Precision & Bias
Statistics

The Peanut Butter (alas, no jelly) of QA

Presentation to AQSSD

by the Monitoring and Quality Assurance Group

August 28, 2003

- **As presented in the Ozone DQO talk, it is important to know the magnitude of**

  measurement bias (systematic deviation from truth), and

  measurement imprecision (random deviations)

  to understand size of decision errors.
- **We also have seen that bias is particularly important.**
- **So how are we going to estimate bias and precision?**

- **What follows is based on several months work involving**
  - **State/Local/Tribal organizations**
  - **EPA (Regional and RTP)**
  - **Contractor support**

A few more details on the rationale for the statistics. "One person's bias is another person's imprecision."

- Depending on temporal and spatial aggregation, one person may say there is bias but another may say there is imprecision.
  - Spatial Example: One site always 10% higher than truth, second site always 10% lower
    - Some say there is bias (either +10% or –10%) and the system is precise
    - Other say there is no bias but the system is imprecise.
  - Temporal Example 1: 10%, 10%, 10%, 10%, 10%, -10%, -10%, -10%, -10%, -10%
    - Some say there is no bias, but there is imprecision.
    - Some say there is great potential for bias, but system is precise.
    - Some say there is potential for bias and imprecision.
  - Temporal Example 2: 10%, -10%, 10%, -10%, 10%, -10%, 10%, -10%, 10%, -10%
    - Some say there is no bias, but there is imprecision.
    - Some say there is potential for bias and imprecision.

- We have the luxury of site specific information so we will not be aggregating over sites. As for aggregating over time, we are taking conservative approach and emphasizing bias since it is so important to the DQO process

How do we estimate P&B?

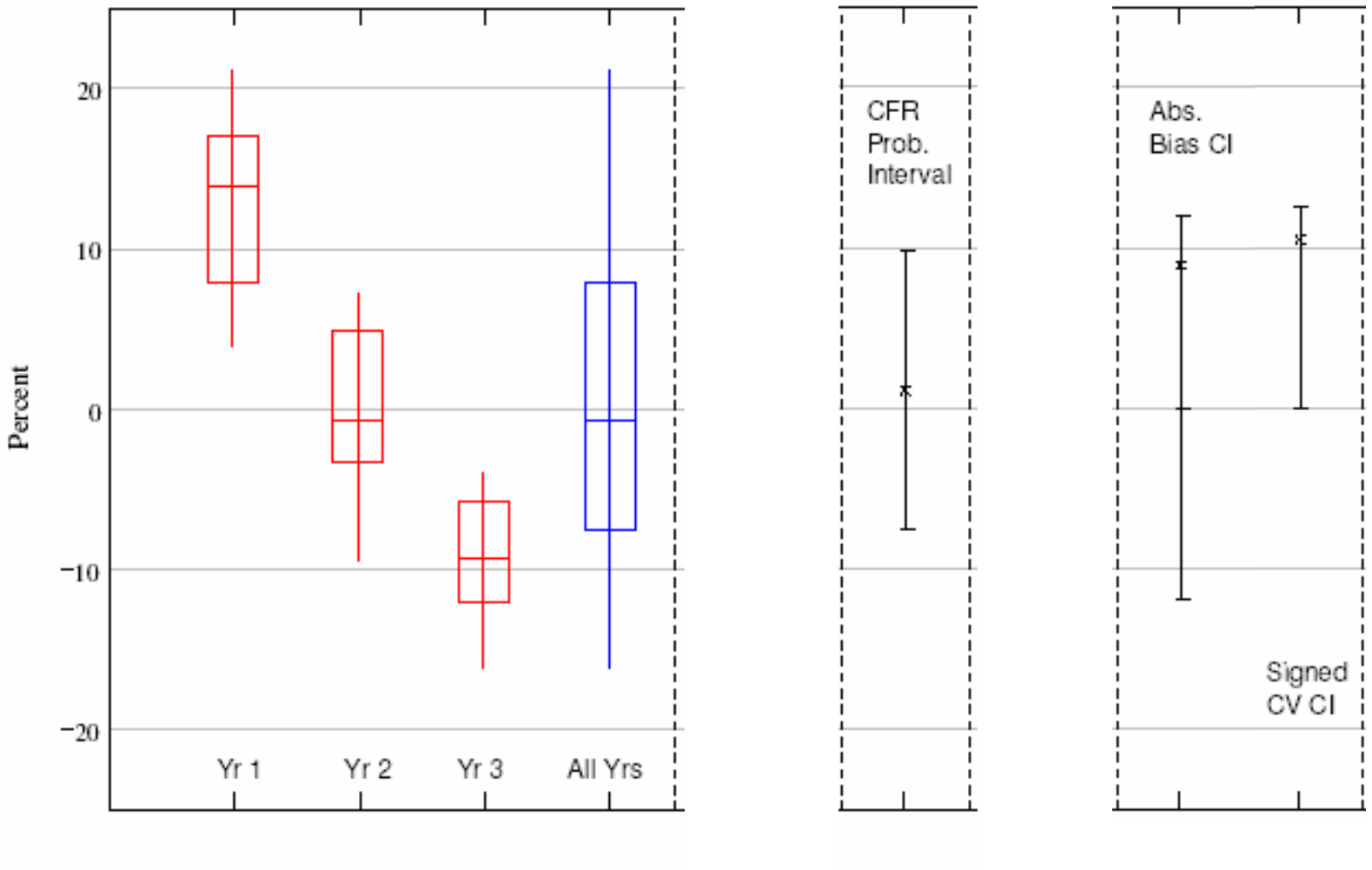
- Bias Estimate:
  - Calculate the percent relative differences: (observed-truth)/truth
  - Take their absolute values.
  - Calculate the average of these absolute values. This is the bias estimate.
  - If the analyzer is generally above the challenging concentration, we say the bias tends to be positive. If the analyzer is generally reading less than the challenging concentration, we say the bias tends to be negative. So we assign a direction to bias in this qualitative sense.

