Data Quality Review (DQR) Data Verification and System Assessment Workshop

Session 14

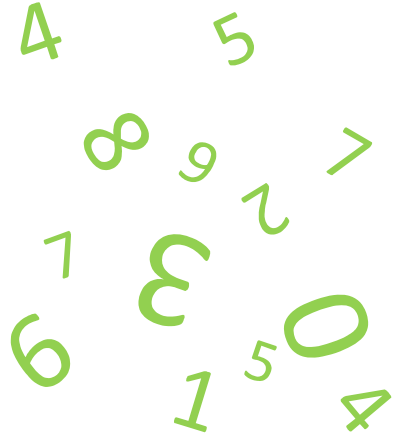
Data Analysis

World Health Organization

USAID
FROM THE AMERICAN PEOPLE

MEASURE Evaluation

The Global Fund



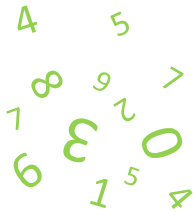
Learning Objective

To become familiar with techniques for data analysis of DV/SA survey results. Specifically, by the end of the session the participants will:

- Learn to populate and use automated data analysis tools (Chartbooks)
- Understand indicator calculation in the preparation of data for use in automated tools
- Understand weighting of data and specifics for weighting of different indicators

SESSION 14

Data Analysis



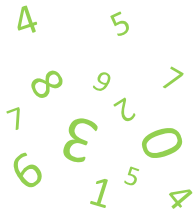
Calculate DV/SA Indicators

SESSION 14

Data Analysis

Calculate indicators from DV/SA data elements

- Once the DV/SA data have been compiled, de-duplicated and cleaned we need to calculate the indicators used in the analysis.
- The Facility and District DV/SA CSPro applications come with a batch edit program file prepared to automatically calculate the indicators.
- Navigate to FACILITY_DV_SA\Batch\Batch_3\Facility DV SA Indicators.bch and double click on the file.
- The left pane of the batch editor shows the data elements in the FACILITY_DV_DA data dictionary.
- The pane on the right shows the editor with the indicators program



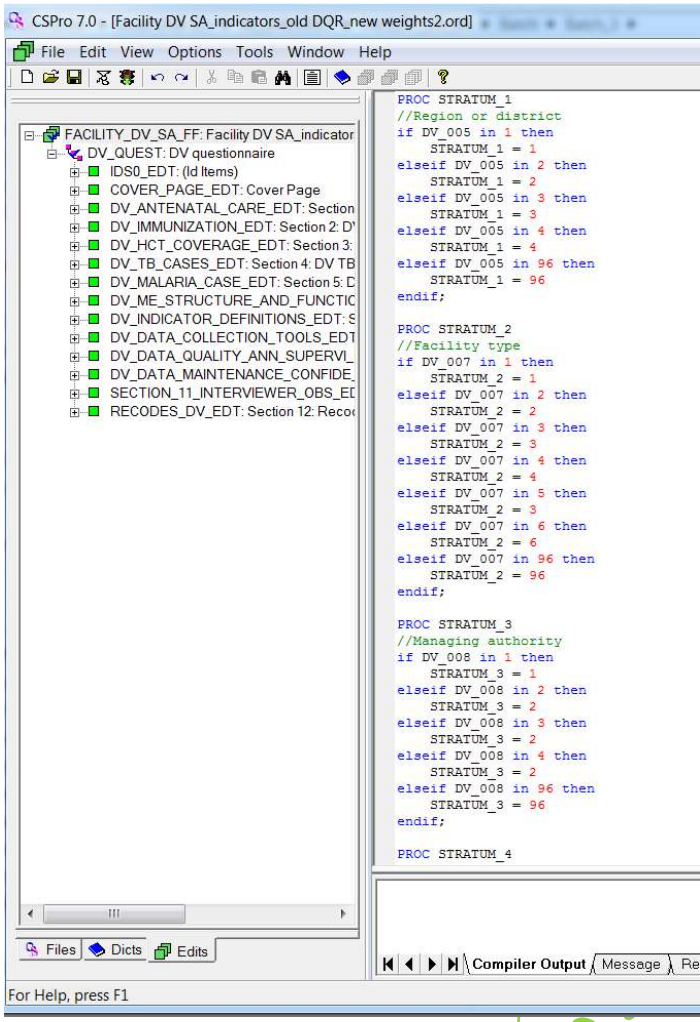
Calculate DV/SA Indicators

SESSION 14

Data Analysis

Calculate indicators from DV SA data elements

- Notice the "STRATUM" variables – these will require local adaptation corresponding to the stratifiers in the sampling methodology.
  - STRATUM 1 is for recodes of the subnational units (e.g. Region or District)
  - STRATUM 2 is for recodes of Facility Type
  - STRATUM 3 is for recodes of Management Authority, and
  - STRATUM 4 is for recodes of Urban/Rural
- Scroll down in the editor pane. Notice the variables for Weight.



