Flint, Michigan
Sequential water data analysis
10 June 2017

John Canar
Chuck Roth
Brian Cooper

FIELDS Group
U.S. EPA, Region 5, Superfund

Outline

1. Residential sequential results (Lead):
   - Round 1 vs. Round 3 (paired homes)
   - Round 1 vs. Round 5 (paired homes)

2. Residential DS (flushed distribution system) results:
   - Lead (all homes)
   - Total Phosphorus (all homes)

3. Residential sequential results (Lead):
   - R1 vs. R3 vs. R4 vs. R5 (all homes)

1a. Residential sequential results (Lead)

Round 1 vs. Round 3 (20 paired homes)

(a measure of how soon things got better)

Statistical notes

Residential sequential results for Rounds 1, 3, and 5 were compared two ways:

- as “paired” data, i.e., assuming the sequential samples are truly paired (dependent) [t-test, Sign, WRS]
- as “independent” data, i.e., assuming that the sequential samples are not paired and tested using two sample tests [ANOVA, KW, Median]

Outcome:

- virtually identical results whether tested as paired or not paired
- paired results shown in slides

Sequential Lead levels by Sampling Event
Two-sample Nonparametric paired tests
Round 1 minus Round 3
Flint, Michigan

The UNIVARIATE Procedure
Variable: Diff_1_3 (R1 – R3)

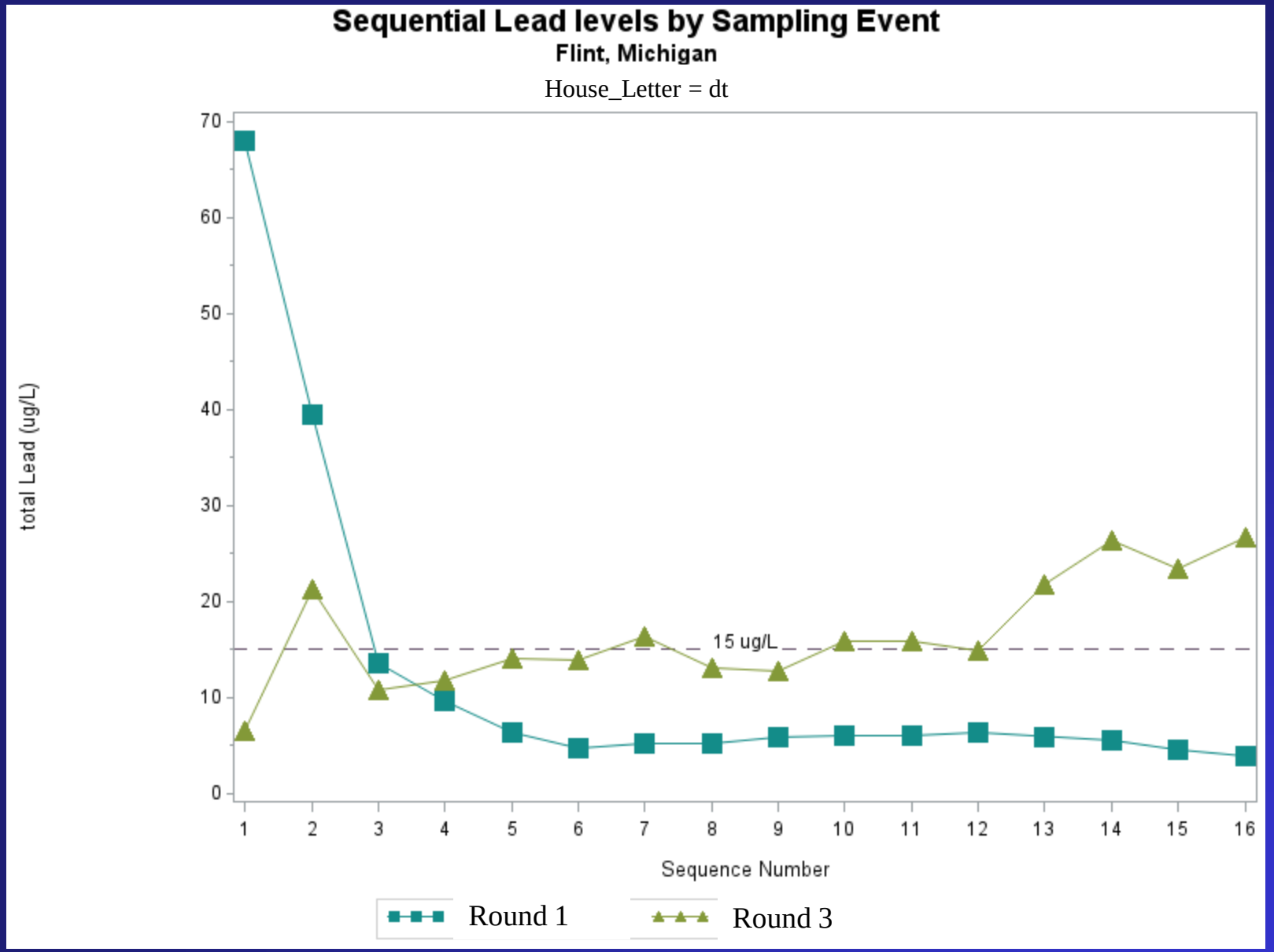
House_Letter = dt

Moments			
N	16	Sum Weights	16
Mean	-4.33375	Sum Observations	-69.34
Std Deviation	20.0243651	Variance	400.975198
Skewness	2.65070055	Kurtosis	8.17938733
Uncorrected SS	6315.1302	Corrected SS	6014.62798
Coeff Variation	-462.0563	Std Error Mean	5.00609128

Basic Statistical Measures			
Location		Variability	
Mean	-4.33375	Std Deviation	20.02437
Median	-8.84000	Variance	400.97520
Mode	-9.84000	Range	84.15000
		Interquartile Range	8.97500

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	-0.8657	Pr > |t|	0.4003
Sign	M	-5	Pr >= |M|	0.0213
Signed Rank	S	-38	Pr >= |S|	0.0492

Significantly HIGHER total Lead in R3


Sequential Lead levels by Sampling Event
Flint, Michigan
House_Letter = dt
total Lead (ug/L)
70
60
50
40
30
20
10
0
15 ug/L
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Sequence Number
Round 1
Round 3

Sequential Lead levels by Sampling Event
Two-sample Nonparametric paired tests
Round 1 minus Round 3
Flint, Michigan

The UNIVARIATE Procedure
Variable: Diff_1_3 (R1 – R3)

