WHO network for HAT elimination

Human African trypanosomiasis: update of the methodological framework for clinical trials

Report of the first meeting of the Development of New Tools subgroup

Geneva, 24 September 2014

World Health Organization

All WHO Library Cataloguing-in-Publication Data:

Human African trypanosomiasis: update of the methodological framework for clinical trials: report of the first meeting of the Development of New Tools subgroup, Geneva, 24 September 2014.

1.Trypanosomiasis, African – drug therapy. 2.Clinical Trials as Topic. 3.Trypanocidal Agents – administration and dosage. 4.Trypanocidal Agents – adverse effects. I.World Health Organization.

ISBN 978 92 4 150883 4 (NLM classification: WC 705)

Printed in (country name)

WHO/HTM/NTD/IDM/2015.5

Contents

1 Introduction 1

1.1 Justification for and objectives of the meeting 1

1.2 Methodological approach of the meeting 1

1.3 Scope of the meeting 2

2 Identification of HAT patients for clinical trials 2

2.1 Screening 2

2.2 Diagnosis: general considerations 3

2.3 Staging 4

2.4 Collaboration between clinical trial teams and national control programmes for identification and case-management of patients 5

2.5 Enrolment criteria by therapeutic target 6

2.6 Use of historical controls 7

3 Assessment of therapeutic efficacy in HAT patients 8

3.1 General considerations and terminology 8

3.2 Evaluation of patients at the end of treatment 9

3.2.1 Evaluation of first-stage patients at the end of treatment 10

3.2.2 Evaluation of second-stage patients at the end of treatment 10

3.2.3 Evaluation of intermediate-stage patients at the end of treatment 11

3.3 Follow-up of patients for efficacy assessment 11

3.3.1 Time-point for final assessment of efficacy (ToC visit) 11

3.3.2 Early prediction of efficacy 13

3.4 Interim follow-up and test-of-cure 16

3.4.1 Evaluation of patients during follow-up 16

3.4.2 Deaths 16

4 Quantification of therapeutic efficacy 20

4.1 Analysis populations 20

4.2 Time windows for assigning patients' follow-up data to planned visits 22

4.3 Handling of missing efficacy data 22

4.4 Efficacy variables 23

4.4.1 Nomenclature 23

4.4.2 Calculation of efficacy 23

5 References 25

6 Annexes 28

Annex 1. Agenda of the meeting 28

Annex 2. List of participants 29

Annex 3. Phase IV studies in HAT treatment development 30

Annex 2. Nomenclature for analysis sets 31

Abbreviations and acronyms

AE	adverse events
CATT	card agglutination test for trypanosomiasis
CSF	cerebrospinal fluid
DSMB	data safety monitoring board
EMA	European medicines agency
EoT	end-of-treatment
ERB	Ethical Review Board
FDA	Food and Drug Administration
g-HAT	gambiense human African trypanosomiasis
HAT	human African trypanosomiasis
IDM	Innovative and Intensified Disease Management
ITT	intention-to-treat
LP	lumbar puncture
mAECT	mini-anion exchange centrifugation technique
mAECT-BC	mAECT on buffy coat
mHCT	micro haematocrit centrifugation technique
mITT	modified intention-to-treat
NECT	nifurtimox–eflornithine combination therapy
NTD	neglected tropical diseases
PLPHA	post lumbar puncture headache
r-HAT	rhodesiense human African trypanosomiasis
RST	rapid serological test
ToC	test-of-cure
WBC	white blood cell
WHO	World Health Organization

1 Introduction

1.1 Justification for and objectives of the meeting

Researchers involved in clinical trials for the evaluation of new treatment modalities for human African trypanosomiasis (HAT), also known as sleeping sickness, face a number of challenges that are rarely, if ever, encountered in this combination in other diseases. Many of these challenges are related to the fact that both the disease and the populations it affects are neglected and that, prior to 2004, there was no background of generally accepted – and ubiquitously feasible – diagnostic and treatment standards for the planning and conduct of clinical evaluation of new treatment modalities for a disease.

In 2004, the World Health Organization (WHO) organized an expert consultation to establish a methodological framework for clinical trials on HAT in order to facilitate collaboration among research actors and comparison of the data obtained by different groups (WHO 2007). The agreed common criteria were applied from that point by the different researchers, which created a new harmony and a collaborative environment.

During the following decade, thanks to renewed research efforts, new diagnostic tools and new knowledge on assessing treatment outcomes became available (WHO 2013), which may allow improvement of the clinical trial methodology. In addition, the number of HAT cases reported annually to WHO has fallen to fewer than 7000, and the disease has been targeted for elimination. These new facts justify an update of some of the criteria adopted in 2004.

This meeting was framed by the WHO Network for HAT Elimination and it convened specifically the sub-group "Development of new tools", with the following objectives:

- To review and discuss how the new knowledge made available since 2004 could impact the implementation of clinical trials; and
- To update the consensus framework for the planning, conduct and analysis of clinical trials in the future in a way that would promote the acquisition of data that can be readily compared and used in meta-analysis.

1.2 Methodological approach of the meeting

As an initial step, the document of the 2004 WHO expert consultation was circulated among the subgroup members who identified points to update and provided their suggestions.

On the basis of this input, the WHO secretariat produced two working documents which were sent to all the subgroup members:

1. An early draft document: based on the 2004 document and already including many of the comments and suggestions received.
2. A list of Key Questions: arising from the experts' proposals. Some questions were already included in the draft, which was indicated. Under each question, four boxes were provided to enter arguments in support, arguments against, positive outcomes

and negative outcomes foreseen. Experts were invited to create new key questions if considered necessary.

Before the meeting, the Key Questions document containing the consolidated input from experts was sent to all the subgroup members in preparation for the discussions at the meeting.

The meeting took place on 24 September 2014 at WHO headquarters in Geneva, Switzerland. The participants were provided with a working draft 2 (showing track-changes from draft 1) at the meeting. The discussions of the meeting were structured around the Key Questions, which were in turn aligned with the text of the draft document.

The discussion points, having reached consensus at the meeting, were integrated in draft 3 of the document which was circulated afterwards among all subgroup members for verification.

The Agenda of the meeting is included as Annex 1 and the List of participants as Annex 2.

1.3 Scope of the meeting

The discussions and conclusions of the meeting were driven by the need to evaluate the efficacy of new treatment regimens, but are in some cases directly applicable or easily adaptable to the evaluation of new diagnostics.

The conclusions focused on the acquisition of data from clinical trials, since data acquired according to common criteria are a prerequisite for any meaningful comparison between the outcomes of different clinical trials.

With the objective of direct comparability of published data on drug efficacy in mind, a framework for analysis and reporting of the efficacy of the treatment regimens under evaluation was also agreed upon.

The group did not discuss the safety evaluation aspect of HAT clinical trials.

This document concerns gambiense HAT (g-HAT). In the case of rhodesiense HAT (r-HAT), the body of clinical evidence is extremely limited and therefore the elements developed here cannot always be applied in studies of clinical products addressed to r-HAT.

2 Identification of HAT patients for clinical trials

2.1 Screening

The card agglutination test for trypanosomiasis (CATT) on whole blood, a serological test for antibody detection, is used by all national sleeping sickness control programmes for mass screening of the population in areas where *Trypanosoma brucei gambiense* is endemic. Its reported specificity is around 0.97 (Checchi 2011).

Newly developed rapid serological test (RSTs) may be used to replace CATT for screening as appropriate. Their reported specificity is between 0.87 and 0.98 (Büscher 2014; SD Bioline HAT test insert).

At the low HAT prevalence currently observed in most HAT foci, serological tests may result in low positive predictive values. Therefore, before inclusion into a clinical trial, all individuals testing positive on serological tests must undergo parasitological examinations to ascertain the presence of the causative parasite.

Molecular tests can be included, depending on feasibility, in the screening algorithm to increase the overall screening sensitivity and thus accelerate enrolment, i.e. by repeating parasitological examinations (which were initially negative) in individuals with positive molecular test results. For serological tests, individuals testing positive by molecular test should undergo parasitological examinations to confirm the presence of trypanosomes before enrolment into a clinical trial can be considered.

2.2 Diagnosis: general considerations

Only patients for whom trypanosomes are seen in body fluids should be included in clinical trials, to ensure that the efficacy of therapies under evaluation is tested in truly infected individuals.

- The most sensitive parasitological test possible should be used, i.e. mini-anion exchange centrifugation technique (mAECT) (Büscher 2009) or mAECT on buffy coat (Camara 2010) or mHCT (Woo 1971) if mAECT is not available for blood examination; and modified single centrifugation for cerebrospinal fluid (CSF) examination (Büscher 2009; Miézan 2000; Mumba Ngoyi 2013a; Mumba Ngoyi 2013b; Mumba Ngoyi 2014). Enlarged lymph nodes should be punctured for direct examination of lymph-node aspirate.
- Parasitological tests should be performed as soon as possible after sample collection to retain maximum sensitivity. Repeated examinations (if possible, over several days) increase the probability of detecting trypanosomes.
- In order to provide complete parasitological baseline characteristics for the patients enrolled in a clinical trial, blood *and* lymph (when punctionable lymph nodes are present) *and* CSF should be examined for parasites at least once, even if the presence of trypanosomes has already been demonstrated in another body fluid. However, in post-approval phase IV studies on therapies with a well-established risk/benefit ratio, it may be considered to waive the baseline CSF examination in patients with trypanosomes seen in another body fluid (see Annex 3).
- Experience in the field has shown that individuals may be incorrectly categorized as parasitological positives. Confirmation of the presence of trypanosomes in each body fluid by a second staff member is mandatory for clinical trials to reduce the risk of including false parasitological positives in the trial and to obtain accurate data on baseline characteristics.
- Digital recording (taking a picture or video with a mobile phone or other device through the microscope) could be used to allow post-hoc documentation and verification of the images and counts, for quality control purposes. In this case, the study protocol must establish clear rules for the handling of any discrepancies between

the image interpretation and the field results, in terms of patient management, patient classification and data analysis.

