DESIGN AND ANALYSIS OF END POINTS IN CLINICAL TRIALS FOR VISCERAL LEISHMANIASIS

by

RAYMOND VICKY OJWANG' OMOLLO

A thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy in Applied Statistics

School of Mathematics, Statistics and Actuarial Science

MASENO UNIVERSITY

ABSTRACT

Visceral Leishmaniasis (VL) is a vector-borne parasitic disease characterized by fever, substantial weight loss, swelling of the spleen and liver, and anaemia. Clinical trials in VL generally take long to conclude due to challenges with patient recruitment and the fact that evidence for definitive cure (DC) can only be seen at least 6 months after treatment completion. Analysis of efficacy at extended follow-up using the triangular test (TT) design is not straight forward in the case of VL trials. In eastern Africa, the combination of Sodium Stibogluconate (SSG) and Paramomycin (PM) is used as a standard treatment for VL while in India, SSG is no longer used due to a poor safety profile particularly cardiotoxicity. The objectives of this study were to develop alternative analyses approaches for DC at extended follow-ups in TT designs, estimate an optimum time point besides 6 months in the analysis of DC and establish the safety of SSG treatment for VL in eastern Africa. Comparisons between the maximum likelihood estimator (MLE), shrinkage estimator (SHE) and probability tree estimator (PTE) in the analysis of DC following the TT design was carried out using bias, root mean square error (RMSE), variances and coverage probabilities of the confidence intervals (CI). An estimate of optimal timing for DC using multi-state models and a review of existing safety data on SSG treatment has been undertaken to assess its association with cardiotoxicity. The results after analysis indicated that Bias, RMSE and variances were low with high coverage probabilities for both the SHE and PTE. The 95% CI of no change in cure status between month 3 and month 6 has a probability of between 98% and 99.9%. Cardiotoxicity was reported in $< 1\%$ of the patients treated with SSG in combination with PM. Both the SHE and PTE are viable alternative approaches in the estimation of efficacy at extended follow-up following TT. There were very few cases in which changes in treatment outcomes occurred between month 3 (M3) and month 6 (M6) follow-up time points. There was not enough evidence to suggest association of SSG use with cardiotoxicity in eastern Africa thus its continued use in combination with PM as a first line treatment for VL is acceptable. Although both the PTE and SHE are good alternatives, analysis approaches for efficacy at extended follow-up to MLE, other analysis approaches of the Bayesian nature need to be explored. Assessment of DC at month 3 gives comparable results to month 6 end point. SSG is still a safe treatment for VL when used in combination with PM but continued monitoring through post-market surveillance is required. This thesis contributes to the improvement in existing knowledge and understanding of the design and analysis methods regarding the conduct of clinical trials in VL.

Chapter 1

INTRODUCTION

1.1 Background to the Study

Visceral Leishmaniasis (VL) is a vector-borne parasitic disease transmitted by the bite of a female phlebotomine sand-fly. The parasite enters macrophages where it multiplies and establishes infection [21]. VL is clinically characterized by fever, hepatosplenomegaly and pancytopenia and is fatal if left untreated [22]. The World Health Organization (WHO) estimates that about 200,000 to 400,000 cases of VL occur every year. In eastern Africa, the annual number of cases is estimated at 30,000 and related deaths at 4,000 [3] with the affected population mostly living in very remote places, where there is limited or no access to health facilities or treatment.

The diagnosis for VL is done by aspiration of the spleen, lymph node or bone marrow with spleen aspirate considered the gold standard. This aspiration requires skilled physicians to be able to perform safely while reading of stained slides also require specially trained laboratory technicians [11] which is hard to achieve in endemic areas and as such not suitable for field use.

In clinical trials for VL, the assessment of final or definitive cure (DC) either as a primary or secondary endpoint is mainly done at 6 months after treatment [22, 43, 34, 36, 17] with some trials assessing it at 12 months post treatment [50]. This is mostly done based on the absence of *Leishmania* parasites in tissue aspirates and or lack of clinical signs and symptoms of the disease. Besides efficacy or cure, assessment of safety is also an important element in the conduct of VL clinical trials. Safety is primarily evaluated based on adverse events reported as well as clinical laboratory assessments during the trial period.

Clinical trials are experiments conducted in humans to compare the effect and value of intervention(s). The intervention could be a therapeutic agent, devise, diagnostic, regimens or procedures. In the case of VL, a number of treatment regimens are evaluated to estimate their success in comparison to existing standard treatments.

An end point in a clinical trial refers to the clinical outcome time point which provides evidence sufficient to fully categorize clinically the effect of a treatment that would support a regulatory claim for that treatment. Thus the choice and timing of the end point whether primary or secondary is crucial to the success of any clinical trial.

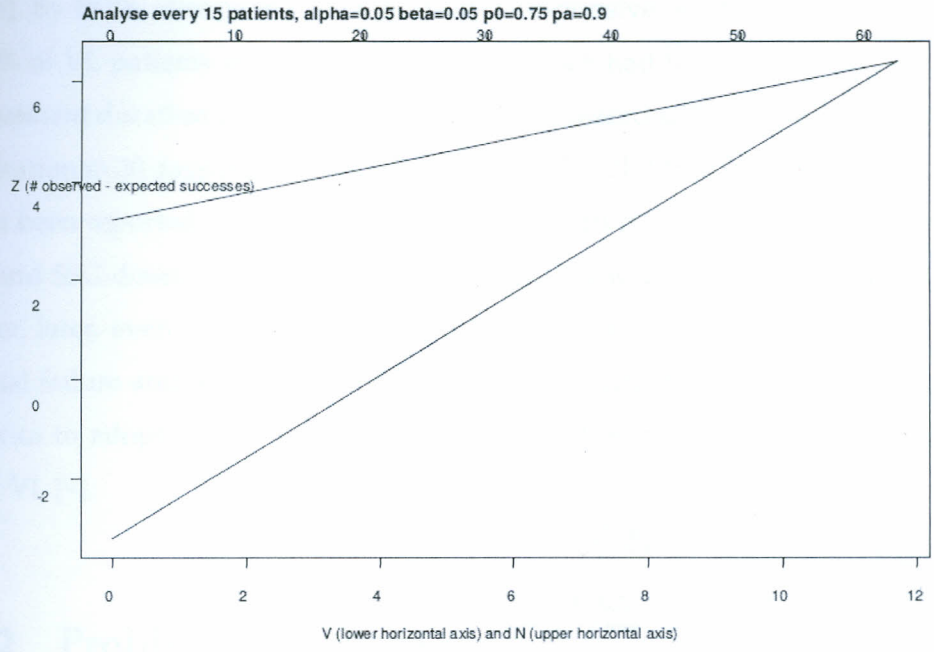
Sequential designs are analyses methods used in clinical trials when one or more interim analyses are performed in a trial. They are different from trials in which sample sizes are fixed in advance with no requirement for interim analysis. This is useful in preventing unnecessary exposure of patients to unsafe new treatments or to existing ones if a new treatment shows significant improvement.

There is great interest in sequential procedures mainly for ethical, economic and administrative reasons [53] as well as the need to satisfy both scientific and statistical constraints. The most compelling reason for monitoring trial data is that ethically, it is desirable to terminate or modify a trial when

evidence has emerged concerning the particular hypothesis of interest. In sequential procedures, the analysis is done as recruitment progresses with the trial being stopped as soon as adequate information is available to reach a decision on stopping the trial for promise (efficacy) or lack of promise (inefficacy). According to International Conference on Harmonization (ICH) E9 statistical principles for clinical trials 4.5: "the goal of an interim analysis is to stop a trial early if superiority is established or a demonstration of relevant treatment difference is unlikely or unacceptable adverse events are present" [1].

Among the sequential designs, is the triangular test (TT) which allows for discrete analyses of data as it accumulates with optimal properties in terms of average sample needed. The TT has a closed continuation region (Figure 1.1.1). Two statistics Z and V are used, where Z is the efficient score and V is Fisher's information for the parameter of interest. Both computed under the null hypothesis with an analysis performed on every group of n patients and Z is plotted against V. The trial is stopped if the sample path crosses a boundary of the continuation region [7]. V is the quantity of information accumulated over the trial period while Z represents the benefit as compared to the null hypothesis. The interim analysis conducted in TT during monitoring only determines when to stop a trial or not but does not provide a complete interpretation of the data specifically when looking at secondary endpoints as is the case in certain diseases like VL when assessing definitive cure [36].

Figure 1.1.1: Triangular Test Design



Source: Omollo et al [36]

The axes of the graph are defined in terms of quantities Z and V [38]:

$$Z = S - Np_0 \tag{1.1}$$

$$V = Np_0(1 - p_0) \tag{1.2}$$

where:

S is the number of patients cured so far in the arm in question

N is the total number of patients included so far in the arm in question

p_0 is the minimum proportion cured which is considered adequate.

In other words, regimens which cure a proportion less than this would not be of interest for further development.

As early as 1940, Sodium Stibogluconate (SSG) had been used as the first line treatment for VL patients in eastern Africa and India. SSG is administered

intravenously (IV) for a duration of 6-10 days at a dosage of 10mg/kg/day [13]. By 1970s, clinicians were reporting unresponsiveness to SSG from over 30% of VL patients in Bihar state in India which had led to the increase of treatment duration and dose. The dose was increased to 20mg/kg/day, and duration to 20 days initially and later to 30 days [52]. Drug resistance to SSG has been reported in India [42], but not in eastern Africa. Studies in Kenya found SSG doses of 10mg/kg/day to be initially well tolerated [12]. However, later, even in Kenya the same dose was likely to be associated with renal failure and so the regime was abandoned, which in turn, led eastern Africa to adopt 20 mg/kg/day of SSG as the standard first-line treatment for VL [9].

1.2 Problem Statement

Sequential methods are gaining prominence in the conduct of clinical trials with the advantage over standard designs being that we can use smaller sample sizes while maintaining the statistical power which is both ethically sound and economical. The decision on whether to stop a trial for inefficacy or not can therefore be made in a timely fashion. The challenge with these designs, particularly the triangular test, lies in how secondary analysis is done given the sequential nature of the design. This is a problem which needs to be addressed.

In VL clinical trials, the assessment of final or definitive cure has mostly been done at 6 months post treatment [22, 43, 34, 31, 46, 48, 47, 45, 59] and in certain instances at much later time points like 12 months [50]. One of the challenges with long follow-up in VL clinical trials has been loss to follow-up which could be associated with the invasive nature of assessing cure. With such long follow-up also comes the challenge of being able to determine at what time point a patient is considered to either have re-infection

or relapse if confirmed as treatment failures. In terms of clinical trial conduct, longer follow-up is known to be inefficient and uneconomical when thinking of the development for new or cheaper treatments for VL [35]. This therefore calls for a review of available clinical trial data to investigate whether the long follow-up time can be shortened to a one single time point before 6 months to asses final cure.

There has not been any formal evaluation of the association of cardiotoxicity with SSG treatment in recent years among VL patients in eastern Africa despite the fact that it is used in combination with Paramomycin (PM) as the first line treatment for VL in the region given that in India, it is no longer in use because of its association with cardiotoxicity in VL patients there [42]. It is therefore critical to find out whether its continued use in eastern Africa may result in similar safety concerns as in India.

1.3 Objectives of the study

The overall objective was to review and determine alternative approaches to the analyses of clinical trial end points for Visceral Leishmaniasis (VL).

The specific objectives of this thesis were to;

1. Develop alternative approaches to the analyses of definitive cure (secondary endpoint) following the triangular test design.
2. Estimate the optimal timing for definitive cure assessment post end of treatment in VL patients.
3. Establish if an association exists between SSG treatment and occurrence of cardiotoxicity among VL patients in eastern Africa.

1.4 Significance of the Study:

This study which primarily relied on data collected from VL clinical trials in eastern Africa has allowed us to

1. Have a better understanding of the methodological issues in group sequential designs particularly the triangular test in phase II trials, specifically the analysis of efficacy at extended follow-up.
2. Estimate the optimal timing for final cure assessment post end of treatment in VL patients.
3. Establish that SSG in combination with PM is a safe treatment for VL when one considers the occurrence of cardiotoxicity among VL patients in eastern Africa.

This thesis contributes to the improvement in existing knowledge and understanding of the design and analyses methods regarding the conduct of clinical trials for VL.

1.5 Basic Concepts

1.5.1 Follow-up:

This refers to the time point(s) when additional patient assessments are done after the end of treatment (EOT). This normally happens at month 1, month 3 or month 6.

