Human Fecal Score: A standardized method for MST data interpretation

Yiping Cao, Ph.D.

Central/Northern California Ocean and Bay
Water Quality Monitoring Group

February 1, 2018

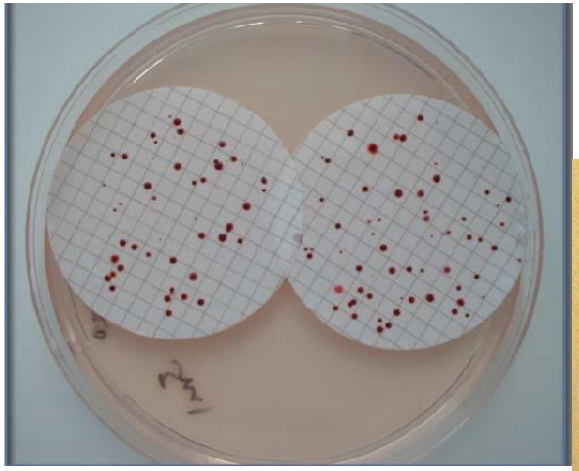
CALIFORNIA WATER BOARDS
State Water Resources Control Board

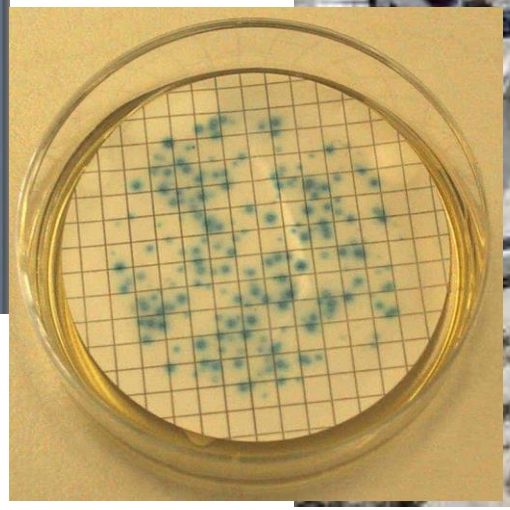
Source Molecular

ilac-MRA ACCREDITED Certificate 4126.01
ISO 17025 Accredited Testing Laboratory

Water Monitoring

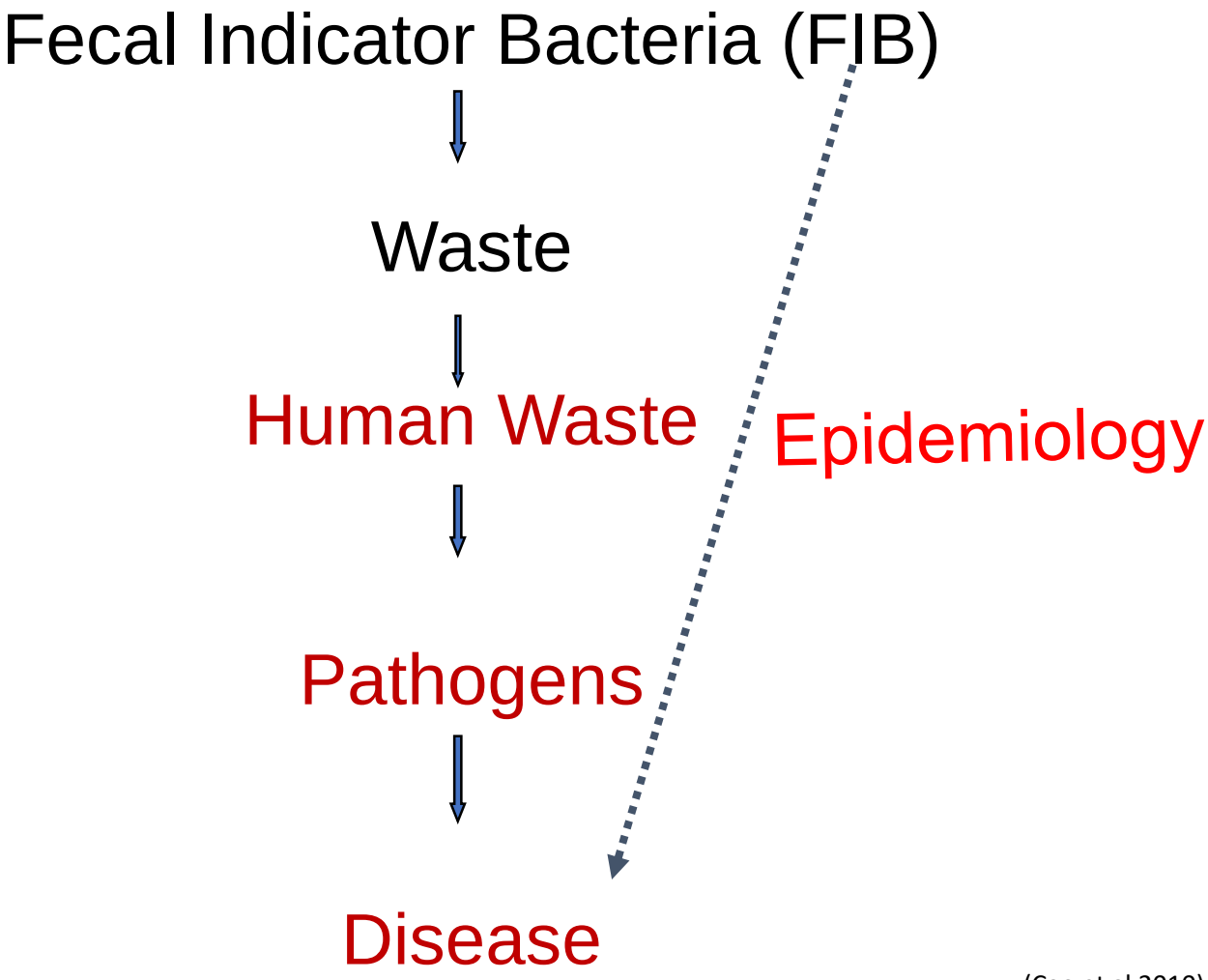
- Fecal Indicator Bacteria (FIB)
  - *Enterococcus* spp.
  - Fecal or total coliforms, *E. coli*