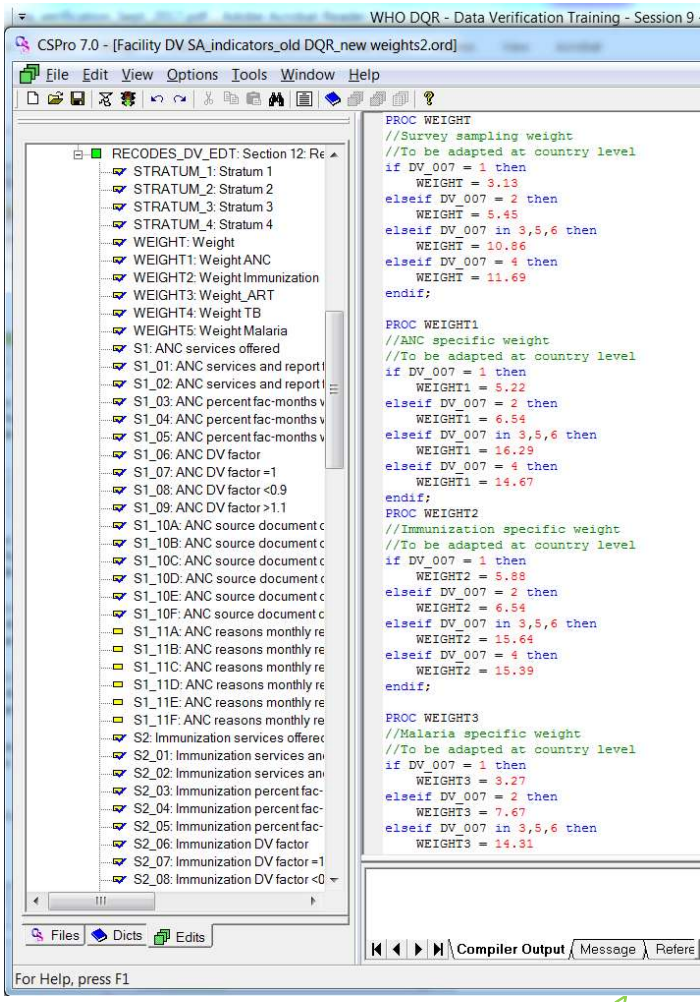
Calculate DV/SA Indicators

SESSION 14

Data Analysis

Program-specific weights

- There is a generic weighting variable called "WEIGHT". This applies weights to sample estimates that apply to system wide indicators (like the System Assessment).
- WEIGHT1, WEIGHT2, WEIGHT3, etc. are program-specific weights.
- Each program-specific weighting variable is derived from the distribution of health facilities in the sample that provide that particular service, and the service volume for the indicator.
- The weights are program-specific since the different health program have different service volume, and not all health facilities offer all the services.



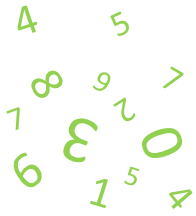
Calculate DV /SA Indicators

SESSION 14

Data Analysis

Weighting of sample estimates

- When we sample health facilities, we will make inferences to the whole population of health facilities based on the information gathered from a small proportion.
- The sample needs should be as similar to the larger population of facilities as possible so that our estimates are precise.
- However, the sample can never be exactly like the larger population of facilities, even if we sample well, since we don't always survey every facility, and sometimes we add other facilities.
- However, we can adjust the sample after the fact to make it more like the larger population of facilities. This adjustment is done through the weighting of the estimates.



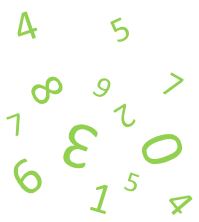
Calculate DV/SA Indicators

SESSION 14

Data Analysis

Weighting of sample estimates

- When we weight the sample estimates we will be adjusting the results based on the several factors;
    1) Distribution of facility type in the country
    2) Non-coverage (facilities that are supposed to offer the service, but don't actually offer the service)
    3) Non-response (facilities that are supposed to report to the HMIS, but don't actually report to the HMIS)
    4) Service volume (i.e. the indicator value for a given month or quarter).
- The WHO Data Quality Review (DQR) Framework document, Module 3 – Data Verification and System Assessment discussing weighting of sample estimates in Annex 5: Sampling Methods and Concerns.
- As in the guide, a grid can be prepared in Excel to facilitate the calculation of the program-specific weights.



Calculate DV/SA Indicators

SESSION 14

Data Analysis

Weighting of sample estimates

Stratum	Facilities in the country (A)	Facilities in the survey sample (B)	Facilities in the survey sample providing vaccination service coverage (C)	Facilities in the sample providing vaccination services & responding to the HMIS (both source and monthly report are available in "Month X") (D)	Probability of sampling each facility by facility type (E= n/N or B/A)	Sampling weight of each facility by facility type (F=1/E)	Factor adjusting for non-coverage of vaccination services (G=B/C)	Factor adjusting for non-response (H=C/D)	Number of vaccinations in "Month X" (excluding vaccinations for months for which either the source document or the report were not available)		Weighted number of vaccinations in "Month X" (adjusted for non-coverage & non-response)		Crude verification factor (VF=I/J)	Weighted verification factor (adjusted for non-coverage & non-response) (VFadj=K/L)
									Recounted in sample (I)	Reported (J)	Recounted in sample (K)	Reported (L)		
General hospitals	185	65	58	48	0.351	2.849	1.121	1.208	2650	3300	10 214	12 719	0.803	0.803
Reference health centres	175	65	56	52	0.371	2.695	1.161	1.077	2650	3300	8918	11 106	0.803	0.803
Health centres	400	130	120	100	0.325	3.077	1.083	1.200	3250	3800	13 000	15 200	0.855	0.855
Health posts	140	50	50	45	0.357	2.801	1.000	1.111	1600	1350	4978	4200	1.185	1.185
Total	900	310	284	245					10 150	11 750	37 110	43 225	0.864	0.859
									Crude verification factor = 10150/11750 =	0.864				