1.5.2 Triangular Test

A trial design allowing for repeated interim analyses, each on a relatively small number of patients, in order to efficiently decide between poorly performing treatments and those showing promise for further investigations.

1.5.3 Definitive Cure

Also known as final cure is the last confirmatory assessment of the effect of treatment. It is mostly done six months after the end of treatment.

1.5.4 Efficacy

The proportion of patients cured by a particular treatment regimen.

1.5.5 Safety

Data showing the tolerability of a given treatment by the patients.

1.5.6 Relapse

Occurs when a patient is classified as cured at an earlier end point (e.g day 28) but then becomes a failure at a later end point (e.g day 210).

1.5.7 Slow response

When a patient clinically improves by the end of treatment (day 28) but has parasites still present which eventually clears at a later end point (e.g day 210) without being given any rescue treatment.

1.6 Outline of the Study

In Chapter 1, an introduction to the study has been done with sections on background to the study, problem statement, objectives of the study, significance of the study and basic concepts. In Chapter 2, a detailed review of existing literature has been undertaken and the methodology surrounding the three objectives covered in Chapter 3. In Chapter 4, the results of the analyses addressing the three objectives have been reported and accompanying discussions on the same. Finally, a summary of the thesis together with recommendations from the study have been highlighted in Chapter 5.

Chapter 2

LITERATURE REVIEW

In this chapter, a detailed review of literature has been undertaken with a focus on three major areas, i.e., the triangular test design (Section 2.1), timing of final cure for VL (Section 2.2) and safety of VL treatments (Section 2.3).

2.1 The triangular test

The triangular test is based on a straight line stopping boundaries approach using a closed continuation region [7]. It is mainly used in non-comparative phase II clinical trials where there is need to determine whether a new treatment is sufficiently effective to deserve further evaluation in phase III or not [36, 7, 49, 37, 8]. It is also possible to use the triangular test in situations where no treatment options are available for new conditions like in the case of HIV-VL co-infection where complete parasite clearance in patients is unattainable.

The primary endpoint in these trials is the proportion of patients responding to treatment with the trial aim being to determine whether the success rate, p is greater than a pre-specified value, p_0, chosen as the largest success rate for which further evaluation in phase III is not worthwhile [7]. In statistical

terms, the two competing hypotheses are $H_0 : p \leq p_0$ verses $H_1 : p > p_0$. Two statistics, Z for the efficient score and V for Fisher's information are used to get the parameter of interest under the null hypothesis, H_0. The triangular test uses a sequential plan defined by these two perpendicular axes and allows an investigator to choose between two statistical hypotheses.

Both the null and alternative hypothesis are expressed in terms of the log odds-ratio statistic $\log\left(p(1-p_0)/p_0(1-p)\right)$ [7]. On average fewer observations are necessary to come to a decision with the same type I and II errors as compared to the standard designs with fixed sample sizes. The sample size is not fixed beforehand but is stochastic. An analysis is performed on every group of n patients and Z plotted against V (Figure 1.1.1) with the trial stopped if the sample path crosses a boundary of the continuation region (information increases with the number of analysis).

Two sample statistics calculated in the Triangular Test are;

$$Z = S - Np_0. \tag{2.1}$$

The coefficient score corresponding to the difference between the number of observed responses (successes), S and the number of expected responses Np_0 for a response rate $p_0 = \theta$ with N being the total number of patients included in the treatment arm. It defines the benefit with treatment under study as compared to the response rate p_0. A positive value corresponds to an improvement while a negative value to lack of improvement.

And

$$V = Np_0(1-p_0). \tag{2.2}$$

The Fisher's information which defines the quantity of information accumulated so far in the trial. Thus Z can be seen as the difference between the observed and number of expected responses under H_0, and V as the variance of Z under H_0 [7].

The boundaries of the triangular region are the lines with the following

equations [38].

$$Z = a + \lambda V \tag{2.3}$$

$$Z = -a + 3\lambda V \tag{2.4}$$

where

$$a = a' - 0.583\sqrt{I} \tag{2.5}$$

$$a' = \frac{2}{\theta_a} \log_e \left(\frac{1}{2\alpha} \right) \tag{2.6}$$

$$\lambda = \frac{1}{4}\theta_a \tag{2.7}$$

$$\theta_a = \log_e \left(\frac{p_a(1-p_0)}{p_0(1-p_a)} \right) \tag{2.8}$$

$$I = np_0(1-p_0) \tag{2.9}$$

n is the number of additional patients between analyses, e.g. 15

p_a is a value of proportion cured high enough to recommend the regimen for further investigation.

The term $0.583\sqrt{I}$ is a correction for discrete inspection of data, I being the increment in V between two consecutive inspections. When the type II error rate β differs from α, a corrected value of θ_1 given by the approximate formula

$$\theta_1' = \frac{\theta_1(2\phi^{-1}(1-\alpha)}{\phi^{-1}(1-\alpha) + \phi^{-1}(1-\beta)} \tag{2.10}$$

can be used based on the assumption that Z follows a normal distribution with mean θV and variance V [7]. For example, for $\alpha = 0.05, \phi^{-1}(1-\alpha) = 1.645$.

Crossing the upper boundary of the region causes the null hypothesis to be rejected. It corresponds to stopping for promise while crossing the lower boundary causes the null hypothesis not to be rejected and corresponds to stopping for lack of promise.

For triangular test, we consider two assumptions about post stopping inspection schedule. One is the 'equal inspections' assumption, that the future sequence would be spaced equally in terms of accumulating information with increments equal to those observed so far. The second is 'one last inspection', that there is only one further inspection, delayed until the latest time consistent with the power requirements of the trial [54].

The analyses of secondary endpoints following a triangular test design is not straight forward as the conventional testing procedures ignores the sequential nature and as such would substantially inflate Type I error and decrease power [27]. There is no consensus on the correct approach to analyze the secondary endpoints, especially when the primary endpoint does not show statistical significance. Additionally, methods for the hypothesis testing of secondary endpoints are scarce [27]. One proposition is to look at the secondary endpoint in terms of events over two consecutive time points, T_1 and T_2 as there is a high possibility of change in patient status between them. To ensure consistency of the estimate of θ at secondary endpoint with primary endpoint, it can be considered in terms of events over two consecutive time points i.e., two probabilities, $p_{T_2} = p_{T_1}r + (1 - p_{T_1})s$ [36].

In VL treatment trials, there are two important efficacy endpoints; initial cure and definitive cure. After completion of treatment, initial cure status is established before the patient is discharged. In a clinical trial with parasitological assessment conducted as a standard at the end of treatment, a patient may have cleared parasites (initial treatment success), have parasites remaining but have improved clinically so that additional treatment is not considered to be required prior to discharge (potential slow responder to treatment), or the patient may have parasites remaining without clinical improvement and so require further treatment prior to discharge (confirmed treatment failure). Treatment failure is confirmed when a patient requires additional rescue treatment. Definitive cure assessment is required after a period of time to confirm treatment success due to the possibility of:

(i) slow response to treatment and

(ii) relapse following an initial treatment success.

Definitive cure is generally assessed six months post end of treatment as a standard. Slow response to treatment is confirmed at six months follow-up with slow response being a subset of treatment success at definitive cure endpoint. Future research direction decisions are actually based on definitive cure.

Under standard practice, efficacy in VL clinical trials is assessed as the proportion of patients responding to treatment to the number recruited. In determining the analyses approaches for the final cure, alternative techniques of efficacy analyses given the nature of the TT design for the primary efficacy endpoint analysis needs to be considered.

In a trial for a new treatment approach for VL in India [49], consideration of the analysis of the outcome at extended follow-up given the sequential nature of the triangular test design was not given and instead only the MLE was used, even though it is known to be biased in such designs [57].

In a phase II randomized trial of VL treatments in Sudan and Kenya, a triangular continuation region was defined for each of the three arms [36]. In this trial, the endpoint used for interim decision making was initial cure using a binary outcome of treatment success or failure based on absence or presence of parasites. This outcome was considered the "best-case" scenario in case of relapses between initial and definitive cure time points so that recruitment of the trial would not continue for poorly performing treatments, that is, initial cure provides a conservative estimate of definitive cure in terms of trial continuation decision making. Initial cure provided a timely evaluation of treatments under study in order to identify poorly performing treatments as early as possible. Use of definitive cure in interim decision-making was deemed not to be feasible due to implications for time to trial

completion and the potential to expose an unacceptable number of patients to ineffective treatments. The day 210 efficacy was estimated by considering it in terms of events over two consecutive time periods: up to day 28 (EOT), and from day 28 to day 210. More specifically, the day 210 efficacy (i.e. percentage cured or p_{120}) is considered as the sum of two probabilities, as follows:

(i) Probability of being cured at day 28 and remaining cured at day 210.

(ii) Probability of not being cured at day 28 but becoming cured at day 210. Each of these probabilities is the product of two terms and enables a point estimate of p_{210} to be determined:

$$p_{210} = p_{28}r + (1 - p_{28})s \tag{2.11}$$

Where:

r is the percentage of those people cured at day 28 who remain cured at day 210.

s is the percentage of those people not cured at day 28 who become cured at day 210.

The confidence interval is estimated via the sampling variance of p_{210}. Using the 'delta method' [5], this is estimated as a function of the sampling variances of p_{28}, r and s; these three quantities being statistically independent. The sampling variance of p_{28} is estimated as the square of its standard error, which in turn is estimated as the width of its 95% confidence interval divided by 2×1.96, using a normal approximation.

The sampling variances of r and s are estimated by considering them as standard binomial proportions.

The above procedure ensures that, for example, if no patient changes cure status between day 28 and day 210 then the point estimates of p_{28} and p_{210}

will be equal (because $r = 1$ and $s = 0$). No change in cure status (success or failure) means that a patient classified as a success at day 28 remains a success at day 210 while the patient classified as a failure remains a failure at day 210. Remember r is the proportion of successes at day 28 who remain successes at day 210 while s is the proportion of failures at day 28 who become successes at day 210. The confidence intervals will, however, generally be unequal. This is because treating the progress between day 28 and day 210 as a separate variable implies additional sampling variation. For example, r has sampling variation even when equal to 1. Moreover, sampling variation is considered on the probability scale, not the log-odds scale as for p_{28}. This also means that the calculated confidence interval could exceed the interval 0-1. In this case, it is truncated at the limits of that interval.

2.2 Timing of final cure for VL

The management of current neglected tropical diseases including VL remain inefficiently and uneconomically long mainly due to inefficient tests of cure and long follow-up times [35]. If trial times were shortened then it would be possible to reach decisions faster in making treatments available to the patients as timely evaluation of VL treatment would be a reality.

In VL treatment and management, the timing for definitive cure is 6 months post treatment which has always been a critical time point in determining definitive cure in most clinical trials [22, 34, 36, 17, 31]. By the very nature of the disease, cure is assessed by parasitological examination through aspiration of body sites (spleen, lymph or bone marrow) which is invasive and painful. There have been variations in how definitive cure is assessed over time with most investigators relying more on clinical judgment rather than aspiration to asses cure at this follow-up time point although the gold standard for VL diagnosis still remains parasitology assessment [59].

One of the challenges with long follow-up in VL clinical trials has been loss to follow-up which could be associated with the invasive nature of assessing cure. With such long follow-up also comes the challenge of being able to determine at what time point a patient is considered to either have re-infection or relapse if confirmed as treatment failures. In the LEAP 0104 trial [33] set up to compare the safety and efficacy of three different treatments for VL in eastern Africa, in a total of 830 patients who were considered treatment successes, that is parasitology negative by end of treatment (EOT), only 1 (0.1%) was classified as a failure by the 6 months follow-up time point while among the 26 who had positive parasitology by EOT (failures), 25 (96%) were considered treatment successes by 3 months (unpublished-data) suggesting the possibility of being able to determine final cure before 6 months.

In terms of clinical trial conduct, longer follow-up could be inefficient and uneconomical when thinking of the development for new or cheaper treatments for VL [35]. This therefore calls for a review of available clinical trial data to investigate whether the long follow-up time is justified or can benefit from a detailed study with several time points for purposes of determining one single time point before 6 months to assess final cure.