(Image: Donna Ferguson, Google)

Basis for Monitoring:
the chain of inference



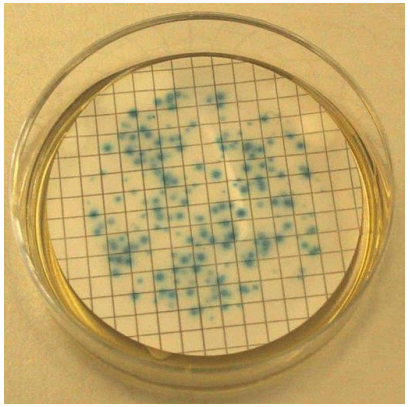
(Cao et al 2010)

Issue: Unspecific

Animals

Human

Plants



Soil,
biofilm

(Image: Donna Ferguson, Google)

Consequences

Mitigation:

- Hinders source abatement
  - Source identification must precede mitigation

TMDL:

- Undermines the basis for water monitoring (i.e. the chain of inference)

Risk Assessment:

  - Not all sources present the same level of risk

Microbial Source Tracking

- There are microbes that are only associated with a given fecal source
  - Host and gut microbes co-evolve
    - Physiological difference of the gut
    - Dietary difference between hosts

- MST provides a set of methods to identify sources of contamination
  - Genetic testing of host fecal markers

(Image: Source Molecular Corp.)

MST Applications

- Site assessment: How bad or good is this site?
  - Among all sites within your jurisdiction?
  - Compared to a reference site with little human activities?
  - Compared to a site with measured health risk via epidemiology studies?
  - Before and after implementing BMP remediation actions?

- Answers should be based on data, using “scientifically sound and statistically defensible approaches”
  - Study design
  - Lab analysis
  - Data interpretation

The Process

Sampling

sample	Cq	Copy per 100ml
1	36.13	?
1	37.41	?
1	36.05	?
2	Non detect	?
2	Non detect	?
2	Non detect	?
3	Non detect	?
3	Non detect	?
3	Non detect	?
4	30.48	19173
4	30.50	18855
4	30.17	23356
…	…	?
…	…	?
…	…	?
n	Non detect	?
n	Non detect	?
n	Non detect	?

qPCR raw data (marker concentrations) from n samples

Lab analysis

Data interpretation
Site assessment

Action at the site

(Image: Yiping Cao, Donna Ferguson, Google)

The Practice

- Best professional judgement
  - Different experts in different projects
- Worries
  - Unintentional bias: inherent subjectivity and implementation variability by experts?
  - Intentional bias: hired gun by discharger or regulator?

