not be required)

This section includes queries on your protocol, study instruments or the informed consent form for which the ERC requires a clarification, and it may not be mandatory for you to make changes to your protocol. Please consider the comments of the ERC and determine if you believe change is needed. If no change is made, the ERC will consider the response. If the judgement of the ERC is that a change should occur, the ERC will promptly notify you.

NIL

C. Suggestions

This section consists of suggestions for alternative scientific or technical approaches or methods for conducting the research but which do not raise critical, ethical issues. These are meant to be helpful to investigators and are presented as suggestions for you to consider incorporating into a revised protocol. No response from you is required for any comment in this section. If, however, you do make changes to the protocol as a result of these suggestions, please submit the revised protocol to the ERC.

NIL

World Health Organization

Research Ethics Review Committee (WHO ERC)

20, AVENUE APPIA – CH-1211 GENEVA 27 – SWITZERLAND – HTTP://INTRANET.WHO.INT/HOMES/RPO/ERC – HTTP://WWW.WHO.INT/RPC/RESEARCH_ETHICS

Protocol ID: ERC.0002635 **Meeting Date**:

Based on the above comments, the Committee has the following recommendation(s) for this proposal:

[ ] The proposal is ***Approved as submitted.*** No modifications are required.

[X] The proposal is ***Conditionally Approved; requires amendments and/or clarifications.*** Final approval is contingent upon an adequate response by the Principal Investigator, to the satisfaction of the reviewers or the Chair on behalf of the ERC.

[ ] The proposal is ***Not approved; requires additional information and/or rewriting.*** A revised version of the proposal should be re-submitted by the WHO responsible staff member as a new submission to the ERC for re-review by Committee.

[ ] The proposal is ***Rejected***. The proposal is ethically unacceptable, for the reasons stated above. The Principal Investigator may submit a new proposal that takes into consideration the ethical issues raised by the Committee. If you do not agree with the Committee's assessment, please feel free to submit an appeal to the Chair of the ERC, through the Secretariat.

NOTE: Final Approval of the Proposal is contingent upon submission of the following:

[ ] Local ethics approval(s) [ ] Other relevant documents

The ERC would like to receive a copy of the recommendations of the local ethics committee when available.

IMPORTANT

1. *Any changes to the proposal or to the attachments (informed consent/study instruments etc.) should be approved by ERC before being implemented.*
2. *The approval for this proposal is valid for a period of one year only.*
3. *Please resubmit this proposal for a Continuing Review at least 2 months before the next re-approval period.*

Chairperson ..... **Date…3rd Sept 2015**.....

Name: Melba Gomes~~/Nigel Rollins /Emilie Alirol~~

FINAL APPROVAL

Amendments and Clarifications to the proposal have been reviewed.
The protocol (Version: 1.2 Date: 21.10.2015) and informed consent Forms (Dated: 21.10.2015) submitted on 21.10.2015
are approved by the ERC

Chairperson .....
MG/NR/EA Melba Gomes
Name .....
23rd October 2015
Date .....

FOR THE SECRETARIAT

Amendments and Clarifications to be reviewed:
[ ] Electronically by ERC
[ ] by Primary reviewers
[X] by Secretariat

Amendments approved /
Clarifications accepted on
Local ERC approval(s) obtained on
N/A (Master protocol)
Relevant Documents submitted on

Comments:

Signature Date 21.10.2015

Oxford Tropical Research Ethics Committee

University of Oxford
Research Services, University Offices
Wellington Square, Oxford OX1 2JD
Tel. +44 (0)1865 (2)82106
E-mail: oxtrec@admin.ox.ac.uk

Ms Astrid Erber 7 August 2015
The Global Health Network
Nuffield Department of Medicine Research Building
Centre for Tropical Medicine
University of Oxford
Old Road Campus, Roosevelt Drive, Oxford

Dear Ms Erber

Full Title of Study: Disease Experiences of Cutaneous Leishmaniasis (CL) Patients

OXTREC Reference: 574-15

Thank you for your email of the 04 August 2015, and for your minimal risk application form. Other documents reviewed were:

Documents:	Version:	Date:
Protocol	V1.1	27/07/15
PI signatory Page		15/07/15

The OxTREC executive team reviewed the above application on the 6 August 2015 and gave approval for this study.