In the two trials conducted by the Leishmaniasis East Africa Platform (LEAP) [22, 33], a total of 24 patients out of 270 (9%) were lost to follow-up by the 6 month follow-up time point in the first case while lost to follow-up was 5% in the second (50 patients out of 972 enrolled). We have therefore sought to establish if significant treatment outcome differences exist between two follow-up time points (month 3 (M3) 3 & M6) and determine how much M3 outcomes are predictive of the M6 outcomes and if there is need to review when definitive cure should be assessed.

Analysis of VL outcome data can be looked at as an event history analysis on VL. In event history analysis, data is obtained by observing individuals over time, focusing on events occurring for the individuals. Thus a typi-

cal outcome data consists of times of event occurrence and types. As such multi-state models provides a relevant framework for event history data [4]. With a multi-state model, the transition states will be from disease to cure, failure or relapse.

2.3 Safety of VL treatments

Sodium Stibogluconate (SSG) monotherapy was a first line treatment for VL in eastern Africa up until 2010 when the WHO recommended its use in combination with Paromomycin (PM) Sulphate for VL following a large phase III study conducted in eastern Africa by the Leishmaniasis East Africa Platform (LEAP) and Drugs for Neglected Diseases initiative (DNDi) [16, 33]. Globally, antimony based treatments, such as SSG, have been the first line treatment for VL despite considerable toxicity [44, 55] and the requirement for 4 weeks hospitalization. As early as 1985, SSG has been associated with cardiotoxicity and sudden death in a minority of patients whilst on treatment [13, 24, 28].

With the recent emergence of resistance to antimony based agents occurring in the Indian subcontinent (Bihar state), their treatment efficacy has dropped to below 60%. However, similar results have not yet emerged in Africa [42, 30, 14] and so the agents still remain effective (above 90% cure rate) in Africa, and Latin America [55].

A number of studies, particularly in Asia, have associated SSG with cardiotoxicity in VL patients [13, 24, 28, 39]. In Nepal, a serious outbreak of cardiotoxicity occurred in 23 patients treated with a generic batch of SSG. Eight patients died (36%), of which 5 deaths were attributed to cardiotoxicity caused by the drug [39]. Between 1994 and 1996, a study on the efficacy and safety profile of 20mg/kg/day of SSG given to VL patients in Bihar for 30 days, as recommended by WHO at the time, owing to unresponsiveness

to the previous shorter duration of 6-10 days was conducted. During this time, more VL patients were seen to suffer from cardiotoxicity caused by antimony based treatments [52]. A total of 4 deaths (5%) occurred due to cardiotoxicity out of 80 patients enrolled in the study. Other cardio anomalies seen were; diminution in height of T-wave by $>$ 1mm (32 patients), diminution in heights of P, R and T-waves (2 patients), sharp inversion of T-wave (7 patients), elevation of ST segment (3 patients), depression of ST segment (3 patients), prolonged QT interval by $> 0.44s$ (6 patients) and cardio arrhythmia (5 patients) [52]. In another trial of 60 patients comparing amphotericin B with SSG, nine patients had cardiotoxicity (15%), 2 of which were fatal [51]. Similar studies have also noted strong evidence to other side effects, such as hepatic and renal disorders [23, 6].

In eastern Africa, a study with 48 patients reported ECG changes in more than 50% of patients treated with 20mg/kg/day of SSG for 20-30 days [13]. However, none were considered associated with the drug. In Sudan, early studies discovered no toxic side-effects from SSG treatment in a large number of VL cases, including six patients with cardio dysfunction [15]. In 2001, a study conducted in Ethiopia concluded that SSG in doses of 20 mg/kg/day for at least 30 days was safe with a rare possibility of Bradycardia [9]. Since then, there has not been any formal evaluation of SSG cardiotoxicity in recent years among VL patients in eastern Africa despite the fact that it is now used in combination with PM as the first line treatment for VL in the region. To justify the continued use of SSG in the treatment and management of patients suffering from VL in eastern Africa, there is need to review recent safety data for VL to identify if there are safety concerns particularly with regards to cardiotoxicity.

Chapter 3

METHODOLOGY

This chapter begins by showing the process of estimating cumulative efficacy at the primary end point, day 28 following the TT design in Section 3.1 and then discusses the methodological approaches used in addressing the three study objectives (Sections 3.2, 3.3 and 3.4).

3.1 Estimation of cumulative, end of trial efficacy for primary end point (day 28)

Crossing the upper boundary of the triangular region implies that the proportion cured is greater than p_0 (H_0 rejected). Crossing the lower boundary implies that H_0 (proportion cured $\leq p_0$) is not rejected and there is power $1-\beta$ to exclude a proportion cured of p_a. Suppose that the null hypothesis concerning the value of a true parameter θ is tested using a sequential procedure. After termination it will be possible to state whether the null hypothesis has been accepted or rejected at some pre-specified level of significance. The maximum likelihood estimate ϑ of θ can be calculated from the data collected, its value will not be altered by the fact that the data was collected sequentially. However the distribution of ϑ will be affected quite

seriously in some cases and it is therefore unwise to use the conventional analysis on data collected from a probability ratio test or triangular test [57].

Therefore the maximum likelihood estimate of the efficacy, i.e., is the number cured divided by the number of patients is biased due to the sequential nature of the trial [57]. To take this into account, the analysis of cumulative efficacy at the primary end point of day 28 follows Bellisant et al [8].

3.2 Estimation of the efficacy outcome at extended follow-up (day 210)

The sequential analysis described above (Section 3.1) depends only on the day 28 efficacy and not day 210 efficacy. However, because the latter (day 210 efficacy) is likely to be highly correlated with the former (day 28), failure to take account of the sequential design in the day 210 analysis could give inconsistent results. For example, if the status of all patients remain the same at day 210 and day 28, then using different methods for analysis would give different efficacy estimates from the same data. This is because the day 28 estimate takes into account the sequential design which, in general would differ from conventional analysis done on the same data at day 210.

For the efficacy outcome estimation for the definitive cure endpoint at day 210, consider a maximum likelihood estimator (MLE), a probability tree estimator (PTE) and a shrinkage estimator (SHE). The MLE is known to be biased in sequential trials, generally overestimating the treatment effect [57], and was used for comparative purposes only.

3.2.1 Maximum likelihood estimator (MLE)

The maximum likelihood estimator (MLE) for the proportion with definitive cure at day 210 is given by

$$\widehat{\pi}_j = \frac{S_j}{n_j}, \quad j = 1,2,3 \tag{3.1}$$

where S_j is the number of successes at extended follow-up and n_j number of patients that received treatment j. A 95% confidence interval for $\widehat{\pi}_j$ is given by

$$\pi_j \pm z_{\alpha/2}\sqrt{\frac{\pi_j(1-\pi_j)}{n_j}}, \quad j = 1,2,3$$

where $z_{1-\alpha/2} = 1.96$ is the $1 - \frac{\alpha}{2}$ percentile of the standard normal distribution.

3.2.2 Probability tree estimator (PTE)

Estimation of the success probability in the initial period, p_{28}, is subject to sequential stopping, while the subsequent follow-up is not. Hence, the motivating trial example proposed estimating the proportion with definitive cure denoted by p_{210} by using a probability tree argument to separate the two periods [36]. More specifically, p_{210} is the sum of the probabilities of two events:

1. Initial cure followed by cure at extended follow-up, and
2. Initial failure followed by cure at extended follow-up.

These probabilities are denoted pr and $(1-p)s$ respectively, i.e., r denotes the conditional probability of cure at day 210 given cure at day 28, and s denotes the conditional probability of cure at day 210 given no cure at day 28.

$$p_{210} = \hat{p}\hat{r} + (1-\hat{p})\hat{s} \tag{3.2}$$

where $\hat{p}$ is the median unbiased estimate of p [58], with $\hat{r}$ and $\hat{s}$ being estimated by maximum likelihood from the 2×2 table of cure status at the two time points (see Table 3.1). The first term, $\hat{p}\hat{r}$ takes into account those patients that relapse, an important case in VL trials. Here, $\hat{r}$ is the proportion of patients cured after treatment (at day 28 in [36]) that do not relapse. The second term, $(1 - \hat{p})\hat{s}$, accounts for slow responders, those that are not cured initially (at day 28) but become cured by follow-up (day 210). Here, $\hat{s}$ is the proportion of slow responders out of those patients that are not cured by day 28. The sampling variance of this estimator is derived as follows;

Starting from

$$p_{210} = p_{28}r + (1 - p_{28})s$$

re-arrange to get

$$p_{210} = s + p_{28}(r - s)$$

Now for the variance we have

$$var(p_{210}) = var(s) + var(p_{28}(r - s)) + 2\,cov(s, p_{28}(r - s)) \tag{3.3}$$

Call the three summands A', B' and C'.

Term A': $var(s) = \sigma_s^2$.

Term B': Again use formula for variance of a product, and independence of r and s to get $var(r - s) = var(r) + var(s) = \sigma_r^2 + \sigma_s^2$.

$$var(p_{28}(r - s)) = p_{28}^2(\sigma_r^2 + \sigma_s^2) + (r - s)^2\sigma_{28}^2 + \sigma_{28}^2(\sigma_r^2 + \sigma_s^2). \tag{3.4}$$

Term C':

$$\begin{aligned} cov(s, p_{28}(r - s)) &= E(sp_{28}(r - s)) - E(s)E(p_{28}(r - s)) \\ &= E(sp_{28}r - sp_{28}s) - E(s)E(p_{28})(E(r) - E(s)) \end{aligned} \tag{3.5}$$

Cancel terms involving + and − product of the three expectations, and replace expectation by observed value of p_{28}:

$$\begin{aligned} &= -p_{28}E(s^2) + p_{28}(E(s))^2 \\ &= -p_{28}\sigma_s^2. \end{aligned}$$

Putting the terms together (again replacing the factor 2 for C'):

$$var(p_{210}) = \sigma_s^2 + p_{28}^2(\sigma_r^2 + \sigma_s^2) + (r-s)^2\sigma_{28}^2 + \sigma_{28}^2(\sigma_r^2 + \sigma_s^2) - 2p_{28}\sigma_s^2.$$

Again gather terms by σ_r^2 and σ_s^2:

$$\begin{aligned} &= \sigma_r^2(p_{28}^2 + \sigma_{28}^2) + \sigma_s^2(1 - 2p_{28} + p_{28}^2 + \sigma_{28}^2) + \sigma_{28}^2(r-s)^2 \\ &= \sigma_r^2(p_{28}^2 + \sigma_{28}^2) + \sigma_s^2((1-p_{28})^2 + \sigma_{28}^2) + \sigma_{28}^2(r-s)^2 \\ Var[p_{210}] &= \sigma_r^2(p^2 + \sigma^2) + \sigma_s^2((1-p^2)^2 + \sigma^2) + \sigma^2(\hat{r} - \hat{s})^2. \end{aligned} \tag{3.6}$$

A 95% confidence interval can then be calculated using

$$p_{210} \pm z_{1-\alpha/2}\sqrt{var[p_{210}]}$$

where $z_{1-\alpha/2} = 1.96$ is the $1 - \alpha/2$ percentile of the standard normal distribution.

Table 3.1: Treatment Outcome by Time point

	Time Point	
	Day 28	Day 210
Success	A	B
Failure	C	D

3.2.3 Shrinkage estimator (SHE)

Shrinkage estimation, which is implicit in Bayesian inference, aims to reduce mean square error and selection bias. In hierarchical modeling the estimate of treatment effect in one particular group "borrows" information about the treatment effect in all other groups, though the groups need not be related [10]. Berry [10] states that "borrowing measurements between

entities that bear no relationship is better than letting them stand alone." We consider a Bayesian probit model of the form,

$$\begin{aligned} P(Y_{i,j} = 1|\theta_j) &= \Phi(\theta_j), \quad i = 1, \ldots, n_j; j = 1, 2, 3 \\ \theta_j|\mu, \tau^2 &\sim N(\mu, \tau^2), j = 1, 2, 3 \\ \mu &\propto 1 \\ \tau^2 &\sim IG(\alpha, \beta), \ \alpha = 2.1, \beta = 0.33 \end{aligned} \tag{3.7}$$

Where $Y_{i,j}$ is the response of the i^{th} patient on treatment j, Φ is the cumulative distribution function (cdf) of the standard normal distribution, and IG denotes the inverse-gamma distribution.