- Molecular tests are not a reliable option for confirmation of diagnosis, because currently their analytical performance is not better than the best parasitological tests (Mumba Ngoyi 2014; Mitashi 2013). However, they can be included, depending on feasibility, in the screening algorithm to increase the overall screening sensitivity (see above).

2.3 Staging

- Staging should be based on parasitological criteria and on the white blood cell (WBC) count in the CSF (WHO 2013).
- Neopterin as a CSF marker for second-stage HAT (Tiberti 2012; Tiberti 2013) in clinical trials is not an option at this time, as the supporting evidence is insufficient.
- Lumbar puncture should be performed using disposable spinal needles and following standard procedures of asepsis.
- Post lumbar puncture headache (PLPHA), a well-recognized complication of LP, can be significantly reduced by using atraumatic needles (Davis 2014). Procedural technique may also play a role, with lower rates of PLPHA observed when (i) the needle bevel is inserted parallel to longitudinal dural fibres (Richman 2006) and (ii) the needle stylet is replaced prior to withdrawing the needle (Strupp 1997).
- A volume of at least 5 ml of CSF should be collected, using two or more tubes. It is important that the first drops be discarded to avoid contamination of the CSF with red blood cells.
- Examination of the CSF for both trypanosomes and WBCs should be initiated immediately after collection (Deisenhammer 2006), within 5 minutes, and be completed within 30 minutes of CSF sampling, as CSF trypanosomes may die (and can thus no longer be detected) and CSF WBCs become deformed or disappear quite rapidly after collection of CSF.

The amount of protein in the CSF should not be taken into account for disease staging as abnormally high levels are not specific for second-stage HAT (WHO 2013).

WBC counts

- Before the WBCs are counted, the CSF should be gently mixed to obtain a homogeneous suspension of cells.
- The use of standardized, single-use counting chambers is advised in order to avoid variations in the volume of CSF, which can occur with incorrectly mounted classical counting chambers.
- The counting chamber should never be filled directly from the lumbar puncture needle while the CSF is being collected.
- When the number of WBCs in a single CSF aliquot is < 30 per µl, the cell count should be performed on a second aliquot and the average of the two cell counts should be used for staging.

- When the CSF contains > 200 red blood cells/µl, the patient should be excluded from the trial unless a non-haemorrhagic CSF sample is obtained later. Where red blood cells are present, a distinction may be made between:
    - a traumatic tap that has damaged a blood vessel: the initial drops (first tube) contain more blood than the last drops (subsequent tubes). In this case the LP can be repeated in 12–24 hours.
    - a suspected subarachnoid haemorrhage: blood is uniformly present in the CSF collected (the same in all tubes). The clinician will decide if and when the LP can be repeated, in function of the overall clinical examination, and the neurological examination in particular.
- Türck solution should not be used for three reasons: (i) as Türck solution lyses red blood cells, it is difficult to assess whether a sample is haemorrhagic; (ii) Türck solution may lyse trypanosomes that were present in the CSF; (iii) normal CSF cell counts are already close to the detection limit for counting chambers, and adding Türck solution increases the volume/dilutes the sample and lowers the accuracy of the cell count.

2.4 Collaboration between clinical trial teams and national control programmes for identification and case-management of patients

The study protocol should be approved by the ethical review board (ERB) of the national sleeping sickness control programme.

Because of the relatively low prevalence of HAT in most endemic areas, large numbers of people need to be screened to enrol the required number of patients in clinical trials. Effective detection of patients is thus best carried out via close collaboration between the clinical trial team and the national control programme and/or the nongovernmental organization conducting HAT control in the area where the trial is taking place. Such collaboration requires considerable flexibility from the collaborating groups. Clinical trial sponsors should ensure that the activities related to the clinical trial do not disturb or weaken the control and surveillance activities planned by the national control programme and/or the nongovernmental organization.

Prior to initiation of the trial, the local authorities should be made aware that patients detected as positive for HAT by a mobile team may be proposed for referral to the local hospital/treating facility for a clinical study. The informed consent procedure will take place at the hospital/treating facility.

During discussions by the clinical trial team and the mobile screening team regarding the criteria for referral or transport of patients to the clinical trial centre for further evaluation, factors such as the local prevalence of the disease must be taken into account. The extent to which patients screened by the local mobile team are further evaluated by that team or in the clinical trial centre will depend on the capacity of the mobile team as well as on the capacity of the clinical trial centre. The clinical trial should not have a negative impact on the normal functioning of the centre and the care of other patients. Whenever possible, a special clinical trial team should assist the local mobile team. Patients referred for the trial but not included in it should benefit from the same indirect advantages (e.g. food).

The planning of the clinical trial should take into account the possible negative impact on involved health structures upon termination of the trial, as the routine operations will continue. Such planning should include provisions to anticipate and minimize this impact. A clinical trial may produce unintentional negative effects when it introduces new benefits to participants, including to both patients and medical staff, which are not sustainable when the study ends. Common examples are: patients enrolled in the clinical trial who receive food during their hospitalization (often extended to all HAT patients even if they are not enrolled); patients who are offered transportation and/or money or gifts as incentives for attending the periodic post-therapeutic controls; and health workers involved in the study who are paid a bonus during the study activities in recognition of the extra work generated by their participation.

2.5 Enrolment criteria by therapeutic target

Uniform enrolment criteria are essential to facilitate interpretation and comparison of data across studies carried out by different researchers and testing different therapies.

Table 1. Criteria for inclusion of HAT patients in clinical trials

Targeted patient population	Criteria
First stage	Trypanosome-positive blood and/or lymph Trypanosome-negative CSF ≤ 5 WBC/µl CSF
Intermediate stage 1	≤ 5 WBC/µl CSF and trypanosomes in the CSF
Intermediate stage 2	Trypanosome-positive blood and/or lymph and/or CSF 6–20 WBC/µl CSF
Second stage	Trypanosome-positive blood and/or lymph and/or CSF > 20 WBC/µl CSF

CSF: cerebrospinal fluid; HAT: human African trypanosomiasis; WBC: white blood cells

Some, but not all, of the patients in the intermediate stages can be cured with first-stage drugs, while second-stage patients need specific therapy. Patients with these intermediate characteristics thus have a response to a given drug under evaluation that does not allow conclusions to be drawn as to its efficacy against first-stage or second-stage HAT. Consequently, such patients should not be included in clinical trials of therapies intended for a specific HAT stage.

Patients in the intermediate stages can be included in trials of therapies intended for both HAT stages. However, complete parasitological and biological baseline characteristics must be established for all participants, and the number of second-stage patients must be sufficient to ensure the statistical power to evaluate efficacy in this subgroup.

Given the characteristics of the currently available therapy for first-stage HAT and the process of elimination of the disease, clinical trials targeting first-stage disease alone are not seen as priority.

2.6 Use of historical controls

The single most important barrier that has risen in the past decade to the conduct of HAT clinical trials is the marked decrease in the number of patients available for enrolment. As an illustration, in the latest ongoing study that is still recruiting (fexinidazole in second-stage HAT), 262 257 individuals were screened in targeted high-risk areas in order to identify 391 HAT cases, of whom 107 met the study criteria and were recruited (nearly 2500 people screened per study participant).

The total number of cases reported globally per year – fewer than 7000 for the past 4 years – is expected to keep decreasing owing to the elimination efforts. The situation has become comparable to that of rare diseases, also called orphan diseases, for which regulatory authorities are open to considering the evidence from less conventional methodological approaches provided that they are unavoidable and appropriately justified (EMA CHMP/EWP/83561/2005).

One such approach is the use of historical controls. The consensus of the expert group was favourable to the use of historical controls, on the basis of the following considerations:

- Historical control trials (HCT) would allow more patients to be enrolled in the experimental arm, thus improving feasibility within realistic timelines and budget, and at a time when patients have become rare and increasingly remote.
- The useful information obtained from each patient identified would be maximized.
- It is known that, for the time being, the active control therapy will be nifurtimox–eflornithine combination therapy (NECT); since NECT presented very high estimates of efficacy, the biases that are typical of HCTs are less likely to play a major role.
- The efficacy data already available for NECT come from well-conducted, well-documented studies that used the same dosage and administration modalities. In both studies the observed efficacy was similar, even if the settings were very different: a randomized controlled trial (Priotto 2007 and 2009) and a prospective effectiveness study in field conditions (NECT-Field, publication in preparation).

The experts pointed out some conditions for the design and conduct of HCT:

- Comparability with the historical data should be maximized by following the same procedures as in the historical studies used as control, including, but not limited to, the enrolment criteria, the efficacy assessment methods, and the classification of outcomes.
- Full documentation of patient characteristics known to have an influence in therapeutic response should be ensured (e.g. mode of case detection, biological parameters – especially WBC counts, general health and nutritional status, comorbidity, concomitant treatments, etc.).

Therefore, the use of historical comparison arms is considered acceptable due to the scarcity of patients, but in that case the studies must be carefully planned and well-executed in order to allow for adequate comparisons.

The preferred design remains the randomized trial which is the most rigorous approach and provides firmer evidence.

3 Assessment of therapeutic efficacy in HAT patients

3.1 General considerations and terminology

The limitations of the current methods for diagnosis and staging of HAT, in particular the low sensitivity with which trypanosomes are detected in blood, lymph and CSF, significantly affect the assessment of the efficacy of therapies under study. The reluctance or refusal of patients to come back for follow-up visits and undergo repeated lumbar punctures further aggravates these difficulties, since no data at all or no CSF data (presence or absence of trypanosomes, or WBC count) may be available for the assessment of the patients' response to the treatment.