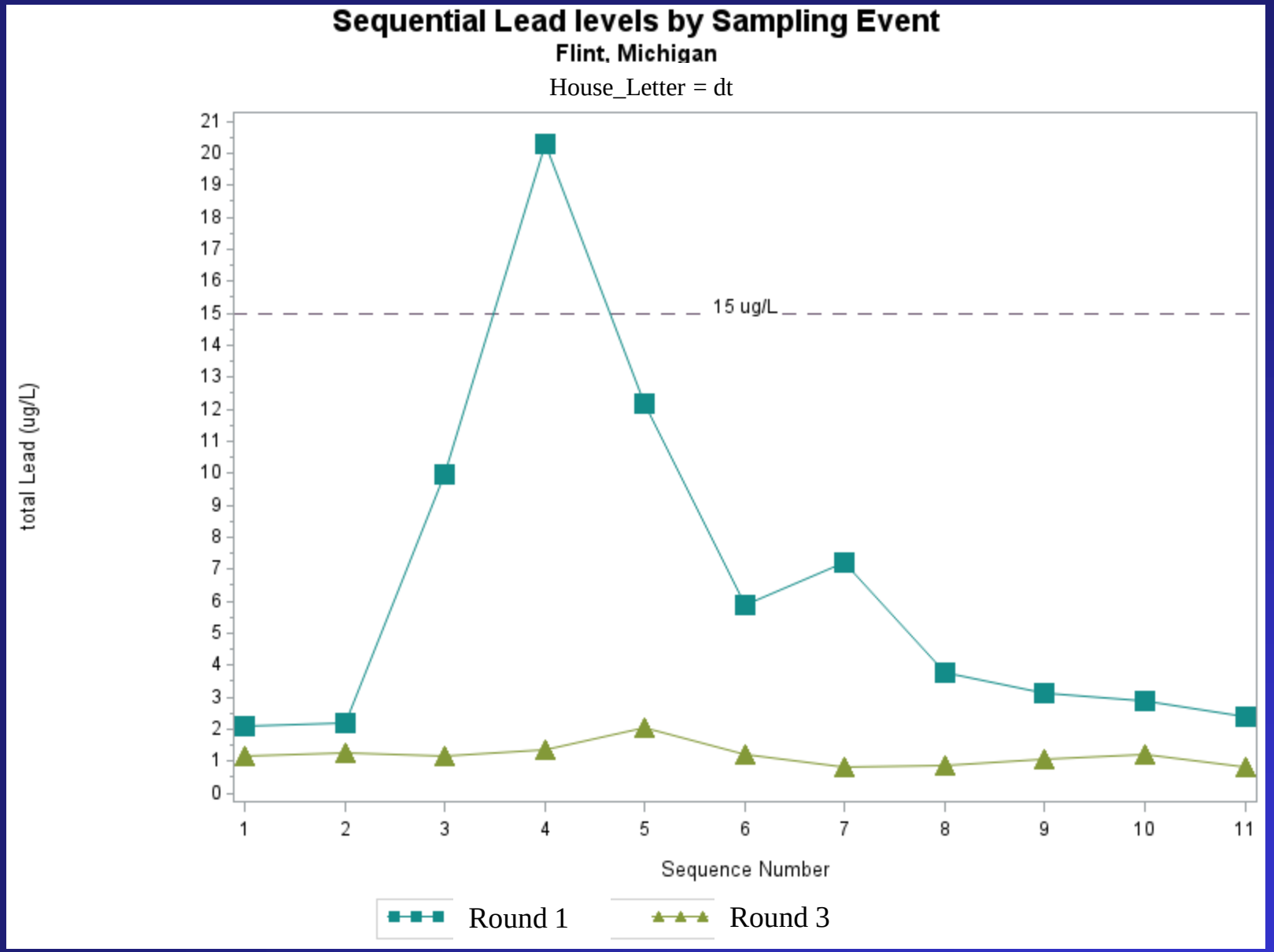
House_Letter = ct

Moments			
N	11	Sum Weights	11
Mean	5.36636364	Sum Observations	59.03
Std Deviation	5.51005857	Variance	30.3607455
Skewness	1.68915974	Kurtosis	2.93678044
Uncorrected SS	620.3839	Corrected SS	303.607455
Coeff Variation	102.677697	Std Error Mean	1.66134517

Basic Statistical Measures			
Location		Variability	
Mean	5.366364	Std Deviation	5.51006
Median	2.920000	Variance	30.36075
Mode	.	Range	17.99000
		Interquartile Range	7.21000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	3.230132	Pr > |t|	0.0090
Sign	M	5.5	Pr >= |M|	0.0010
Signed Rank	S	33	Pr >= |S|	0.0010

Significantly lower total Lead in Seq_3


Sequential Lead levels by Sampling Event
Flint, Michigan
House_Letter = dt
total Lead (ug/L)
15 ug/L
Sequence Number
Round 1
Round 3

House_Letter	Student's t	Sign test	Signed Rank test	Outcome
dt	NS	S	S	R1 < R3
df	S	S	S	R1 < R3
bs	S	0.0574	S	R1 > R3
bf	NS	NS	NS	NA
ct	S	S	S	R1 > R3
by	NS	S	S	R1 > R3
bc	NS	NS	NS	NA
n	S	S	S	R1 < R3
t	S	S	S	R1 > R3
am	S	S	S	R1 > R3
w	S	S	S	R1 > R3
ak	S	S	S	R1 > R3
bh	NS	NS	NS	NA
dj	NS	NS	NS	NA
dg	NS	NS	NS	NA
dk	S	NS	S	R1 < R3
g	S	S	S	R1 < R3
m	S	S	S	R1 > R3
cd	S	S	S	R1 < R3
cf	NS	NS	NS	NA

Notes:

NS = not statistically significant

S = statistically significant

NA = not applicable

#.#### = p-value (used when slightly above 0.05)

Outcome:

40% statistically lower

30% no significant difference

30% statistically higher

1b. Residential sequential results (Lead)

Round 1 vs. Round 5 (15 paired homes)

(a measure of how things are now compared to the beginning)

Sequential Lead levels by Sampling Event
Two-sample Nonparametric paired tests
Round 1 minus Round 5
Flint, Michigan

The UNIVARIATE Procedure
Variable: Diff_1_5 (R1 – R5)