Sampling

Lab analysis

Data Interpretation

The BPJ Exercise

- Assess variability in MST data interpretation

Create a simulated data set
(26 site, 20 sample/site)

Ten experts rank the sites 1 to 26 regarding relative levels of human fecal contamination

- Experts: research scientists and water quality managers
  - from the federal government
  - a public research agency
  - academic
  - a wastewater treatment agency

- Assessment: Compare ranks among pairs of experts via correlation

(Cao et al 2014)

The BPJ Exercise

- Assess variability in MST data interpretation

Two iterations

- 1st iteration: no prior discussion among experts

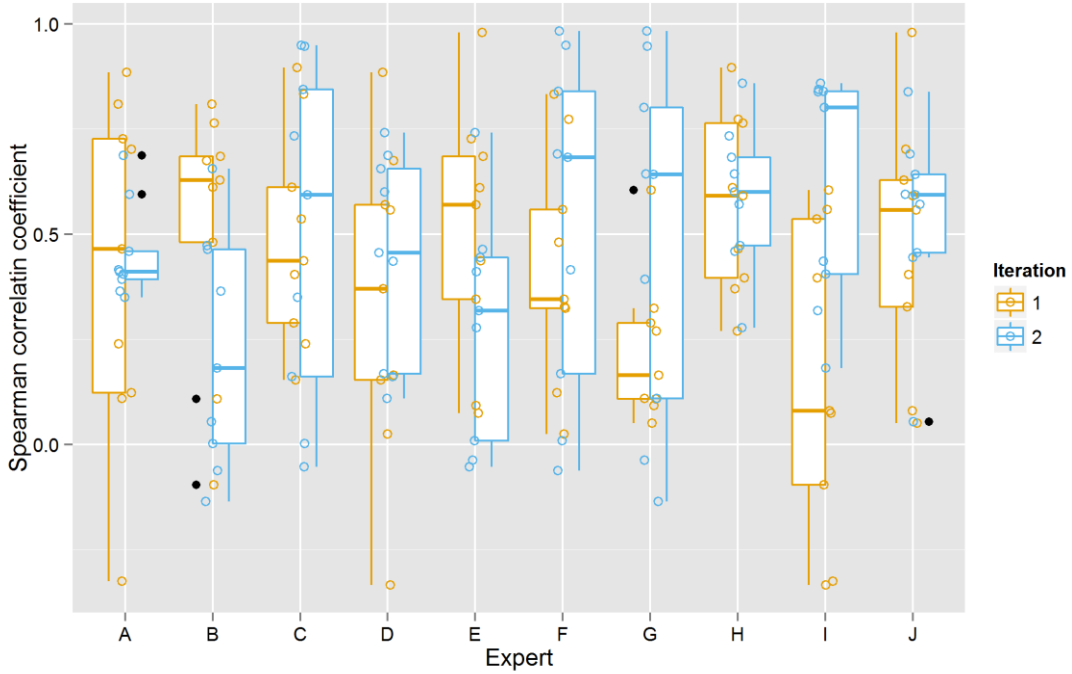
  r= - 0.33 to 0.98 (avg: 0.41)

- 2nd iteration: experts agreed to a set of principles before ranking

  r= - 0.14 to 0.98 (avg: 0.47)

(Cao et al 2014)

BPJ Interpretation Highly Inconsistent



- Experts' interpretation of the same data were highly variable
  - 1st iteration: r= - 0.33 to 0.98 (avg: 0.41)
  - 2nd iteration: r= - 0.14 to 0.98 (avg: 0.47)

So, how well does BPJ work? – not so well
Are we right to worry? - yes

(Cao et al 2014)

Motivation for Human Fecal Score

- BPJ exercise conclusion: a standardized mathematically defined objective approach is needed!

- Team:
  - SCCWRP: Drs. Yiping Cao, John Griffith, Steve Weisberg
  - USEPA: Drs. Orin Shanks, Mano Sivaganesan, Catherine Kelty
  - Stanford: Drs. Ali Boehm, Dan Wang

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

A human fecal contamination score for ranking recreational sites using the HF183/BacR287 quantitative real-time PCR method

(Cao et al 2018)

Human Fecal Score (HFS): Simple

- Simple
  - Site average concentration of HF183 marker
  - One number to characterize the extent of human fecal pollution at a site

Human Fecal Score

=

55

(n samples, 3n data points, for some we don't even have a number for)

sample	Cq	Copy per 100ml
1	36.13	?
1	37.41	?
1	36.05	?
2	Non detect	?
2	Non detect	?
2	Non detect	?
3	Non detect	?
3	Non detect	?
3	Non detect	?
4	30.48	19173
4	30.50	18855
4	30.17	23356
…	…	?
…	…	?
…	…	?
n	Non detect	?
n	Non detect	?
n	Non detect	?

