CENTER for PRODUCE SAFETY

Center for Produce Safety 2009 RFP
Final Report due October 31, 2011

Project Title

Epidemiologic analysis and risk management practices for reducing *E. coli* in irrigation source water supplies and distribution systems

Project Period

October 1, 2009 – September 30, 2011

Principal Investigator

Rob Atwill, D.V.M., Ph.D.
Western Institute for Food Safety & Security
School of Veterinary Medicine
University of California, Davis
Email: ratwill@ucdavis.edu
Phone: 530-754-2154

Co-Principal Investigators

Kenneth W. Tate
Department of Plant Sciences, University of California, Davis
kwtate@ucdavis.edu, 530-754-8988

Michele Jay-Russell
Western Institute for Food Safety & Security, University of California, Davis
mjay@ucdavis.edu, 530-757-5756

Linda Harris
Western Institute for Food Safety & Security and Dept. of Food Science & Technology
University of California, Davis
ljharris@ucdavis.edu, 530-754-9485

Objectives

1. Working in close collaboration with the California and Arizona LG produce industry and allied organizations, finalize the master data file for statistical and epidemiological analyses of objectives 2 through 5.
2. Determine environmental, geographical, structural and operational risk factors for the occurrence of generic *E. coli* in irrigation water supplies. We will also determine the influence of different diagnostic methods on measured *E. coli* levels.
3. Identify predisposing environmental, structural, and operational risk factors associated with generic *E. coli* exceedances in irrigation water supplies.
4. Determine the ability of difference mitigation measures to reduce the reoccurrence of an *E. coli* exceedance in irrigation water supplies.
5. Develop more efficient irrigation water sampling plans for low- to high-*E. coli* risk source water supplies.

FINAL REPORT

Background

Our aim was to conduct a detailed analysis of irrigation water monitoring data currently being collected by the produce industry in order to characterize the occurrence of generic *E. coli* (herein referred to as *E. coli*) for different regions within California, different sources of water, and across all seasons. An additional aim was to identify risk management practices and environmental factors associated with *E. coli* in irrigation water supplies. Finally, at the request of various growers, we evaluated the benefits and risks of changing the current sampling rate and/or volume of water being tested per sample as an alternative monitoring strategy that would better match the sampling effort to the risk level of the water supply (from low to high risk of *E. coli*) and maximize the ability of growers to detect occurrences of *E. coli*.

Research Methods and Results

Two datasets were compiled from data provided by members of the produce industry (growers, processors, laboratories), with each dataset representing numerous growers from throughout California from various sources of irrigation water. For the first dataset (n=44,249, herein referred to as Set 1), we established three tiers of analysis based on the thoroughness of the information about each water sample and our confidence in the information provided. Tier 1 included all water samples with at least a minimum of basic information about the sample (date, city, water source). Tier 2 was a subset of tier 1 data, tier 3 was a subset of tier 2, with tier 3 having the most complete information about the water sample.

A second dataset (n=15,486, herein referred to as Set 2) was generated by targeting a different cross-section of the produce industry, but the possibility existed that some water samples were present in both Set 1 and Set 2 due to confidentiality of the data (i.e., when grower location was not revealed). Similar to Set 1, these data were from across California growing regions covering a wide range of farming operations, water sources and locations. Analysis of these data show similar trends in *E. coli* occurrence as the Set1.

All data were further categorized into four growing regions (Figure C1, Appendix C) and four water sources (well, canal, reservoir, other). Using logistic and mixed-effects Poisson regression, the association between *E. coli* and the above mentioned variables (date, region, water source), along with environmental variables such as air and soil temperature, wind speed, precipitation, groundwater depth, were examined. Mixed effects models were used due to the issue of repeated sampling of specific irrigation water locations and the possibility of non-independent data. Statistical significance was determined using a forward stepping algorithm, with p-value ≤ 0.05 for inclusion of the variable in the final regression models.

SET 1 DATA

The overall mean concentration of *E. coli* was similar between tier one and two datasets, but tier three, which had the highest data quality, also had significantly higher concentrations of *E. coli* compared to tiers one and two (Table 1).

Table 1. Mean *E. coli* concentrations (MPN/100mL) between three tiers of analysis.

Tier	No. samples	Mean	Standard Deviation
1	44,249	11.41[a]	109.43
2	41,083	11.61[a]	109.61
3	17,788	14.42[b]	121.78

[a] Mean *E. coli* not significantly different, $p>0.05$; [b] Mean *E. coli* in tier 3 greater than tier 1 or tier 2, $p<0.01$

TIER 1 ANALYSES

The majority of water samples (79% (35,093/44,249)) contained no detectable *E. coli*, and equally important, only 0.86% (380/44,249) of water samples exceeded the single sample maximum (SSM) of >235 MPN/100mL for foliar application of irrigation water, and less than 0.43% of samples exceeded the SSM for non-foliar application of >576 MPN/100mL (Fig 2). This indicates that for tier 1 data the occurrence of an exceedance in California irrigation water supplies was very rare between the years of February 2007 through September 2010. It is important to note that this low occurrence of exceedances was derived from industry data representing a large number of growers from throughout produce growing regions of California, and includes various water sources and all four seasons.

The prevalence of water samples with any level of detectable *E. coli* (MPN ≥1 /100mL) varied between water sources (Table 2). About 8% of well samples had detectable *E. coli* compared to 86% and 48% of canal and reservoir samples, respectively. The occurrence of *E. coli* varied significantly between regions, with higher concentrations of *E. coli* in the Desert region compared to the South or North Central Coast ($p<0.01$). There were higher concentrations in the South Central Coast compared to the North Central Coast ($p<0.01$). *E. coli* concentrations in the Central Valley did not vary significantly from any other region ($p>0.05$).

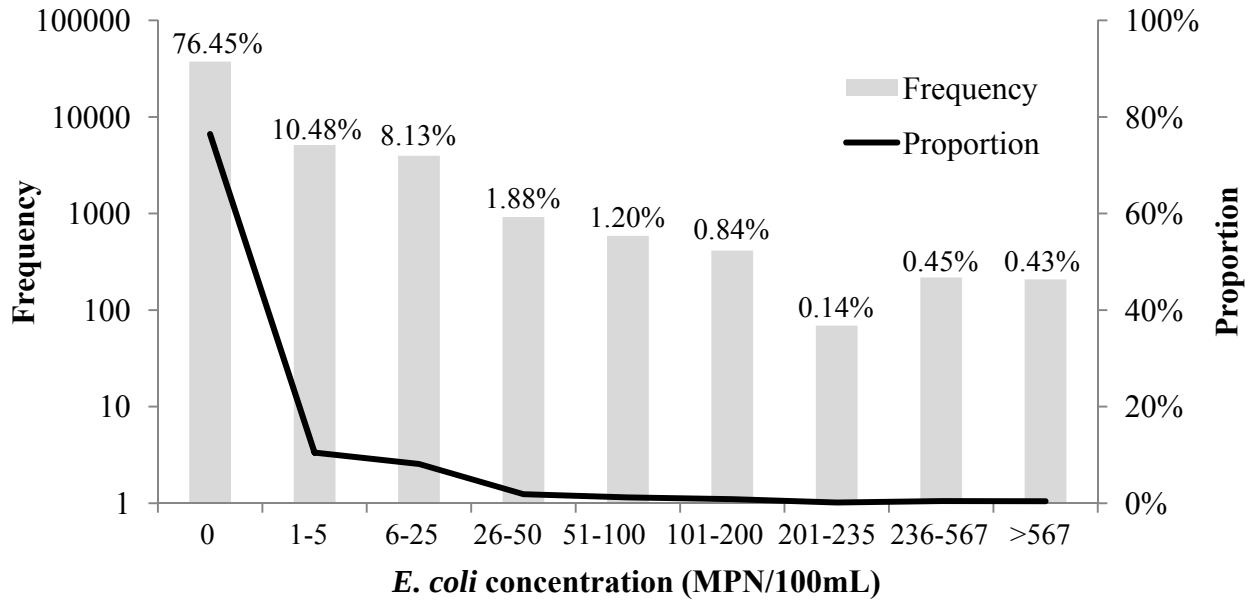


Figure 1. Frequency distribution of *E. coli* concentrations for all regions and water sources.