Stratum	Total Number of vaccinations reported counts in HMIS in "Month X" (A)	Weighted number of vaccinations in "Month X" (adjusted for non-coverage and non-response)		Adjusted verification factor (VFadj = D=B/C)	Analytical weight (E=A/C)	Weight factor by HMIS counts (F=D x E)
		Recounted (B)	Reported (C)			
General hospitals	16 170	10 214	12 719	0.803	1.270	1.020
Reference health centres	13 860	8918	11 106	0.803	1.247	1.001
Health centres	11 550	13 000	15 200	0.855	0.760	0.650
Health posts	4620	4978	4200	1.185	1.100	1.304
Total	46 200	37 132	43 225	0.859	4.377	3.975
		Adjusted verification factor	0.859	Adjusted verification factor weighted by HMIS reported counts [F/E]	0.908	

- Each indicator will have a separate calculation and grid
- The inputs come from the master facility list and the sample DV/SA data file
- Once the sample weights are calculated enter them into the appropriate place in the indicator batch file.
- Weight1 = ANC, WEIGHT2 = Immunization, WEIGHT3 = HIV/AIDS, etc.
- Use the Excel Tool designed to facilitate the calculation of survey weights.

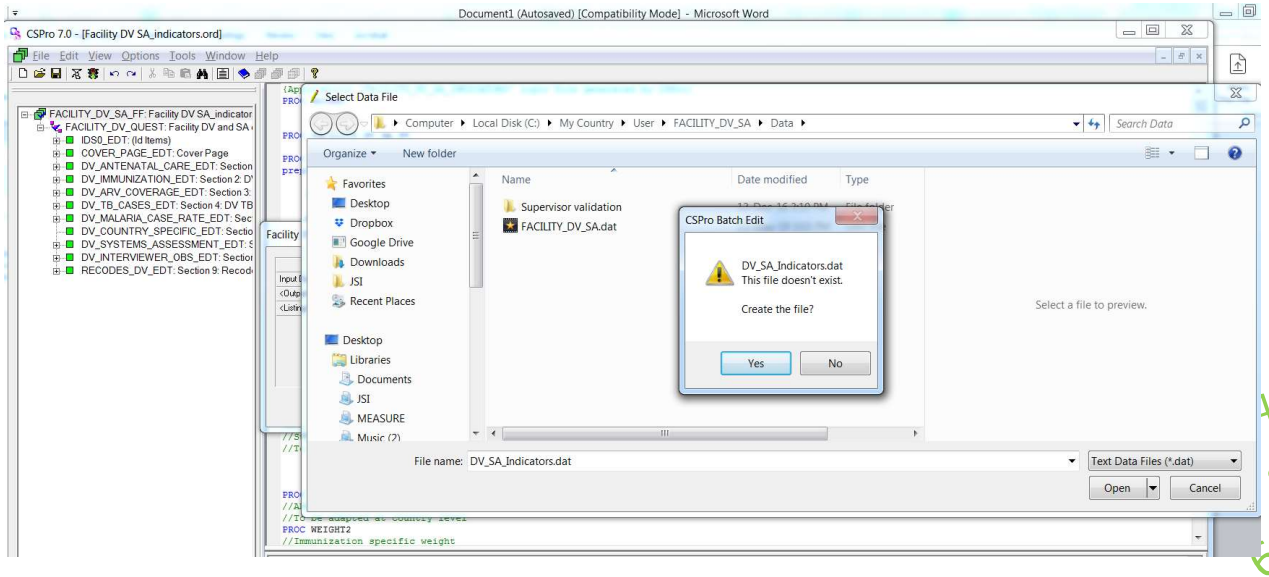
Calculate DV/SA Indicators

SESSION 14

Data Analysis

Run the indicator batch file

- Once the STRATUM variables are updated with local adaptations and the sample weights are added, run the indicator batch file by clicking on the stoplight, or select "run" from the file pull-down menu.
- Other local adaptations, such as survey type and essential medicines, will also need edited in the batch file.)
- Click "Yes" when asked to create the file.