These factors have certainly contributed to the use of different criteria to determine relapse by different investigators and to the introduction of a category of "suspected relapse" and a category of "patients requiring close follow-up" in many trials for patients with clinical signs and/or CSF WBC counts suggesting a relapse but without parasitological confirmation.

The meeting considered these issues and concluded as follows:

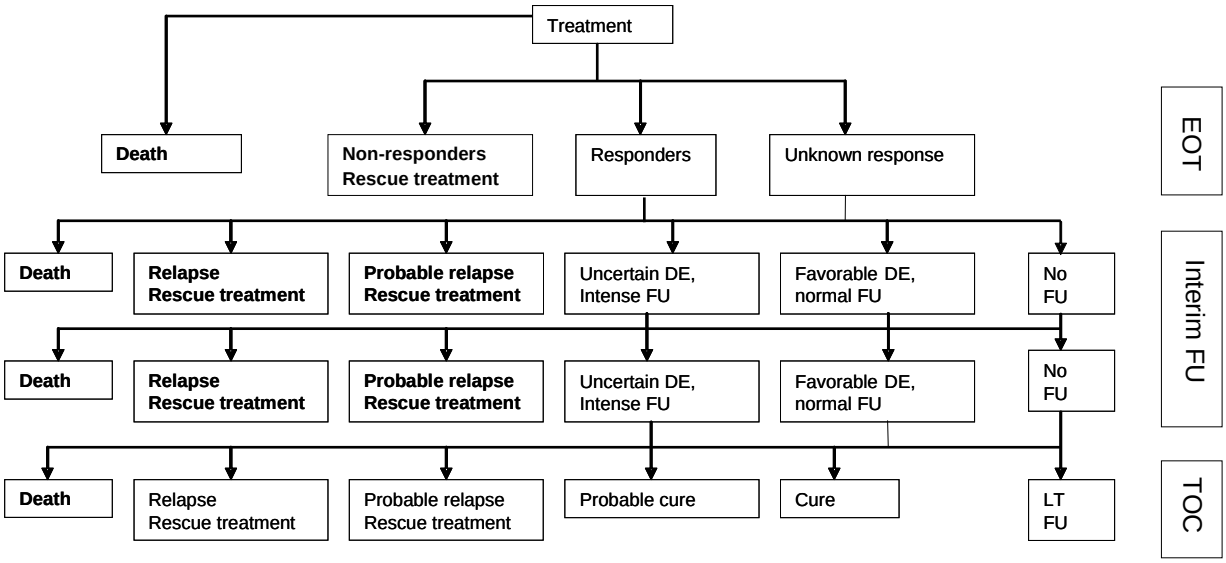
1 For clinical trials, an elaborate classification system is required for patients during follow-up. Figure 1 provides an overview of the agreed classification of patients at different times after treatment: end of treatment (EoT), during interim follow-up, and at the test-of-cure (ToC) visit. The criteria for each classification are provided in section 3.4 for patients with first-stage and second-stage HAT, respectively.

2 There are few hard data that could serve as the basis for objective determination of response status in the absence of parasitological evidence. Consequently, the criteria in the following tables refer to the investigator's judgement when objective criteria are not available.[1]

3 There is a need for all institutions conducting clinical trials or treatment programmes to assemble comprehensive databases on the clinical and laboratory characteristics of all patients at each follow-up. Eventually these can be pooled and analysed for evidence-based, objective criteria for the categorization of response status

[1] The judgement of the investigator could be replaced in a specific study protocol by objective criteria agreed upon between sponsor(s) and investigator(s).

Figure 1. Terminology for assessment of HAT patients after treatment



Bold font End-point reached before ToC visit

DE	Disease evolution
EoT	End-of-treatment evaluation within 1–2 days, latest 14 days, after the end of treatment
FU	Follow-up
Interim FU	Interim follow-up: at 3 (depending on preliminary knowledge about the efficacy of the experimental therapy), 6 and 12 months after treatment. The data collected at the latter could be used for early prediction of efficacy.
LP	Lumbar puncture
LT FU	Lost to follow-up
ToC	Test-of-cure (final efficacy assessment) 18 months after treatment

3.2 Evaluation of patients at the end of treatment

- An "end-of-treatment (EoT) evaluation" is performed within 1–2 days of the end of treatment (at the latest within 14 days of the end of treatment). It consists of parasitological examinations on blood and lymph (when punctionable lymph nodes are present) and/or CSF examination depending on the type of study.
- From the efficacy assessment standpoint, the lumbar puncture at EoT has only the objective to detect trypanosomes, because the WBC counts are not informative of the response to treatment at this time (Dumas 1978). There may be other study objectives, such as pharmacology, for planning a lumbar puncture.
- The EoT evaluation, including a CSF examination, is particularly necessary in phase II trials for proof-of-concept which are subject to futility analysis, whereas in phase III and IV trials, when usually there is preliminary reassuring efficacy information on the experimental therapy, the EoT CSF examination is of little interest because finding trypanosomes is highly unlikely, and the patient can be spared the inconvenience of a lumbar puncture.
- In cases where preliminary information calls for more caution, a study protocol may plan a stepwise approach, where an initial number of patients enrolled have an EoT CSF examination, which subsequently can be phased out if results are reassuring.

3.2.1 Evaluation of first-stage patients at the end of treatment

At the EoT evaluation of first-stage patients, CSF examination is not regarded as necessary, except for non-responders (parasites in the blood and/or lymph) for whom it is necessary to determine the stage of the disease and thus the appropriate rescue treatment.

Depending on the survival of the patient during treatment and the outcome of the EoT evaluation, as applicable, patients will be classified as summarized in Table 2.

Table 2. Criteria for classification of first-stage patients at the EoT evaluation

Category	Patient characteristics/Criteria	Action
Death	Within 30 days of treatment start: temporally related to treatment	Classify according to 3.4.2. Record as serious adverse event
Non-responders	Evidence of trypanosomes in blood and/or lymph	Rescue treatment according to disease stage determined via LP
Responders	No evidence of trypanosomes in blood and/or lymph	Scheduled for the next follow-up visit per protocol
Uncertain responders	No evidence of trypanosomes in blood and/or lymph but persistence of clinical signs and symptoms	Perform LP: if trypanosome seen or > 20 WBC, class as non-responder; otherwise regard as responder.
Unknown response	No EoT evaluation data are available	All attempts should be made to evaluate these patients as soon as possible

EoT: end-of-treatment; LP: lumbar puncture

3.2.2 Evaluation of second-stage patients at the end of treatment

At the EoT evaluation, the protocol should plan for an assessment of the presence of trypanosomes in the blood and lymph nodes (when feasible). Lumbar puncture for CSF examination should be planned in phase II trials or when preliminary efficacy information is not sufficiently reassuring (see above). Because CSF WBC counts may not have normalized at that time, they should not be taken into account for classification as responder or non-responder (Dumas 1978). Depending on the survival of the patient during treatment and the outcome of the EoT evaluation, as applicable, patients will be classified as shown in Table 3.

Table 3. Criteria for classification of second-stage patients at the EoT evaluation

Category	Patient characteristics/Criteria	Action
Death	Within 30 days of treatment start: temporally related to treatment	Classify according to 3.4.2. Record as serious adverse event
Non-responders	Evidence of trypanosomes in any body fluid	Rescue treatment as per study protocol
Responders	No evidence of trypanosomes in any body fluid	Scheduled for the next follow-up visit
Unknown response	No EoT evaluation data are available	All attempts should be made to evaluate these patients as soon as possible

EoT: end-of-treatment

3.2.3 Evaluation of intermediate-stage patients at the end of treatment

The evaluation criteria are the same as in second-stage patients.

3.3 Follow-up of patients for efficacy assessment

- Parasitological examinations on blood and lymph (when punctionable lymph nodes are present) *and* CSF examination should be performed at each follow-up visit after the EoT evaluation.
- The full spectrum of parasitological tests, including the most sensitive tests available (mAECT-BC or mAECT, or exceptionally mHCT if mAECT is not available) for blood, and modified single centrifugation for CSF, should be used to increase the sensitivity of detection of relapse (Miézan 2000, Mumba Ngoyi 2013). Provisions for confirmation of trypanosome presence by a second staff member should be included in the protocol to reduce the risk of false-positive results.
- Up to now, there is no marker in blood that is useful for follow-up.
- Molecular tests are not useful for efficacy assessment, since DNA of the parasite can be found in the CSF of 20% of cured patients after 24 months (Deborggraeve 2011a).
- Given the relatively low HAT prevalence in most endemic areas, the probability that patients diagnosed as relapses are actually re-infections is likely to be small. Until validated field-usable methods allow an unambiguous distinction between re-infection and relapse, all patients who are trypanosome-positive during follow-up will be classed as relapses.

3.3.1 Time-point for final assessment of efficacy (ToC visit)

Systematic follow-up after treatment is no longer recommended in HAT control programmes, essentially by virtue of the confirmed high effectiveness of current first-line treatment for both first- and second-stage HAT. At the end of treatment, patients are encouraged to present themselves when clinical symptoms suggestive of HAT do appear. Follow-up, including CSF examination, focuses on symptomatic patients. This is based on the recommendations of the 2013 meeting of the WHO Expert Committee on Control and Surveillance of Human African Trypanosomiasis.

These recommendations apply only for routine treatment with well-established therapies and not for clinical trials of new drugs or new treatment regimens.