Table 2. Mean *E. coli* concentrations across regions by water source.[†]

Water Source		Mean *E. coli* (MPN/100mL)	No. samples
N. Central Coast		**6.75**[a,b,c]	**23,919**
	Well	3.70	20,052
	Reservoir	55.25	1,529
	Other	1.13	2,329
S. Central Coast		**11.90**[a,b,c]	**12,799**
	Well	6.53	10,272
	Reservoir	109.29	588
	Other	10.15	1,928
Central Valley		**15.34**[c]	**454**
	Well	18.76	246
	Canal	21.76	76
	Other	5.27	125
Desert		**26.02**[a,c]	**7,077**
	Well	16.67	473
	Canal	30.03	5,413
	Reservoir	10.92	1,125
	Other	21.59	66
Grand Total		**11.41**	**44,249**

[†] Sources of water from each region with less than 20 water samples are not shown due to industry concerns about the representativeness of the data (canal samples from NCC and SCC; reservoirs in the CV).
[a] Significant difference between Desert and Central Coast regions, $p<0.01$.
[b] Significant between Central Coast regions, $p<0.01$.
[c] Significant between Central Valley and other regions, $p>0.05$.

TIER 2 ANALYSES

E. coli concentrations (MPN/100mL) varied significantly across season for most of the regions. Levels of *E. coli* were highest in the two Central Coast regions during the fall season (Sept-Nov) ($p<0.01$), while concentrations were highest in the Desert ($p<0.05$) region during the summer (June-Aug) (Table 3). We did not detect significant seasonal differences in *E. coli* concentration in the Central Valley, in part due to the smaller sample sizes from that region.

Combining all the data for tier 2, the occurrence of SSM exceedances for foliar application ranged from 0.45% to 1.65% across season, with highest proportion of exceedances occurring in summer and especially during the fall (Table 4). Nonetheless, similar to tier one results, SSM exceedances appear to be uncommon during most of the year, with the exception of fall where about 1 in 60 water samples had >235 MPN/100mL.

Using logistic regression to identify factors associated with the odds of an exceedance, well water samples had about a 3.6 and 3.0 times higher odds of exceedance during summer and fall, respectively, compared to exceedances in winter (Table C1, Appendix C). The odds of an exceedance in reservoir water samples were 6.6 and 15 times higher during summer and fall, respectively, compared to exceedances in winter. The odds ratios for the seasonal patterns were just the opposite for canal water, whereby the odds of an exceedance was lower in summer and fall compared to winter, but when mean air temperature is added to the calculation the overall risk of an exceedance was much higher in summer than winter. The association between 24-hour air temperature, 24-hour wind speed and the odds of an exceedance were not consistent between water sources. Adjusted for season, mean air temperature was negatively associated with the odds of an exceedance

for reservoirs. If there is increased turnover of reservoir water during warmer days due to higher crop demands for irrigation water, it is possible that the lower residence time for water in the reservoir resulted in lower *E. coli* contamination. In contrast, there was a positive association between mean air temperature and the odds of an exceedance for canal sources (from mostly desert region), which functioned to substantially increase the calculated risk of an exceedance during summer and fall season. Adjusted for season, wind speed was negatively associated with the odds of an exceedance for well water sources but not significant for reservoir and canal sources. If higher wind velocity occurring over several days can increase crop demands for irrigation water, then the increased use of the irrigation well will result in minimal residence time for irrigation water in the well infrastructure and thereby generate water representative of ambient groundwater conditions. In our experience groundwater from properly constructed wells has very low levels of *E. coli* for this region of California. These associations between environmental factors and the odds of exceedance are in the process of being more carefully analyzed to determine the best fitting regression model for these complex relationships between *E. coli*, source of water, and environmental factors.

Table 3. Mean *E. coli* concentrations across regions by season. †

Regions	Mean *E. coli* (MPN/100mL)	No. samples
N. Central Coast	**6.69**[a]	**24,233**
Winter	1.82	4,831
Spring	3.59	7,078
Summer	8.80	7,154
Fall	12.51	5,170
S. Central Coast	**13.62**[b]	**9,243**
Winter	5.14	1,787
Spring	6.28	2,225
Summer	15.41	3,025
Fall	25.81	2,206
Central Valley	**15.51**	**442**
Winter	0.93	80
Spring	21.91	116
Summer	29.63	118
Fall	5.81	128
Desert	**25.01**	**7,385**
Winter	20.66	2,475
Spring	21.25	998
Summer	36.82	866
Fall	29.59	3,046
Overall	**11.86**	**41,303**

† Winter = Dec-Feb, Spring = Mar-May, Summer = June-Aug, Fall = Sept-Nov.
[a] Significantly different from S. Central Coast and Desert, $p<0.01$
[b] Significantly different from Desert, $p<0.01$

Table 4. Distribution of SSM exceedances (*E. coli* >235 MPN/100mL) across seasons.

Season	No. samples	%
Winter		
<235	9,038	99.55
≥235	41	0.45
Spring		
<235	10,316	99.55
≥235	47	0.45
Summer		
<235	11,032	99.07
≥235	104[a]	0.93
Fall		
<235	10,332	98.35
≥235	173[b]	1.65

[a] Significantly higher than winter and spring, $p<0.01$
[b] Significantly higher than all other seasons, $p<0.01$

TIER 3 ANALYSES

Average concentrations of *E. coli* were significantly higher for data used in Tier 3 analyses than in the previous two tiers (Table 1). Models for the occurrence and exceedance of *E. coli* were similar between all three levels of analysis, though coefficients varied slightly. As with other tiers of analysis, exceedances were rare, with 1.1% of all 17,788 samples >235 MPN/100mL. Concentrations of *E. coli* were significantly greater in reservoirs than water taken from wells on the same property in the Central Coast regions ($p<0.01$) (Table 5). The odds of an exceedance for foliar application was about three times higher for water taken from a reservoir compared to water taken from a nearby well on the same ranch, and also for water sampled during summer and fall compared to winter (Table 6). Twenty-four hour cummulative precipitation was positively associated with the likelihood of an exceedance in wells and reservoirs ($p<0.05$) (Figure C2, Appendix C). Hence, risk of an exceedance is higher during conditions of heavy rain fall during the winter and spring along the Central Coast regions (Table 7, Appendix C), yet overall the risk of an exceedance is much lower than during summer and fall (Figure C2, Appendix C). Similar assocations were found for the occurrence of any *E. coli* >0 MPN/100mL, except that the odds of detecting *E. coli* was negatively associated with windspeed for these pairs of wells and reservoirs (Table C2, Appendix C).

Table 5. Comparison of average *E. coli* concentrations between paired well and reservoir samples across seasons in North and South Central Coast regions.

	Mean *E. coli* concentrations (MPN/100mL)			
Seasons	**Well**	**Reservoir**	**Average Difference**	**% Increase**
Winter	1.0	18.5	17.5	1850
Spring	9.8	21.5	11.7	220
Summer	19.4	77.6	58.2	400
Fall	20.8	65.4	44.6	315
Overall Average	**13.9**	**50.6[a]**	**36.7**	**365**

[a] Significantly greater, $p<0.01$

Table 6. The odds of detecting an *E. coli* exceedance (MPN/100mL >235) in well-reservoir pairs in North and South Central Coast regions as a function of cumulative 24-hr precipitation and season.

Covariates		Odds Ratio	p-value	95% Confidence Interval	
Water source	Well[a]	1.00	--	--	--
	Reservoir	2.86	0.001	1.52	5.37
Season	Winter [a]	1.00	--	--	--
	Spring	0.68	0.577	0.18	2.60
	Summer	2.84	0.045	1.02	7.90
	Fall	2.91	0.045	1.03	8.25
Cum. 24 hr. precip (mm)		1.15	0.030	1.01	1.30

[a] Referent category, hence OR=1.0.

Table 7. Average 24-hour cumulative precipitation (mm) in the North and Central Coast regions across all seasons.

