



Welcome to today's presentation: Data Quality: Looking Back and Moving Forward

My name is Michael Costa and I'm from the SPHERE/Abt team. I'm here with our presenters, Ellie Coombs, also from SPHERE/Abt, and Maria Jackson-Hittle, from WRMA/CSR.

Grantees and providers did an excellent job in the RSR's first year! 89% of required providers submitted client-level data files. During this presentation, we want to help you use your 2009 experience and some Technical Assistance tools to look at the quality of your data. This can help you improve your report for Calendar Year 2010. It will also help you improve your data collection for Calendar Year 2011.

The ultimate goal is to use these lessons to improve data quality as we move into the 2011 data collection period. **We want to have the highest quality data possible for 2011 because these data will stand alone. HAB will not collect RDR data reports for the services you provide in 2011.** We know you want Congress to see the true quality of your work and this depends on presenting accurate data.

Data quality improvement is like other quality improvement: not about pointing fingers – we want to help you analyze your data, spot and deal with challenges. We want the data to tell your story how it should be told.

Remember! 2011 RSR data quality is even more important because we are phasing out the RDR that year – so we are putting all our eggs in the RSR basket!

Now, I'll hand the presentation over to Ellie.

Outline of Today's Presentation

- The Importance of Data Quality
- Looking Back at 2009 Data Quality Issues
- Moving Forward
 - Data Validation
 - Completeness Reports

Ryan White Services
Report

We will cover three things today:
Why is data quality so important?

What does 2009 data quality show?

What can you do?

Quality of Data Versus Quality of Care

- Two different things
 - Quality of data = Accuracy and timeliness of your data
 - Quality of care = What your program is doing compared with what it should be doing
- We are looking at:
 - Quality of data
 - Not quality of care

Ryan White Services
Report

Quality of Data looks to see if data accurately reflect the services you deliver.

Quality of Care compares the services that were actually delivered to a external measure of quality.

Let's look at some examples: The HAB performance measure on CD4 tests requires at least 2 tests spread out during the year.

Assume you only capture data on CD4 tests that are done by your agency, and half of the CD4 tests are sent to an outside lab. If you have 100 patients and half of the patients' tests are not in your report, it will look like you did not do ANY CD4 tests for half of your patients. You quickly dismiss this issue as a data quality issue. However, some of your clients may not be getting their CD4 tests and you can't detect how many.

If you capture data on all of the CD4 tests done in your agency or referred by your agency, your very best possible score could be 100%. If it falls below that, you have an opportunity for improvement in the quality of your care.

If you don't know if you have quality data, you don't know where to focus your improvement.

Good quality data can help you improve quality of care, poor quality data cannot.



To illustrate the importance of data quality, here is a representation of your program – a masterpiece!



Here is how the RDR would represent your program. Although you can see the general structure of the program, you are missing the details. The RDR provides duplicated data. It does not provide so much clinical information. It cannot tell you clearly about the care the bulk of your patients receive.



Here's your program through the lens of the RSR. It only covers the funded patients, so it may lack some color. However, because it shows you all the clinical data on a funded patient, the picture is more complete.



And finally, here's your program through the lens of the RSR with poor data quality. Your program appears distorted; these data do not reflect the true quality of the services you provide.

Outline of Today's Presentation

- The Importance of Data Quality
- **Looking Back at 2009 Data Quality Issues**
- Moving Forward
 - Data Validation
 - Completeness Reports

Ryan White Services
Report

Now you know why the quality of your data is so important.

In this section, we will look at some data quality indicators that stood out in 2009. These are not fine grained data review techniques.

The view of the data is from 50,000 feet. We can pick out some large items with some surety. The accuracy of smaller items is not as good. We depend on you to look at the items we identify and respond appropriately.

2009 Data Quality Issues

- Completeness
- Consistency between RSR/RDR client counts
- Clinical data completeness

We are fortunate to have three teams to help with data quality:
SAIC
SPHERE/Abt
WRMA/CSR.

So, what were the 2009 data quality issue?

Many of you recall our initial review of your completeness reports.
Second, we saw if the RSR data is consistent with the data you submitted in the RDR.
Finally, we took a second, closer, look at the data completeness for your clinical data.

Everyone Got a Completeness Report

- A Completeness Report is available on the RSR System within the EHB
- Each provider who submitted a client-level data XML file has a report
- Each grantee can obtain reports with data from one or multiple providers

You can download your Completeness Report from the RSR-System on the EHB. Every provider that submitted client-level data will get a Completeness Report. Grantees will also get a report for one or multiple providers.