Requirements for the assessment of efficacy in clinical trials are different from those for efficacy assessment in HAT control programmes. The WHO expert consultation of 2004 had examined post-therapeutic data available at that time, and had concluded that a follow-up of 18 months was sufficient for comparative evaluations of efficacy in randomized trials for new drugs. The observations based on the data considered then (WHO 2004) can be summarized as follows:

- quantification of efficacy 24 months after treatment is not as robust as would be desirable, since at that time data are missing for 20–80% of patients;
- at least 70% of relapses diagnosed by 24 and 36 months after treatment have already occurred within 18 months;
- 40–100% of relapses were detected within 12 months after treatment.

It was noted that the variability in the proportion of patients diagnosed as relapsed at certain time-points is likely related to a number of factors including, but not limited to:

- The drugs and drug regimens used; for example, relapses after eflornithine treatment tend to occur later than after melarsoprol (Pépin 2000, Priotto 2008);
- The criteria for relapse used by different control programmes and in different studies;
- The number of follow-up investigations per patient and the percentage of patients who have follow-up examinations at different time-points after treatment;
- The frequency and timing at which the follow-up examinations took place, i.e. at a sufficient number of time-points to allow detection of relapse shortly after it becomes detectable;
- Biological factors (e.g. differences in inclusion criteria, disease status at baseline);
- Technical aspects of follow-up (methods employed, experience of investigators and technicians).

The data considered for the 2004 expert consultation were predominantly composed of studies and field programs that had made little investment in follow-up. These data show a follow-up attendance that drops sharply after 6 months and, as expected, less relapses are detected at each consecutive time-point (Schmid 2004, Schmid 2005, Balasegaram 2006, Cross 2006, Balasegaram 2008, WHO 2007).

In the past decade, new data became available from clinical trials which invested more effort in ensuring a good follow-up (Priotto 2006, Bisser 2007, Priotto 2009, NECT-Field publication in preparation). These studies, which were not included in the 2004 WHO expert consultation, reveal a different picture:

- Post-treatment follow-up data were available for 93% of patients at 18 months and for 71–92% at 24 months.
- Only half of the relapses had been detected by 12 months, and the other half had been diagnosed at 18 months or later (in studies that planned 24 months follow-up).

These new data confirm the feasibility and the importance of good follow-up of HAT clinical trial participants for 18 months.

For a timely diagnosis of relapse, patients should have follow-up evaluations at 3 (depending on preliminary data), 6, 12 and 18 months after treatment. The choice of 3 months follow-up (where the objective is to ascertain the presence of trypanosomes only, as WBC counts have an uncertain interpretation at that time) will depend on preliminary knowledge on the efficacy of the therapy under evaluation. To ensure data comparability between trials, clinical studies requiring assessments at other time-points should add these time-points, rather than replace the standard evaluations at 3, 6, 12 and 18 months after treatment.

To ensure that a majority of patients have follow-up data acquired 18 months after treatment, it is suggested that the study protocol contains plans to:

- For several months after the 18 months' time-point, continue the efforts to localize and evaluate patients who did not present for the 18 months assessment.
- Ensure that patients and their family retain a positive souvenir of their stay in the hospital, by providing basic comfort such as hygiene, safe water and food, and making sure that all staff involved remain sympathetic and helpful at all times.
- Ensure that patients and their family are educated about the disease, its treatment and the importance of completing the post-therapeutic follow-up, particularly in clinical trials.
- Collect detailed data about the location of the patients and their social circle, to allow effective tracing of those not attending the follow-up appointments.
- Involve the community leaders of the patient's village.
- Send reminder messages to the patients and to the community leaders via the local means (letters transported by community members, SMS by mobile phone, radio).
- Profit from HAT mobile units that may be able to perform the examinations at the patient's village or nearby.
- Have the study staff visit personally the village of patients who fail to attend despite all of the above measures, and approach not only the patient but also the family and community leaders.

One motion in the meeting proposed that the time-point for final assessment of efficacy in clinical trials be changed to 12 months. However, after discussion and consideration of the available data, a consensus was not reached on this point. It was agreed that this concept be reviewed again by experts in the light of an ad-hoc data analysis to be carried out later.

3.3.2 Early prediction of efficacy

Although the efficacy of HAT therapy cannot be appropriately assessed without a minimum of 18 months follow-up, the predictive value of the CSF WBC count at 6 and 12 months has been described (Mumba Ngoyi 2010; Priotto 2012), and may allow for some decision-making regarding a clinical trial before the full follow-up data have been obtained. This would typically be the role of a Data Safety Monitoring Board (DSMB).

However, it must be kept in mind that these predictive value studies used either small prospective datasets with partly unconfirmed outcomes, or larger retrospective datasets selected with stricter criteria, but originating from routine field programmes, which are generally of lower quality than datasets from prospective studies. Also, both datasets had

a heavy component of melarsoprol-treated patients, in areas with high melarsoprol-failure rates. Patients relapsing after being treated with melarsoprol showed higher WBC counts at 6 and 12 months before being diagnosed with relapse, which allowed a differentiation of them (with some degree of error) from patients who were eventually cured. Of note also is that relapsing patients who had received combination treatments showed elevated CSF WBC counts only at 18 months; therefore at 12 months they were still indistinguishable from cured patients (Priotto 2012).

At 6 and 12 months the WBC counts may be useful in order to visualize early the poor efficacy of an experimental therapy, but less so a difference (or absence of a difference) with the active comparator, which nowadays is likely to be a drug combination therapy.

The predictive value of WBC counts at 6 and 12 months can help in deciding about:

- Early termination of inclusions into a trial, when the predicted relapses exceed a pre-specified threshold.
- Launch of a new trial, based on the anticipated results, therefore gaining some time in the larger process of therapeutic development.
- Early advice for first-tier regulatory authority processes designed towards simplified approval of drugs for neglected diseases (e.g. EMA Art. 58, US FDA orphan drugs, new Swissmedic).

For early prediction at 6 months, a cell count cut-off can be chosen to privilege either the sensitivity or the specificity of the prediction (Figure 2). For example, if the objective is to make a conservative estimation of the proportion of patients that may relapse, then patients with > 5 WBC/µl are classified as failure, with 84% sensitivity (of all true relapses, 84% will be identified), and 60% specificity (of all truly cured, 60% will be identified).

Figure 2: Performance of CSF WBC count at 6 months for the prediction of relapse

Cut-off of CSF leucocytes count at 6 months	Sensitivity	Specificity	Correctly classified	Likelihood Ratio	
				Positive	Negative
>5	83.98%	60.21%	64.93%	2.11	0.27
>10	76.24%	80.41%	79.58%	3.89	0.29
>20	59.67%	93.49%	86.77%	9.17	0.43

CSF: cerebrospinal fluid; WBC: white blood cells

(Source: Priotto 2012)

With the two-step algorithms (Figure 3) a prediction can be made using the CSF WBC counts at 6 and 12 months. The expression "5-50-20" means: at 6 months ≤ 5 WBC/mL equals cure and ≥ 50 WBC/mL equals relapse; at 12 months a single cut-off distinguishes cure (≤ 20 WBC/mL) from relapse.

Again, the choice of algorithm depends on the objectives of such prediction, to privilege either sensitivity or specificity. For example, the 5-50-20 algorithm will identify 87% of all true relapses, while the 5-20-15 algorithm will identify 90%, and so on.

Figure 3. Comparison of several two-step (6 and 12 months) algorithms for the prediction of relapse

Algorithm	n	Sensitivity	(95%CI)	Specificity	(95%CI)	LR+	(95%CI)	LR−	(95%CI)	False cured	(95%CI)	% classified at 6 months
5-50-20[a]	213	94.4	(86–98)	97.8	(94–100)	42.20	(13.8–129.3)	0.06	(0.02–0.15)	1.9	(0.5–4.9)	66.2
5-50-20[b]	2190	87.4	(85–90)	97.7	(97–98)	37.84	(26.4–54.3)	0.13	(0.11–0.16)	4.5	(3.6–5.5)	66.4
5-20-20	2190	89.4	(87–92)	92.0	(90–93)	11.12	(9.2–13.4)	0.12	(0.09–0.14)	3.8	(3.0–4.7)	74.1
5-20-15	2190	90.0	(88–92)	91.7	(90–93)	10.87	(9.0–13.1)	0.11	(0.09–0.14)	3.6	(2.8–4.5)	74.1
5-20-10	2190	90.1	(88–92)	90.8	(89–92)	9.78	(8.2–11.6)	0.11	(0.09–0.14)	3.5	(2.8–4.4)	74.1
5-30-15	2190	89.4	(87–92)	95.4	(94–97)	19.52	(15.2–25.1)	0.11	(0.09–0.14)	3.8	(3.0–4.7)	70.1
5-40-20	2190	87.7	(85–90)	97.1	(96–98)	29.90	(21.7–41.1)	0.13	(0.10–0.15)	4.4	(3.5–5.4)	67.8
5-40-15	2190	88.7	(86–91)	96.8	(96–98)	27.29	(20.2–36.9)	0.12	(0.10–0.14)	4.0	(3.2–5.0)	67.8
5-40-10	2190	89.5	(87–92)	95.2	(94–96)	18.80	(14.7–24.1)	0.11	(0.09–0.12)	3.7	(2.9–4.7)	67.8

[a]As reported by Mumba 2010, "algorithm C": includes deaths as treatment failures and patients with incomplete follow-up;
[b]Same algorithm tested with our dataset including only laboratory-confirmed outcomes. LR: Likelihood ratio. False cured: fraction of patients that are wrongly classified as cured by the algorithm.

(Sources: Mumba 2010; Priotto 2012)

Given these predictive limitations, and with the current therapeutic arsenal, final conclusions about efficacy cannot be reached with 12 months of follow-up, and clinical trials should complete the 18 months of follow-up of their cohort in order to reach conclusive results of efficacy assessment.

