University of Zurich[UZH]

ECPM
European Center of Pharmaceutical Medicine

"Role of Clinical Assessment Technologies (CAT) in Developing New Medicines"

Developing New Medicines – An Introduction

MARCH 26, 2014

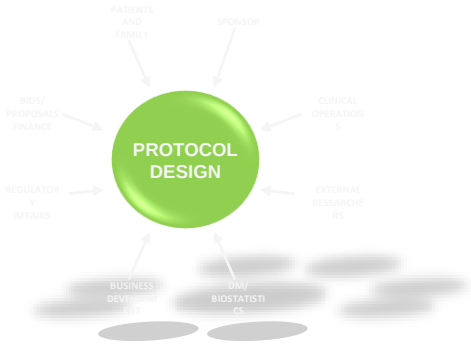
Zoran M Pavlovic MD
Psychiatrist
Director Scientific Affairs and Clinical Assessments

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

Running Successfully Clinical Trials

Feasibility

Protocol Design



Clinical Assessment Technologies

Medical Monitoring

Medical Writing

Clinical Assessment Technologies and Rater Training Goals

1) **Main goal of** CAT and Rater training is to **improve the efficacy of subjective outcome assessments**
2) **Experience of raters is assessed as part of the initial site evaluation** and is considered crucial to the selection of appropriate investigative sites
3) **Rater knowledge of study-specific scales is tested throughout the training program** to confirm that key concepts and study-specific rating conventions are understood
4) **In-study monitoring of diagnostic and assessment data** is a natural extension of the original rater training efforts to ensure that the appropriate patients are being enrolled in the study and that all raters continue to administer and score assessment measures in the same manner as well as insuring that protocol guidelines are maintained

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

CAT and Training of the Project Team

- Worldwide Clinical Trials (WCT) considers comprehensive training of the clinical project team to be critical to the successful execution of any clinical trial

- Together with members of the Medical and Scientific Affairs team, CAT members provide thorough training on the indication as well as the specific assessment scales

- Important issues are addressed throughout the training of sites and project team members, such as recruitment, retention, clinician/research subject relationship, and other factors related to placebo response and how to minimize their impact during the study

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

CAT Activities

- **Program Planning Activities**
- CAT will develop a study-specific rater experience survey
- The goal is to determine in advance which sites have appropriately trained staff for participation the study
- Only appropriately qualified raters will be eligible to participate in the rater training program

- **Scale Management Services**
- CAT will research and obtain any copyright permission and licensing agreements applicable, for all assessment measures
- CAT will also prepare rater workbooks, formatting rating scales and compiling each visit with instructions for the order of administration

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

CAT Activities

- **Clinical Assessment and Rater Training**

- There is a need to standardize the information disseminated to the raters relating to indication definition and assessment guidelines in order to reduce misinterpretation of diagnostic and/or assessment measures

- The focus of rater training is to assess experience, knowledge and applied skill in order to align raters within the study criteria Important patient related issues are addressed throughout the training, such as recruitment, retention, responses to questions about placebo and the clinician/patient relationship

- Potential raters must meet three criteria in order to be considered a certified rater:
  - Experience
  - Knowledge
  - Skill

- For all raters alignment adult education criteria of 80% is used as the pass rate, so that all raters need to have 80% or more assessments aligned to their peers and the gold standard consensus panel

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

CAT Activities

- **Investigators Meeting (IM) Services**
- It is important that training at the IM be interactive and thought-provoking
- CAT will develop study specific training on research alliance with attention to recruitment and retention of subjects while decreasing placebo response
- CAT will utilize audience response systems (ARS) to confirm understanding of the training and encourage discussion through multiple choice questions that are immediately scored and displayed after the voting
- This system may be utilized throughout the IM to confirm understanding of all presentations
- Video presentations with patients interviews are used as a part of the testing process

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

CAT Activities

- **Web-based Training of New Raters following IM**

- Not all raters from the sites will be able to attend the scheduled IM
- It is also expected that sites will experience some turnover in raters and need to have new raters trained during the study or will need to have back-up raters trained in order to manage multiple subjects at their sites
- In our experience, for studies of this type and duration, it is not uncommon for sites to have at least one two more raters trained during the course of the study, in addition to the raters originally trained at the IM
- CAT will develop a study-specific training and assessment website utilizing WCT's state-of-the-art web portal platform
- Identical training materials will be used for new raters as were used for training of raters at the IM
- Each new rater will be provided with a unique log-in and password combination after his/her experience is approved based on the completed rater experience survey
- Approximately two-weeks after the IM, the training will be available for any raters who did not attend the IM
- Didactic presentations and videos will be distributed and reviewed with a member of the clinical team after the potential new rater is approved

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

CAT Activities

- **Monitoring of clinical assessments**

- The monitoring of clinical assessments during the start-up phase of a clinical trial is very important to ensure that all raters are conducting neuropsychiatric measures in the same manner across sites as well as to optimize enrolment

- Worldwide Clinical Trials will monitor assessment of source documentation for the screen and baseline visit of the first enrolled patient at every site

