

**Feasibility Study to Assess the DAIDS/OCSO
Clinical Site Monitoring Process Improvement
Program (CSM-PIP)**

Final Report

Judith A. Hautala

Oren Grad

Samuel Thomas

December 31, 2012

National Institute of Allergy and Infectious Diseases

National Institutes of Health

Contract #01A-0408601

Order #HHSN272201100290U

Executive Summary

DAIDS/OCSO Clinical Site Monitoring Overview

Clinical site monitoring for the Division of Acquired Immune Deficiency Syndrome (DAIDS) of the National Institute of Allergy and Infectious Diseases (NIAID) is managed through the DAIDS Office of Clinical Site Oversight (OCSO). DAIDS clinical trials subject to this monitoring oversight include approximately 100 active protocols conducted across five primary clinical trials networks,¹ encompassing approximately 500 domestic and international sites and involving 20,000–25,000 subjects on study annually. The majority of clinical trial sites are monitored quarterly with the specific protocols and subjects for which records are monitored determined primarily by either the protocol risk-based algorithm described in Chapter 1 of this report or by protocol-specific monitoring plans that supersede the algorithm for certain protocols. The monitoring data are collected centrally by the DAIDS-Enterprise System (DAIDS-ES) and the results reviewed and responded to primarily through OCSO Program Officers working directly with sites.

In 2008, OCSO initiated a clinical site monitoring process improvement program (CSM-PIP). The first of these initiatives, Monitoring More Recent Data (MRD), was designed to focus monitoring efforts on more recent data² and was implemented at all sites in January 2010. The second initiative, designated Risk-Based Protocol Monitoring Plans (RBPMP), established standardized methods to rank protocols according to risk and assigned increased levels of monitoring to higher risk protocols.³ This initiative was implemented at all sites in April 2010. A third initiative, Risk-Based Site Performance Monitoring (RBSPM), designed to allocate increased monitoring resources to sites judged to be at higher risk of poor performance has not yet been implemented. All of these initiatives were designed with the goal of improving site performance in terms of a

¹ AIDS Clinical Trials Group (ACTG), HIV Prevention Trials Network (HPTN), HIV Vaccine Trials Network (HVTN), Microbicide Trials Network (MTN) and International Maternal Pediatric Adolescent AIDS Clinical Trials (IMPAACT).

² MRD originally limited record review beyond informed consent and eligibility to data collected only in the 12 months prior to the monitoring visit but, at the request of the Food and Drug Administration (FDA), that was increased in January 2011 to data collected in the three years prior to the monitoring visit.

³ Higher risk protocols are those for which the potential for harm to study subjects from errors or the potential for errors to adversely affect data quality are judged to be higher.

NIAID CONFIDENTIAL

reduced incidence of errors, either for all protocols or for those protocols where the consequences of errors might be greater.

Purpose of Feasibility Study

In September 2011, the IDA Science and Technology Policy Institute (STPI) was tasked by DAIDS/OCSO with analyzing the feasibility of conducting an integrated process and outcome evaluation to determine whether the first two CSM-PIP initiatives have had an impact on site performance. The feasibility study was intended to address various topics, including the evaluation questions to be asked, the data that would be required to answer those questions, and a plan for conducting the evaluation if it were deemed feasible. In addition, the study was to provide an analysis of best practices for site monitoring in public and private sector clinical trials through a comprehensive literature review and discussions with individuals responsible for monitoring practices.

Analysis Methodology

The feasibility analysis involved seven primary activities.

1. Analysis of documents concerning the DAIDS clinical trials program and the OCSO monitoring program
2. Interviews with DAIDS/OCSO staff and contractors involved with site monitoring to identify monitoring data and information that could be used to indicate an improvement in site performance
3. Interviews with DAIDS clinical research site staff to gain their perspective on the monitoring process and the impact of CSM-PIP
4. Interviews with representatives from the DAIDS Network Statistics and Data Management Centers and the HVTN Evaluation Committee to identify additional data and information that might indicate an improvement in site performance
5. Analysis of the algorithm used to determine which protocols and subject records are reviewed at each monitoring visit
6. Analysis of the structure of the monitoring data collected by DAIDS-ES
7. Analysis of DAIDS-ES site monitoring data to determine if these data would be useful in establishing whether site performance changed with implementation of CSM-PIP

NIAID CONFIDENTIAL

Analysis of DAIDS-ES Site Monitoring Data

An exploratory analysis of existing DAIDS/OCSO monitoring data was performed to determine whether these data can be used to assess the impact of CSM-PIP on site performance. The basic performance metric for the analysis was the number and character of the “observations” recorded by the monitors. These observations are assigned by the monitors to one of 25 “codes” that classify observed anomalies in the site’s adherence to protocol or in the completeness or accuracy of submitted data. Each observation in the data set provided by DAIDS-ES was characterized by site ID, protocol ID, site visit start date, and observation code. Each observation was then assigned by STPI researchers to one of the five DAIDS clinical trial networks based on the protocol ID.

The time period for the analysis was January 2008 through June 2012, separated into the following stages.

Period 0A	Jan-Dec 2008	Baseline
Period 0B	Jan-Jun 2009	DAIDS-ES selection algorithm implemented
Period 1	Jul-Sept 2009	MRD ⁴ implemented at domestic pilot sites
Period 2	Oct-Dec 2009	MRD ⁴ implemented at international pilot sites
Period 3	Jan-Mar 2010	MRD ⁴ implemented at all sites
Period 4	Apr-Dec 2011	RBPMP implemented at all sites
Period 5	Jan 2011-Jun 2012	MRD ⁵ and RBPMP implemented at all sites

Theoretical Concerns

From the outset, it was recognized that there were several potential concerns with the ability of this type of data analysis to provide valid measures of the impact of CSM-PIP on site performance.

Calculating Error Rates

The first concern was that the observation data provide error *counts* rather than error *rates*. To calculate actual error rates, one needs a suitable denominator by which error counts can be divided in order to calculate how many errors occur *relative to the number of opportunities for an error to occur*. In principle, the number of opportunities for an

⁴ From July 2009 through December 2010, the MRD intervention specified that review of on-study visits was generally limited to the 12 months prior to the monitoring visit.

⁵ After January 2011, the MRD intervention was modified at the request of FDA such that review of on-study visits was extended to three years prior to the monitoring visit.

NIAID CONFIDENTIAL

error to occur is related to the number of protocol events reviewed by the site monitors which would be the ideal denominator. However, capturing this parameter is not feasible as it would consume time otherwise used to evaluate data elements for errors and thus reduce the number of data elements reviewed. Therefore, two proxy denominators, number of subjects on study and time, were investigated for the data analyses presented in this report.

Sampling Bias

The second concern was that both of the implemented CSM-PIP interventions changed the sampling frame for the subject records reviewed. MRD changed the time period for which records are sampled for review and RBPMP changed the degree to which different protocols are sampled for review. As a result, if there are pre-existing, systematic variations in error rates over time within or across protocols or pre-existing, systematic variations in error rates across different protocols, examining error rates before and after CSM-PIP implementation could lead to an apparent change in site performance even where there has been no actual change, simply because records with different characteristics were selected for review. A number of the exploratory analyses of DAIDS-ES monitoring data described in this report were conducted to assess whether such systematic variation was present in the data set.

Attribution

The third concern was whether any changes observed in error rates after implementation of the CSM-PIP interventions could with confidence be attributed to the interventions. Although the phased introduction of MRD and RBPMP provide at least some opportunity for internal experimental controls, these controls may be inadequate if the actual effect is small relative to the random variation across time, sites and protocols or if the time required to observe the effect is long compared to the implementation timeline. Because of these uncertainties and the many other un-controlled and non-measurable influences described in this report, attribution of any observed change in error rates to the CSM-PIP interventions would need to be based on an informed, subjective assessment that takes into account other qualitative information about the behavior of the system.

Findings

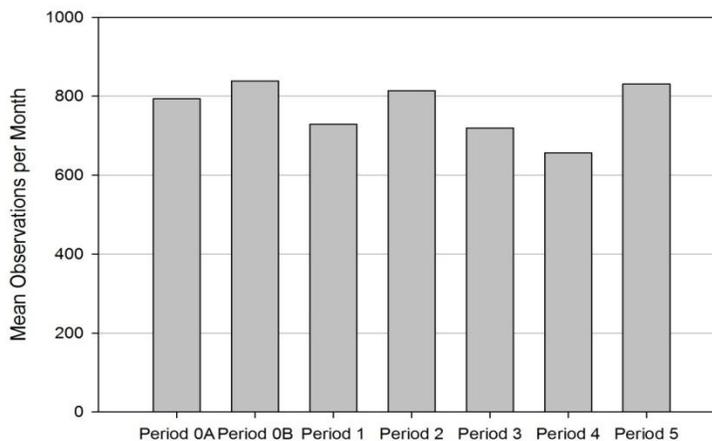
Overall Errors

As shown in the figure on the next page, the overall number of observations across all five DAIDS clinical trial networks from January 2008 to June 2012 is reasonably stable with a mean of 769 ± 68 observations per month. This result on the surface would

NIAID CONFIDENTIAL

indicate that CSM-PIP has not to date had any substantial effect on site performance as measured by monitoring results. However, because of the theoretical concerns noted above and the potential for other confounding variables, several more detailed analyses were conducted across networks, protocols and sites to further investigate this conclusion. These data also reveal that the absolute level of observations across the system is low. Given that the monitoring program encompasses 500 sites and 100 protocols, the observation level is on average only 1.5 observations per site per month and 7.5 observations per protocol per month.

Mean Observations per Month for All Networks Combined

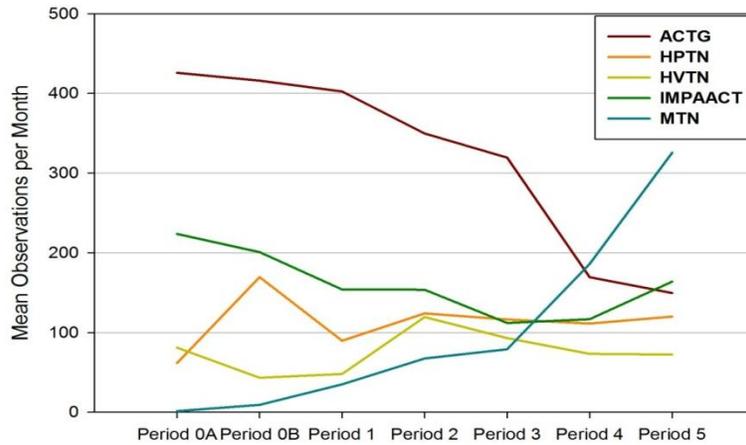


Errors by Network

As shown below, observation counts per month for HPTN, HVTN and IMPAACT are reasonably constant over the 54 month period, in line with the aggregated data. In contrast, the observations per month for ACTG decrease substantially while those for MTN increase to a similar degree, balancing out to give a constant level of observations on an aggregated basis.

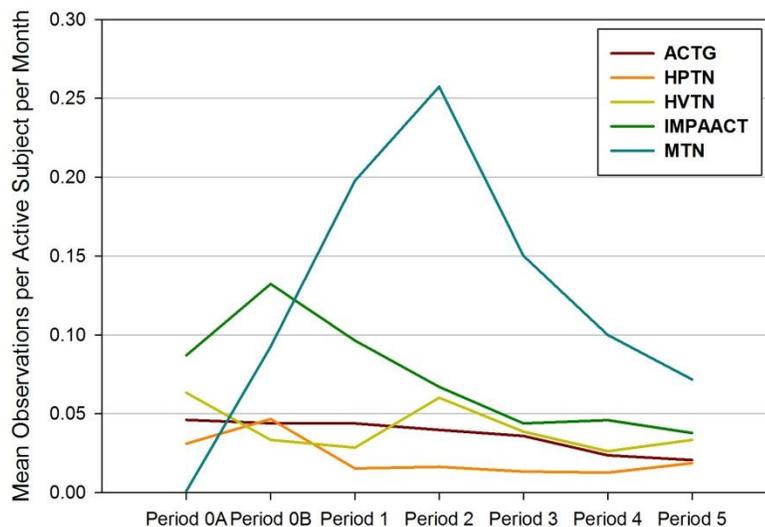
NIAID CONFIDENTIAL

Mean Observations per Month by Network



However, when the data is normalized for the number of subjects on study in the different periods, the results are somewhat different (see below). In this case, observations per month per subject are reasonably constant for HPTN, HVTN and ACTG while IMPAACT decreases by ~50%. MTN presents the most interesting pattern in that observations per month per subject fell 70% from January 2010 through June 2012 while the total subjects increased almost 10 fold. Prior to that, the dramatic spike in observations per month per subject during the second half of 2009 may reflect the period that sites were just coming on line when errors would be expected to be greater.

Mean Observations per Active Subject per Month by Network



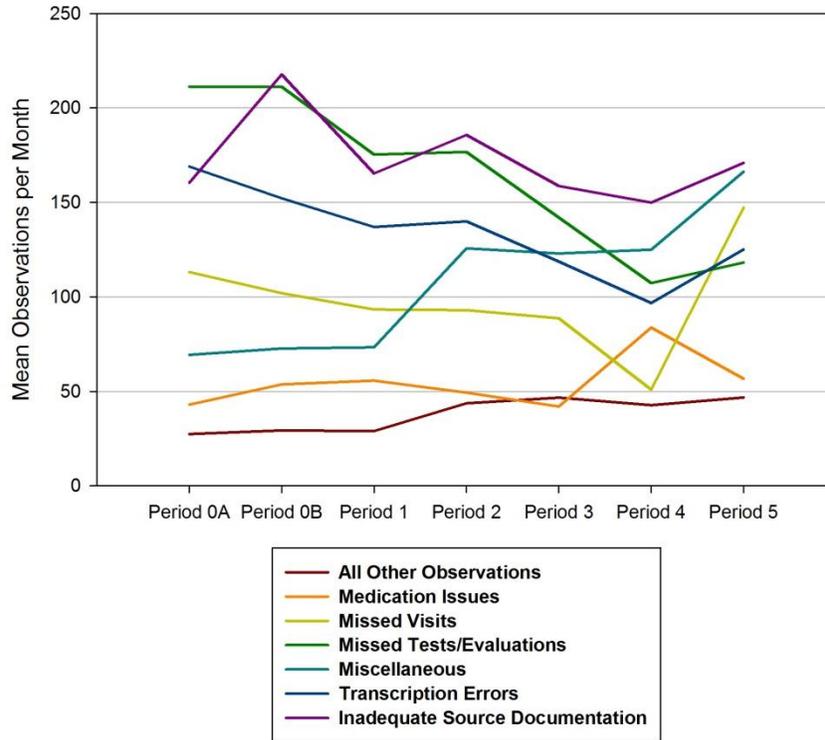
NIAID CONFIDENTIAL

This analysis by network illustrates the complexity of factors potentially affecting error rates that will likely confound attempts to distinguish an effect of the CSM-PIP interventions. Moreover, the fact that in either analysis one sees a decrease in errors only for a single network makes any attribution of the change to CSM-PIP difficult unless there is a substantive reason why the effect would be different for different networks.

Analysis by Observation Code

Across the system and over the 54 month period analyzed, it was noted that approximately 95% of the observations appeared in 12 of the 25 codes tracked by the monitors. The figure below shows the pattern of these observations on a per month basis over the 54 months. It is worth noting that missed tests/evaluations, transcription errors and inadequate source documentation are the most prevalent of these major observation codes, accounting for approximately 60% of total observations. Although two of these codes, missed tests/evaluations and transcription errors, decreased substantially over the time period analyzed (50% and 30% respectively), it is not clear why the CSM-PIP interventions would have a specific effect on the incidence of these particular errors. However, these data (and the analysis by network in Chapter 2, Section E.3 of this report) demonstrate that routine analysis of errors by observation code could be valuable in identifying error trends that might warrant further investigation.

Mean Major Observations per Month



The second analysis by observation code targeted those observations viewed by DAIDS/OCSO and the clinical trial networks as the most serious. As shown in Figures 6 and 7 of this report, serious observations represent only a small fraction of the total observations for all networks (2-4% in the final January 2011-June 2012 period) and show similar trends over time by network as the total observation codes. Therefore, analysis of these serious codes also did not demonstrate an effect from the CSM-PIP interventions on site performance.

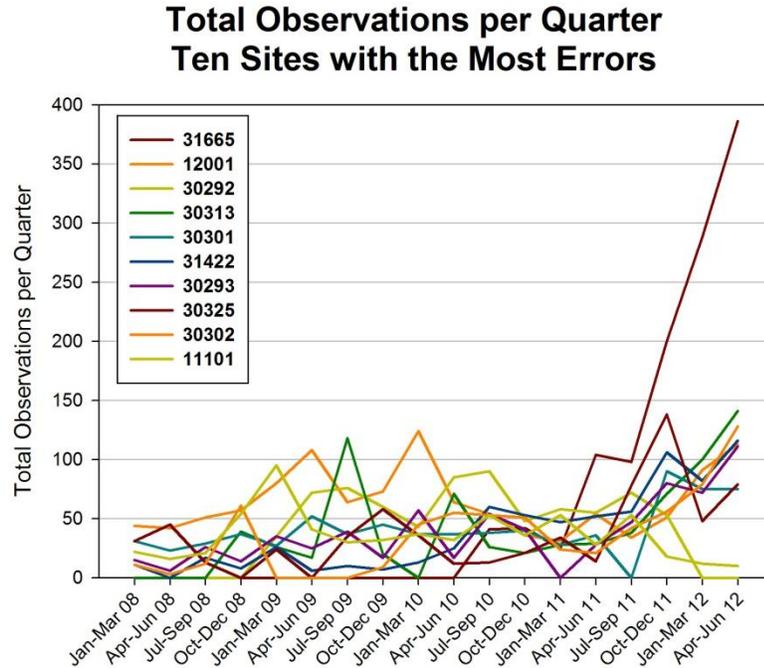
Analysis by Site

Although analysis of changes in observations by network over time did not show promise for the use of monitoring data to assess the impact of CSM-PIP, a network based analysis aggregates data from many individual sites and could potentially obscure an effect of CSM-PIP at a site specific level. STPI researchers therefore analyzed monitoring data for the 10 sites which had the largest number of total observations during the 54 month analysis period. The pattern of observations per quarter⁶ shown below

⁶ The data is expressed per quarter to reflect the fact that an individual site would be visited once per quarter and therefore a per month value is not as relevant on a site specific basis.

NIAID CONFIDENTIAL

indicates that, with the exception of the last three quarters of data for one site, the observations per quarter for all of these sites varies between 0 and 100 with no distinct patterns over time or by site.

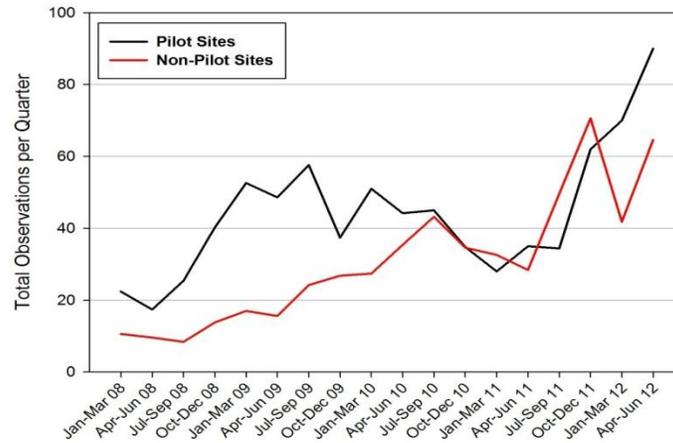


Because five of these 10 most error prone sites were MRD pilot sites, the aggregated data for the five MRD pilot sites was compared with that for the remaining sites to determine if there was any evidence of an early effect of MRD at the MRD pilot sites. As shown below,⁷ there is no clear evidence that MRD was having an early effect at the pilot sites. Moreover, the increase observed over the latter part of the analysis period might indicate that some factor or factors was causing error rates to increase at these sites which might obscure any simultaneous effect of CSM-PIP in reducing error rates.

⁷ Site 31655 was eliminated from the non-MRD pilot site data as the observations at this site were so divergent.