The treatment effects are assumed to be drawn from a normal distribution with mean μ and variance τ^2. Our prior distribution for μ is non-informative. We cannot, however choose a non-informative prior for τ, as there are only three treatment groups, and this causes convergence issues for the Markov chain Monte Carlo (MCMC) algorithm required to fit the model. The choice of prior for τ is known to be difficult when the number of group sizes is small [19]. An inverse-gamma distribution for τ^2 is a convenient choice in terms of facilitating a relatively simple MCMC algorithm. We then decide to use the prior knowledge that a range of treatment effect sizes ($\theta_j s$) of $\Phi^{-1}(0.9) - \Phi^{-1}(0.75) \approx 0.6$ was very plausible, but a range of effect sizes twice this large was fairly unlikely. For a normal distribution with variance 0.3 , 42% of the probability density lies within 0.3 of the mean, and 73% of the density lies within 0.6 of the mean. It therefore seemed a reasonable and pragmatic choice to give τ^2 a prior mean of 0.3. This implies that $\beta = 0.3(\alpha - 1)$. The larger the choice of α, the smaller the variance for τ^2, since

$$Var(\tau^2) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \text{ for } \alpha > 2.$$

If the variance chosen is too small, the degree of shrinkage or "borrowing" is effectively fixed in advance, i.e., it does not depend on how "spread out" the treatment effect estimates are. On the other hand, if the variance of

τ^2 is too large, the MCMC will run into convergence issues. A value of $\alpha = 2.0$ was found via a process of trial and error, such that the model gave acceptable performance in a range of simulated scenarios. The posterior means of the $\theta_j s$ were used for point estimation and posterior 95% credible intervals were used for the interval estimation (i.e. their performance as confidence intervals was assessed).

3.2.4 Bias and efficiency of the estimation methods

To compare the performance of the SHE, PTE and the MLE, simulation was performed in order to calculate bias, root mean square error (RMSE), length of 95% confidence intervals and coverage probabilities. Four different scenarios for the true proportion of successes were considered at the initial time point:

1. Scenario 1: all treatments unpromising ($p_1 = p_2 = p_3 = 0.75$)
2. Scenario 2: all treatments promising ($p_1 = p_2 = p_3 = 0.90$)
3. Scenario 3: one treatment promising ($p_1 = p_2 = 0.75, p_3 = 0.90$)
4. Scenario 4: linear relationship between efficacy and treatment ($p_1 = 0.75, p_2 = 0.825, p_3 = 0.90$)

And four different cases for the change in patient status between the initial period and follow-up:

1. Case 1: no relapse, no slow responders ($q = 1, s = 0$)
2. Case 2: no relapses, 33% slow responders ($q = 1, s = 0.33$)
3. Case 3: 25% relapses, no slow responders ($q = 0.75, s = 0$)
4. Case 4: 25% relapses, 33% slow responders ($q = 0.75, s = 0.33$)

Where q and s were defined as in the definition of the PTE. For each combination of the above, simulations with 10^4 iterations were performed, using a design with three interim analyses after every 15 patients are recruited.

The bias and RMSE of the three estimators were calculated after selecting the "best performing" treatment of each iteration. The "best performing" treatment refers to the treatment with the largest estimated success probability at follow-up. The bias and RMSE from selecting the best treatment were calculated using the following formulae

$$bias = b_p(Q_s) = E_p(Q_s - p'_s) \tag{3.8}$$

$$\begin{aligned} RMSE_p(Q_s) &= \sqrt{MSE_p(Q_s)} = \sqrt{E_p\left((Q_s - p'_s)^2\right)} \\ &= \sqrt{Var_p(Q_p) + b_p^2(Q_s)} \end{aligned} \tag{3.9}$$

Where $s \in (1, 2, 3)$ is the index of the selected treatment, Q_s is the estimator used and p' is the true value of the efficacy at follow-up.

Additional data from the LEAP 0208 trial [36] has been used in calculating the variances of the efficacy estimates (MLE, PTE & SHE) at extended follow-up (day 210) and comparisons made on which estimator provides the least variance.

3.3 Timing for definitive cure

To achieve this, one can consider multi-state models which are probability models that describe the random movement of a subject between a series of states in continuous or discrete time. Multi-state models are increasingly being used to model the natural history of chronic, viral or infectious diseases like cancer, diabetes and HIV among others in patients as well as to characterize patient follow-up under varied treatment protocols [20] and can thus be extended to VL.

Nowhere has multi-state models been used in VL characterization but with an interest in estimating the transition rates in VL, it can be adopted to look at the different disease states within the treatment and management framework for VL. This could be initial cure, slow response, definitive/final cure or relapse. In VL studies where assessment time points are discrete, interval censored data are apparent and methods of inference in multi-state models could potentially be useful in determining at what time point the subjects change their state which then can inform how relevant the current timing for final cure is.

The transition probability from state S_i to S_j is written as

$$p_{ij}(t) = P\left(X_{t+1} = S_j | X_t = S_i\right) \tag{3.10}$$

i.e., given the present state of the system, $X_n = S_i$, the future of the system is independent of the past. If the Markov property, i.e., the independence of the future from the past given the present holds and the transition probabilities do not depend on t, we say that the process is a (time) homogeneous Markov Chain [26].

3.3.1 Definition for follow-up

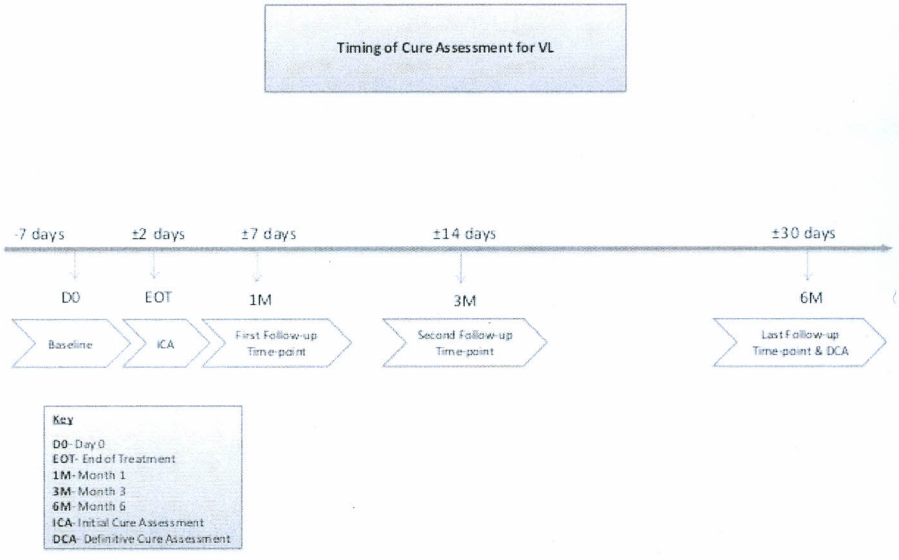
The day of follow-up assessments has been calculated from EOT. Therefore follow-up at M3 is 90 days from EOT while that at M6 is 180 days from EOT. Since it was not possible to have all patients come for follow-up on the exact visit day, a window period around follow-up time point has been allowed for;

1. A visit is considered to be at EOT (end of treatment) if it occurs on EOT ($\pm$2 day)
2. A visit is considered to be at M3 (day 90) if it occurs on day 90 ($\pm$14 days)

3. A visit is considered to be at M6 (day 180) if it occurs on day 180 (±30 days)

For purposes of this analysis, anything falling outside the window period above was classified as missing and has been excluded from the final analysis. The extent of missing data from the above criteria has been produced to help in understanding gaps arising from such trials.

Figure 3.3.1: Timing of Cure Assessment for VL



Source: Researcher

In VL trials, the screening duration could be up to 7 days before administration of the first treatment, at the end of treatment (EOT), the efficacy assessment can be done within ±2 days of treatment completion while follow-up assessments could occur anywhere between ±7 days for month 1 or ±14 days form month 3 to ±30 days for month 6 of the expected follow-up assessment day.

3.3.2 Modeling transitions between states

In modeling the transition between two states, 1 (treatment success) and 2 (treatment failure), a first order autoregressive model (AR(1)) has been used. An AR(1) model for the probability that an individual j is in state 1 at time t, $p_{t,1}$ is

$$\log\left(\frac{p_{t,1}}{1-p_{t,1}}\right) = \alpha + x_{t,1}\beta + \gamma y_{t-1,1} + u_1 \tag{3.11}$$

Where

α is an intercept term

γ is the effect of the state occupied at $t-1$ on the log odds of being in state 1 at t

$u_1 \sim N(0, \delta_u^2)$ is an individual specific random effect.

We model

$$\begin{aligned} p_{t,1} &= Pr(\text{state 1 at start of } t) \\ &= Pr(Y_{t,1} = 1) = Pr(y_{1,2} = 1). \end{aligned}$$

Suppose we fix $x_{t,1} = 0$ and $u_1 = 0$.

The probability of moving from state 1 to 2 is

$$\begin{aligned} Pr(y_{t,1} = 0 | y_{t-1,1} = 1) &= 1 - Pr(y_{t,1} = 1 | y_{t-1,1} = 1) \\ &= 1 - \frac{\exp(\alpha+\gamma)}{1+\exp(\alpha+\gamma)}. \end{aligned} \tag{3.12}$$

While the probability of moving from state 2 to 1 is given by

$$Pr(y_{t,2} = 1 | y_{t-1,2} = 0) = \frac{\exp(\alpha)}{1+\exp(\alpha)}. \tag{3.13}$$

Since same individuals can contribute observed time at risk to more than one state, observations made relate to the number of events and total time at risk rather than the number of individuals [56].

3.3.3 Predictive model

We have come up with a probability value, $p_{1,2}$ (and its 95% confidence interval) defined as the probability of no change in cure status between two time points 1, 2 (i.e EOT and M3 or M3 and M6) to determine how much of M6 outcomes can be predicted by M3 outcomes or M3 by EOT outcomes. No change in cure status between time points implies that if one is a success at EOT, then they remain a success at M3 or if one is a failure at M3 then they remains a failure at M6.

Where

$$p_{1,2} = \begin{cases} -ve(T_{1,2}) \\ +ve(T_{1,2}) \end{cases} . \tag{3.14}$$

A negative ($-ve$) refers to being a treatment success (no VL parasites) at time 1 and remaining a treatment success at time 2 while a positive ($+ve$) refers to being a treatment failure (VL parasites present) at time 1 and remaining a treatment failure at time 2.

In coming up with $p_{1,2}$, we have assumed treatment success (clinical cure) by M3 in instances where parasitology was not done. This is because parasitology was only being done at these time points when clinically indicated.

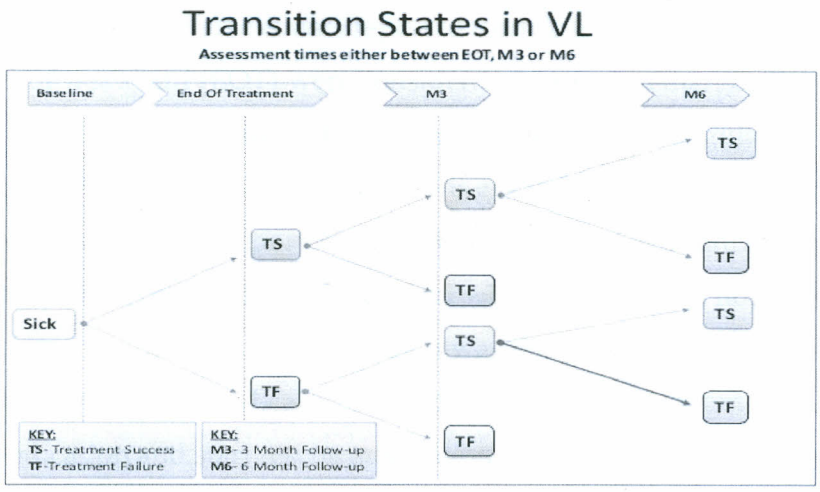


Figure 3.3.2: VL Transition States

Source: Researcher

3.4 Safety of SSG treatment in VL

Cardiotoxicity is based on abnormalities detected through Electrocardiogram (ECG) examinations. Patients enrolled in the study were assessed at baseline for any major physical and electrocardiographic abnormalities. Patient's hematology and biochemistry were also measured weekly during treatment to identify any major deviation from baseline in clinical and laboratory values. ECG and audiometric evaluations were scheduled for baseline, day 14, end of treatment and follow-up at 3 and 6 months. The trial sites in Kenya performed additional ECGs weekly to gather additional data owing to the use of generic SSG. ECGs were performed with the patient resting supine on the bed using a portable self reporting ECG machine (Cardiofax, Model ECG 9620, Nihon Kohden) provided to each site prior to trial start. Trial site investigators monitored ECG examinations and made their judgment on whether or not an ECG was normal or abnormal; and if abnormal, whether the abnormality was clinically significant. Other adverse events were reported from the start of treatment until end of follow-up at

6 months. Those adverse events reported during treatment and 30 days after end of treatment (EOT) were considered treatment emergent. Adverse events were coded according to Medical Dictionary for Regulatory Activities (MedDRa) version 11 [32] and were graded as mild, moderate or severe. Site investigators reported any serious adverse events immediately (or within 24 hours) to a medical coordinator who in turn would notify the data and safety monitoring board (DSMB). It was the site investigators decision, with advice from the Medical Coordinator on whether a patient should stop trial medication and start rescue medication with liposomal amphotericin B (up to 30mg/kg over 10 days, total dose) or not. Adverse drug reactions were also noted according to relation of an adverse event to the treatment given or disease.