HFS: Complete

- Uses all data
  - non-detect
  - detected but not quantifiable
  - Quantifiable

Human Fecal Score

=

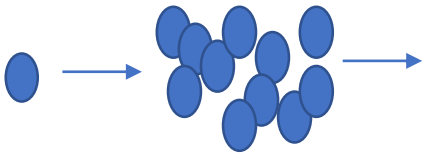
55

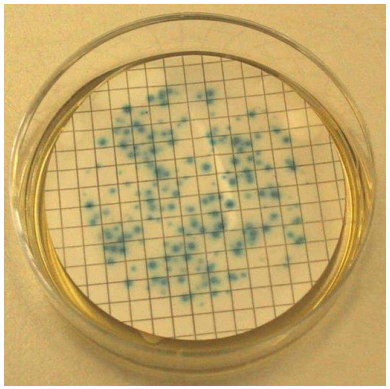
(Can't average non-number, e.g. ?)

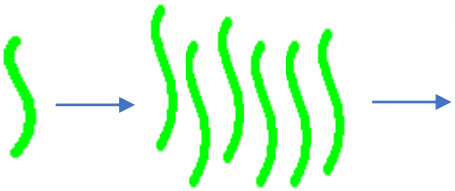
sample	Cq	Copy per 100ml
1	36.13	?
1	37.41	?
1	36.05	?
2	Non detect	?
2	Non detect	?
2	Non detect	?
3	Non detect	?
3	Non detect	?
3	Non detect	?
4	30.48	19173
4	30.50	18855
4	30.17	23356
…	…	?
…	…	?
…	…	?
n	Non detect	?
n	Non detect	?
n	Non detect	?

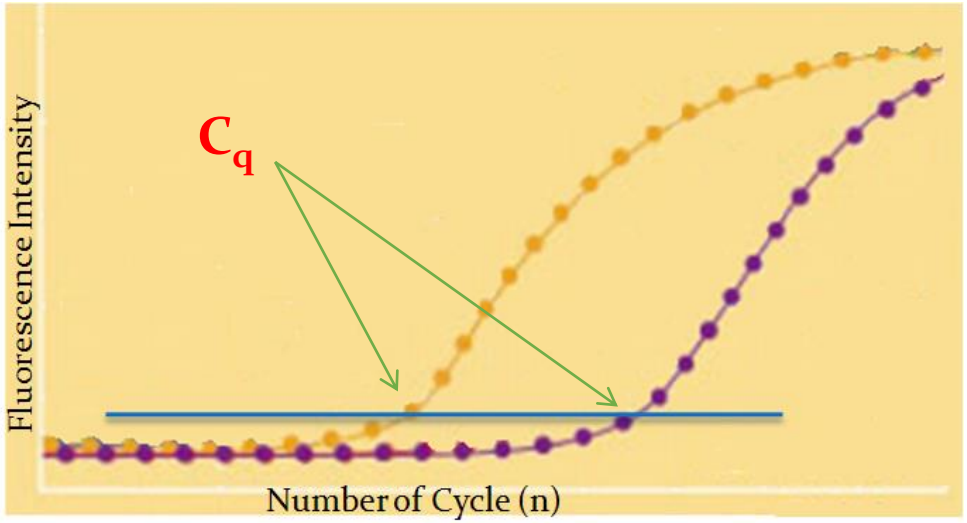
How do qPCR quantify?