NIAID CONFIDENTIAL

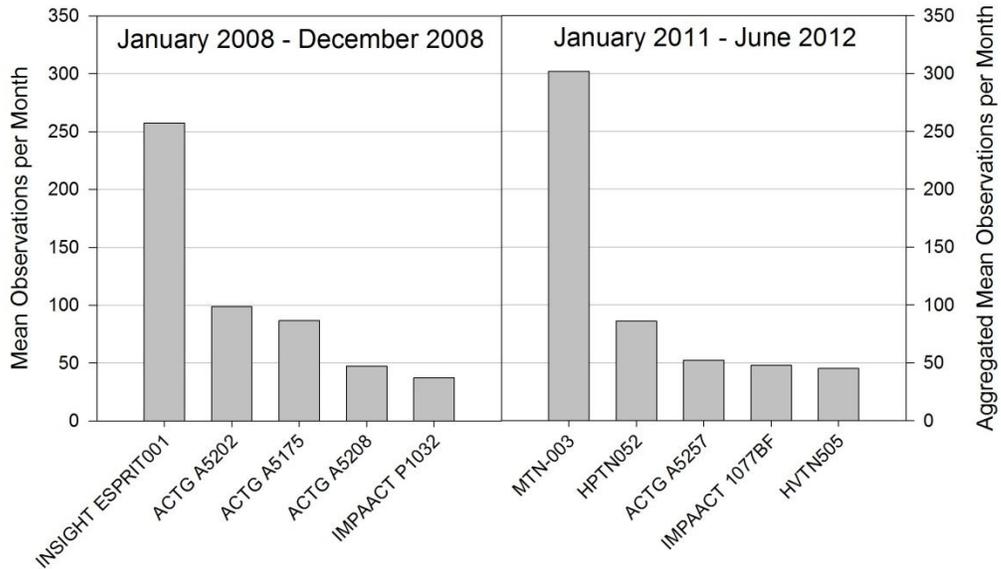
Total Observations per Quarter MRD Pilot Sites vs. Non-Pilot Sites



Analysis by Protocol

The final approach that was investigated for detecting an impact of CSM-PIP on error rates was to examine observations by protocol on the premise that the CSM-PIP interventions might have different effects on different protocols. Because the overall level of errors per protocol across the system is low (7.5 per protocol per month), the analysis focused on those protocols with the largest number of errors. The figure below shows the data for the five protocols with the most errors in the first and last analysis periods (January–December 2008 and January 2011–June 2012).

**Mean Observations per Month
Five Protocols with the Most Errors**



Although each period had one active protocol with a substantial number of observations per month, the other four protocols had 50-100 observations per month on average. This number is probably too low to expect implementation of CSM-PIP to have effect on a sufficiently large number of sites so show an observable effect in the total error rate even for these most error prone protocols. Therefore, protocol specific analysis of error rates is not likely to be any more useful for evaluation than the other types of analyses.

Stakeholder Perceptions of CSM-PIP Value

In the course of this study, STPI researchers gained certain insights on the value of the CSM-PIP interventions from discussions with DAIDS/OCSO Program Officers, PPD staff involved in DAIDS clinical site monitoring, and staff from DAIDS clinical research sites who are involved with the monitoring process.

The Program Officers had a generally positive view of MRD. In their view, it makes observations more relevant to current site operations and they perceive sites to be more diligent because more current data, for which they can be held more directly responsible, is being monitored. The only effect they noted from RBPMP was that interventional trials appear to have increased errors because these trials receive more of the monitoring effort. Although concentrating more monitoring resources on sites that are more likely to have errors through RBSPM is viewed as logical, there were concerns with the static ranking of site risk as the degree of risk at a site changes over time as the site grows, takes on new studies, has staff turnover, etc.

NIAID CONFIDENTIAL

PPD monitoring staff viewed MRD as having minimal impact because, from their perspective, most records assigned for review are more recent records in any event. The primary advantage identified for RBPMP was that it set monitoring goals for each study. This practice allows protocols to be removed from the pool when their goals have been met. In the past, monitors continued to randomly review subject records on all protocols to a theoretical level of 100% which spread monitoring resources much more sparsely across protocols. PPD staff shared the Program Officer view that focusing monitoring effort on those sites where errors are more likely to occur will have the greatest effect on site performance but also shared the concern that the risk status of a site is not static but changes over time.

Clinical research site staff viewed MRD as making monitoring a more positive experience for sites because errors can be acted on and corrected. From the sites' perspective, they have not seen any changes since implementation of RBPMP and they had no comment on RBSPM. They did comment that error rates are not driven by monitoring effort but by factors such as starting new protocols and training new staff.

Feasibility of Process and Outcome Evaluation

Based on the analysis in this report, STPI researchers have concluded that it is not feasible or warranted to conduct a formal process and outcome evaluation of the CSM-PIP interventions implemented to date based on analysis of objective monitoring data. Furthermore, although it would be possible to collect qualitative information on the value of the CSM-PIP interventions through systematic interviews with PPD site monitors and clinical research site staff, such a process may not be warranted. The reasons for these conclusions are as follows.

Error Rate Trends

No objective effect on error rates is discernible to date in either overall error rates across the system or when evaluated separately by network, observation code, site or protocol. It is, of course, possible that the CSM-PIP interventions have not been in place sufficiently long for their effect to be manifested. In that case, repeating the analyses described in Chapter 2 periodically over time might reveal an effect. However, a formal process and outcome evaluation would still not be warranted. The effect could be demonstrated simply by DAIDS/OCSO staff performing the analyses periodically and determining if a reduction in error rates is occurring.

Confounding Variables

The most important confounding factor is wide variability in the number and complexity of protocol events associated with different protocols and hence the number of opportunities for an error to occur. Unfortunately, as explained in Chapter 2, the effort

NIAID CONFIDENTIAL

involved in obtaining events reviewed as an error rate denominator would be substantial and difficult to justify given that other factors could further confound any analysis of error rates. The second confounding variable is that protocol events related to prevalent observations can change in frequency over time. Again, determining for every protocol whether changes in these prevalent observations were due to changes in the number of protocol events or site performance would not be justified.

The third confounding variable is that the most important factors affecting error rates are the number of new protocols being opened, the number of new sites coming on line, staff training, staff turnover, etc. Although changes such as CSM-PIP might result in more effective corrective actions to improve site performance, such improvements would often be overshadowed by these other factors which might either decrease or increase error rates. Because all of these variables are site or protocol specific, sorting out the variables and correcting for them would have to be done for each site/protocol pairing which would be an unreasonable amount of effort simply to demonstrate with error rates that CSM-PIP did or did not have an effect. The fourth confounding variable is that changes in monitoring practice not related to the CSM-PIP interventions can shift error rates. A potential example of this with regard to A16 observations is described in Chapter 2.

Low Levels of Errors

The data in Chapter 2 demonstrate that the absolute level of observations across the system is low with 1.5 observations per site per month and 7.5 observations per protocol per month. Thus, it might be difficult to see a change in errors due to CSM-PIP or any other intervention because of a signal to noise problem. This is particularly true when one considers that accounting for many of the confounding variables noted above would require examining error rates on a site and protocol specific basis.

Lack of Correlation between Errors and Protocol Risk

Because RBPMP shifts monitoring effort to the protocols where the consequences of errors are judged likely to be greatest, the potential for examining error rates for high-risk versus low-risk protocols was considered as a means to evaluate the impact of this intervention. However, for the reasons described in Chapter 2, Section E.5 of this report, it was not feasible to perform an exploratory analysis of this type and the number of confounding variables renders such an analysis unlikely to be useful for evaluation.

Little Added Value from Stakeholder Interviews

STPI researchers considered the option of conducting a formal qualitative evaluation of CSM-PIP through stakeholder interviews. The logical rationale behind the CSM-PIP initiatives is sound and interviews might reveal whether the theoretical

NIAID CONFIDENTIAL

advantages are seen in practice. For example, it was clear from the stakeholder interviews described in Chapter 3 of this report that MRD was viewed positively by both the monitors and site staff. Therefore a comprehensive set of interviews to more robustly identify the advantages might provide a qualitative evaluation of MRD effectiveness. However, because the value of MRD to sites and monitors appears to be obvious and well-accepted, such an exercise would probably not be warranted.

With regard to RBPMP, one could potentially interview DAIDS Program Officers, PPD monitors and clinical research site staff to identify specific corrective actions that were taken with regard to high-risk protocols which in the view of the various stakeholders provided an important improvement in either patient safety or data quality that would not have occurred without identification of the problem through enhanced monitoring. While such an interview based evaluation is feasible, because the prevalence of such events is likely to be low and the fidelity with which they could be identified by stakeholders is highly uncertain, such an interview based evaluation is not recommended.

Best Practices in Clinical Trials Monitoring

Based on a comprehensive literature review and discussions with individuals responsible for monitoring practices, the following conclusions were reached by STPI researchers concerning best practices for site monitoring in public and private sector clinical trials. Sponsors employ diverse methods to monitor their clinical trials, reflecting differences in interpretation of regulatory requirements, different degrees of risk-aversion, and variations in resource availability. While several authoritative entities have developed recommendations for best monitoring practices, these recommendations are based largely on expert consensus rather than empirical evidence.

Nevertheless, the following points of consensus were identified by STPI researchers with regard to best practices.

- Risk based monitoring that targets research sites, protocols or data fields which are more critical or more error-prone is gaining acceptance
- Central monitoring can be cost-effective and efficient at identifying certain types of errors
- Experts generally agree that 100% source data verification is not necessary for most trials
- Quality management should be built into the design and conduct of all clinical trials.

NIAID CONFIDENTIAL

Recommendations for Monitoring Improvements

Although not specifically requested by DAIDS/OCSO, STPI researchers, in the course of this feasibility study, identified certain recommendations for potential improvements in the monitoring program which are described briefly below.

Develop Management “Dashboard” of Monitoring Results

The DAIDS-ES system currently collects detailed real time data on monitoring results which is captured by observation code, site, protocol, site visit date and observation date. Based on the analyses of monitoring data performed for this feasibility study, STPI researchers recommend that these data be converted into a routine, user friendly series of reports, data tables or graphs stratified by site, protocol, network and observation code and updated in near real time (e.g., monthly). These could be made available through a “dashboard” portal and include both current and historical DAIDS-ES data.

DAIDS/OCSO Program Officers could use the dashboard to track every site and every protocol for the most recent monitoring results and for any error trends over time. The dashboard could also be used to target detailed Program Officer review of site monitoring reports only to those sites and protocols where errors warranting follow-up action appeared in the automated reports which should save considerable time. Such an automated DAIDS-ES reporting system would also provide DAIDS and network management the ability to easily track errors over time for both the overall system and for the various networks. Negative trends could be investigated for underlying causes and positive trends could be investigated as a possible source of best practices.

Implement Adaptive Risk-Based Site Performance Monitoring

All DAIDS monitoring program stakeholders agree that targeting monitoring efforts to those sites most likely to make errors would be an important process improvement. However, rather than attempting to implement RBSPM as a static system, STPI researchers recommend that the need to focus monitoring efforts on the most risky sites should be a driving force behind developing the automated monitoring data reporting system described in the previous recommendation. If such an automated system were in place, sites that are experiencing a high level of errors could be readily identified and more intense monitoring efforts directed at those sites. But such a system would also allow identification of sites whose performance was beginning to decline and those which had improved their performance and therefore no longer need an increased level of monitoring.

NIAID CONFIDENTIAL

Adapt Monitoring Effort to Trial Activity

DAIDS monitoring visits are driven by a “one size fits all” approach that targets the review of records for 20 subjects (i.e. patient IDs or PIDs) at each site each quarter, independent of the number of subjects on study or the number of active protocols at the site. STPI researchers recommend that monitoring effort be deployed in response to trial activity rather than on this fixed schedule. In this approach, each site would be examined quarterly to determine, based on the protocols and subjects on study in each protocol, how many PIDs need to be reviewed at that site in the upcoming visit to maintain the required IC/EC and full record review percentage monitoring goals. The number of monitors and days on site could then be established based on that number of PIDs. Such an adaptive approach to monitoring would help ensure that the detection of errors was in line with the opportunity for errors (i.e., the number of protocol events that occurred) at the various sites.

Although this approach should not change the absolute level of the overall DAIDS monitoring effort, it would require two process changes. First, the PID selection algorithm would need to incorporate the total subjects on each protocol at each site, the total PID reviews required to meet the monitoring goals for each protocol at each site, and the number of PID reviews already completed for each protocol at each site. These data would be used to determine the number of PIDs to be reviewed during a visit to maintain the monitoring goals. Secondly, the monitors would need to move from a stable schedule of four-day quarterly visits to each site to a more irregular schedule of trips with varying duration.

Integrate Monitoring and SDMC Data

Integration of monitoring data with SCMC data, enabled by development of the automated monitoring data reporting system described in the first recommendation, would facilitate identification of sites or protocols with both monitoring and SDMC data quality issues. In addition, a spike in negative results in the real time SDMC data for particular sites could alert Program Officers to focus greater monitoring attention on those sites.

Eliminate Monitoring Duplication of SDMC Data Quality Review

Two of the most prevalent errors in the monitoring data supplied to STPI were missed tests and missed visits, representing 21% and 13% of total observations. However, missed tests and missed visits are tracked by the SDMCs in real time for 100% of subjects and the sites are contacted immediately for resolution. Because there appears to be no added value to re-identifying these errors for a small percentage of subjects often long after the event occurred, STPI researchers recommend that identifying missed visits and missed sites be eliminated from the monitoring requirements.

NIAID CONFIDENTIAL

Prioritize PID Review for High-Risk Protocols

In current practice, the order of PIDs on the list given to the monitors is not prioritized. Because DAIDS has developed a process for identifying high-risk protocols that are to receive a more intense level of review, STPI researchers recommend that the protocol order on the PID list should be high-risk protocols first followed by medium- and low-risk protocols.

Eliminate Record Reviews Not Counted Toward Monitoring Goals

The current DAIDS-ES PID selection algorithm is designed such that not all PID records reviewed actually count toward meeting the monitoring goals for full record review. In every round of the algorithm for a given protocol at a given site, the entire pool of IC/EC reviewed PIDs is available for selection for ongoing record review. The result is that when the full record review monitoring goal for a protocol is met, there are a number of partially reviewed PIDs from that protocol which do not count toward meeting the monitoring goal. Although review of these “extra” PIDs provides value in reviewing more records for errors, that value is distributed randomly across protocols.

There are two potential solutions to this problem that STPI researchers recommend for DAIDS consideration. The first would be to restrict ongoing record review to PIDs previously selected for ongoing review. This change in practice would have two advantages. It would allow increased monitoring goals as all PID reviews would count toward the goal. In addition, the number of PIDs reviewed for each protocol would remain correlated with the risk level of a protocol as opposed to being random. There would, however, be some disadvantages. The most important is that the sites would know in advance which PIDs will be reviewed. The second disadvantage is that the records reviewed would tend to be for PIDs enrolled earlier in the life of the protocol at the site and that might introduce a bias.

A preferred, but clearly more complicated, approach would be to select PIDs randomly in groups of study visits to meet the ongoing review monitoring goal for each group of study visits as described in Chapter 6, Section G of this report. This change would have the two advantages described above but it would also have a greater diversity of individual PIDs reviewed and would not suffer from the sites knowing in advance which PIDs are likely to be reviewed. It would still, however, suffer from the disadvantage of reviewing PIDs enrolled earlier in the life of the protocol at the site.

Contents

Introduction.....	1
1. DAIDS/OCSO Clinical Site Monitoring	5
2. Analysis of Monitoring Data	15
3. Stakeholder Perceptions of CSM-PIP Value	37
4. Feasibility of Process and Outcome Evaluation	41
5. Best Practices in Clinical Site Monitoring	47
6. Recommendations for Monitoring Improvements	59
Appendix Stakeholders Interviewed	65
References.....	67

Introduction

In 2008, the Division of Acquired Immune Deficiency Syndrome (DAIDS) of the National Institute of Allergy and Infectious Diseases (NIAID) through its Office of Clinical Site Oversight (OCSO) initiated a clinical site monitoring process improvement program (CSM-PIP). The first of these initiatives, Monitoring More Recent Data (MRD), was designed to focus monitoring efforts on more recent data⁸ and was implemented at all sites in January 2010. The second initiative, designated Risk-Based Protocol Monitoring Plans (RBPMP), established standardized methods to rank protocols according to risk and assigned increased levels of monitoring to higher risk protocols. This initiative was implemented at all sites in April 2010. A third initiative, Risk-Based Site Performance Monitoring (RBSPM), designed to allocate increased monitoring resources to sites judged to be at higher risk of poor performance has not yet been implemented. All of these initiatives were designed with the goal of improving site performance in terms of a reduced incidence of errors, either for all protocols or for those protocols where the consequences of errors might be greater.

In September 2011, the IDA Science and Technology Policy Institute (STPI) was tasked by DAIDS/OCSO with analyzing the feasibility of conducting an integrated process and outcome evaluation to determine whether the first two CSM-PIP initiatives have had an impact on site performance. The feasibility study was designed to address various topics including the evaluation questions to be asked, the data that would be required to answer those questions, and a plan for conducting the evaluation if it were deemed feasible. In addition, the study was to provide an analysis of best practices for site monitoring in public and private sector clinical trials through a comprehensive literature review and discussions with individuals responsible for monitoring practices.

The DAIDS clinical trial program involves approximately 100 active protocols conducted across five primary clinical trials networks⁹, encompassing approximately 500 domestic and international sites and involving approximately 20,000- 25,000 subjects on

⁸ MRD originally limited record review beyond informed consent and eligibility to data collected only in the 12 months prior to the monitoring visit but, at the request of the Food and Drug Administration (FDA), that was increased in January 2011 to data collected in the three years prior to the monitoring visit.

⁹ AIDS Clinical Trials Group (ACTG), HIV Prevention Trials Network (HPTN), HIV Vaccine Trials Network (HVTN), Microbicide Trials Network (MTN) and International Maternal Pediatric Adolescent AIDS Clinical Trials (IMPAACT).

NIAID CONFIDENTIAL

study annually. The majority of clinical trial sites are monitored quarterly with the specific protocols and subjects for which records are monitored determined primarily by either the protocol risk-based algorithm described in Chapter 1 of this report or by protocol-specific monitoring plans which supersede the algorithm for certain protocols. The monitoring data are collected centrally by the DAIDS-Enterprise System (DAIDS-ES) and the results reviewed and responded to primarily through OCSO Program Officers working directly with sites.

The feasibility analysis involved seven primary activities

1. Analysis of DAIDS/OCSO documents concerning the scope and nature of both the clinical trials program and the monitoring program as currently implemented including templates, Standard Operating Procedures, model reports, etc.
2. Interviews with DAIDS/OCSO staff and contractors involved with site monitoring to identify existing monitoring data/information that might be used to determine if there is improvement in site performance due to CSM-PIP and/or to identify data/information that should be collected in order to determine if site performance is improving.
3. Interviews with DAIDS clinical research site staff to gain information their perspective on the monitoring process and the impact, if any, that CSM-PIP has had on the process.
4. Interviews with representatives from the DAIDS Network Statistics and Data Management Centers and the HVTN Evaluation Committee to determine if these entities, which are not directly involved in monitoring activities, might nonetheless have data/information on site performance which could be used to identify improvements due to CSM-PIP.
5. Analysis of the operation of the algorithm used to determine which protocols and subject records are reviewed at each monitoring visit.
6. Analysis of the structure of the monitoring data collected by DAIDS-ES.
7. A pilot analysis of DAIDS-ES site monitoring data from January 2008–June 2012 to determine if such data would be useful in establishing whether site performance changed with implementation of CSM-PIP.

This report consists of six chapters. The first chapter provides a description of the DAIDS/OCSO clinical site monitoring program, including the CSM-PIP initiatives, as essential context for understanding the findings and conclusions in the remainder of the report. The second chapter describes the results of analyzing the monitoring data in DAIDS-ES and the potential that such data analyses can provide information on the effect of the CSM-PIP initiatives on site performance. Chapter 3 summarizes perceptions gathered from stakeholders on the value of the CSM-PIP initiatives. Chapter 4 addresses

NIAID CONFIDENTIAL

the feasibility of conducting a process and outcome evaluation of the effect of the two CSM-PIP initiatives implemented to date based on the analyses presented in Chapters 2 and 3. Chapter 5 describes current best practices for clinical site monitoring based on a literature review and discussions with individuals responsible for monitoring practices in other organizations. Finally, although not specifically requested by DAIDS/OCSO, Chapter 6 presents recommendations for potential improvements in the monitoring program. These recommendations arose directly from the comprehensive, multi-faceted analyses of both current monitoring practices and the monitoring data currently collected and STPI researchers considered it important to share these insights as part of this report.

1. DAIDS/OCSO Clinical Site Monitoring

DAIDS/OCSO clinical site monitoring is primarily accomplished through onsite visits, which are conducted by Pharmaceutical Product Development (PPD), a contract research organization, and overseen by OCSO Program Officers. Beginning in 2009, DAIDS/OCSO began to implement a series of process improvement initiatives aimed at improving site performance.

A. Overview

1. Scope of OCSO Monitoring Responsibilities

DAIDS/OCSO oversees over 500 domestic and international clinical research sites that accrue subjects to approximately 100 active clinical trial protocols at any given time. Most of the trials monitored by OCSO are sponsored by one of the five DAIDS clinical trials networks:

- AIDS Clinical Trials Group (ACTG)
- HIV Prevention Trials Network (HPTN)
- HIV Vaccine Trials Network (HVTN)
- Microbicide Trials Network (MTN)
- International Maternal Pediatric Adolescent AIDS Clinical Trials Group (IMPAACT)

In addition to trials sponsored by these networks, OCSO monitors some out-of-network trials conducted in partnership with other organizations, depending on the specific agreement of the partnership, as well as some investigator initiated trials.

