Sampling and Data Collection

This additional resource is intended to supplement any documentation about performance measurement in this Workbook, including Project Step 1: Review, Collect and Analyze Project Data on page 82, and Program Step II: Establishing Performance Measurement Systems on page 57.

Information and Toolbox examples in this section explain how to:

- Construct a population sample
- Design a data collection tool
- Assign and train abstractors
- Validate results
Construct A Population Sample.

In many situations it is unrealistic and inefficient to collect data from every patient file. Data sampling allows teams to make inferences about a total patient population based on observations of a smaller subset of that group (the sample), saving both time and resources during data collection. Most facilities utilize random sampling, wherein medical records are drawn from the total population in such a way that each time a record is selected every record in the population has an equal opportunity to appear in the sample.

The goal of sampling is to reduce the amount of work involved in chart reviews for data collection, analysis and reporting. However, if the facility’s information system can access all the data that are needed for the indicators, a project team can skip ahead to the task: "Design a data collection tool," on page 147.

Defining Selection Criteria

Criteria determine who will be included in the sample. Examples of selection criteria include:

- **Location.** What facilities within the HIV care system will be included?
- **Gender.** Does the indicator apply to men, women, or both?
- **Age.** Does the indicator have particular age limits?
- **Patient conditions.** Does the indicator apply to all HIV-infected patients or is a specific diagnosis required? Is a confirmation of the diagnosis required, or is an empiric diagnosis acceptable? Do certain conditions make the patient ineligible?
- **Treatment status.** How many visits are required for eligibility? Must the patient currently be in treatment? Must the treatment have occurred within a specific time frame?

In the National HIVQUAL Project, the sample population includes male and female HIV patients who were active during a calendar year (for example, 1/1/2005 to 12/31/2005). Active cases are defined as those with at least two medical visits during the study period, with at least one visit during the last six months during the review period. Patients who expired before the end of the review period, but meet the visit criteria, are eligible for inclusion.

Additional Resource

Identifying Eligible Cases

Project teams will need to separate out the medical records that are eligible for measurement based on the selection criteria. The reporting capabilities of the facility’s computer system will determine how much work will need to be done by hand. When querying the facility management information system, the answers written for “defining your population focus” will need to be completed. For example, a system might specify all male and female HIV-infected patients over the age of 17 who made at least two visits between 1/1/2005 and 12/31/2005, with one of those visits in the last six months of the study period from 7/1/2005 to 12/31/2005. If the system will provide only a list of names for at least two visits within the year, but not specify within the last six months of the study period, then the list is reviewed and those names that did not have a visit within the last six months crossed off. Likewise, if the system will provide only a list of patients seen during the study period, then the list is carefully reviewed to identify the eligible cases.

To calculate the total number of eligible patients, you need to first divide the entire case list into male and female lists to ensure sufficient numbers for the GYN indicator. These lists will now provide you with two numbers, the total number of eligible males and the total number of eligible females. (The total eligible population is the sum of the two.)

Identifying Sample Size

The minimum sample size for an accurate measurement is based on the number of eligible cases. Some facilities use pre-existing sample tables to determine a project’s minimum sample size based on their own requirements.

The Toolbox on page 145 shows the HIVQUAL Sample Size Chart. It indicates the minimum number of records to be pulled for chart review for men and women based on the total number of eligible cases. The maximum number of records to be reviewed is 107 though it depends on the facility’s caseload. To ensure that women are properly represented in the GYN portion of the review, this table oversamples women. For example, if an agency has 55 eligible female patients and 90 eligible male patients (145 total eligible patients), then the minimum number of female records to be reviewed is 39. The number of male records to be reviewed is determined by subtracting the number of female records to be reviewed from the total number of records: 64 total records-39 female records=25 male records.
Toolbox:
HIVQUAL Sample Size Chart

This chart is based on a 90% confidence interval with an error width of 16% when using the minimum number of records.