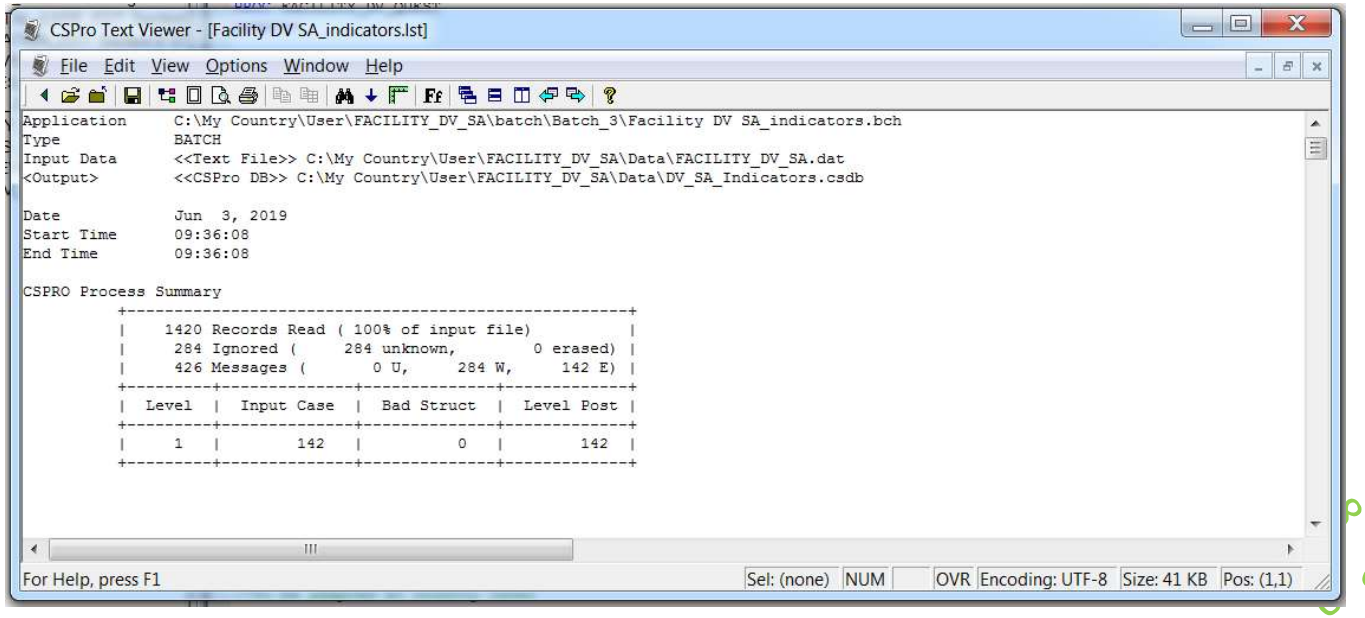
Calculate DV/SA Indicators

SESSION 14

Data Analysis

Run the indicator batch file

- A file called Facility DV/SA Indicators.csdb was created and saved to the FACILITY_DV_SA\Data folder.
- Check the Facility DV SA_Indicators.lst to ensure everything went as it was supposed to.



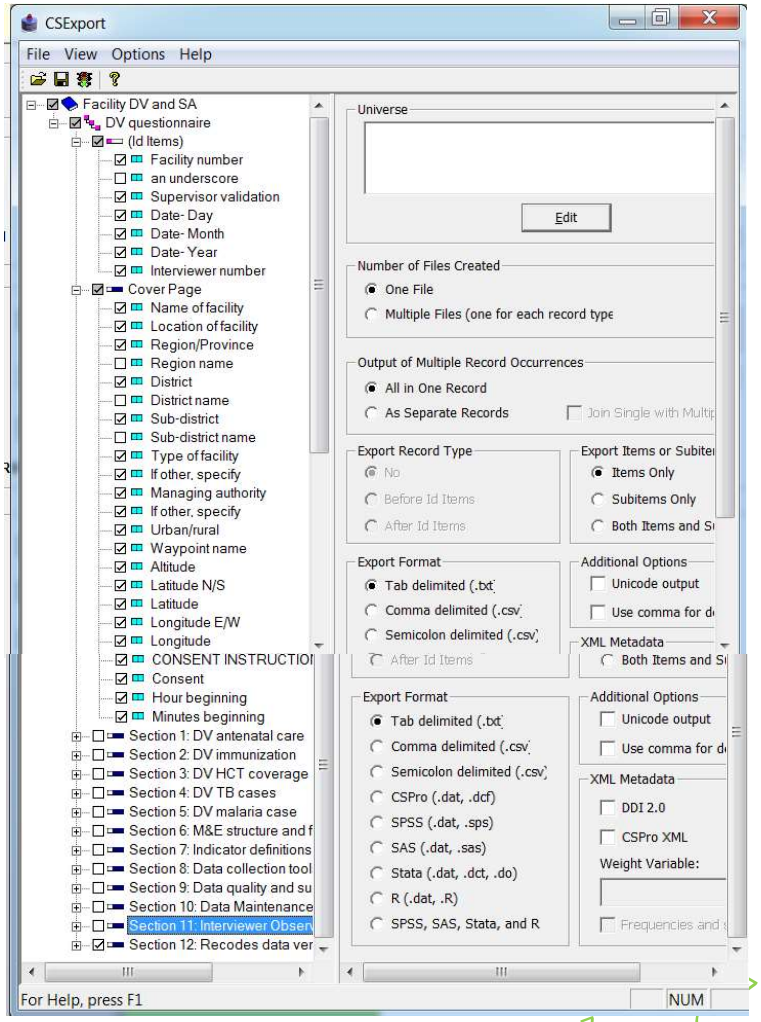
Export DV/SA Indicators

SESSION 14

Data Analysis

Export the indicator file to .txt

- Open CSPro, go to the Tools menu and select Export.
- Specify the FACILITY_DV_SA data dictionary when prompted.
- Click on ID Items to select all data elements. De-select "an underscore"
- Click on Cover Page to select all items.
- Click on Section 12: Recodes Data Verification to select all data elements
- Leave all other data elements unchecked.
- Use the default settings on the right and ensure the file is exported as a text file (tab delimited). Select "run".