2. Organization of OCSO

DAIDS OCSO consists of four branches:

- Monitoring Operations Branch
- Asia and the Americas Branch
- Africa and the Domestic Partners Branch
- Pharmaceutical Affairs Branch

NIAID CONFIDENTIAL

The Monitoring Operations Branch is broadly responsible for oversight of monitoring operations. Program Officers within the two regional branches are responsible for overseeing a specific cohort of domestic and international clinical research sites. Program Officers review monitoring reports and work with site leadership to improve site performance based on the errors reported. In addition, Program Officers are responsible for ensuring that their assigned sites are adequately monitored and have some discretion to increase or decrease monitoring intensity on protocols conducted at their sites. The Pharmaceutical Affairs Branch oversees pharmacy operations, which is not directly relevant to clinical site monitoring.

3. Monitoring Program Fundamentals

Monitors generally visit sites on a quarterly basis, although sites with little activity may be visited less often but never less than once per year. If there are serious problems at a site, it may be visited more frequently, although in practice this rarely occurs. During a monitoring visit, the monitor reviews a pre-determined selection of subject records. Depending on the protocol and subject, this review may only involve confirming that the subject met the eligibility criteria and has properly documented informed consent or may involve reviewing all records thorough the date of the monitoring visit that have not been previously reviewed.

Full record review consists primarily of confirming the following: protocol compliance in terms of study visits, tests, and proper drug administration; reporting and documenting all clinical events; and correspondence between case report forms and medical records or other source documents. During a site visit, monitors may also review regulatory files, pharmacy records, laboratories, site operations, and conduct other special assignments but 75-80% of the monitoring visit is spent on subject record review.

Each monitoring visit generally involves four days (32 work hours) onsite and targets record review for 20 subjects. It is estimated that typically monitors complete review for about 80% of the assigned subjects. On occasion, multiple monitors are sent on a single visit.

At the end of each visit, the monitor briefs site staff on the findings and may assist in developing any necessary corrective action plans. After the visit, the written site monitoring report is uploaded to DAIDS-ES and transmitted to the DAIDS Program Officer and site staff. The Program Officer works with the site staff to resolve any substantial issues and/or develop a corrective action plan which must be approved by the Program Officer.

NIAID CONFIDENTIAL

4. Other DAIDS Processes Affecting Site Performance

In addition to the OCSO monitoring program, site performance can also be affected by feedback on data quality from the Network Statistics and Data Management Centers (SDMCs) and the results of the annual Network Performance Evaluation Committee review of each site. The DAIDS SDMCs conduct a wide variety of automated checks to identify data quality issues such as delinquencies in data submission, missing data, data entered on subjects not enrolled in the study, values outside the acceptable range for a data field, and data indicating illogical form completion. Any problems in data quality thus identified are communicated in real time to the sites for resolution.

Although these SDMC data quality checks primarily examine factors not covered by the monitoring program, a deficiency in data quality could signal a decrease in performance at a site which might also be reflected in the monitoring data in terms of lack of proper documentation of clinical events or the lack of correspondence of reported data with source documents. It is also worth noting that two items, missed visits and missed tests, are picked up by both the SDMC data quality review and the monitoring program. Since the SDMCs identify these errors for 100% of the sites and pursue resolution in real time, the added value of monitoring these items is unclear.

The Network Performance Evaluation Committee annual review of each site includes factors such as data timeliness, data quality, and resolution of identified errors, which are provided by the SDMC, as well as “adherence to protocol” factors which rely on monitoring data. The most important monitoring data for the Network Evaluation Committee reviews are reported to be conformance with eligibility criteria, proper informed consent, problems with drug administration and unreported or inadequately documented clinical events, endpoints or adverse events.

B. Process Improvement Program

In 2009, OCSO began to pilot a series of process improvements to its clinical site monitoring program (often referred to as CSM-PIP) with the goal of utilizing monitoring resources more effectively to improve site performance as evidenced by reduced error rates. The process improvement program included three initiatives. The first initiative focused monitoring efforts on the review of more recent data. The second initiative deployed greater monitoring effort on protocols judged to be of higher risk. The third initiative is to monitor more intensely those sites which are judged to be at higher risk of making errors. The first two of these initiatives have been fully implemented while the third is still in the design stage.

1. Monitoring Recent Data (MRD)

Monitoring Recent Data (MRD) was the first process improvement initiative to be implemented. The goal of MRD is to ensure that monitoring findings reflect recent site

NIAID CONFIDENTIAL

performance as correction of recent errors is more likely to improve future performance. MRD was developed partially in response to complaints from sites that monitors were reviewing older records and, therefore, sites could not take any corrective action for many of the observations.

When launched in the first quarter of 2010,¹⁰ MRD restricted record review to subject visits within 12 months of the monitoring visit although eligibility, enrollment, and informed consent data were reviewed regardless of when these data were collected. Responding to FDA concerns that MRD may miss important data, OCSO revised MRD in early 2011 to include all visits for subjects enrolled within the last three years. For subjects enrolled three years prior to the monitoring visit, only informed consent and eligibility criteria are reviewed.

Protocol specific monitoring plans (described in Section B.4 of this chapter) may supersede MRD and require review of a different scope of records. In addition, the first two subjects enrolled on each protocol at a site are monitored through the duration of the study or through the target monitoring visit, if applicable. MRD never limits record review for the first two enrolled subjects.

2. Risk-Based Protocol Monitoring Plans (RBPMP)

The goal of RBPMP is to monitor the “most risky” protocols more intensively while monitoring the “less risky” protocols less intensively. In implementing this initiative, DAIDS developed a standard set of criteria to evaluate protocol risk, which reflects both the potential for harm to study subjects from errors and the potential for errors to adversely affect data quality. These criteria include factors such as trial phase, IND status, potential toxicity, protocol complexity, and others.

Under RBPMP, which was implemented in September 2010, protocols receive a risk rank from level 1 (highest risk) to level 3 (lowest risk) which is assigned by the protocol medical officer based on the standard risk criteria. Some protocols, generally observational studies, are assigned a risk level of 0 and are not monitored at all.¹¹ The risk ranking of a protocol directly determines the degree to which the protocol is monitored in terms of the percentage of subjects that are reviewed for eligibility criteria (EC) and informed consent (IC) as well as the percentage that are subject to full record review. “Monitoring goal” ranges for the various risk levels have been set by DAIDS/OCSO as shown in Table 1.

¹⁰ MRD was first piloted in 19 sites from July 2009–December 2009.

¹¹ Risk level 0 is synonymous with risk level 99, which is seen in some documentation.

NIAID CONFIDENTIAL

Table 1. Monitoring Goal Ranges by Protocol Risk Level

Risk Level	Review for IC & EC	Full Record Review
Level 1	50-100%	25-50%
Level 2	20-50%	15-30%
Level 3	10-20%	10-20%
Level 0	0%	0%

Protocols that do not have protocol specific monitoring plans have as a monitoring goal the minimum percentage of the range for the assigned risk level. Protocols with protocol specific monitoring plans may have a monitoring goal of up to the maximum of the range for the assigned risk level. Protocol specific monitoring plans may not assign a monitoring goal outside of the range specified by the assigned risk level for the protocol.

In addition to defining the monitoring goal, protocol risk ranks determine how often protocols are selected for monitoring (see Section C.2 of this chapter).

3. Risk-Based Site Performance Monitoring (RBSPM)

The goal of RBSPM, which has not yet been implemented, is to monitor more intensely those sites judged to be at higher risk of making errors. DAIDS piloted a site ranking process in which they had Program Officers use standard documentation on site performance to rank the risk level for sites with which they were not familiar. The Program Officers had difficulty assigning site risk based on the documentation and refinements to the site risk criteria and risk ranking process are under discussion.

4. Protocol Specific Monitoring Plans

Protocol specific monitoring plans allow for customizing the intensity and duration of monitoring to meet the needs of individual protocols. These plans may stipulate one or more of the following: identify a target end date for monitoring; set a specific monitoring goal within the range for the risk level of the protocol; override MRD restrictions on the extent of record review; stipulate increased monitoring resources (such as the presence of multiple monitors per site visit); and provide special instructions, such as assigning different monitoring goals for infected and uninfected subjects. About 20-25% of DAIDS clinical trials have protocol specific monitoring plans.

C. Selection of Subject Records for Monitoring

The process and criteria used for selecting the subject records to be monitored form the foundation of the DAIDS/OCSO clinical site monitoring process and it is this selection of records that has been impacted by the process improvement initiatives. It is therefore essential to understand the selection process and criteria before and after these changes. In the DAIDS/OCSO system, the subjects selected for record review are

NIAID CONFIDENTIAL

referred to as “PIDs” (patient IDs) and this terminology will be used in this report for convenience.

1. Selection of PIDs Prior to CSM-PIP and the DAIDS-ES Algorithm

Prior to the initiation of CSM-PIP, the PPD monitor made the decision as to which active protocols would be reviewed at a given site visit and how many PIDs would be reviewed for each study both for IC/EC only and for full record review. There were no set criteria for which protocols should be monitored except that “never monitored” studies always had the highest priority. This meant that a protocol would always be reviewed at the first monitoring visit after it became active at the site but the number of PIDs reviewed for that protocol was at the discretion of the monitor. There were also no established monitoring goals in terms of percentage of PIDs reviewed and in theory 100% of the PIDs on a protocol at a site could have been reviewed for both IC/EC and full record review through completion of the study. The number of protocols and PIDs specified was driven largely by the number of records that the monitor believed could be reviewed during the scheduled duration of the site visit.

Once these selections were made, a Site Visit PID List was generated. The PID list was constructed so that reviews of records for deaths and withdrawals, follow up issues from previous site visits, and PIDs for which records were not available at a previous site visit were handled first, independent of the protocol in which they occurred. The next priority was the PIDs randomly generated for “never monitored” protocols up to the number specified by the monitor. The randomly generated PIDs for the other protocols selected by the monitors were next on the list grouped by protocol (Random Review PIDs). After this list of Random Review PIDs was generated the monitor had the option to add or remove PIDs primarily driven by protocol specific directions from DAIDS. There was also the option to add “custom PIDs” which might be added at the discretion of the monitor to satisfy a special monitoring need.

Once on site, the monitors completed as many of the PIDs on the list as possible given the scheduled duration of the visit. PIDs whose records were not reviewed due to time constraints simply reentered the pool for potential random selection for a future site visit. Although the number of PIDs reviewed at a given site visit varied widely, it typically was in the range of 10 PIDs.

2. Selection of PIDs Post CSM-PIP and the DAIDS-ES Algorithm

Just prior to implementation of the CSM-PIP interventions, DAIDS developed an algorithm to select PIDs for review at each monitoring visit. The algorithm selects protocols for review and then PIDs within each protocol. The algorithm is now mainly based on protocol risk level and protocol specific monitoring plans, although MRD does have an effect.

NIAID CONFIDENTIAL

The only PID review assignments that are not governed by the algorithm are PIDs with follow up issues from previous site visits and PIDs for which records were not available at a previous site visit, both of which take precedence in review priority over the PIDs generated by the algorithm. After the PIDs randomly generated by the algorithm are reviewed, the monitors then review any custom PIDs which they have selected to meet protocol specific monitoring goals or other special needs and to complete any special assignments given to them by the DAIDS/OCSO Program Officers. There are, in addition, several scenarios in which the algorithm can be manually overridden by a DAIDS Program Officer or a PPD monitor.

a. Protocol Selection

For a given monitoring visit, the algorithm first selects between three and six protocols for review. Eligible protocols are those with a monitoring status of either “pending” or “on-going” that have not yet reached their monitoring goals¹² or been manually removed from the pool by a DAIDS Program Officer. Protocols are selected based on their assigned risk levels and the date of last review. At a minimum,¹³ one risk level 1 (high risk) protocol is selected for every site visit, one risk level 2 (moderate risk) protocol is selected for every second visit, and one risk level 3 (low risk) protocol is selected for every third visit. For sites with multiple protocols at a given risk level, priority is given to protocols that have the oldest date of their last monitoring review. The overall life of the protocol at the site is not considered.

Once three protocols have been selected, the algorithm determines how many PIDs are eligible for review (including for both IC/EC and full review) from those protocols combined. If the number of PIDs eligible for review is greater than or equal to 20, no additional protocols are selected for review. If the number of eligible PIDs is fewer than 20, additional protocols (up to six) are selected until there are at least 20 eligible PIDs.

b. Specific PID Selection

Once protocols have been selected for review, the algorithm selects PIDs from each protocol for one of three categories of review.

1. **Follow-up through completion.** Records for the first two enrolled PIDs on a protocol at a site are reviewed through the target study visit for that protocol every time the protocol is reviewed at the site.
2. **IC/EC.** PIDs whose records are reviewed for at least informed consent and eligibility.

¹² Either specified by their risk level or a protocol specific monitoring plan

¹³ Unless the site has no eligible protocols at the respective risk levels

NIAID CONFIDENTIAL

3. **General follow-up.** PIDs whose records are reviewed through the date of the monitoring visit, or the target study visit for the protocol, whichever is earlier.

To select individual PIDs for review, the algorithm starts with the first category, “follow-up through completion,” and places the first two enrolled PIDs from each protocol being reviewed on the monitoring assignment list unless these PIDs have been reviewed through the target study visit. Next, the algorithm selects one PID from each protocol that has not had IC/EC review. Finally, the algorithm selects one PID from each protocol that has already completed IC/EC review to have full record review. If at this point fewer than 20 PIDs have been selected, the algorithm picks one more PID from each protocol that has not had IC/EC review. If the number of PIDs selected is still fewer than 20, the algorithm selects one PID from each protocol for full record review. If the number of PIDs is still less than 20, the algorithm cycles between selecting PIDs for IC/EC and full record review. If at any point 20 PIDs have been selected, the algorithm ceases PID selection and finalizes the assignment list.

It is worth noting that the algorithm specifically limits the PIDs selected for follow-up review to those PIDs that have already received IC/EC review thus limiting the eligible PIDs to only a specific subset of the total PIDs. PIDs selected for IC/EC review may, at the discretion of the monitor, have all their records through the monitoring visit (or the target study visit) reviewed if the monitoring goal for full record review has not been met and thus they become part of the follow-up review from the outset. However, in some cases, the monitor will decide that a PID selected for IC/EC review has too many records available and will skip those records and move on to the records of another PID on the list.

Once the monitoring goals for either IC/EC or full record review have been met, no further PIDs will be selected for that category of review. In addition, PIDs with a status of “Off Study” as of one year prior to the monitoring visit who have had IC/EC review will not be selected for general follow up review. Finally, MRD influences the algorithm in that PIDs enrolled more than three years ago are excluded from the general follow up PID pool although they remain eligible for IC/EC review.

c. Final Algorithm Output

The final output of the PID selection algorithm is a list of 20 PIDs to be reviewed by the monitor onsite. The list is organized by protocol number and then by PID number. Protocols are not prioritized for review by risk level and neither protocols nor PIDs are prioritized in any other systematic way. The list does indicate whether each PID was selected for IC/EC or full review. Both types of PIDs are displayed on the same list.

NIAID CONFIDENTIAL

d. Tracking Progress toward Monitoring Goals

Tracking progress toward reaching the monitoring goals for each protocol is the responsibility of the DAIDS Program Officers. In order to track monitoring progress, Program Officers study the Comprehensive Record Review Summary (CRRS) report for each protocol. These reports provide information on the number and percentage of PIDs reviewed for IC/EC and full record review through the target study visit. The CRRS report also provides information on the total number of PIDs reviewed. In other words, it provides information on the number of PIDs whose records have been partially reviewed. Another resource available to Program Officers is the generic PID report, which lists all PIDs reviewed at each visit for a given site and indicates through which study visit the PID file was monitored.

Only PIDs whose records are reviewed through the target study visit for the protocol count toward the monitoring goal. However, there is no priority given in selecting PIDs for general follow up review to those PIDs who have had a previous general follow up review which did not reach the target study visit. Therefore, it is highly likely that many PID records are reviewed that do not count toward achieving the monitoring goals. Although the data is available, DAIDS has never determined the number and percentage of these partially reviewed PIDs once the monitoring goal has been reached. Therefore, the extent of this untargeted monitoring effort is not known.

e. Manual Algorithm Overrides

Some situations require either DAIDS Program Officers or PPD monitors to override the DAIDS-ES PID selection algorithm. These situations arise for two reasons.

1. The automated algorithm lacks critical information
2. Monitoring on a certain protocol is deemed sufficient and complete.

Examples of the first case include protocols with specific monitoring plans that rely on clinical data or that evolve through the life of the protocol. DAIDS-ES does not know the schedule of events for protocols and does not have access to clinical data. For example, a protocol that requires a different level of monitoring for infected patients compared with uninfected patients will require manual selection of PIDs for review. In these cases, the PPD monitor will create a custom PID list. Custom PID lists are not transmitted to DAIDS and are not entered into the DAIDS-ES system. The only information transmitted to DAIDS-ES is the list of PIDs actually reviewed.

In the second scenario, DAIDS Program Officers can, after reviewing the monitoring status of a protocol, decide that a protocol has had sufficient monitoring and no longer needs regular review even though the pre-selected monitoring goals have not been met. The Program Officer will then manually remove the protocol from the pool from which the algorithm selects protocols and PIDs.

2. Analysis of Monitoring Data

A. Introduction

An analysis of DAIDS/OCSO monitoring data was performed to determine whether these data can be used to assess the impact of the CSM-PIP interventions on site performance as reflected by error rates. If such an assessment was determined to be feasible, the goal was to define metrics and analytic procedures that could be used to conduct such an assessment. If such an assessment were determined not to be feasible, it was important to understand the reasons why, and to assess possible alternative approaches.

The analysis was performed in two stages. First, the general characteristics of the available data were determined through discussions with DAIDS/OCSO program staff and DAIDS-ES staff. This allowed provisional definition of possible metrics relevant to site performance and development of a conceptual framework for analysis. Second, building on this conceptual analysis, a number of exploratory analyses were conducted using monitoring and subject accrual data provided by DAIDS. These analyses enabled STPI researchers to characterize the available data in a more concrete and detailed way and to draw conclusions about the feasibility of using these data as the basis for an integrated process and outcome evaluation of the impact of the first two CSM-PIP initiatives on site performance.

B. Conceptual Framework

The conceptual framework for the analysis was built around three fundamental components: a basic performance metric for clinical sites, the rollout schedule for the CSM-PIP interventions, and a set of comparisons/controls intended to enable assessment of the impact of the interventions.

The fundamental unit of data relevant to site performance that is recorded by site monitors and compiled centrally in DAIDS-ES is the observation code. Observation codes perform two related functions. The first function is to provide a tally of the number of observed anomalies in a clinical site's adherence to protocol or in the completeness or accuracy of data submitted by the site through a study's case report forms (CRFs). The second function is to classify the observed anomalies into a set of qualitative categories. Thus, the rate of occurrence of observation codes, which can be interpreted as an error rate for the site, was chosen as the basic site performance metric. In principle, a variety of more specialized metrics can be derived from the observation code data, including rates

NIAID CONFIDENTIAL

of particular types of error (occurrence of certain observation codes), rates of errors on particular protocols, etc.

The time period for analysis was defined as January 2008 through June 2012, to encompass a baseline period, both before and after introduction of the DAIDS-ES PID selection algorithm, as well as the phased introduction of the CSM-PIP interventions. The overall analysis interval was broken into several observation periods corresponding to the following stages.

Period 0A	Jan-Dec 2008	Baseline
Period 0B	Jan-Jun 2009	DAIDS-ES PID selection algorithm implemented
Period 1	Jul-Sept 2009	MRD ¹⁴ implemented at domestic pilot sites
Period 2	Oct-Dec 2009	MRD ¹⁴ implemented at international pilot sites
Period 3	Jan-Mar 2010	MRD ¹⁴ implemented at all sites
Period 4	Apr-Dec 2011	RBPMP implemented at all sites
Period 5	Jan 2011-Jun 2012	MRD ¹⁵ and RBPMP implemented at all sites

CSM-PIP was not implemented such that there was a control group of sites where the monitoring process was not changed. However, with regard to the roll out of MRD, the implementation sequence did result in a group of “natural” intervention cohorts.

- All domestic pilot sites
- All international pilot sites
- All other sites

In principle, one might compare error rates among these cohorts, especially within Periods 1, 2 and 3, when MRD was being rolled out. The clinical sites could also be stratified along other dimensions, such as clinical trial network or domestic versus foreign location. Finally there is the opportunity to examine error rates across time, within individual cohorts and for the population of clinical sites as a whole.

