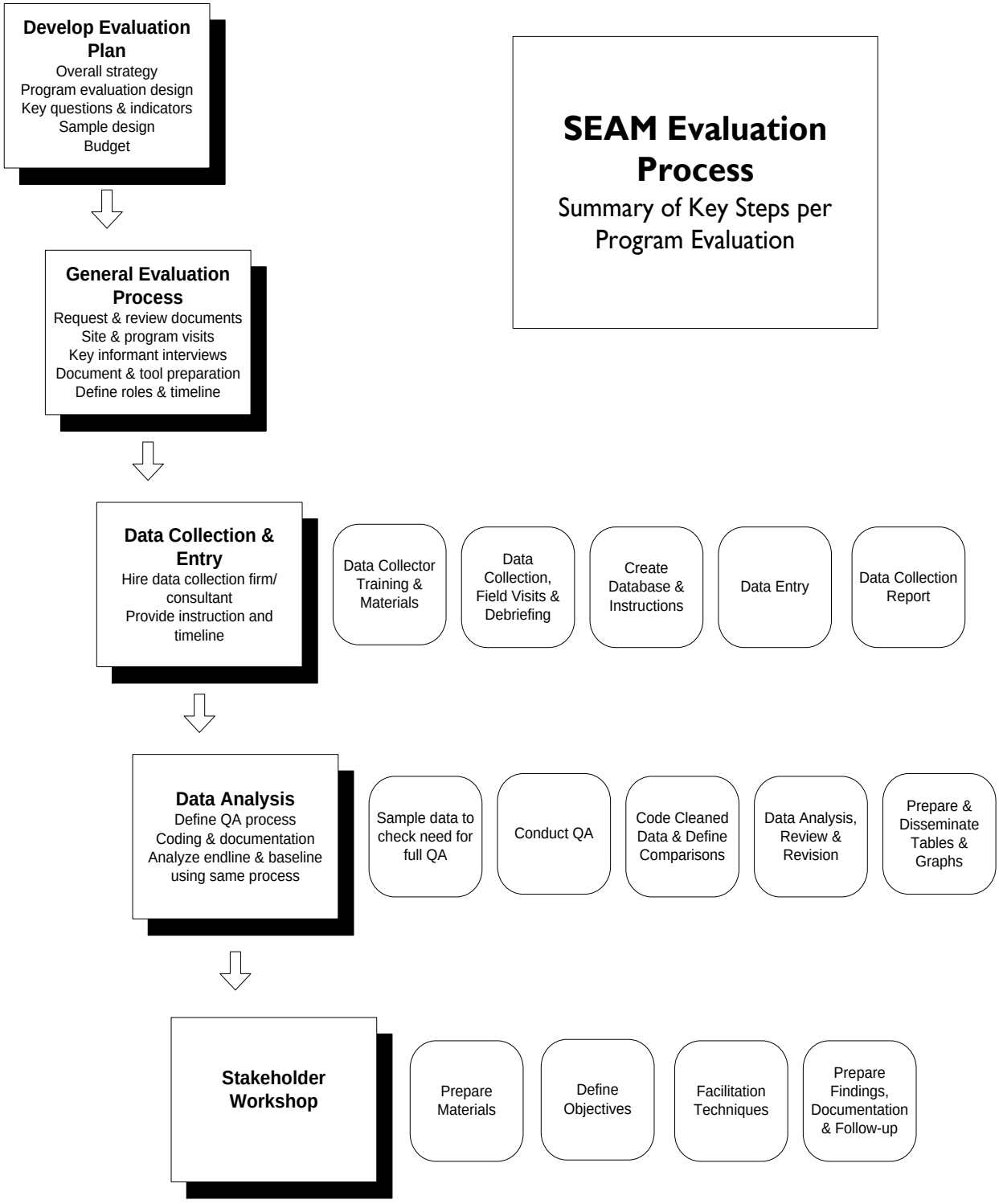
SEAM Evaluation Process - DRAFT

Activities, Issues, Level of Effort

Below is a description of steps taken in the SEAM final evaluation process. The text includes a set of evaluation steps, a description of activities undertaken to complete each step, and an estimated level of effort per program using the CAREshop evaluation as an example.



Develop Evaluation Plan:

Activity:
Evaluation plans and initial indicators were developed for baseline assessments. At the start of the endline survey process these plans were updated with program information. The indicator lists were reformulated to review the results in terms of the various factors that influence access to medicines.

Plans were developed for the following programs:

- Tanzania ADDO
- Tanzania Prime Vendor
- Tanzania QA
- Ghana CAREshops
- Ghana CPS
- El Salvador
- Kenya SHEF (developed later on special request)

Issues:
In some cases no final baseline report or findings were available, thus indicators and sample groups were selected based on consultation with program staff and best available information. In some cases these were revised when the evaluation got underway.