We discussed these at the All Grantee Meeting and had a series of webcasts about them. We also did some email follow up with grantees when the completeness report was unusual.

The Main Indicator in the Report is the Completeness Rate

- The percentage of required clients that were reported, including Unknown
- Example
 - Total number of eUCIs = 100
 - Required for #64 (Pregnant) = 22 (female)
 - Reported = 15
 - $15/22 * 100 = 68\%$

Let's look at what the Completeness Report tells us –

It tells us how often your client data file contained an appropriate response to a question.

The main indicator in the Completeness Report is the Completeness Rate – which is the percent of required clients for which you reported, including Unknown.

As an example, say you reported 100 unduplicated clients. 22 of these clients are female. Therefore, the number of required clients for data element #64 (whether the client is pregnant or not) is 22. However, you only reported the pregnancy status for 15 women. Therefore, your completeness rate is 68%.

Completeness

- Complete
- Almost perfect
- Not so complete

Regarding completeness, all providers submitted information on all required clients for seven data elements – predominantly demographic data elements, such as enrollment status, ethnicity and poverty level.

Almost all providers submitted information on almost all required clients for several demographic data elements – birth year, race and geo unit code – in addition to clinical data elements related to women and screened for Hep B and Tuberculosis since HIV Diagnosis.

Some data elements were more of a challenge – these were usually elements you had not reported individually before: these include many of the screening variables, viral load and CD4 count, and client services – including ambulatory services.

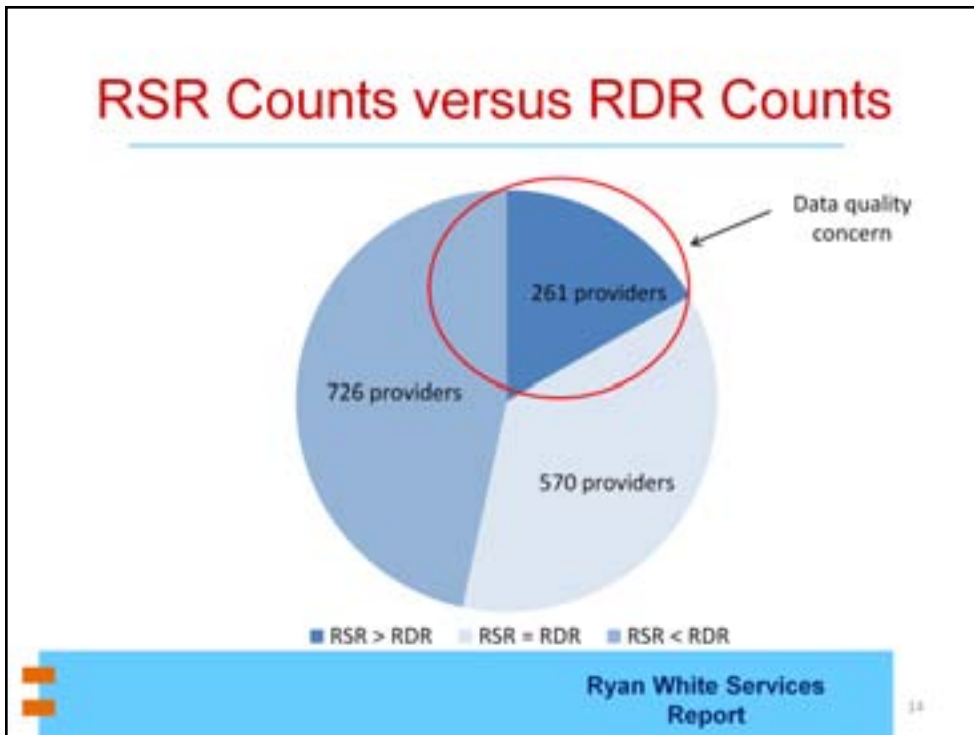
Now, I'd like to hand the presentation off to Maria.

RSR and RDR Relationships

- Differences in reporting scope:
 - RDR Preferred Scope: Clients who received a service eligible for Ryan White funding
 - RSR Scope: Clients who received a service that was funded, at least partially, through the Ryan White Program
- Therefore, at the provider level, expect RSR unduplicated client counts less than or equal to RDR unduplicated client counts

Our second data quality indicator involves the relationship between RSR counts and RDR counts. Because of reporting scope, we consider it MOST LIKELY that RSR unduplicated client counts will be less than or equal to RDR unduplicated client counts.

This is because the reports have different client eligibility requirements. Within the RDR, clients who received a service eligible for Ryan White funding are included. Within the RSR, clients who received a service that was funded, at least partially, through the Ryan White Program are included.