Season	Cumulative precipitation (mm)			
	Mean	Std. Dev.	Min	Max
Winter	0.97	3.08	0.00	22.30
Spring	0.23	1.12	0.00	13.90
Summer	0.01	0.12	0.00	2.60
Fall	0.07	0.50	0.00	6.50

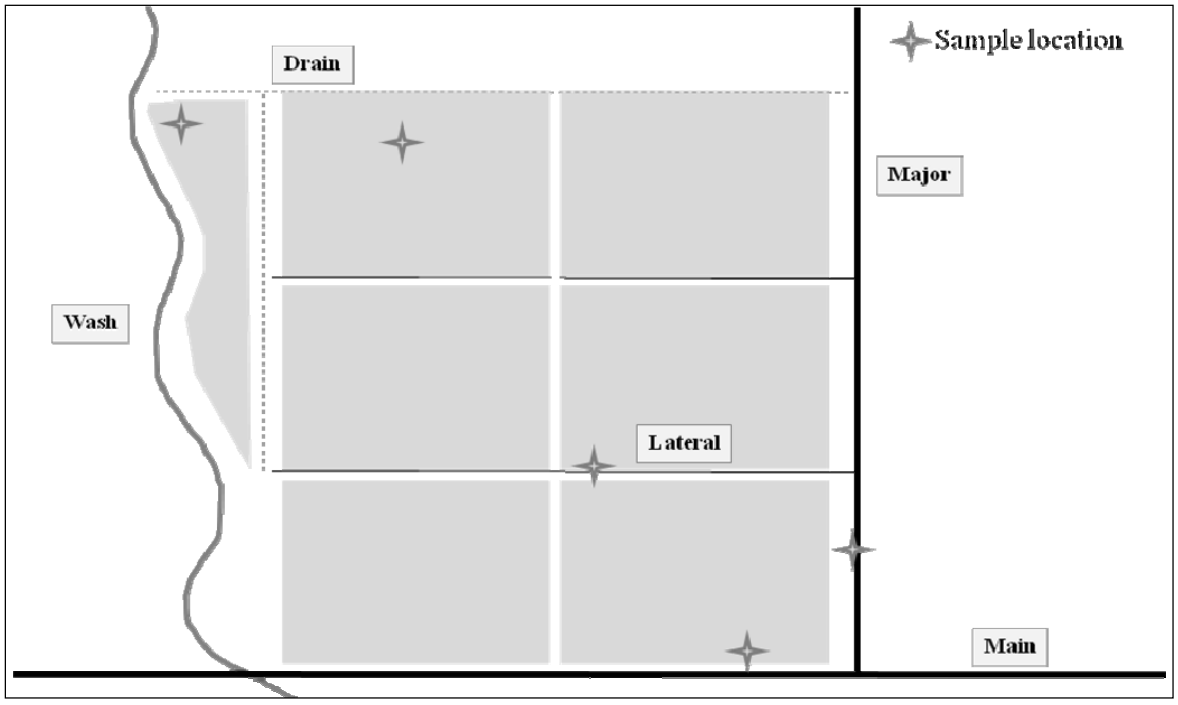


Figure 2. Schematic showing the theoretical location of main, major, lateral, drian, and wash. Water samples were taken adjacent to these locations based on the information provided in the confidential dataset. The schematic is not intended to indicate that every field has a wash or drain adjacent to it; rather the figure indicates where the various locations could be if present.

IRRIGATION CANAL ANALYSES WITHIN TIER 3

Figure 2 above is a simplified schematic to help those readers who are unfamiliar with irrigation systems to better understand what is meant by a main or major canal, a lateral canal or drain, and to show where a wash might theorctically be located when it occurs in a real-world setting. Samples, as indicated for the Tier 3 data, were designated with these various terms (main, major, etc) based on the data provided in the dataset. For example, a sample that was taken nearest to the drainage end of the field would be considered nearer to the drain than to the major canal at the opposite end of the field, even though the canal was the source of water.

The odds of an *E. coli* exceedance (>235 MPN/100mL) was about 12 times greater in summer compared to winter (Table 8). Interestingly, this higher risk of an exceedance occurred during the time of year when leafy green produce is for the most part not grown; instead, the growing season in the desert regions of California occurs primarily during fall through spring which is a time when the risk of an exceedance was much lower. In other words, microbial water quality was at its worst during a time when leafy green produce was not grown, and water quality was at its highest during the growing season. The odds of an exceedance was also associated with sample location along the irrigation canal network. For example, samples taken near major or lateral canals were about 40 to 70% less likely to have an exceedance than those samples taken near a main canal. In contrast to risk factors for the odds of an exceedance, the odds of having any *E. coli* in a water sample were higher for locations other than the main canal. For example, the odds of detecting any *E. coli* from a lateral canal was nearly 10 times higher than water from the main canal (Table C3, Appendix C). Feedlot density was not significantly associated with the odds of an exceedance (Table 8), but was associated with the occurrence of any *E. coli* in a sample (Table C3, Appendix C).

Table 8. The odds of an *E. coli* exceedance (>235 MPN/100mL) in canal water associated with different locations along the irrigation canal network, adjusted for feedlot density and season.

Covariates		Odds Ratio	p-value	95% Confidence Interval	
Season	Winter [a]	1.00	--	--	--
	Spring	1.90	0.65	0.116	31.07
	Summer	11.91	0.03	1.345	105.52
	Fall	4.86	0.13	0.622	37.93
Nearest Canal Network Part[†]					
	Main Canal[b]	1.00	--	--	--
	Major Canal	0.43	0.44	0.050	3.70
	Lateral Canal	0.29	0.26	0.033	2.48
	Drain	0.66	0.78	0.038	11.67
Feedlot Count w/in 5 Km[†]		0.90	0.59	0.600	1.33

[a] Winter season set as the referent category, hence OR=1.0; [b] Main canal as referent category, hence OR=1.0
[†] These variables were not significant in the exceedance model but were retained to allow comparison with the occurrence model (Table C3).

SET 2 DATA

Data comprising Set 2 (n=15,486) were from throughout California from different growing regions and from all four seasons of the year. The overall mean *E. coli* concentration was slightly lower in Set 2 data (9.4 MPN/100mL) than in Set 1 data (11.4 MPN/100mL). While a greater proportion of samples from Set 2 had detectable levels of *E. coli* (27%) compared to Set 1 (21%), only 0.71% of the samples exceeded the SSM of >235 MPN/100mL for foliar application (Figure 3). In addition, only 0.19% of samples exceeded the SSM for non-foliar application of >576 MPN/100mL. Concentrations of *E. coli* were greatest in the summer and fall ($p<0.01$) for the entire data set (Table C4, Appendix C).

While exceedances for foliar application were rare in well samples (~0.2%) and non-existent in reservoir samples, they were more frequent in canal samples (2.5%, $p<0.01$). Exceedances were more common in the summer and fall (Table 9). Exceedances were not significantly more likely in samples where the point of entry (POE) was a gate (canal water) or sprinkler head (all water types) than at a valve ($p>0.05$). The proportion of water samples that exceeded the SSM for foliar application ranged from 0% to 2.8% depending on the POE (Table C5, Appendix C).

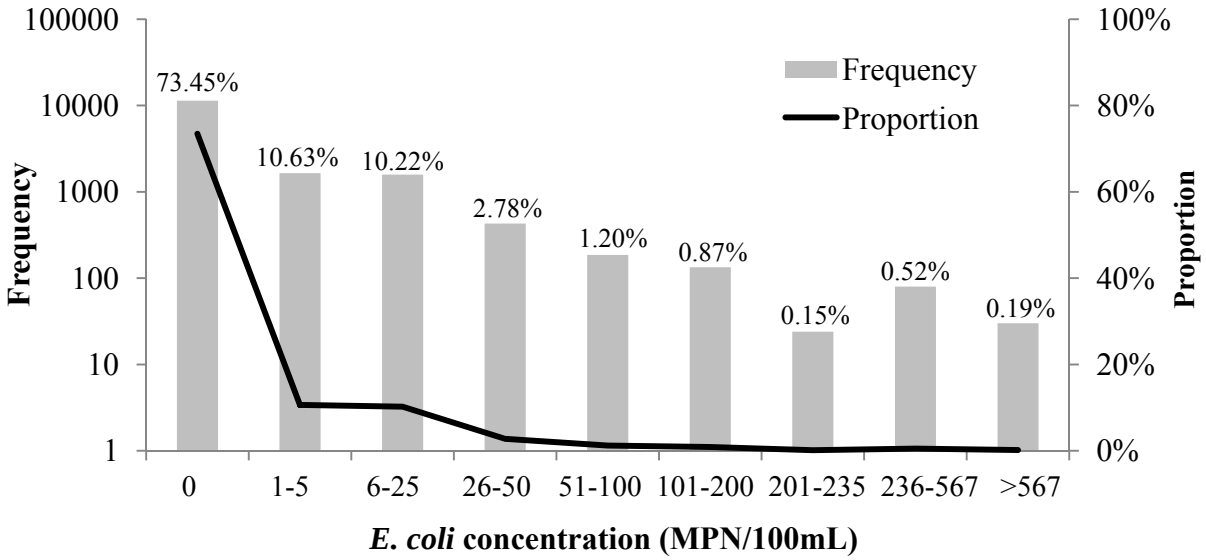


Figure 3. Frequency distribution of *E. coli* (MPN/100mL) across all regions and seasons for Set 2 data.