Evaluation plan level of effort: This activity took place in February/March 2004, and involved CPM and SEAM program management, consultant, SEAM program staff, and support staff. An estimated level of effort **per program** (using CAREshops as an example):

- CPM and SEAM program management: 1-2 days per person (overall discussions, plan review, conference calls with program staff)
- SEAM program staff: .5-1 day per person (review plan, revisions)
- Consultant: 1 day per person
- Support staff: 2-3 days per person (writing plan, revisions, budget)

Sample Design & Selection

Activity:
The evaluation plan included the basic elements of the comparison groups and sampling plan. The sample design included the following steps:

- Select sample groups (intervention & control)
    - Check that each group has similar characteristics (i.e. started intervention at similar times, same type of facility, etc)
- Determine sizes of sample groups
- Obtain complete and accurate account of all possible facilities in each group, and conduct random sample from each.
    - If a selected facility is closed or not at the location indicated during data collection, have a plan in advance for where to send data collectors in these cases. One option is to select at random and in advance some alternate facilities.

After the evaluation was completed, a sample comparison summary was prepared for each evaluation, showing the totals by study and by form. For example:

Regions/ Comparison Groups	Baseline Sample Size	Endline Sample Size
Total		

Number of facilities for which data were collected, by topic area/form										
Form	Price & Availability		Malaria		URTI		Registration		Satisfaction	
Study	Base line	End line	Base line	End line	Base line	End line	Base line	End line	Base line	End line
Intervention Group										
Control Group										

Issues:

- Sampling frames used at baseline were not always clear, and in some cases we were given incorrect information on what numbers and types of facilities had been visited. This was discovered at the time of the analysis and corrected for the endline documentation.
- In the baseline and in the practice sessions for the endline it was found that addresses for facilities were at times incorrect.
- Sampling for simulated clients was an issue and handled differently in the three drug outlet interventions depending on program preference and budget.
    - In the ADDO endline study the malaria simulated client was conducted at all facilities. Two weeks later the shops were visited again to conduct the URTI

simulated client scenario only. This was the easiest approach but was costly because of the extra time involved for data collectors.

- In the CAREshop endline study, the sampling for the simulated client scenarios was conducted separately. Two simulated client scenarios were conducted in the course of the endline study, one for malaria and one for upper respiratory tract infection. The shops were divided so that each facility would only receive one simulated client. However, since the sample size goal for each group was approximately 50, this would limit the simulated client to only 25 scenarios per scenario. To supplement the simulated client scenarios, 5 additional shops were randomly selected in each sample group and assigned only the simulated client scenario. No other data were collected at those additional shops. This approach caused some confusion among data collectors, and in fact some of the facilities slated to receive the URTI scenario actually received the malaria scenario.
- In Kenya, all facilities received both the malaria and URTI scenarios during the same visit, but from different data collectors and before other data collection forms were administered. Malaria scenarios were administered first. This may have limited the validity of the URTI scenario responses if shop keepers became aware that the scenario was a test or unusual.

Sampling design level of effort:
Preparation and implementation of the sampling design was incorporated into the evaluation planning process and into the data collector training. As such, it is not broken out as a separate level of effort.

However, in Kenya the selection of the control group required a field visit to identify locations of suitable control facilities. This visit was carried out prior to data collection and required 1-2 weeks of time, including preparation.

Baseline Data Collection Process (CAREshop example):

Baseline evaluation activity:
The CAREshops evaluation process began with a baseline evaluation that covered a key set of indicators and gathered information used for project planning. The baseline assessment was carried out in August 2003 in study and control facilities in Eastern and Volta regions (as noted above). The baseline evaluation included the following five areas of study:

- Stock availability & price
- Storage area adequacy survey
- Client exit interview (to assess purchasing practices)
- Malaria simulated client
- Household survey

Baseline issues:
Limited documentation was available for baseline processes. Forms and databases were not readily available, and in some cases incorrect forms or information were provided, such as the incorrect tracer list or the incorrect baseline database. To increase reliability, all baseline data were reanalyzed at endline.

Baseline level of effort:
Unknown for CAREshops.