3.4.1 Treatments

Paromomycin Sulphate (Gland Pharma, India) is the same formulation as Paromomycin base: the base is the active ingredient whereas PM, the manufactured product (produced by fermentation), is a Sulphate salt. PM was administered intramuscularly (IM) and administered to five clinical trial sites (KEMRI, Kenya; Kassab, Sudan; Gondar and Arba Minch, Ethiopia and Amudat, Uganda) using Batch number FB501X and KT701X with expiry date of January 2008 and July 2009 respectively. At the start of the trial only proprietary branded SSG (Pentostam) from GlaxoSmithKline was licensed for use. However, by mid-2007, generic SSG (manufacturer Albert David, India) was approved and administered IM except in Kenya where it was administered intravenously (IV) with batch number 4P12004, 4P12009, 4P12010, 5P12036, 5P12037, 5P12038, 5P12042, 5P12044, 6P12001, 6P12007, 6P12008, 8P12004, 8P2010. SSG dosage was capped at 850 mg/day for all countries except Sudan. Rescue medication for the trial was liposomal amphotericin B, (manufactured as Ambisome, Gilead, USA) and provided ac-

cording to national dosage guidelines for each country with batch numbers.

3.5 Available data

In addressing the objectives outlined in Section 1.3, permission was granted by the sponsor (Drugs for Neglected Diseases Initiative - DNDi) and the country principal investigators (PIs) of Leishmaniasis East Africa Platform (LEAP) in Kenya, Sudan, Ethiopia and Uganda to use data from the following three clinical trials which have been conducted in accordance with the requirements of International Conference on Harmonization - Good Clinical Practice (ICH-GCP) [18].

1. A three-arm multicentre, open-label, randomized, controlled trial comparing three treatment regimens (sodium Stibogluconate (SSG) verses SSG & Paramomycin (PM) verses PM for VL in East Africa [22, 33]. Also referred to as the LEAP 0104 trial where 0104 refers to January 2004 when the protocol was developed.

2. An open-label, 2 arm, non-inferiority, multicentre randomized controlled trial to determine the optimal single-dose treatment with AmbisomeÂ® [17, 25]. Also referred to as the LEAP 0106 trial where 0106 refers to January 2006 when the protocol was developed.

3. A phase II randomized, parallel arm, open-labeled trial to assess the efficacy of each of the three regimens: liposomal amphotericin B with sodium Stibogluconate (SSG), Liposomal amphotericin B with Miltefosine and Miltefosine alone. The primary endpoint was at day 28 with secondary endpoint at day 210 [36]. Also referred to as the LEAP 0208 trial where 0208 refers to February 2008 when the protocol was developed.

Data analyses has been done using R [40] and STATA [41] software's.

Chapter 4

RESULTS AND DISCUSSIONS

In this chapter, the results as per the study objectives have been presented followed by discussions.

4.1 Efficacy at extended follow-up (day 210)

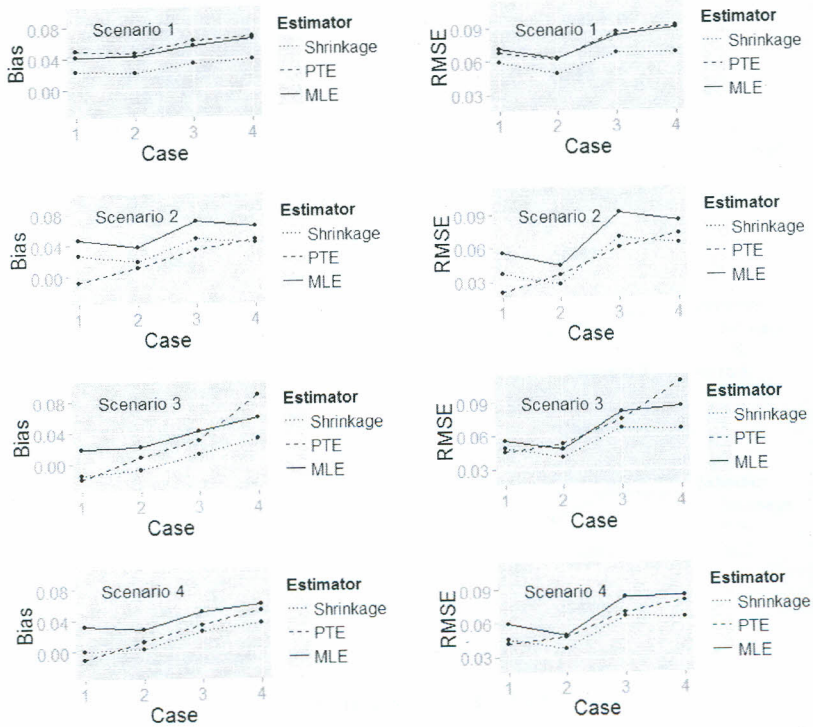
4.1.1 Bias and efficiency of the estimators

A plot of the bias and RMSE under each combination of the true success probability and change in patient status scenarios, with the latter represented on the x-axis is given in Figure 4.1.1. This shows that in the majority of cases, the SHE performs better in terms of reducing both bias and RMSE than PTE and MLE.

We also calculated 95% confidence intervals (CIs) and recorded the length and coverage probabilities. Figure 4.1.2 shows the length of 95% CIs and the coverage probability for the three estimators (x-axis as before). Whilst the length of the 95% CIs are similar for all the three estimators, the coverage probability of the SHE is greater, meaning that the 95% CIs contain the true

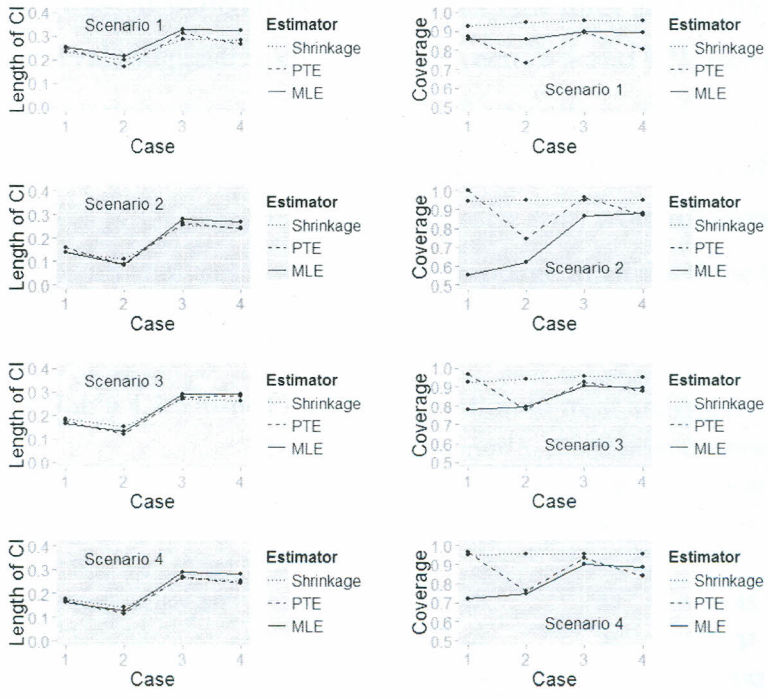
value of interest more often.

Figure 4.1.1: Bias and RMS



Source: researcher

Figure 4.1.2: Length of 95% CIs and coverage probability



Source: researcher

4.1.2 Variances of the estimators

Table 4.1: Efficacy at D28: Intention to treat analysis

	Treatment A	Treatment B	Treatment C
Final number of patients (N)	51	49	51
Final number cured (S)	47	46	45
Maximum likelihood (biased) estimate of proportion cured (S/N)	0.92	0.94	0.88
C^*	11.5	11.1	11.5
θ_a	1.10	1.10	1.10
p value 0.043	0.042	0.043	
unbiased estimate of θ (θ_{28})	0.61	0.62	0.61
(95% confidence interval, θ_L to θ_U)	(-0.09, 1.29)	(-0.09, 1.31)	(-0.09, 1.29)
unbiased estimate of proportion cured	**0.85**	**0.85**	**0.85**
p_{28} **(95% confidence interval)**	**(0.73, 0.92)**	**(0.73, 0.92)**	**(0.73, 0.92)**

Accounting for the sequential trial design and repeated interim analyses, cure at D28 was calculated to be 85% in each of the three arms with 95% confidence intervals suggesting cure to be between 73% and 92% (see appendix V for computations using STATA [41]). The results in Table 4.1, Table 4.2 and Table 4.3 are from the LEAP 0208 trial [36] in which a total of three interim analyses (after every 15* patients in each arm) were undertaken with the third interim analyses leading to trial stoppage in all the three treatment arms.

Table 4.2: Cure at day 210 (Intention to treat analysis)

	Treatment A	Treatment B	Treatment C
Proportion cured (p_{28}, Table 4.1)	0.85	0.85	0.85
Number of patients with non-missing cure status at both days 28 and 210 (N_{28})	51	49	51
Number cured at day 28	47	46	45
of whom still cured at day 210	46	39	37
as a proportion (s)	0.98	0.85	0.82
Number not cured at day 28	4	3	6
of whom became cured at day 210	1	1	1
as a proportion (r)	0.25	0.33	0.17
Proportion cured at day 210 (p_{210})	0.87	0.77	0.72
Standard error of p_{210}	0.052	0.067	0.063
95% confidence interval for p_{210}	0.77 – 0.97	0.64 – 0.90	0.60 – 0.85

Relative to day 28, treatment A retains its efficacy better than the other arms because a higher proportion remained cured. Some patients also become cured at day 210 (having not been cured at day 28), although this factor is less important in determining the overall proportion cured at day 210. See appendix VI for calculation of day 210 efficacies using STATA [41]).

Table 4.3: Variances of the Day 210 Efficacy Estimators

		Point Estimate	95% C.I.	Variance
MLE	Treatment A ($n = 51$)	0.92	0.81 - 0.98	3.686
	Treatment B ($n = 49$)	0.82	0.68 - 0.91	7.347
	Treatment C ($n = 51$)	0.75	0.60 - 0.86	9.686
PTE	Treatment A ($n = 51$)	0.87	0.77 - 0.97	0.003
	Treatment B ($n = 49$)	0.77	0.64 - 0.90	0.004
	Treatment C ($n = 51$)	0.72	0.60 - 0.85	0.004
SHE	Treatment A ($n = 51$)	0.90	0.81 - 0.96	0.002
	Treatment B ($n = 49$)	0.82	0.71 - 0.91	0.002
	Treatment C ($n = 51$)	0.77	0.65 - 0.86	0.003

MLE=Maximum Likelihood Estimate; PTE=Probability Tree Estimate; SHE=Shrinkage Estimate

Discussions

The triangular test is a popular choice of sequential test due to its low expected sample size across a wide range of potential treatment effect sizes, and also for historical reasons, as very good approximations could be used to generate stopping boundaries without a computer.

Stopping boundaries and sample sizes can be found to match specified Type I and Type II error rates. For normally distributed data, the error rates are achieved exactly. For many types of non-normal data [58], they are achieved approximately. The aim was to derive methods for estimation of the outcome variable (e.g. cure) at a time point later than that used for the stopping rule. In a trial for a new treatment approach in Indian VL [49], consideration of the analysis of the outcome at extended follow-up given the sequential design used was not made and instead the MLE was used, which we know to be a biased estimator.