Of the 1557 providers who submitted an RDR and an RSR, about half had RSR counts lower than RDR counts – as expected. Almost 600 providers had counts exactly equal. This is quite possible if all the provider’s clients received at least one funded service. Or if the provider only reports funded scope on the RDR.

However, 261 providers have RSR counts greater than RDR counts – which strongly suggests data quality concerns.

What Might Be Causing This Problem?

- Software challenges
 - eUCI
 - File merge
- Scope of service
 - Clients who did not receive a funded service
 - Clients who received a service ineligible for client reporting (e.g. HIV testing)

We followed up with a few providers to investigate from where the problem is coming.

The first problem identified involves the eUCI. If a provider submits multiple files, each created from a different system, and those systems do not generate the eUCI the same, then records will not be merged appropriately. This could stem from different eUCI generation algorithms or inconsistent reporting of the UCI's underlying data elements.

We also found issues with the RDR that made its numbers too low. In one case, the provider was not merging files correctly from all sub-providers, thus excluding clients eligible for reporting.

Also, we found instances where clients that should not have been included in the RSR were. For example, clients who did not receive a funded services, clients who received a service ineligible for client level reporting, such as HIV testing, or clients who were seen by the umbrella agency and not the provider itself.

We have been following up with the grantees for the 80 providers with differences greater than 10%. We sent emails to the grantee of record noted in the RSR. We have had some great success in identifying important data quality issues.

Provider Name	Element	Unique	Required	Known		Missing	
				#	%	#	%
EUCALYPTUS CENTER	AmbulatoryService	492	401	281	70%	70	30%
EUCALYPTUS CENTER	CD4Test	492	401	328	82%	73	18%
EUCALYPTUS CENTER	FirstAmbulatoryCareD	492	401	397	99%	0	0%
EUCALYPTUS CENT	Pregnant	492	120	97	81%	0	0%
EUCALYPTUS CENT	PrenatalCare	492	8	8	100%	0	0%
EUCALYPTUS CENT	PrescribedArvMedicab	492	8	6	75%	2	0%
EUCALYPTUS CENT	PrescribedHaart	492	401	350	87%	51	0%
EUCALYPTUS CENTER	PrescribedPcpProphyla	492	401	365	91%	36	0%
EUCALYPTUS CENTER	PrescribedCervicalPapSe	492	120	120	100%	0	0%
EUCALYPTUS CENTER	ScreeningProvided	492	401	363	91%	38	0%
EUCALYPTUS CENTER	ScreenedTB	492	401	326	81%	75	0%
EUCALYPTUS CENTER	ScreenedHepatitisB	492	401	326	81%	75	0%
EUCALYPTUS CENTER	VaccinatedHepatitisC	492	401	64	16%	337	0%
EUCALYPTUS CENTER	ViralLoadTest	492	401	0	0%	73	18%
HAIGHT MEMORIAL HOSPI	AmbulatoryService	167	160	160	100%	0	0%
HAIGHT MEMORIAL HOSPI	CD4Test	167	160	46	29%	114	71%

Now I'm going to turn to an enhanced version of the completeness report – which is depicted on the slide. There are two main changes:

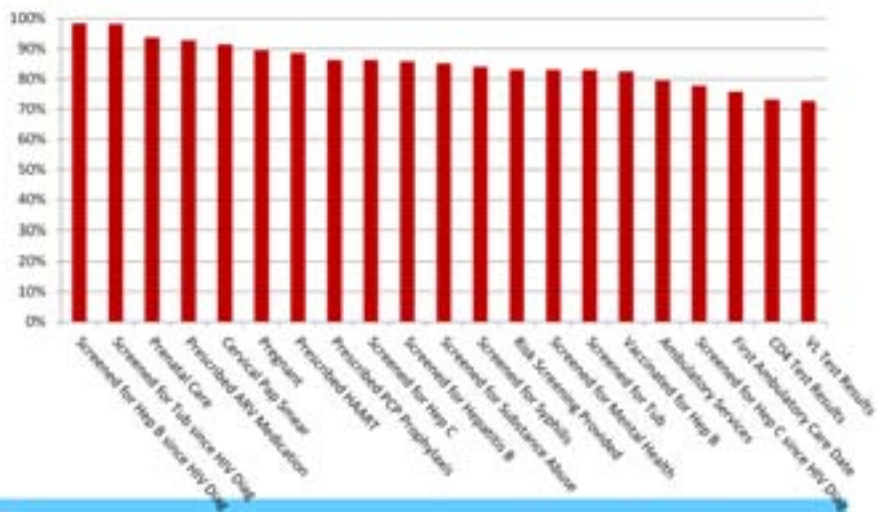
First, we decided to look more closely at the clinical questions, Numbers 46 to 66. These are the questions that ONLY Outpatient Ambulatory Medical Care providers must answer.