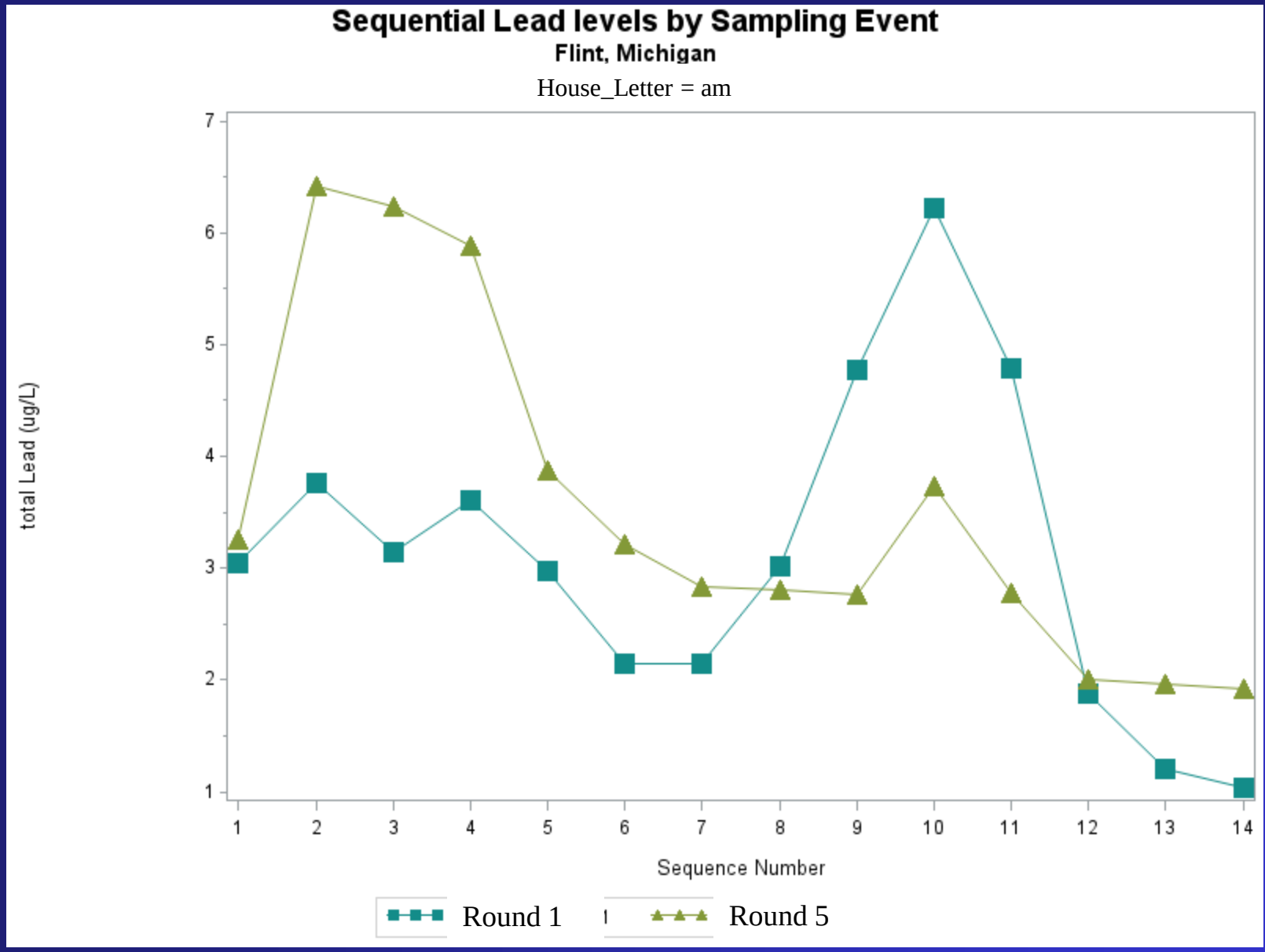
House_Letter = am

Moments			
N	14	Sum Weights	14
Mean	-0.4264286	Sum Observations	-5.97
Std Deviation	1.69389223	Variance	2.86927088
Skewness	0.31153172	Kurtosis	-0.451862
Uncorrected SS	39.8463	Corrected SS	37.3005214
Coeff Variation	-397.22766	Std Error Mean	0.45271174

Basic Statistical Measures			
Location		Variability	
Mean	-0.42643	Std Deviation	1.69389
Median	-0.72000	Variance	2.86927
Mode	.	Range	5.56000
		Interquartile Range	1.28000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	-0.94194	Pr > |t|	0.3634
Sign	M	-3	Pr >= |M|	0.1796
Signed Rank	S	-19.5	Pr >= |S|	0.2412

NO sig. differences


Sequential Lead levels by Sampling Event
Flint, Michigan
House_Letter = am
total Lead (ug/L)
Sequence Number
Round 1
Round 5

House_Letter	Student's t	Sign test	Signed Rank test	Outcome
df	NS	NS	NS	NA
bs	S	S	S	R1 > R5
bf	S	S	S	R1 > R5
bc	S	0.0574	S	R1 > R5
z	S	S	S	R1 < R5
a	NS	NS	NS	NA
n	S	S	S	R1 > R5
am	NS	NS	NS	NA
w	S	S	S	R1 > R5
bh	NS	S	0.0569	R1 < R5
dk	S	S	S	R1 > R5
at2	S	S	S	R1 > R5
m	S	S	S	R1 > R5
cd	S	S	S	R1 > R5
cf	S	S	S	R1 > R5

Notes:

NS = not statistically significant

S = statistically significant

NA = not applicable

#.#### = p-value (used when slightly above 0.05)

Outcome:
67% statistically lower
20% no significant difference
13% statistically higher

Summary residential sequential results (Lead)

Comparing sequential sampling R1-R3 vs. R1-R5:

- the percentage of homes with statistically lower sequential Lead levels increased from 40% to 67%. Hence two thirds of the homes show a statistical decrease in Lead levels.
- the percentage of homes with statistically higher sequential Lead levels decreased from 30% to 13%.
- regardless of sampling event, two thirds of the homes have no sequential Lead levels above 15 ug/L. One third of the homes have one or more sequential Lead levels above 15 ug/L.

