# 5. Occurrence and Predictive Model

## 5.1 Introduction

EPA determined from the Six-Year Review data[1] that systems differ considerably with respect to the observed occurrence of total coliform (TC) and *E. coli* (EC) in both routine and repeat sampling. Some of this variability can be explained by the type, size, water source, and disinfection practices of the systems. However, even among systems having the same characteristics, some systems were found to rarely experience positive assays, while others often did.

EPA developed a model to describe TC and EC occurrence in water systems for the current TCR (baseline for this EA) and the proposed Revised TCR (RTCR) alternatives. The model generates estimates of reduced TC and EC occurrence based on requirements of the proposed RTCR options to perform assessments and corrective actions not explicitly required in the current TCR. (In addition, the models take into account reductions attributable to the implementation of the Ground Water Rule (GWR), which will be effective in December 2009.)

The model developed by EPA has two major, separate components. The first model component focuses on the distribution of the routine and repeat TC and EC hit rates within each category of water systems. These distributions reflect variation both within and among specific groups of systems. The second component of the model uses the TC and EC occurrence distributions to predict changes in TC and EC occurrence over time due to the changes in monitoring frequency and the new assessment and corrective action requirements resulting from the GWR and proposed RTCR. The changes in TC and EC occurrence are then compared with the expected TC and EC occurrence over time from the current regulations (TCR and GWR) to assess the impacts of the proposed RTCR.

Occurrence modeling focuses on small PWSs (those serving ≤4,100 people) because (a) there are adequate data to support development of the occurrence distributions that are needed for the predictive components of the model, (b) these systems are known to have higher occurrence levels than larger systems, and (c) these systems, in particular, are subject to changes in the monitoring requirements under the proposed RTCR. Systems that are not affected by changes in the baseline routine monitoring requirements are: a) All PWSs using SW; b) all CWSs, which must monitor monthly regardless of system type or size; and c) GW NCWSs serving >1,000 people. Chapter 3 of this EA provides further detail on requirements under each regulatory scenario by system type and system size.

The occurrence model produces outputs over a 25-year time horizon[2] that include national estimates of TC and EC occurrence in PWSs across the U.S., and the resulting Level

[1] A discussion of the use of data from the Six-Year Review of SDWA regulations is presented in Chapter 4 of this EA.
[2] The model includes an additional 5 years prior to the commencement of proposed RTCR monitoring to account for effects of the GWR after it is fully implemented, for a total of 30 modeled years.

1 and Level 2 assessments[3] and corrective actions performed. These outputs are used to characterize reduced exposure to the potential contamination under each regulatory scenario considered, as presented in Chapter 6 of this EA, and to estimate incremental costs for PWSs across the U.S., as described in Chapter 7 of this EA.

The two major components the occurrence model used for systems serving ≤4,100 people are described in sections 5.2 and 5.3 of this chapter.

Although not included in the modeling described in sections 5.3 and 5.4, this EA also estimates the expected levels of occurrence for larger systems (i.e., serving 4,101 to 33,000 people) using violations data from the Safe Drinking Water Information System-Federal Version (SDWIS/FED) (USEPA, 2007); the selection of this data source is explained section in 4.3.4.2 of this EA. EPA did not develop estimates of indicator reduction for the largest systems (i.e., serving >33,000 people) because: a) occurrence data are scant for systems serving >4,100, for reasons described in section 5.4; and b) these systems rarely have sufficient numbers of positive assays for TC or EC to trigger the systems into nonacute or acute violations which is not expected to notably change under the proposed RTCR. For the largest systems (serving >33,000 people), which generally have more resources and more sophisticated operations, EPA expects no notable change in monitoring, corrective actions, or indicator occurrence and so they are not modeled for this EA.

A complete description of the methods for developing the occurrence estimates as described above is found in the remaining sections of this chapter:

**Section 5.2** Describes development of parameter estimates $\alpha, \beta$ for use in the proposed RTCR occurrence model.

**Section 5.3** Describes the occurrence model used for systems serving ≤ 4,100 people.

**Section 5.4** Describes the how occurrence is characterized for systems serving 4,101 – 33,000 people.

[3] The AIP and Alternative Analysis require PWSs to complete either a Level 1 or Level 2 assessment following triggers specified in Exhibit 3.1 of this EA. The purpose of Level 1 and 2 assessments is to identify the presence of sanitary defects and deficiencies in distribution system coliform monitoring practices, similar to an annual inspection but more focused on determining the cause of a contamination event. In order to complete a Level 2 assessment the PWS must perform a more detailed examination of the system than for a Level 1 assessment, including its monitoring and operational practices.