- Precision Estimate:
  - Start with the percent relative differences from above.
  - Do NOT take their absolute values.
  - Calculate the standard deviation of those percent differences. This is the precision estimate.

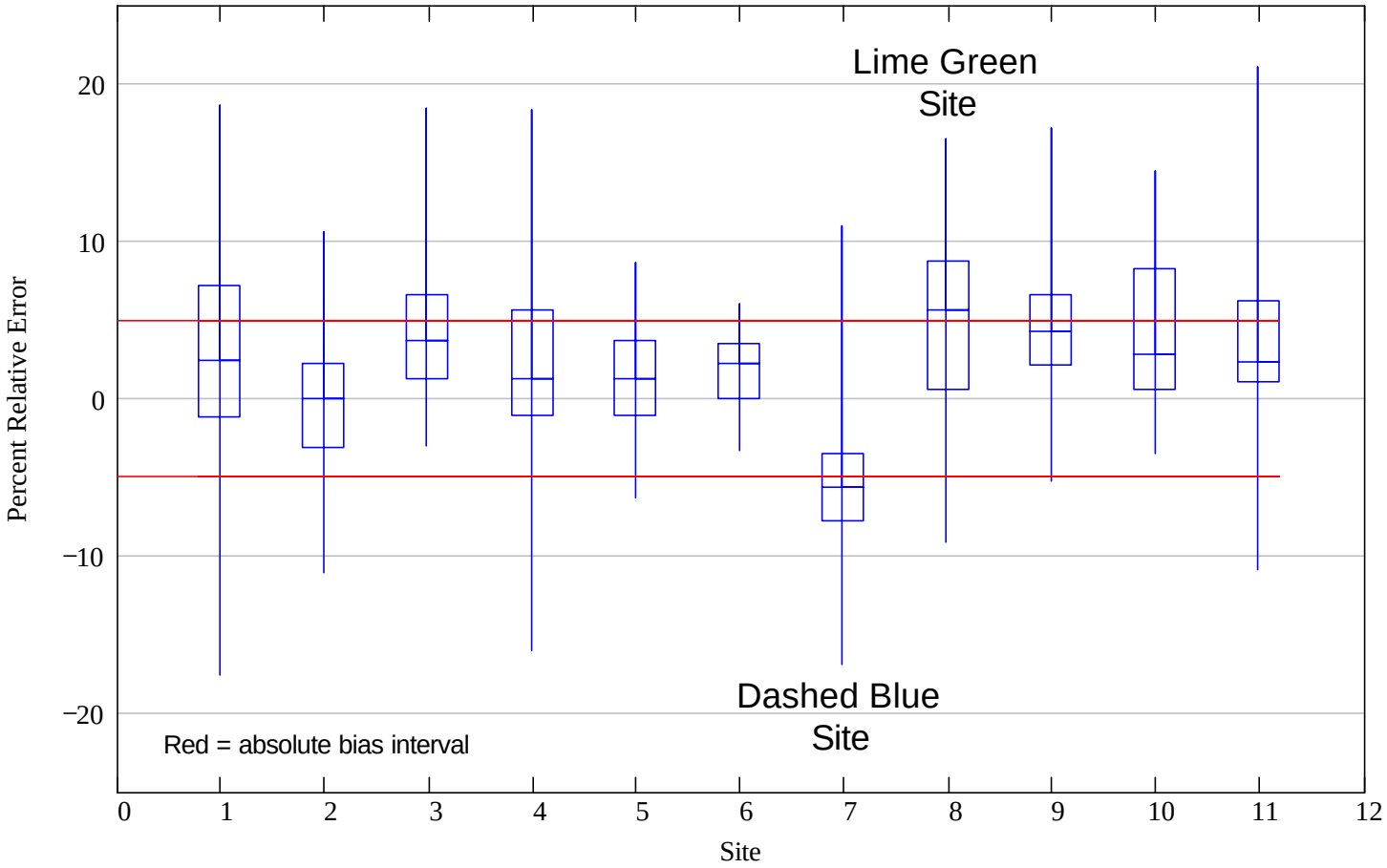
Now some examples
to help all this make intuitive sense.

How does this differ from what we currently do?

- **Current approach is like the person who sees imprecision instead of bias, that is, precision is emphasized.**



Percent Relative Differences in Ozone for Sites in “Variable” Reporting Organization, Aggregated for Each Site



Site	Bias	Prec
1	2.7%	7.0%
2	-0.4%	4.5%
3	4.2%	3.7%
4	2.6%	6.2%
5	1.4%	3.2%
6	1.7%	2.5%
7	-5.6%	4.7%
8	4.6%	6.3%
9	4.8%	5.2%
10	4.0%	4.4%
11	3.7%	5.8%

Several sites display systematic behavior. Also, spread is generally large.

And a few last details on the statistics...

- The DQO goals discussed in the previous presentation are not for the point estimates (the average bias and the precision).
- The goals are for the upper bounds on the confidence intervals for the point estimates. That is, what value are we pretty sure the bias is less than and what value are we pretty sure the precision is less than?
  - Student's t distribution used to estimate upper bound for absolute bias
  - Chi-square distribution used to estimate upper bound for precision
- These intervals are highly influenced by sample size. The more numbers, the closer the upper bounds are to the point estimates. The fewer the numbers, the wider the spread between the upper bounds and point estimates.
- Which provides us with QA flexibility.
  - The sites that have consistently small differences between the observed and true concentrations don't need to check so often, like those in the "typical" rep org or Site 5 in the "variable" rep org.
  - The sites that have inconsistencies, like most of the sites in the "variable" rep org but particularly the "Lime Green" and "Dashed Blue" sites, should have more checks and QA scrutiny to understand better their behavior.
  - So we free up resources from high quality sites and devote those resources to the sites with potentially lower quality data.