Summary residential sequential results (Lead)

Plumbing: no LSLs, but don't know between water main and house
Lead/Phosphorus: DS samples don't explain differences

		R1 => R3	R1 => R5	One or more samples >15ug/L			DSPhosphorus (mg/L)			DS Lead (ug/L)		
House_Letter	**Disposition**	**Stat. Signifigance**	**Stat. Signifigance**	**R1 > 15ug/L**	**R3 > 15ug/L**	**R5 > 15ug/L**	R1_P	R3_P	R5_P	R1_Pb	R3_Pb	R5_Pb
	Did Not Improve											
df	*One of five final events show significant decrease (dk).*	R1 < R3	R1 = R5	Above	Above	Above	0.61	0.96	1.02	1.2	0.97	1.1
dk	*DK had one sample (seq #17) >15ug/L in R5. DG was*	R1 < R3	R1 > R5	Above	Above	Above	0.54	0.97	1.37	5.25	4	1.8
dg	*below 15ug/L except for sequence #1 for both R1*	R1 = R3		Above	Above		0.93	1.01		0.5	0.5	
dt	*and R3 (>50ug/L).*	R1 < R3		Above	Above		1.13	0.92		1.5	2.29	
a			R1 = R5	Above		Above	1.04		0.99	6.27		3.3
	Improved											
bs	*All final sequences were significantly lower with all*	R1 > R3	R1 > R5	Above	Above	Below	1.38	0.92	1.08	2.1	1.7	1.1
n	*samples below 15 ug/L.*	R1 < R3	R1 > R5	Above	Above	Below	1.01	1.05	1.02	1.41	2.08	0.54
w		R1 > R3	R1 > R5	Above	Below	Below	1.09	0.96	0.99	2.45	0.5	0.5
ak		R1 > R3		Above	Below		0.98	0.96		0.5	0.5	
by		R1 > R3		Above	Below		1.15	0.15		0.5	0.5	
ct		R1 > R3		Above	Below		1.07	0.99		0.581	12.4	
dj		R1 = R3		Above	Below		1.12	0.97		0.5	0.5	
at			R1 > R5	Above		Below	1.09		1.01	3.18		1.9
	Low and Stayed Low											
am	*5 of the 11 tests showed no change or significant decrease. The*	R1 > R3	R1 = R5	Below	Below	Below	1.09	1.02	0.84	0.718	0.73	0.77
bc	*results suggest random variabilty, as opposed to changes due to*	R1 = R3	R1 > R5	Below	Below	Below	0.9	0.89	0.99	0.697	0.5	0.5
bf	*conditions in the plumbing (given all of the results were*	R1 = R3	R1 > R5	Below	Below	Below	1.27	1.02	1.02	1.69	2.67	1.4
cf	*relatively low and stayed low).*	R1 = R3	R1 > R5	Below	Below	Below	1.22	1.03	1.19	0.5	0.93	0.61
m		R1 > R3	R1 > R5	Below	Below	Below	1.04	1.02	1.15	1.33	0.89	0.5
g		R1 < R3		Below	Below		1.2	1.01		1.51	1.58	
t		R1 > R3		Below	Below		1.14	1.1		0.5	0.5	
	Mixed											
z	*Z final sequential samples >15ug/L for 4 samples (sample#3-6)*		R1 < R5	Below		Above	1.2		1.02	0.5		0.6
cd	*CD final sequential samples all >15ug/L*	R1 < R3	R1 > R5	Below	Above	Below	1.14	0.94	0.99	0.61	3.13	2.01
bh	*BH had one R5 sample >15ug/L (sample #1>25 ug/L)*	R1 = R3	R1 < R5	Below	Below	Above	1.36	1.06	1	0.5	0.52	1.1

2. Residential DS (flushed distribution system) results

(Lead and Phosphorus)

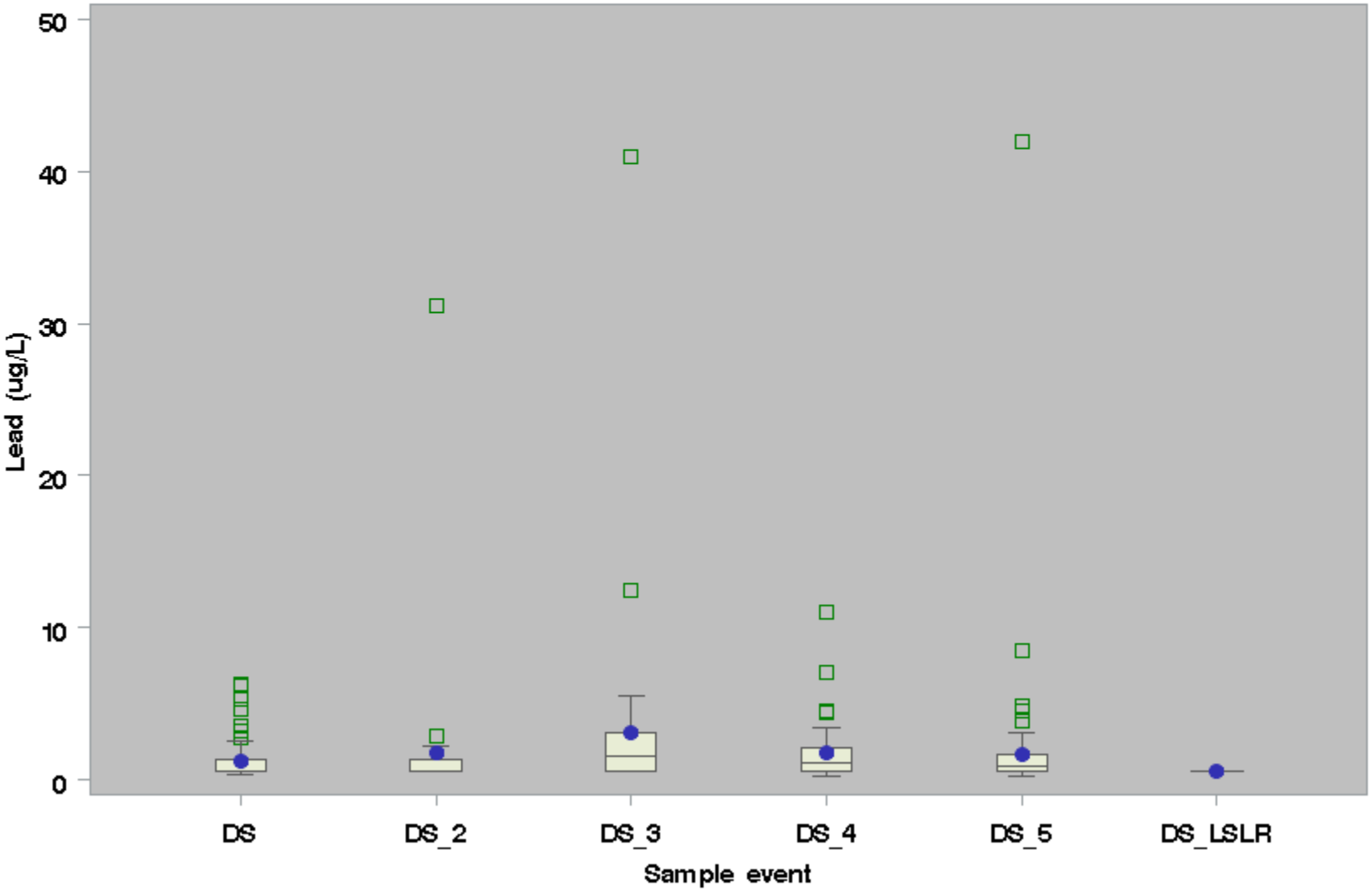
Lead

Boxplots of Lead levels by Sampling event

Flint, Michigan

N	73	35	42	48	102	4
Mean	1.19	1.80	3.06	1.71	1.65	0.50
Q2	0.52	0.56	1.59	1.08	0.92	0.50
Range	5.92	30.7	40.5	10.8	41.8	0.02