Endline Data Collection Process (CAREshop example):

Endline evaluation activity overview:
The CAREshops evaluation incorporated many of the indicators collected in the baseline assessment. Some additional areas of study were assessed in the endline study (indicators and data collection tools are discussed in more detail below). The endline study included the following five areas of study:

- Stock availability & price
- Client exit interview (to assess satisfaction)
- Malaria simulated client
- Upper respiratory infection simulated client
- Product registration status survey

The endline data collection and analysis required a set of preparatory and participatory activities, including:

- Request & review documents and reports from the project
- Conduct site visits to GSMFEL, CAREshops and LCS
- Review performance monitoring data gathered in supervisory visits
- Conduct in-depth interviews of key actors in Volta, Eastern & Accra
- Hire/contract and plan with local data collection firm/consultants/coordinators
- Collect, review & revise baseline forms as available in preparation for the endline. Develop new forms and instructions as needed. Review with field staff and solicit expert advice as needed.
- Develop summary timeline & roles

Endline evaluation preparatory work level of effort:

- Consultant in country: 1-2 weeks (meet staff, site visits, document preparation, document review, interviews, logistics, travel time, etc)
- Consultant off-site: up to 1 week (document preparation, tool preparation & revision, logistics, etc)
- Data collection consultant: 2-3 months, not full time (Contract for prep work, training, data collection, data entry, and report)
- In-country program staff: 1-2 weeks (assist with above activities)
- Support staff: 1-2 weeks (assist with above activities)

Endline data collector training activity:
Data collector training materials were prepared based on the evaluation plan and data collection forms. The hired data collection firm was involved in planning for the training event and working with local country staff to coordinate. For CAREshops, the data collector training and tool testing was held in October 2004 in collaboration with GSMFEL, MSH and the local data collection coordinators. It was three days in length including practice in the field using the forms.

- Data Collector Training Objectives:
    - Build capacity to collect data accurately

- Provide hands-on practice in filling the forms
- Ensure consistency in data collection by giving everybody the same instructions and information
- Assign roles and responsibilities of each team member

Data collector training issues:

- Based on an assessment of the skills of the data collectors, 18 data collectors were selected to participate in the field study. Not all were determined to have the skill or time to commit to the data collection. Data collectors were split into six teams, a team leader was designated, and the team was assigned a list of facilities from the sample facilities. Packets with forms, per diem, logistics information and other essentials were provided to each team and data collector.
- The testing of the forms in the field during the training resulted in some changes to the data collection forms. It was necessary to allow for time for revision of the forms and copying on the last day of the training.

Endline data collector training level of effort:

- Consultant: 1 week (prepare for training, conduct training, follow-up with data collectors in field in cooperation with data collection firm, document training and data collection activities)
- Data collection consultant: see LOE above; 1 week for training component
- Data collectors: 3 days (18)
- In-country program staff: up to 1 week (assist with above activities)
- Support staff: 1-2 weeks (prepare training materials, revise as needed, assist with above activities)

Endline data collection & debriefing activity:

CAREshops data were collected from October 25 – November 5, 2004. During field activities regular scheduled calls were held with teams to review issues and provide updates. In some cases field visits to teams were made to verify and reinforce data collection approaches.

In-country data collection coordinators were requested to organize debriefings with data collectors upon returning from the field. Debriefings served to identify any problems with the data collection process, the need to return to certain facilities for follow-up, and any information about facilities and the process that would be relevant to data analysis. A debriefing session was held on November 12, 2004 with each data collection team. Debriefing notes and other related information were summarized by the data collection coordinator and provided to MSH.

The debriefings were documented and included information on the following:

- *Reconciliation of facilities visited and forms collected* - to verify that the control and intervention groups have a sufficient number of each form collected
- *Facilities where data collection was interrupted or cancelled* – substitute facilities were identified where necessary (see mention in sampling plan above)
- *Issues with forms* – for documentation and lessons learned

Endline data collector training/debriefing level of effort:

- Consultant: 1-2 days (available for input on issues that arise while data collectors are in the field)
- Data collection consultant: see LOE above; up to 1 week for field visits and other follow-up as needed, 1 week for debriefing and debriefing report
- Data collectors: 15 days (2 weeks for data collection, 1 day for debriefing)
- In-country program staff: 1-2 days (assist with above activities)
- Support staff: 1-2 days (assist with above activities)