Table 9. Odds of detecting an *E. coli* exceedance (>235 MPN/100mL) associated with season, sample point of entry (POE) and growing region.

Covariates		Odds Ratio	p-value	95% Confidence interval	
Season	Winter [a]	1.00	--	--	--
	Spring	3.19	0.043	1.04	9.83
	Summer	10.86	0.001	4.49	26.26
	Fall	5.57	0.001	2.39	12.95
POE	Valve [b]	1.00	--	--	--
	Gate	1.68	0.541	0.32	8.90
	Sprinkler	1.86	0.458	0.36	9.61
	Other	12.88	0.017	1.57	105.89
Region	N. Central Coast [c]	1.00	--	--	--
	S. Central Coast	3.77	0.286	0.33	43.17
	Central Valley	11.27	0.039	1.13	112.80
	Desert	59.20	0.001	15.50	226.14

[a] Winter set as referent category, hence OR=1.0
[b] Valve as referent, hence OR=1.0
[c] N. Central Coast as referent, hence OR=1.0

Occurrence of *E. coli* follows a similar pattern to previous analyses: the odds of detecting any *E. coli* in a sample was twice as likely in the fall compared to the winter (Table C6, Appendix C). The occurrence of *E. coli* varied between different points of entry when pooled across all water types and region (Figure 4, Table C6). The POE with higher proportion of positive samples were gate, reservoir, furrow, and pipe (Figure 4), but undoubtedly these bacterial levels are being driven by a multitude of factors besides POE. Samples taken at a pipe or gate, as recorded in the database, were two to three times more likely to have detectable *E. coli* compared to samples taken at a valve (Table C6).

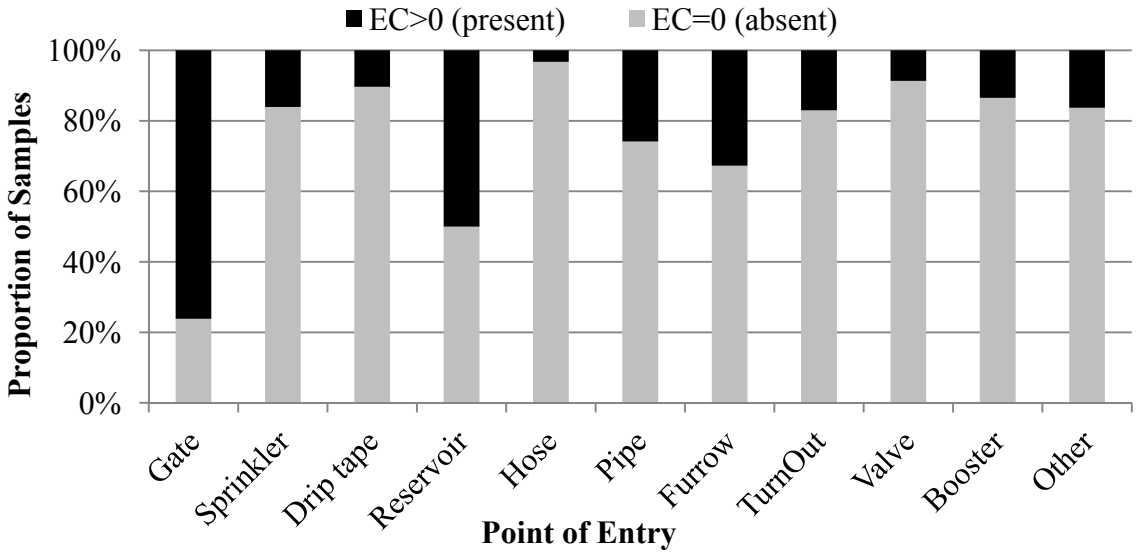


Figure 4. Proportion of samples with *E. coli* above detection limit (>0 MPN/100mL) across a variety of points of entry.

ANALYSIS OF IRRIGATION SAMPLING PLANS

As indicated in the above analyses, the overall concentration of *E. coli* and odds of an exceedance are associated with a large number of factors, resulting in a wide range of risk levels for this indicator bacterium. Growers have inquired whether the frequency of sampling can be modified to better match these risk levels. One question that has been posed is whether locations exhibiting a low risk of *E. coli* detection can be sampled less frequently under the assumption that a location's occurrence of *E. coli* are relatively stable over time. While it possible that sites with low risk of *E. coli* detection pose a similar low risk for pathogen occurrence (this is an unproven assumption), reductions in the sampling frequency will in general reduce the probability of correctly classifying a site's risk level for *E. coli*. In other words, accurately detecting rare events necessarily requires higher rates of sampling to find these rare events, while detecting common events is often easier due to their commonality. In addition, as the variability of *E. coli* counts increases, say during summer or fall when the odds of an exceedance increase significantly for many regions and water sources, a higher rate of sampling is typically needed to correctly classify a location's risk level. If the variability of counts of *E. coli* is low, in general fewer samples are needed to correctly classify a location's risk level. These general conclusions are supported by the analyses shown below.

A series of simulations were conducted to better understand the consequences of changes in sampling frequency, and to explore the possible benefits of increasing the sampling volume per assay to offset a reduced sampling frequency. To conduct these simulations, we created a series of risk levels to classify water quality (Table C7, Appendix C). We used concentration data from the North and South Central Coastal region and from the Desert region of California (Set 1, Tier 2 data set) to create occurrence scenarios and evaluate the effect of sampling frequency on the probability of correctly assigning a scenario to its risk category (1 to 7, Table C7). One thousand simulated seasons were used to calculate the detection probabilities shown in Figure 5 (right-hand panels). The left-hand panels in Figure 5 are regional

concentration profiles, generated by fitting a lognormal distribution to the historical data for a region's season and water source.

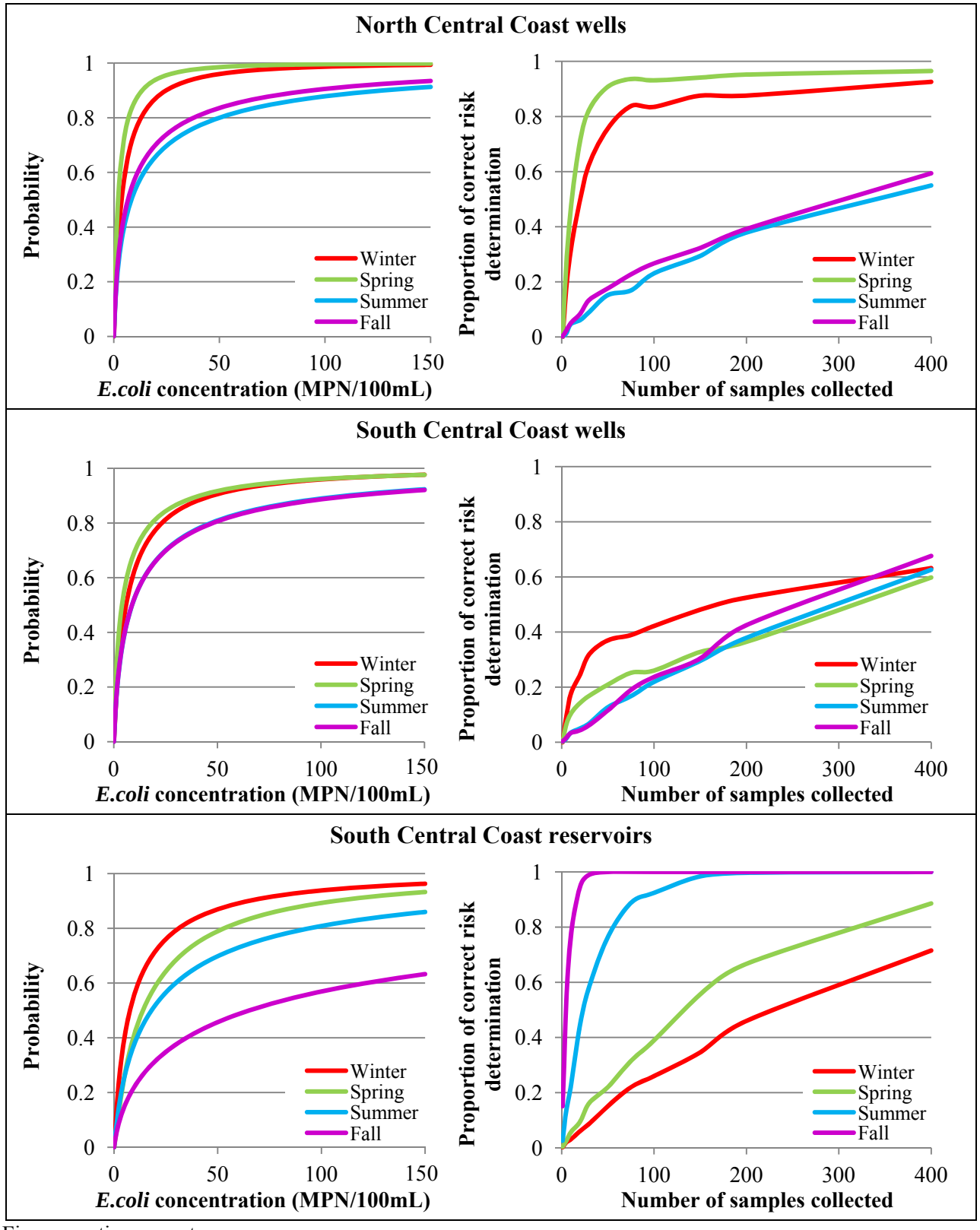


Figure continues next page.