C. Methodological Challenges

Critical examination of the conceptual framework from a theoretical perspective revealed a variety of potential challenges in constructing metrics that would be valid measures of the impact of the CSM-PIP interventions on site performance. This section

¹⁴ From July 2009 through December 2010, the MRD intervention specified that review of on-study visits was generally limited to the 12 months prior to the monitoring visit.

¹⁵ After January 2011, the MRD intervention was modified at the request of FDA such that review of on-study visits was extended to three years prior to the monitoring visit.

NIAID CONFIDENTIAL

briefly describes these challenges and the approaches taken in the exploratory data analysis to understand their potential impact and, where possible, mitigate them.

1. Calculating Error Rates

Although the site performance metric was defined as the error rate, the information compiled by site monitors and recorded in DAIDS-ES is just the observation codes themselves—in effect, error *counts* rather than error *rates*. What is missing is a suitable denominator by which error counts can be divided in order to calculate how many errors occur *relative to the number of opportunities for an error to occur*.

In principle, in the context of site monitoring, the number of opportunities for an error to occur is related to the number of protocol events¹⁶ (e.g., patient consent and enrollment actions; medication doses, procedures, tests, and other clinical events; data collection items specified by the protocol) reviewed by the site monitors. Thus, the ideal denominator for an error rate calculation would be the number of protocol events reviewed by the monitors. However, this parameter is not captured by the monitors. This is understandable, as doing so would impose a substantial operational burden, consuming time otherwise used to evaluate data elements for errors and thus reducing the number of data elements that can be reviewed.

STPI researchers investigated several “proxy denominators” in terms of their availability and the degree to which they could be used in place of the number of protocol events reviewed. In theory, study protocols could be analyzed in conjunction with the number of subjects on study to determine the number of scheduled protocol events for a given observation period. However, such an analysis is impractical for at least three reasons. First, it would be labor intensive. Second, the number of protocol events that actually occur frequently departs from the study plan in unpredictable ways as a function of study enrollment and subject and study site behavior. Third, site monitors review only a subset of actual protocol events and this fraction also varies unpredictably.

A more realistic proxy denominator is simply the number of subjects on study at the site during the monitoring visit as these data are readily available. However, the number of subjects on study is a reasonable proxy only for protocol events tied directly to the number of subjects such as informed consent. Due to variation in study design and patient follow-up, the number of other protocol events per subject varies substantially across and even within protocols. Nevertheless, in some the exploratory analyses in this report, the utility of this proxy denominator was investigated.

¹⁶ Most of these protocol events actually define at least two actions in which there may be the opportunity for an error of commission or omission to occur—the clinical or administrative event itself must actually happen and that event must be recorded correctly and completely in the CRF.

NIAID CONFIDENTIAL

A third alternative proxy denominator is time, on the assumption that the total number of protocol events reviewed across the system as a whole remains reasonably steady over time. This parameter is conveniently available without requiring special data collection, and is the denominator used for most of the analyses presented in this report.

2. Sampling Bias

Because monitor time on site is limited, the monitors cannot exhaustively review all subject records and all protocol events associated with those subjects. They necessarily examine only a sample from the universe of subjects and protocol events. Introduction of the new DAIDS-ES PID selection algorithm and the CSM-PIP interventions amounted to a change in sampling approach compared to the monitor directed sampling employed previously. The PID selection algorithm substantially standardizes the selection of protocols and PIDs reviewed across all sites, while MRD eliminates the review of older records within a protocol and RBPMP changes which protocols are reviewed and the percentage of PIDs reviewed within each protocol.

This change in the sampling approach creates a challenge in using errors rates to analyze the impact of the CSM-PIP interventions on site performance. If there are pre-existing, systematic variations in error rates over time either within a protocol or across protocols, a change in the sampling method from a time perspective can lead to an apparent change in site performance even where there has been no actual change, simply because records with different characteristics have been selected for review. Accordingly, it is important to assess any trends in error rates over time before and after the intervention rather than to rely only on assessment at single time points pre- and post-intervention. Similarly, if there are pre-existing, systematic variations in error rates across different protocols, changing the intensity of review for certain protocols can also lead to an apparent change in site performance when there has been no actual change. A number of the exploratory analyses described in this report were conducted to assess whether such systematic variation was present in the data set.

3. Attribution

One of the greatest challenges in using error rates to assess the impact of the CSM-PIP interventions on site performance is the issue of attribution. In principle, the phased introduction of MRD and RBPMP provide at least some opportunity for internal experimental controls pre- and post-CSM-PIP implementation which might assist in assessing attribution. However, these controls may be inadequate if the actual effects of the intervention are small relative to random variation over time and across sites and protocols. Additionally, detecting a “signal” may be difficult if the time required for the intervention to have its effect is long compared to the phase-in timeline. For example, there may not have been sufficient time for the MRD pilot sites to demonstrate a change

NIAID CONFIDENTIAL

in error rates before MRD was implemented at all sites, rendering a comparison of sites with and without MRD uninformative. Because of the many un-controlled and non-measurable influences on error rates, any attribution of a change in error rates to the CSM-PIP interventions would need to be based on an informed, subjective assessment that combines objective data on observed error rates with what is known qualitatively about the behavior of the system.

D. Exploratory Analyses

1. Purpose

As explained in Section C of this chapter, there are theoretical concerns about whether it would be possible to calculate accurate error rates from clinical site monitoring data and whether observed changes in error rates are real or could be attributed to the CSM-PIP interventions. However, from the theoretical analysis alone the actual extent of any biases and confounders, and whether they would be truly debilitating, could not be known.

Accordingly, and as a central component in analyzing the feasibility of a formal process and outcome evaluation, a series of exploratory analyses of DAIDS-ES monitoring data were conducted. The primary purpose of this analysis was two-fold: (a) to determine whether any distinct overall trends could be observed that might plausibly be interpreted as effects of CSM-PIP and (b) whether there are any underlying patterns in the data that validate or dispel the theoretical concerns.

2. Data

DAIDS/OCSO and the DAIDS-ES contractor provided two Excel files with data for the exploratory analysis. The first of these files, which will be referred to here as the Monitoring Data File, is a complete file of observation codes recorded by site monitors during the period from January 2008 through June 2012. For each observation, several fields were provided that characterized the observation itself or the site monitoring visit during which it was recorded. The fields used in the exploratory analysis were the following:

- Site ID
- Protocol ID
- Site visit start date
- Observation code
- Observation description

NIAID CONFIDENTIAL

Each observation was also assigned by STPI researchers to one of the five DAIDS networks based on the protocol ID. As each network uses a unique protocol ID format, a simple alphanumeric sort on the protocol IDs enabled them to be separated into blocks of protocols by network. To avoid double counting, in the minority of cases where two IDs were assigned to a protocol, indicating shared sponsorship of a study, the first listed ID was used to assign the protocol to a single network.

The second file, which will be referred to here as the Subjects on Study Data File, tracked patients on study by month during the January 2008 through June 2012 interval. The fields used in the exploratory data analysis were the following:

- Protocol ID
- Protocol network
- Month
- Year
- Number of subjects actively on study for specified month/year

Two “corrections” were applied to the data provided by DAIDS-ES prior to analysis. The first was to remove A15 observation codes that were used during the MRD pilot to simply document records that were excluded from review because of the new selection criteria. The second correction was to exclude monitoring data for INSIGHT network studies. Monitoring of these studies was discontinued after Period 0B, and thus before the CSM-PIP interventions were implemented.

3. Analysis Approach

The general approach for the exploratory analyses was to group or stratify the observation code entries in different ways and search for patterns that might constitute evidence for CSM-PIP impact or shed light on underlying characteristics of networks, sites, and protocols that might affect the interpretation of such patterns. Trends in observation code counts per month and in the mix of reported codes were examined over time and analyzed for variation across networks and for selected protocols and sites.

The “backbone” of many of the analyses was a sorting of reported observation codes into the observation periods defined in Section B of this chapter. Although the Monitoring Data File included both the site visit start date and the date when the observation code was recorded, for simplicity each individual observation was assigned by STPI researchers to an observation period based on the site visit start date. Few site monitoring visits straddled observation period boundaries; the fraction of observation codes affected by this simplification was 0.6%.

NIAID CONFIDENTIAL

Observation codes were first analyzed individually, and then alternative approaches to grouping the codes were examined. A distinct subgroup of codes designated “serious codes” was defined based on the codes designated as “Significant Findings Requiring Resolution and Follow-up” in the Standard Operating Procedure (SOP 13) used by DAIDS/OCSO Program Officers in overseeing the monitoring of individual clinical research sites (see Table 2). These codes also correspond closely to the codes considered by the HVTN Network Evaluation Committee to constitute serious errors.

Table 2. Serious Code Categories

Description	Codes
Informed Consent Violations	A1/A51
Enrollment Violations	A2
Unreported Clinical Endpoint	A8
Death, Problems with Reporting	A10
Dosing/Dispensation/Administration Problem	A11/A21
Unreported Clinical Events	A20
Unreported SAE/EAE	A61
Unreported AE	A62
Confirmatory Inadequate Source Documentation	A71

SOP 13 also designates certain additional codes which require resolution and follow-up only if their occurrence is widespread or numerous. These were not included as “serious codes” in our analyses because of the complexity of factoring in frequency of occurrence at every site. In addition, the A1A/A1B and A2A/A2B codes were merged into the parent A1 and A2 codes to maintain continuity across the time series.

E. Findings

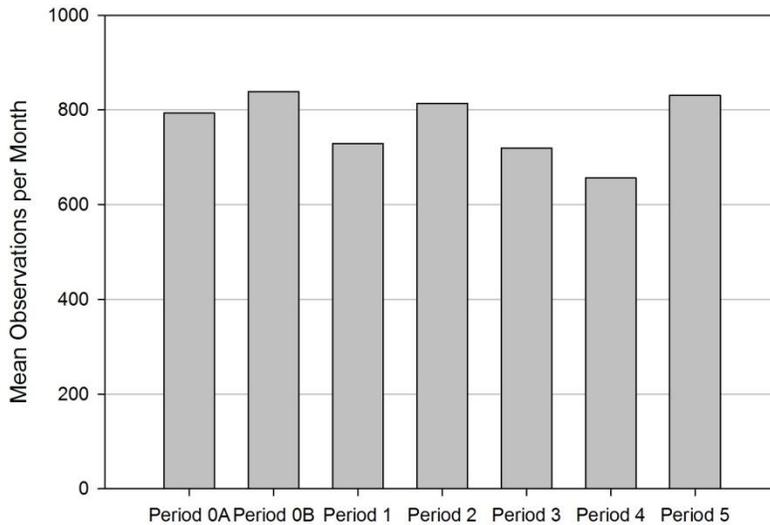
1. Overall Errors

As shown in Figure 1, the overall number of observation counts per month across all five DAIDS clinical trial networks from January 2008 to June 2012 is reasonably stable with a mean of 769 ± 68 observations per month. These data on the surface would indicate that the introduction of the CSM-PIP interventions has not to date had any measurable effect on overall error rates and hence site performance. However, because as noted above, these data represent *error counts* rather than *error rates* and there has been a change in the sampling frame, this type of aggregated analysis might obscure real changes in error rates occurring at particular sites or in particular protocols.

Moreover, because of the large number of sites and protocols involved in the DIADS clinical trial system, there may be other confounding variables affecting error

rates for particular networks, sites or protocols that would obscure any reduction in error rates due to the CSM-PIP interventions. Therefore, a number of more detailed analyses were conducted in an attempt to gain further evidence on the factors that might affect error rates overtime and across networks, protocols and sites. This information will be important in determining how the monitoring data might need to be stratified in order to be used as a basis for evaluating the impact of the CSM-PIP interventions.

Figure 1. Mean Observations per Month for All Networks Combined

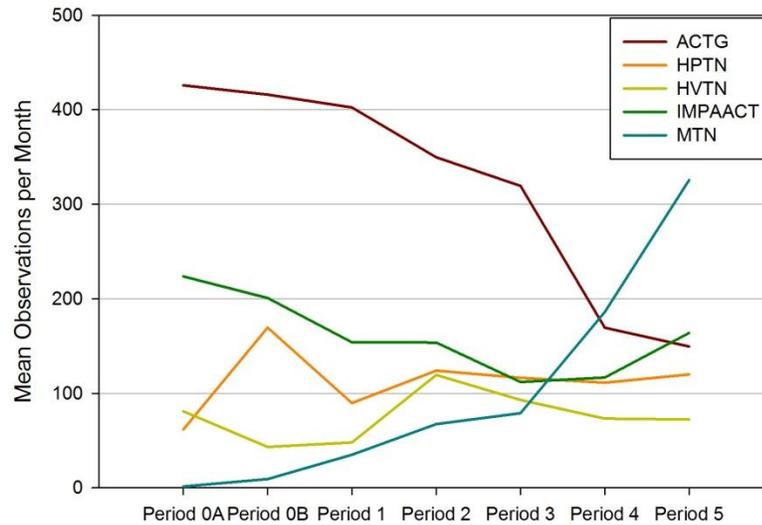


2. Errors by Network

Figure 2 shows the observation counts per month for the five networks over the seven periods analyzed. HPTN, HVTN and IMPAACT are reasonably constant over the 54 month period, in line with the aggregated data. However, the observations per month for ACTG decrease substantially while those for MTN increase to a similar degree, balancing out to give a constant level of observations on an aggregated basis.

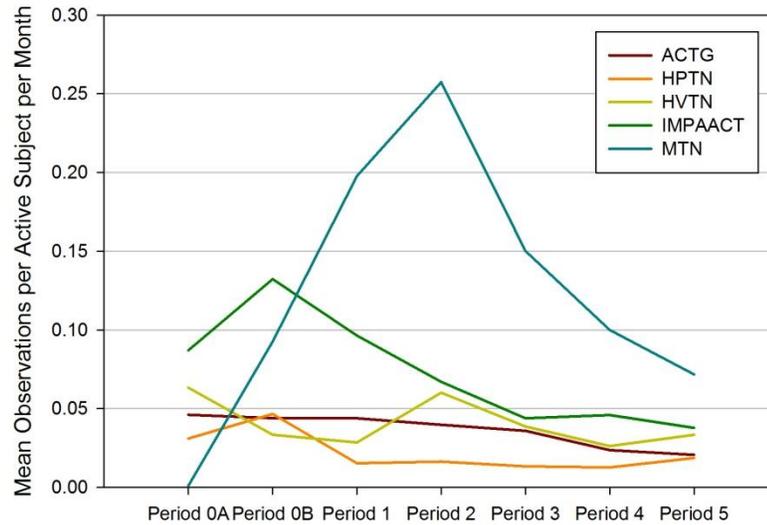
NIAID CONFIDENTIAL

Figure 2. Mean Observations per Month by Network



However, when the data for the five networks is normalized for the number of subjects on study (one of the proposed proxy denominators), the results are somewhat different. In this case, the observations per month per subject for HPTN, HVTN and ACTG are reasonably constant over the seven periods (see Figure 3) while IMPAACT decreases by ~50%. MTN presents the most interesting pattern in that observations per month per subject fell 70% from January 2010 through June 2012 while the total subjects increased almost 10 fold. Prior to that, the dramatic spike in observations per month per subject during the second half of 2009 may reflect the period that sites were just coming on line when errors would be expected to be greater.

Figure 3. Mean Observations per Active Subject per Month by Network



This analysis by network illustrates the complexity of factors affecting error rates that will confound attempts to distinguish an effect of the CSM-PIP interventions. Based on observations per month, one might conclude that there was an effect on ACTG site performance that should be investigated to determine if it was caused by the CSM-PIP interventions or other factors. However, if you examine observations per month per subject as a way to correct total observations for the number of PIDs or records reviewed at sites from the various networks (which might or might not be an accurate proxy denominator for the actual protocol events reviewed), one might conclude that IMPAACT was the network potentially affected.

In any event, the fact that in either analysis one sees a decrease in errors only for a single network (assuming the changes for MTN reflect the ramp up of study activity) makes any attribution of the change to CSM-PIP difficult unless there is a substantive reason why the effect would be different for different networks. Nevertheless, this analysis demonstrates the potential of using monitoring data in a systematic fashion to identify improvements in performance for certain networks that might warrant an investigation of the reasons as a potential best practice or new intervention for other sites.

3. Analysis by Observation Codes

Although the analysis of changes in overall observations over time did not show promise for the use of monitoring data to assess the impact of CSM-PIP, STPI researchers decided to conduct a more detailed analysis by observation code. This seemed justified given the complexity of observations of different character identified by

NIAID CONFIDENTIAL

the monitors and the resulting potential that an effect might be seen in some of these performance areas and not others.

a. Major Codes

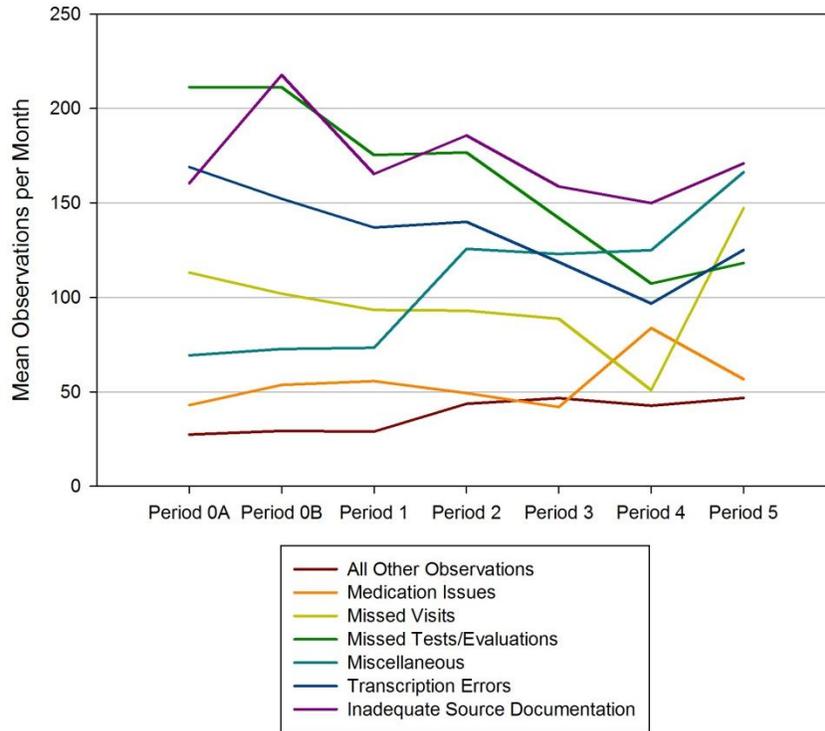
In analyzing the data to determine overall observations per month over time for the different networks, it was noted that across the system and over the 54 month period, approximately 95% of the observations appeared in 12 of the 25 codes tracked by the monitors. Because the errors were concentrated in these specific observation codes (hereinafter referred to as “major observation codes”), it was of interest to determine if the overall pattern of observations could be further understood by examining changes in these codes individually.

Table 3 lists the major observation codes by category and the definitions developed for these categories by STPI researchers while Figure 4 shows the pattern of these observations across the system on a per month basis over the 54 month period. The first item worth noting is that missed tests/evaluations, transcription errors and inadequate source documentation are the most prevalent of the major observation codes, accounting for approximately 60% of total observations. Two of these codes, missed tests/evaluations and transcription errors, decreased substantially over the time period (50% and 30% respectively) while inadequate source documentation errors were reasonably constant.