Example Table Showing Flexibility

Sample Sizes to Confidently Conclude **1**

(overall error rate controlled to less than 10 %)

Bias<=7 and Precision<=7

	Bias													
Prec	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5
2.0	4	4	4	4	4	4	4	4	4	4	5	7	13	45
2.5	5	5	5	5	5	5	5	5	5	5	7	10	19	69
3.0	6	6	6	6	6	6	6	6	6	6	8	13	26	98
3.5	7	7	7	7	7	7	7	7	7	8	11	17	35	133
4.0	9	9	9	9	9	9	9	9	9	9	13	21	45	173
4.5	12	12	12	12	12	12	12	12	12	12	16	26	56	218
5.0	18	18	18	18	18	18	18	18	18	18	19	32	69	269
5.5	31	31	31	31	31	31	31	31	31	31	31	38	83	325
6.0	68	68	68	68	68	68	68	68	68	68	68	68	98	386
6.5	265	265	265	265	265	265	265	265	265	265	265	265	265	453

3-Year Ozone Site Precision and Bias Point Estimates

Percentile	Precision	Bias	N
25%	1.9	1.6	262
50%	2.8	2.3	524
75%	3.7	3.3	786
90%	4.8	4.5	944

1049

Percentiles of 3-year (99-2001) Site Level Bias and Precision Upper Confidence Intervals

CO

	abs_bias	CV
P_0	0.495518	0.885291
P_10	1.32391	1.531072
P_20	1.730389	1.92744
P_30	2.008613	2.275285
P_40	2.318792	2.522563
P_50	2.760372	3.01155
P_60	3.182354	3.391786
P_70	3.678674	3.880518
P_80	4.407137	4.530607
P_90	5.366907	5.777561
P_100	19.28201	35.49346

NO2

	abs_bias	CV
P_0	0.494946	0.845987
P_10	2.085199	2.304727
P_20	2.691062	3.110572
P_30	3.245284	3.669459
P_40	3.665825	4.289854
P_50	4.12546	4.742549
P_60	4.630721	5.273939
P_70	5.17034	5.948534
P_80	5.898041	6.50689
P_90	6.802439	7.518919
P_100	32.68663	34.8422

SO2

	abs_bias	CV
P_0	0	0
P_10	1.602741	1.864202
P_20	1.858238	2.314077
P_30	2.334023	2.675836
P_40	2.641022	3.095242
P_50	3.114478	3.565135
P_60	3.559878	3.980958
P_70	4.032764	4.520087
P_80	4.650041	5.102396
P_90	5.545477	6.054677
P_100	13.55546	22.35704

O3

	abs_bias	CV
P_0	0.40894	0
P_10	1.229341	1.522103
P_20	1.642159	1.955228
P_30	1.979491	2.329164
P_40	2.380341	2.697468
P_50	2.72313	3.118359
P_60	3.05972	3.52292
P_70	3.50187	4.017527
P_80	4.041713	4.553211
P_90	5.00315	5.431491
P_100	24.04165	26.22982

What's Next?

- **Have concurrence with proposal from:**
  - **National QA Workgroup, AQSSD/AQTAG, old guard**
- **Proposal may be discussed at next SAMWG meeting (Oct)**
- **Modify CFR to reflect this new philosophy (site-specific estimates and flexibility) and these new statistics**
- **Develop regular reports available from AQS so Monitoring and QA personnel have the tools they need to oversee Bias and Precision**
  - **box-and-whisker plots for each rep org showing the behavior of each site**
  - **summary bias and precision statistics for each site, including sample size**

Example Table Showing Flexibility

Sample Sizes to Confidently Conclude
(overall error rate controlled to less than 10 %)