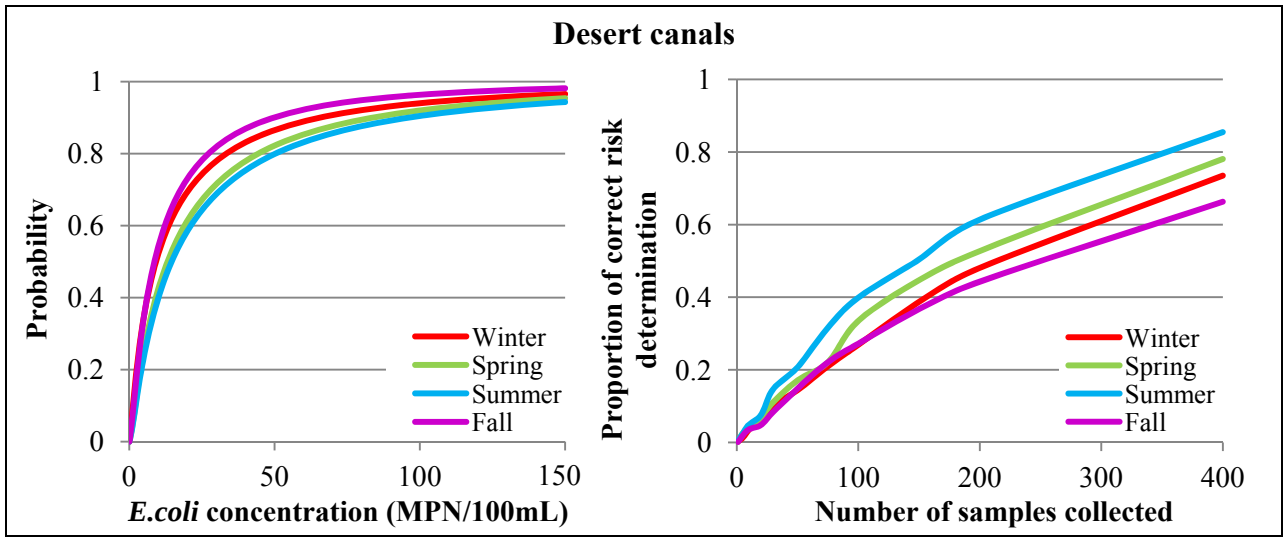


Figure 5. Probability of correctly classifying a location into its actual risk level (categories 1-7, Table C7) as a function of the number of samples collected over a season scenario. Left column: statistical distribution of historical data by region, water source, and season, fitted with a lognormal distribution. Right column: probability of correct water quality estimation, based on the sampling simulation. Lines may not appear smooth due to interpolation and simulation variability.

These simulations indicate that when the amount of *E. coli* across a season is clustered around a single concentration (i.e. when the variability in concentration is low), a relatively small number of samples are needed to correctly estimate water quality with high probability. For example, relative to other scenarios, fewer samples are needed during winter and spring for North Central Coast wells. Conversely, in scenarios with occasional higher concentrations, it is more difficult to capture these rare events with a sampling plan and thus a significantly higher number of samples are needed to correctly classify water quality into a risk category 1 through 7. The sampling frequencies shown in Figure 5 are quite high relative to the monthly sampling frequency being used by many growers, indicating that reducing the rate of water monitoring may substantially reduce the probability of a grower to correctly classify a location's water quality.

Rather than advocate for a higher rate of sampling per season than is currently being practiced by the produce industry, we explored the impact of increasing the sample volume per assay as an alternative to increasing the sampling frequency. Given the lack of data for 100mL grab samples taken across a single day for all regions, season, and sources of water, for the time being we assumed that the variability of *E. coli* in 100mL grab samples is the same within a day as across the season. This assumption may not be true for all regions, seasons, and sources of water; hence, results should be interpreted with caution until better data is made available to verify or modify these assumptions. The results in Figure C3 (Appendix C) show that a single 100mL grab sample has a low probability to detect existing levels of *E. coli*. This is a particular problem for monitoring wells along the Central Coast where the use of single 100mL grab sample has a less than 50% probability of detection for *E. coli* given its low background levels. As sample volume per assay is increased to 1 liter or more (i.e., combine ten or more 100mL grab samples at once), the probability of detecting *E. coli* rapidly increases for most of the regions, seasons, and sources of water. If the produce industry were to support this alternative sampling strategy, given its potentially lower cost compared to the cost of increased sampling frequency, it would eventually require a revision of the design of regulatory water testing protocols, such as the technology currently being used by IDEXX Laboratories. We propose that this strategy be thoroughly evaluated and if proven to be a viable, sensitive, and cost-effective alternative to the single 100mL grab sample, we believe market forces will drive commercial innovation to develop inexpensive 1-liter assays. Adoption of, for example, a 1-liter assay should generate considerable cost savings for the produce industry as a strategy to allow reduction in the sample frequency yet maintain

adequate sensitivity to detect *E. coli*. Sampling water at critical periods of the growing season and during periods of high water-quality vulnerability, taken together with higher volumes per assay, could form the basis of a more strategic and focused protocol for monitoring irrigation water quality. In addition, when sampling along a canal network from main and majors down to the laterals, it may be possible to reduce the number of sampling locations if the concentration of *E. coli* is highly correlated along these canal networks, but this assertion would need to be carefully evaluated prior to implementation.

Outcomes and Accomplishments

With assistance from numerous organizations, processors, and diagnostic laboratories, along with a lot of hard work by individual growers and their food safety staff, two datasets were compiled that together represent a very large number of produce growers from throughout California from various sources of irrigation water and for all seasons of the year. The first dataset was comprised of 44,249 water samples and a second dataset of 15,486 samples was generated by targeting a different cross-section of the produce industry. Together this represents about 60,000 data points that has allowed us to determine background levels, industry trends, and risk factors for the occurrence of *E. coli* and the odds of an exceedance in irrigation water typical for the California produce industry.

We have characterized the level of *E. coli* and the odds of an exceedance for foliar application for four major produce regions of California for the majority of sources of irrigation water for each of the four seasons of the year. Various risk factors were identified for the occurrence of *E. coli* and the odds of an exceedance were calculated for wells, reservoirs, and canal water sources. We also examined the effect of sampling frequency using the standard 100mL grab sample to properly classify a location as to its risk level of *E. coli*. Finally, we determined the positive effect of increasing the sample volume per assay to detect *E. coli* in irrigation water as a means to improve the ability to detect *E. coli* in irrigation water and to generate cost savings for the produce industry if the sampling frequency is to be reduced for some regions, seasons, and/or sources of water.

Summary of Findings and Recommendations

Findings

Two datasets were generated. For the first dataset (Set 1, n=44,249) we established three tiers of analysis based on the thoroughness of the data. Tier 1 included all water samples that had a minimum of basic information (date, city, water source). Tier 2 was a subset of tier 1 data, tier 3 was a subset of tier 2, with tier 3 having the most complete information about the water sample. A second dataset (Set 2, n=15,486) was generated from a different cross-section of the produce industry. Similar to Set 1, these data were from across California growing regions covering a wide range of farming operations, water sources and locations.

For Tier 1 data from Set 1, the majority of water samples (79% (35,093/44,249)) contained no detectable *E. coli,* and only 0.86% (380/44,249) of water samples exceeded the single sample maximum (SSM) of >235 MPN/100mL for foliar application. Less than 0.43% of samples exceeded the SSM for non-foliar application of >576 MPN/100mL. This indicates that exceedances in California irrigation water supplies were rare between the years of February 2007 through September 2010. The prevalence of water samples with any level of detectable *E. coli* (MPN ≥1 /100mL) varied between water sources. About 8% of well samples had detectable *E. coli* compared to 86% and 48% of canal and reservoir samples, respectively.