Table 3. Major Observation Code Categories

Description	Codes
Medication Issues	A11/A111/A21/A121
Missed Visits	A13/A23
Missed Tests/Evaluations	A14/A24
Miscellaneous	A16
Transcription Errors	A17
Inadequate Source Documentation	A71/A72
All Other Codes	A1/A2/A10/A12/A22/A15/A20/A30/A51/A61/A62/A8/A9

Figure 4. Mean Major Observations per Month



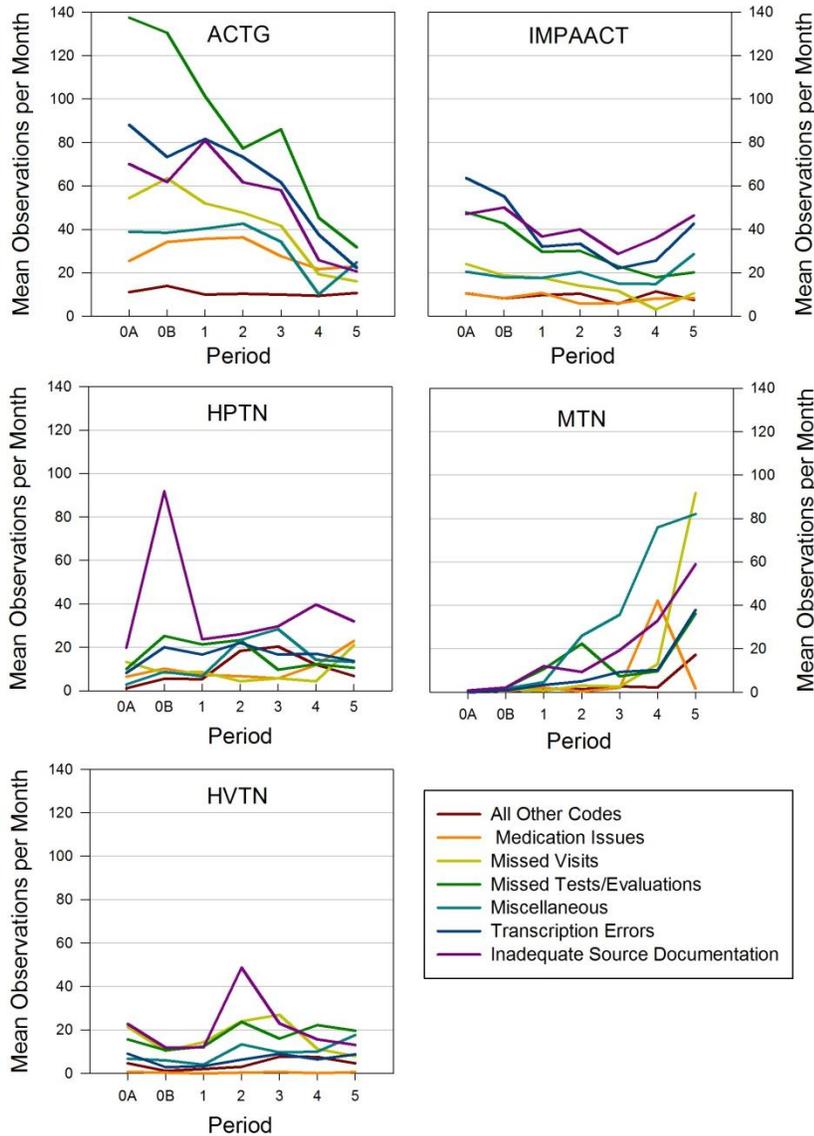
Of the remaining major codes, medication issues and missed visits show only minor variation over the 54 month period while the code for recording any observation not covered by another code (A16) increased over two-fold. Although it is not clear why the CSM-PIP interventions would have a specific effect on the incidence of missed tests/evaluations or transcription errors, these data show how routine analysis of errors by observation code could be valuable in identifying error trends that might warrant further investigation.

An analysis of the counts per month of these major observations by network was also conducted to determine if there was any meaningful variation (see Figure 5). As might have been expected, HPTN, HVTN and IMPAACT showed little variation over the 54 month period for any of the codes. The only exceptions are a substantial spike in inadequate source documentation errors in the January-June, 2009 period for HPTN and a somewhat smaller spike in the same codes for HVTN in October-December of 2009. The MTN analysis indicates that the overall increase in observations per month over time is primarily concentrated in three areas, missed visits, inadequate source documentation and the uncharacterized A16 observations while other errors which might also have been expected to increase with an increase in study activity, such as transcription errors, do not increase. Finally, this analysis indicates that the overall decrease in errors per month for ACTG is primarily driven by a decrease in missed tests/evaluations followed by a lesser

NIAID CONFIDENTIAL

impact from a reduction in transcription errors, inadequate source documentation and missed visits all of which might be affected by changes in study activity.

Figure 5. Mean Major Observations per Month by Network



These results thus do not provide any additional support for the ability to use monitoring data to assess the impact of CSM-PIP as there is not a consistent decrease in errors for any particular observation code across all of the networks. However, this data analysis shows how a standardized, reasonably real-time (i.e., monthly), routine analysis of observation counts stratified by network and observation code could be valuable in identifying error trends that should be investigated and in understanding the basis for any overall change in monitoring observations.

b. Serious Codes

Although reducing errors of all types is important for overall quality management of clinical trial conduct, it is the reduction of those errors viewed by DAIDS and the Networks as most serious which would have the greatest benefit. Therefore, STPI researchers also stratified analysis of the observation data to look only at these “serious codes” (see Section D.3 of this chapter). As shown in Figure 6, the percentage of errors in these 11 “serious codes” varied across the networks during the period analyzed. MTN and HVTN have the lowest percentages (~5% on average over the time period) while HPTN, IMPAACT and ACTG are 8-10% on average. HPTN also has one severe spike of almost 20% in the early months of 2010. However, looking at the trend over time, the percentage of serious errors is decreasing overall and during the period January 2011 through June 2012, the percentage was 2-4% for all networks.

Figure 6. Serious Observations as Percentage of Total Observations by Network

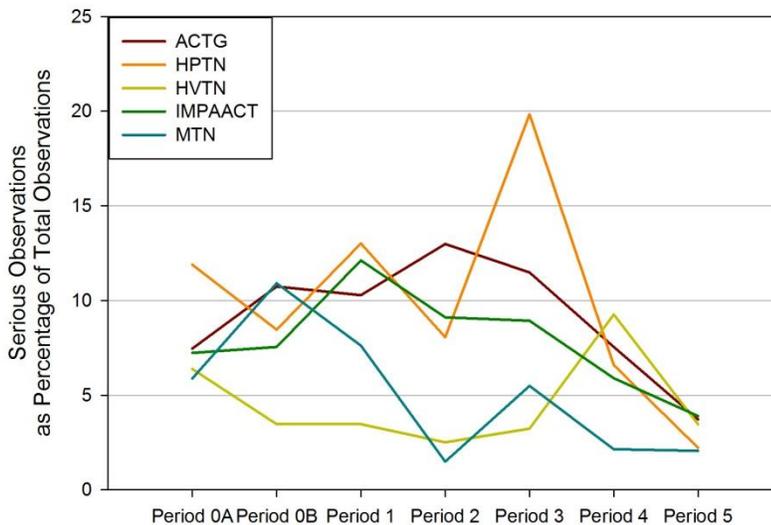
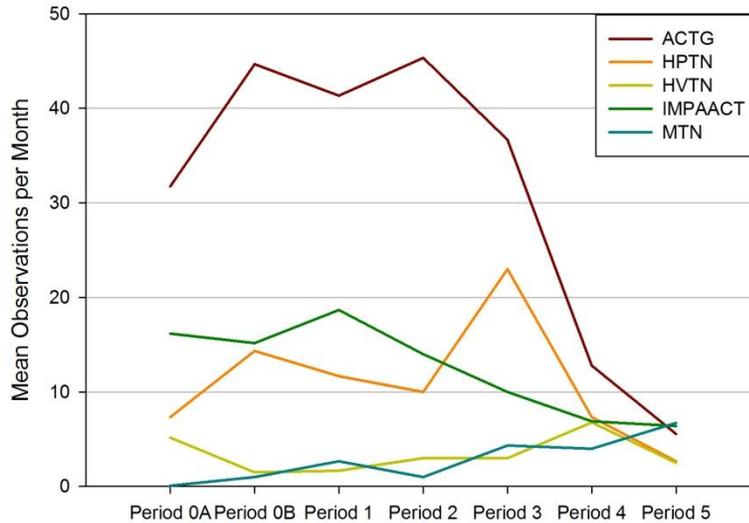


Figure 7 shows the variation in these serious observations per month for the five networks over the 54 month analysis period. For HPTN, HVTN and IMPAACT, the serious observations per month are reasonably consistent over time as was the case for total observations per month. For MTN, the serious observations per month increase slightly over time but not nearly to the extent found for total observations and the absolute level is lower than for any other network except HVTN. ACTG follows a similar pattern to that seen for total observations per month although the decrease in serious observations is almost 90% compared to a 75% decrease in total observations.

Figure 7. Mean Serious Observations per Month by Network



Taken together, these data show a substantial decrease in the overall serious observations per month across the networks as well as a decrease in serious errors as a percentage of total observations. In theory, these decreases could be indicative of an impact from the CSM-PIP interventions due to the time coincidence although it is hard to justify that MRD and RBPMP would disproportionately reduce serious errors while not affecting the overall observations per month.

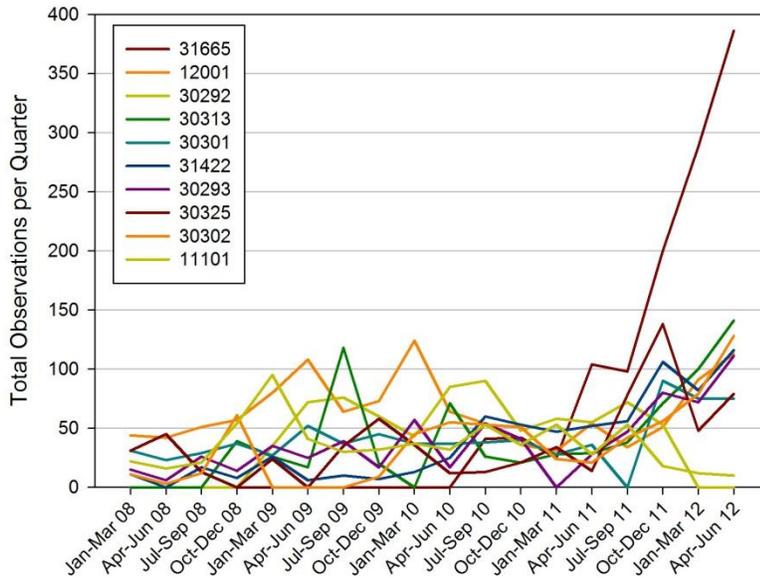
4. Analysis by Site

Although the analysis of changes in observations by network over time did not show promise for the use of monitoring data to assess the impact of CSM-PIP, this network based analysis aggregates data from many individual sites and could potentially obscure an effect of CSM-PIP at a site specific level. STPI researchers therefore decided to conduct an analysis of the 10 sites which had the largest number of total observations during the 54 month analysis period. All of these sites are in Africa except for one which was missing from the reference table and could not be identified. These sites were chosen because the potential to see a change in error rates would be greatest if the sites had a large number of total observations. No attempt was made to determine if these sites represent poor performers or whether they simply have a high level of activity and hence a higher level of monitored records and thus a higher total number of observations. Fortunately, five of these sites were also ones selected for the MRD pilot.

NIAID CONFIDENTIAL

The total number of observations for these 10 sites over the 54 month period ranged from 1184 to 643 with the pattern of observations per quarter¹⁷ shown in Figure 8. With the exception of the last three quarters of data for one site (31655), the observations per quarter for all of these sites varies between 0 and 100 with no distinct patterns over time or by site. In order to determine if these data would give any evidence of an early effect of MRD at the MRD pilot sites, the aggregated data for the five MRD pilot sites was compared with the remaining sites as shown in Figure 9.¹⁸

**Figure 8. Total Observations per Quarter
Ten Sites with the Most Errors**

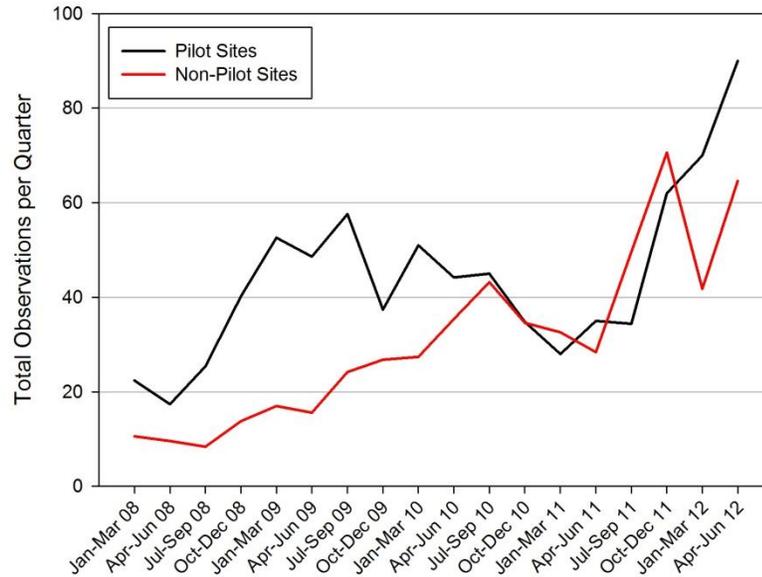


¹⁷ The data is expressed per quarter to reflect the fact that an individual site would be visited once per quarter and therefore a per month value is not as relevant on a site specific basis.

¹⁸ Site 31655 was eliminated from the non-MRD pilot site data as the observations at this site were so divergent.

NIAID CONFIDENTIAL

**Figure 9. Total Observations per Quarter
MRD Pilot Sites vs. Non-Pilot Sites**

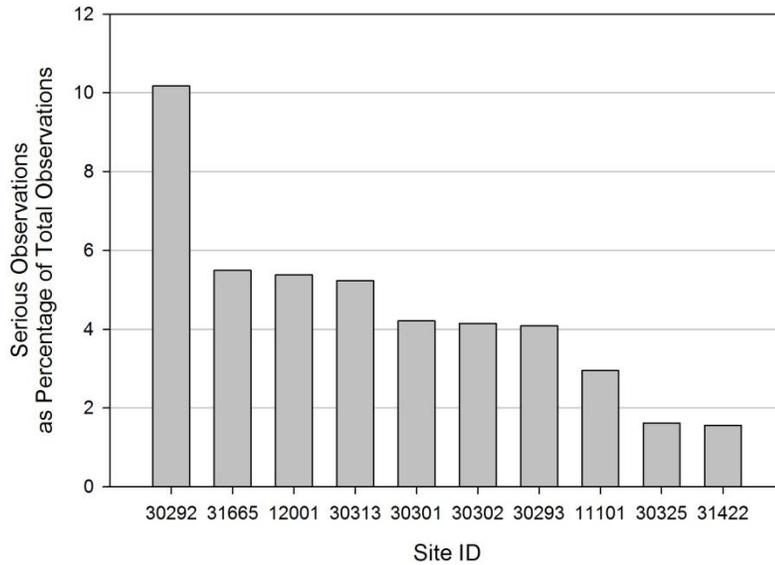


This analysis yields some interesting findings. The first is that although the pilot-MRD sites do show a slight decrease in observations per quarter between July and December 2009 compared to the other sites, their absolute error levels were higher than other sites prior to July 2009 so the attribution of this decrease to CSM-PIP is difficult to make. The second is that the observations per quarter for both the pilot MRD sites and the other sites decreased slightly after July 2010 and then began to increase in mid 2011. This increase over the latter part of the analysis period might indicate that some factor or factors was causing error rates to increase at these sites which might obscure any simultaneous effect of CSM-PIP in reducing error rates.

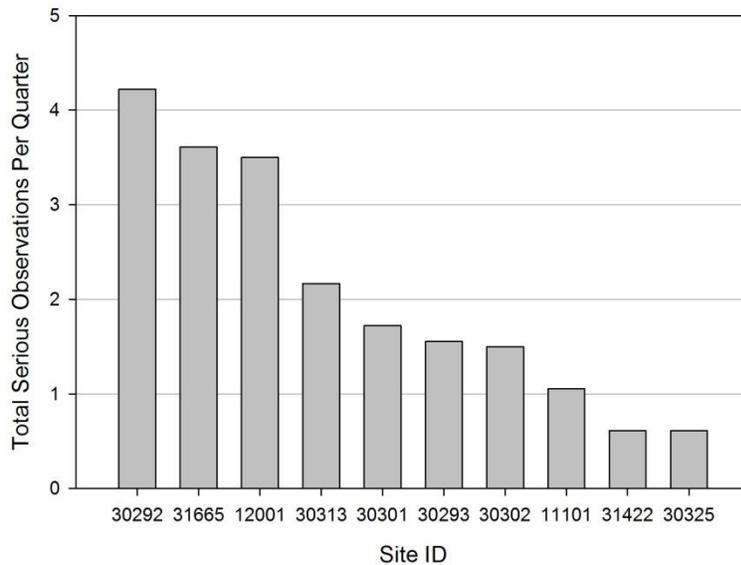
These sites were also analyzed to determine the absolute number of serious errors that occur on a per quarter basis. This was considered relevant because the most important goal of CSM-PIP should be to reduce these serious errors and, if those errors are already at a low level on a per site basis, then the likelihood of being able to detect such a change in the monitoring data is low. Figure 10 shows that the overall percentage of serious errors at these sites across the analysis period was in the range of 2-5% except for site 30292 for which 10% of the errors were in the serious codes. This is similar to the overall prevalence of serious codes across the system. In terms of absolute numbers, this translates into a range of serious observations per quarter per site from 0.6 to 4.2 (see Figure 11). Since observations per quarter for the remaining sites will be much lower, the level of serious errors identified per visit at each site is likely to be so low as to render the detection of a change due to CSM-PIP difficult if not impossible.

NIAID CONFIDENTIAL

**Figure 10. Serious Observations as Percentage of Total Observations
Ten Sites with the Most Errors**



**Figure 11. Total Serious Observations per Quarter
Ten Sites with the Most Errors**



5. Analysis by Protocol

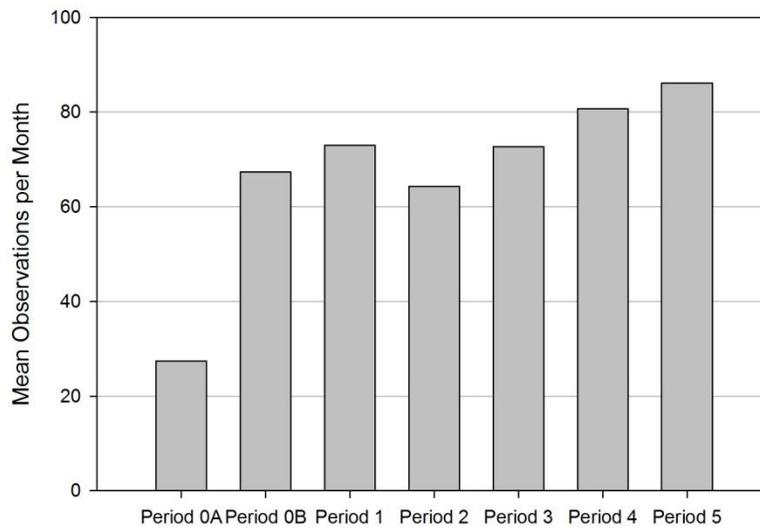
The final approach that was investigated for detecting an impact of CSM-PIP on error rates was to examine observations by protocol on the premise that the CSM-PIP interventions might have different effects on different protocols. The pattern of

observations for individual protocols was examined from three perspectives. The first was changes over time during the analysis period. The second was the patterns observed for protocols with the largest number of observations during the period. The third was examining whether there were changes in observations according to risk level.

a. Change over Time

There was only one protocol, HPTN052, which had monitoring data throughout the 54 month analysis period. As shown in Figure 12, with the exception of January-December 2008, the observations per month for HPTN052 are reasonably constant over time. Moreover, when normalized for the number of subjects on study, even the January-December 2008 data are in line with the other periods (see Figure 13). When analyzed by major observation codes, there is still no discernible trend in error rates over time with the observations per month in various code categories varying up and down over the course of the analysis period (see Figure 14). Therefore, even when looking at a single protocol which might have fewer confounding variables, there is no evidence of reduced error rates due to CSM-PIP.