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

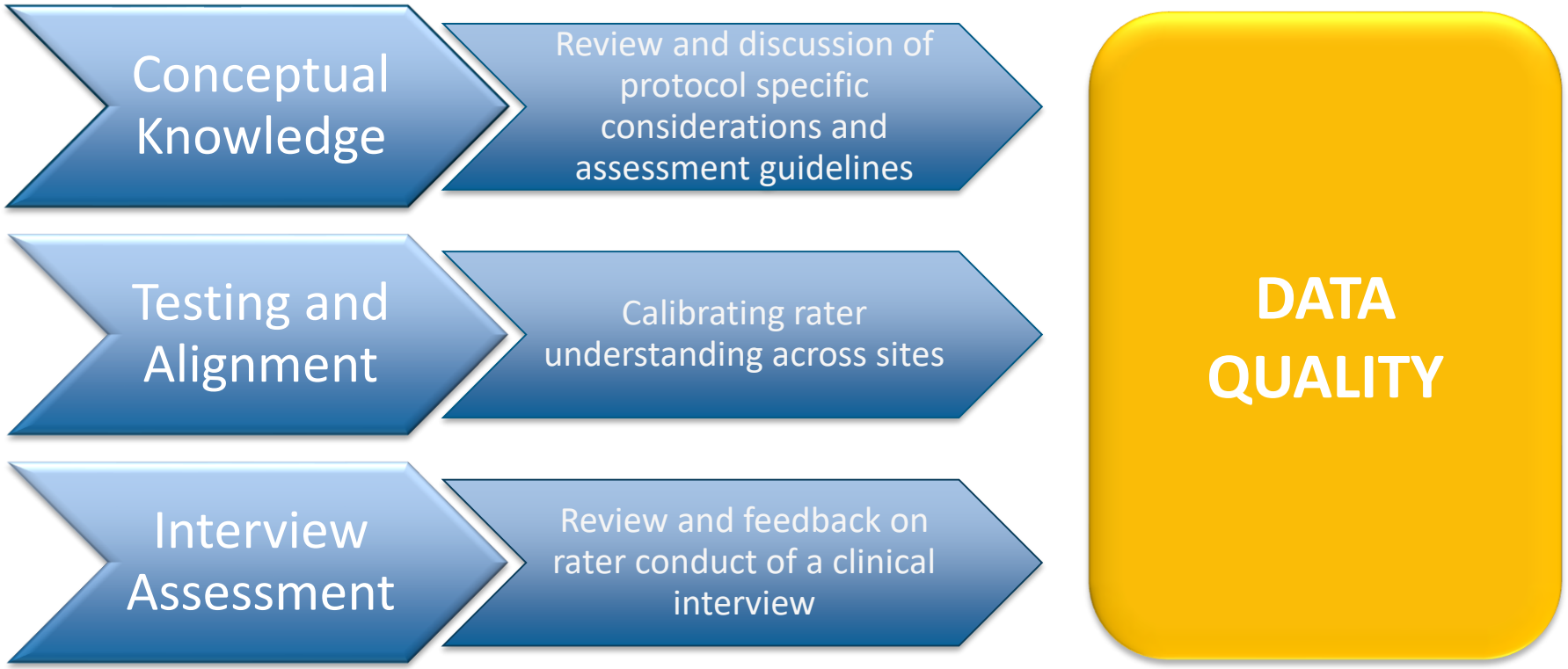
Clinical Assessment Technologies

- Site Vetting
  - PSSV – Assessment Experience
  - Ongoing Motivational Visits
- Training
  - Sponsor
  - CRA
  - Sites
    - Investigators' Meeting
    - Web –based
    - Interactive In-study
- Data Efficacy
  - Source Documents
    - Certified Translations
    - Assessment Workbooks
  - Monitoring Scoring In-Study

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

Clinical Assessment Technologies

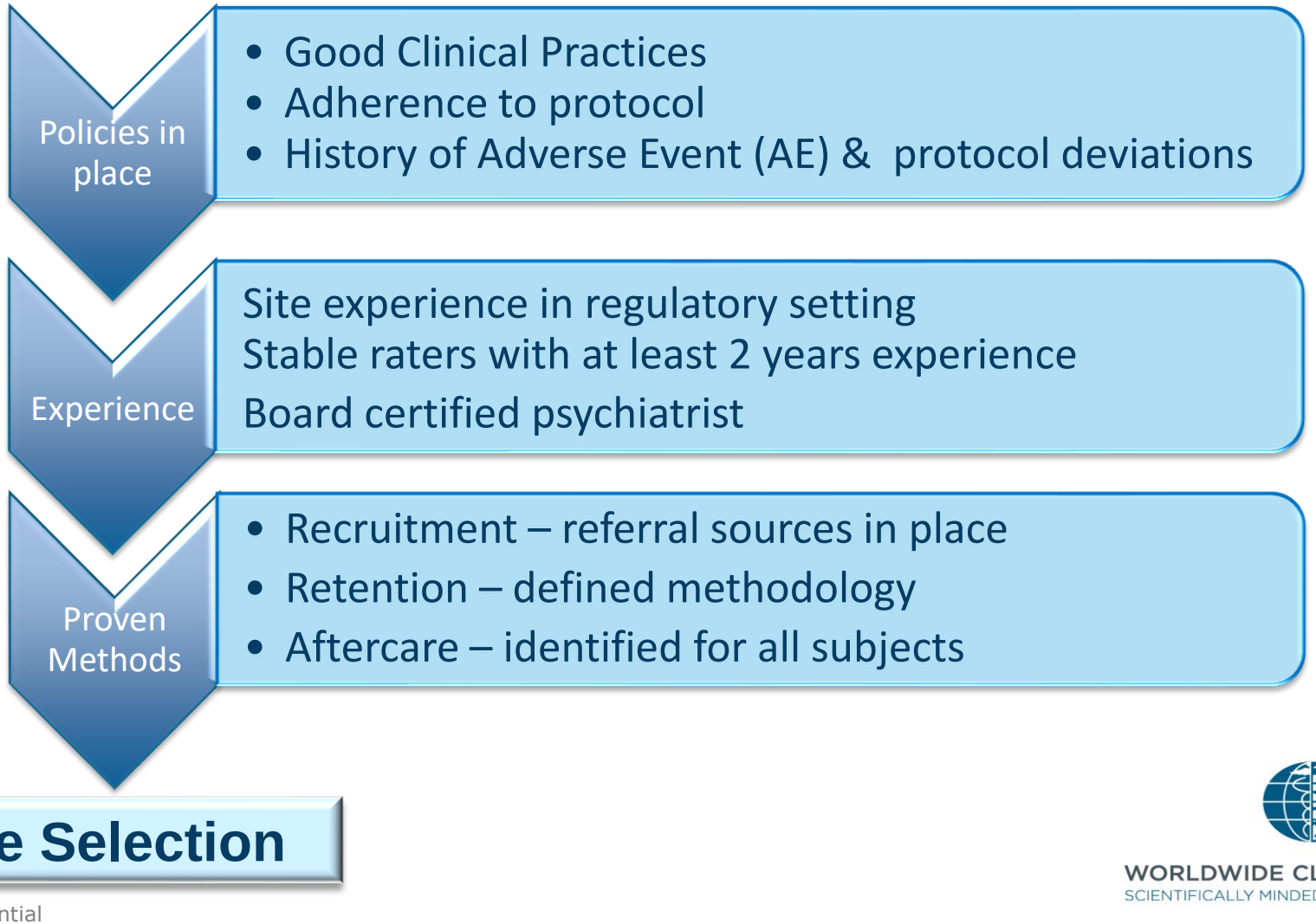
Rater Training and Data Quality Assurance



WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

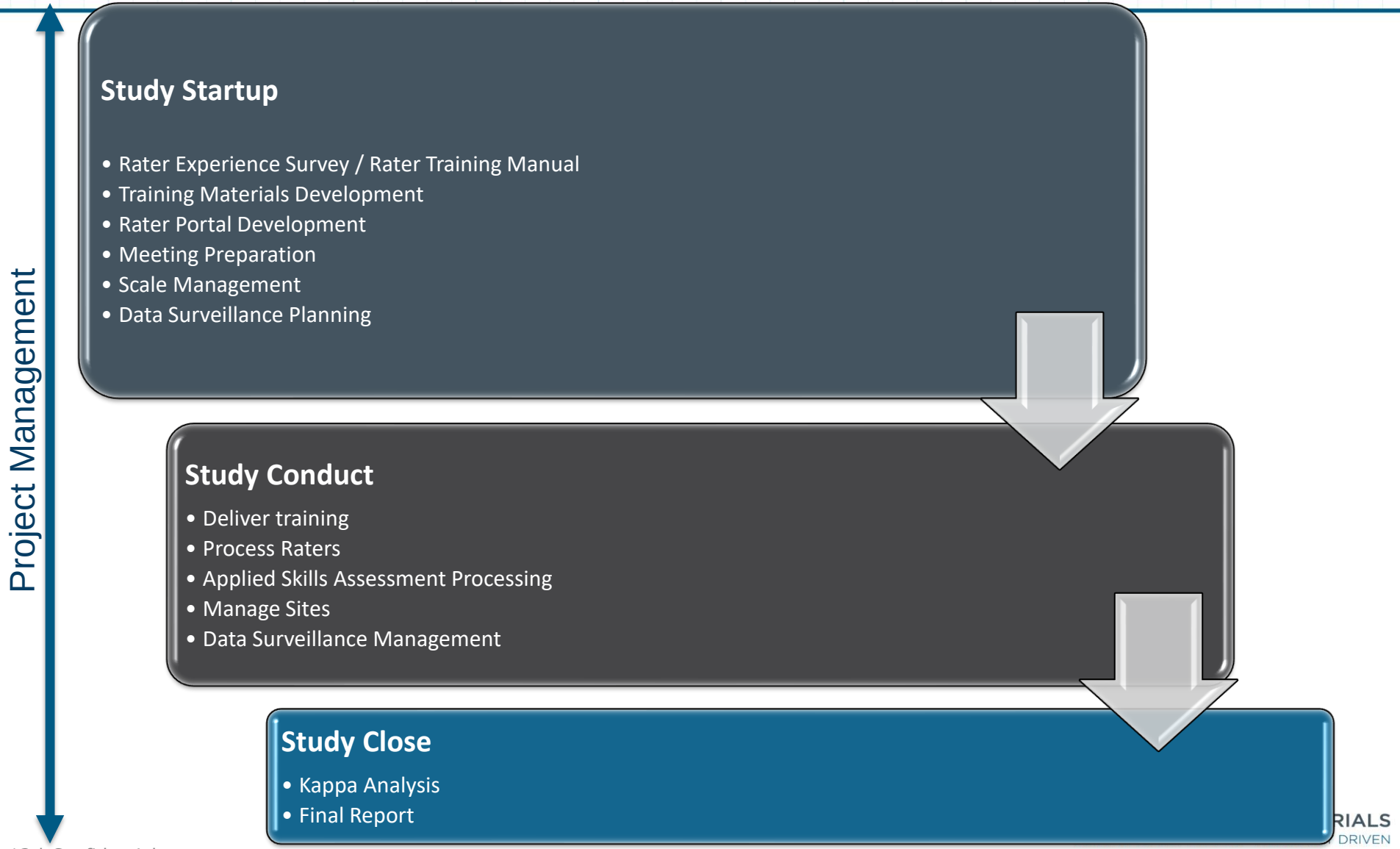
Site Vetting

CAT helps establish criteria for site selection and often attends PSSVs for new sites



WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

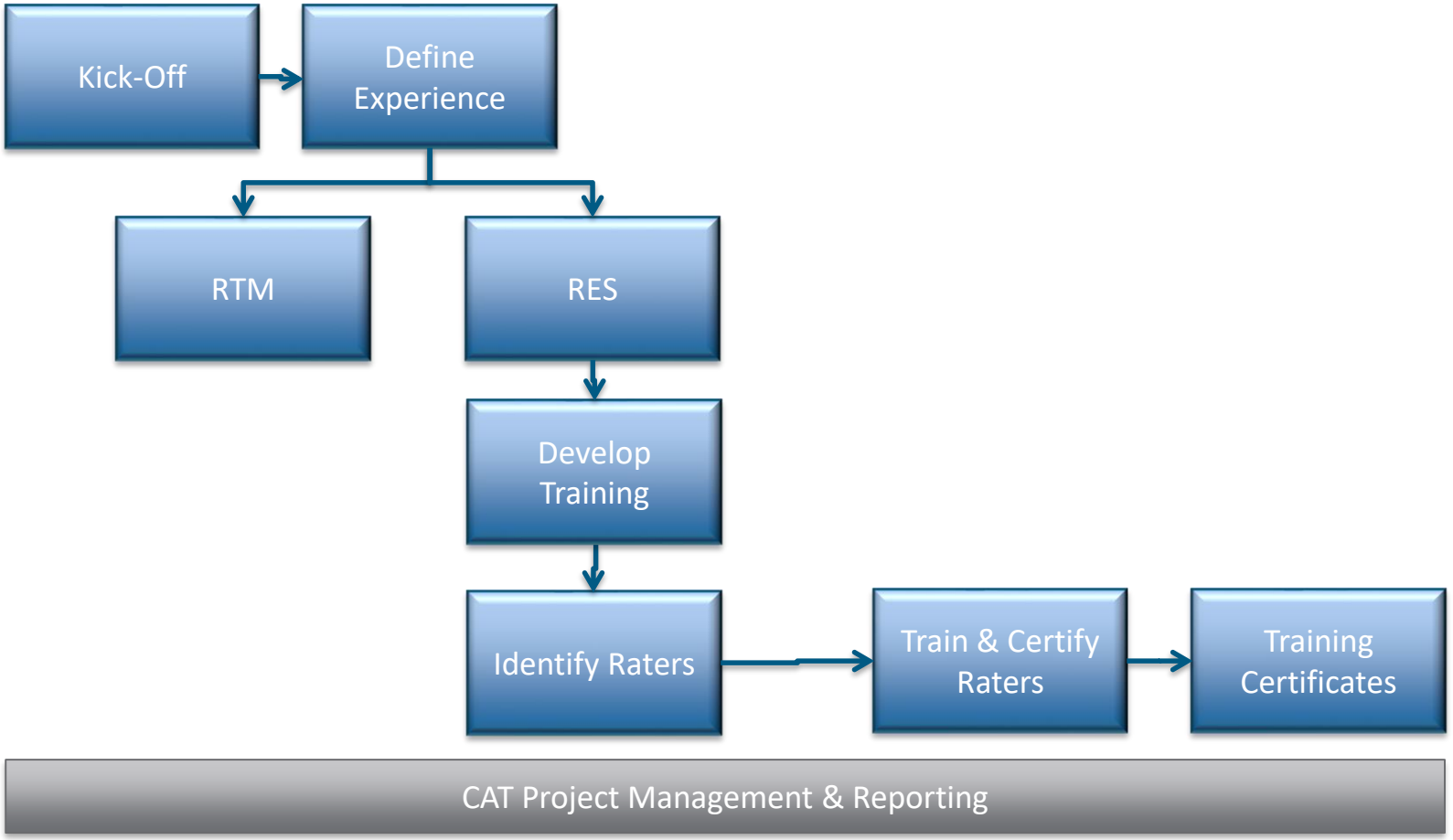
CAT Project Phases



Typical Rater Training Deliverables

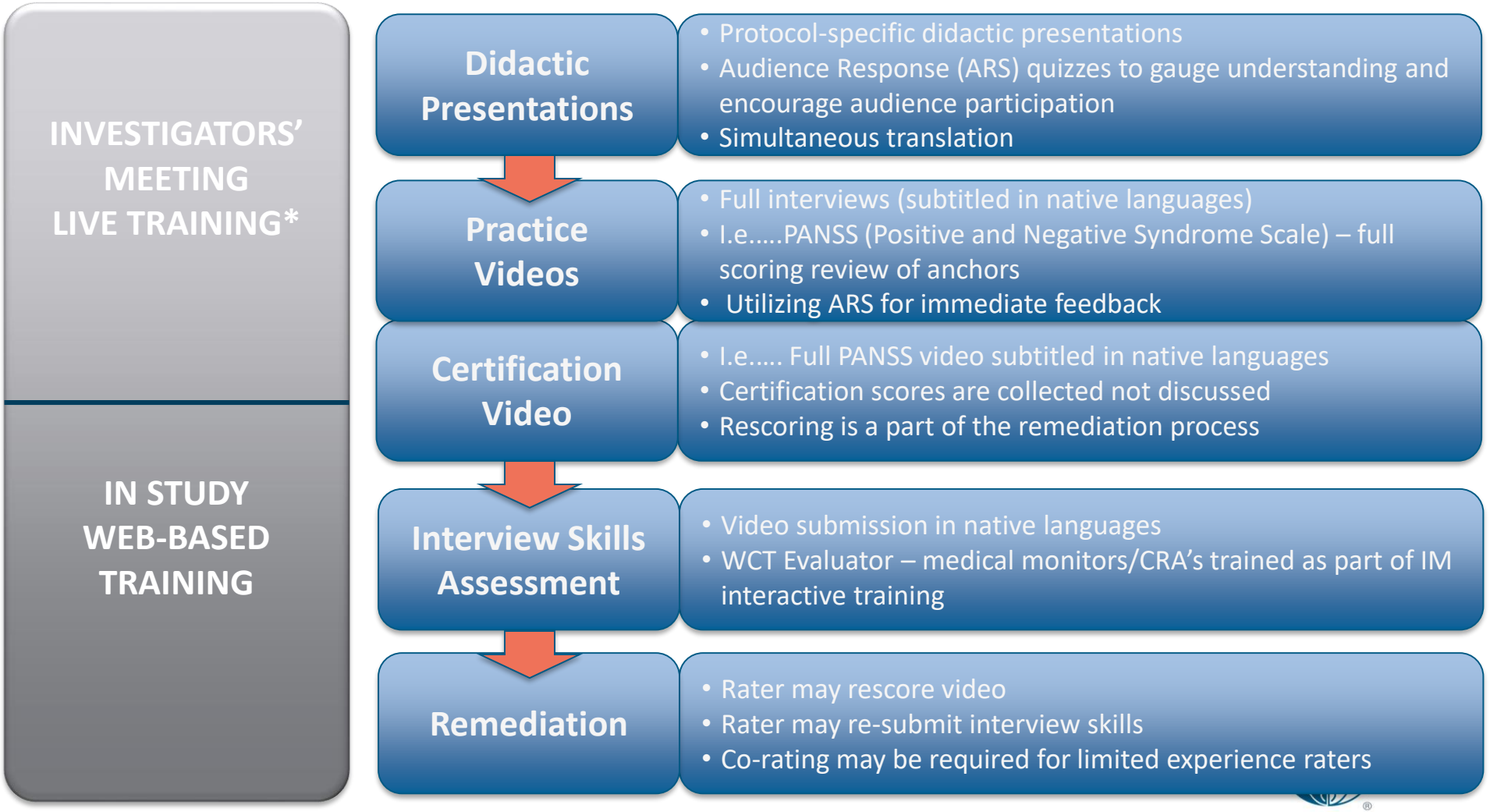
Item	Description
Rater Training Manual (RTM)	RTM documents the training program and rater requirements for the study. It includes the experience requirements, rater roles, scales for training & certification, remediation requirements
Rater Experience Survey (RES)	The RES is used to collect the experience of the potential rater. The RES is customized for each study to provide the ability to assess the raters against the experience requirements of the study. All raters are required to complete an RES.
Training Didactics	The didactics are customized with protocol specific details for the study. They are created for each designated scale for training. Other didactics may include Rater Training Overview and CRA Training.
Patient Videos	Patient videos may be used in the training video. Real subjects are used when possible. Used for training and certification.
Rater Tracker	The rater tracker is used to document all the details of the rater, ie… experience, role, training completion, and certificate date. This information is used to create the Rater Status Report.
Rater Status Report	This report is distributed to CRA and Sponsor to detail the status of the sites and raters. Included information is: Site, Rater, Scale and Status.
Training Certificates	Raters are required to have a training certificate for all qualification scales for the study prior to rating subjects. This document should be maintained in the Regulatory binder at the site.
Rater Training Portal	The training portal is developed for studies to provide a means for in-study rater to do their training. It is customized for each study and provides the ability to generate the Training Certificates.
Kappa Analysis	The Kappa will show the inter rater reliability of the raters for the given training and is by scale. Often required for RoW but, not for US.
Final Report	The Final Report will document the Rater Training program. It should include the program delivery specifics and the final listing of raters qualified for the study.