The application of the CSF WBC counts for early decision-making is exemplified below in Tables 4 and 5.

Table 4. Early prediction of therapeutic efficacy using 6 months of follow-up (example of 10 cells cut-off)

Category	Patients with 6 months follow-up data	Analysed as:
Death	See section 3.4.2 for details	Failure or excluded
Relapse	Trypanosomes detected in any body fluid.	Failure
Predicted relapse	> 10 WBC/µl CSF	Failure
Predicted cure	≤ 10 WBC/µl CSF	Success
Unpredictable	No CSF data (no LP or haemorrhagic CSF)	Excluded

The performance values for this prediction criteria are shown in Figure 2.

Table 5. Early prediction of therapeutic efficacy at 6 and 12 months of follow-up (example of algorithm "5-50-20")

Category	Patients with 6 months follow-up data	Analysed as:
Death	See section 3.4.2 for details	Failure or excluded
Relapse	Trypanosomes detected in any body fluid	Failure
Predicted relapse	> 50 WBC/µl CSF	Failure
Predicted cure	≤ 5 WBC/µl CSF	Success
Uncertain evolution	6-49 WBC/µl CSF	Excluded
Unpredictable	No CSF data (no LP or haemorrhagic CSF)	Excluded
Patients with 12 months follow-up data		
Death	See section 3.4.2 for details	Failure or excluded
Relapse	Trypanosomes detected in any body fluid.	Failure
Predicted relapse	> 20 WBC/µl CSF	Failure
Predicted cure	≤ 20 WBC/µl CSF	Success
Unpredictable	No CSF data (no LP or haemorrhagic CSF) and classed Uncertain or Unpredictable at 6 months	Excluded

The performance values for this prediction criteria are 87.4 sensitivity and 97.7 specificity as in Figure 3.

3.4 Interim follow-up and test-of-cure

3.4.1 Evaluation of patients during follow-up

To ensure timely diagnosis of relapse, interim follow-up evaluations, including parasitological examination of the blood, lymph (when feasible) and CSF, should be performed at 3 or 6 months (depending on preliminary knowledge on the efficacy of the therapy under evaluation) and at 12 months. Additional follow-up visits should be scheduled if indicated clinically or when the patient is classified as having "uncertain evolution".

At the ToC visit 18 months after treatment, all parasitological examinations should be performed as well.

The agreed patient classification is shown in Tables 6 and 7.

This classification rules leave some of the criteria for diagnosing patients with "probable relapse" (resulting in rescue treatment) or "uncertain evolution" (resulting in close follow-up) open to the "opinion of the investigator". However, it may be possible, depending on the study, to introduce more objective criteria elements in the study protocol.

3.4.2 Deaths

As patients may die during treatment, shortly after completing treatment or after early treatment termination due to adverse events, in the hospital or at home shortly after

discharge, there is need of a standard benchmark for computing "treatment-emergent" deaths in a way that data are comparable between treatment arms in a given study and across different studies. The benchmark used in several HAT trials in the past (Burri 2000, Priotto 2006, Checchi 2007, Bisser 2007, Priotto 2009) was 30 days after treatment start. Deaths occurring within that period (or resulting from an event that started within that period) were regarded as temporally related to treatment.

It is proposed therefore that clinical trials compute deaths occurring within 30 days after the start of treatment (or resulting from an event that started within that period) as temporally related to treatment, or treatment-emergent.

Regarding causality, patients who died during treatment or follow-up should be categorized based on the likely or definite cause of death:

- Human African trypanosomiasis;
- Adverse events classified as possibly, probably or definitely related to HAT treatment;
- Causes unrelated to HAT or treatment for HAT;
- Unknown or medically unclear causes.

Before classification of a patient as having died due to unknown causes, attempts should be made to determine the cause of death. Interviews of family members and neighbours or local health-care workers (sometimes referred to as "oral autopsy") may provide the likely cause of death. "Suggestive questioning" should be avoided. Standardization of techniques for oral autopsy between clinical trials should be discussed.

Table 6. Criteria for classification of first-stage HAT patients during follow-up

Category	Interim follow-up visits	Action	Analysis
Death	Treatment-emergent: within 30 days after the start of treatment (or resulting from an event that started within that period). During follow-up: clinical picture compatible with HAT, OR unclear.	Death report, based on medical staff or family.	Failure
Death	During follow-up: clearly other causes than HAT, AND last categorized as "favourable evolution"	Death report, based on medical staff or family.	Success
Relapse	Trypanosomes detected in any body fluid.	Rescue treatment as per study protocol.	Failure
Probable relapse	Trypanosome-negative, > 20 WBC/µl CSF not haemorrhagic, unlikely due to other causes. Patient who refuses lumbar puncture OR whose CSF sample is haemorrhagic AND who in the opinion of the investigator requires immediate rescue treatment based on a marked deterioration in clinical condition unlikely due to causes other than HAT.	Rescue treatment as per study protocol. Attempt to convince the patient to undergo LP prior to classification as probable relapse.	Failure
Uncertain evolution	Trypanosome-negative, 6–20 WBC/µl, non-haemorrhagic CSF sample Patient who refuses lumbar puncture OR has haemorrhagic CSF sample AND no marked clinical deterioration compared to the previous evaluation AND who in the opinion of the investigator should have an additional close follow-up examination.	Additional follow-up after 1–3 months with blood, lymph and CSF exams and clinical evaluation.	Carried forward as failure
Favourable evolution	Trypanosome-negative patient whose CSF sample is not haemorrhagic with ≤ 5 WBC/µl	Scheduled for the next follow-up visit as per protocol.	Carried forward as success
	ToC visit at 18 months		
Death	See above (see also section 3.4.2)		See above
Relapse	Trypanosomes detected in any body fluid.	Rescue treatment as per study protocol.	Failure
Probable relapse	Trypanosome-negative, > 20 WBC/µl CSF not haemorrhagic, unlikely due to causes other than HAT. Patient who refuses lumbar puncture OR has haemorrhagic CSF sample AND who in the opinion of the investigator requires rescue treatment because of a marked clinical deterioration that is unlikely due to another disease than HAT.	Rescue treatment as per study protocol. Attempt to convince the patient to undergo LP prior to classification as probable relapse.	Failure
Probable cure	Trypanosome-negative patients with 6–20 WBC/µl CSF in a non-haemorrhagic sample. Patient who refuses lumbar puncture OR whose CSF is haemorrhagic AND who in the opinion of the investigator does not require rescue treatment because of satisfactory clinical condition or symptoms attributed to a disease other than HAT.	The routine follow-up at 24 months (not required for clinical trial data purposes) may be done at the clinical trial centre or at the nearest national control programme facility.	Success
Cure	Trypanosome-negative patients whose CSF sample is not haemorrhagic with ≤ 5 WBC/µl.		Success

Table 7: Criteria for classification of second-stage HAT patients during follow-up

Category	Interim follow-up visits	Action	Analysis
Death	Treatment-emergent: within 30 days after the start of treatment (or resulting from an event that started within that period). During follow-up: clinical picture compatible with HAT, OR unclear.	Death report, based on medical staff or family.	Failure
Death	During follow-up: clearly other causes than HAT, AND last categorized as "favourable evolution"	Death report, based on medical staff or family.	Success
Relapse	Trypanosomes detected in any body fluid.	Rescue treatment as per study protocol	Failure
Probable relapse	Trypanosome-negative patient who, in the opinion of the investigator, requires rescue treatment because of a CSF WBC count suggestive of relapse (e.g. > 50/µl) or presents a marked deterioration in clinical condition unlikely due to a disease other than HAT.	Rescue treatment as per study protocol Attempt to convince the patient to undergo LP prior to classification as probable relapse.	Failure
Uncertain evolution	Trypanosome-negative patient who, in the opinion of the investigator, requires a close follow-up examination because of a rising CSF WBC count (e.g. 21–50/µl) or a deterioration in clinical condition that might or might not be due to HAT.	Additional follow-up after 1–3 months with evaluation of blood, lymph and CSF	Carried forward as failure
Favourable evolution	Trypanosome-negative patient whose CSF sample is not haemorrhagic with ≤ 20 WBC/µl OR 21–50 WBC/µl CSF and decreased from previous values.	Scheduled for the next follow-up visit as per protocol	Carried forward as success
	ToC visit at 18 months		
Death	See above (see also section 3.4.2)		See above
Relapse	Trypanosomes detected in any body fluid.	Rescue treatment as per study protocol	Failure
Probable relapse	Trypanosome-negative, > 20 WBC/µl CSF not haemorrhagic, unlikely due to other causes. Trypanosome-negative patient who refuses lumbar puncture OR whose CSF is haemorrhagic AND who, in the opinion of the investigator, requires rescue treatment because of a marked deterioration in clinical condition unlikely due to a disease other than HAT.	Rescue treatment as per study protocol Attempt to convince the patient to undergo LP prior to classification as probable relapse.	Failure
Probable cure	Trypanosome-negative patient who refuses lumbar puncture OR whose CSF sample is haemorrhagic AND whose clinical condition is satisfactory OR whose clinical status is unlikely due to HAT.	The routine follow-up at 24 months (not required for clinical trial data purposes) may be done at the clinical trial centre or at the nearest national control programme facility	Success
Cure	Trypanosome-negative patient whose CSF sample is not haemorrhagic with ≤ 20 WBC/µl.		Success

CSF: cerebrospinal fluid; HAT: human African trypanosomiasis; ToC: test-of-cure; WBC: white blood cell; haemorrhagic CSF sample: >200 RBC/µl

4 Quantification of therapeutic efficacy

Ideally, all subjects enrolled in a clinical trial or treatment protocol/series should comply with all inclusion/exclusion criteria, be treated exactly as planned in the protocol and have a complete set of follow-up data. In practice, this is never the case, and thus the question arises as to which patients and which data should be included in a particular analysis. To ensure an unbiased analysis of the data, the analyses to be conducted should be clearly specified before initiation of the study, i.e. in the study protocol.