Approval is given for the first five years and is subject to receiving the local ethical approval (if this approval has not yet been received).

Any subsequent changes to the application must be submitted to the Committee as an Amendment. This should include a letter to give the reasons for the proposed modifications and all revised documents with changes tracked.

Please ensure that you submit a completed Annual Report form on every anniversary of this approval and a final End of Study Report. The relevant forms can be found on the OxTREC website: http://www.tropicalmedicine.ox.ac.uk/oxtrec.

Yours sincerely

Mary Warrell

Dr Mary Warrell
OxTREC Chairman

Direct Line Tel: +44 (0)1865 (2)82106
Email: oxtrec@admin.ox.ac.uk
Web: www.admin.ox.ac.uk/rso/

8.7. Appendix 7: Background parameters informing the sampling strategy

	Brazil	Burkina-Faso	Colombia	Colombia	Ethiopia	Iran	Morocco	Peru	Tunisia
Institution/Study site	Centro de Pesquisa René Rachou (CPqRR), Fundação Oswaldo Cruz (FIOCRUZ) Minas Gerais	Centre MURAZ, Bobo-Dioulasso, Burkina Faso	Centro Internacional de Entrenamiento de Investigaciones Médicas (CIDEIM), Cali	Programa de Estudio y Control de Enfermedades Tropicales (PECET), Medellin	Ayder Referral Hospital, College of Health Sciences, Mekelle University, Mekelle	Molecular Dermatology Research Center, Shiraz University of Medical Sciences, Shiraz	National School of Public Health, Rabat, Morocco	Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia, Lima	Institut Pasteur de Tunis, Tunis
Area where the study is conducted	Belo Horizonte	Bobo-Dioulasso and Ouagadougou	Cali and Tumaco	Leishmaniasis Recovery Center, Boyacá, for military patients from North-West and South	Mekelle	Shiraz and vicinity	Sefrou, Moulay Yacoub	Lima, San Martin, Cusco	Sidi-Bouzid and Gafsa
Number of cases in country per year, as of 2014 (WHO 2016)	19402	100	11433	11433	342 (grossly under-reported)	21148 (underreported)	*L. tropica* 2095/*L. major* 460	5888	3368 (investigators estimate up to 10,000)
Most prevalent *Leishmania* species in the study area	*L. (Viannia) braziliensis*	*L. major*	*L. (Viannia) anamensis*	*L. (Viannia) braziliensis, L. (Viannia) panamensis*	*L. aethiopica*	*L. major, L. tropica*	*L.tropica*	*L. (Viannia) braziliensis and L. (Viannia) peruviana*	*L. major*
Age category of patients	adults	adults	adults, children	adults	adults, children	adults, children	adults, children	adults	adults
Gender ratio estimates in the study area (F:M)	40:60=F:M	50:50=F:M	35:65=F:M	20:80=F:M	45:55=F:M	~ 50:50=F:M	60:40=F:M	20:80=F:M	~ 50:50=F:M
Transmission in the study area (emerging or stable foci)	stable	emerging (Bobo-Dioulasso), stable (Ouagadougou)	stable	stable	both	stable	stable	stable	both
Setting (rural or (peri-) urban)	peri-urban	rural (Bobo-Dioulasso), peri-urban (Ouagadougou)	rural	rural or peri-urban	rural, peri-urban	rural and urban	rural, peri-urban	rural	rural, peri-urban
Transmission (seasonal or year-round)	year-round	seasonal (September to November)	year-round	year-round	year-round	seasonal (Peak: September-March)	seasonal (November-April)	*L. (Viannia) braziliensis* seasonal (Peak: January to June); *L. (Viannia) peruviana* year-round	seasonal (September-March)
Average number of lesions observed in the study area	1-3	> 5	1-2	70% = 1	1-3	few to many	1-2 (*L. tropica*)	1-2	1-2
Most common type of lesion observed in the study area	70% ulcer	ulcer	80% ulcer	80% ulcer	nodular plaque; infiltrated	mostly ulcerated nodule or plaque	70-85% ulcer	80% ulcer	90% ulcer
Treatment (type/dose)	Sb^V IV 1st line; amphotericin B; (liposomal amphotericin B) (Sb^V IL)	local antiseptics (Bobo Dioulasso), Sb^V IM in Ouagadougou)	Sb^V IM 1st line; miltefosine or amphotericin B 2nd line	Sb^V IM 1st line; pentamidine or miltefosine 2nd line	Sb^V IL; Sb^V IM; cryotherapy; IM pentamidine (when supplied)	cryotherapy, IL and systemic glucantime	Sb^V IL and antiseptics or Sb^V IM (depending on lesions numbers, size and localisation)	Sb^V IV 1st line; if failure amphotericin B	Sb^V IL and antiseptics