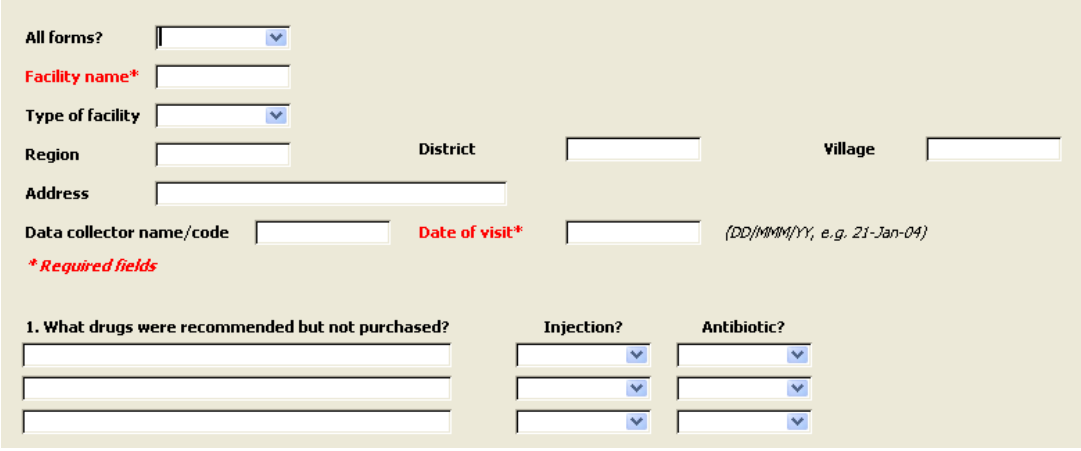
Endline Data Entry

Data entry activities:

Data entry databases were created using Access, and in a few cases, Excel. In general, a master database was created for each evaluation activity, and within the database a separate sub-database was created for each form, as pictured below:

Data entry supervisors were provided with detailed written instructions and screenshots on how to navigate and enter data into the databases.

Where possible, drop down boxes were provided in the databases to standardize answers, reduce spelling mistakes, and facilitate data entry (e.g. fields for type of facility, region, yes/no questions, and satisfaction scale questions).



Validation formulae were used where appropriate to increase the quality of the data entry (e.g. data collectors could only enter a range of 0-31 for the number days a tracer item was out of stock for the month of January, and any value above 31 was accepted).

If necessary, a conference call was made with the data collection/entry supervisors to go over the data collection and entry process and answer any outstanding questions. In Ghana and Tanzania, this was facilitated by in-country MSH staff.

Data entry issues:

- Data entry occurred shortly after data collection. Databases were prepared in advance of data collection, but had to be revised based on changes to forms that were made during data collector training.
- In some cases, the data entry personnel made changes to databases that had to be corrected at a later date. Communication needs to be open with the data collection firm to put a process in place when changes need to be made to a database so that errors do not result.
- Registration data collection was more intensive than planned. After the registration information was entered into the database, the products had to be checked against the TDA or Pharmacy Board registration database, if available, to verify registration status. We could not confirm how up-to-date the registration database was. There were delays in receiving the information from the national authorities, which were then input into the databases. This process needs to be planned for in advance.

Endline data entry level of effort:

- Consultant: up to 1 day (work with data collection firm in-country to review & practice on databases, pre- or post-training)
- Data collection consultant & data entry personnel: see LOE above; 1- 1 ½ months (for data entry, depending on number of forms and number of data entry personnel)
- In-country program staff: 1-2 days (assist with checking on portions of the data, such as medicine registration status)
- Support staff: up to 2-3 weeks (develop & test databases, assist with training on databases, assist with problems and queries as needed)

Endline Data Quality Assurance Checking

Quality assurance activities:
A quality assurance process was implemented to check the quality of data entry and to identify any missing or questionable data. All entries were checked against the original data collection forms. All queries were sent to MHS/Ghana and the data collection coordinator. In addition, the baseline databases were checked for any data quality questions, but the original baseline data collection forms were unavailable for double-checking of data entry.

Data collection coordinators in-country sent copies and/or originals of the data collection forms to the MSH/Arlington office. They were asked to keep a set for themselves. This facilitated checking of values on both ends for data entry QA. While this was an expense for the field office, it helped to provide a way to check any specific entries that looked incorrect.

Data in the databases were double checked against the copies of the data collection forms. A temporary employee was hired to perform this task for all SEAM program evaluations.

Facility information (e.g. facility name, facility type, region) for each form was verified, and corrected where necessary. This was done because in some cases the field staff did not keep track of how many facilities were visited, the locations, or the facility type, and as a result the information was often entered inconsistently. For example, a facility was entered as an ADDO when it was really a Duka la Dawa Baridi. By checking the facility information across forms as a first step, the facility identifiers could be checked for consistency, and counts could be made on the total number of forms collected and the total number of facilities surveyed. Any discrepancies were researched and corrected or documented.

Data collection/entry supervisors were contacted to address any queries that arose from the process of verifying the data in the databases

QA documentation process:
Notes were made directly on the copies of the data collection forms to keep track of any data collection errors, as well as changes made to the original data as a result of the QA process (e.g. standardizing facility identifiers). In addition, detailed notes were made to keep track of the data validation process for each form, meaning a different document was created to keep track of the price form, registration form, etc. These notes included the following elements:

- General comments (# forms by study group, etc)
- Corrections made to the databases when data were not consistent with the data collection forms
- Queries about specific issues (inconsistent data, data collection issues, missing information and data)
    - Notes on any queries directed to the field staff, and their responses
    - Clarifications, decisions and actions made to the database (these were highlighted to keep track of what was pending and what had been resolved)

The QA process ended when all pending questions were reviewed and answered, and updates were made to each database. If clarification of any question was not possible, in most cases the information was left in the database as it appears on the form. In few cases, information that was inconsistent or redundant was deleted and a note of this was made.