The major sources of potential bias identified by the meeting included criteria for:

- Inclusion or exclusion of patients in particular analyses;
- How to deal with missing follow-up or final efficacy evaluation data;
- How to include those patients in the analysis of efficacy at a specific time-point during the study whose follow-up data had not been obtained within the time periods planned in the protocol.

Analyses and reporting of efficacy according to a common nomenclature and criteria would facilitate the direct comparison of data reported for different clinical studies and treatment protocols/series.

The discussions and the resulting definitions and procedures for data analysis were based on the guidelines for statistical analysis issued by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH, www.ich.org).

4.1 Analysis populations

Definition of criteria for the inclusion/exclusion of patients for specific analyses, i.e. definition of "analysis sets" or "analysis populations", should be included in the protocol.

The nomenclature outlined in Annex 4 is commonly used and was proposed by the meeting.

Criteria for defining analysis sets for efficacy are provided in Table 8. Considering that in the first clinical trial of a new therapy, the percentage of subjects who were parasite-free at the EoT may be the primary efficacy variable, a "modified full analysis set" has been defined.

Table 8. Analysis sets for quantification of therapeutic efficacy in HAT patients

Analysis set	Patients included
Intention-to-treat set (Full analysis set)	All patients enrolled in the study who received at least one dose of study medication
Per-protocol set	All patients enrolled in compliance with the inclusion/exclusion criteria AND who received the protocol-defined medication OR for whom treatment was discontinued because of adverse events (including death) AND who had a ToC visit assessment or reached a protocol-defined end-point (death, non-responder, relapse, probable relapse) earlier.
Modified intention-to-treat set	All patients enrolled in the study who received at least one dose of study medication AND for whom efficacy data are available either as end-point (cure, probable cure, non-responder, relapse, probable relapse, death) or as partial follow-up (uncertain evolution, favourable evolution).
Safety-analysis set	All patients enrolled in the study who received at least one dose of study medication

ToC: test-of-cure; For trials with other efficacy end-points (e.g. trials evaluating a new drug or new drug combination for the first time), these criteria need to be adapted accordingly.

The efficacy indicators (variables) to be calculated for the analysis sets are described in Table 11.

The minimum amount of treatment required for inclusion in the per-protocol and the modified-intention-to-treat (m-ITT) population is drug-specific. Table 9 provides the minimum amount for currently available therapies used as comparators (to be specified in the protocol). The meeting stressed that the most effective regimens currently in use should be used as comparators for new therapies.

Table 9. Minimum amount of HAT treatment required per-protocol

Drug	Standard treatment regimen	Minimum amount of treatment for per-protocol population
Melarsoprol for *T. b. gambiense*	2.2 mg/kg per day IV x 10 days	8 days of treatment without interruption or up to one treatment interruption for ≤ 2 days
Eflornithine	400 mg/kg per day IV x 14 days	12 days of treatment without interruption or up to two treatment interruptions for up to two successive doses
NECT	Eflornithine 400 mg/kg/day IV every 12h x 7 days; nifurtimox 15 mg/kg/day orally every 8h x 10 days	13 doses of eflornithine. If interruptions last > 24 h, add doses for compensation depending on the moment of the interruption.
Pentamidine	4 mg/kg per day IM x 7 days	Up to one treatment interruption of ≤ 2 days

IM, intramuscular; IV, intravenous; NECT, nifurtimox–eflornithine combination therapy

4.2 Time windows for assigning patients' follow-up data to planned visits

As seen in the practice of HAT field trials, it is unlikely that all patients will undergo the post-therapeutic evaluations closely at the protocol-defined times.

To ensure an unbiased data analysis, tolerance time windows need to be defined (prior to the data becoming available for analysis) that determine into which planned visit the data from a given follow-up visit are to be slotted. These time windows need to cover the whole follow-up period in a contiguous way, such that the data from all subjects can be included in the analysis, no matter when the actual visit dates.

The meeting proposed the use of the time windows described in Table 10.

Table 10. Slotting of actual follow-up visit time-points for efficacy data analysis

Follow-up visit planned	Tolerance window
End of treatment	1–30 days after end of treatment[a]
3 months	2–4 months after end of treatment
6 months	5–9 months after end of treatment
12 months	10–16 months after end of treatment
18 months	17–21 months after end of treatment
24 months	≥ 22 months after end of treatment[b]

[a] In all cases, the time period begins on the first day of the first month and lasts until the last day of the last month; [b] In clinical trials planning 24 months of follow-up

For a clinical trial which includes patients who were not evaluated for ToC until well after 18 months, a differentiation between patients having their ToC visit at 18 months or at 24 months after treatment is in general not necessary.

4.3 Handling of missing efficacy data

Experience prior to 2004 shows that the number of patients with missing follow-up data for at least one time-point can be large, and that this number increases with time after treatment. However, in the latest HAT trials the follow-up data have been complete for most patients, owing to a more developed follow-up strategy including, but not limited to, active tracing.

To ensure comparability of data from trials conducted by different investigators/sponsors, the way in which patients with missing ToC data are included in the efficacy analysis should be uniform. Tables 6 and 7 summarize the proposed categorization of patients with or without a ToC evaluation and who have not reached a protocol-defined end-point (e.g. death, probable relapse, relapse) before the ToC visit.

In the per-protocol analysis, patients with incomplete follow-up are excluded. But in the intention-to-treat analysis (ITT) and the modified-ITT, some of these patients are included, who at their last control were categorized as "uncertain evolution" or "favourable evolution". In the absence of ToC data, for the ITT and mITT analyses their last assessment is carried forward as "failure" or "success", respectively.

4.4 Efficacy variables

4.4.1 Nomenclature

A uniform use of terms to refer to measures of therapeutic efficacy (as proposed in Table 11) in combination with uniform calculation of efficacy measures (Table 12) will facilitate comparison between data reported from different trials and treatment protocols.

Table 11. Nomenclature for different measures of therapeutic efficacy in HAT

Variable	Treatment effect referred to
Treatment fatality rate	Quantifies the therapeutic efficacy in terms of deaths attributable either to lack of curative effect or to toxicity. Subjects who died from unknown causes are included (as for treatment failure rate).
Treatment failure rate	Quantifies failure to cure the patient independent of whether this failure is attributable to lack of efficacy or to toxicity of the medication. Treatment failure should also include those patients who died from unknown causes (and for whom it is thus not certain that they did not die because of lack of efficacy or toxicity of the drug under evaluation) and those who discontinued treatment owing to toxicity of the drug.
Relapse rate	Quantifies the therapeutic efficacy via the number of patients who were alive and trypanosome-negative at the end of treatment, but were diagnosed as "relapse" or "probable relapse". Depending on the objectives of the analysis, the calculation of the overall relapse rate (i.e. including "relapse" and "probable relapse") may be complemented by the calculation of "parasitologically-confirmed relapse rate".
Cure rate	Quantifies the therapeutic efficacy via the number of patients who were classified as "cure" or "probable cure". Depending on the objectives of the analysis, the calculation of the overall cure rate (i.e. including "cure" and "probable cure") may be complemented by the calculation of "laboratory-confirmed cure rate" and "probable cure rate".
Response rate	Quantifies the therapeutic efficacy via the number of patients who were found to be parasite-free at the end of treatment. This measure is of particular importance for the initial studies of new treatment or new combinations of established treatments.

4.4.2 Calculation of efficacy

A series of efficacy variables (or indicators) to be calculated was defined on the basis of requirements for regulatory submissions for new drugs and comparative interpretation of results of different studies. Which of the efficacy variables defined below should be chosen as the primary efficacy variable will depend on the objectives of the study (e.g. pivotal versus dose-finding study).

The patients to be included in the nominator and denominator for the different measures of efficacy are provided in Table 12. Definitions of and criteria for the different response categories in the nominator are provided in section 3 and Figure 1.

Each of these variables can be calculated using different denominators, i.e. for different analysis populations as well as for different time-points after treatment (e.g. treatment

failure rate at EoT or at a specified time-point during interim follow-up), depending on the objectives of the analysis.

Table 12. Description of therapeutic efficacy variables in HAT

Variable	Nominator	Denominator
Treatment fatality rate	Sum of patients who, within 30 days of treatment start: ▪ died during treatment (likely) due to HAT ▪ died during treatment (likely) due to treatment-related AE ▪ died from unknown or medically unclear causes	Safety analysis set
Treatment failure rate	Sum of patients who: ▪ died during treatment (likely) due to HAT ▪ died during treatment (likely) due to treatment-related AE ▪ died during follow-up (likely) due to HAT ▪ died during follow-up (likely) due to treatment-related AE ▪ died during treatment or follow-up due to unknown or medically unclear causes ▪ were non-responders at EoT visit ▪ relapsed ▪ probably relapsed ▪ were discontinued from study treatment due to treatment-related AE (and not cured) ▪ have no follow-up data	ITT set Per-protocol set
Overall relapse rate	Sum of patients who: ▪ relapsed ▪ probably relapsed	ITT set Per-protocol set Modified ITT set
Parasitologically-confirmed relapse rate	Patients who: ▪ relapsed	ITT set Per-protocol set Modified ITT set
Overall cure rate	Sum of patients who: ▪ were cured ▪ were probably cured	ITT set Per-protocol set Modified ITT set
Parasitologically-confirmed cure rate	Sum of patients who: ▪ were cured	ITT set Per-protocol set Modified ITT set
Response rate	Patients: ▪ who were responders at EoT visit	Per-protocol set ITT set

AE: adverse events**;** EoT: end-of-treatment; HAT: human African trypanosomiasis; ITT; intention-to-treat;

5 References

Balasegaram M, Harris S, Checchi F, Hamel C, Karunakara U. (2006) Treatment outcomes and risk factors for relapse in patients with early-stage human African trypanosomiasis (HAT) in the Republic of the Congo. Bull World Health Organ;84(10):777–82.