Modified from (Erber et al., 2018)

8.8. Appendix 8: Interview topic guide

A. NARRATIVE SECTION

Can you start by telling me your story about CL?

B. SYMPTOMS AND TREATMENT

Seeking treatment, why now? Types of treatment, knowledge about treatments, efficacy of treatment, side effects, satisfaction with treatment, barriers to treatment (geographical, financial, cultural).

1. What did you did do when you first found a lesion?

2. Tell me about the first time you noticed your lesion.

3. How did you feel when you noticed that you have CL?

4. How do you think you got it?

    Do you know how to prevent it?

5. What types of treatment did you look for?

6. What else did you try?

    Medical, traditional, spiritual, self treatment

    What, when, where, why, who advised you?

7. Did the treatment work?

8. Can you tell me a bit more about the treatment?

9. How did/do you feel about the treatment?

Cost, side effects, what happened, duration of treatment, number of cycles

10. How easy was it to get treatment?

11. What happened after treatment?

Follow up? Completion of treatment

12. Do you know any other treatments for CL?

Other treatments that are not available?

13. What do you think would be the best treatment for CL?

Place of treatment, pills/IV/duration of treatment

What is/was most inconvenient in your current/past treatment?

14. Who do you think should be treated?

+probing

C. EVERYDAY LIFE

Economic, social, emotional, psychological experiences, disability, limitations, work, relationships, effect on community, stigma, positive experiences, support, community.

14. Can you describe your lesion?

15. How has the lesion changed your life? (alternative) What was your life like when you had CL?

16. How did people respond to your disease?

Immediate family, friends, colleagues, community, health professionals

Positive and negative responses

17. What help did you get?

Who? What type of help? What helped you most?

Help before, during and after treatment

18. Could you tell me how CL has affected you in your work?

Modify depending on employment (carers, etc)

Impact on salary/income if applicable

19. Is there anything you can't do or no longer do because of your disease?

D. EXPECTATIONS/CONCERNS

What does cure mean to you? Feelings about treatments, cosmetic effects/scars, will I have a relapse, will I experience later complications

20. How do you think your skin condition is going?

21. What did you hope for when you started the treatment?

Treatment in general, traditional/medical/self-treatment

22. What does healing mean to you?

23. Some people are concerned about scars. How do you feel about scars?

Can you tell me a bit more about your scar? (if relevant)

24. What do you think might happen in the future about your condition?

Are you afraid that the skin condition come back (relapse/late complications). Do you think being in treatment will prevent it from coming back? Do you think you can prevent the scar with therapy?