QA Issues:

- In some cases there was a delay in receiving forms or for other reasons QA did not happen immediately after data entry. Due to the time lag between data collection/entry and QA, it was hard to verify some of the information or ask data collectors for their recollection. Data collectors need to be advised in advance that they may be contacted at a future date for clarification.
- Data collection firms need to keep detailed track of each facility visited by comparison group and by form. Because this type of information from data collectors turned out to be unavailable or unreliable, an extra step of verifying all facility information was added to the QA process.
- The QA process uncovered significant errors in data entry. In the case of ADDO satisfaction entries, almost 50% of all forms had at least one and up to three errors in data entry. These issues led to the hiring of a temporary employee to provide assistance on QA, at extra cost, with no ramifications for the data entry firm. Provisions need to be incorporated into contracts and into the evaluation process to allow time, money, and incentives/penalties for major data entry quality problems.
- In the future, provide data collection firms with specific instructions for data quality assurance but maintain an internal mechanism for sampling and checking for data entry quality.

Endline QA level of effort:

- Consultant: up to 3 days (when in-country would assist with follow-up, would assist in making decisions around questions with the forms)
- Data collection consultant & data entry personnel: see LOE above; up to 3 days (to copy & send forms and databases, to follow-up on data entry questions)
- In-country program staff: up to 1-2 days (to follow-up on data entry questions, assist with copying & sending forms and databases as needed)
- Support staff: up to 2-3 weeks (QA process, documentation, follow-up and corrections)

Endline Data Analysis

Data analysis activities:
Excel was used for all the data coding and data analysis. Data entered in Access were exported to Excel after the appropriate QA had been done. Typically, there was only one Excel data analysis file per form used in an assessment. Corresponding baseline and endline assessments have separate data analysis files, however, all the final summary tables comparing the two datasets were kept in the endline data analysis files for each form.

All baseline indicator data were reanalyzed and in some cases recoded at the time of the endline to ensure consistency of calculation.

A data analyst prepared the analysis and preliminary tables once data were coded. These tables and data analysis calculations were reviewed by other staff and the external evaluator. When data analysis was finalized, tables and graphs were prepared and disseminated for review in Word format.

The following summary covers the basic data analysis elements, key issues that came up during the analysis, a summary of key steps for each area of study, and an estimated level of effort for data analysis using the CAREshops program as an example.

A typical data analysis file contained the following tabs and elements:

Notes or assumptions – This worksheet documented any queries, changes, additions, or corrections made to the dataset or analyses.

The original dataset (post QA) - In the case of the CPS medical record review, for example, data were entered directly into Excel. Although there was one worksheet per facility in the data entry file, these data were consolidated into one worksheet in the data analysis file. The original data entry worksheets were also included in the data analysis files, however they were hidden to avoid confusion. Instructions on how to display them were included in the notes worksheets.

Data removed from the analyses - For example, in one study two facilities were visited that should not have been part of a particular assessment. The data for these facilities could not be included in the analyses, so they were removed from the final dataset and documented in a separate worksheet as deleted data.

Decision matrice*s* – These document the criteria used to code STG decisions for simulated client data.

The final dataset used for the analyses – the cleaned and coded data

Pivot tables – Data were run using pivot tables based on the final dataset. These tables can be refreshed if changes are made to the final dataset.

Summary tables containing indicator results - Pivot tables were used to calculate many of the figures in these tables, however the summary tables present the data in a more useful format.

Graphs - Especially in later assessments, graphs were linked back to the summary tables so that they automatically update if any changes are made to the summary tables.

General data analysis issues:

Problems that caused delays, affected the quality of the analyses, or resulted in a loss of data:

Incorrect or conflicting facility type and region information – For example, an ADDO was coded as a DLDB, or vice versa.

Incomplete or missing data on forms - Data collectors often did not complete all the fields on a data collection form, or did not provide pertinent information in the comments section when comments were warranted.

Incomplete data entry – For example, half of the data for the qualitative questions had not been entered for an entire set of satisfaction forms.

Spelling (e.g. drug and facility names) – Pivot tables cannot effectively summarize data that are not standardized. For example, if cotrimoxazole had been spelled incorrectly in some cases, a pivot table would count the number of instances where each spelling occurred instead of providing a total count for all cotrimoxazole. Therefore, misspelled items had to be identified and standardized prior to data analysis (manually or using the "find and replace" function in Excel).