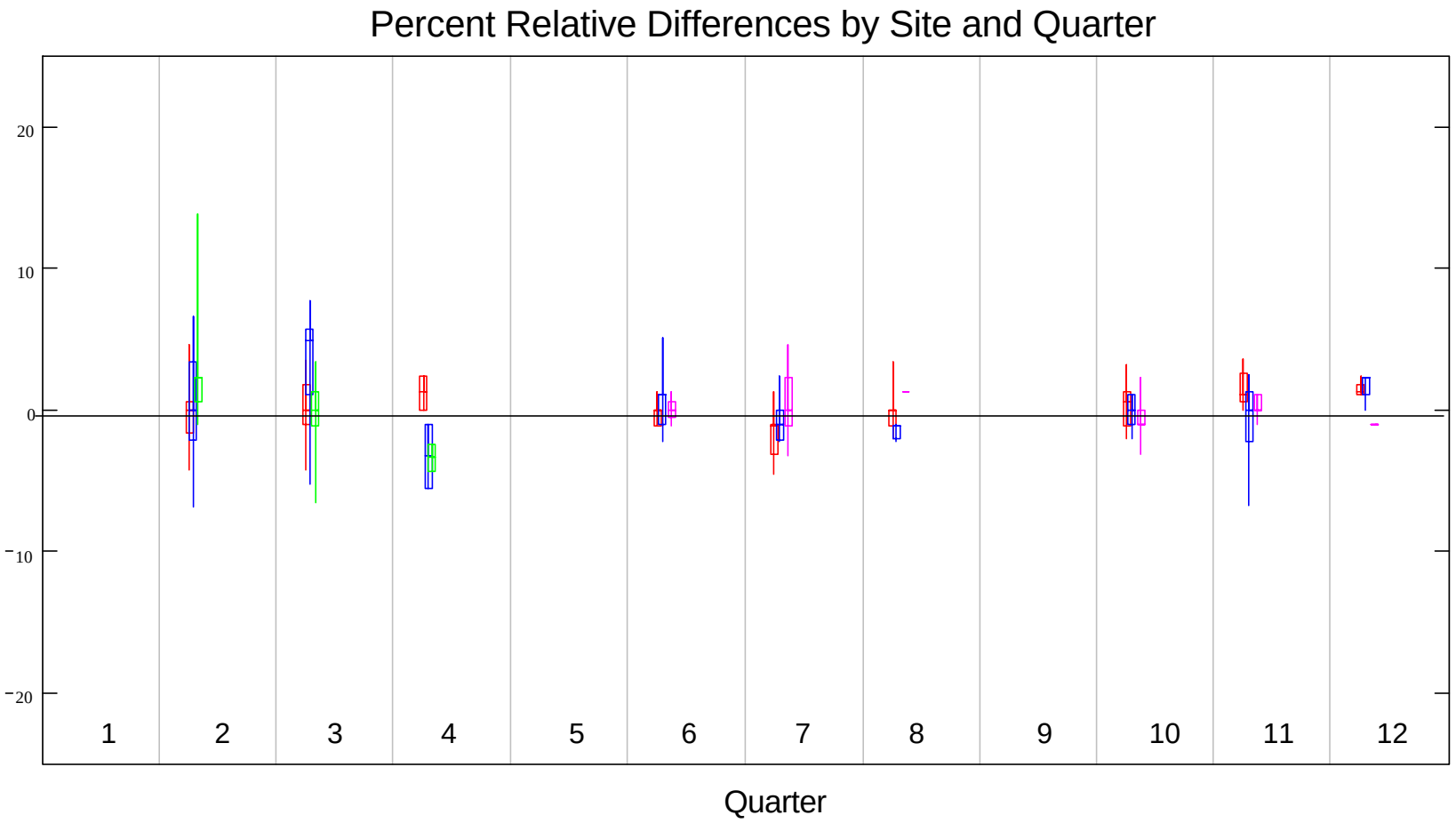
Bias<=5 and Precision<=5

	Bias									
Prec	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5
2.0	5	5	5	5	5	5	5	7	13	45
2.5	7	7	7	7	7	7	7	10	19	69
3.0	10	10	10	10	10	10	10	13	26	98
3.5	17	17	17	17	17	17	17	17	35	133
4.0	36	36	36	36	36	36	36	36	45	173
4.5	137	137	137	137	137	137	137	137	137	218

How does this differ from what we currently do? (continued)

- **Current approach is like the person who sees imprecision instead of bias, that is, precision is emphasized.**
- **Aggregation in current approach is at reporting organization, not at each site. That is, the summary statistics mix the sites together and thus makes it harder to see if one or two sites behaves differently than the rest.**