25. Some people believe CL can come back if you don't get treatment. What do you think about this?

26. If you had the authority, what would you change for people with CL?

Messages to others

27. Finally, can I ask what does CL mean to you?

Knowledge of CL, attitudes towards CL

8.9. Appendix 9: Delphi consensus survey

8.9.1. Introduction and background

In the following, a survey among HCPs and researchers working with CL patients will be proposed. The aim of this study is to address clinical trial design, and consideration of outcomes considered important by patients.

An overview of a Delphi-type survey over three stages to obtain consensus on outcomes and eligibility criteria reported in the literature, as well as outcomes brought up during the patient interviews, will be described as follows. The description excludes the data analysis plan. The broader aim of this work is harmonization of outcomes and eligibility criteria to be used in clinical trials of treatments for CL.

Consensus-development methods are generally considered to be appropriate to determine the extent to which experts agree on a certain topic when only insufficient research-based evidence exists. Their aim is to guide group opinion towards agreement, and they are increasingly used in clinical guideline development (Murphy et al., 1998). In general, three main methodological approaches for achieving consensus have been used in the health field: The Delphi method, the nominal group technique or expert panel, and the consensus development conference (J. Jones & Hunter, 1995; Murphy et al., 1998). The Delphi method elicits opinions from a panel of participants with relevant expertise on a certain topic. Thus, Delphi participants are termed 'experts' (defined as a panel of 'informed individuals' (McKenna, 1994)) or 'panellists'. Using sequential paper or online questionnaires, Delphi aims at achieving consensus among a group of panellists: They provide controlled feedback in each round, based on results from previous rounds. After several iterations, there is typically a tendency to converge towards consensus (Hasson, Keeney, & McKenna, 2000; Sumsion, 1998). Whereas the classic Delphi technique, developed by the RAND institute in the 1950s, had 4 rounds, more recent Delphi surveys seem to prefer two or three rounds (Hasson et al., 2000). Delphi was chosen because, in contrast to the nominal group technique and the expert panel, participants do not need to meet face to face. Therefore, it is rather inexpensive, and the group members can remain anonymous, thus allowing freedom of expression and avoiding

dominance effects of certain persons over others often experienced in face-to-face meetings (Prinsen et al., 2014). Three rounds of feedback allow participants to modify their opinion in reaction to the group's results. This Delphi will be conducted *via* an online questionnaire, as the aim was for a large number of participants in different countries to be involved, representing a wide range of institutions and organizations. Computer literacy, and access to computing facilities in order to complete an online questionnaire may be a limiting factor (discuss afterwards), particularly in rather remote areas in LMICs without suitable internet connection. Previous experience with online Delphi systems was positive with regards to accessibility in LMIC settings (E. Allen, personal communication).

8.9.2. Study design

The Delphi method allows for integration of qualitative and quantitative data. The proposed study follows a mixed methods design incorporating outcomes identified during the patient interviews as qualitative components (Creswell, 2011): The questionnaire for the first Delphi round aims at asking participants to score variables identified from the literature and patient interviews, as well as gathering additional input from panellists. Any items proposed in addition will then be added (grouped to concepts, if necessary) to the questionnaires for the second round. The purpose of the third round is to obtain consensus on the relevant items. By asking panellists to score outcomes brought up by patients and additional items brought up during the first round with regards to their importance from a clinical perspective, we are thus aiming at transforming qualitative data.. This method of data transformation was chosen to put the patients' experience from a rich real life context back in a clinical context, as the overall aim of this study was to inform the evidence base in a specific field. However, as noted above, this will not compromise the intrinsic validity of the patient-derived outcomes.

A review of the relevant literature is typically conducted at the beginning of a Delphi (Figure 20). In this case, the study will be mainly informed by two Cochrane Collaboration systematic reviews (González et al., 2009, 2008), subsequent literature reviews (Monge-Maillo & López-Vélez, 2013;

Olliaro et al., 2018; Reveiz et al., 2013) and the outcomes identified through the patient interviews in this study. Published guidance papers (Olliaro et al., 2018, 2013) provide a consensus statement of a group of experts, and will guide questionnaire design.