Incorrectly recorded price information - Misunderstandings on price were often not apparent until data collection was over, which increased the time required for the quality assurance process as well as the potential for loss of data. For example, if the data collectors recorded the price for one pack of paracetamol but did not record the number of pills in the pack, the price per unit figure would appear as an outlier as it would be much greater than expected. Similarly, in some cases contraceptive pills were recorded as 28 pills or as one pack, requiring standardization. This increased the amount of double checking and QA documentation, as well as the number of queries to local resources.

Data collected for forms of a drug that weren't specified on the tracer list - For example, a data collector would record data for the tablet form of a tracer item when the tracer list specifies that the tracer item is in suspension form.

Incorrect or inconsistent recording of pack size units for tracer items - For example, a data collector would record the pack size of a 100ml bottle of a tracer item as 1 unit when the tracer list specifies that 1 unit is equivalent 500ml bottle (this was a frequent problem for suspensions and oral contraceptive pills). These had to be standardized.

Determining the generic names for obscure local brand name drugs – For simulated client data, all the antibiotics and antimalarials had to be identified by local counterparts who were familiar with local brand names. In some cases there was space for this on the data collection forms, but the data collectors did not know the generic names and were unable to fill the information in. Ascertaining the generic names of medicines required extensive support from country staff.

Fields that are left blank, or marked as unavailable or not applicable – The formulas used to calculate indicators had to be adjusted to account for missing data. It is helpful to train data collectors and data entry personnel to use one standard approach for missing information – either blank or NA but not zero – otherwise multiple approaches will be applied.

Poor photocopies of the original forms - Records in pencil were often illegible and it was often difficult to get revised copies mailed or faxed.

Lengthy turnaround time on queries to local resources

Data analysis issues specific to the baseline assessments:
Problems that caused delays, affected the quality of the analyses, or resulted in a loss of data.

Original baseline data collection forms (filled out) were not available – All attempts made to locate the forms were unsuccessful. As such, questions about baseline data in the databases could not be answered or verified.

Poor recording keeping/documentation on the QA and data analysis processes for most of the baseline assessments – It was difficult to locate the baseline final raw data and data analysis files. When the files were finally received, it was unclear how much QA had been done on the data, and whether the data and analyses contained in those files were the most current (none of the files were accompanied by any documentation on the QA or analyses). In some cases where coding was required, there were no files available with the original data coding, so the baseline had to be re-coded to ensure consistency. In some cases, the wrong files were sent, but this was not discovered until data analysis began.

Length of time it took to obtain baseline files – In some instances the baseline data were received after the endline data collection and analyses were completed. This increased the length of time it took to complete the baseline-endline analyses, and reduced the amount of time available for queries to local resources regarding the baseline files.

Quality of baseline data – The baseline data could not be validated (see exceptions noted below) without the original data collection forms. One way to estimate the quality of the baseline data is to assume that the prevalence of errors in the baseline data were similar to what was found in the endline data through the QA process, though this method is probably an underestimation. Many of the data entry quality control steps put in place for

the endline assessments (e.g. designing databases to minimize data entry error, and providing detailed written instructions for data entry) were not present for the baseline assessments, so it is likely that the frequency of data entry errors would be greater than estimated. Regardless, the baseline-endline analyses are comparing unvalidated baseline data to clean endline data, which needs to be taken into consideration when interpreting the results of the baseline-endline analyses.

Correcting errors in the baseline analyses – In rare cases, it was possible to check the quality of the baseline data and make the necessary corrections without the original forms. This was the case for the price data – provided that the data collectors recorded data in the appropriate columns, and provided details where necessary, the units for each tracer item could be standardized. (In many cases they had not been. For example prices for a quantity of 10 tablets had been combined in some cases with prices for 1 tablet. This was identified and had to be re-coded and standardized at the time of the endline analysis.)

Standardizing the baseline data for the purpose of the baseline-endline analyses - All baseline data were reanalyzed to ensure that they went through the same QA (to the extent possible), formatting, analysis, and documentation process as the endline data. The following are examples of how the baseline data were standardized:

- Raw data and data analysis worksheets were arranged and formatted in the same manner as the endline files.
- "Notes" or "Assumptions" worksheets were included in the revised baseline data files to document any queries, corrections, or alterations.
- Outliers were corrected or removed based on the same criteria used in the endline analyses.
- Dose instructions for antimalarials and antibiotics dispensed during simulated client visits were evaluated based on the same STG criteria used in the endline analyses.
- Units for tracer items in the price per unit analysis were standardized to match the units used in the endline analysis.