<table>
<thead>
<tr>
<th>TOTAL ELIGIBLE POPULATION</th>
<th>MINIMUM TOTAL RECORDS</th>
<th>CHARTS TO PULL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 20</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>21 - 30</td>
<td>24</td>
<td>31</td>
</tr>
<tr>
<td>31 - 40</td>
<td>30</td>
<td>39</td>
</tr>
<tr>
<td>41 - 50</td>
<td>35</td>
<td>46</td>
</tr>
<tr>
<td>51 - 60</td>
<td>39</td>
<td>51</td>
</tr>
<tr>
<td>61 - 70</td>
<td>43</td>
<td>56</td>
</tr>
<tr>
<td>71 - 80</td>
<td>46</td>
<td>60</td>
</tr>
<tr>
<td>81 - 90</td>
<td>49</td>
<td>64</td>
</tr>
<tr>
<td>91 - 100</td>
<td>52</td>
<td>68</td>
</tr>
<tr>
<td>101 - 119</td>
<td>57</td>
<td>74</td>
</tr>
<tr>
<td>120 - 139</td>
<td>61</td>
<td>79</td>
</tr>
<tr>
<td>140 - 159</td>
<td>64</td>
<td>83</td>
</tr>
<tr>
<td>160 - 179</td>
<td>67</td>
<td>87</td>
</tr>
<tr>
<td>180 - 199</td>
<td>70</td>
<td>91</td>
</tr>
<tr>
<td>200 - 249</td>
<td>75</td>
<td>98</td>
</tr>
<tr>
<td>250 - 299</td>
<td>79</td>
<td>103</td>
</tr>
<tr>
<td>300 - 349</td>
<td>82</td>
<td>107</td>
</tr>
<tr>
<td>350 - 399</td>
<td>85</td>
<td>111</td>
</tr>
<tr>
<td>400 - 449</td>
<td>87</td>
<td>113</td>
</tr>
<tr>
<td>450 - 499</td>
<td>88</td>
<td>114</td>
</tr>
<tr>
<td>500 - 749</td>
<td>94</td>
<td>122</td>
</tr>
<tr>
<td>750 - 999</td>
<td>97</td>
<td>126</td>
</tr>
<tr>
<td>1000 - 4999</td>
<td>105</td>
<td>137</td>
</tr>
<tr>
<td>5000 or more</td>
<td>107</td>
<td>139</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TOTAL ELIGIBLE FEMALES</th>
<th>MINIMUM FEMALE RECORDS</th>
<th>CHARTS TO PULL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 20</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>21 - 30</td>
<td>24</td>
<td>31</td>
</tr>
<tr>
<td>31 - 40</td>
<td>30</td>
<td>39</td>
</tr>
<tr>
<td>41 - 50</td>
<td>35</td>
<td>46</td>
</tr>
<tr>
<td>51 - 60</td>
<td>39</td>
<td>51</td>
</tr>
<tr>
<td>61 - 70</td>
<td>43</td>
<td>56</td>
</tr>
<tr>
<td>71 - 80</td>
<td>46</td>
<td>60</td>
</tr>
<tr>
<td>81 - 90</td>
<td>49</td>
<td>64</td>
</tr>
<tr>
<td>91 - 100</td>
<td>52</td>
<td>68</td>
</tr>
<tr>
<td>101 - 119</td>
<td>57</td>
<td>74</td>
</tr>
<tr>
<td>120 - 139</td>
<td>61</td>
<td>79</td>
</tr>
<tr>
<td>140 - 159</td>
<td>64</td>
<td>83</td>
</tr>
<tr>
<td>160 - 179</td>
<td>67</td>
<td>87</td>
</tr>
<tr>
<td>180 - 199</td>
<td>70</td>
<td>91</td>
</tr>
<tr>
<td>200 - 249</td>
<td>75</td>
<td>98</td>
</tr>
<tr>
<td>250 - 299</td>
<td>79</td>
<td>103</td>
</tr>
<tr>
<td>300 - 349</td>
<td>82</td>
<td>107</td>
</tr>
<tr>
<td>350 - 399</td>
<td>85</td>
<td>111</td>
</tr>
<tr>
<td>400 - 449</td>
<td>87</td>
<td>113</td>
</tr>
<tr>
<td>450 - 499</td>
<td>88</td>
<td>114</td>
</tr>
<tr>
<td>500 - 749</td>
<td>94</td>
<td>122</td>
</tr>
<tr>
<td>750 - 999</td>
<td>97</td>
<td>126</td>
</tr>
<tr>
<td>1000 - 4999</td>
<td>105</td>
<td>137</td>
</tr>
<tr>
<td>5000 or more</td>
<td>107</td>
<td>139</td>
</tr>
</tbody>
</table>
Constructing The Sample