8.9.3. Sampling

A probabilistic sampling approach does not seem appropriate, and a downside with Delphi, in general, is that it is difficult to verify the representativeness of the selection of panellists. We are aiming for the study population to reflect users of the resulting recommendations (Prinsen et al., 2014). In order to enhance the credibility and acceptance of the guideline, panellists were selected from a diverse range of institutions as well as geographical areas, in line with the *Leishmania* species and patient spectrum that this study intends to cover.

To recruit panellists, a purposive sampling approach was used to select appropriate participants, concomitantly with recruiting collaborators for the patient interviews study as outlined in 3.2.2 Establishment of the collaboration and protocol development. In addition, principal investigators of studies included in and excluded of the Cochrane Collaboration systematic reviews (González et al., 2009, 2008), as well as the PIs of studies registered at www.clinicaltrials.gov, were contacted *via* e-mail and invited to participate. 'Snowball sampling' was used to identify further participants: Invited panellists, and persons responding to the *Call for expression of interest* (WHO/TDR, 2015) were given the opportunity to invite whomever else they felt should be included as relevant.

The panellists will remain anonymous towards each other for the entire duration of the study in order to facilitate expression of opinion free from group domination, existing hierarchies or experience. If they wished so, their identity will be disclosed after study closure, and they will be included *via* group authorship in the resulting publication(s).

The Delphi panellists should have the relevant experience and expertise with CL intervention trials, or treatment of CL patients. Expertise was defined by either experience in CL clinical trials, or in the treatment of CL patients, or authorship of multiple publications in the field. Individuals

replying to the *Call for expression of interest* (WHO/TDR, 2015) were asked to specify their professional role, affiliation and experience with CL.

Panellists were chosen from the following groups:

a) Researchers/Clinical investigators with experience in clinical trials

b) Health workers (nurses or doctors) treating CL patients

No guidelines regarding sample size calculations for Delphi studies were found, in line with existing literature (Murphy et al., 1998; Prinsen et al., 2014; Sumsion, 1998). Akins et al. (2005) have shown that the sample size in a Delphi process does not have statistical relevance if the inclusion criteria are stringent. We assumed that a higher number of panellists will increase the reliability of group judgement as well as the significance and the acceptance of the results in the research community, as a broader spectrum of opinions are represented (Murphy et al., 1998; Prinsen et al., 2014; Sumsion, 1998). Considerations for a limit to the maximum sample size include technical limitations of the online survey system and data analysis (Mandimika et al., 2013).

We have not performed sample size calculations, and we have not set a limit to the maximum number of panellists, as this study would like to be as broad and inclusive as possible and does not expect any major technical limitations. 114 panellists, with a few exceptions currently working, or having worked in the past, in endemic areas, registered their interest in participating, and will receive the first questionnaire. We are aiming for a response rate of 70% for each round, as recommended by (Sumsion, 1998), in order to ensure the rigor of the Delphi process. We therefore expect 50-60 panellists to complete the last round.

Relevant variables were extracted from suitable literature on CL (González et al., 2009, 2008; López-Carvajal et al., 2018; Olliaro et al., 2018, 2013) for questionnaire development for the first round. The different areas of expertise among the 2 groups of panellists were reflected in two different versions of the questionnaire (Figure 21): One for PIs and researchers working with CL clinical trials, the other one for nurses and physicians treating CL patients.