Export DV/SA Indicators

SESSION 14

Data Analysis

Open the exported data file

- Open a blank workbook in Microsoft Excel
- Go to File, Open, and browse to the FACILITY_DV_SA\Data folder. Make sure ALL FILES is selected in the browse window. Select the Facility DV/SA Indicators.txt file that has just been exported from CSPro.
- In the text import wizard:
    1. Make sure delimited is selected
    2. Under delimiter, select Tab
    3. Under column data format, select general
- The txt file is now open in Microsoft Excel. Click on File -> Save as and change the save as type to Excel Workbook. The file is now saved as an XLS file.
- If your dataset contains a decimal comma instead of a decimal point, you will need to replace the commas with points in order for the chartbook to work properly. To do this, select all the data using Ctrl+A, then Ctl+F to open the Find/Replace window. In the "Find" box type , and in the "Replace" box type . then click on replace all.

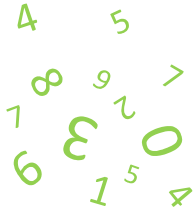
Export DV/SA Indicators

SESSION 14

Data Analysis

Check The Stratum And Weights Columns Are Complete

- Select the header row and then in the menu ribbon click on Data -> Filter
- On column with the header STRATUM_1, right click on the small icon in the right side of the cell. Scroll down and there will be a list of response values. Check that this list does not include "Blanks"
- Repeat for columns STRATUM_2, STRATUM_3, STRATUM_4, and WEIGHTS
- If there are blanks, please return to the data processing steps in CSPro to ensure all questions are complete. If there are no blanks, please move on to the next step.



Populate the DV/SA Chartbook

SESSION 14

Data Analysis

Check that column headers match between the chartbook and the exported data

- Select the header row of the exported indicator data and select copy
- Open a new, blank workbook and the in cell A1, right click and select paste/transpose
- Open the FACILITY DV/SA chartbook. Go to the blue tab called **Indicators**. Select the header row and select copy.
- Go back to the blank workbook, click on the cell B1, right click, and select paste/transpose
- In cell C1 type the following: =exact(A1,B1)
- Hover the mouse over the lower right-hand corner of cell C1 until a plus sign appears, then double click so that the formula is dragged to the bottom of the sheet.

Populate the DV/SA Chartbook

SESSION 14

Data Analysis

Check that column headers match between the chartbook and the exported data

- Check that all cells in column C say true (or that the column headers match). If FALSE appears anywhere in column C, please edit the exported data file to add or subtract columns to match the chartbook columns.
- Some versions of the chartbook have three data elements each for latitude and longitude of the health facility. The FACILITY_DV_SA data dictionary only has two each. You may need to add additional column for both latitude and longitude to the Facility DV/SA Indicators.xls in order to reconcile the difference in columns between the two files.
- Once the column headers match, go to the FACILITY DV/SA chartbook and click on the red tab called INSTRUCTIONS.
- Scroll down to row 41 and replace the text in <> quotes so that the name of the country and year of the assessment automatically shows on the output printouts.

Populate the DV/SA Chartbook

SESSION 14

Data Analysis

Copy/paste exported data into chartbook

- Go to the FACILITY DV/SA chartbook, and click on the blue tab called **Indicators**. Highlight all data except for the header row and delete the existing data.
- Go to the exported data file and select all data except for the header row, then select copy.
- Go back to the FACILITY DV/SA chartbook **Indicators** tab, click on cell A2, and then click paste. This should paste the data from the exported data file to the chartbook.
- The next tab, **Weighted_Data** will automatically update based on the information entered on the **Indicators** tab. The weighted_data tab is where the indicators are multiplied by the weights. The analytical outputs are based on the weighted data in this worksheet.

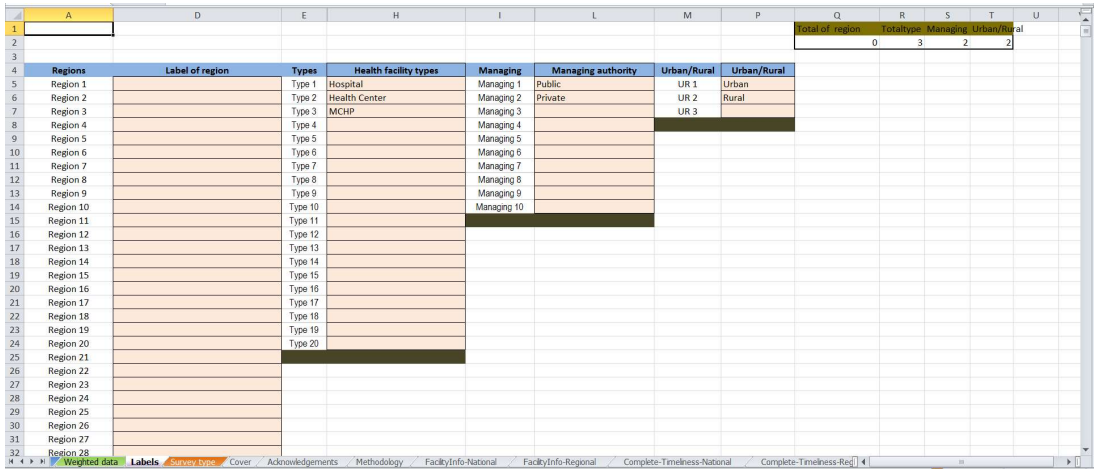
Populate the DV/SA Chartbook

SESSION 14

Data Analysis

Enter information on labels tab

- Go to the FACILITY DV/SA chartbook, and click on the purple tab called **Labels**.
- In column D, Label of region, enter the labels for each region in your survey. These should correspond to the order created for STRATUM_1 in the data processing step. Please note, the column is labelled "Region", but this can correspond to any administrative unit such as zone, district, etc.
- In column H, Health facility types, please enter the labels for each facility type in your survey. These should correspond to the order created for STRATUM_2 in the data processing step.