High Level Rater Training Process



WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

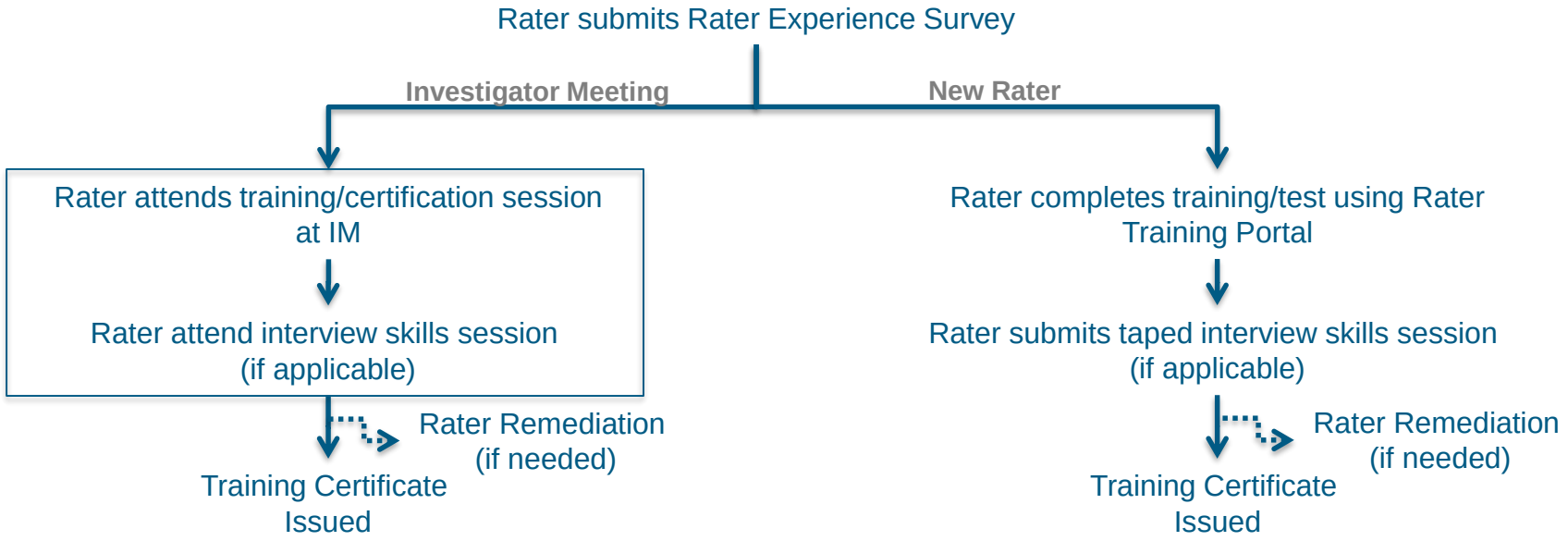
Rater Training Delivery



*Training may be recorded for in-study "new" raters

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

Rater Certification Process



Experience + Training + [Test + ASA] = Certified

Note: Each site must have at least 1 certified rater for all the scales; estimate 2 raters per site
Note: Certificates provided to Site, CRA, and Sponsor as applicable

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

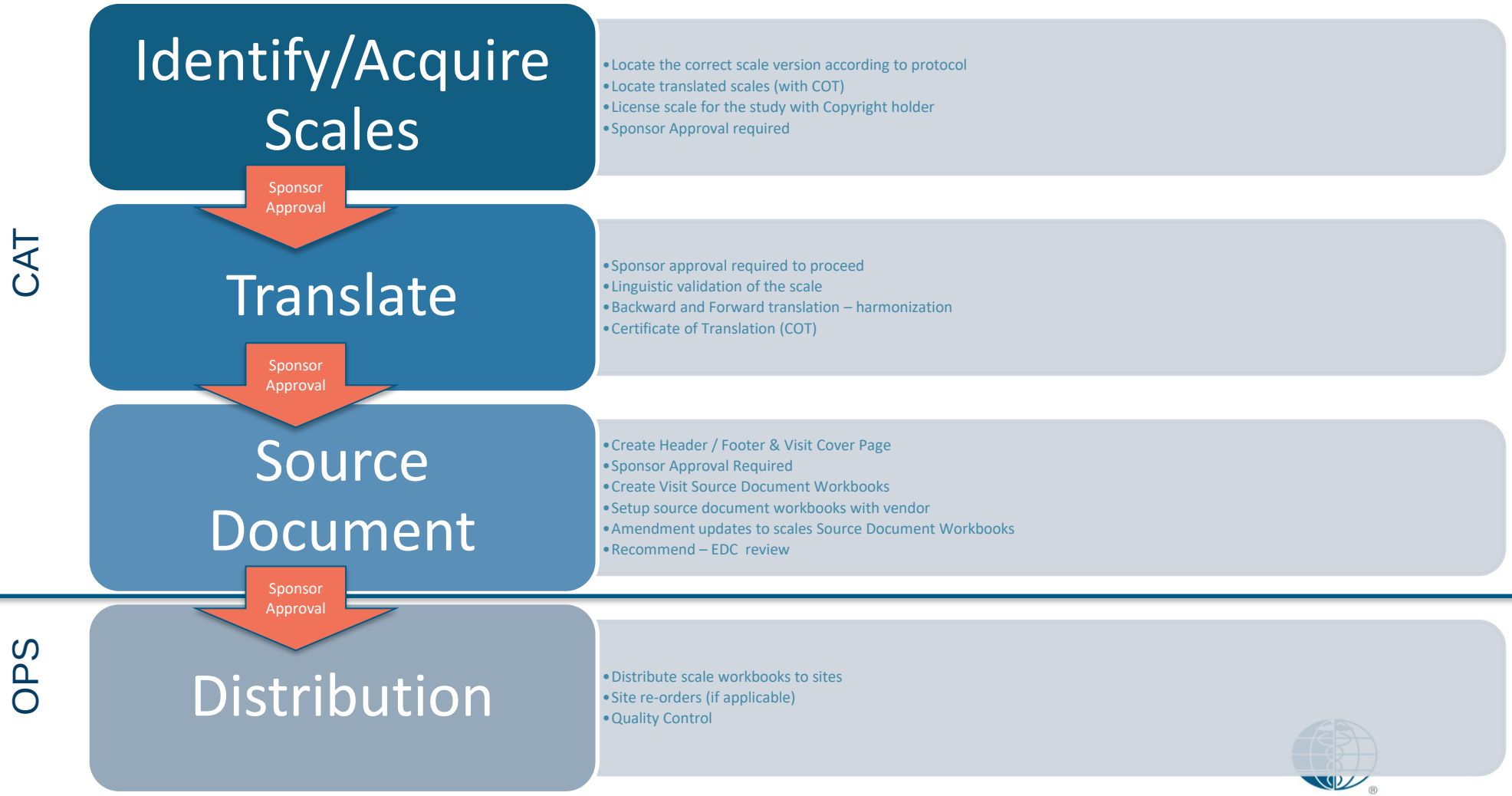
Applied Skills Assessment (ASA)