For Tier 2 data from Set 1, *E. coli* concentrations (MPN/100mL) varied significantly across season for most of the regions. Levels of *E. coli* were highest in the two Central Coast regions during the fall season (Sept-Nov), while predicted concentrations were highest in the Desert region during the summer (June-Aug). We did not detect significant seasonal differences in *E. coli* concentration in the Central Valley. The occurrence of SSM exceedances for foliar application ranged from 0.45% to 1.65% across season, with highest proportion of exceedances occurring in summer and especially during the fall. When stratified by water

source, well and reservoir water samples had a higher odds of exceedance during summer and fall, respectively, compared to exceedances in winter. The odds ratios for the seasonal patterns were just the opposite for canal water, whereby the odds of an exceedance was lower in summer and fall compared to winter, but when mean air temperature was added to the calculation the overall risk of an exceedance was much higher in summer than winter. Adjusted for season, mean air temperature was negatively associated with the odds of an exceedance for reservoirs. In contrast, there was a positive association between mean air temperature and the odds of an exceedance for canal sources (from mostly desert region), which functioned to substantially increase the calculated risk of an exceedance during summer and fall season. Adjusted for season, wind speed was negatively associated with the odds of an exceedance for well water sources but not significant for reservoir and canal sources.

For Tier 3 data from Set 1, exceedances were also rare, with 1.1% of all 17,788 samples having >235 MPN/100mL. Concentrations of *E. coli* were significantly greater in reservoirs compared to wells on the same property in the Central Coast regions. In addition, the odds of an exceedance for foliar application was about three times higher for water taken from a reservoir compared to water taken from a nearby well on the same ranch. Similarly the odds of an exceednace was three time higer for water sampled during summer and fall compared to winter. Twenty-four hour cummulative precipitation was positively associated with the likelihood of an exceedance in wells and reservoirs. For canals in the California desert, the odds of an *E. coli* exceedance (>235 MPN/100mL) was about 12 times greater in summer compared to winter. This higher risk of an exceedance occurred during the time of year when leafy green produce was for the most part not grown, hence, microbial water quality was at its best during the period when leafy green produce was being grown.

For Set 2 data (n=15,486), 0.71% of the samples exceeded the SSM of >235 MPN/100mL for foliar application and only 0.19% of samples exceeded the SSM for non-foliar application of >576 MPN/100mL. Concentrations of *E. coli* were greatest in the summer and fall for the entire data set. While exceedances for foliar application were rare in well samples (~0.2%) and non-existent in reservoir samples, they were more frequent in canal samples (2.5%). Exceedances were more common in the summer and fall. The proportion of water samples that exceeded the SSM for foliar application ranged from 0% to 2.8% depending on the point of entry.

Using simulation methods to evaluate the advantages and disadvantages of alternative irrigation sampling plans, we determined that when the amount of *E. coli* across a season had minimal variability for a location, a relatively small number of samples were needed to correctly estimate water quality with high probability. Conversely, in scenarios with occasional high concentrations (i.e. sporadic spikes of *E. coli*), it is more difficult to capture these rare events with a sampling plan and thus a higher number of samples are needed to correctly classify water quality for a location. Reducing the rate of water monitoring under these high variance conditions may substantially reduce the probability that a grower correctly classifies their location's water quality. As an alternative, the grower can potentially increase the volume processed per assay and substantially increase the likelihood of detecting *E. coli* in irrigation water, especially for well samples. Volumes of one liter or more appear to capture most of the benefit of using a higher volume, hence, private biotechnology firms might consider partnering with the produce industry to develop the necessary platform that conveniently and cost-effectively processes one liter samples similar to how 100mL are processed today.

Recommendations

Based on two large datasets totaling about 60,000 data points that represent a very large number of produce growers from throughout California from various sources of irrigation water and for all seasons of the year, the rate of exceedances was uncommon for most locations and sources of irrigation water. Using this low rate of exceedance as a justification to reduce the rate of monitoring for locations with persistent high water quality should be an active discussion point between produce industry, allied associations, scientists familiar with epidemiological and risk-based approaches, and regulatory agencies. To inform such a discussion it

would be prudent to better understand the linkages or lack thereof between generic *E. coli* and the various foodborne pathogens of concern to the produce industry. It is possible that under certain specific risk factor scenarios there are linkages between high levels of indicator *E. coli* and pathogens such as *E. coli* O157:H7 and *Salmonella*, but in the absence of such risk factors minimal correlation likely exists for generic *E. coli* and pathogens associated with foodborne illness. This project has begun to identify some of these potential risk factors, but prospective studies need to be conducted in close collaboration with individual growers to verify proximity, structural and environmental risk factors for not just generic *E. coli*, but produce food safety pathogens.

In our opinion, key to developing more cost-effective water quality monitoring plans for produce food safety is to adopt the use of larger volumes of water per assay. For many regions and sources of water such as wells the use of 100mL volumes does not provide reliable estimates of generic *E. coli*. We propose that the U.S. produce industry thoroughly evaluate the use of a standardized 1-liter assay that can simultaneously increase the accuracy of monitoring data and possibly set the foundation for more site-specific rates of sampling based on risk levels for bacterial contamination. As the industry moves in this direction, market forces will likely drive commercial innovation to develop an inexpensive 1-liter assay. Lastly, sampling water at critical periods of the growing season and during periods of high vulnerability for microbial water quality, when taken together with higher volumes per assay, could form the basis of a more strategic and focused protocol for monitoring irrigation water quality and insuring produce food safety.

APPENDICES

Appendix A. Publications and Presentations

Atwill, E.R. 2010. Epidemiologic analysis and risk management practices for reducing E. coli in irrigation source water supplies and distribution systems. 2010 Fresh Summit: Produce Marketing Association. Orlando, Florida. October 15-17.

Atwill, E.R. 2011. Irrigation water and produce food safety. Washington Tree Fruit Research Commission and Northwest Horticultural Council. Yakima, Washington. January 26 and 27.

Atwill, E.R. 2011. WIFSS research on preharvest food safety. FDA Advisory Board Meeting. College Park, Maryland. April 5-6.

Atwill, E.R. 2011. Mining Industry Data, panel speaker. 2nd Annual Conference, Center for Produce Safety. Orlando, Florida. June 27-28.

Appendix B. Budget Summary

CATEGORY	BUDGETED	EXPENDED	REMAINING
SALARY	172,641.00	125,771.75	46,869.25
SUPPLIES AND EXPENSE	56,265.00	35,502.4	20,762.6
TRAVEL	19,300.00	5,317.79	13,982.21
EMPLOYEE BENEFITS	57,238.91	44,697.22	12,541.69
Totals:	**300,000.00**	**227,214.86**	**72,785.14**