Simulation shows that the SHE based on a Bayesian probit model outperforms the MLE and a PTE in terms of reducing bias and RMSE in most

scenarios (see Figure 4.1.1). For all estimators, bias was smaller when no relapses occurred (case 1 and 2 of Figure 4.1.1). The RMSE was smaller for all estimators when no relapses occurred (cases 1 and 2 of Figure 4.1.1) and for the SHE when there were slow responders (cases 2 and 4 of Figure 4.1.1).

The SHE performed consistently well in terms of the coverage probability of 95% CIs; the PTE performed better when there were no slow responders (cases 1 and 3 of Figure 4.1.2), while the MLE performed better when there were relapses (cases 1 and 3 of Figure 4.1.2). It is well known that for binary data, the MLE CI based on the normal approximation only performs well for large sample sizes n. Agresti states that "the actual coverage probability usually falls below the nominal confidence coefficient, much below when π is near to 0 or 1", where π is the true success proportion [2]. We would not recommend using the MLE for estimation of efficacy at follow-up based on these findings in any of the scenarios considered. The variances for the PTE and SHE were much lower than that from MLE demonstrating that they are good estimators for outcome following TT design (see Table 4.3).

Whilst the SHE performs best, it is difficult to implement in comparison to PTE and MLE. In particular, choosing the prior distribution for the variance random effects is a subtle task. The SHE would also be unsuitable for use in trials of a single treatment arm and so the PTE may provide a suitable alternative.

In the estimation of efficacy at extended follow-up following a triangular test design, the SHE is preferable. The PTE would provide an alternative for use in one arm trials or when the SHE is not possible due to computational complexities.

4.2 Timing for definitive cure

4.2.1 LEAP 0104

Table 4.4: Assessment Days: LEAP 0104

	EOT[a]	Day 90	Day 180
Seen within window period	925	532 (57.6%)	612 (66.2%)
hline Seen outside window period	22	249	317
Parasitology Done[b]	924	532	612
Parasitology not done[b]	1	196[c]	0
Received rescue treatment[b]	61	9	15

[a] End of Treatment (day 18 if SSG & PM; day 22 if PM alone; day 31 if SSG alone)

[b] Among those seen within the window period

[c] assumed to be parasite negative since parasitology not indicated

There were about 58% of the patients coming for day 90 (month 3) visit within the acceptable window period from EOT with about 66% doing the same at day 180 (month 6).

Multi-state Model:

Table 4.5: Person visits by treatment outcome - LEAP 0104

Outcome	**Overall**	**Between**	**Within**
Treatment success	1072 (98.0%)	366 (97.9%)	98.5%
Treatment failure	22 (2.0%)	22 (5.1%)	61.4%
Total	**1094 (100%)**	**388 (103.7)** ($n = 374$)	**96.4%**

Table 4.6: Transition Probabilities - LEAP 0104

Outcome	**Treatment success**	**Treatment failure**	**Total**
Treatment success	706 (99.0%)	7 (1.0)	713 (100)
Treatment failure	7 (100)	0(-)	7 (100)
Total	**713 (99.0)**	**7 (1.0)**	

There were 1072 person visits of data in which the treatment outcome was a success and 22 person visits where it was a failure (2% of our data). A total of 366 patients ever had a treatment success while 22 ever had a treatment

failure with a total of 388 ever having either. In our data set however, we only have 374 patients which means that there were patients who sometimes had successes and failure at other times (see Table 4.5).

Conditional on a patient ever having a success, 98.5% of their observations had a success outcome. Similarly, conditional on ever having a failure, 61.4% of a patient's observation had a failure. Both percentages are a measure of the stability of the two outcomes indicating more stability for the success outcome when compared to failure with the overall stability of outcomes being 96.4% (see Table 4.5).

At each visit, some 99% of treatment success in the data remained treatment successes at the next visit with the other 1% being failures at the next visit. Although treatment successes had a 1% chance of becoming failures at the next visit the treatment failures had 100% chance of becoming successes at the next visit (see Table 4.6).

Standard predictions:

Table 4.7: Treatment Outcomes: EOT to M3 - LEAP 0104

End of Treatment	**Day 90 (3 months) Follow-up**		
Negative: 359 (96%)	Negative	354[a]	98% [96.8-99.5]
	Positive	5[b]	
Positive: 15 (4%)	Positive	0	100% [78.2-100¥]
	Negative	15[c]	
Day 90 (3 months) Follow-up	**Day 180 (6 months) Follow-up**		
Negative: 369 (99%)	Negative	367[d]	99% [98.1-99.9]
	Positive	2[e]	
Positive: 5 (1%)	Positive	1[f]	80% [28.2-99.5]
	Negative	4[g]	

[a] A total of 8 cases received rescue treatment; [b] All the 5 patients had received rescue by M3

[c] Eight patients were given rescue treatment; ¥ One sided 97.5% confidence interval

[d] A total of 14 cases received rescue treatment; [e] All the 2 patients had received rescue by M6;

[f] 1 patient was given rescue treatment; [g] All the 4 patients were given rescue treatment by M6;

For the timing analyses, only patients seen within all the visit window periods have been considered ($n = 374$). Among the 359 patients who had cleared parasites (negative) by EOT, only 5 (1.3%) became positive by the

M3 follow-up assessment. This implies that, there was a 98% probability of a patient with no parasites at EOT remaining parasite free by M3 follow-up while of the 369 patients who had cleared parasites (negative) by M3, only 2 (0.5%) became positive by the M6 follow-up assessment implying that there was a 99% probability of a patient with no parasites at M3 remaining parasite free by M6 follow-up (Table 4.7).

Characteristics of patients who change cure status: End of Treatment to Month 3

Five patients who were negative at EOT became positive by M3 (all were put on rescue treatment at the M3 visit assessment). They all had very high parasite counts at baseline with exception of one case (count of 1×10^6). They all had very large spleens at baseline (>6cm) which either reduced marginally or increased in size by the M3 assessment time point.

Fifteen patients who were positive at EOT but became negative by M3 assessment with only 8 of them receiving rescue treatment (between days 23 and 66) before the M3 assessment. The other seven had significant reduction in parasite counts between baseline and EOT.

Characteristics of patients who change cure status: M3 to M6

Two patients who had no parasites by M3 became positive by M6. They both looked like relapse cases as they were both negative by EOT as well. They were put on rescue treatment at the M6 visit assessment. In both cases the spleen sizes had increased by at least two fold between the M3 and M6 visits. Four patients who were positive by M3 then became negative by M6 (all received rescue treatment). In all the 4 cases, there were marked reductions in spleen sizes.

Table 4.8: Assessment Days - LEAP 0106

	EOT[a]	Day 90	Day 180
Seen within window period	121	60	81
Seen outside window period	0	40	32
Parasitology Done[b]	121	60[c]	81
Parasitology not done[b]	0	0	0
Received rescue treatment[b]	36	1	2

[a] End of Treatment (day 30)

[b] Among those seen within the window period

[c] Assumed to be parasite free if parasitology is not clinically indicated

4.2.2 LEAP 0106

From this data set, a total of 45 patients had complete follow-up data on all assessment time points.

Table 4.9: Person visits by treatment outcome-LEAP 0106

Outcome	Overall	Between	Within
Treatment success	127 (98.5%)	44 (97.8%)	98.9%
Treatment failure	2 (1.5%)	2 (4.4%)	75.0%
Total	**129 (100%)**	**46 (102.2)** [$n = 45$]	**97.8%**

Table 4.10: Transition Probabilities - LEAP 0106

Outcome	Treatment success	Treatment failure	Total
Treatment success	83 (98.8%)	1 (1.2)	84 (100)
Treatment failure	-	-	-
Total	**83 (98.8)**	**1 (1.2)**	

Multi-state Model

There were 127 person visits of data in which the treatment outcome was a success and 2 person visits where it was a failure (1.5% of our data). Forty four patients ever had a treatment success while 2 ever had a treatment failure with a total of 46 ever having either. In our data set however, we only have 45 patients which means that there was one who sometimes had a success and a failure at other time (see Table 4.5).

Table 4.11: Treatment Outcomes: EOT to M3 - LEAP 0106

End of Treatment	**Day 90 (3 months) Follow-up**		
Negative: 44 (98%)	Negative	43[a]	98% [88-99]
	Positive	1[b]	
Positive: 1 (2%)	Positive	0	-
	Negative	1[c]	
Day 90 (3 months) Follow-up	**Day 180 (6 months) Follow-up**		
Negative: 44 (98%)	Negative	43[d]	98% [88-99]
	Positive	1[e]	
Positive: 1 (2%)	Positive	0	-
	Negative	1[f]	

[a] A total of 2 cases received rescue treatment; [b] 1 patient had received rescue by M3
[c] 1 patient was given rescue treatment; [d] A total of 2 cases received rescue treatment
[e] 1 patient had received rescue by M3; [f] 1 patient was given rescue treatment;

Conditional on a patient ever having a success, 98.9% of their observations had a success outcome. Similarly, conditional on ever having a failure, 75% of a patient's observation had a failure. Both percentages are a measure of the stability of the two outcomes indicating more stability for the success outcome when compared to failure, with the overall stability of outcomes being 97.8% (see Table 4.5).

At each visit, some 98.8% of treatment success in the data remained treatment successes at the next visit with the other 1.2% being failures at the next visit. Although treatment successes had a 1% chance of becoming failures at the next visit there were no treatment failures (see Table 4.6).

Standard predictions:

From this data set, a total of 45 patients had complete follow-up data on all assessment time points. Among the 44 patients who had cleared parasites (negative) by EOT, only 1 (2.2%) became positive by the M3 follow-up assessment indicating a 98% probability of a patient with no parasites at EOT remaining parasite free by M3 follow-up while among the 44 patients who had cleared parasites (negative) by M3, only 1 (2.2%) became positive by the M6 follow-up assessment implying a 98% probability of a patient with no parasites at M3 remaining parasite free by M6 follow-up.

Characteristics of patients who change cure status: End of Treatment to Month 3

One patient was negative at EOT but became positive by M3. The patient had increased spleen size between the two time points (3 cm to 5 cm). One patient had a positive outcome at EOT which became negative by M3 (upon receipt of rescue at the end of treatment assessment). It had a reduction in its spleen size from 7cm to 2cm.

Characteristics of patients who change cure status: M3 to M6

One patient was negative at M3 but became positive by M6. The patient had increased spleen size between the two time points and looks like a relapse case. Was given rescue treatment at M6 visit assessment. One patient had a positive outcome at M3 which became negative by M6 (upon receipt of rescue in between M3 and M6).

Discussions

One of the interesting results from this review is the significant number of patients in the two clinical trials in East Africa who are not being assessed within the expected follow-up time points. This could be as a result of several factors ranging from the nature of the disease itself to the patient population as well as other factors which are still unknown but are related to patient motivation to attend follow-up assessments after treatment completion.

In LEAP 0104 trial, out of the 972 patients enrolled only 532 (55%) were assessed within the expected 3 month follow-up visit window with a slight improvement (63%) of the same at the 6 month follow-up time point. In the LEAP 0106 trial, the scenario is not any different (about 50% at 3 months and 67% at the 6 months time points). This is an important finding related

to the conduct of VL trials in East Africa and could provide valuable learning points towards improving follow-up and justifies the need even if to a smaller extent to evaluate and validate shorter follow-up in patients for better management.

From these results, it is clear that there are very few cases in which changes in treatment outcomes occur between M3 and M6 follow-up time points. The 95% confidence interval on no change in cure status between M3 and M6 has a probability of between 98% and 99.9%. We therefore suggest that a reduction in the follow-up from 6 month to 3 months does not significantly change the expected treatment outcome. Under very limited circumstances (about 0.5%) do patients change cure status between M3 and M6 (see Table 4.6 and Table 4.10). It is therefore possible to predict this limited change based on other patient characteristics like increase in organ size (e.g spleen or liver) after cure.

In terms of transition probabilities from the multi-state model, the change in treatment outcomes between visits is just about 1% in-case of treatment success at the first time point. The treatment outcomes also have high overall stability (>97%) as such can be relied on to form the basis for a review of the follow-up time point in the assessment of definitive cure for VL. Generally, few patients changed their cure status between EOT, M3 and M6 follow-ups with those changing from success to failure all having slightly higher spleen sizes at the points they are declared failures. Nearly all patients who become successes after failure had received rescue treatment and just a few who did not receive any rescue (considered slow responders) had significant reduction in their spleen sizes.