Regions	Label of region	Types	Health facility types	Managing	Managing authority	Urban/Rural	Urban/Rural
Region 1		Type 1	Hospital	Managing 1	Public	UR 1	Urban
Region 2		Type 2	Health Center	Managing 2	Private	UR 2	Rural
Region 3		Type 3	MCHP	Managing 3		UR 3	
Region 4		Type 4		Managing 4			
Region 5		Type 5		Managing 5			
Region 6		Type 6		Managing 6			
Region 7		Type 7		Managing 7			
Region 8		Type 8		Managing 8			
Region 9		Type 9		Managing 9			
Region 10		Type 10		Managing 10			
Region 11		Type 11					
Region 12		Type 12					
Region 13		Type 13					
Region 14		Type 14					
Region 15		Type 15					
Region 16		Type 16					
Region 17		Type 17					
Region 18		Type 18					
Region 19		Type 19					
Region 20		Type 20					
Region 21							
Region 22							
Region 23							
Region 24							
Region 25							
Region 26							
Region 27							
Region 28							

Total of region: 0 | Totaltype: 3 | Managing: 2 | Urban/Rural: 2

Weighted data | Labels | Survey type | Cover | Acknowledgements | Methodology | FacilityInfo-National | FacilityInfo-Regional | Complete-Timeliness-National | Complete-Timeliness-Reg

Populate the DV/SA Chartbook

SESSION 14

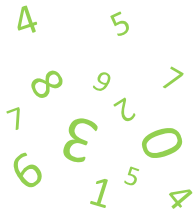
Data Analysis

Enter information on labels tab

- In column L, Managing authority, please enter the labels for each managing authority in your survey. These should correspond to the order created for STRATUM_3 in the data processing step.
- In column P, Urban/Rural, please enter the labels for each urban/rural category in your survey. These should correspond to the order created for STRATUM_4 in the data processing step.

Enter information on survey type

- Go to the FACILITY DV/SA chartbook, and click on the orange tab called **Survey type**.
- Fist click on the button 1- Check weighted data. This automatically extends or retracts the formulas in the **Weighted_Data** tab to match the number of rows of data the in the **Indicators** tab.



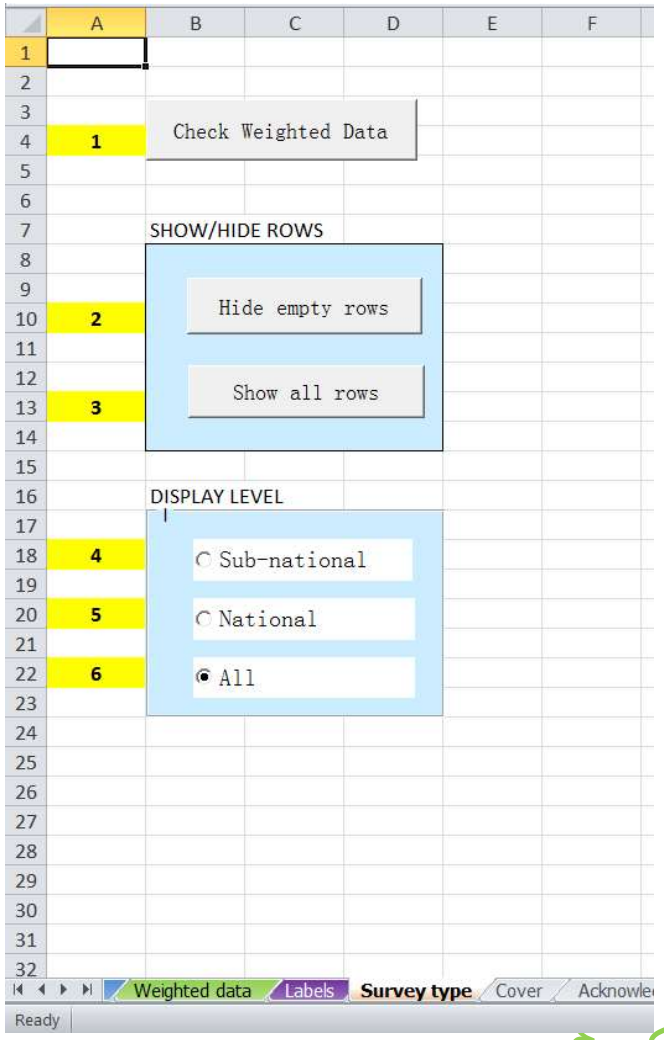
Populate the DV/SA Chartbook

SESSION 14

Data Analysis

Enter information on survey type

- Then click on the button select "Hide empty rows". This will hide the extra rows that will not be used.
- It may take up to 5 minutes for Excel to run this command. Please wait for Excel to finish the task.
- If you do not change the labels tab again, then you should not need to run this step even if the data is updated.
- However, if you change the labels you will need to click "Show all rows" then "Hide empty rows" to update the chartbook.