Grow cells vs. “grow” DNA



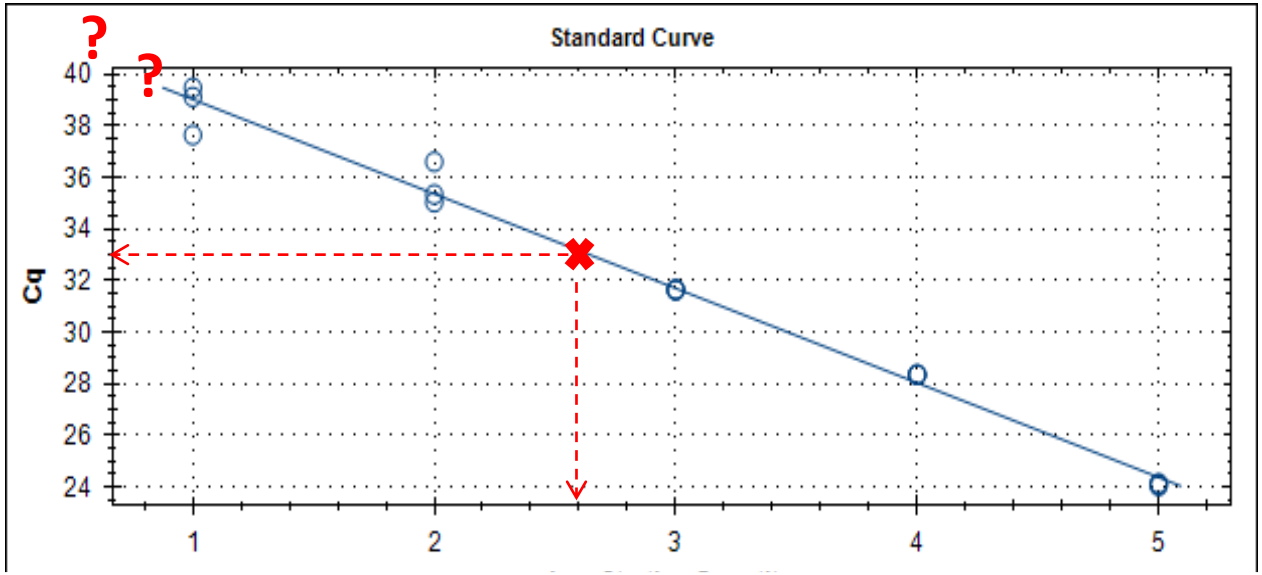






qPCR Range of Quantification (ROQ)

- Within Range: Cq linearly inversely relates to marker concentration
- At low concentration: no more linear relationship
  - Can't quantify using the standard curve
- Non-detect: no quantification



Log Marker Concentration

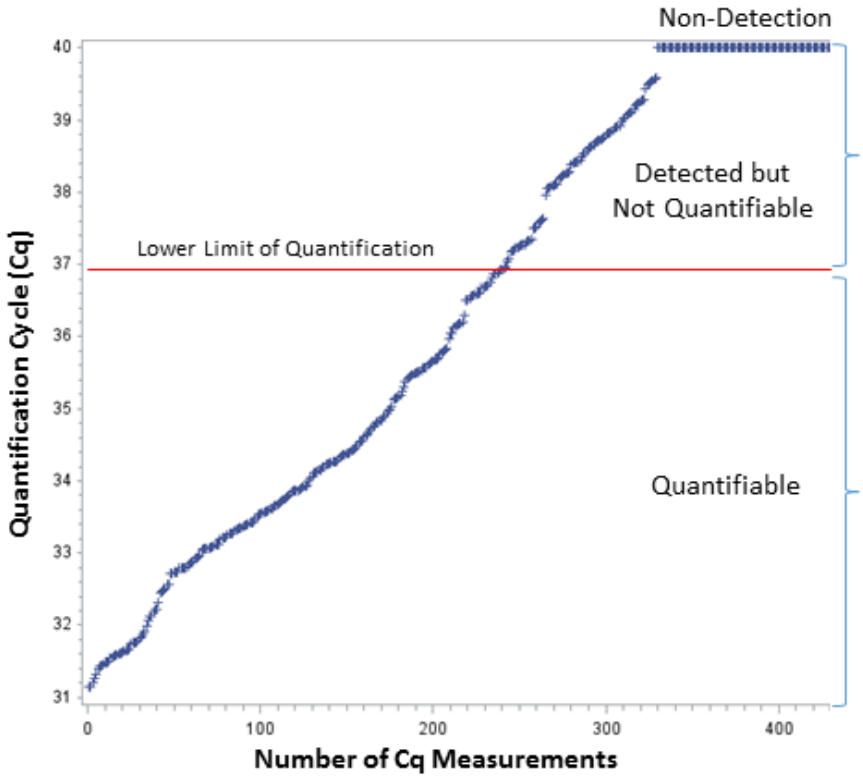
- Previous “solutions”
  - Ignore non-detect and detected but not quantifiable
  - Arbitrarily assign a number
    - 0, DL/2, DL, LLOQ …
  - Force standard curve outside ROQ
  - Statistics for censored data
    - Not applicable in most cases
- HFS: use underlying Poisson distribution to estimate ?’s outside ROQ

(Can’t average non-number, e.g. ?)

sample	Cq	Copy per 100ml
1	36.13	?
1	37.41	?
1	36.05	?
2	Non detect	?
2	Non detect	?
2	Non detect	?
3	Non detect	?
3	Non detect	?
3	Non detect	?
4	30.48	19173
4	30.50	18855
4	30.17	23356
…	…	?
…	…	?
…	…	?
n	Non detect	?
n	Non detect	?
n	Non detect	?