Balasegaram M, Young H, Chappuis F, Priotto G, Raguenaud ME, Checchi F. (2008) Effectiveness of melarsoprol and eflornithine as first-line regimens for gambiense sleeping sickness in nine Médecins Sans Frontières programmes. Trans R Soc Trop Med Hyg;103(3):280–90. doi:10.1016/j.trstmh.2008.09.005.

Bisser S, N'Siesi FX, Lejon V, Preux PM, Van Nieuwenhove S, Miaka Mia Bilenge C et al. (2007) Equivalence trial of melarsoprol and nifurtimox monotherapy and combination therapy for the treatment of second-stage *Trypanosoma brucei gambiense* sleeping sickness. J Infect Dis;195(3):322–9.

Büscher P, Mumba Ngoyi D, Kaboré J, Lejon V, Robays J, Jamonneau V et al. (2009) Improved models of mini anion exchange centrifugation technique (mAECT) and modified single centrifugation (MSC) for sleeping sickness diagnosis and staging. PLoS Negl Trop Dis;3(11):e471. doi:10.1371/journal.pntd.0000471.

Büscher P, Gilleman Q, Lejon V. (2013) Novel rapid diagnostic test for sleeping sickness. New Engl J Med; 368:1069–70.

Büscher P, Mertens P, Leclipteux T, Gilleman Q, Jacquet D, Mumba-Ngoyi D et al. (2014) Sensitivity and specificity of HAT Sero-*K*-SeT, a rapid diagnostic test for serodiagnosis of sleeping sickness caused by *Trypanosoma brucei gambiense*: a case-control study. Lancet Glob Health;2:e359–63. doi.org/10.1016/S2214-109X(14)70203-7.

Burri C, Nkunku S, Merolle A, Smith T, Blum J, Brun R. (2000) Efficacy of new, concise schedule for melarsoprol in treatment of sleeping sickness caused by *Trypanosoma brucei gambiense*: a randomised trial. Lancet;355: 1419-25. doi.org/10.1016/S0140-6736(00)02141-3.

Camara M, Camara O, Ilboudo H, Sakande H, Kaboré J, N'Dri L et al. (2010) Sleeping sickness diagnosis: use of buffy coats improves the sensitivity of the mini anion exchange centrifugation test. Trop Med Int Health;15(7):796–9. doi:10.1111/j.1365-3156.2010.02546.

Checchi F, Chappuis F, Karunakara U, Priotto G, Chandramohan D. (2011) Accuracy of five algorithms to diagnose gambiense human African trypanosomiasis. PLoS Negl Trop Dis;5(7):e1233. doi:10.1371/journal.pntd.0001233

Cross P, Doua F, Jaffar S. (2006) The risk factors for relapse among patients with African trypanosomiasis in Daloa, Cote d'Ivoire. Trop Doct;36(2):90–3.

Davis A, Dobson R, Kaninia S, Espasandin M, Berg A, Giovannoni G, Schmierer K. (2014) Change practice now! Using atraumatic needles to prevent post lumbar puncture headache. Eur J Neurol;21(2):305–11. doi:10.1111/ene.12307.

Deborggraeve S, Lejon V, Ali Ekangu R et al. (2011a) Diagnostic accuracy of PCR in gambiense sleeping sickness diagnosis, staging and post-treatment follow-up: a 2-year longitudinal study. PLoS Negl Trop Dis; 5(2):e972. doi:10.1371/journal.pntd.0000972.

Deborggraeve S, Lejon V, Ali Ekangu R, Ngoyi DM, Pyana P, Mulunda JP et al. (2011b) How reliable is PCR for diagnosis, staging and follow-up of gambiense sleeping sickness? Trop Med Int Health;16(S1):168.

Deisenhammer F, Bartos A, Egg R, Gilhus NE, Giovannoni G, Rauer S et al. (2006) Guidelines on routine cerebrospinal fluid analysis. Report from an EFNS task force. Eur J Neurol;13(9):913–22.

Dumas M, Girard PL. Human African trypanosomiasis (sleeping sickness). (1978) In: Vinken PJ, Bruyn GW, eds. Handbook of clinical neurology. Vol 35. Infections of the nervous system. Part III. Amsterdam; North-Holland Publishing Company:67–83.

Lejon V, Büscher P. (2013) Stage determination and follow-up. In: Cattand P, Louis FJ, Simarro PP, eds. Sleeping Sickness Lectures. Gémenos; Association contre la Trypanosomiase en Afrique:215–234.

Lejon V, Mumba Ngoyi D, Boelaert M, Büscher P. (2010) A CATT negative result after treatment for human African trypanosomiasis is no indication for cure. PLoS Negl Trop Dis;4:e590.

Lejon V, Roger I, Mumba Ngoyi D, Menten J, Robays J, N'siesi FX et al. (2008) Novel markers for treatment outcome in late stage *Trypanosma brucei gambiense* trypanosomiasis. Clin Infect Dis;47:15–22.

Lumbala, C, Bisser S, Nguertoum E, Flevaud L, Jacquet, D, Büscher P et al. (2013) Development and evaluation of a rapid screening test for sleeping sickness. Annales Africaines de Médecine;6(Suppl. 1):49.

Lutje V, Seixas J, Kennedy A. (2013) Chemotherapy for second-stage human African trypanosomiasis. Cochrane Database Syst Rev;6:CD006201. doi:10.1002/14651858.CD006201.pub3.

Miézan TW, Meda AH, Doua F, Djè NN, Lejon V, Büscher P. (2000) Single centrifugation of cerebrospinal fluid in a sealed pasteur pipette for simple, rapid and sensitive detection of trypanosomes. Trans R Soc Trop Med Hyg;94(3):293.

Mitashi P, Hasker E, Mumba Ngoyi D, Pyana PP, Lejon V, Van der Veken W et al. (2013) Diagnostic accuracy of Loopamp *Trypanosoma brucei* detection kit for diagnosis of human African trypanosomiasis in clinical samples. PLoS Negl Trop Dis;7:e2504.

Mumba Ngoyi D, Lejon V, Pyana P, Boelaert M, Ilunga M, Menten J et al. (2010) How to shorten patient follow-up after treatment for Trypanosoma brucei gambiense sleeping sickness? J Infect Dis;201(3):453–63. doi:10.1086/649917.

Mumba Ngoyi D, Menten J, Pyana PP, Büscher P, Lejon V. (2013a) Stage determination in sleeping sickness: comparison of two cell counting and two parasite detection techniques. Annales Africaines de Médecine; 6(Suppl. 1):79.

Mumba Ngoyi D, Menten J, Pyana PP, Büscher P, Lejon V. (2013b) Stage determination in sleeping sickness: comparison of two cell counting and two parasite detection techniques. Trop Med Int Health;18:778–82. doi:10.1111/tmi.12102.

Mumba Ngoyi, D, Ali Ekangu R, Mumvemba Kodi MF, Pyana PP, Balharbi F, Decq M et al. (2014) Performance of parasitological and molecular techniques for the diagnosis and surveillance of gambiense sleeping sickness. PLoS Negl Trop Dis;8(6):e2954. doi:10.1371/journal.pntd.0002954.

Pépin J, Khonde N, Maiso F, Doua F, Jaffar S, Ngampo S et al. (2000) Short-course eflornithine in Gambian trypanosomiasis: a multicenter randomized controlled trial. Bull World Health Organ;78(11):1284–95.

Priotto G, Chappuis F, Bastard M, Flevaud L, Etard J-F. (2012) Early prediction of treatment efficacy in second-stage gambiense human African trypanosomiasis. PLoS Negl Trop Dis;6(6):e1662.

Priotto G, Kasparian S, Ngouama D, Ghorashian S, Arnold U, Ghabri S et al. (2007) Nifurtimox-eflornithine combination therapy for second-stage *Trypanosoma brucei gambiense* sleeping sickness: a randomized clinical trial in Congo. Clin Infect Dis; 45(11):1435–42.

Priotto G, Kasparian S, Mutombo W, Ngouama D, Ghorashian S, Arnold U et al. (2009) Nifurtimox-eflornithine combination therapy for second-stage African *Trypanosoma brucei gambiense* trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial. Lancet;374(9683):56–64. doi:10.1016/S0140-6736(09)61117-X.

Priotto G, Pinoges L, Fursa IB, Burke B, Nicolay N, Grillet G et al. (2008) Safety and effectiveness of first line eflornithine for *Trypanosoma brucei gambiense* sleeping sickness in Sudan: cohort study. BMJ;336(7646):705–8. doi:10.1136/bmj.39485.592674.BE.

Richman JM, Joe EM, Cohen SR, Rowlingson AJ, Michaels RK, Jeffries MA et al. (2006) Bevel direction and postdural puncture headache: a meta-analysis. Neurologist; 12(4):224–8.

Schmid C, Nkunku S, Merolle A, Vounatsou P, Burri C. (2004) Efficacy of 10-day melarsoprol schedule 2 years after treatment for late-stage gambiense sleeping sickness. Lancet;364(9436):789–90.