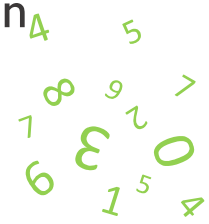
Populate the DV/SA Chartbook

SESSION 14

Data Analysis

Enter information on survey type

- Next, specify if the survey is to have national level results or sub-national level results. In the display level, please select either Sub-national or National based on the sampling methodology chosen for the survey.
- The national option activates sheets that have national estimates as well as stratification by facility type, managing authority, and urban/rural.
- The sub-national option activates sheets that include the same information as the national option, but also includes stratification by region.
- Please note, if you are interested in sub-national information by region, you will need a chartbook for each region.
- The chartbook should now be ready for review. Please click on each tab in the chartbook to review all tables and charts.



Review the Data in the Chartbook

SESSION 14

Data Analysis

General facility information

Availability of services and reporting data status

Percentage of facilities providing each health service, by facility type, managing authority, and urban/rural (N=507)

	S1	S2	S3	S4	S5	
	ANC	DTP3/PENTA3	Currently on ART	Notified cases of TB	Malaria cases	Total number of facilities
Facility type						
Hospital	71%	51%	20%	12%	76%	59
Health Center	87%	67%	0%	7%	80%	15
MCHP	89%	83%	2%	1%	72%	381
Managing authority						
Public	79%	71%	4%	11%	67%	302
Private	77%	71%	3%	14%	64%	205
Urban/Rural						
Urban	78%	70%	5%	15%	67%	418
Rural	78%	75%	1%	1%	62%	89
Total	**78%**	**71%**	**4%**	**12%**	**66%**	**507**

- The Facility_Info tab has information about the availability of service provision by indicator.
- This is further disaggregated by Facility Type, Management Authority, and Location.
- Also included is the proportion of facilities reporting to the HMIS, and the availability of source documents and reports.

Review the Data in the Chartbook

SESSION 14

Data Analysis

Data quality: timeliness and completeness

Data element completeness

Amongst facilities providing a specific service and reporting data, percentage of facility-months that have data for the following indicators in their monthly reports, by facility type, managing authority, and urban/rural

	S1_05	S2_05	S3_05	S4_05	S5_05
	ANC (N=373)	DTP3/PENTA (N=347)	Currently on ART (N=14)	Notified cases of TB (N=58)	Malaria cases (N=307)
Facility type					
Hospital	65%	62%	100%	100%	70%
Health Center	61%	53%	-	100%	70%
MCHP	87%	88%	50%	100%	77%
Managing authority					
Public	86%	87%	100%	84%	78%
Private	81%	82%	50%	92%	73%
Urban/Rural					
Urban	82%	85%	86%	88%	75%
Rural	92%	86%	-	-	80%
Total	**84%**	**85%**	**86%**	**88%**	**76%**

- The Complete-Timeliness tab has information about the completeness of the DV data (% of facility-months with available data), by indicator.
- Also included is Completeness of information on TB minimum set of variables, and the by Facility type, Management authority, and location.

Review the Data in the Chartbook

SESSION 14

Data Analysis

Data quality: data verification factor

Data verification factor

Facility level data verification factor, by facility type, managing authority, and urban/rural

	S1_06	S2_06	S3_06	S4_06	S5_06
	ANC (N=282)	DTP3/PENTA (N=277)	Currently on ART (N=10)	Notified cases of TB (N=45)	Malaria cases(N=214)
Facility type					
Hospital	1.02	0.97	1.00	0.53	1.02
Health Center	1.01	0.96	-	0.00	1.00
MCHP	1.06	0.99	1.00	0.52	1.00
Managing authority					
Public	1.04	1.00	1.00	0.76	1.01
Private	1.09	0.98	1.00	1.98	0.99
Urban/Rural					
Urban	1.06	0.99	1.00	1.33	1.00
Rural	1.03	1.01	-	-	1.00
Total	**1.06**	**0.99**	**1.00**	**1.33**	**1.00**

- The Facility DV-Factor tab has the verification factor for each indicator, disaggregated by Facility type, Management authority, and location.
- The number of facilities that contribute to the calculation is indicated for each program area in the header row.

Review the Data in the Chartbook

SESSION 14

Data Analysis

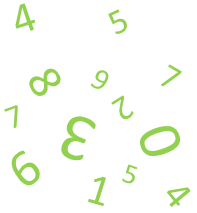
Data quality: ANC data verification factor

Facility level data verification factor

ANC Facility level data verification factor indicators, by facility type, managing authority, and urban/rural (N=282)

	S1_06	S1_07	S1_08	S1_09	
	National verification factor	% of facilities for which source data exactly match reported data	% of facilities that over-report by more than 10% (V.F. < 0.9)	% of facilities that under-report by more than 10% (V.F. > 1.10)	Total number of facilities providing the service and reporting data that have all required source records and reports
Facility type					
Hospital	1.02	86%	0%	5%	21
Health Center	1.01	40%	0%	0%	5
MCHP	1.06	64%	4%	4%	256
Managing authority					
Public	1.04	59%	5%	5%	172
Private	1.09	74%	3%	2%	110
Urban/Rural					
Urban	1.06	64%	4%	3%	228
Rural	1.03	67%	4%	7%	54
Total	**1.06**	**65%**	**4%**	**4%**	**282**

- Program-specific data verification tabs are available for each indicator, with:
  - the verification factor,
  - % of facilities with exact match between source documents and reports,
  - % of facilities over-reporting by more than 10%, and
  - % of facilities under-reporting by more than 10%.