Due to challenges with patient follow-up, only a third of the data set have been used in this analyses but this is assumed to be representative of the patient population of VL trials conducted in East Africa.

4.3 Safety of VL treatments in eastern africa: SSG

A total of 702 patients with clinical and parasitological diagnosis of VL (parasites demonstrated by microscopy in stained tissue samples) were included. Patients with any serious underlying diseases (cardio, renal or hepatic) or previous history of cardio arrhythmia were excluded from the study [33].

4.3.1 Safety profiling

Electrocardiogram Assessments

Examinations for ECG at baseline were carried out for all patients across all sites (Table 4.12). A total of 26 patients did not have any results for ECG assessments at end of treatment. 12 patients treatment was stopped and put on rescue medication, 1 patient show lack of response to treatment and put on rescue medication, 3 patients withdrew consent during treatment, 3 patients died without being put on rescue medication and 7 patients remained in study but did not have ECG assessed at EOT, of which two patients reported as missing values (each of PM and Combination), the other five remained in the trial with two patients reporting abnormal but insignificant ECG values at baseline (each on SSG and Combination).

Less than 1% of patients had a clinically significant ECG abnormality at EOT, all in the combination arm (2 patients) - these were T-wave inversion in V1-V4 suggestive of anterior ischemia and arrhythmia. However, abnormal ECG values returned to normal before 3 months or 6 months follow-up (Table 4.13).

During treatment one patient had a clinically significant ECG abnormality on day 7, a male patient from Sudan, taking 1340mg per day of SSG was diagnosed with an abnormal ECG, although he had no complaints. On ex-

amination he had no systemic abnormalities and had a regular weight of 68kg for a 27 year old, baseline ECG was abnormal but clinically insignificant (no QT interval was provided). QT /QTc interval was 452/504ms - considered a serious adverse event. While other laboratory parameters are in normal ranges, treatment was stopped on day 7 and daily ECGs were done to monitor patient progress. On day 9 and 10 QT/QTc interval returned to normal. The patient had no further ECG evaluations while on rescue medication (Table 4.14). Approximately 0.4% (3 out of 702 patients) of all patients developed abnormal and clinically significant ECG at any time during treatment.

Adverse events

The most direct method to detect cardiotoxicity is through the reporting of adverse events (non-serious and serious). A total of 456 patients experienced one or more adverse events during the trial, of which 386 patients had events that occurred during the treatment emergent period. Cardio-based disorders were found in 6 patients: 2 on SSG and 4 on Combination.

On the Combination arm, two patients suffered from sinus Bradycardia during follow-up and were defined as unrelated to the drug. Two patients on 500mg/day and 350 mg/day of SSG respectively, suffered from mild T-wave inversions and prolonged QT at the EOT, both already described previously. The first patient (Male, 26kg at baseline) was found to have T-wave inversion in V1-V4 suggestive of anterior ischemia in a mild form on day 18. Later, the patient developed unrelated chronic conjunctivitis during follow-up. The second patient (male, 17.5 kg at baseline) had prolonged QT intervals at EOT on day 18 as well as mild arrhythmia detected during ECG assessment at EOT. Both patient ECGs returned to normal during follow-up.

A patient on the SSG arm (male, 52kg at baseline) suffered from mild palpitations during treatment but had clinically insignificant abnormal ECG. The patient also suffered from mild urinary tract infection, headache, rhini-

tis, pyrexia and pneumonia during treatment. A patient with mildly prolonged QT was classified as having clinically insignificant ECG value on day 21. The patient also experienced other adverse events such as injection site pain, urinary tract infection, malaria, and pneumonia. Both patients ECG returned to normal by EOT.

Serious adverse events of any type were experienced by approximately 4% of patients overall. A total of 30 SAE's were reported. cardiotoxicity was reported in only one patient (male, 54kg at baseline) in the SSG arm (Table 4.16), who was given a dosage of 850mg per day. On day 5 of treatment, the patient developed moderate dyspepsia, and by day 11 he developed severe cardiotoxicity, which was possibly related to the SSG treatment. The patient died the same day with no rescue treatment provided. ECG at baseline was abnormal but clinically insignificant and no further ECG readings were taken.

Table 4.12: ECG examinations by centre

Day	Gondar $n = 45$	Arba Minch $(n = 45)$	Kassab $(n = 442)$	KEMRI $(n = 146)$	Amudat $(n = 24)$	Total $(n = 702)$
0	45	45	442	146	24	702
7	0	0	439	143	3	585
14	43	45	419	137	17	661
21[χ]	0	0	27	55	2	81
End of Treatment [φ]	43	45	430	138	20	676
90	0	43	0	133	20	196
180	48	45	418	136	24	661

[χ] only SSG arm were evaluated at day 21 for ECG.

[φ] End of treatment for each arm is different: Combination (17 days), SSG (30 days), PM (21 days)

The effect of antimony based treatment, such as SSG, on ECG readings has been in question since the 1940s, when it was found to cause toxicity in patients being treated for schistosomiasis [29]. The first study to look at ECG changes in leishmaniasis patients treated with SSG in eastern Africa, was done in 1985 [12]; which recommended that "ECG assessments are only necessary on dosages above 20 mg/kg/day for more than 20 days, and should be carried out every 3-4 days." It was noted that over a long duration and

Table 4.13: Overall ECG during treatment and follow-up

	SSG	SSG+PM	PM
Total patients assessed at day 0	251	246	205
Number of patients normal* at baseline	251	246	205
Abnormal at end of treatment	0	2 (0.8)$^{\varphi}$	0
Abnormal at 6 months follow-up	0	0	0

* normal is the combination of normal or clinically insignificant ECG readings at day 0

$^{\varphi}$ One patient was shown to have abnormal ECG but clinically insignificant values since baseline. Other patient developed abnormal values gradually during treatment.

Table 4.14: ECG shifts between baselines, weekly assessments to EOT

ECG Shifts	Day 7				Day 14				Day 21				End of treatment			
Baseline	Normal	ACI	ACS	ND	Normal	ACI	ACS	ND	Normal	ACI	ACS	ND	Normal	ACI	ACS	ND
Normal ($n = 222$)	60	97	1*	64	102	105	-	12	11	17	-	52	104	110	1	3
ACI ($n = 480$)	86	341	0	50	117	337	-	14	20	36	-	111	131	326	1	6
ACS ($n = 0$)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ND ($n = 0$)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Absolute values. ACI - Abnormal, Clinically Insignificant; ACS - Abnormal Clinically Significant; ND - Not Done ECG

* patient had prolonged QT interval during day 7 ECG defined as severe by clinician. Treatment was stopped and patient put on rescue medication. No further ECG examinations were done. Patient was provided a dose of 1340 mg per day of SSG on the combination arm.

Table 4.15: Number of patients experiencing adverse events

	SSG	Combination	PM	Total
Number randomised	251	246	205	702
Number of Patients with $\geq$ 1 AE				
At any time	172 (68.5)	158 (64.2)	126 (61.5)	456 (65.0)
Treatment emergent†	147 (58.6)	132 (53.7)	107 (52.2)	386 (55.0)
During follow-up	78 (31.1)	72 (29.3)	54 (26.3)	204 (29.1)
Number of AE*:				
Total	370 (41.0)	281 (31.2)	251 (27.8)	902
Treatment emergent†	268 (40.5)	208 (31.4)	186 (28.1)	662
During follow-up	102 (42.5)	73 (30.4)	65 (27.1)	240
Adverse drug reaction‡	227 (37.2)	281 (31.2)	251 (27.8)	611
Unrelated to study drug	143 (49.1)	94 (32.3)	54 (18.6)	291
Cardio disorders	**2 (0.8)**	**4 (1.6)**	**0 (0.0)**	**6 (8.5)**

AE=Adverse event; Data are $n(\%)$ of patients randomised

* Not all patients experienced an AE whereas some patients experienced more than one AE

† Treatment emergent defined as onset being between day 1 of treatment and 30 days post end of treatment, inclusive.

‡ Adverse drug reaction recorded as unlikely, possible or probable relation to study drug

with high doses of SSG treatment, the body tends to accumulate Sb (antimony) in the system leading to toxicity, and in particular can be associated with changes in corrected QT interval, arrhythmias and in S and T wave inversions [13, 12]. Later in 2001, a study in Ethiopia assessed the risk of car-

Table 4.16: Adverse events: Cardio

Arm	Dosage (mg/day)	Preferred term	Intensity	Treatment Emergent	Relation to drug$^{\chi}$	Time of AE*
Combination	SSG: 850 PM: 800	Sinus Bradycardia	Moderate	No	Unlikely	94
Combination	SSG: 380 PM: 300	Sinus Bradycardia	Mild	No	Not related	113
Combination	SSG: 500 PM: 400	Electrocardiogram T-wave inversion	Mild	Yes	Unlikely	18
Combination	SSG: 350 PM: 187.5	Electrocardiogram QT prolongation	Mild	Yes	Possible	18
SSG	SSG: 670	Electrocardiogram QT prolongation	Mild	Yes	Possible	21
SSG	SSG: 850	Palpitations	Mild	Yes	Possible	9

* calculated as time in days after start of treatment

$^{\chi}$ADR=Adverse drug reaction, recorded by investigator as Probable, Possible or Unlikely relation to study drug

Table 4.17: Number of patients experiencing SAE's

	SSG	Combination	PM	Total
Number Randomised	251	246	205	702
Number of patients with SAE*:				
Total	9 (3.6)	13 (5.3)	8 (3.9)	30 (4.3)
Treatment Emergent†	7 (2.8)	13 (5.3)	7 (3.4)	27 (3.9)
During Follow-up	2 (0.8)	0 (0.0)	1 (0.5)	3 (0.4)
Adverse drug reaction‡	5 (2.0)	11 (4.5)	6 (2.9)	22 (3.1)
Unrelated to study drug	4 (1.6)	2 (0.8)	2 (1.0)	8 (1.1)
Cardio disorders	**1 (0.4)**	**1 (0.4)§**	**0 (0.0)**	**2 (0.3)**

Data are $n(\%)$ of patients randomised

SAE=Serious Adverse Event, AE=Non-serious adverse event

* No patients experienced more than one SAE

† Treatment emergent defined as onset being between day 1 of treatment and 30 days post end of treatment, inclusive.

‡ Adverse drug reaction recorded as unlikely, possible or probable relation to study drug.

§ Patient experienced prolonged QT interval but coded under investigation using MedDRA. Patient had abnormal ECG results on day 7 and was immediately put on rescue medication

diotoxicity related to antimonial treatments of VL patients in south Ethiopia and found no significant toxicity. From the data of 702 patients, on either PM Monotherapy, SSG monotherapy or a combination of SSG and PM, of

which only three patients, all on the combination arm reported abnormal ECG changes at any time during treatment. Cardio disorders were found in eight patients (including the 3 patients with abnormal ECG assessments) during treatment, two of which were severe (one death due to cardiotoxicity and one prolonged QT interval which resulted in stopping of treatment and patient receiving rescue medication). Other cardio disorders observed were Sinus Bradycardia, T-wave inversion, mild QT interval prolongation and palpitations.

Chapter 5

SUMMARY, CONCLUSIONS AND RECOMENDATIONS

In this chapter, a presentation of the summary of study findings (Section 5.1), conclusions from the study (Section 5.2) and recommendations based on the results of the study (Section 5.3) have been made.

5.1 Summary

The objectives of the study were to develop alternative approaches to the analyses of definitive cure following the triangular test design, estimate the optimal timing for definitive cure assessment post end of treatment in VL patients and establish if an association exists between SSG treatment and occurrence of cardio-toxicity among VL patients in eastern Africa.

In the first objective, two alternative analyses approaches i.e the PTE and SHE are viable alternative approaches in the estimation of extended follow-up besides MLE. They both give low bias and RMSE values and have high coverage probabilities (see Section 4.1 Efficacy at Extended follow-up (day

210)).

In the second objective, there is a very low probability of change in cure status between M3 and M6 with the 95% C.I of no change in cure status between M3 and M6 having a probability of between 98% and 99.9%.

In the third objective, the review of available data on SSG use in eastern Africa shows that there is not enough evidence to suggest that SSG is associated with cardio-toxicity.

5.2 Conclusions

The PTE and SHE are viable alternative approaches in the estimation of extended follow-up following the triangular test design in VL clinical trials.

There is no significant change in the cure status between the month 3 and month 6 end points in the assessment definitive cure.

There is not enough evidence to suggest association of SSG use with cardio-toxicity in eastern Africa.