HFS: Based on statistics

- Two different quantification mechanisms
  - Executed by Bayesian models, integrating data uncertainty



MPN Range
Quantification by MPN approach

ROQ Range
Quantification by calibration model

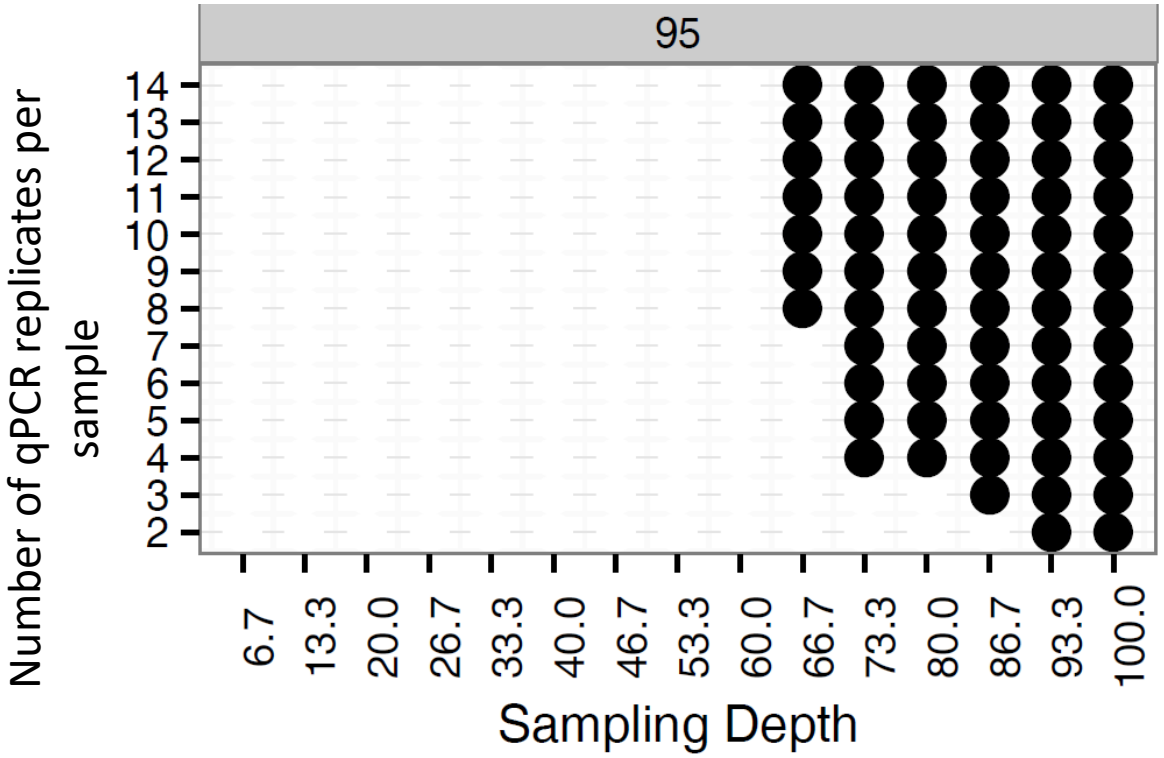
Human Fecal Score
(weighted average)

$$\mathrm{Log_{10}\,HFS} = W \cdot \mathrm{Log_{10}}\,C_1 + (1-W) \cdot \mathrm{Log_{10}}\,C_2$$

(Cao et al 2018)