Review the Data in the Chartbook

SESSION 14

Data Analysis

ANC reasons for discrepancy between source data and reported data, by facility type, managing authority, and urban/rural (N=373)

	S1_10A	S1_10B	S1_10C	S1_10D	S1_10E	
	No discrepancy	Arithmetic errors	Transcription errors	Missing documents during report preparation	Missing documents during survey implementation	Total number of facilities providing the service and reporting data
Facility type						
Hospital	74%	8%	3%	5%	18%	38
Health Center	67%	25%	17%	0%	33%	12
MCHP	80%	16%	19%	5%	8%	322
Managing authority						
Public	77%	17%	20%	5%	9%	225
Private	82%	12%	14%	5%	12%	148
Urban/Rural						
Urban	80%	17%	16%	5%	11%	309
Rural	75%	8%	27%	5%	6%	64
Total	**79%**	**15%**	**17%**	**5%**	**10%**	**373**

Also included in the indicator-specific DV tabs are;

- The reasons for discrepancy between the source documents and reports, and…
- The reasons for missing monthly reports

ANC reasons for missing monthly reports, by facility type, managing authority, and urban/rural (N=373)

	S1_11A	S1_11B	S1_11C	S1_11D	S1_11E	S1_11F	
	All reports available	Submitted report cannot be located now	Do not have trained staff to report	No reporting form was available	Interruption in servie delivery in one or more of the selected months	Other reasons	Total number of facilities providing the service and reporting data
Facility type							
Hospital	47%	47%	47%	47%	47%	47%	38
Health Center	17%	17%	17%	17%	17%	17%	12
MCHP	51%	51%	51%	51%	51%	51%	322
Managing authority							
Public	45%	45%	45%	45%	45%	45%	225
Private	55%	55%	55%	55%	55%	55%	148
Urban/Rural							
Urban	48%	48%	48%	48%	48%	48%	309
Rural	56%	56%	56%	56%	56%	56%	64
Total	**49%**	**49%**	**49%**	**49%**	**49%**	**49%**	**373**

Review the Data in the Chartbook

SESSION 14

Data Analysis

Systems assessment

Percentage of facilities that report health data to a MOH reporting system with tracer items for data management, by facility type, managing authority, and urban/rural (N=396)

	SA_01	SA_02	SA_03	SA_04	SA_05	SA_06	SA_07			
	Trained staff	Guidelines	No stock outs of taly sheets, registers, and reporting forms	Received supervision and written feedback including on data quality	Analyze and use data	Mean availability of items	All items	Overall score	Total number of facilities that report health data to a MOH reporting system	Total number of facilities
Facility type										
Hospital	43%	43%	43%	43%	43%	43%	43%	31%	42	59
Health Center	15%	15%	15%	15%	15%	15%	15%	13%	13	15
MCHP	48%	48%	48%	48%	48%	48%	48%	43%	340	381
Managing authority										
Public	43%	43%	43%	43%	43%	43%	43%	34%	239	302
Private	52%	52%	52%	52%	52%	52%	52%	40%	157	205
Urban/Rural										
Urban	45%	45%	45%	45%	45%	45%	45%	35%	327	418
Rural	52%	52%	52%	52%	52%	52%	52%	40%	69	89
Total	**46%**	**46%**	**46%**	**46%**	**46%**	**46%**	**46%**	**36%**	**396**	**507**
	0%	0%	0%	0%	0%	0%	0%	0%		
	100%	100%	100%	100%	100%	100%	100%	100%		

- The System Assessment is color-coded for ease of interpretation.
- A red-to-green color gradient is used to color-code cells based on their value. Green for the best scores, red for the worst. The two bottom rows represent the legend.
- System assessment indicators are also presented disaggregated by Facility Type, Management Authority, and Location (Urban/Rural).

District DV/SA

SESSION 14

Data Analysis

District level data analysis

- All the data analysis procedures are also available for the District DV SA.
- There is a District DV/SA indicator batch file to create indicators for analysis of district level results. The CSPro DV/SA application is found in the DISTRICT_DV_SA folder and the District DV/SA Indicator batch file is found in DISTRICT_DV_SA\Batch\Batch_3.
- The District DV/SA Chartbook works exactly like the Facility DV/SA Chartbook.
- However, since Districts are not sampled (all districts will be visited) they should be unweighted (i.e. each district has a weight of "1").

Slide 27

AS4 New district chartbook has some nuances and extra settings -- might be important to talk about this
Ashley Sheffel, 12/9/2019

D.B.1 Maybe we should discuss - not sure which ones you mean.
David Boone, 4/24/2020

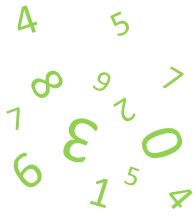
Questions

District level data analysis

- Describe the purpose of the indicator batch file.
- What is the purpose of weighting of survey estimates from DV/SA? How are the weights derived? How are the weights applied in CSPro?

SESSION 14

Data Analysis



Practice

SESSION 14

Data Analysis

Practice

- Practice creating the indicator batch file from the raw Facility DV/SA data.
- Edit the Facility DV/SA Indicators batch file and run the program.
- Export the CSPro Indicators table to text file and import to Excel.
- Copy and paste the data into a blank copy of the Facility DV/SA Chartbook and configure the Chartbook for use, adding the labels.
- Review the data and discus your findings in small groups.
- You have 85 minutes.