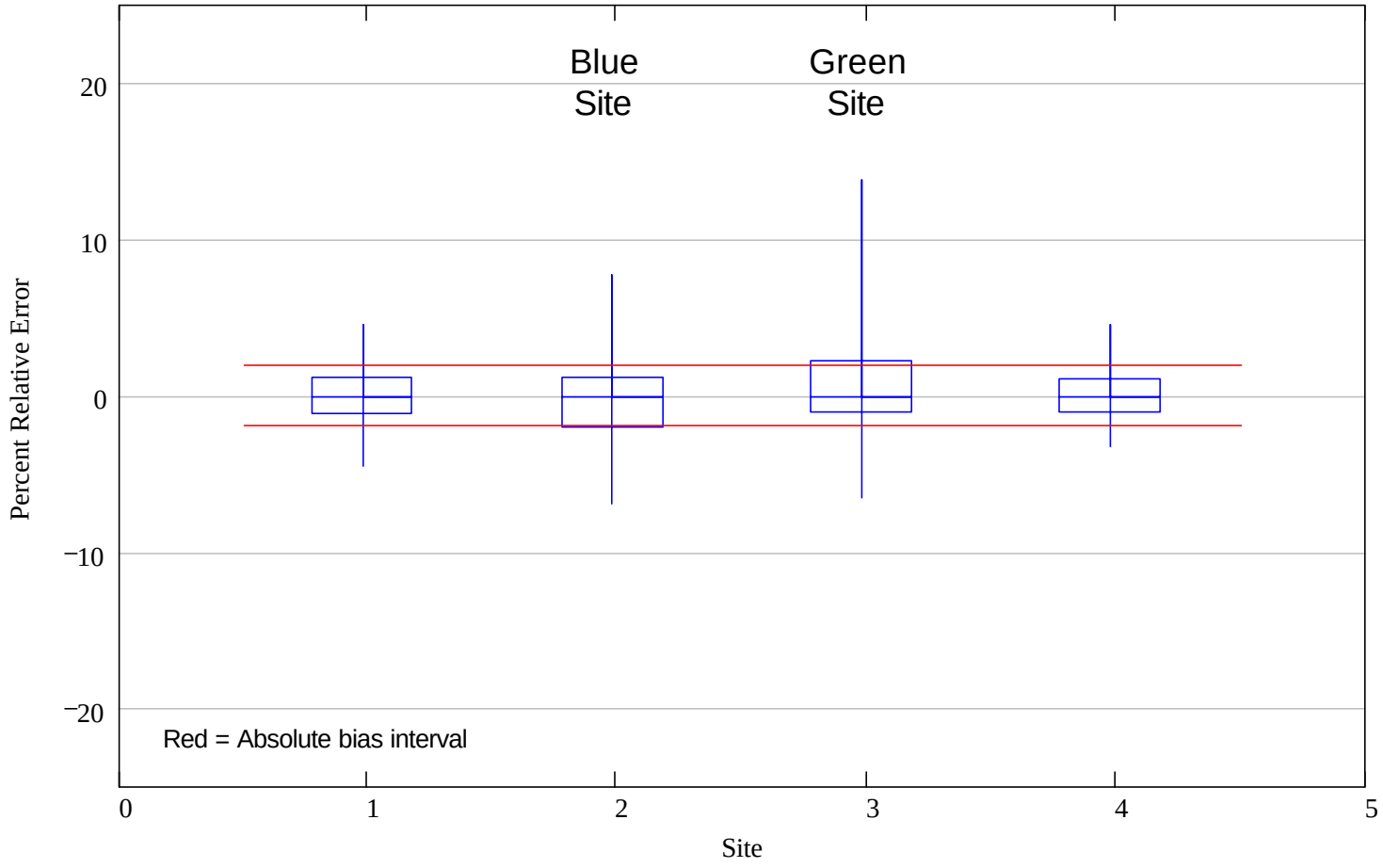
Percent Relative Differences in Ozone for Sites in “Typical” Reporting Organization



Green and Blue sites more scatter.

Red and Purple sites less scatter.