Median



Lead

R1

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	6.27
99%	6.27
95%	4.62
90%	2.50
75% Q3	1.33
50% Median	0.52
25% Q1	0.50
10%	0.50
5%	0.50
1%	0.35
0% Min	0.35

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.35	69	3.50	59
0.35	25	4.62	19
0.50	73	5.25	60
0.50	72	6.17	5
0.50	71	6.27	35

R2

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	31.20
99%	31.20
95%	2.90
90%	1.99
75% Q3	1.30
50% Median	0.56
25% Q1	0.50
10%	0.50
5%	0.50
1%	0.50
0% Min	0.50

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.5	104	1.96	81
0.5	103	1.99	82
0.5	98	2.16	91
0.5	97	2.90	105
0.5	96	31.20	100

R3

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	41.000
99%	41.000
95%	5.500
90%	4.900
75% Q3	3.130
50% Median	1.585
25% Q1	0.500
10%	0.500
5%	0.500
1%	0.500
0% Min	0.500

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.5	146	4.90	130
0.5	140	5.35	109
0.5	138	5.50	141
0.5	136	12.40	122
0.5	134	41.00	121

R4

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	11.00
99%	11.00
95%	4.50
90%	3.40
75% Q3	2.04
50% Median	1.08
25% Q1	0.50
10%	0.50
5%	0.32
1%	0.20
0% Min	0.20

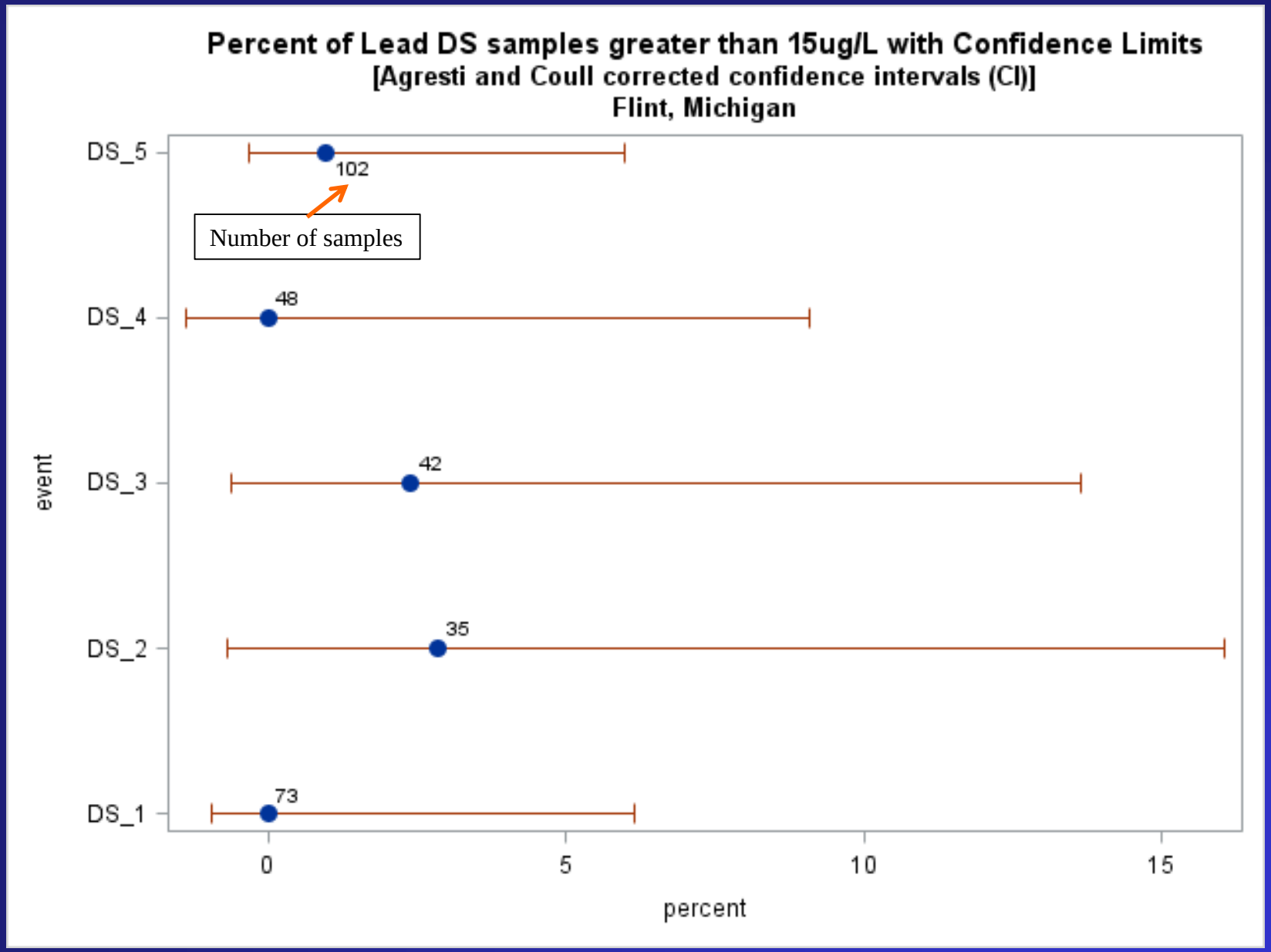
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.20	157	3.4	196
0.26	164	4.4	187
0.32	193	4.5	173
0.45	186	7.0	185
0.50	194	11.0	169

R5

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	42.000
99%	8.500
95%	3.100
90%	2.480
75% Q3	1.700
50% Median	0.915
25% Q1	0.500
10%	0.500
5%	0.500
1%	0.260
0% Min	0.250

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.25	273	3.80	279
0.26	274	4.48	254
0.27	272	4.87	271
0.38	284	8.50	236
0.48	283	42.00	261