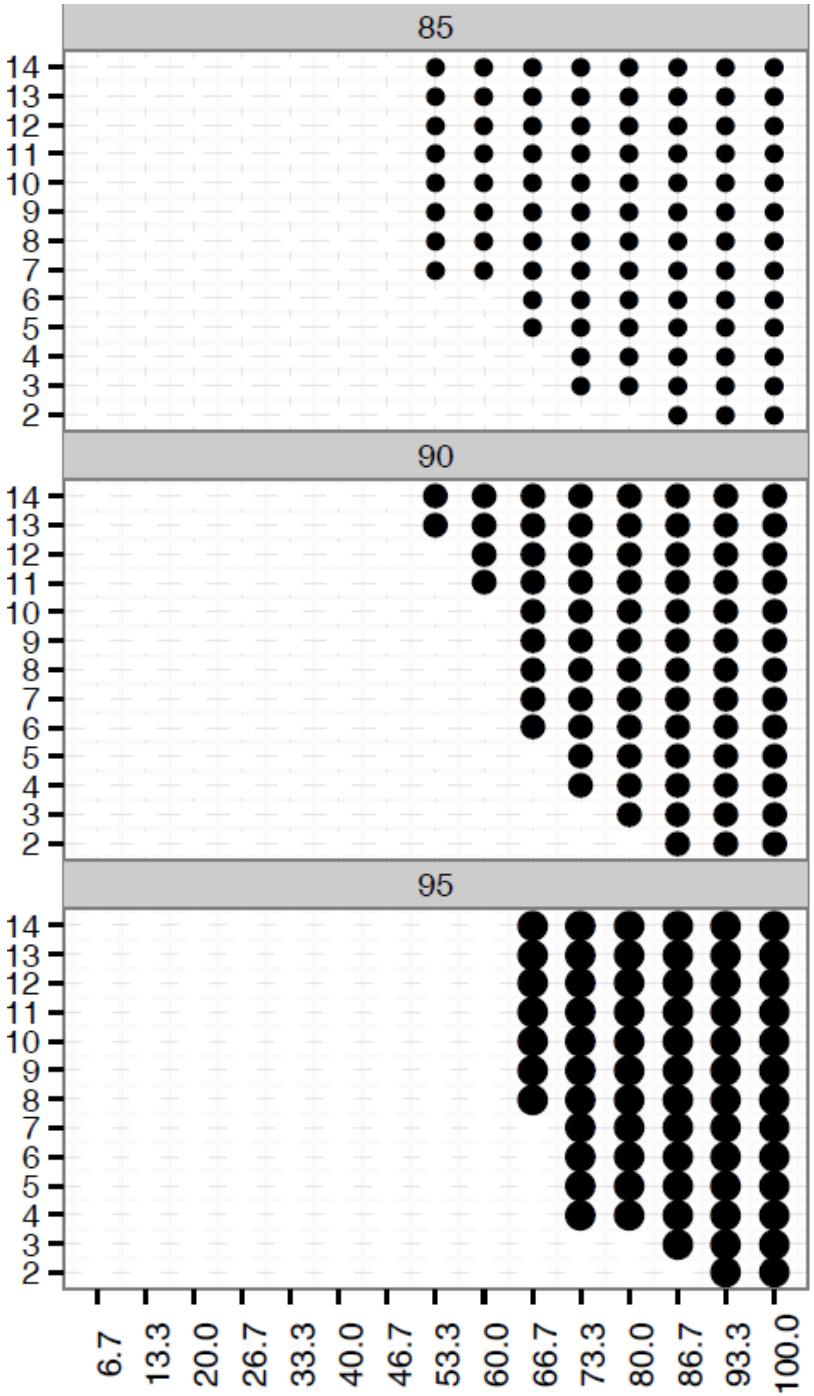
HFS allows sampling design optimization

- Certainty accepted by managers and/or regulators
- Trade-off between sample size and qPCR replication



(Cao et al 2018)

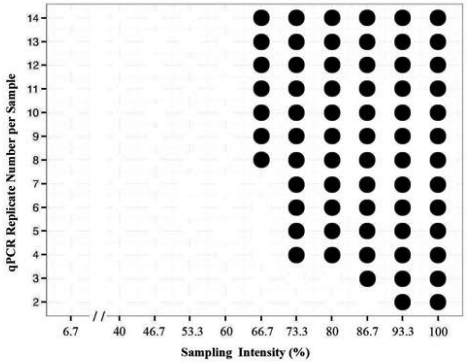
Willing to accept different chances of getting the right answer?



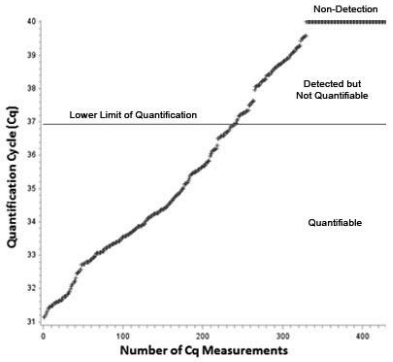
HFS Application: Prioritizing Remediation