Data analysis of price data:

1. Price data were standardized so that the comparison units were identical for a given tracer item (e.g. 1 tablet/capsule, 1 pack of 28 oral contraceptive pills, 1 bottle of 100ml).

2. Any data collected for incorrect dosage forms or strengths of a tracer item (for example suspension versus tablet or 200mg versus 500mg) were excluded from the analyses.

3. Price per unit data were calculated and outliers were identified.

4. Outliers were checked against the data collection forms to see if they were the result of errors in the data entry (in which case the outliers could be easily corrected by referencing the original data collection forms).

5. Data collection/entry supervisors were contacted to help resolve or evaluate any remaining outliers.

6. Rarely, outliers were excluded from analysis, however, the raw data for these outliers can be found in separate worksheets in those data analysis files.

Data analysis of availability data

If a tracer item was found to be expired it was considered "Not Available". In cases where expiry data were collected, data were checked to make sure that tracer items were correctly marked as unavailable if they were expired (though price analyses might include data for tracer items that were expired or otherwise unavailable).

Data analysis of simulated client data - malaria and URTI

1. Columns were added to the raw data to categorize the drugs that were dispensed or recommended but not purchased. The categories of interest were:

   Malaria
   - 1st line antimalarial sold
   - 2nd line antimalarial sold
   - 3rd line antimalarial sold (if applicable)
   - antibiotic sold
   - other drug sold
   - # of drugs sold during the encounter
   - antimalarial recommended but not sold
   - antibiotic recommended but not sold
   - # of drugs recommended but not sold during the encounter

   URTI
   - antibiotic sold
   - antimalarial sold
   - other drug sold
   - # of drugs sold during the encounter
   - antimalarial recommended but not sold
   - antibiotic recommended but not sold
   - # of drugs recommended but not sold during the encounter

2. Columns were added to categorize the dispensing of medicines according to the following criteria:

   Malaria
   - exactly according to STG (for age indicated in scenario)
   - consistent with STG (accounts for weight fluctuations – not found to be a useful metric)
   - sold in sufficient quantities to be able to comply with STG dose, regardless of whether or not the dose instructions were correct according to STG (not found to be a useful metric)

- ***Note:*** *The STG criteria used to code these columns are documented in the data analysis files. They are specific to each country's STGs.*

URTI

- exactly according to STG for pneumonia
- consistent with STG for pneumonia (accounts for weight fluctuations – not found to be a useful metric)
- in sufficient quantities according to the STG for pneumonia, regardless of whether or not the dose instructions were correct (not found to be a useful metric)
- ***Note:*** *The STG criteria used to code these columns are documented in the data analysis files and were only used in the ADDO analysis. The country's specific STGs were applied.*

3. When the simulated client datasets were exported from Access to Excel, the default format for the data was one row, or record, per drug dispensed. (For those facilities where the simulated client did not receive medication, only one row of data was generated for each facility. For those facilities where the client did receive medication, one record was generated for *each* drug that was dispensed at *each* of those facilities, e.g. if an attendant at one facility sold 5 items, the resulting Excel file would have 5 rows of data for that facility.) As the indicators for the simulated client scenarios evaluated data on the level of patient encounters, the exported dataset was collapsed so that all the information on the drugs dispensed during an encounter was summarized in a single record according to the methods below:
    - If the facility dispensed a drug of interest (an antimalarial or an antibiotic), the record for that drug was the one kept for that facility
    - If a facility dispensed both an antimalarial *and* an antibiotic, the record for the antimalarial was the one kept for that facility, and the appropriate column was marked indicating that an antibiotic was also dispensed
    - ***Note:*** *There was only one occasion where two antimalarials were dispensed by the same facility (CAREshop assessment), and notes on how the data were handled can be found in the data analysis file for this assessment*

    The resulting dataset contained only one record per facility and was used to calculate the indicators of interest. All simulated client data analysis files contain the original data, as well as an additional worksheet containing the data summarized by patient encounter.

4. Columns were added to the worksheet containing only one record per facility to categorize dispensing and referral practices according to the following criteria:
    - an antimalarial was dispensed and the client was referred
    - an antimalarial was dispensed, but the client was not referred
    - no antimalarial was dispensed, but the patient was referred
    - no antimalarial was dispensed, and the patient was not referred
    - at least one drug was dispensed and the client was referred
    - at least one drug was dispensed, but the client was not referred

- no drugs were dispensed, but the patient was referred
- no drugs were dispensed, and the patient was not referred

5. A column was added to the worksheet containing only one record per facility to indicate facilities where dispensers asked the simulated client about *both* of following:
    - the symptoms of the child
    - other medications the child may have taken

Data analysis of registration data

1. Data were checked to make sure that all the items surveyed were appropriate for analysis (e.g. medical supplies, herbal products, and other non-drug items were removed – a list of these items, if applicable, can be found in the registration data analysis files as "deleted items").