Lead


Percent of Lead DS samples greater than 15ug/L with Confidence Limits
[Agresti and Coull corrected confidence intervals (CI)]
Flint, Michigan
DS_5
DS_4
DS_3
DS_2
DS_1
event
102
48
42
35
73
Number of samples
0
5
10
15
percent

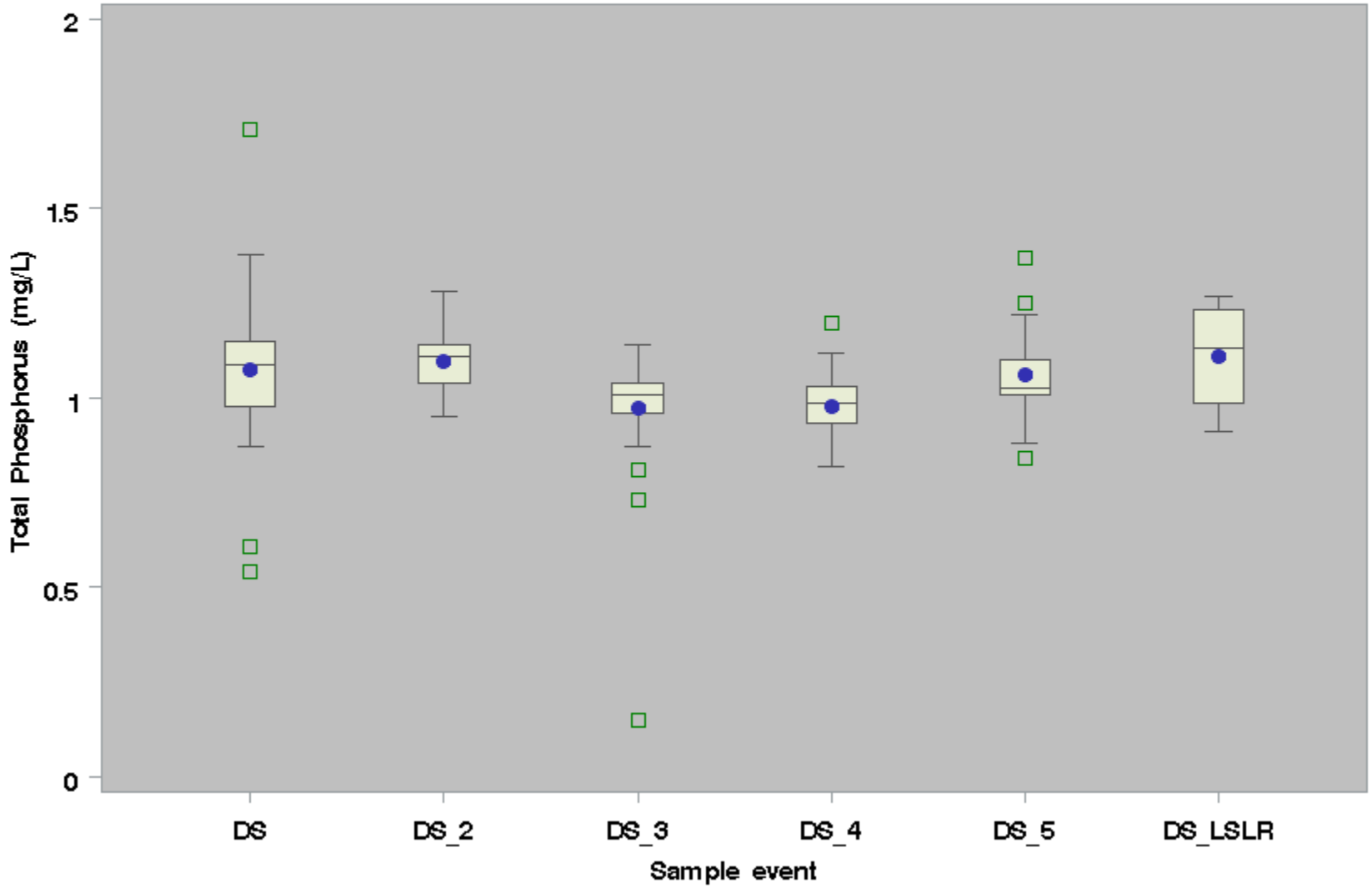
Phosphorus

Boxplots of Total Phosphorus levels by Sampling event

Flint, Michigan

N	73	35	42	48	42	4
Mean	1.07	1.10	0.97	0.98	1.06	1.11
Q2	1.09	1.11	1.01	0.99	1.03	1.13
Range	1.17	0.33	0.99	0.38	0.53	0.36

Median



Phosphorus

R1

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	1.71
99%	1.71
95%	1.30
90%	1.22
75% Q3	1.15
50% Median	1.09
25% Q1	0.98
10%	0.90
5%	0.87
1%	0.54
0% Min	0.54

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.54	60	1.27	15
0.61	7	1.30	54
0.87	41	1.36	52
0.87	28	1.38	13
0.88	58	1.71	14

R2

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	1.28
99%	1.28
95%	1.20
90%	1.17
75% Q3	1.14
50% Median	1.11
25% Q1	1.04
10%	1.00
5%	0.99
1%	0.95
0% Min	0.95

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.95	84	1.15	108
0.99	81	1.17	77
1.00	107	1.19	105
1.00	76	1.20	92
1.03	99	1.28	93

R3

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	1.14
99%	1.14
95%	1.10
90%	1.06
75% Q3	1.04
50% Median	1.01
25% Q1	0.96
10%	0.89
5%	0.81
1%	0.15
0% Min	0.15

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.15	123	1.06	137
0.73	112	1.08	130
0.81	121	1.10	127
0.87	148	1.10	143
0.89	124	1.14	111

R4

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	1.200
99%	1.200
95%	1.100
90%	1.070
75% Q3	1.030
50% Median	0.985
25% Q1	0.935
10%	0.870
5%	0.840
1%	0.820
0% Min	0.820