**Figure 12. Mean Observations per Month
Protocol HPTN 052**



NIAID CONFIDENTIAL

Figure 13. Mean Observations per Active Subject per Month for Protocol HPTN 052

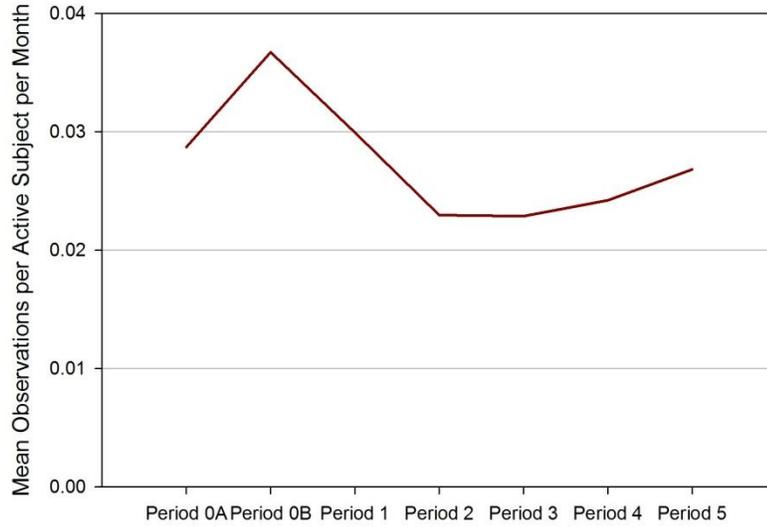
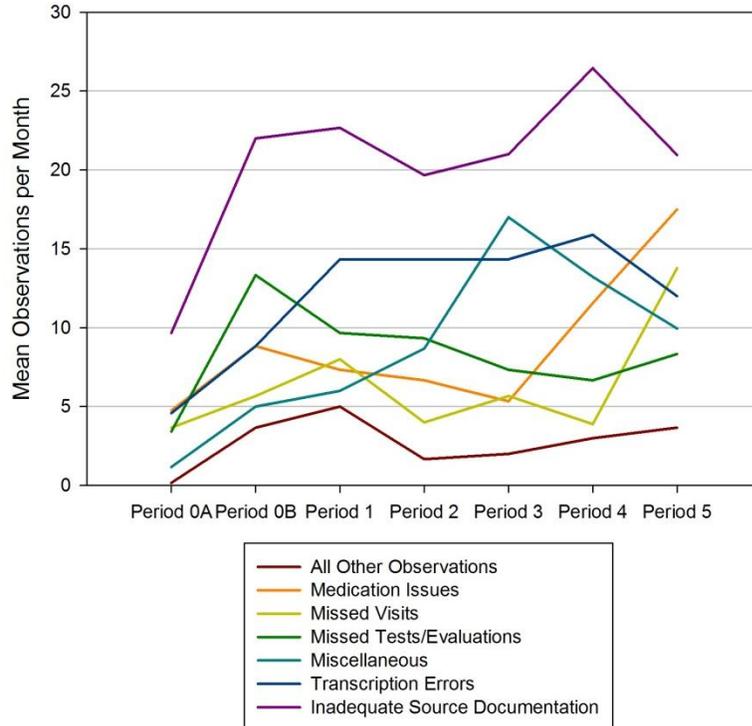


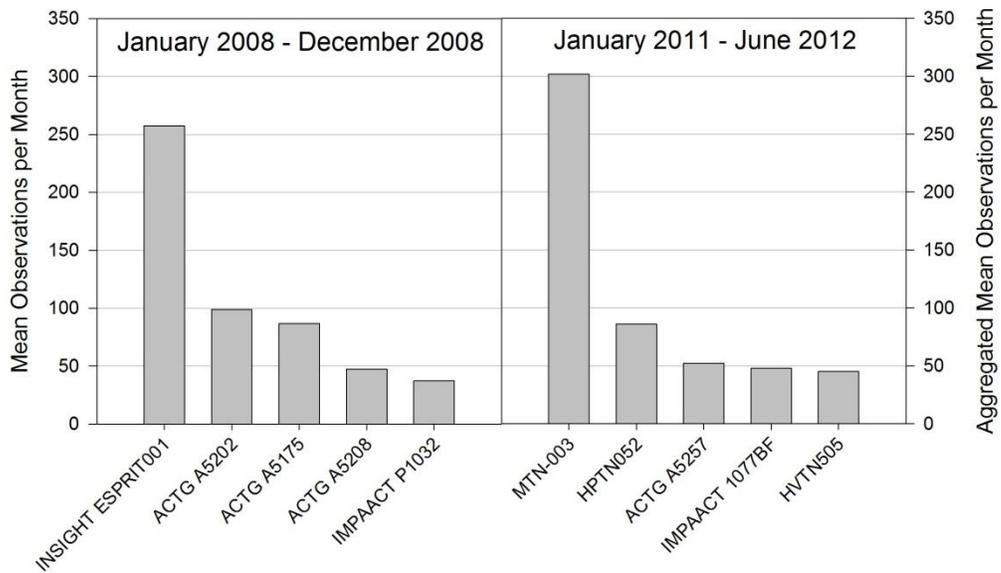
Figure 14. Mean Major Observations per Month Protocol HPTN 052



b. Protocols with the Most Errors

The theoretical basis for analyzing error rates for protocols with the largest number of errors over the entire analysis period is that the level of errors in these protocols might be sufficiently large to detect a difference due to CSM-PIP. Figure 15 presents the total observations per month for the five most error prone protocols in the period January–December 2008 and January 2011–June 2012.¹⁹ Although each period had one active protocol with a substantial number of observations per month, most protocols had 50-100 observations per month on average. This number is probably too low to expect implementation of CSM-PIP to have enough of an effect on a sufficiently large number of sites to show an observable effect in the total error rate even for these protocols with the largest number of errors observed. Therefore, protocol specific analysis of error rates is not likely to be more useful than the other types of analyses.

**Figure 15. Mean Observations per Month
Five Protocols with the Most Errors**



c. Variation with Protocol Risk

The rationale behind the RBPMP CSM-PIP intervention was to shift monitoring effort to the protocols where the consequences of errors were likely to be greatest. Therefore, if effective, over time one might expect the error rates for these protocols to decrease compared to less risky protocols which are monitored less intensely. However,

¹⁹ Although INSIGHT monitoring data was not included in any of the other analyses in this report, because this was the protocol with the largest number of errors in Period 0A, it is included here.

NIAID CONFIDENTIAL

early in implementation, one might expect error rates for these high-risk protocols to actually increase due to the higher intensity of monitoring and then to decrease over time as protocol specific corrective actions are taken with regard to the identified errors. In contrast, the error rates for protocols judged less risky would probably decrease initially due to lower monitoring intensity and then either remain constant or increase due the reduced monitoring intensity. Given that the overall level of monitoring is expected to remain constant, RBPMP itself is unlikely to reduce overall error rates and any analysis of rates would have to be performed as a comparison between high-risk protocols and lower risk protocols.

Unfortunately, the retrospective data set used for the other analyses in this report was not sufficient for STPI researchers to compare a set of high-risk and low-risk protocols, all of which were initiated after the introduction of RBPMP and which could be tracked for error rates over the lifetime of the protocol. The protocols in the data set provided were not coded as high or low risk and, more importantly, there may not have been sufficient elapsed time on a sufficiently large number of protocols to perform a robust analysis. Nevertheless, the factors that would affect the feasibility of such an analysis could be analyzed and are described in Chapter 4 of this report.

3. Stakeholder Perceptions of CSM-PIP Value

In the course of the feasibility study, STPI researchers gained certain insights on the value of the CSM-PIP interventions from discussions with DAIDS/OCSO program staff, PPD staff involved in DAIDS clinical site monitoring, and staff from DAIDS clinical research sites who are involved with the monitoring process. Although the primary purpose of these discussions was to understand the details of the monitoring process and data or information that might be useful for a CSM-PIP process and outcome evaluation, STPI researchers did elicit perceptions of the value of the CSM-PIP initiatives. It is important to note that for the PPD monitoring staff and the clinical research site staff only 1-2 individuals were interviewed for this feasibility study. Thus the views expressed may not reflect a consensus view of these stakeholders but only provide a guide for potential future investigation.

A. DAIDS/OCSO Program Officers

MRD was generally viewed positively by the Program Officers. In their view, MRD makes the errors observed more relevant to current site operations and the perception is that sites are more diligent because more current data, for which they can be held more directly responsible, is being monitored. Program Officers have also tended to raise more issues from monitoring reports after MRD because the errors are more current and relevant. Another advantage cited was that the review of each monitoring report takes less time after MRD because fewer records are reviewed. A disadvantage cited was that it was harder to view site longitudinally for a pattern of errors. The revised version of MRD with review of the last three years versus last 12 months was viewed as better from this perspective. There was, however, no perception that site performance had improved to any meaningful extent since implementation of MRD.

Perceptions of Program Officers with regard to risk based protocol monitoring (RBPMP) were much more neutral. They do not perceive that protocols ranked as high risk are more error prone. They did note that error rates increased on interventional trials after RBPMP introduction because they were more the focus of monitoring. With regard to the potential introduction of risk-based site performance monitoring (RBSPM), there were mixed views. Although concentrating monitoring resources on sites that are more likely to have errors is viewed as logical, there were concerns about implementation. The first concern was that Program Officers believe site risk is already taken into account in distributing monitoring resources. The second concern was that risk ranking sites in a

NIAID CONFIDENTIAL

static fashion would be ineffective as the degree of risk at a site changes over time as the site grows, takes on new studies, has staff turnover, etc. It was also noted that, unlike the other two CSM-PIP interventions, RBSPM is likely to increase workload for the Program Officers.

B. PPD Monitoring Staff

Implementation of MRD was viewed by the PPD staff as having minimal impact. One reason is that, from their perspective, most records assigned for review are more recent records in any event and therefore not affected by MRD. This is particularly true since the conversion from 12 months to three years of records to be reviewed. PPD staff did however note the advantages that more PIDs may be reviewed because of MRD and that the errors identified are more relevant for improving site performance.

There were differing views among the PPD staff interviewed as to the value of RBPMP. Some felt it had little effect on monitoring effort as record review was the same. Others felt it had an impact not because it focused on high-risk protocols but because it set monitoring goals for each study which allows protocols to be removed from the pool when their goals had been met. In the past, monitors continued to randomly review PIDs on all protocols to a theoretical level of 100% which spread monitoring resources much more thinly across protocols. They did, however, note that risky protocols may be more complex and thus more difficult and time consuming to monitor. This could result in RBPMP reducing the number of PIDs that can be reviewed in a site visit.

RBSPM was viewed as being a good idea as focusing monitoring effort on those sites where errors are more likely to occur will have the greatest effect on site performance. However, they shared the Program Officer concerns that the risk status of a site is not static but changes over time.

A final note is that although PPD staff did not view the CSM-PIP interventions as having a particularly positive effect, they did state that prior to CSM-PIP only 6-10 PIDs were reviewed per visit whereas now approximately 16 of the 20 PIDs selected for review at a site are completed. This would indicate that either CSM-PIP or some other simultaneous intervention has simplified record review so that more PIDs are covered.

C. Clinical Research Site Staff

The clinical research site staff members interviewed stated that MRD has made monitoring a more positive experience for sites because errors can be acted on and corrected. They also perceive that monitors now have more time to check records more thoroughly and make specific suggestions for improvement. In addition, as with PPD staff, they perceive that monitors are reviewing a larger percentage of the PIDs on their list since implementation of MRD. From the sites' perspective, they have not seen any

NIAID CONFIDENTIAL

changes since implementation of RBPMP and they had no comment on RBSPM. One other item of note mentioned by the site staff was that error rates are not driven by monitoring effort but by factors such as starting new protocols and training new staff.

4. Feasibility of Process and Outcome Evaluation

Based on the analysis in this report, STPI researchers have concluded that it is not feasible or warranted to conduct a formal process and outcome evaluation of the CSM-PIP interventions implemented to date based on analysis of objective monitoring data. Furthermore, although it would be possible to collect qualitative information on the value of the CSM-PIP interventions through systematic interviews with PPD site monitors and clinical research site staff, such a process may not be warranted. The reasons for these conclusions are described below.

A. Evaluation Based on Monitoring Data

1. Error Rate Trends

The DAIDS CSM-PIP interventions have all been designed with the goal of allocating a finite level of monitoring resources more effectively so as to improve site performance in terms of a reduced incidence of errors. Since the first two initiatives were fully implemented by January 2011, and sites are monitored quarterly, it was reasonable to ask whether there had been a reduction in error rates over the six quarters that had elapsed by June 2012 compared to a baseline period going back to January 2008. However, as demonstrated by the data presented in Chapter 2 of this report, Analysis of Monitoring Data, no such effect is discernible to date in either overall error rates across the system or when evaluated separately by network, observation code, site or protocol.

It is, of course, possible that the CSM-PIP interventions have not yet been in place long enough for their effect to be manifested. In that case, repeating the analyses described in Chapter 2 periodically over time might reveal an effect. But even if that were the case, a formal process and outcome evaluation would not be warranted as the effect could be demonstrated simply by DAIDS/OCSO staff performing the analyses periodically and determining if a reduction in error rates is occurring.

2. Confounding Variables

The most important conclusion from the analysis in Chapter 2 is that the variability of the monitoring data across networks, sites and protocols is so large and affected by so many different factors that it is unreasonable to expect to see a reduction in error rates due specifically to the CSM-PIP interventions, even if they occur, because of the number

NIAID CONFIDENTIAL

of other factors that could counteract or obscure such an effect. This is demonstrated by the widely different patterns of error rates over time for the five networks either in terms of observations per month (Figure 2) or observations per month per subject (Figure 3). Some networks have stable error rates while others show substantial decreases or increases, undoubtedly due to the interplay of a number of factors.

The first complicating factor is the varying nature of protocols within and across networks in terms of the number and complexity of protocol events and hence the number of opportunities for an error to occur. If more protocol events must be reviewed per PID at a site due to the nature of the protocols, then there will naturally be increased errors (providing monitor time is available to review the larger number of events) even if site performance remains the same. However, if site performance is improving due to a change like CSM-PIP, then the increased number of protocol events reviewed might still obscure any improvement when looking at error rates. Unfortunately, as explained in Chapter 2, the effort involved in obtaining events reviewed as an error rate denominator would be substantial and difficult to justify given that other factors could further confound any analysis of error rates.

The second complicating factor is that error rates can go up or down if protocol events that lead to a high percentage of the total errors are changing substantially. For example, in the analysis of error rates by observation code, the number of missed tests per month decreased from 200 to 100 between January 2008 and June 2012 (see Figure 4). If this decrease is because the number of tests performed has decreased by half, then site performance in this area has not really improved, the opportunities for error have simply decreased. Again, sorting this out for every protocol in order to identify which error rate changes truly reflected site performance would not be justified.

The third complicating factor is that, in discussion with STPI researchers, all stakeholders in the DAIDS monitoring system stated that the most important factors affecting error rates are variables such as the number of new protocols being opened, the number of new sites coming on line, staff training, staff turnover, etc. Even though changes in the deployment of monitoring resources, such as CSM-PIP, might result in more effective corrective actions to improve site performance, such improvements would often be overshadowed by these other factors that might either decrease or increase error rates. Because all of these variables are site or protocol specific, sorting out the variables and correcting for them would have to be done for each site/protocol pairing which would be an unreasonable amount of effort just to demonstrate with error rates that CSM-PIP did or did not have an effect.

The fourth complicating factor is that other changes in monitoring practice can shift error rates. An example of this may be the increasing incidence of A16 observations over the 54 month analysis period described in Chapter 2. If this increase is due to change in monitoring practice to capture a larger number of uncharacterized errors as A16

NIAID CONFIDENTIAL

observations as opposed to ignoring them, then this would artificially inflate error rates compared to before the monitoring practice change. STPI researchers have no evidence that this is what happened but only present this an illustration of what might happen.

3. Low Levels of Errors

Another conclusion from the analysis in Chapter 2 is that error rates are often too low to reasonably expect to see a difference due to CSM-PIP. This can be illustrated by two examples.

In the discussion with various DAIDS stakeholders from OCSO, PPD, sites and the networks, it was clear that of the observation codes noted by monitors, about 50% are considered serious and it is those on which you want to improve site performance. Improvement in the remaining 50% would be of much less consequence in demonstrating the value of CSM-PIP. Fortunately for clinical trial quality and subjects, those serious errors represented only 5-10% of the observations made by monitors in the early part of the STPI analysis period and by the final period, January 2011-June 2012, had fallen to 2-4%.

The result is that serious errors per month are now in the single digit range for even the most error prone protocols across all sites and in the low single digit range for the most error prone sites. Thus, it would be difficult to see a change in serious errors due to CSM-PIP or any other intervention because of the signal to noise problem. One might, of course, consider whether the decrease in the serious error rate by the end of the analysis period was in fact due to CSM-PIP. The problem with this conclusion is that STPI researchers could not identify any reason why the CSM-PIP interventions would preferentially reduce the rate of serious errors without a similar effect on other errors which is not demonstrated by the data.

As described in Section A.2 of this chapter, the only approach for dealing with some of the confounding variables would be to examine error rates on a site and protocol specific basis. Besides being labor intensive, such an approach would also likely suffer from a signal to noise problem because the number of errors at a given site and in a given protocol is low. For example, the 10 sites with the largest number of errors in the system, out of over 100 sites, have an average of 50 observations per quarter over the 54 month period analyzed of which only 1-4 per quarter are serious errors. The remaining sites by definition have many fewer. When this low level of observations per site is spread over all the protocols open at a site, the number probably becomes vanishingly small for most sites. Thus, it is probably not reasonable to expect to be able to see an effect from CSM-PIP on a site specific basis. In terms of observations per month per protocol across all sites, the 10 most error prone protocols also have low rates in the range of 50-100. Given this low level, it is also unreasonable to expect to see an effect on the overall protocol error rate from interventions operating at each individual site where the protocol is active.

4. Lack of Correlation between Errors and Protocol Risk

As described in Chapter 2, Section E.5 of this report, focusing monitoring effort more intensely on protocols where the consequences of errors are more serious either for subject safety or data quality would not be expected to impact overall error rates. Nevertheless, it might be possible to determine an effect from RBPMP by determining if error rates on high-risk protocols decrease over time compared to medium- and low-risk protocols. However, because error rates on a protocol specific basis are likely to vary substantially due to varying protocol complexity and the number of protocol events that must be reviewed, it is not clear whether it will be possible to discern differences in the trends in error rates over time by comparing all high-risk protocols with all medium- and all low-risk protocols. It also may be the case that corrective actions on a specific protocol may or may not have a more general effect on errors of that type in other high-risk protocols. As a result, any observed decrease in error rates may be protocol specific. For all these reasons, discerning an effect of RBPMP on error rates may require analyzing error rates over time separately for each protocol which would be a labor-intensive undertaking.

There are two other theoretical issues with attempting to identify an effect from RBPMP from protocol specific monitoring data. First, there may be insufficient time for a corrective action due to an error observed in a protocol at a particular site to have an effect on error rates for that protocol at the site before the study has been closed. Second, although correcting the error may have a beneficial effect in terms of reducing the immediate consequences of the error, the effect may not translate to other PIDs or study events within the protocol rapidly enough to be observed in general error rates for that protocol. Third, as noted above, the number of observations per site per protocol is generally so low that it will be difficult to identify a difference due to implementation of RBPMP because of a signal to noise problem.

B. Evaluation Based on Stakeholder Interviews

Despite the conclusions above that it will be difficult, if not impossible, to evaluate the impact of CSM-PIP interventions based on monitoring data, it might be possible to conduct a formal qualitative evaluation of their value through stakeholder interviews. The logical rationale behind the CSM-PIP initiatives is sound and interviews might reveal whether the theoretical advantages are seen in practice.

The effect of MRD in concentrating monitoring on more recent records is designed to identify errors that can be more readily corrected and may therefore improve future site performance. However, for all the reasons cited above, the degree to which correcting individual errors at specific sites will actually lead to sufficient improvement in future site performance to be observed in error rates is questionable. Nevertheless, it was clear from the stakeholder interviews described in Chapter 3 that MRD was viewed positively

NIAID CONFIDENTIAL

by both the monitors and site staff. Therefore a comprehensive set of interviews to more robustly identify the advantages might provide a qualitative evaluation of MRD effectiveness. However, because the value of MRD to sites and monitors appears to be obvious and well-accepted, such an exercise would probably not be warranted.

The rationale behind RBPMP is not to improve overall site performance but to improve performance on those specific protocols where the consequences of errors are judged to be more serious. If there is an overall decrease in error rates for these protocols, then the assumption is that the errors with serious consequences will be reduced pro rata. However, for the reasons noted in Section A.4 of this chapter, it may be difficult to demonstrate that overall error rates on these protocols are decreasing. One possible alternative approach would be to interview DAIDS Program Officers, PPD monitors and clinical research site staff to identify specific corrective actions that were taken with regard to high-risk protocols which in the view of the various stakeholders provided an important improvement in either subject safety or data quality that would not have occurred without identification of the problem through enhanced monitoring. While such an interview based evaluation is feasible, because the prevalence of such events is likely to be low and the fidelity with which they could be identified by stakeholders is highly uncertain, such an interview based evaluation is not recommended.

5. Best Practices in Clinical Site Monitoring

A. Introduction

Clinical trial sponsors monitor their studies to meet their ethical obligations to study participants and future patients who may rely on study results and to meet the requirements of regulatory authorities. While the need for monitoring is widely accepted, what constitutes adequate monitoring continues to evolve.

Monitoring practices vary across clinical trial sponsors. Pharmaceutical companies have historically taken a conservative approach to monitoring, with frequent on-site monitoring visits and often 100% source data verification (SDV), out of concern that alternative monitoring strategies may be considered inadequate by regulatory authorities and thus may result in delay or rejection of submissions for marketing approval. On the other hand, government and academic sponsors of clinical studies often lack the resources for such comprehensive monitoring. One survey of monitoring practices found that on-site monitoring visits are performed on only approximately 31% of academic- and government-sponsored trials, compared with 80-89% of privately-sponsored trials (Morrison et al. 2011).

Regulatory bodies, industry groups, and others have each proposed guidelines for acceptable monitoring practices (ICH guideline for good clinical practice (E6) 1996; Food and Drug Administration 2011; Knatterud et al. 1998; Usher 2010). These existing guidelines are generally based on expert consensus rather than empirical evidence although a few promising randomized studies are underway designed to empirically evaluate monitoring methods (Journot et al. 2011). However, until empirically determined best practices are developed, sponsors of clinical trials must individually balance the evolving consensus on effective and appropriate monitoring approaches with available resources.