Using the number of eligible female patients, determine the minimum number of female records needed from the HIVQUAL Sample Size Chart (page 145). To determine the number of male records you need, subtract the minimum number of female records from the total minimum number of records.

Select the required number of records required from the total eligible records using a random process. Be aware that each medical record should have an equal chance of being included in the sample. Most spreadsheet programs offer random number tables.

The Tool box on page 148 provides a sample random number table and calculation instructions.

Notes
Design A Data Collection Tool.

Data collection consists of the tools and techniques used to collect baseline and follow-up data, typically from patient medical records. For example, a project team may collect data on how many patients received PCP prophylaxis in the previous year. The final percentage is a starting point for process improvement and a baseline for future measurements.

The goal of a data collection tool is to help obtain the most current, complete and accurate information possible. If data are collected with little regard for accuracy, project results are jeopardized.

Developing The Tool

To ensure that data are collected as intended, detailed instructions to guide abstractors through the data collection process need to be written. Agreement needs to be found on what data are needed and why. Tools such as data entry forms and procedure checklists help lower the margin for collection error.

Those responsible for the review process should first create a paper or electronic document listing the review criteria. The form should include eligibility criteria, and question/response parameters that define the ‘yes’ and ‘no’ responses, and N/A conditions. The form should be straightforward and concise to facilitate accurate data collection as well as any future re-measurement of data.

The following guidelines are useful when developing a data collection tool:

1. Decide on the evaluation questions and information needed to answer those questions.
2. Choose a format for collecting data such as questionnaire, discussion guide, interview, etc. consistent with the types of questions to be asked on the tool. For example, “Was a Pap smear performed on this female patient between 12/31/2005 and 1/1/2005?”
3. Agree on the interpretation of what the responses mean (‘yes’, ‘no’, ‘NA’).
4. Create results for each question. This will provide feedback on whether the results are meaningful. If not, revise the questions or responses or consider eliminating it.
5. Develop and field test a data collection sheet.

Additional Resource

You can download a software program, called HIVQUAL3, at www.hivqual.org at no cost. The software that includes indicators based on clinical practice guidelines provides an efficient means of measuring and reporting clinical performance.
Toolbox: Random Number Table

To use random numbers for case selection, imagine that the list of eligible patients either male or female, not both, has been numbered sequentially, starting with one. Based on your sample size, select those cases whose assigned numbers correspond with the random numbers listed in the appropriate random number calculation. If a random number exceeds the final number on the list, continue counting sequentially from the top of the list, including only those cases that have not yet been chosen. An example follows on page 150.