## 5.2 Modeling of Current TC and EC Hit Rate Distributions for Systems Serving up to 4,100

### 5.2.1 Distributional Model and Notation

The Six-Year Review data demonstrate that, even among systems of similar size and type, there are some which rarely find routine samples to be TC-positive and others that often find TC-positive samples. For example, among the very small (serving <101) transient noncommunity water systems (TNCWSs) using nondisinfected ground water, 47 systems reported ten or more routine TC positives, while assaying 40 or fewer samples during 2005. Among the same set of small TNCWSs with nondisinfected ground water are 60 that assayed 20 or more routine samples, finding all of them TC-negative. Clearly, routine TC-positive rates vary considerably within this set of systems. Significantly different fractions positive can be found among any basic subsets of systems (defined by system type, water type, and population served, as described in section 5.2.2) and for any type of assay (both routine and repeat TC and both routine and repeat *E. coli*).

Within a group of similar systems, pRTTC, which is defined as the probability that a routine sample will test positive, is assumed to vary as a beta random variable. The beta distribution is commonly used to model varying probabilities because it is limited to values between 0 and 1 and can assume a wide variety of shapes. EPA used a beta distribution to describe how probabilities of virus-positive samples vary among virus-contaminated wells in the Economic Analysis for the Ground Water Rule (USEPA 2006a).

Samples that test negative for TC are generally not tested further for the presence of *E. coli* (EC), but TC-positive samples are. pRTEC, which is defined as the probability that a TC-positive routine (RT) sample will also test EC-positive, is assumed to vary among similar systems as another beta random variable, but with parameters that can be quite different from those of the routine TC samples. Although systems with frequent routine TC positives necessarily generated more EC assays, the EC positive rates for these systems did not appear to be significantly different than the EC positive rates of systems with rare TC positives. The variables pRTTC and pRTEC are therefore modeled as independent beta-distributions.

Similar assumptions are extended to the repeat positive probabilities, pRPTC and pRPEC. Like pRTTC and pRTEC, these are assumed to be independently beta-distributed among sets of similar systems. A process for determining "sets of similar systems" is described in sections 5.2.3 and 5.2.4. The process is informed by system-specific data that are reduced from the sample-specific data that were described in Chapter 4 of this EA, as discussed in the following section.

### 5.2.2 Data Reduction

The statistical model assumes that each of a system's positive routine TC measurements results from some unobserved probability (pRTTC). This probability may vary from month-to-month and season-to-season. However, in this model each result, whether positive or negative, is assumed to be an independent trial with positive probability pRTTC. Each of these is called a Bernoulli trial. The total number positive during a year (KRTTC) is therefore a binomial random variable, whose parameters are the positive probability (pRTTC) and the total number of samples assayed (NRTTC). What we observe

for the system are the number of positive samples (KRTTC) and the number of samples assayed (NRTTC). The order of the positives and negatives is unimportant, so the routine TC data needed for any system can be reduced to these two values: NRTTC and KRTTC. The leading character "N" designates the number of samples tested and "K" designates the number positive.

Similarly, the routine EC data for a system can be reduced to the number of routine TC-positive samples that were assayed for EC (RTEC) and, of those, the number testing positive. For many systems, both NRTEC and KRTEC will be zero because they encountered no routine TC positives during 2005 (KRTTC = 0), and EPA assumes that in the case of a TC-, no EC assay is performed.

Finally, the data for repeat samples for a system can be reduced to four numbers: NRPTC, KRPTC, NRPEC, and KRPEC. Again, the leading characters, "N" and "K", designate numbers of samples assayed and numbers positive, respectively. "RP" designates that these are repeat samples, while TC and EC designate assays for TC and EC, respectively. For many systems, all four of these numbers are zero because no routine samples were TC positive.

It is important to note that only TC-positive samples are assayed for EC. The positive probabilities for EC are therefore conditional, that is, they apply only to samples that have tested TC positive. All TC-negative samples are assumed to be EC-negative. The overall (unconditional) probability that a routine sample will be EC-positive is the product pRTTC*pRTEC and the overall (unconditional) probability that a repeat sample will be EC-positive is the product pRPTC*pRPEC.

The reduced dataset includes eight integer values (for NRTTC, KRTTC, NRTEC, KRTEC, NRPTC, KRPTC, NRPEC, KRPEC) for each of approximately 93,000 systems. Although none of the NRTTC are zero (all systems assayed at least one routine TC sample during 2005), the great majority of the other values are zeros.

### 5.2.3 Basic Subsets of Systems

The 92,747 systems serving ≤4,100 people in the Six-Year Review dataset (which represent approximately 147,000 total systems in this size range) can be classified by system type, source water type and disinfection status, and population served. The three types of systems are community (CWS), transient noncommunity (TNCWS) and nontransient noncommunity (NTNCWS). Three water source types are surface water (SW), which is always disinfected, nondisinfected ground water (NondisGW) and disinfected ground water (DisGW). Three important size categories are small systems (serving ≤100 people), systems serving 101 to 1,000 people, and systems serving 1,001 to 4,100 people. Systems serving more than 4,100 people are not included here, for reasons discussed in Chapter 4 of this EA and earlier in this chapter. In total, there are 27 possible categories that each PWS may belong to (3 system types x 3 source water types x 3 population served sizes). Each PWS may only be in one of the 27 categories. Exhibit 5.1 lists the 27 basic subsets and corresponding numbers of systems in the dataset.