Moreover, because of increasing product complexity and more stringent regulatory demands, the cost and time required for pharmaceutical product development is steadily rising. This has led pharmaceutical companies to seek ways to simplify and streamline the development process including clinical trial monitoring. This effort has been aided by rapid advances in information and communication technologies and by adapting quality improvement ideas derived from the management sciences.

As part of this effort to improve the quality and efficiency of clinical trials, the Food and Drug Administration (FDA) and Duke University have founded the Clinical Trials Transformation Initiative (CTTI), a public-private partnership that now includes over 60

NIAID CONFIDENTIAL

member organizations from government, academia, and industry. One of the initial activities conducted by CTTI was a project to assess current monitoring practices and recommend improvements.

B. Regulatory Requirements for Monitoring

Both domestic and international regulatory authorities have issued statements or guidelines on the need for adequate monitoring. In 1988, the FDA issued industry a “Guideline for the Monitoring of Clinical Investigations” which was modestly revised in 1998.²⁰ In this guidance, FDA described “principles [that] are not legal requirements but represent a standard of practice that is acceptable to FDA”. These principles included frequent visits to clinical trial sites as well as review of “a representative number of subject records and other supporting documents with the investigator’s reports”. However, standards for visit frequency or for adequate sampling of records were not specified.

In 1996, the International Conference on Harmonisation (ICH E6) released guidelines for good clinical practice (GCP), which outlined standards of conduct for clinical trials. Regarding monitoring, ICH E6 states that sponsors are responsible for adequately monitoring clinical trials and that

The sponsor should determine the appropriate extent and nature of monitoring. The determination of the extent and nature of monitoring should be based on considerations such as the objective, purpose, design, complexity, blinding, size and endpoints of the trial. In general, there is a need for on-site monitoring, before, during, and after the trial; however, in exceptional circumstances, the sponsor may determine that central monitoring in conjunction with procedures such as investigators’ training and meetings, and extensive written guidance can assure appropriate conduct of the trial in accordance with GCP. Statistically controlled sampling may be an acceptable method for selecting the data to be verified (*ICH guideline for good clinical practice (E6) 1996*).

However, the ICH E6 does not specify clear standards for when central monitoring would be considered appropriate or for what constitutes statistically adequate sampling.

The ICH E6 guidelines, which have not been updated since the release in 1996, were developed through a consensus process between industry and government regulators. Other key stakeholders, such as academic clinical researchers, did not participate in the process. The guidelines have been criticized as vague, outdated, unnecessarily burdensome, and based on opinions rather than evidence (Funning et al.

²⁰ This guidance is no longer in effect and has been removed from the FDA website. However, it may still be viewed at <http://www.ahc.umn.edu/img/assets/19826/Clinical%20monitoring.pdf> (accessed Nov. 20, 2012).

NIAID CONFIDENTIAL

2009; Grimes et al. 2005; Lang, Cheah, and White 2011; Lörstad 2004; Abraham and Reed 2001). Moreover, Abraham and Reed (2001) argued that the ICH guidelines lowered regulatory standards in some countries and increased the vulnerability of regulatory processes to “capture by industrial interests.” No studies have empirically demonstrated advantages to patient safety or data quality from adherence to ICH E6 guidelines (Grimes et al. 2005).

In addition to its design flaws, GCP compliance is a major driver of the cost of clinical trials. A survey by Funning and colleagues (2009) found that GCP-related activities constituted about 50% of the cost of all phase III trials in Sweden. Half of the cost of GCP activities was attributed to source data verification. Further, 71% of respondents disagreed with the assertion that GCP compliance increases the quality and scientific credibility of clinical trials.

Recently, FDA released a new draft guidance encouraging alternative, risk-based approaches to monitoring (Food and Drug Administration 2011). The new draft guidance acknowledges that adoption of electronic data capture (EDC) systems enables remote execution of many monitoring activities, including source data verification, traditionally accomplished through on-site monitoring. The draft guidance encourages adoption of central monitoring techniques: “FDA encourages greater reliance on centralized monitoring practices than has been the case historically, with correspondingly less emphasis on on-site monitoring” (Food and Drug Administration 2011). The European Medicines Agency (EMA) has issued a paper similarly encouraging increased adoption of new, risk-based monitoring approaches (European Medicines Agency 2011).

While the new FDA and EMA guidance documents apparently endorse alternatives to intensive on-site monitoring, they do not provide prescriptive formulas for sponsors. Sponsors must thus continue to judge for themselves the best monitoring methods for their trials. In the face of such ambiguity, the most conservative approach is still to conduct extensive on-site monitoring. However, some have argued that intensive on-site monitoring is not in fact required for regulatory compliance (De 2011; Baigent et al. 2008),

C. Types of Monitoring

Sponsors employ many methods to ensure patient safety and data validity in their trials. Traditionally, exhaustive on-site monitoring by independent visiting monitors has been considered the most important component of clinical trial quality assurance. However, alternative methods, such as central monitoring and risk-based statistical sampling for on-site monitoring, are increasingly being used. The most common monitoring methods are discussed below.

1. On-Site Monitoring

Monitoring through frequent site visits, often with near 100% source data verification, is the predominant method of monitoring clinical trials. One survey found 87% of trial sponsors always perform on-site monitoring visits (Morrison et al. 2011). On-site monitoring has even been called the “standard of practice” in quality management of clinical trials (Journot et al. 2011).

A major reason for the acceptance of on-site monitoring is that in most cases, on-site visits are the only way for sponsors to review data not recorded in case report forms (Brosteanu et al. 2009). However, monitors often spend time on-site reviewing data that could be monitored remotely (De 2011). Even informed consent documents can be reviewed centrally with proper electronic data capture systems (Baigent et al. 2008). In addition, reduced and targeted on-site monitoring guided by risk assessments and central statistical review are viewed by some to be as effective as traditional, broad-coverage on-site monitoring (Bakobaki et al. 2012; De 2011). Ongoing studies are attempting to validate the reliability of such methods (Brosteanu et al. 2009; Journot et al. 2011). On the other hand, Tantsyura and colleagues (2010) argue that monitoring approaches with reduced source data verification may be appropriate for large, late phase trials, while 100% source data verification may be feasible and optimal for smaller, early phase trials.

a. Risk-Based Site Monitoring

One variation of on-site monitoring is risk-based or risk-adapted monitoring. These monitoring strategies identify either risky research sites or protocols for increased monitoring or target certain data fields which are more critical or more error-prone than others for increased scrutiny (De 2011). Risk-based monitoring programs reduce on-site monitoring costs (Grieve 2012); however, in many cases their effectiveness has not been validated and whether they are as effective as traditional on-site monitoring is unknown (Brosteanu et al. 2009). Nonetheless, about 78% of trial sponsors report conducting risk assessments for at least some protocols before developing monitoring plans (Morrison et al. 2011).

Responding to the need for validated alternative monitoring methods, two randomized studies—ADAMON and OPTIMON—are comparing risk-adapted methods to traditional on-site monitoring. The risk assessment method used in the OPTIMON study has already been validated (Journot et al. 2011). The ADAMON and OPTIMON studies are discussed in more detail in Section E of this chapter.

A second risk-based approach is targeted source data verification (De 2011). This approach focuses monitoring effort on high-risk data fields in source documents. High-risk data fields may be either those that are prone to errors or those in which an error would have critical impact. Another approach to targeted source data verification

NIAID CONFIDENTIAL

involves random sampling of CRF data fields for monitoring. The random data field method may reduce on-site monitoring costs, but its effectiveness has not been validated.

b. Adaptive Site Monitoring

Adaptive site monitoring uses data from central monitoring (discussed below) to determine whether and when particular research sites require on-site monitoring visits. Unplanned monitoring visits prompted by central review of trial data may be used to supplement or replace routine site visits. While only about 27% of trial sponsors use adaptive monitoring strategies (Morrison et al. 2011), it has been suggested that central statistical monitoring should always guide on-site monitoring activities (Baigent et al. 2008).

2. Central Monitoring

Remote monitoring of study data from a central location can be a useful supplement to, or even replacement for, on-site monitoring of multicenter trials. Central monitoring can be used to inform site monitoring visit schedules, as discussed above, but can also be used to directly identify many types of errors.

In principle, where source documents can be faxed to a central location, source data can be verified centrally. (Baigent et al. 2008; De 2011). This approach may be applied to selected items such as informed consent. However, it is generally not practical to transmit the full range of source documentation from a site. Accordingly, central monitoring strategies generally rely on automated and ad hoc analyses of routinely collected data. Central statistical monitoring is particularly effective at revealing fraudulent data by checking for unusual data patterns; one case report demonstrated successful detection of fraud by central review after site visits and independent audits both found no abnormalities (The ESPS2 Group 1997).

It is important to note that important insights can be obtained not only from patterns of clinical endpoints collected per protocol, but also from information about process (e.g., timeliness or completeness of reporting) and even structure (e.g., descriptive information about a sites geographic or institutional setting and staffing). Correlations among these variables can be used to develop risk indicators that can guide decisions about selective on-site monitoring.

Recently, one retrospective study sought to determine the potential for central monitoring to replace on-site monitoring (Bakobaki et al. 2012). The authors analyzed one year's worth of data from an AIDS microbicide trial to determine the proportion of 268 monitoring findings that could have been identified through central review. Twenty-eight percent of the monitoring findings were found in the central database and the authors concluded that an additional 67% of findings could have been identified had

NIAID CONFIDENTIAL

central checks been in place during the study. Thus, the study found that central monitoring could potentially have identified about 95% of the findings from on-site monitoring visits. Larger, prospective studies are needed to confirm these results.

Currently, about 83% of clinical trial sponsors use some form of centrally available data to evaluate performance of research sites, while only 12% use centralized monitoring to replace on-site visits (Morrison et al. 2011).

a. Approaches to Central Monitoring

There are two predominant approaches to central monitoring: key risk indicators (KRI) and central statistical monitoring. In the KRI approach, trial-specific indicators of quality, such as rates of reporting serious adverse events and aspects of protocol compliance, are identified on a trial specific basis. Centrally collected data fields are then developed for these indicators which can be monitored routinely. Deviations in the KRIs alert study leaders to potential problems at research sites. The primary disadvantage of KRIs is the need to develop or refine indicators for each study (Elsa et al. 2011).

Only limited information on KRI systems is publicly available as many sponsors develop proprietary systems. However, at a CTTI project meeting, a Novartis representative described that company's work on a system of key risk indicators called the Trend and Pattern Alert System (TAPAS).²¹ This system is conceived not as an automated substitute for on-site monitoring, but as part of an integrated quality management system that combines and balances centralized and site-based review activities to optimize quality while deploying resources efficiently.

As an alternative to KRI, central statistical monitoring examines all centrally available data elements, applying multivariate analyses to identify likely errors and potential fraud. While less specific than the KRI approach, the same central statistical monitoring plan can be used for multiple trials without substantial modification (Venet et al. 2012).

b. Electronic Data Capture

Some form of electronic data capture (EDC) is essential to make central monitoring cost-effective. The advantages of EDC over paper data collection include faster processing time, improved data integrity, and cost savings (Welker 2007). One modeling study found EDC to be 49-62% cheaper than paper collection systems (Pavlovic, Kern,

²¹ "Clinical Data Mining as a basis for an Information-Based Clinical QA Program", slide presentation to the CTTI by C. Grant Simmons, October 2010. (<https://www.ctti-clinicaltrials.org/website-administration/documents/Simmons-%20Clin%20data%20mining%20and%20risk%20based%20audit%20-%20CTTI%20Oct%202010.pdf>, accessed Nov. 21, 2012.)

and Miklavcic 2009). However, there are also barriers to implementing EDC. User motivation, inadequate technical support, access to technology (for trials in remote or less developed locations), regulatory hurdles, and implementation costs all hinder adoption (Welker 2007).

3. Alternative Monitoring Methods

Some sponsors use other, less common monitoring methods. For example, Chilengi and colleagues (2010) describe a cost-effective in-house system of monitoring in an academic setting. The approach is to train and employ the research center's own clinical staff as monitors for trials which are independent of their personal work activities. The authors explain that their monitoring system is more akin to mentoring and claim that it has been well-received by staff and has increased the quality of their trials. While such alternative monitoring approaches are uncommon, they may present workable solutions for certain types of sponsors.

D. Role of Trial Design, Training and Oversight in Ensuring Quality

Several factors not part of monitoring per se need to be considered in the overall quality management of clinical trials. The first is trial design as not only does the basic integrity of clinical trials depend on good design (Williams 2006), but quality assurance mechanisms may be most effective when built into trial protocols from the beginning (Clinical Trials Transformation Initiative 2012). This principle is often called "quality by design." In this spirit, CTTI work groups have been developing principles for a more comprehensive concept of quality by design that cuts across the full scope of clinical trial design and implementation.

Appropriate training of personnel, including clinical research site staff as well as monitors and auditors, is also critical (Williams 2006). Periodic retraining, including training in response to errors identified, helps ensure that protocols are followed and mistakes are noticed and corrected.

Oversight bodies also play an important role in quality management. Exact titles and configurations may vary, but in principle these include the Trial Management Committee (TMC), Trial Steering Committee (TSC), and Data Management Center (DMC). According to Baigent and colleagues (Baigent et al. 2008), the TMC is responsible for ensuring adequate systems for monitoring and detecting errors before trial initiation, while the TSC provides strategic advice to ensure appropriate data analysis. The DMC reviews unblinded data during the trial to determine if any protocol amendments are needed or if the trial needs to be stopped for safety concerns. Califf and colleagues (2003) noted that there are no common standards for DMCs, which results in variability in detection of safety issues.

NIAID CONFIDENTIAL

Finally, it should be noted that DMC's routinely compile information that is broadly relevant to quality management generally and assessment of site performance in particular. Automated completeness and consistency checks of submitted data, as well as information on timeliness of submission of required data elements, are used to provide performance improvement feedback to sites. Timeliness, appropriateness and effectiveness of site response to such feedback can be assessed as a further measure of site performance. Coordination of DMC oversight activities with independent monitoring is a potentially important aspect of comprehensive approaches to clinical trial quality management.

E. Empirical Studies of Monitoring Effectiveness

Several initiatives are underway to empirically determine the effectiveness of various types of monitoring. Two of these initiatives are discussed below.

1. ADAMON Project

The ADAMON²² (adapted monitoring) project is an ongoing prospective randomized controlled study comparing the effectiveness of a trial-specific adapted monitoring system to a more traditional intensive on-site monitoring approach. At the end of the study, results will be compared through 100% monitoring audits of both approaches. The ADAMON project will continue through November 2012. When complete, the ADAMON study may provide insight into the validity of one approach to adaptive monitoring.

The primary evaluation criterion adopted by the ADAMON project for comparing the two monitoring approaches is the impact on the following types of errors.

- Patient consent obtained either with gross negligence or not at all
- Safety-relevant or efficacy-relevant inclusion and exclusion criteria disregarded without prior agreement of the study director
- A serious adverse event (SAE) reported too late, inadequately, with gross negligence, or not at all
- Primary endpoint of the clinical trial avoidably obtained with negligence, inadequately, or not at all
- Significant departure from the protocol-specified treatment or protocol-specified observation or follow-up observation, without any compelling medical reasons

Secondary evaluation criteria include:

²² More information on the ADAMON project can be found on their website: http://www.adamon.de/ADAMON_EN/Projectdescription.aspx, accessed Nov. 20, 2012

NIAID CONFIDENTIAL

- Frequency of serious or critical violations of GCP not identified by monitoring.
- Frequency of serious or critical violations of GCP, individually by category

2. OPTIMON Project

Similar to the ADAMON study, the OPTIMON (optimization of monitoring) project is an ongoing study comparing two monitoring strategies: a traditional intensive on-site monitoring process and a monitoring regimen adapted to patient risk. While the OPTIMON study was ongoing, the group has published a validated risk-assessment scale for ranking protocols according to the risk posed to study participants (Journot et al. 2011). Validation of the OPTIMON risk-assessment scale involved independent ranking of 200 protocols by 49 experts. Risk-adapted monitoring plans for each risk level were also developed. Once full results of the OPTIMON study are published, it may be a useful reference for sponsors of clinical trials attempting to implement risk-adapted monitoring programs.

F. Previous Recommendations for Monitoring Best Practices

Beyond the guidelines provided by regulatory authorities, several independent groups have also recommended guidelines for acceptable clinical trial monitoring practices including the Society for Clinical Trials, the Pharmaceutical Research and Manufacturers of America (PhRMA), and CTTI (Knatterud et al. 1998; Usher 2010; Clinical Trials Transformation Initiative 2010). Each set of recommendations is based primarily on expert consensus and not empirical evidence.²³ Indeed, the recommendations from the Society for Clinical Trials begin with the acknowledgement, “although we have no hard evidence [...]” (Knatterud et al. 1998). The highlights of each of set of recommendations are described below.

a. Society for Clinical Trials

The guidelines for quality assurance recommended by the Society for Clinical Trials stress the importance of preventive measures, such as adequate staff training and well-defined eligibility criteria, to reduce the number of errors committed. The Society also recommends source data verification, either centrally or through on-site visits, of about 5-10% of patients on study. Further, the authors recommend monitoring the first few enrolled patients on each study at each site to improve early detection of errors and to provide a chance for additional training if necessary. The guidelines state that on-site monitoring is preferred because of the opportunity for monitors to provide feedback to site staff, as well as the chance to examine facilities. Finally, the Society recommended

²³ While CTTI did conduct a survey of monitoring practices, their recommendations are based on expert opinions.

NIAID CONFIDENTIAL

that journals publishing clinical trial results require the publication of quality management strategies to increase the focus on quality assurance.

b. PhRMA BioResearch Monitoring Committee

The PhRMA monitoring recommendations focus on trials with industry sponsors although PhRMA asserts that its guidelines are appropriate for government and academic sponsors as well. In general, PhRMA advocates a flexible approach where monitoring plans are tailored to the needs of individual studies. The recommendations also encourage the use of risk-based approaches that focus on the most important study parameters. The PhRMA committee further suggests that a variety of monitoring approaches may be appropriate and no single method should necessarily be used for every trial.

While PhRMA acknowledges the importance of verifying source data, they believe the focus should be on procedural and systematic problems at research sites that may affect the overall results of the study. Accordingly, central monitoring is viewed by PhRMA as useful because it enhances the ability to detect such critical errors. The PhRMA committee also suggests that site initiation visits are important to ensure sites adhere to protocol from the outset. After the initial visit, the authors recommend that on-site monitoring be driven by the initial monitoring plan as well as findings from central review of study data.

The PhRMA recommendations consider traditional on-site monitoring to be acceptable as well, assuming that all findings are followed up and corrective action plans are developed for significant errors. Finally, PhRMA advocates dialogue with regulatory agencies to allow regulators to provide sponsors with feedback regarding the adequacy of their quality management systems and to speed the review of pharmaceutical products.

c. Clinical Trials Transformation Initiative

The primary CTTI recommendations on monitoring revolve around the principle of building quality into the scientific and operational design and conduct of clinical trials and are couched in terms of general approaches rather than specific procedures:²⁴

- Focus on what matters
 - “Quality” is defined as the absence of errors that matter (i.e., errors that have a meaningful impact on patient safety or interpretation of results)

²⁴ “Results and Recommendations: Effective and Efficient Monitoring as a Component of Quality Assurance in the Conduct of Clinical Trials”, Clinical Trials Transformation Initiative (<https://www.ctti-clinicaltrials.org/website-administration/documents/Monitoring%20Project%20Results%20and%20Recommendations%20FINAL%20090311.pdf>, accessed Nov. 21, 2012).

NIAID CONFIDENTIAL

- Determine what matters for a specific trial
- Develop a quality management plan
 - Initiate plan in parallel with protocol development
 - Focus on areas of highest risk for generating errors that matter
 - Seek regulatory review of plan
- Assess performance on important parameters
 - Prospectively measure error rates on important parameters
 - Tailor monitoring approach (e.g., site visits, central, statistical) to the trial design and key quality objectives
- Improve training and procedures
 - Focus on measured parameters
- Report findings of quality management approach
 - Include issues found, actions taken, impact on analysis and interpretation of results
 - Incorporate into regulatory submissions and publications
 - Encourage inclusion in International Committee of Medical Journal Editors (ICMJE) requirements

G. Conclusion

Sponsors employ diverse methods to monitor their clinical trials, reflecting differences in interpretation of regulatory requirements, different degrees of risk-aversion, and variations in resource availability. While several authoritative entities have developed recommendations for best monitoring practices, these recommendations are based largely on expert consensus rather than empirical evidence because few controlled studies have compared the effectiveness and efficiency of different monitoring strategies.

Nevertheless, some points of consensus can be derived with regard to best practices.

- Risk-based monitoring that targets research sites, protocols or data fields which are more critical or more error-prone is gaining acceptance
- Central monitoring can be cost-effective and efficient at identifying certain types of errors
- Experts generally agree that 100% source data verification is not necessary for most trials

NIAID CONFIDENTIAL

- Quality management should be built into the design and conduct of all clinical trials.