- Applied Skills Assessment (ASA) is a test of the raters' ability to elicit information using a Mock Interview
- Mock Interview
  - A standardized subject scenario
  - Patient Actors will attend the IM or a designee at the site
  - Individual interview sessions ~ 10 - 20 minutes
- IM raters will conduct the ASA at the IM while new raters will be required to submit a video taped interview
  - Program may be designed for CRA to conduct interview skills at the site and report results to CAT
- Raters that do not pass their ASA will receive a clinical tutorial and may be required to submit an additional ASA as part of remediation

Note: All materials including Video cameras will be provided to the sites for new raters

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

Scale Management



*Operations to provide regulatory submission plan (ongoing)

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

Maximizing Efficacy through Surveillance

Real time data reviews for continuous monitoring of sites throughout study. The focus is ~~Clinical~~ Management ~~not error~~ detection.

- Comprehensive program of clinical monitoring
- Source document review for enrollment confirmation – Screening and Baseline
- Correlations and consistency reviews visit-to-visit
- Monitoring for scoring patterns and trends
- Rater confirmation (change and qualified)
- Clinical contacts/feedback and re-training with raters

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

Typical Data Surveillance Deliverables

Item	Description
Surveillance Methodology (SM)	The surveillance methodology documents the data surveillance program for the study. Details include the type of reviews to be conducted, method of receiving the data, and communication plan. This may be incorporated into the RTM for some studies.
Source Document Tracker	The source document tracker is used to manage the subject visits being monitored. It will contain all the pertinent information about the subject visit including the clinical findings.
Sponsor Detail Report	This report is created from the source document tracker and simply formats the data for the sponsor to more easily view it by subject.
Proportion of Issues Report	The proportion of issues shows what issues are more concerning based upon the proportion of occurrence. This is a higher level report at the issue level and is cumulative for the entire study.
Final Report	The Final Report will document the surveillance program. It should include a summary of the findings in the study.

Types of Surveillance

Programs are designed specifically around the type of study and the concerns of the study.

Item	Description
Enrollment Confirmation	Review and assess the subject visit for subject enrollment into study. May review both screen and baseline visits. Also work closely with Medical Monitor as necessary.
Trend Analysis	Use data edits to monitor the raters performance by assessing the scales for trending.
Site Monitoring	Similar to Trend Analysis; but at an aggregate – can see clustering of scores. Helpful to see if a site could be inflating scores at screening.
Independent Subject Scoring	Monitor the raters by having a independent rater also score the interview and comparing the scores. Using video tapes allow the monitoring of interview skills on an ongoing basis.
Objective Scale Monitoring	Objective scales may be monitored for clinical accuracy, i.e… ADAS-cog

CAT Position Papers

The Impact of Site Characteristics on Efficacy Measures

Bethanne Friedmann, PsyD[1] , Erin B. Kornsey, MS[1], Kathryn Dawson, PhD[1], Henry J. Riordan, PhD[1], Michael F. Murphy, MD, PhD[1], Neal R. Cutler, MD[2]
[1] WORLDWIDE CLINICAL TRIALS, King of Prussia, PA, [2] WORLDWIDE CLINICAL TRIALS, Beverly Hills, CA

Abstract

Introduction
The success of psychiatric trials hinges on the quality of data collected by sites. Data errors and discrepancies between the key efficacy measures of a double-blind, randomized, placebo-controlled, multicenter, phase II clinical trial designed for adult attention deficit/hyperactivity disorder (ADHD), were closely monitored by a panel of clinical experts.

Methods
Data "flags" were based on various scales including the Clinical Global Impression of Severity (CGI-S), Clinical Global Impression of Improvement (CGI-I), Conners' Adult ADHD Rating Scale – Observer: Screening Version (CAARS-O:SV), and Conners' Adult ADHD Rating Scale –Self Report: Short Version (CAARS-S:S). Reports were generated twice a week based on information sites electronically entered since the last report was sent. These reports contained discrepancies between the scales, data entry errors, and rater errors. The clinical manager contacted the sites to gather more information about these data flags via telephone and email.

Results
Sixteen sites with randomized subjects for the trial and were included for analysis (mean number of flags 11.38, standard deviation 12. 47). The overall number of data flags per site was negatively correlated with both the number of subjects screened (r = - 0.45) and number of subjects randomized per site (r = -0.19). Prior studies have suggested that the number of flags per site were proportional to the number of subjects enrolled per site. Because there is evidence this was not the case, sites were divided into 2 categories: sites with a high rate of flags per randomized subject and sites with a low rate of flags per randomized subject. A significant difference in the mean number of subjects randomized by the data flag rates was determined (t = 2.43, p = 0.03) with a higher mean number of randomized subjects for site with lower flag rates. The severity of the data flags were also assigned numeric ratings from 1 (least severe, e.g. data entry error, missing data needs to be entered; n = 116), 2 (moderately severe, e.g. possible incorrect rater completing assessment; n= 59), to 3 (most severe, possible scale discordance; n = 7). This has important implications for the appraisal of sites with lower patient screen and randomization numbers.

Conclusion
By investigating and tracking the frequency and severity of the various flags over the course of the study it is possible to enhance the overall quality of data and ultimately lead to increased effect sizes.

Background

- Data errors and discrepancies between the key efficacy measures in double-blind, randomized, placebo-controlled trials detract from the quality of data and ability to detect separation from placebo.
- It was hypothesized that high enrolling sites could be susceptible to high error rates.
- Alternatively, a low number of data related errors may reflect the expertise gained from repeatedly administering the assessments at the high enrolling sites.

Methods

- Reports were generated twice a week based on new electronically entered data. These reports listed discrepancies between the scales, data entry errors, and rater errors (data "flags").
- Flags were based on efficacy measures including the CGI-S, CGI-I, CAARS-O:SV, and CAARS-S:S.
- The clinical manager contacted the sites to gather more information about the data flags and when necessary to re-educate the site.
- The severity of flags was divided into three levels.