Random Number Table

Eligible Cases=21-30; Minimum Total Records=24
1, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30
Eligible Cases=31-40; Minimum Total Records=30
1, 2, 3, 4, 5, 6, 9, 11, 12, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 31, 33, 34, 35, 37, 40, 41
Eligible Cases=41-50; Minimum Total Records=35
1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 14, 15, 17, 20, 22, 23, 24, 28, 30, 31, 32, 34, 35, 37, 38, 39, 40, 42, 43, 45, 46, 47, 48, 50
Eligible Cases=51-60; Minimum Total Records=39
1, 2, 3, 4, 5, 6, 8, 9, 12, 14, 16, 17, 19, 21, 26, 27, 28, 29, 30, 31, 32, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 47, 48, 52, 54, 55, 56, 59, 60
Eligible Cases=61-70; Minimum Total Records=43
1, 2, 4, 7, 9, 10, 12, 13, 14, 16, 19, 22, 24, 25, 26, 27, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 47, 48, 52, 54, 55, 56, 57, 60
Eligible Cases=71-80; Minimum Total Records=46
2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 24, 27, 30, 32, 33, 36, 37, 41, 44, 48, 49, 50, 53, 54, 55, 56, 57, 59, 60, 62, 63, 66, 69, 70, 72, 76, 77, 79, 80
Eligible Cases=81-90; Minimum Total Records=49
2, 3, 5, 6, 9, 10, 11, 12, 13, 16, 17, 19, 20, 23, 26, 29, 34, 35, 38, 42, 43, 44, 46, 48, 50, 52, 54, 56, 57, 58, 58, 59, 60, 61, 63, 64, 65, 66, 68, 69, 70, 71, 73, 76, 77, 80, 81, 84, 86
Eligible Cases=91-100; Minimum Total Records=52
1, 2, 6, 9, 10, 13, 14, 15, 19, 22, 26, 28, 30, 32, 33, 34, 35, 37, 38, 40, 42, 43, 44, 45, 46, 47, 48, 51, 54, 55, 56, 58, 59, 60, 61, 70, 71, 74, 75, 79, 80, 81, 85, 86, 88, 92, 93, 94, 96, 97, 98, 99
Toolbox: Random Number Table Calculations

A team pulls a sample out of a total of 46 eligible records (20 male and 26 female patients).
The random number calculation for females is based on the random number table as follows:
Total eligible female records=26; Minimum female records=24
Random number calculations: 1, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29

To begin, the team numbers the list of eligible female records from 1 to 26 and selects the first round of records as indicated by the random number table calculations: 1, 4, 5, 6, 7, 9, 10, 11, 13, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, and 26. Since the next random number is 27, the team continues counting from the top of the record list—but this time excludes those records that have already been selected.
Assign And Train Abstractors.

Data abstractors are those HIV staff that collect measurement data from the sample.

Assigning Abstractors

At a minimum, staff members chosen to be abstractors should be familiar with medical records and relevant terminology. If possible, providers should be engaged in a peer-review process to broaden the sense of ownership and learn from each other. Other factors to consider are availability and comfort level working with data.

The collection process should have clearly defined roles. A single staff member might oversee the data collection process while others remain available to answer questions that arise.

Since data extraction from charts can be tedious for one person, many HIV facilities find it helpful to assign a few HIV clinical staff to do this activity together for a few hours at a time. The benefits of this approach are:

- Completing the task in a timely manner.
- Applying the same interpretation to answering the questions.
- Expanding understanding of both the process of data extraction and identification of issues involved.

Training Abstractors

Whether the project team collects the data or other staff is involved, some time for an informal training session should be scheduled as close to the actual time of data collection as possible. A designated staff member should begin by reviewing the basic purpose of performance measurement and the specific clinical aspects of the subject under review. Abstractors should be given a ‘walk through’ of the data collection process. Patient eligibility criteria and where to find required information in patient charts should be clearly indicated.

Additional Resource

For guidance in teaching small groups about the characteristics of a well-designed data collection system, see the HIVQUAL Group Learning Guide “Data Collection” exercise. The exercise could also be used to assess the strengths and weaknesses of an existing data collection system. You can download this publication at www.hivqual.org.
## Toolbox: Data Collection Tool

### ARV Management (Eligibility: All Patients on ARV Therapy)

For each trimester (every 4 months), how do you assess the patient’s stability?