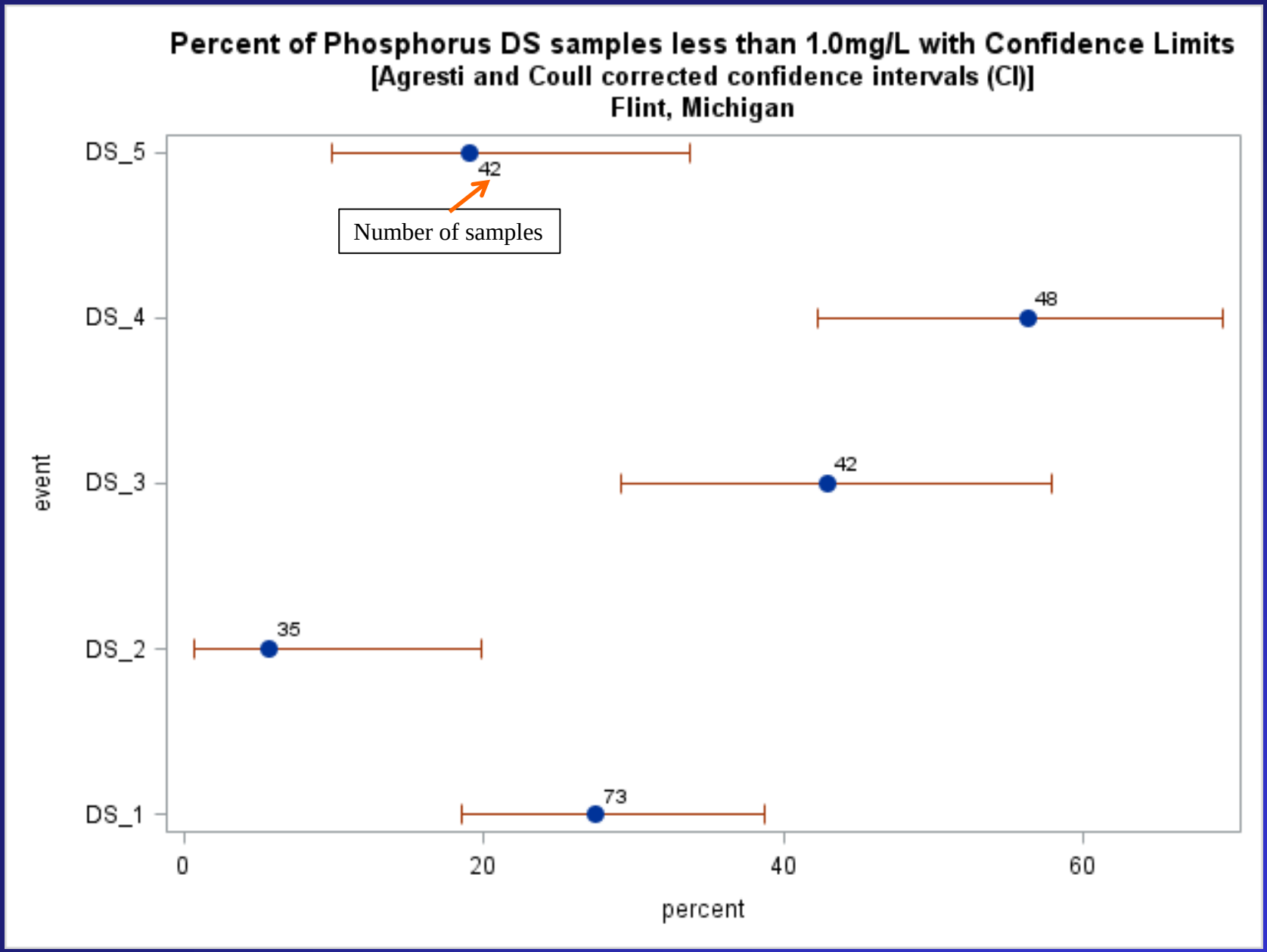
Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.82	173	1.07	192
0.84	188	1.09	190
0.84	182	1.10	163
0.86	154	1.12	183
0.87	166	1.20	151

R5

Quantiles (Definition 5)	
Quantile	Estimate
100% Max	1.370
99%	1.370
95%	1.220
90%	1.190
75% Q3	1.100
50% Median	1.025
25% Q1	1.010
10%	0.990
5%	0.980
1%	0.840
0% Min	0.840

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.84	214	1.19	240
0.88	220	1.20	234
0.98	230	1.22	199
0.99	239	1.25	238
0.99	218	1.37	226

Phosphorus


Percent of Phosphorus DS samples less than 1.0mg/L with Confidence Limits
[Agresti and Coull corrected confidence intervals (CI)]
Flint, Michigan
DS_5
DS_4
DS_3
DS_2
DS_1
event
42
Number of samples
48
42
35
73
0
20
40
60
percent

Summary residential DS (flushed distribution system) results:

Lead has a very low percent of values above 15 ug/L, and small confidence limits. Only three sampling events, DS_2, DS_3, and DS_5, had concentrations above 15 ug/L and that occurred only once for each sampling event. Hence, it appears that elevated levels of Lead in homes is highly unlikely to come from the water main, i.e., much more likely to come from interior plumbing and/or the service line.