Table 1. Description of Flag Severity

Least Severe	Moderately Severe	Most Severe
• Basic data entry errors, missing data • For example: CGI-S and CGI-I were not entered; one CAARS-O:SV subject left blank.	• Possible un-blinding of rater • For example: rater's initials on CAARS-O:SV and CAARS-S:S were SMI. The CAARS INV rater was blinded to all other measures and therefore, should not have completed the CAARS –S:S.	• Possible scale discordance; rater inflation • For example: from baseline to week 14 CAARS-INV decreased from 48-91; CAARS –S:S decreased from 58-83. CGI-S remained 4 from baseline.

Results

- Analysis was conducted on sites with randomized subjects (n=16). The mean number of flags for the sites was 11.38 (sd = 12. 47) with a total number of 182 flags for the current study.

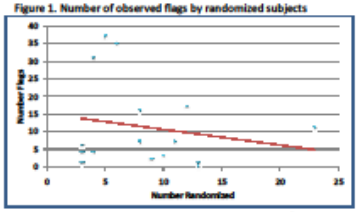
Table 2. Frequency of Randomized Subjects and Flags per Site

Site	Number Randomized	Number of Flags	Low Severity (1)Flags	Moderate Severity (2)Flags	Severe Flags
1	8	16	3	13	0
2	3	1	1	0	0
3	6	35	28	4	3
4	13	1	0	0	1
5	3	4	4	0	0
6	23	11	11	0	0
7	10	3	3	0	0
8	8	7	2	5	0
9	9	2	1	1	0
10	3	6	0	6	0
11	12	17	17	0	0
12	5	37	21	16	0
13	13	0	0	0	0
14	4	31	19	11	1
15	11	7	3	3	1
16	4	4	3	0	1

Results continued

- The overall number of data flags per site was negatively correlated with both the number of subjects screened (r = - 0.45) and number of subjects randomized per site (r = -0.19).

Figure 1. Number of observed flags by randomized subjects



- Sites were divided into 2 categories: high incidents of flags (as defined by ≥ 2 flags/randomized subject) and low incidents of flags per randomized subject.
- There was a significantly larger mean number of randomized subjects associated with 11 sites with a low flag rate (9.9±5.75) when compared to the 5 sites with the high flag rate (5.2±1.92) (t = 2.43, p = 0.03).
- There was no difference in the distribution of the severity of the flags by the number of randomized subjects (p = 0.871).

Table 3. Severity of Flags by Number of Randomized Subjects

	Least Severe	Moderately Severe	Most Severe
<8 Randomized	6 (40%)	4 (31%)	3 (23%)
≥8 Randomized	7 (54%)	4 (31%)	2 (15%)

Conclusions

- Sites with the highest enrollment had the least number of flags which is counter to some past findings and our expectations.
- The distribution of severity of these flags did not differ by number of randomized subjects using a median split.
- Importantly, the number of flags can be decreased during the course of trial with expert rater feedback.
- More research is needed to determine if sites with the lowest number of flags, in particular the lowest number of severe flags, were better able to detect drug-placebo differences than those with higher number of flags.

No conflicts of interest exist in the research and development of this poster

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

CAT Position Papers

Site vs. Remote Inter-Rater Reliability of the PANSS and Information Demand

Bethanne Friedmann, PsyD[1], Henry Riordan, PhD[1], Erin Kornsey, MS[1], Evan Braxton[1], Michael F. Murphy, MD, PhD[1], Neal R. Cutler, MD[2]
[1]Worldwide Clinical Trials, King of Prussia, PA, [2]Worldwide Clinical Trials, Beverly Hills, CA

Abstract

Psychiatric clinical trials utilize subjective measures to assess efficacy. Training and alignment of raters is crucial to ensure adequate inter-rater reliability and eventual study success. Most training is conducted at study start-up and focus on raters' ability to score video recordings within an acceptable standard. While this method is proven to align raters, it does not assess a rater's ability to conduct an interview, nor does it prevent rater drift during the study. There are numerous monitoring methods that assess interview skills as well as decrease rater drift.

Videotaping/videoconferencing promote continuous inter-rater reliability by permitting expert independent ratings of the same subject and immediate feedback on scoring and interview skills. Accurate scoring of the Positive and Negative Syndrome Scale (PANSS) requires rater observation, interviewing the patient (clinical interview) and collecting information from a family member over the previous week (informant information). However, remote raters may not have access to all of the information needed to assess all scale items resulting in poor validity and reliability.

For this study, six raters completed and videotaped twenty-eight ratings of the Structured Clinical Interview for the Positive and Negative Syndrome Scale (SCI-PANSS). These videos were rated by expert same-language raters. Results demonstrated only a moderate agreement between the site raters and expert raters for total PANSS scores. PANSS items were divided into three categories based on the type of information needed to score each item. Average Intraclass Correlation Coefficient (ICC) suggested only poor reliability between site raters and expert raters for items based solely on informant information (ICC =.04); fair reliability was found for items scored using the clinical interview and informant information (ICC = .376). Results indicate there was a moderate reliability for items scored solely from the clinical interview (ICC = .403). This data underscores the demand for all available source of information for remote raters in order to ensure valid and reliable PANSS assessments.

Background

The PANSS[1] is a 30-item, 7-point interview-based assessment utilized to measure symptoms of psychosis in a variety of psychiatric disorders, such as schizophrenia, bipolar disorder, and depressive psychosis. It is routinely used in psychopharmacological studies to measure change in psychotic symptoms. It is divided into seven positive (P1-P7), seven negative (N1-N7), and 16 general items (G1-G16) associated with the symptoms of psychosis.

Utilization of the structured clinical interview for the PANSS (SCI-PANSS) increases the validity of the information obtained. The worst symptoms from the past seven days were assessed. Scoring is based on information obtained during an interview with the patient and corroborated by an informant. Collateral information can come from primary care hospital staff and family members, in addition to behavioral observations by the rater during the clinical interview.

To ensure the validity of data, site raters (SR) were monitored by expert raters (ER) via video recorded interviews. We hypothesized that lack of access to the informant would differentially affect PANSS items based on the amount of collateral information needed for accurate rating.