Appendix C. Additional Tables/Figures

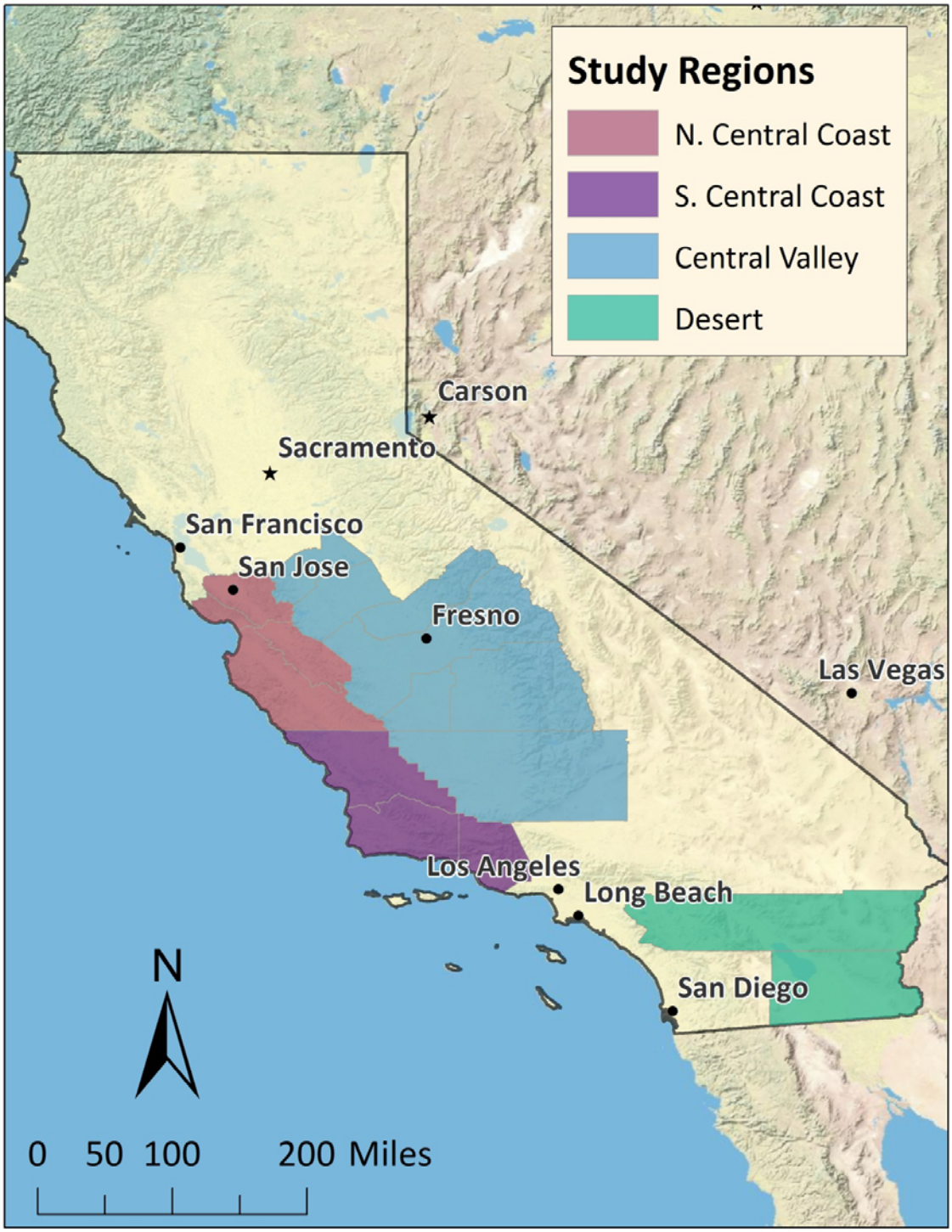


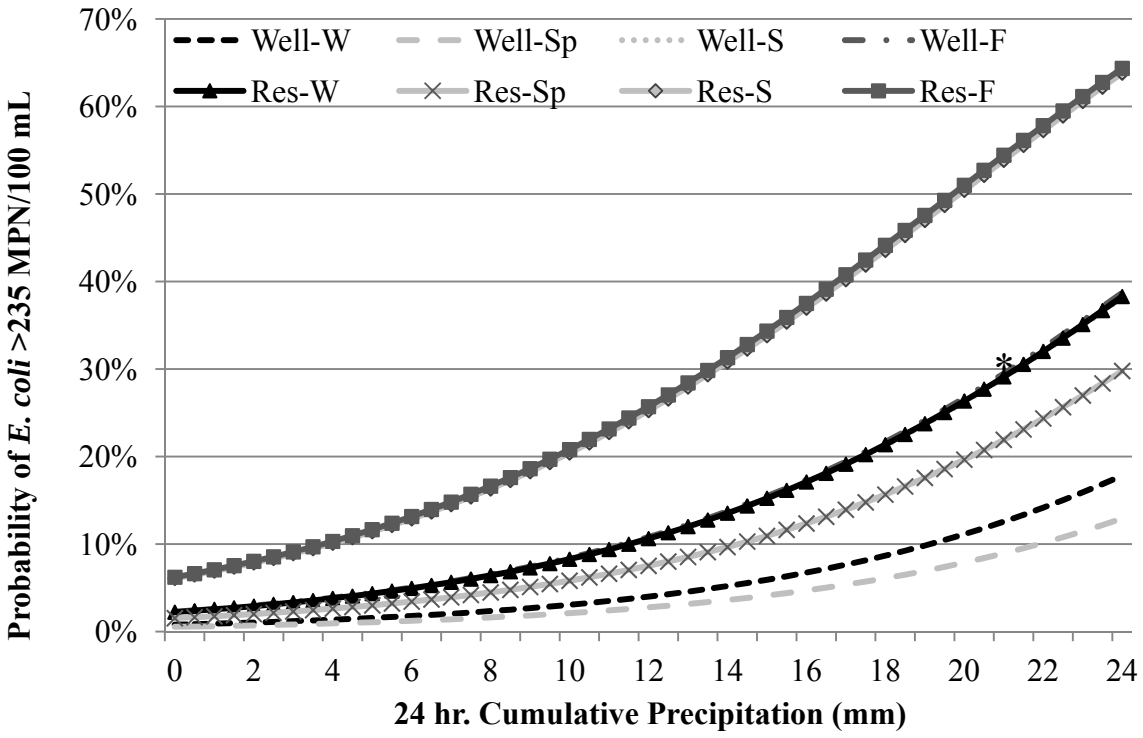
Figure C1. Map of California displaying the four geographical regions used to designate region for this project.

Table C1. The odds of an *E. coli* exceedance (*E. coli* >235 MPN/100mL) associated with water source, season, air temperature and wind speed.

Covariates		Odds Ratio	p-value	95% Confidence Interval	
Well					
Season	Winter [a]	1.00	--	--	--
	Spring	1.84	0.113	0.87	3.91
	Summer	3.60	0.001	1.66	7.81
	Fall	2.86	0.007	1.33	6.14
Mean 24 hr. air temperature (°C) [†]		1.01	0.598	0.96	1.07
Mean 24 hr. wind speed (m/s)		0.70	0.008	0.54	0.91
Reservoir					
Season	Winter [a]	1.00	--	--	--
	Spring	0.92	0.880	0.32	2.67
	Summer	6.63	0.001	2.87	15.30
	Fall	14.83	0.001	7.01	31.36
Mean 24 hr. air temperature (°C)		0.93	0.001	0.89	0.97
Mean 24 hr. wind speed (m/s) [†]		0.89	0.298	0.71	1.11
Canal					
Season	Winter [a]	1.00	--	--	--
	Spring	0.83	0.666	0.35	1.97
	Summer	0.10	0.001	0.03	0.36
	Fall	0.43	0.051	0.18	1.00
Mean 24 hr. air temperature (°C)		1.15	0.001	1.10	1.20
Mean 24 hr. wind speed (m/s) [†]		0.95	0.582	0.78	1.15

[a] Winter season set as the referent category, hence OR=1.0; the odds of SSM exceedance for foliar application for other seasons are compared against the amount of exceedances occurring during winter.

[†] These variables were not significant in the season-specific model, but were retained in the model in order to have equivalent models for each of the three water sources.



*Wells during the Summer and Fall have almost the identical probability of exceedance as reservoirs in the Winter.

Figure C2. Probability of *E. coli* exceedance (E. *coli* >235 MPN/100mL) in well-reservoir (Res) pairs associated with 24-hour cumulative precipitation across seasons; Winter (W), Spring (Sp), Summer (S) and Fall (F).

Table C2. The odds of detecting any *E. coli* (>0 MPN/100mL) associated with well-reservoir pairs, season and wind speed.

Covariates		Odds Ratio	p-value	95% Confidence Interval	
Reservoirs		3.55	0.000	2.87	4.41
Season	Winter [a]	1.00	--	--	--
	Spring	1.78	0.001	1.25	2.54
	Summer	3.21	0.001	2.28	4.53
	Fall	2.16	0.001	1.54	3.04
Mean 24 hr. wind speed (m/s)		0.83	0.006	0.72	0.95

[a] Winter season set as the referent category, hence OR=1.0.

Table C3. The odds of detecting any *E. coli* (>0 MPN/100mL) in canal water associated with different locations along the irrigation canal network, adjusted for feedlot density and season.

Covariates		Odds Ratio	p-value	95% Confidence Interval	
Season	Winter [a]	1.00	--	--	--
	Spring	2.29	0.000	1.57	3.34
	Summer	1.45	0.092	0.94	2.23
	Fall	1.39	0.028	1.04	1.86
Nearest Canal Network Part					
	Main Canal[b]	1.00	--	--	--
	Major Canal	11.67	0.001	7.22	18.85
	Lateral Canal	9.63	0.001	6.57	14.10
	Drain	3.26	0.001	1.63	6.54
	Wash	5.84	0.001	3.47	9.83
Feedlot Count w/in 5Km		1.16	0.032	1.01	1.34

[a] Winter season set as the referent category, hence OR=1.0
[b] Main canal as referent category, hence OR=1.0

Table C4. Average *E. coli* concentrations (MPN/100mL) across seasons by water source.