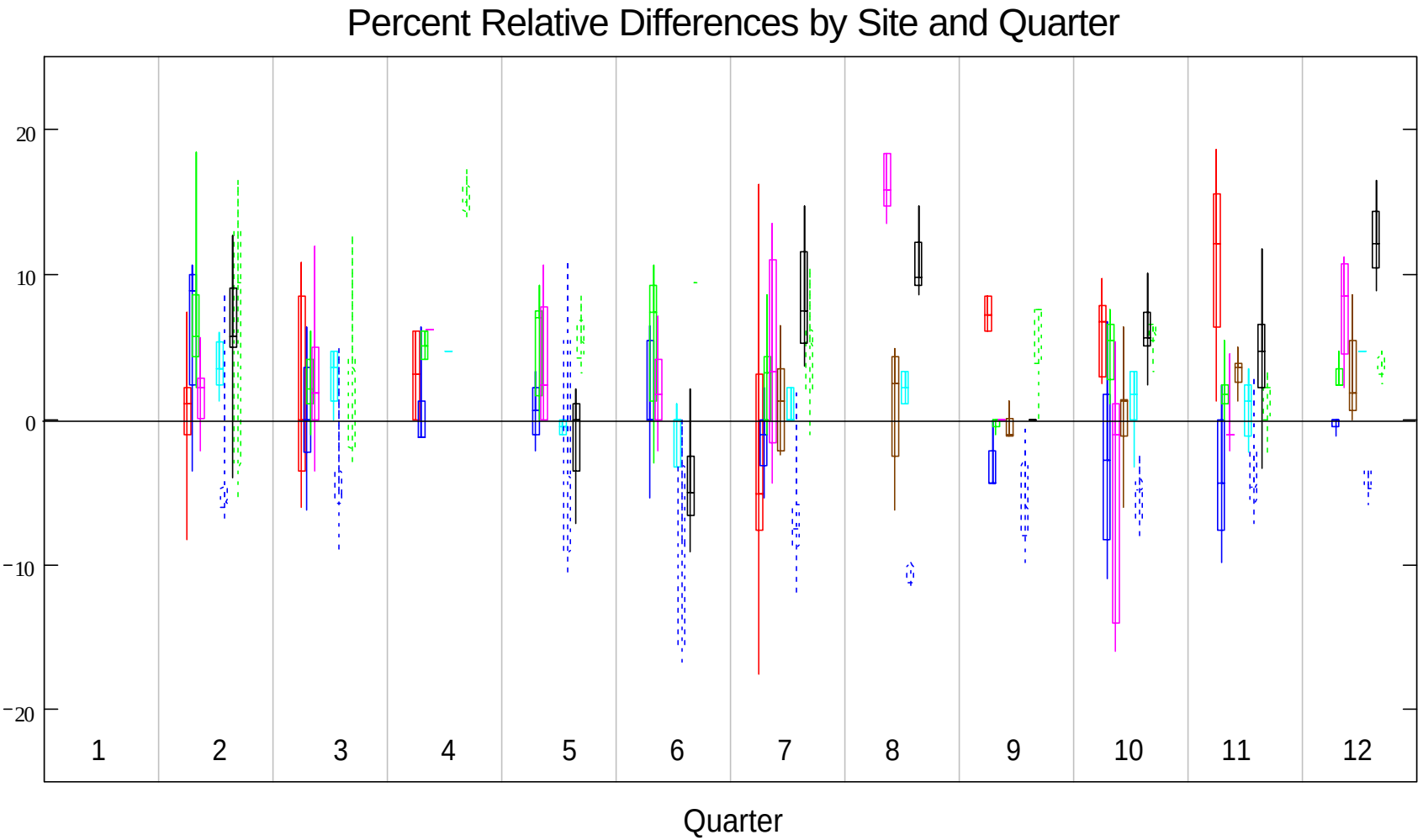
Percent Relative Differences in Ozone for Sites in “Typical” Reporting Organization, Aggregated for Each Site



Site	Bias	Prec
1	-0.1%	2.0%
2	0.1%	3.0%
3	0.6%	4.2%
4	-0.1%	1.5%

1. Bias nearly 0 for each site.
2. Precision tight but 1 site more variable than others.

Percent Relative Differences in Ozone for Sites in “Variable” Reporting Organization



Lime Green high all the time. Dashed Blue low most of the time.