**Exhibit 5.1 Basic Classifications of Systems Used for Occurrence Modeling**

| Subset | SystemType | Water Type | Population Served Group | Number of Systems in 6-Year Review Data |
|---|---|---|---|---|
| 1 | TNCWS | SW | ≤100 | 838 |
| 2 | TNCWS | SW | 101-1000 | 358 |
| 3 | TNCWS | SW | 1001-4100 | 30 |
| 4 | NTNCWS | SW | ≤100 | 114 |
| 5 | NTNCWS | SW | 101-1000 | 172 |
| 6 | NTNCWS | SW | 1001-4100 | 33 |
| 7 | CWS | SW | ≤100 | 484 |
| 8 | CWS | SW | 101-1000 | 2,034 |
| 9 | CWS | SW | 1001-4100 | 1,699 |
| 10 | TNCWS | DisGW | ≤100 | 8,112 |
| 11 | TNCWS | DisGW | 101-1000 | 3,007 |
| 12 | TNCWS | DisGW | 1001-4100 | 119 |
| 13 | NTNCWS | DisGW | ≤100 | 1,577 |
| 14 | NTNCWS | DisGW | 101-1000 | 1,780 |
| 15 | NTNCWS | DisGW | 1001-4100 | 252 |
| 16 | CWS | DisGW | ≤100 | 3,662 |
| 17 | CWS | DisGW | 101-1000 | 7,678 |
| 18 | CWS | DisGW | 1001-4100 | 2,707 |
| 19 | TNCWS | NondisGW | ≤100 | 32,028 |
| 20 | TNCWS | NondisGW | 101-1000 | 10,217 |
| 21 | TNCWS | NondisGW | 1001-4100 | 217 |
| 22 | NTNCWS | NondisGW | ≤100 | 4,026 |
| 23 | NTNCWS | NondisGW | 101-1000 | 3,264 |
| 24 | NTNCWS | NondisGW | 1001-4100 | 152 |
| 25 | CWS | NondisGW | ≤100 | 3,788 |
| 26 | CWS | NondisGW | 101-1000 | 3,755 |
| 27 | CWS | NondisGW | 1001-4100 | 644 |
| | | | **TOTAL** | **92,747** |

Note: Water Type includes disinfection status, as informed by information from SDWIS-FED. Systems with no indication of disinfection status in SDWIS-FED were assumed to not disinfect (Nondis).