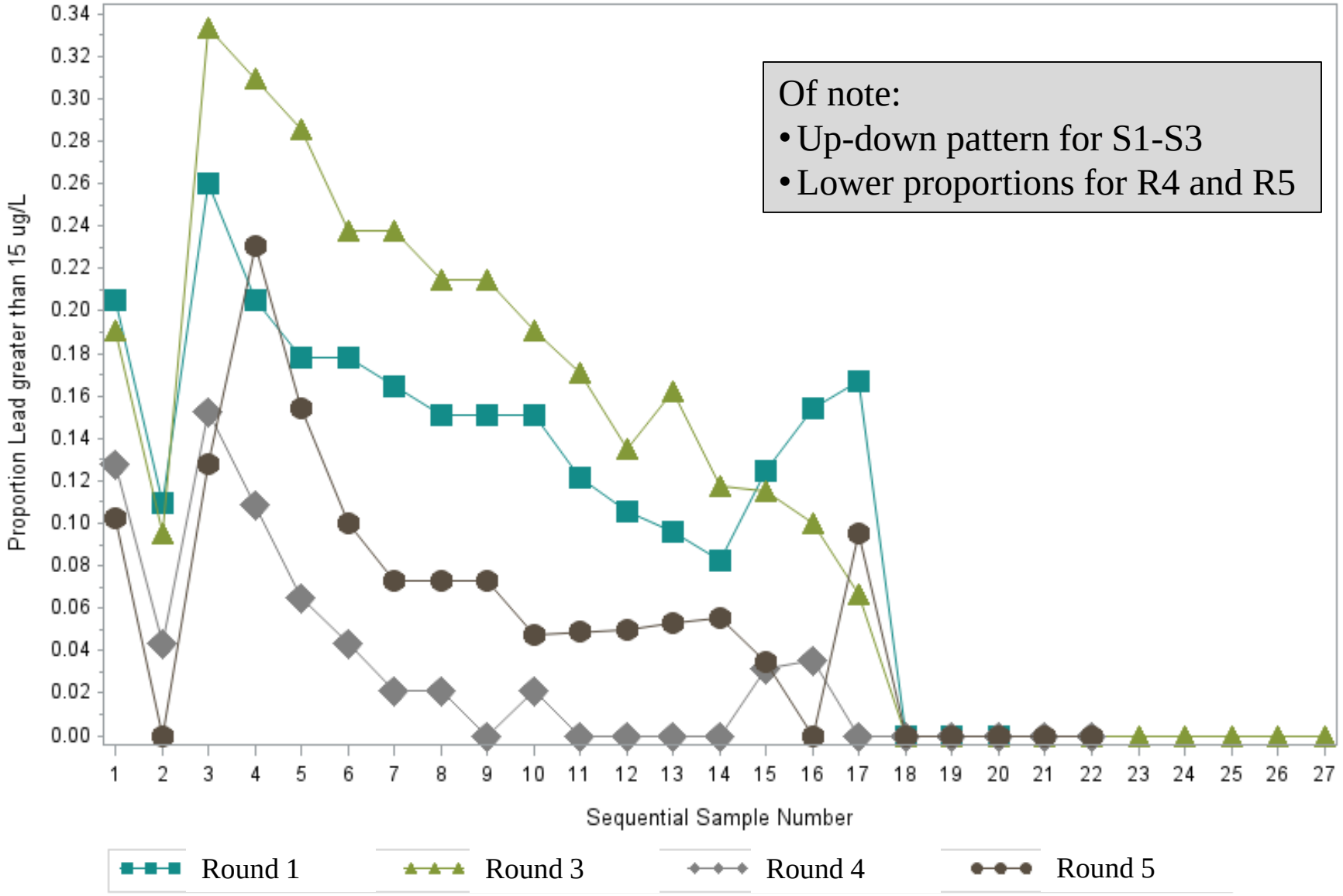
Compared to Lead, Phosphorus has a higher percent of values with concentrations less than 1.0 mg/L. (Where 1.0 mg/L is the minimum residual phosphorus concentration requirement set by the State for the City of Flint distribution system samples.) The percent of Phosphorus samples with concentrations less than 1.0 mg/L by sampling event ranges from 6 to 56%.

3. Residential sequential results (Lead)

R1 vs. R3 vs. R4 vs. R5 (all homes)

Proportion of Lead Sequential samples greater than 15 ug/L

[by Sequential sampling event]
Flint, Michigan



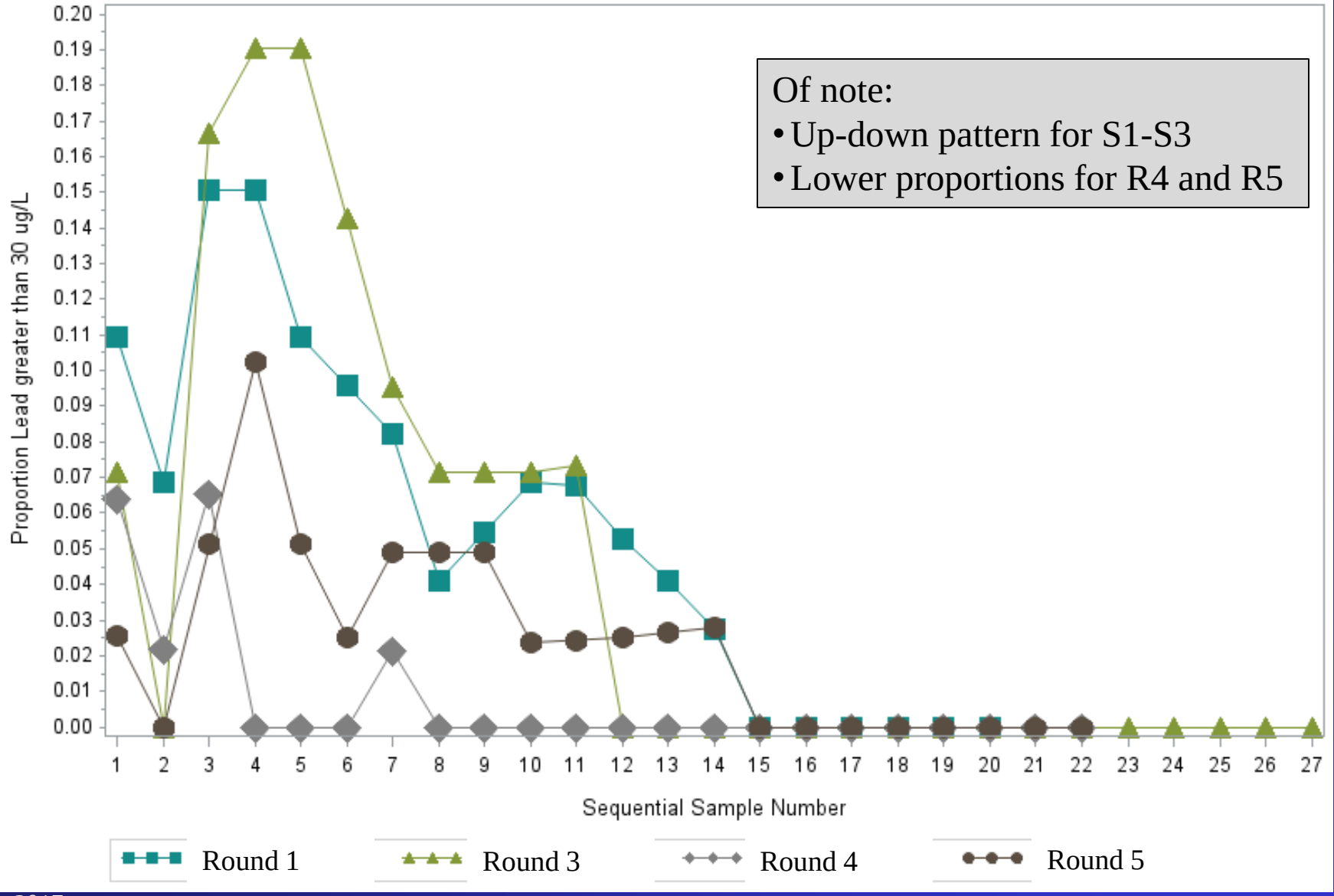
Of note:

- Up-down pattern for S1-S3
- Lower proportions for R4 and R5

Proportion of Lead Sequential samples greater than 30 ug/L

[by Sequential sampling event]

Flint, Michigan



Of note:

- Up-down pattern for S1-S3
- Lower proportions for R4 and R5