- **Stable:** Was a viral load performed within each trimester?  
  - Yes  
  - No

- **Unstable:**
  - If ARV medication was changed, was a viral load performed within 8 weeks?  
    - Yes  
    - No
  - If ARV medication was stopped, were decision and clinical follow up documented within three months?  
    - Yes  
    - No
  - If ARV medication unchanged, was justification documented?  
    - Yes  
    - No

If medication was started or changed in this trimester, did the patient get treatment education?  
- Yes  
- No

### Adherence to ARV Therapy (Eligibility: All Patients on ARV Therapy)

Was adherence discussed with the patient each trimester (every 4 months)?

- Yes: Adherence discussion documented
- No: No documentation of adherence discussion

Was an adherence problem identified?

- Yes (Was the adherence problem addressed?)
  - Yes  
  - No
- No

### HIV Specialist Care

Was the patient seen by an HIV Specialist at minimum every 4 months?

- Yes: Patient was seen at least every 4 months by an HIV Specialist.
- No: Patient was not seen by an HIV Specialist.

### CD4 Cell Count

Was a CD4 count performed within each trimester (at minimum 3 times/12 months)?

- Yes: A CD4 count was documented in each trimester.
- No: A CD4 count was not documented in each trimester.

### Viral Load

Was a viral load performed within each trimester (at minimum 3 times/12 months)?

- Yes: A viral load was documented in each trimester.
- No: A viral load was not documented in each trimester.

### Lipid Screening (Eligibility: All Patients on ARV Therapy)

Was a Lipid Profile Done?

- Yes: Lipid profile was done within the last 12 months.
- No: Lipid profile was not done within the last 12 months.

### PCP Prophylaxis

Did the patient with fewer than 200 CD4 cells and no sustained CD4 cell increase > 200 during the last 6 months of the review period receive PCP prophylaxis?

- Yes: The patient received PCP prophylaxis.
- No: The patient did not receive PCP prophylaxis.

### MAC Prophylaxis (Eligibility: CD4 cells < 50 and no sustained CD4 cell increase to > 50 during the last 3 months of review period)

Did the patient receive MAC prophylaxis?

- Yes: The patient received MAC prophylaxis. (CBC done within 6 months?)
  - Yes
  - No
- No: The patient did not receive MAC prophylaxis.
**Toolbox: Data Collection Tool...Continued**

**TB Screening** (Eligibility: HIV- infected patients without a history of previous TB treatment or a history of a positive PPD)

- Was PPD testing performed (placed and read within 72 hours) in the past 12 months?
  - Yes: PPD screening was performed (placed and read).
  - No: PPD screening was not performed (or it was placed but results not documented).

**Hepatitis C (HCV) Screening**

- Was the patient’s hepatitis C status known?
  - Yes: HCV screening positive
    - Was patient’s hepatitis A status known?
      - Yes, seropositive
        - Was alcohol counseling documented within the last 12 months?
          - Yes
          - No
      - Yes, seronegative (Was patient offered hepatitis A vaccination?)
        - Yes
        - No
    - No
  - Yes: HCV screening negative
  - No: No screening performed

**Ophthalmology Exam** (Eligibility: CD4 cells < 50)

- Was an ophthalmology exam documented within the last 12 months?
  - Yes: An exam was done within the last 12 months.
  - No: An exam was not done within the last 12 months.

**Annual Pelvic Exam**

- Was a pelvic exam performed within the past 12 months?
  - Yes: A pelvic exam was recorded.
    - Pap done (Abnormal?)
      - Yes
      - No
    - Gonorrhea culture done?
    - Chlamydia screening done?
  - No: A pelvic exam was not recorded within the past 12 months.

**Annual Syphilis Serology**

- Was a syphilis serology performed within the past 12 months?
  - Yes: A syphilis serology was performed. (If yes, was a confirmation test performed?)
    - Yes
    - No
  - No: A syphilis serology was not performed within the past 12 months.