Schmid C, Richer M, Bilenge CM, Josenando T, Chappuis F, Manthelot CR et al. (2005) Effectiveness of a 10-day melarsoprol schedule for the treatment of late-stage human African trypanosomiasis: confirmation from a multinational study (IMPAMEL II). J Infect Dis;191(11):1922–31.

Strupp M, Brandt T. (1997) Should one reinsert the stylet during lumbar puncture? N Engl J Med; 336:1190–8.

SD Bioline HAT [test insert LEF53EN / Rev. 1310-00]. World's first rapid diagnostic test for African trypanosomiasis
(https://ensur.invmed.com/ensur/broker/ensurbroker.aspx?code=LEF53EN&cs=26780809).

Tiberti N, Hainard A, Lejon V, Courtioux B, Matovu E, Enyaru JC et al. (2012) Cerebrospinal fluid neopterin as marker of the meningo-encephalitic stage of *Trypanosoma brucei gambiense* sleeping sickness. PLoS ONE;7(7):e40909. doi:10.1371/journal.pone.0040909.

Tiberti N, Lejon V, Hainard A et al. (2013) Neopterin is a cerebrospinal fluid marker for treatment outcome evaluation in patients affected by *Trypanosoma brucei gambiense* sleeping sickness. PLoS Negl Trop Dis;7(2):e2088. doi:10.1371/journal.pntd.0002088.

WHO. (2007) Recommendations of the Informal Consultation on issues for clinical product development for human African trypanosomiasis. Geneva, Switzerland, 9–10 September 2004. Geneva: World Health Organization (http://whqlibdoc.who.int/hq/2007/WHO_CDS_NTD_IDM_2007.1_eng.pdf?ua=1).

WHO. (2013) Control and surveillance of human African trypanosomiasis. Report of a WHO Expert Committee. Geneva: World Health Organization (WHO Technical Report Series 98; http://apps.who.int/iris/bitstream/10665/95732/1/9789241209847_eng.pdf).

Woo PTK (1971). Evaluation of the haematocrit centrifuge and other techniques for the field diagnosis of human trypanosomiasis and filariasis. Acta Trop;28(3):298–303.

6 Annexes

Annex 1. Agenda of the meeting

Time	Topic	Detail	Presenter
09:00–09:15	**Welcome**	- Addresses by WHO	ADG/HTM Director DPC/AFRO Director NTD/HQ
09:15–09:30	**Introduction**	- Presentation of the meeting - Why to update the methodological framework for HAT clinical trials	Coordinator IDM/HQ P.P. Simarro
09:30–10:30	**Identification of patients for clinical trials (1)**	- Diagnosis and staging	All
10:30–10:45	**Coffee break**		
10:45–11:15	**Identification of patients for clinical trials (2)**	- Enrolment criteria by therapeutic target	All
11:15–12:30	**Efficacy assessment**	- General considerations and terminology - End-of-treatment evaluation	All
12:30–13:30	**Lunch**		
13:30–14:30	**Follow-up for efficacy assessment (1)**	- Time-point for final assessment of efficacy (ToC) - Early prediction of efficacy	All
14:30–15:30	**Follow-up for efficacy assessment (2)**	- Interim follow-up and test-of-cure	All
15:30–15:45	**Coffee break**		
15:45–17:00	**Quantification of efficacy**	- Analysis populations - Handling of missing efficacy data - Calculation of efficacy	All
17:00–17:30	**Review of consensus points**	- Collaboration between clinical trial teams and national control programmes - Key text modifications	All

Annex 2. List of participants

Name	Institution	Country
Anne C. Moore (Chairperson)	Member of the WHO NTD Strategic and Technical Advisory Group Division of Parasitic Diseases and Malaria – Center for Global Health United States Centers for Disease Control and Prevention, Atlanta	USA
Antoine Tarral	Head of HAT Clinical Programme Drugs for Neglected Diseases initiative (DNDi), Geneva	Switzerland
Christian Burri	Head, Department of Medicines Research Swiss Tropical and Public Health Institute, Basel	Switzerland
David Wesche	Integrated Development – Global Health Programme Bill & Melinda Gates Foundation	USA
Joseph Ndung'u	Head, Neglected Tropical Diseases Programme – Foundation for Innovative New Diagnostics (FIND), Geneva	Switzerland
Justin Masimango Mbaruku	Deputy Director Programme National de Lutte contre la Trypanosomiase Humaine Africaine – Ministère de la Santé Publique	Democratic Republic of the Congo
Marleen Boelaert (remote partic.)	Head, Unit of Epidemiology and Control of Tropical Diseases – Institute of Tropical Medicine, Antwerp	Belgium
Nathalie Strub	Medical Director Drugs for Neglected Diseases initiative (DNDi) , Geneva	Switzerland
Philippe Büscher (remote partic.)	Head, Parasite Diagnostics Unit – Department of Parasitology – Institute of Tropical Medicine, Antwerp WHO Collaborating Centre for Research and Training on HAT diagnostics	Belgium
Sonja Bernhard	Pharmaceutical Medicine Unit Swiss Tropical and Public Health Institute, Basel	Switzerland
Susan Linna	Portfolio and Platform Lead Strategy, Planning & Management Global Health Program Bill & Melinda Gates Foundation	USA
Sylvain Biéler	Senior Scientific Officer Neglected Tropical Diseases Programme – Foundation for Innovative New Diagnostics (FIND), Geneva	Switzerland
Sylvie Bisser	INSERM 1094, Neuroépidémiologie tropicale, Limoges	France
Veerle Lejon	Director of Research Unité Mixte de Recherche UMR 177 – Intertryp Institut de Recherche pour le Développement, Montpellier	France
	WHO	
Annette Kuesel	Scientist – Intervention and Implementation Research (HQ/HTM/TDR/IIR)	
Gerardo Priotto	Medical Officer – Human African Trypanosomiasis Programme HQ/HTM/NTD/IDM	
Hiroki Nakatani	Assistant Director-General – HQ/HTM/HMA	
Jean Jannin	Coordinator – HQ/HTM/NTD/IDM	
Jose Ramon Franco Minguell	Medical Officer – Human African Trypanosomiasis Programme HQ/HTM/NTD/IDM	
Pere Perez Simarro	Medical Officer in-charge of Human African Trypanosomiasis Programme (HQ/HTM/NTD/IDM)	

Annex 3. Phase IV studies in HAT treatment development

Phase IV studies are an important phase of treatment development. The "real-world" effectiveness of a new therapy as evaluated in an observational, non-interventional trial in life-like settings will complement the efficacy data that arise from a randomized controlled trial.

Such effectiveness studies become crucial when there are important differences between the controlled environment of phase III trials and the "real-world" routine administration of the treatment. In the case of the new therapies currently under evaluation for HAT, these studies are particularly important, because the passage from in-hospital injectable administration to oral therapy implies a new risk of "real-world" effectiveness linked with possibly suboptimal adherence to the treatment schedule.

Phase IV studies will also allow for further characterizing the true safety profile of a drug.

For these type of studies on HAT therapies for which the risk/benefit ratio has been well-established, some of the methodological constraints can be relaxed.

In particular, for molecules that have been recognized as sufficiently effective in both stages of the disease in phase III studies and have a safety profile comparable to that of current first-stage treatments, it may be considered to enrol patients without performing a lumbar puncture, if trypanosomes have been seen in blood or lymph examinations. Follow-up criteria should be clarified in this case.

The omission of lumbar punctures may have an enormous impact in facilitating HAT diagnosis and treatment, and in community participation.

Annex 2. Nomenclature for analysis sets

Analysis set	Comments
Intention-to-treat set **(Full analysis set)**	The "intention-to-treat" or "full analysis set" includes, ideally, all subjects randomized in a clinical trial (or treated within a treatment protocol/case series). Exclusion of randomized subjects is acceptable only under circumstances where there is no danger of biasing the analysis outcome through this exclusion (e.g. the patient did not receive a single dose of treatment, or there are no follow-up data available that would allow the effect of the drug in that patient to be assessed). The intention-to-treat analysis is regarded as the best way to assess the effect of a treatment policy, i.e. it approximates the effect of a specific treatment policy in "real life". Consequently, for the full analysis set or intent-to-treat analysis, subjects are analysed as if they had received the treatment to which they were randomized, even if they received only a single dose of treatment, or actually received another treatment than the one to which they were randomized.
Per-protocol set	The "per-protocol set" ideally includes subjects who have been enrolled and treated as planned in the protocol and for whom final efficacy data are available, i.e. subjects who reached a protocol defined end-point (e.g. discontinuation of treatment owing to treatment-related adverse events, death, relapse) before the end of treatment or before the end of the protocol-defined follow-up period, and patients for whom efficacy data at the test-of-cure visit are available. The results of the per-protocol analysis will allow to draw conclusions as to the validity of the hypothesis underlying a clinical trial, i.e. that the drug under investigation, if administered as planned to a specific type of patient (defined via the inclusion and exclusion criteria in the protocol) will have a certain efficacy (e.g. percentage of subjects cured) specified in the protocol, usually in the sample size justification, with a level superior to the comparator drug, or non-inferior to the comparator drug.
Modified intention-to-treat set	In some cases a "modified intention-to-treat" set may be defined up-front, which also includes patients who have received only a pre-specified minimum amount of treatment (expected to have the minimum level of efficacy that the drug should have to perform its planned role in patient management/disease control) but not the full planned treatment, and/or patients who did not comply with specific requirements in the protocol (e.g. certain inclusion/exclusion criteria, concomitant treatment, concomitant diseases).
Safety analysis set	The "safety analysis set" includes all subjects who received at least one dose of the drug. This allows the safety profile of the drug to be fully characterized among the patients being treated, including adverse events which result in treatment discontinuation after only one or few doses (e.g. allergies, other symptoms of intolerance).


9 789241 508834