Water Source	Mean *E. coli* MPN/100mL	N
Winter	**2.31**	**3964**
Well	1.14	3357
Canal	13.28	348
Reservoir	0.76	122
Other	4.33	137
Spring	**4.05**	**3443**
Well	0.92	2173
Canal	10.28	1112
Reservoir	1.59	98
Other	5.73	60
Summer	**12.00[a]**	**3469**
Well	2.76	2644
Canal	54.59	622
Reservoir	0.12	112
Other	4.26	91
Fall	**17.51[a,b]**	**4607**
Well	2.51	2372
Canal	36.04	2064
Reservoir	0.26	71
Other	2.74	100
Total	**9.39**	**15483**

[a] Significantly greater than winter and spring, $p<0.01$
[b] Significantly greater than summer, $p<0.01$

Table C5. Proportion of samples with exceedances (>235 MPN/100mL) and *E. coli* occurrences (MPN/100mL>0) across several points of entry (POE) as reported by growers.

POE	EC≤235	EC>235	EC=0	EC>0
Gate	97.2%	2.8%	23.8%	76.2%
Sprinkler	99.7%	0.3%	83.9%	16.1%
Drip tape	100.0%	0.0%	89.7%	10.3%
Reservoir	100.0%	0.0%	50.0%	50.0%
Hose	100.0%	0.0%	96.7%	3.3%
Pipe	100.0%	0.0%	74.2%	25.8%
Furrow	100.0%	0.0%	67.3%	32.7%
Turnout	100.0%	0.0%	83.0%	17.0%
Valve	99.07%	0.03%	91.3%	8.7%
Booster	100.0%	0.0%	86.5%	13.5%
Other	98.4%	1.6%	83.7%	16.3%

Table C6. Odds of detecting any *E. coli* (>0 MPN/100mL) associated with season, water source, sample point of entry (POE) and growing region.

Covariates		Odds Ratio	p-value	95% Confidence Interval	
Season	Winter [a]	1.00	--	--	--
	Spring	0.84	0.027	0.72	0.98
	Summer	1.12	0.141	0.96	1.30
	Fall	1.99	0.000	1.76	2.26
Water source	Well [b]	1.00	--	--	--
	Reservoir	0.51	0.002	0.33	0.79
	Canal	0.80	0.686	0.27	2.38
	Other	1.35	0.073	0.97	1.87
POE	Valve [c]	1.00	--	--	--
	Gate	3.37	0.001	2.70	4.20
	Sprinkler	1.02	0.774	0.88	1.18
	Drip tape	0.71	0.029	0.52	0.97
	Reservoir	2.18	0.322	0.46	10.27
	Hose	0.19	0.006	0.06	0.61
	Pipe	2.23	0.001	1.71	2.90
	Furrow	1.12	0.437	0.84	1.48
	Turnout	1.33	0.155	0.90	1.96
	Booster	1.91	0.056	0.98	3.73
	Other	0.95	0.846	0.55	1.63
Region	N. Central Coast [d]	1.00	--	--	--
	S. Central Coast	1.21	0.118	0.95	1.53
	Central Valley	0.67	0.052	0.45	1.00
	Desert	7.44	0.001	2.48	22.34

[a] Winter set as referent category, hence OR=1.0
[b] Well as referent category, hence OR=1.0
[c] Valve as referent, hence OR=1.0
[d] N. Central Coast as referent, hence OR=1.0.

Table C7. Definition of *E. coli* concentration categories for simulation purposes. A location is assigned to one of seven risk categories based on its historical monitoring data.

Risk category	Lower boundary (MPN /100mL)[a]	Upper boundary (MPN /100mL)[a]	Concentration variable used to determine risk category
1	0	0	All samples are non-detects
2	0	20	5-sample geometric mean
3	20	50	5-sample geometric mean
4	50	126	5-sample geometric mean
5	126	235	5-sample geometric mean. Type 1 exceedance
6	235	576	Maximum individual sample. Type 2 exceedance
7	576	3000[b]	Maximum individual sample. Type 3 exceedance

[a] Each risk category is defined as above the lower boundary, and below or equal to the upper boundary.
[b] The maximum value in the data set is 2420 MPN/100mL.

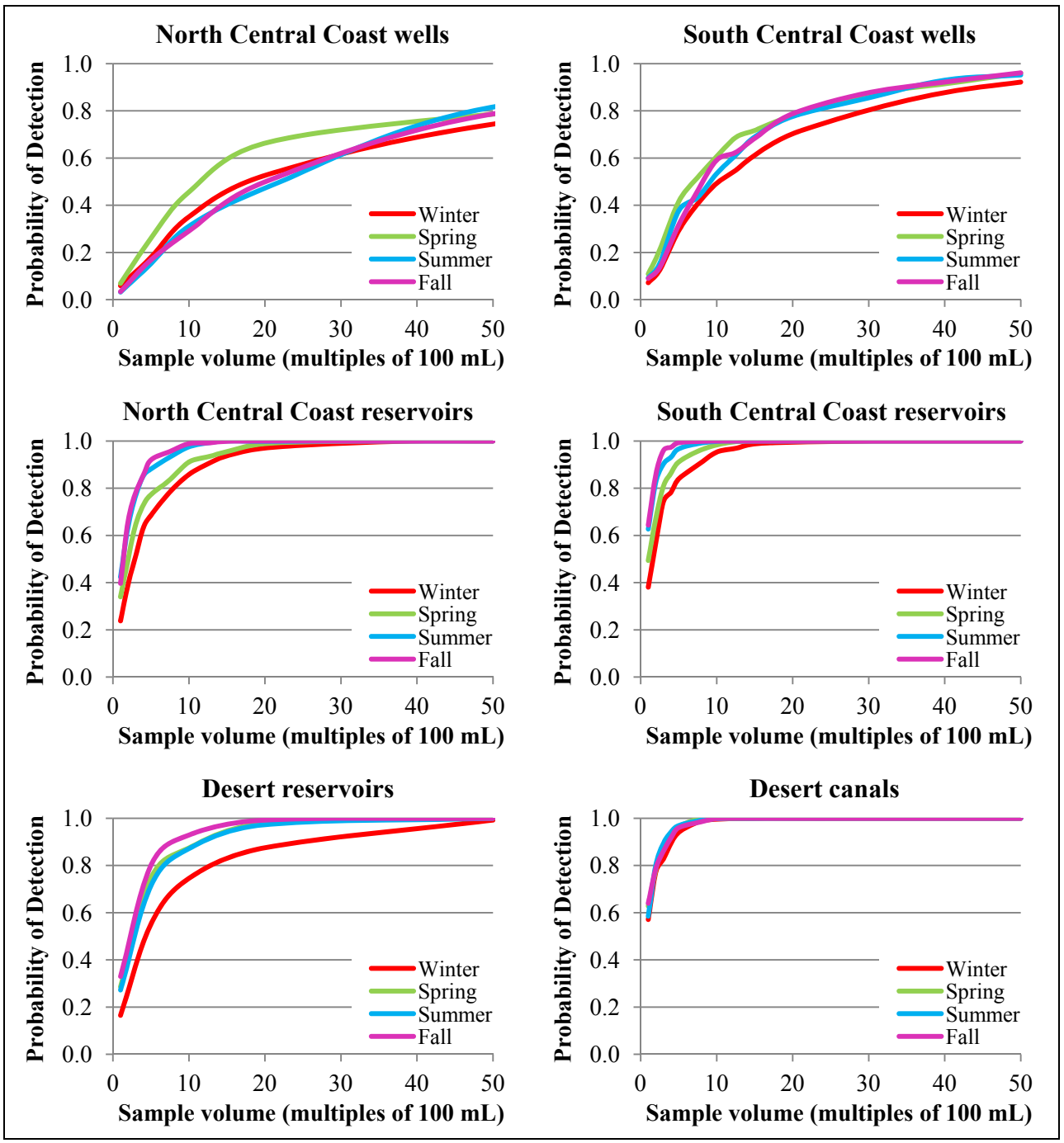


Figure C3. Probability to detect *E. coli* in a water sample as a function of the number of 100mL grab samples taken at one time (i.e., a water sample is comprised of 1 or more 100mL grab samples). These simulations assume that the variability of *E. coli* in 100mL grab samples is the same within a day as across the season, which may not be true for all regions, seasons, and sources of water; hence, results should be interpreted with caution until better data is made available to verify or modify these assumptions and results. Lines may not appear smooth due to interpolation and simulation variability.