2. Data were checked to make sure that the registration and approval status of the drugs surveyed were entered consistently and correctly. For example, data were sorted to show the drugs listed by generic name. In some cases a drug from a certain manufacturer was listed as registered, but in another entry it had been left blank. This was re-checked with the field staff and corrected as appropriate.

3. Pivot tables were run and data were summarized according to the following criteria where applicable (different categories were used in different assessments):
    - Registered
    - Unregistered
    - Approved
    - Unapproved
    - Locally manufactured
    - Country of origin

Data analysis of stock out data

Stock out data were entered into Excel, not Access, to facilitate data analysis.

The total number of days possible for each tracer item to be out of stock (i.e. the total number of days in the time period surveyed) was adjusted to account for months where data were missing or unavailable (e.g. data were mistakenly not recorded, or the data collectors did not have access to the stock out records for that month).

Care was taken in data collector training to make sure that "0" or zero was entered only when there were no days out of stock for the tracer item, not for cases where no information was available.

Malaria medical record review data (Ghana CPS/DTC evaluation only)

1. The following columns were added to the data to characterize the drugs prescribed:
    - Was an antimalarial prescribed?

- What type of antimalarial was prescribed (i.e. what was the generic name)?
- Was the antimalarial an injectable?

2. The drugs prescribed were categorized as follows:
   - Chloroquine
   - Amodiaquine
   - S&P
   - Artemether derivative
   - Other antimalarial
   - Other drug

3. The prescriptions for drugs of interest were coded as being correct or incorrect according to national STGs and according to institutional STGs (which were not identical).

4. A new worksheet was created that summarized the raw data by patient encounter

5. A new worksheet was created that summarized the data by patient encounter (meaning that there was only one entry per encounter that contained all of the coding summarized above)

Hypertension medical record review data (Ghana CPS/DTC evaluation only)

1. The following columns were added to the data to characterize the drugs dispensed:
   - Was an antihypertensive prescribed?
   - What type of antihypertensive was prescribed (i.e. what was the generic name)?

2. The drugs prescribed were categorized as follows:
   - Atenolol
   - Bendrofluazide
   - Lisinopril
   - Methyldopate
   - Nifedipine
   - Propanolol
   - Other antihypertensives
   - Other drug

3. The prescriptions for drugs of interest were coded as being correct or incorrect according to national STGs and according to institutional STGs (which were not identical).

4. A new worksheet was created that summarized the raw data by patient encounter

5. A new worksheet was created that summarized the data by patient encounter.

Endline data analysis level of effort per program:

- Consultant: up to 1 week (guide analysis, define comparisons, review tables and graphs, conduct qualitative analysis as needed)
- Data analyst: 2-3 weeks assuming 2-3 days minimum per database to be analyzed (standardization/corrections, analysis, prepare & revise tables)
- Support staff: 2-3 weeks assuming 2-3 days per database (coding, review analysis and tables, prepare graphs and summary documents for field staff)

Stakeholder workshop

A stakeholder workshop was held following data analysis to provide an opportunity for a participatory review of the preliminary findings. The goals of the workshops were to come to a consensus on the findings and to discuss next steps. Workshop materials prepared in advance included the full set of tables and graphs prepared during data analysis. Workshops were generally held over a two-day period.

Workshop steps included

- Assist local counterparts with presentations on the program
- Prepare presentation with overview of workshop and evaluation
- Prepare handouts with all tables and graphs produced in the analysis
- Facilitate break-out sessions to identify key findings and reasons for key findings (fish-bone analysis used in some cases)
- Facilitate break-out sessions on next steps (force-field analysis used in some cases)
- Prepare summary of workshop discussions and findings

The workshop generated requests for additional data analysis and figures, requiring follow-up data analysis and coding. These revised materials were provided to program staff in the weeks following the workshops. In addition, findings were summarized from the workshop presentations, discussions, and small group presentations.

Stakeholder workshop & follow-up level of effort per program:

- Consultant: up to 1 week (travel time, facilitation, prepare materials & presentations, prepare summary)
- In-country program staff: 3-4 days (workshop logistics & preparation, prepare materials, participate in workshop)
- Data analyst: up to 1-2 weeks depending on number of requests following the workshop (conduct additional analysis, prepare & revise tables)
- Support staff: up to 1-2 weeks depending on number of requests following the workshop (travel time if attending workshop, workshop documentation, conduct & review additional analysis, prepare & revise tables, communicate and follow-up on changes)
- Workshop participants: 2-3 days (travel time & workshop participation)