Methods

PANSS items were divided into three categories according to the type of information needed to score these items appropriately (see Table 1). One category that has two items (N4 and G16) contains items based solely on informant report (I); the second category has 16 items (P2, N1, N3, N5-7, G1-4, G9-13 and G15) that utilize information gathered during the clinical interview (C); the third category is composed of the 12 remaining items that utilize information from both informant and clinical interview (IC).

CATEGORY	ITEMS
Informant Information Items (I)	N4, G16
Clinical Interview Items (C)	P1, P3, P4, P5, P6, P7, N2, G5, G6, G7,G8, G14
Both Informant Information and Clinical Interview (IC)	P2, N1,N3, N5, N6, N7, G1, G2, G3, G4, G9, G10, G11, G12, G13, G1

Table 1. PANSS categories divided by information sources.

This study used six raters from a multinational Phase IIa clinical trial investigating the effects of an antipsychotic agent on individuals diagnosed with schizophrenia who were hospitalized for an acute exacerbation of psychotic symptoms. The primary efficacy of the clinical trial was a change in PANSS score from baseline to week four.

The three expert raters used in this study were considered key opinion leaders in schizophrenia in Russia and Ukraine. They all held medical degrees and university appointments.

Table 2 depicts the experience of the site raters. All six raters were psychiatrists. Three raters had 3-5 years experience administering the PANSS, and three raters had more than 6 years experience administering the PANSS. All raters had used the PANSS multiple times in the past two years (three raters administered the PANSS 26-50 times; three raters administered the PANSS more than 50 times.)

NUMBER OF TRIALS CONDUCTED IN THE PAST TWO YEARS Mean (Standard Deviation)	RESEARCH EXPERIENCE (YEARS) Mean (Standard Deviation)	CLINICAL EXPERIENCE (YEARS) Mean (Standard Deviation)
3.4 (1.14)	5.4 (1.67)	7.8 (5.54)

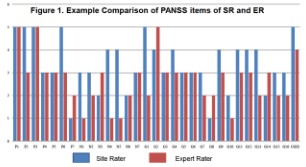
Table 2. Site Raters' Experience with Schizophrenia.

Methods (cont)

The site raters were trained on how to conduct the PANSS at an investigators meeting in Europe where they participated in an interactive review of anchors and scoring conventions which was simultaneously translated into their native language. Certification was based on raters' ability to score a video-recorded patient interview. The video interview was conducted in English and subtitled in their native language. Site and expert raters were aligned to a consensus scoring of the certification videos and had to achieve greater than or equal to 80 percent concordance prior to being approved for this trial.

Twenty-eight video recordings of site raters PANSS interviews were obtained from visits 3 and 9. All site raters utilized the SCI-PANSS. Each PANSS interview was conducted in a local language and rated separately by an in-country expert rater. All site raters were instructed to provide a summary of corroborating information. Expert raters were blinded to the site rater's scores and did not have direct access to informants; all information had to come from the videotape, including site rater narratives of informant information. The expert raters provided feedback on the interview with their scores for comparison to the site raters' scores. Figure 1 demonstrates an example comparison between a site rater and expert rater by individual PANSS items.

Figure 1. Example Comparison of PANSS items of SR and ER



The ICC was utilized as a measure of the reliability of ratings between site and expert raters and may be conceptualized as the ratio of between-groups variance to total variance.

The ICC evaluates the level of agreement between raters in measurements, when the measurements are interval in nature. This method is better than ordinary correlation as more than two raters can be included, and there is a correction for correlations between raters that becomes apparent when the range of measurement is large. The coefficient represents concordance, where 1 is perfect agreement and 0 is no agreement at all.

Results

ICC by category are shown in Table 3. Average ICCs suggested poor reliability between site raters and experts for items based on informant report only (ICC I =.04); fair reliability (ICC IC=.376) for combined items, and moderate reliability (ICC C = .403) for items gathered based solely upon clinical interview. Despite these qualitative differences there were no significant statistical differences between these three information categories (x^2=1.68, p=.43) when comparing ICCs.

CATEGORY	INTRACLASS CORRELATION (ICC)
Informant Information Items (I)	.04
Clinical Interview Items (C)	.403
Both Informant Information and Clinical Interview (IC)	.376

Table 3. Intraclass Correlations (ICC) by Category.

According to interpretation conventions by Fleiss (1981)[2], the ICC for the overall PANSS score showed only moderate agreement between site and expert raters (ICC = .438, p >.05). The majority of individual PANSS items ICCs (60%) fell in the moderate range (.40–.59) of reliability while 20% of the site rater versus expert rater ICCs fell in the good to excellent range (0.60–1) of reliability, and 20% fell in the poor range (<0.20). No readily apparent pattern of ICCs was noted based on positive/negative/general psychopathology composites. Rater demographics such as prior PANSS experience and rating performance at the investigator meeting did not appear to be related to site versus expert ICCs.

Conclusions

The PANSS assessment is the cornerstone of efficacy analysis for a majority of clinical trials. Raters play a significant role in obtaining this data. This is typically site raters, but the use of remote raters is increasing. A multitude of studies fail to demonstrate change in symptoms based on PANSS scoring. It is critical that sources of variability among raters is minimized in order to ensure that PANSS assessments are accurate.

Videotaping / videoconferencing promote continuous inter-rater reliability by permitting expert independent ratings of the same subject and immediate feedback on scoring and interview skills. However, remote raters may not have access to all of the information needed to assess all scale items, resulting in poor validity and reliability.

By limiting what PANSS information remote expert raters had access to, this study showed that incomplete information, whether from informants or from the clinical interview, altered inter-rater reliability. Not surprisingly, the lack of clinical interview information most reduced inter-rater reliability. But the fact that the majority of individual PANSS items ICCs fell only in the moderate range of reliability between site and remote raters, suggests there may be overall reliability issues with remote raters that could affect PANSS scoring. Reliability may be improved by using the informant questionnaire for the PANSS (IQ-PANSS) as opposed to site rater narratives. Further study of remote rater PANSS scoring variability is warranted.

References

Kay S. R., Opler L.A., Fiszbein A. 2006. Positive and Syndrome Scale (PANSS) Technical Manual. New York. Multi-Health Systems.

Fleiss J.L. 1981. Statistical Methods for Rates and Proportions. 2nd ed. Wiley, New York.

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED · MEDICALLY DRIVEN

Questions

WORLDWIDE CLINICAL TRIALS
SCIENTIFICALLY MINDED • MEDICALLY DRIVEN

Thank You

WORLDWIDE CLINICAL TRIALS

SCIENTIFICALLY MINDED • MEDICALLY DRIVEN