5.3 Recommendations

From the study results, the following recommendations can be made

1. That both the PTE and SHE are viable analysis approaches for efficacy at extended follow-up when the triangular test design is used in the conduct of VL clinical trials. Other analysis approaches particularly those of Bayesian nature needs to be explored and comparisons made with the PTE and SHE on the efficiency.
2. Month 3 as an end point for assessment of definitive cure can be used

as a replacement for the month 6 endpoint but this needs to be validated in well-designed trial with additional endpoints at month 4 or month 5.

3. The current first line treatment for VL in eastern Africa is a safe treatment in the current dosage, however there is need to continuously monitor its safety in wider usage particularly through a post-market surveillance study in the region.

REFERENCES

[1] Agency, E. (2006). September 1998 ICH topic E 9 statistical principles for clinical trials step 5 note for guidance on (CPMP/ICH/363/96). *EMEA*.

[2] Agresti, A. (2002). *Categorical Data Analysis*. Wiley & Sons.

[3] Alvar, J., Vélez, I., Bern, C., et al. (2012). Leishmaniasis worldwide and global estimates of its incidence. *PLoS One*, 7:e35671.

[4] Andersen, P. and Keiding, N. (2002). Multi-state models for event history analysis. *Stat. Methods Med. Res.*, 11:91–115.

[5] Armitage, M. and Berry, G. (2001). *Statistical Methods in Medical Research. 4th ed.* Oxford: Blackwell Scientific Publications.

[6] Balzan, M. and Fenech, F. (1992). Acute renal failure in visceral leishmaniasis treated with sodium stibogluconate. *Trans. R. Soc. Trop. Med. Hyg.*, 86:515–516.

[7] Bellissant, E., Benichou, J., and Chastang, C. (1990). Application of the triangular test to phase i1. *Cancer Clin. Trials.*, 9:907–917.

[8] Bellissant, E., Bénichou, J., and Chastang, C. (1996). The group sequential triangular test for phase ii cancer clinical trials. *Am. J. Clin. Oncol.*, 19:422–430.

[9] Berhe, N., Abraham, Y., Hailu, A., et al. (2001). Electrocardiographic findings in ethiopians on pentavalent antimony therapy for visceral leishmaniasis. *East Afr. Med. J.*, 78:608–610.

[10] Berry, D., Broglio, S., Groshen, K., and Berry, S. (2013). Bayesian hierarchical modeling of patient subpopulations: E_cient designs of phase ii oncology clinical trials. *Clin. Trials*, 10:720–734.

[11] Boelaert, M. (2007). Evaluation of rapid diagnostic tests: visceral leishmaniasis. *Nat. Microbiol*, pages S30–S39.

[12] Chulay, J., Bhatt, S., Muigai, R., et al. (1983). A comparison of three dosage regimens of sodium stibogluconate in the treatment of visceral leishmaniasis in kenya. *J. Infect. Dis.*, 148:148–155.

[13] Chulay, M. and Spencer, H. (1985). Electrocardiographic changes during treatment of leishmaniasis with pentavalent antimony (sodium stibogluconate). *Am. J. Trop. Med. Hyg.*, 34:702–709.

[14] Croft, S. (2001). Monitoring drug resistance in leishmaniasis. *Trop. Med. Int. Health*, 6:899–905.

[15] de Beer, P., el Harith, A., Deng, L., et al. (1991). A killing disease epidemic among displaced sudanese population identified as visceral leishmaniasis. *Am. J. Trop. Med. Hyg.*, 44:283–289.

[16] DNDi. Drugs for neglected diseases initiative, (n.d.). `http://www.dndi.org`. Accessed: 2015-07-10.

[17] Edwards, T., Omollo, R., Yifru, S., et al. (2011). Single-dose liposomal amphotericin B (AmBisome®) for the treatment of visceral leishmaniasis in east africa: study protocol for a randomized controlled trial. *Trials*, 12:66.

[18] FDA. Guideline for good clinical practice E6 (R1), 1996. `http://www.fda.gov/cder/guidance/index.html`. Accessed: 2015-07-06.

[19] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Analysis*, 1 No, 3:515–534.

[20] Guihenneuc-Jouyaux, C., Richardson, S., and Longini, I. (2000). Modeling markers of disease progression by a hidden markov process: application to characterizing cd4 cell decline. *Biometrics*, 56:733–741.

[21] Hailu, A., Musa, A., Royce, C., and Wasunna, M. (2005). Visceral leishmaniasis: New health tools are needed. *PLoS Med.*, 2:590–594.

[22] Hailu, A., Musa, A., Wasunna, M., et al. (2010). Geographical variation in the response of visceral leishmaniasis to paromomycin in east africa: a multicentre, open-label, randomized trial. *PLoS Negl. Trop. Dis.*, 4:e709.

[23] Hepburn, H., Siddique, I., Howie, A., and Beckett, G. (1993). Hepatotoxicity of sodium stibogluconate in leishmaniasis. *Lancet*, 342:238–239.

[24] Hepburn, N. C., Nolan, J., Fenn, L., et al. (1994). Cardiac effects of sodium stibogluconate: myocardial, electrophysiological and biochemical studies. *Q. J. Med.*, 87:465–472.

[25] Khalil, E. A. G., Weldegebreal, T., Omollo, R., et al. (2014). Safety and efficacy of single dose versus multiple doses of AmBisome® for treatment of visceral leishmaniasis in eastern africa: A randomised trial. *PLoS Negl. Trop. Dis.*, 8:e2613.

[26] Konstantopoulos, T. Markov chains and random walks, (2009). `http://www2.math.uu.se/~takis/L/McRw/mcrw.pdf`. Accessed: 2015-07-10.

[27] Liu, A., Tan, M., Boyett, J., and Xiong, X. (2000). Testing secondary hypotheses following sequential clinical trials. *Biometrics*, 56:640–644.

[28] Maheshwari, D., Seth, A., Kaur, S., et al. (2011). Cumulative cardiac toxicity of sodium stibogluconate and amphotericin B in treatment of kala-azar. *Paedatric Infect. Dis. J.*, 30:180–181.

[29] Mainzer, M. and Krause, F. (1940). Changes of the electrocardiogram appearing during antimony treatment. *Trans. R. Soc. Trop. Med. Hyg.*, 33:405–418.

[30] Maltezou, H. (2010). Drug resistance in visceral leishmaniasis. *J. Biomed. Biotechnol*, 2010:617521.

[31] M.K. Wasunna and, J. R., Mbui, J., et al. (2005). A phase ii dose-increasing study of sitamaquine for the treatment of visceral leishmaniasis in kenya. *Am. J. Trop. Med. Hyg.*, 73:871–876.

[32] MSSO. MedDRA, (2013) MedDRA MSSO. `http://www.meddramsso.com/subscriber_new_subscribers.asp`. Accessed: 2015-07-10.

[33] Musa, A., Khalil, E., Omollo, R., et al. (2012). Sodium stibogluconate (SSG) and paromomycin combination compared to SSG for visceral leishmaniasis in east africa: A randomised controlled trial. *PLoS Negl. Trop. Dis.*, 6:e1674.

[34] Musa, A., Younis, B., Fadlalla, A., et al. (2010). Paromomycin for the treatment of visceral leishmaniasis in sudan: a randomized, open-label, dose-finding study. *PLoS Negl. Trop. Dis.*, 4:e855.

[35] Olliaro, P., Vaillant, M., Sundar, S., and Balasegaram, M. (2012). More efficient ways of assessing treatments for neglected tropical diseases are required: Innovative study designs, new endpoints, and markers of effects. *PLoS Negl. Trop. Dis.*, 6:e1545.

[36] Omollo, R., Alexander, N., Edwards, T., et al. (2011). Safety and efficacy of miltefosine alone and in combination with sodium stibogluconate and liposomal amphotericin B for the treatment of primary visceral leishmaniasis in east africa: study protocol for a randomized controlled trial. *Trials*, 12:166.

[37] Piloquet, H., Berger, B., Rochat, F., et al. (2009). supplementation with probiotics in very-low-birth-weight preterm infants: a randomized , double-blind , placebo-controlled trial 1 - 4. *American Journal of Clinical Nutrition*, 89:1828–1835.

[38] Ranque, S., Badiaga, S., Delmont, J., and Brouqui, P. (2002). Triangular test applied to the clinical trial of azithromycin against relapses in plasmodium vivax infections. *Malaria Journal*, 7.

[39] Rijal, K., Chappuis, F., Singh, R., et al. (2003). Sodium stibogluconate cardiotoxicity and safety of generics. *Trans. R. Soc. Trop. Med. Hyg.*, 97:597–598.

[40] RStudio Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.

[41] StataCorp (2003). *Statacorp Stata Statistical Software: Release 13.1*. StataCorp LP, Texas.

[42] Sundar, S. (2001). Drug resistance in indian visceral leishmaniasis. *Trop. Med. Int. Health*, 6:849–854.

[43] Sundar, S., Agrawal, N., Arora, R., et al. (2009). Short-course paromomycin treatment of visceral leishmaniasis in india: 14-day vs 21-day treatment. *Clin. Infect. Dis.*, 49:914–948.

[44] Sundar, S. and Chakravarty, J. (2010). Antimony toxicity. *Int. J. Environ. Res. Public Health*, 7:4267–4277.

[45] Sundar, S., Chakravarty, J., Agarwal, D., et al. (2008a). Safety of a preformulated amphotericin b lipid emulsion for the treatment of indian kala-azar. *Trop. Med. Int. Health*, 13:1208–1212.

[46] Sundar, S., Jha, T., Thakur, C., et al. (2003). Single-dose liposomal amphotericin B in the treatment of visceral leishmaniasis in india: a multicenter study. *Clin. Infect. Dis.*, 37:800–804.

[47] Sundar, S., Jha, T., Thakur, C., et al. (2007). Injectable paromomycin for visceral leishmaniasis in India. *new england journal*, 356 (25):2571–2581.

[48] Sundar, S., Mehta, H., Chhabra, A., et al. (2006). Amphotericin B colloidal dispersion for the treatment of indian visceral leishmaniasis. *Clin. Infect. Dis.*, 42:608–613.

[49] Sundar, S., Rai, M., Chakravarty, J., et al. (2008b). New treatment approach in indian visceral leishmaniasis: single-dose liposomal amphotericin b followed by short-course oral miltefosine. *Clin. Infect. Dis.*, 47:1000–1006.

[50] Thakur, C., Kumar, A., Mitra, D., et al. (2010). Improving outcome of treatment of kala-azar by supplementation of amphotericin B with physiologic saline and potassium chloride. *Am. J. Trop. Med. Hyg.*, 83:1040–1043.

[51] Thakur, C. and Narayan, S. (2004). A comparative evaluation of amphotericin b and sodiumantimony gluconate, as first-line drugs in the treatment of indian visceral eishmaniasis. *Annals of Tropical Medicine and Parasitology*, 98 (2):129–138.

[52] Thakur, C., Sinha, G., Pandey, A. K., et al. (1998). Do the diminishing efficacy and increasing toxicity of sodium stibogluconate in the treatment of visceral leishmaniasis in bihar, india, justify its continued use as a first-line drug?, an observational study of 80 cases. *Ann. Trop. Med. Parasitol.*, 92:561–569.

[53] Todd, S. (2007). A 25-year review of sequential methodology in clinical studies. *Stat. Med.*, pages 237–252.

[54] Todd, S. and Whitehead, J. (1996). Point and interval estimation following a sequential clinical trial. *Biometrika*, 83(2):453–461.

[55] van Griensven, J., Balasegaram, M., Meheus, F., et al. (2010). Combination therapy for visceral leishmaniasis. *Lancet Infect. Dis.*, 10:184–194.

[56] Welton, N. and Ades, E. (2005). Estimation of markov chain transition probabilities and rates from fully and partially observed data: un-

certainty propagation, evidence synthesis, and model calibration. *Med. Decis. Making*, 25:633–645.

[57] Whitehead, J. (1986). On the bias of maximum likelihood estimation following a sequential test. *Biometrika*, 73:573.

[58] Whitehead, J. (1997). *The Design and Analysis of Sequential Clinical Trials, Revised, 2nd Edition*. Wiley & Sons.

[59] WHO. Control of the leishmaniases, geneva, 2010. `http://whqlibdoc.who.int/trs/WHO_TRS_949_eng.pdf`. Accessed: 2015-07-10.