**Annual Discussion of Substance Use**

- Was substance use discussed with the patient during the past 12 months?
  - Yes: Substance use was discussed.
  - Current user (within 6 months)
    - If a substance was injected, was safer injection/needle exchange addressed?
      - Yes
      - No
    - Patient in treatment during review period?
      - Yes
      - No
    - If no:
      - Treatment discussed; no referral made
      - Treatment discussed; referral made
      - Treatment not discussed
  - Past user only (last use over 6 months); Prevention/ongoing treatment discussed?
    - Yes
    - No
  - No current use (within 6 months) or past use (over 6 months) identified
  - No: Substance use was not discussed with the patient during the past 12 months.
**Toolbox:**

Data Collection Tool...*Continued*

**Mental Health Screening**

- Was a mental health screening performed within the last 12 months? □ Yes □ No
- Was a need for mental health referral identified? □ Yes □ No
- Was referral to a mental health provider made? □ Yes □ No

**Annual Discussion of Tobacco Use**

- Was tobacco use discussed with the patient during the past 12 months?  
  □ Yes: Tobacco use was discussed. □ No: Tobacco use was not discussed.

**Annual Dental Exam**

- Was a dental exam performed within the last 12 months?  
  □ Yes: A dental exam was performed on ___ /___ /____.  
  □ No: A dental exam was not performed.

**Comprehensive Case Management Assessment**

- Was a comprehensive case management assessment of client needs performed within 30 days of initial client contact? □ Yes □ No

**Case Management Service Plan Development**

- Was a case management service plan developed within 45 days of initial client contact?  
  □ Yes: Service plan developed within 45 days of initial client contact.  
  □ No: No documentation that service plan was developed within 45 days of initial client contact.

**Case Management Follow-Up on Service Plan Goals and Referrals**

- For each identified client need area, is there follow-up regarding service plan goals every 120 days until achieved? □ Yes □ No
- For each identified client need area, is there follow-up regarding service plan referrals every 120 days until achieved? □ Yes □ No

**Case Management Coordination of Services**

- Is coordination of services documented (i.e., case conference or progress notes, other activities involving coordination of services) on at least a quarterly basis? □ Yes □ No
Conducting A Dry Run

If possible, a dry run of the measurement process should be conducted to assess its efficiency and to ensure that the abstractors are able to get the information they need from the sample before starting the full-scale measurement. Using two to three records from the total sample, abstractors should collect the required information as explained during the training session. The project team or quality committee should meet with the abstractors to review the results and collect feedback about the process. Based on this debriefing, the measurement can be tweaked as needed. An additional dry run of the revised process may be conducted depending on the changes made.

Notes
Validate Results.

It is good practice to validate the data collection results to ensure the results are correct. An inter-rater reliability testing procedure can accomplish this.

Reliability is defined as the consistency of an instrument’s measurements when used under the same conditions with the same subjects. The reliability of a measure is important for ensuring comparability of results over time. Frequently used in medical record abstraction, inter-rater reliability is defined as the reliability between two or more abstractors reviewing the same records.

Consider the following simple examples of inter-rater reliability testing. If one person is responsible for chart review, identify another person to look at approximately 8-10 charts to assess the chart abstraction process using consistent project data collection tools. If multiple people are responsible for chart review, then each can check a small sample (2-4) of the others’ charts. Apply the same process for the data entry task.

Based on abstractor feedback and reliability data, the quality management committee and/or improvement project team should make any necessary changes to the data collection process before the next measurement effort. This may require that team members modify the process and its related forms, retrain abstractors, and/or reconsider the chosen indicator as a critical aspect of care.

Notes

Additional Resource

The following resource will provide you additional information on reliability testing and statistical analysis beyond the scope of this document: Agresti, Alan. An Introduction to Categorical Data Analysis, Wiley Series in Probability and Statistics. Applied Probability and Statistics.