HUMAN FECAL SCORE FOR SITE RANKING

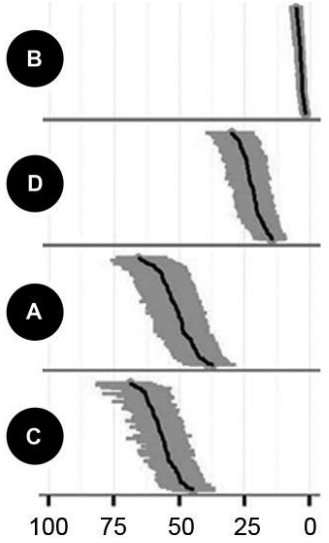
Polluted Sites



Customized Field Sampling and qPCR Solutions



All qPCR Data Included

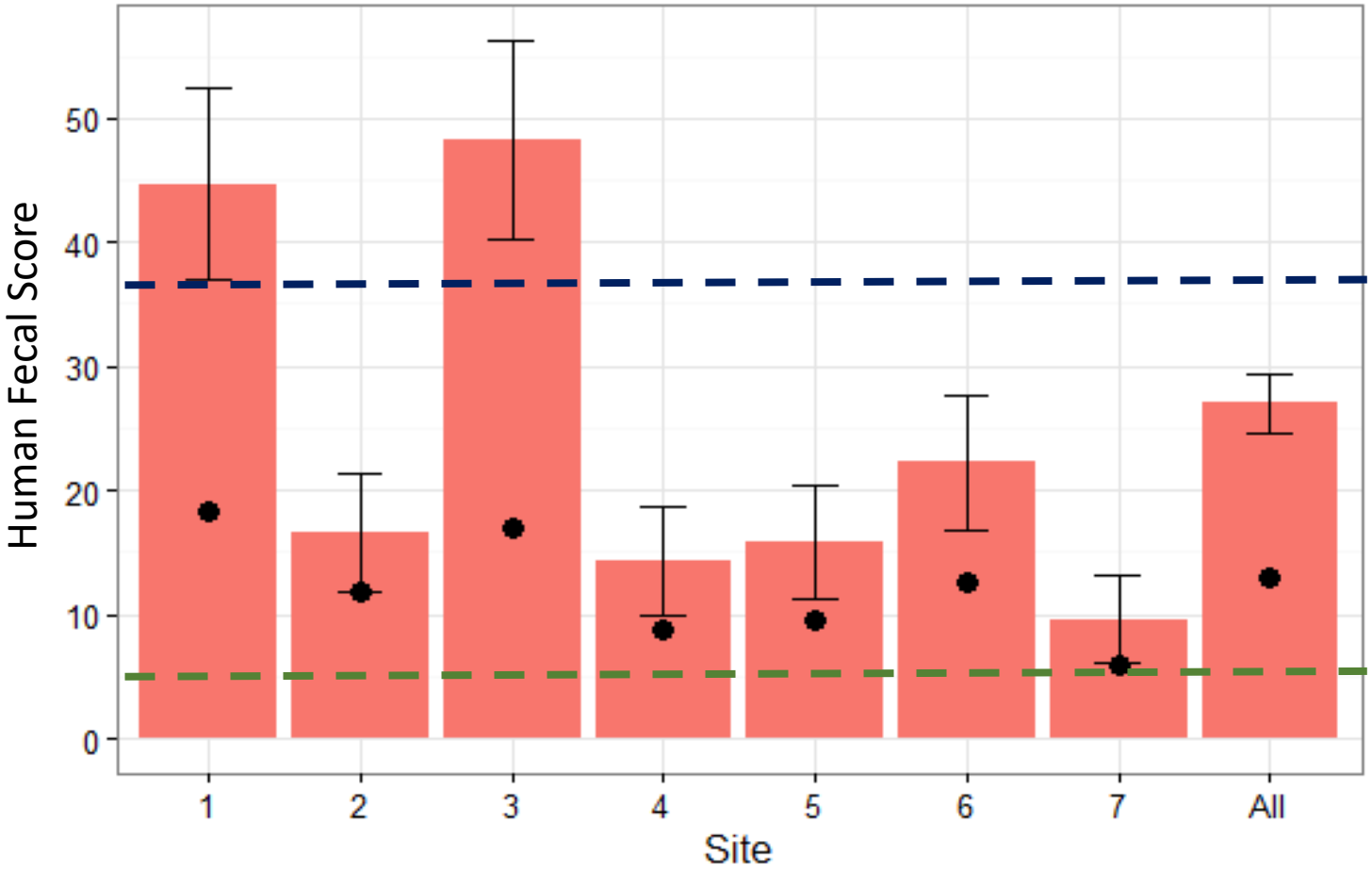


Human Fecal Score
(Copies per 100mL 95% BCI)

STANDARDIZED PROCEDURE

(Figure: adapted from Cao et al 2018)

HFS Application: Real Site Scores



HFS Summary

- Simple
- Respect data
  - Use everything
  - Add nothing
  - Respect underlying data distribution
  - Integrate uncertainty in data
- Objective
  - Mathematically defined
  - Build on formulas instead of narratives
- Standardization
  - Use the U.S. EPA standard HF183 qPCR method
  - Sampling design

Sampling

Lab analysis

Data Interpretation

Implications for water quality management

- HFS describes a standardized method for characterizing human fecal pollution level at a site

- General:
  - Other markers: Cow Fecal Score, Gull Fecal Score
  - Other technology: digital PCR

- Potential applications
  - BMP effectiveness
  - Rank sites
  - CSO consent decree compliance
  - QMRA site eligibility

Thank you!

Yiping Cao
Vice President of Technology
ycao@sourcemolecular.com

Source Molecular

ilac-MRA
ACCREDITED
Certificate 4126.01
ISO 17025 Accredited
Testing Laboratory