In addition, we break out Unknowns from Reported. The modified Completeness Report contains the number of unique clients, number of required clients, and the number and percent of Known, Unknown and Missing values for those clinical data elements.

We provide the percent of Unknowns, so you can be aware of them and gradually reduce them over time, if possible. We recognize that some level of Unknown is normal. For example, some Hispanics may not identify with any race and therefore, not report it. The provider must often resort to Unknown.

What we are looking for is data that stands out because it is different.

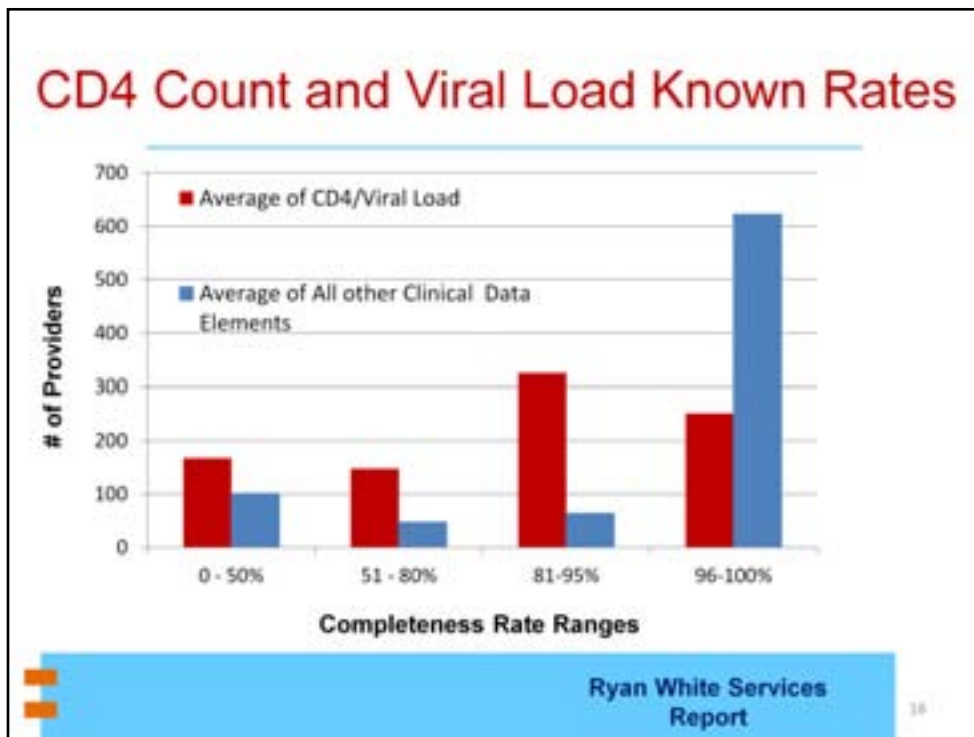
Overall, Known Rates Are High



Ryan White Services
Report

17

Even excluding Unknowns from reported, a look at the overall provider performance shows some very fine results. For most data elements, the Known Rate average is above 80%.



We did find that two important data elements stood out: CD4 count and viral load. This graph presents the number of providers that fell into various known rate ranges (0-50%, 51-80%, 81-95% and 96-100%) for two groups of clinical data elements. We lumped CD4 count and viral load as one group and all other clinical data elements as another group. You will see that a large # of providers, almost 600, had between 96 and 100% for all other clinical data elements. That was not the case with CD4 count and viral load – where more providers fall within the 81-95% category.

When analyzing completeness rates for clinical measures, look for two trends. What if a group of data elements has the same percentage as missing? For example, 10 elements have missing rates of 7%. What does this indicate? Mostly likely, data for a group of clients is left out of the report. Why would this happen? Maybe these clients are seen at a satellite clinic, whose systems do not communicate with the main clinic.

Or, what if several data elements, say for example CD4 count and viral load, are 100% Missing. Does this mean that this clinical provider is not rendering any CD4 or Viral Load tests? Very doubtful! More likely, the provider is having problems marrying its lab data with its other clinical data. Maybe the provider conducts the labs in-house, but the data are stored in a different system; maybe they refer the labs out and don't track them. Whatever the data quality problem, I'm sure the provider *wants* to demonstrate that their clients *are* getting needed services.

If you are experiencing either of these problems, contact us, so we can work out a solution.