6. Recommendations for Monitoring Improvements

Although not specifically requested by DAIDS/OCSO, STPI researchers, in the course of this feasibility study, identified certain recommendations for potential improvements in the monitoring program which are described briefly below.

A. Develop Management “Dashboard” of Monitoring Results

The DAIDS-ES system currently collects detailed real time data on monitoring results which is captured by observation code, site, protocol, site visit date and observation date. However, these data are not routinely converted into a convenient, user friendly series of reports, data tables or graphs that could be used by DAIDS/OCSO to track and manage site performance. Based on the analyses of monitoring data performed for this feasibility study, STPI researchers recommend that a series of standard reports be developed that present both current and historical DAIDS-ES data stratified by site, protocol, network and observation code. These reports could be updated in near real time (e.g., monthly) and made available through a user-friendly “dashboard” portal.

The primary recipients of these reports would be the DAIDS/OCSO Program Officers who could use the data to track every site and every protocol both for the most recent monitoring results and for any error trends in specific sites or protocols over time. The reports could even be designed to facilitate comparisons across sites and protocols. The reports would also enable Program Officers to identify substantial changes in specific types of observations which might warrant targeted corrective actions at certain sites or for individual protocols at all sites. Certain events, such as a sudden spike in one of the observation codes corresponding to a serious error, could be flagged by the system so that the Program Officer could see easily if the data indicated a situation that might warrant immediate attention.

Such a series of automated DAIDS-ES reports could be used to target detailed Program Officer review of site monitoring reports only to those sites and protocols where errors warranting follow-up action appeared in the automated reports. From the data analyzed in Chapter 2, it is clear that the number of errors identified for each protocol during each site visit is typically small. Therefore detailed review of each site monitoring report would not be necessary if the overall level and types of errors could be reviewed “at a glance” for each protocol at the site and no concerns were raised.

An automated DAIDS-ES reporting system would also provide DAIDS and network management the ability to easily track errors over time for both the overall system and for

the various networks. Negative trends could be investigated for underlying causes and positive trends could be investigated as a possible source of best practices.

B. Implement Adaptive Risk-Based Site Performance Monitoring

All DAIDS monitoring program stakeholders agree that targeting monitoring efforts to those sites most likely to make errors would be an important process improvement. However, developing a reliable process for the risk ranking sites was viewed as difficult primarily because site risk is not a static characteristic but changes over time in response to a number of different variables. Therefore, rather than attempting to implement such a static system, STPI researchers recommend that the need to focus monitoring efforts on the most risky sites should be a driving force behind developing the automated monitoring data reporting system described in the previous section.

If such an automated system were in place, sites that are experiencing a higher level of errors than other sites could be readily identified and more intense monitoring efforts directed at those sites. More importantly, the real time nature of the data would alert Program Officers to sites whose performance was beginning to decline. This could result not only in increased monitoring attention but also immediate interaction with the site to identify and potentially correct the problem. Finally, automated real time data would also allow the ready identification of sites which had improved their performance and therefore no longer needed an increased level of monitoring as well as sites with long term high performance which might need less intense monitoring than sites with average performance. Because the extent of the DAIDS monitoring effort is necessarily constrained by available funds, it is important to be able to shift resources rapidly to where they will have the greatest benefit.

C. Adapt Monitoring Effort to Trial Activity

DAIDS monitoring visits are driven by a “one size fits all” approach that targets the review of 20 PIDs at each site each quarter independent of the number of subjects on study or the number of active protocols at the site. Even if the total number of monitor days on site annually must remain constant due to resource constraints, STPI researchers recommend that those days be deployed in response to trial activity rather than on a fixed schedule. Such an approach should be possible now that each protocol has a defined percentage of PIDs that must be reviewed for IC/EC and for full record review. For example, each site could be examined quarterly to determine, based on the protocols and subjects on study in each protocol, how many PIDs need to be reviewed at that site in the upcoming visit to maintain the required IC/EC and full record review percentages. The number of monitors and days on site could then be established based on that number of PIDs.

NIAID CONFIDENTIAL

If the number of PIDs needing review to maintain the percentages is small, then it might be appropriate to skip the next quarterly visit to that site and review more PIDs at the subsequent visit. This would not result in the site permanently having a six month visit cycle but only be a one-time adjustment unless the trial activity remained low. If the number of PIDs to be reviewed were large, multiple monitors could be sent to the site or the visit extended. Because the number of PIDs that eventually will need to be reviewed for each protocol at each site to meet the monitoring goals is a fixed number, this should not change the absolute level of the overall DAIDS monitoring effort but simply adjust the way that effort is deployed.

Such an adaptive approach to monitoring would help ensure that the detection of errors was in line with the opportunity for errors (i.e., the number of protocol events that occurred) at the various sites. However, implementation of such a system would require two process changes. The first is to incorporate into the PID selection algorithm a calculation involving the total subjects on each protocol at each site, the percentage monitoring goals for each protocol, the total PID reviews required at the site to meet the goals for each protocol, and the number of PID reviews already completed at the site for each protocol to determine the PIDs that need to be reviewed at the upcoming visit to maintain the monitoring goals. Because all of these data are available electronically, it might not be too difficult to modify the algorithm to institute this change. The larger process change would be to transition the monitors from a stable schedule of four-day quarterly visits to each site to a potentially more irregular schedule of trips with varying duration.

D. Integrate Monitoring and SDMC Data

Automated analysis would also facilitate the integration of monitoring data with adverse event and data quality results from the network SDMCs. Such integration would allow management of the various networks and trials to more readily identify any sites or protocols that were having serious quality issues as evidenced by problems in both the SDMC and the monitoring data. In addition, a spike in negative results in the real time SDMC data for particular sites could alert Program Officers to focus greater monitoring effort on those sites.

E. Eliminate Monitoring Duplication of SDMC Data Quality Review

Two of the most prevalent errors in the monitoring data supplied to STPI were missed tests and missed visits, representing 21% and 13% of total observations. However, recording these observations is largely duplicative of data quality checks conducted by the DAIDS SDMCs. Missed tests and missed visits are tracked by the SDMCs in real time for 100% of subjects and the sites are contacted for resolution. Thus, the added value of re-identifying these errors for only a small percentage of the enrolled

NIAID CONFIDENTIAL

subjects and often long after the event occurred is not clear. STPI researchers therefore recommend that the requirement to identify missed visits and missed sites be eliminated from the monitoring procedure. The resulting savings in monitor time could then be redeployed for more value added activities, such as more intense monitoring of high-risk sites.

F. Prioritize PID Review for High-Risk Protocols

Based on discussions with monitoring staff, it appears that the order in which PID records are reviewed once the monitor is on site is not prioritized. The list is constructed alphanumerically by protocol and the monitor begins at the top of the list and works down. These discussions also indicated that on average only 80% of the PIDs on the list are actually reviewed during the visit so that protocols at the bottom of the list might not be reviewed. Because DAIDS has developed a process for identifying high-risk protocols that are to receive a more intense level of review, STPI researchers recommend that the protocol order on the PID list should be high-risk protocols first followed by medium- and low-risk protocols. Without this change, there is the potential that, during a given monitoring visit, PIDs from low- and medium-risk protocols would be reviewed while those from a high-risk protocol were missed because that protocol was serendipitously at the bottom of the list.

G. Eliminate Record Reviews Not Counted Toward Monitoring Goals

The current DAIDS-ES PID selection algorithm is designed such that not all PID records reviewed actually count toward meeting the monitoring goals for full record review. PIDs selected for ongoing record review are limited to those PIDs that have already been selected for IC/EC review. Thus if the monitoring goal for IC/EC is 50% and the full record review goal is 25%, the PIDs selected for ongoing record review are a subset of the 50% selected for IC/EC review. However, the understanding of STPI researchers is that in every round of the algorithm for a given protocol at a given site, the entire pool of IC/EC reviewed PIDs is available for ongoing record review. Thus, over multiple monitoring visits, the algorithm continues to select randomly from those PIDs until it has randomly chosen enough PIDs with records through the target visit to meet the full record review monitoring goal for that protocol.

The result of this practice is that when the full record review monitoring goal for a protocol is met, there are a number of partially reviewed PIDs from that protocol which have undergone record review through their last study visit but which have not reached the target visit and therefore do not count toward meeting the monitoring goal. Although review of these “extra” PIDs provides value in reviewing more records for errors, that value is distributed randomly across protocols. For example, a low-risk protocol might actually have more total PIDs reviewed than a high-risk protocol simply because by

NIAID CONFIDENTIAL

chance the algorithm picked up more PIDs at earlier stages of the protocol that were reviewed in part but did not reach the target visit before the monitoring goal was met.

There are two potential solutions to this problem that STPI researchers recommend for DAIDS consideration. The first would be that once a sufficient number of IC/EC reviewed PIDs have been picked to meet the full record review monitoring goal for that protocol, no new PIDs would be selected for ongoing record review unless additional subjects are enrolled at the site which would increase the total required PID number. Each monitoring visit, only new records for those previously selected PIDs, which will ultimately contribute to meeting the monitoring goal, will be reviewed unless more subjects are enrolled and thus new PIDs must be selected for full record review.

This change in practice would reduce the number of total PIDs reviewed and the number of records that would require review during a given visit for each protocol. This would allow DAIDS to increase the monitoring goals for the various risk levels so that more PIDs were required to be reviewed through the target visit. The advantage of this approach is that the number of PIDs reviewed for each protocol would remain correlated with the risk level of a protocol as opposed to being random. There would, however, be some disadvantages to this approach. The first is that the site would know in advance many of the PIDs to be reviewed and therefore might make special efforts to ensure those records were correct which might not reflect actual error rates overall. The second is that the records reviewed would tend to be for PIDs enrolled earlier in the life of the protocol at the site and that might introduce a bias.

A preferred, but clearly more complicated, approach would be that, to meet the full record review monitoring goal, records for different groups of study visits within the protocol would have to be reviewed for the number of PIDs represented by the monitoring goal. For example, in a high-risk protocol with 12 study visits, in advance of each monitoring visit, IC/EC reviewed PIDs would be separated into those having completed four, eight and 12 study visits. Up to 50% of the IC/EC reviewed PIDs in each group (i.e., up to 25% of total PIDs) would be selected randomly and their records reviewed for visits 1–4, 5–8, and 9–12, respectively. Once the 25% goal was met for any group of study visits, no more PIDs would be selected from that group. The end result would be that 25% of the PIDs on the protocol would be reviewed for each of the study visits but because of the random selection process, the different study visit groups should have at least some differences in the specific PIDs reviewed.

This change in practice would reduce the number of records that would require review during a given visit for each protocol which would allow the monitoring goals of the various risk levels to be increased. Moreover, this approach has some other advantages. First, it should increase the total number of PIDs reviewed compared to the first alternative described above. Second, the sites would have no more advance knowledge of which PIDs might be reviewed than they do under the current algorithm.

NIAID CONFIDENTIAL

This approach may still have a bias toward review of earlier enrolled PIDs as they will be overrepresented in the later study visit groups. However, that might be preferable to the random distribution of partially reviewed PIDs under the current system especially because the monitoring goals for each protocol risk level could be increased without increasing the overall monitoring effort.

Appendix

Stakeholders Interviewed

The STPI researchers interviewed the following stakeholders for the purpose of information gathering for this study.

DAIDS OCSO Program Officers (PO)

- Bill Sachau, OCSO PO
- Dara Potter, OCSO PO
- Gail Tauscher, OCSO PO

PPD Management

- Michelle Blanchard Vargas, Program Director
- Bambra Stokes, Senior Project Manager
- Mary Glessner, Lead Site Monitor
- Karen Hufham, Clinical Team Manager

Clinical Site Research Staff

- Joan Coetzee, Clinical Research Site Coordinator, IMPAACT, Cape Town, South Africa
- Cheryl Marcus, Clinical Trials Unit Coordinator, University of North Carolina Chapel Hill

Industry Representatives

- Peter Schiemann, Managing Partner, Wilder & Schiemann, formerly Global Head of Quality Risk Management, Roche
- Grant Simmons, Global Head Quality & Compliance Excellence, Novartis

Network Evaluation Committees

- Danielle Harden, HVTN Network Evaluation Committee Co-Chair
- Lindsey Baden, HVTN Network Evaluation Committee Co-Chair

NIAID CONFIDENTIAL

Statistics and Data Management Centers

- Linda Marillo, Chief Data Manager, IMPAACT and ACTG DMCs, Frontier Science Foundation
- Kris Coughlin, Chief Data Manager, IMPAACT and ACTG DMCs, Frontier Science Foundation
- Thula Weisel, Deputy Director of the Data Management Unit, Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Research Center
- Lynda Emel, Senior Project Manager in the Project Management Unit, Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Research Center
- Melita Romasco, Data Operations Manager, Data Operations Group, Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Research Center
- Maija Anderson, Clinical Affairs Safety Associate in the Clinical Affairs Group, Statistical Center for HIV/AIDS Research and Prevention, Fred Hutchinson Cancer Research Center

References

- Abraham, John, and Tim Reed. 2001. "Trading Risks for Markets: The International Harmonisation of Pharmaceuticals Regulation." *Health, Risk & Society* 3 (1):113–128. doi: 10.1080/713670172.
- Baigent, Colin, Frank E Harrell, Marc Buyse, Jonathan R Emberson, and Douglas G Altman. 2008. "Ensuring Trial Validity by Data Quality Assurance and Diversification of Monitoring Methods." *Clinical Trials* 5 (1):49–55. doi: 10.1177/1740774507087554.
- Bakobaki, J. M., M. Rauchenberger, N. Joffe, S. McCormack, S. Stenning, and S. Meredith. 2012. "The Potential for Central Monitoring Techniques to Replace On-Site Monitoring: Findings from an International Multi-centre Clinical Trial." *Clinical Trials* 9 (2):257–264. doi: 10.1177/1740774511427325.
- Brosteanu, O., P. Houben, K. Ihrig, C. Ohmann, U. Paulus, B. Pfistner, G. Schwarz, A. Streng-Hesse, and U. Zettelmeyer. 2009. "Risk Analysis and Risk Adapted On-Site Monitoring in Noncommercial Clinical Trials." *Clinical Trials* 6 (6):585–596. doi: 10.1177/1740774509347398.
- Califf, R. M., M. A. Morse, J. Wittes, S. N. Goodman, D. K. Nelson, D. L. DeMets, R. P. Iafrate, and J. Sugarman. 2003. "Toward Protecting the Safety of Participants in Clinical Trials." *Control Clinical Trials* 24 (3):256–71. doi: S0197245603000059 [pii].
- Chilengi, R., G. Ogetii, and T. Lang. 2010. "A Sensible Approach to Monitoring Trials: Finding Effective Solutions In-House." *WebmedCentral Clinical Trials* 1 (10).
- Clinical Trials Transformation Initiative. 2010. "Developing Effective Quality Systems in Clinical Trials: An Enlightened Approach." Summary of an expert meeting held October 13–14, 2010.
- . 2012. "Workshop on Quality Risk Management: Making Clinical Trials Fit for Purpose." Bethesda, MD.
- De, S. 2011. "Hybrid Approaches to Clinical Trial Monitoring: Practical Alternatives to 100% Source Data Verification." *Perspectives in Clinical Research* 2 (3):100–4. doi: 10.4103/2229-3485.83226 PCR-2-100 [pii].
- Elsa, V. M., H. C. Jemma, L. Martin, and A. Jane. 2011. "A Key Risk Indicator Approach to Central Statistical Monitoring in Multicentre Clinical Trials: Method Development in the Context of an Ongoing Large-Scale Randomized Trial." *Trials* 12 Suppl 1:A135. doi: 1745-6215-12-S1-A135 [pii] 10.1186/1745-6215-12-S1-A135.
- European Medicines Agency. 2011. "Reflection Paper on Risk Based Quality Management in Clinical Trials." London.

NIAID CONFIDENTIAL

- European Stroke Prevention Study Group. 1997. "European Stroke Prevention Study 2: Efficacy and Safety Data." *Journal of the Neurological Sciences* 151, Supplement:S1–S77.
- Food and Drug Administration. 2011. "Guidance for Industry: Oversight of Clinical Investigations: A Risk-Based Approach to Monitoring." Rockville, MD.
- Funning, Sandra, Anders Grahnén, Karin Eriksson, and Åsa Kettis-Linblad. 2009. "Quality Assurance within the Scope of Good Clinical Practice (GCP)—What Is the Cost of GCP-Related Activities? A Survey Within the Swedish Association of the Pharmaceutical Industry (LIF)'s Members." *The Quality Assurance Journal* 12 (1):3–7. doi: 10.1002/qaj.433.
- Grieve, Andrew P. 2012. "Source Data Verification by Statistical Sampling: Issues in Implementation." *Drug Information Journal* 46 (3):368–377. doi: 10.1177/0092861512442057.
- Grimes, D. A., D. Hubacher, K. Nanda, K. F. Schulz, D. Moher, and D. G. Altman. 2005. "The Good Clinical Practice Guideline: A Bronze Standard for Clinical Research." *Lancet* 366 (9480):172–4. doi: S0140-6736(05)66875-4 [pii]10.1016/S0140-6736(05)66875-4.
- ICH guideline for good clinical practice (E6)*. 1996. Media, PA: Barnett International/Parexel Clinical Training Group.
- Journot, V., J. P. Pignon, C. Gaultier, V. Daurat, A. Bouxin-Metro, B. Giraudeau, P. M. Preux, J. M. Treluyer, S. Chevret, V. Plattner, C. Thalamas, S. Clisant, P. Ravaud, G. Chene, and Grp Optimon Collaborative. 2011. "Validation of a Risk-Assessment Scale and a Risk-Adapted Monitoring Plan for Academic Clinical Research Studies: The Pre-Optimon Study." *Contemporary Clinical Trials* 32 (1):16–24. doi: 10.1016/j.cct.2010.10.001.
- Knatterud, G. L., F. W. Rockhold, S. L. George, F. B. Barton, C. E. Davis, W. R. Fairweather, T. Honohan, R. Mowery, and R. O'Neill. 1998. "Guidelines for Quality Assurance in Multicenter Trials: A Position Paper." *Control Clinical Trials* 19 (5):477–93. doi: S0197-2456(98)00033-6 [pii].
- Lang, Trudie, Phaik Yeong Cheah, and Nicholas J. White. 2011. "Clinical Research: Time for Sensible Global Guidelines." *The Lancet* 377 (9777):1553–1555.
- Lörstad, Mats. 2004. "Data Quality of the Clinical Trial Process: Costly Regulatory Compliance at the Expense of Scientific Proficiency." *The Quality Assurance Journal* 8 (3):177–182. doi: 10.1002/qaj.288.
- Morrison, B. W., C. J. Cochran, J. G. White, J. Harley, C. F. Kleppinger, A. Liu, J. T. Mitchel, D. F. Nickerson, C. R. Zacharias, J. M. Kramer, and J. D. Neaton. 2011. "Monitoring the Quality of Conduct of Clinical Trials: A Survey of Current Practices." *Clinical Trials* 8 (3):342–9. doi: 8/3/342 [pii] 10.1177/1740774511402703.
- Pavlovic, I., T. Kern, and D. Miklavcic. 2009. "Comparison of Paper-Based and Electronic Data Collection Process in Clinical Trials: Costs Simulation Study." *Contemporary Clinical Trials* 30 (4):300–316. doi: 10.1016/j.cct.2009.03.008.

NIAID CONFIDENTIAL

- Tantsyura, Vadim, Imogene Grimes, Jules Mitchel, Kaye Fendt, Sergiy Sirichenko, Joel Waters, Jim Crowe, and Barbara Tardiff. 2010. "Risk-Based Source Data Verification Approaches: Pros and Cons." *Drug Information Journal* 44 (6):745–756. doi: 10.1177/009286151004400611.
- Usher, Roland W. 2010. "PhRMA BioResearch Monitoring Committee Perspective on Acceptable Approaches for Clinical Trial Monitoring." *Drug Information Journal* 44 (4):477–483.
- Venet, David, Erik Doffagne, Tomasz Burzykowski, François Beckers, Yves Tellier, Eric Genevois-Marlin, Ursula Becker, Valerie Bee, Veronique Wilson, Catherine Legrand, and Marc Buyse. 2012. "A Statistical Approach to Central Monitoring of Data Quality in Clinical Trials." *Clinical Trials*. doi: 10.1177/1740774512447898.
- Welker, J. A. 2007. "Implementation of Electronic Data Capture Systems: Barriers and Solutions." *Contemporary Clinical Trials* 28 (3):329–36. doi: S1551-7144(07)00002-X [pii]10.1016/j.cct.2007.01.001.
- Williams, George W. 2006. "The Other Side of Clinical Trial Monitoring: Assuring Data Quality and Procedural Adherence." *Clinical Trials* 3 (6):530–537. doi: 10.1177/1740774506073104.