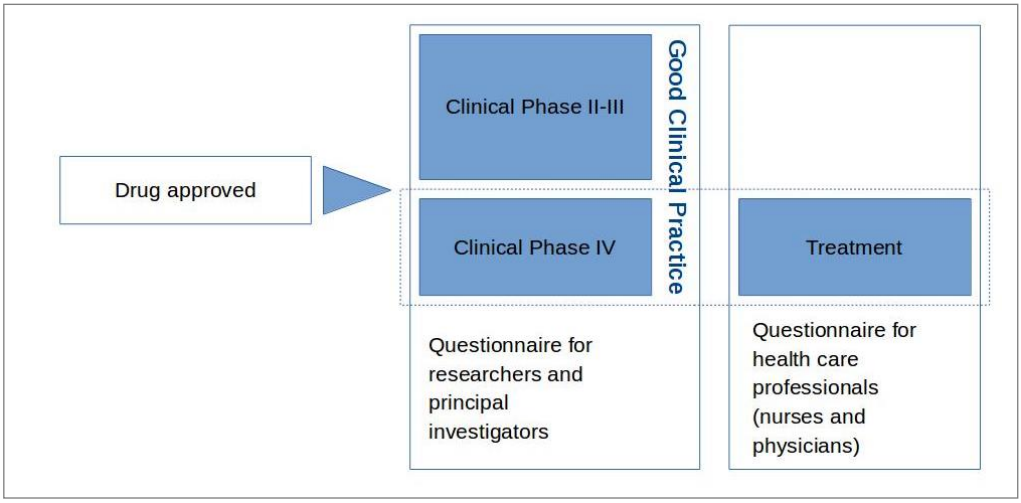


Figure 21: **Schematic overview of assignment of Delphi questionnaire versions** to clinical trial and treatment contexts. Questions related to use of drugs in pre-approval phases of clinical trials (phases II and III) are only part of the questionnaire version for researchers and PIs, whereas questions related to post-approval use (phase IV and treatment context) form part of both questionnaires (dotted line).

The prerequisite for completing the researchers' questionnaire is familiarity with the principles of Good Clinical Practice (GCP). The researchers' version of the questionnaire distinguishes between the different phases of clinical trials in patients (phases II, III and IV). The subset of questions asking for application of an approved drug within a Phase IV (post-marketing surveillance) study is considered as equivalent questions regarding treatment of patients with a licensed drug by HCPs.

The questionnaire (work in progress, not shown) consists mainly of closed questions corresponding to the possible options found in the literature, with additional free text boxes to justify choices and to add further variables that are deemed relevant. Previous experience (E. Allen, personal communication) has shown the importance of careful piloting in order to obtain feedback on the wording, and whether the questions are well understood, unambiguous and complete (Sumsion, 1998). The questionnaire will be piloted with a small number of health care professionals/researchers over three rounds, and was adjusted accordingly after each round. The

questionnaire also contains information on the geographical region and the parasite species, if known.

8.9.4. Data collection

In the first round, the questionnaire will be made accessible to participants *via* a web-based survey, e.g. using LimeSurvey® or equivalent, with low technological requirements. Panellists will be asked to answer questions and rate items, that is, to assign a score corresponding to their importance from a clinical perspective on a scale from 1-10 (Likert Scale), as well as to make suggestions for additional items. Responses from the first round will be aggregated. In a second round, a report containing the scores of individual items from the first round, as well as suggested additional items, will fed back to all panellists, along with their own individual responses from the first round. Panellists will be asked to review, and if they wish, to modify their individual responses from the first round in the light of the group's responses, and to rate the additional items brought up. This is a quantitative step with closed questions only.

For the third round of the Delphi, results of the group response, along with the panellists' individual responses from the previous round, were sent out to the participants. The panellists will be asked to modify their responses, if they wish, with the aim of resolving any remaining disagreement.

For each round, we were aiming for a response rate of 70%, as recommended by (Sumsion, 1998), in order to ensure the rigor of the Delphi process. Consensus will be defined as having no less than 70% of participants selecting the majority option (Mandimika et al., 2013).

8.9.5. Ethical considerations

The Delphi method itself was assessed as minimal risk, and will not need to undergo ethical clearance *via* the Oxford Tropical Medicine Ethics Committee (OxTREC). Confidentiality issues will be addressed by ensuring the anonymity of respondents towards other participants, and

towards the research team. The participants' identity will be included only in the study report, or the publication of the results, if they wish. Consent to participate in the Delphi study will be assumed as soon as panellists completed and submit the questionnaires, after having been informed of the use of the data