Email Outreach

- Contact grantees with providers that do not meet certain thresholds
- Grantees inform providers
- HAB, TA contractors, and grantees assist providers to resolve issues

We established two thresholds for data follow up:

For CD4 count and Viral Load, we will follow up with grantees that were 95% and below.

For the remaining clinical questions, we will follow up with grantees that were 50% and below completion.

We will send emails from the SPHERE/Abt team to the data contact in your 2009 RSR.

Back to you, Ellie.

Outline of Today's Presentation

- The Importance of Data Quality
- Looking Back at 2009 Data Quality Issues
- **Moving Forward**
 - Data Validation
 - Completeness Reports

Now, we'll discuss the tools and support available to help you evaluate your 2010 and 2011 data going forward.

Data Validation

- What is it?
- What will be used?
- When will we know?

Data is valid when it is consistent, both internally and externally. HAB is calling the process of checking for data consistency “Data Validation.”

Your data are internally NOT consistent when the answer to one question creates a condition that excludes another answer for the same client during the same reporting period. For example, if your answer to question 64 (Was the client pregnant?) was NO, then your answer to question 66 (was the client prescribed antiretroviral therapy to prevent maternal-to-child transmission?) CANNOT be YES.

You will find that data can be inconsistent in relation to items outside the reporting period. If the reporting period is Calendar Year 2011, then test dates such as 11/25/2010 or 2/10/2012 are NOT consistent.

HAB is currently finalizing its data validation checks and should have them available for you within a month or so. In addition, we have started incorporating alerts into 2010 reporting; we’ll expand on this during a webcast on the 28th.

HAB Plan of Action for Data Validation

- Provide validation for data collected in 2011 – to grantees, software vendors, other Ryan White stakeholders
- Encourage RSR software vendors to roll out validation for data collected in 2011

HAB realizes that the best place to stop data inconsistencies is at the point of data entry. The last thing we want to do is go back and edit data in order to get a successful submission! This means we plan to work with vendors so that these checks are rolled out while 2011 data is collected.

Completion Reports

- Useful
- Earlier
- More specific

In 2010, we plan to provide you more useful and more specific feedback on your completeness. And many providers and grantees have requested that the reports are provided earlier. That is our goal!

Calculate Group 1 Performance Measures with CLD

Measure	Denominator	Numerator
ARV therapy for pregnant women	Current gender (#7) Was the client pregnant (#64)	Was the client prescribed ARV (#66)
CD4 T-cell count	Outpatient/ambulatory care visit dates (#48)	Value and test date of CD4 cell counts (#49)
HAART	HIV/AIDS status (#12) Outpatient/ambulatory care visit dates (#48)	Was the client prescribed HAART (#52)
Medical visits	Outpatient/ambulatory care visit dates (#48) (at least one)	Outpatient/ambulatory care visit dates (#48) (at least two)
PCP prophylaxis	Value and test date of CD4 cell counts (#49)	PCP prophylaxis (#51)

Report

24

Group 1 performance measures can also help clinical providers evaluate their data. This slide contains which client-level data elements are necessary to calculate the measures. For example, the CD4 count measure calculates the % of clients with HIV/AIDS who had 2 or more CD4 T-cell count test in the measurement year. To calculate this measure, you would use as the denominator the total number of clients with an outpatient/ambulatory visit. Those are the clients for which you have information on CD4 count. As the numerator, you'd use the # of clients for which you actually had a CD 4 count value and test date for.

Answer the Following Questions

- Which percentages are lower and higher than expected?
- Why?
 - Is my data system not **capturing** the data necessary to properly reflect my program?
 - Is my data system not **reporting** the data necessary to properly reflect my program?
- How can I improve my system to better capture and report data?

Once you have calculated your performance measures, you can ask yourself if any of these numbers seems strange to you – too high or too low. If so, you might be seeing a data quality issue. Check to see if your data system is not capturing the necessary data. Or maybe the problem lies with reporting; you have the data in your system, but you are not extracting it correctly. Finally, you can work with us to fix those problems for better data quality in the future.

Technical Assistance Resources

- Contact SPHERE/Abt:
 - RSR.TA@sphereinstitute.org
- TARGET Center Website
 - <http://www.careacttarget.org/rsr.asp>
- Ryan White HIV/AIDS Program Data Support
 - 888.640.9356
 - Available 9 a.m. to 5:30 p.m. ET, Monday through Friday
 - ryanwhitedatasupport.wrma@csrincorporated.com
- HRSA Call Center
 - 877.Go4.HRSA (877.464.4772)
 - Available 9 a.m. to 5:30 p.m. ET, Monday through Friday
 - CallCenter@HRSA.gov