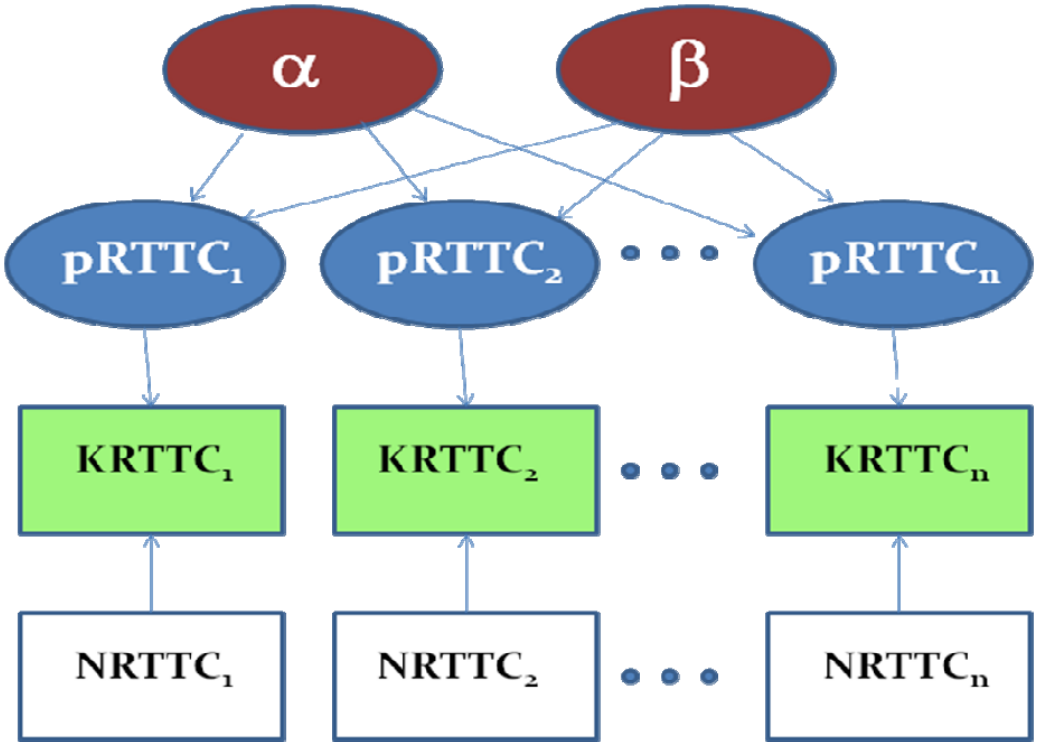
### 5.2.4 Estimation Methodology

The methodology described in this section applies to all four kinds of measurements, routine TC (RTTC), routine *E. coli* (RTEC), repeat TC (RPTC) and repeat *E. coli* (RPEC). However, to simplify the presentation, the description will be expressed only in terms of RTTC. Results presented in section 5.6 show that the final system classifications used to model TC (both routine and repeat) are quite different from the final subsets used to model EC occurrence.

For any individual subset of systems from Exhibit 5.1 or grouping of these subsets, two methods were used to estimate parameters of the beta-distributed RTTC: maximum likelihood estimation and Bayesian Markov Chain Monte Carlo (MCMC) sampling. The MCMC samples are used to test whether two basic subsets have equal average positive rates and to check the maximum likelihood estimates (MLEs). This check was conducted to ensure that the algorithm used to find the MLE had not stopped too soon.

Both the MLE and MCMC methodologies require computation of the likelihood function, L(data | $\alpha,\beta$), where $\alpha$ and $\beta$ are parameters of beta-distributed pRTTC. For a specific system in a set of similar systems, the number of positive RTTC samples is a binomial random variable with parameters NRTTC and pRTTC, which are the numbers of RTTC samples assayed by the system and the system's unobserved probability of a positive, respectively. The probability of a positive, pRTTC, is a beta random variable, whose parameters are $\alpha$ and $\beta$. Exhibit 5.2 is a directed graph, showing the model structure. At the top are the two high-level parameters ($\alpha$ and $\beta$) and arrows from them showing their influence on the distribution of the system-specific parameters, the positive probabilities or pRTTCs. Finally, the Exhibit shows how the pRTTCs, together with the numbers of samples assayed (NRTTCs), influenced the numbers estimated to be positive (KRTTCs).

**Exhibit 5.2 Directed Graph of Model Used for RTTC Occurrence**




This same model could have been equally well represented with all of the arrows pointing up to show the direction of inference, starting with the data (NRTTCs and KRTTCs), which

inform estimates of the unobserved positive probabilities (pRTTCs), and then those informing estimates of the high level parameters ($\alpha$ and $\beta$).

### 5.2.4.1 Maximum Likelihood Estimation

For a single system in a subset, if pRTTC were known, the likelihood of the system's RTTC data (KRTTC and NRTTC) would be given by the binomial probability mass function:

$$\text{dbinom(KRTTC,NRTTC,pRTTC)} = \text{NRTTC!} * \text{pRTTC}^{\text{KRTTC}} * (1\text{-pRTTC})^{\text{NRTTC-KRRTC}}/\text{KRTTC!(NRTTC-KRTTC)!}$$

But pRTTC is not known, nor can it be observed explicitly. It is, therefore, treated as a beta random variable. The expected value of the likelihood function is obtained by integrating the product of the beta probability density function and the binomial probability mass function. The beta probability density function is:

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot \text{pRTTC}^{\alpha-1}\cdot(1-\text{pRTTC})^{\beta-1}$$

And the expected likelihood is:

$$\int_0^1 \frac{\text{NRTTC!}}{\text{KRTTC!}\cdot(\text{NRTTC}-\text{KRTTC})!}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot \text{pRTTC}^{\alpha-1+\text{KRTTC}}\cdot(1-\text{pRTTC})^{\beta-1+\text{NRTTC}-\text{KRTTC}}\,d\text{pRTTC}$$

Because both NRTTC and KRTTC are constant (observed data), the function only depends on the factors involving pRTTC, alpha, and beta, so the factorial terms in the above may be ignored when seeking to maximize the likelihood function. This simplifies the computation considerably, and the resulting integral has a closed-form solution, which is:

$$\frac{\Gamma(\alpha+\beta)\cdot\Gamma(\alpha+\text{KRTTC})\cdot\Gamma(\beta+\text{NRTTC}-\text{KRTTC})}{\Gamma(\alpha)\cdot\Gamma(\beta)\cdot\Gamma(\alpha+\beta+\text{NRTTC})}$$

Logarithms were used to avoid numerical overflow/underflow, and Microsoft Excel's Solver add-in was used to find parameters ($\alpha$ and $\beta$) that maximize the sum of the system-specific log-likelihoods.

### 5.2.4.2 Markov Chain Monte Carlo Sampling

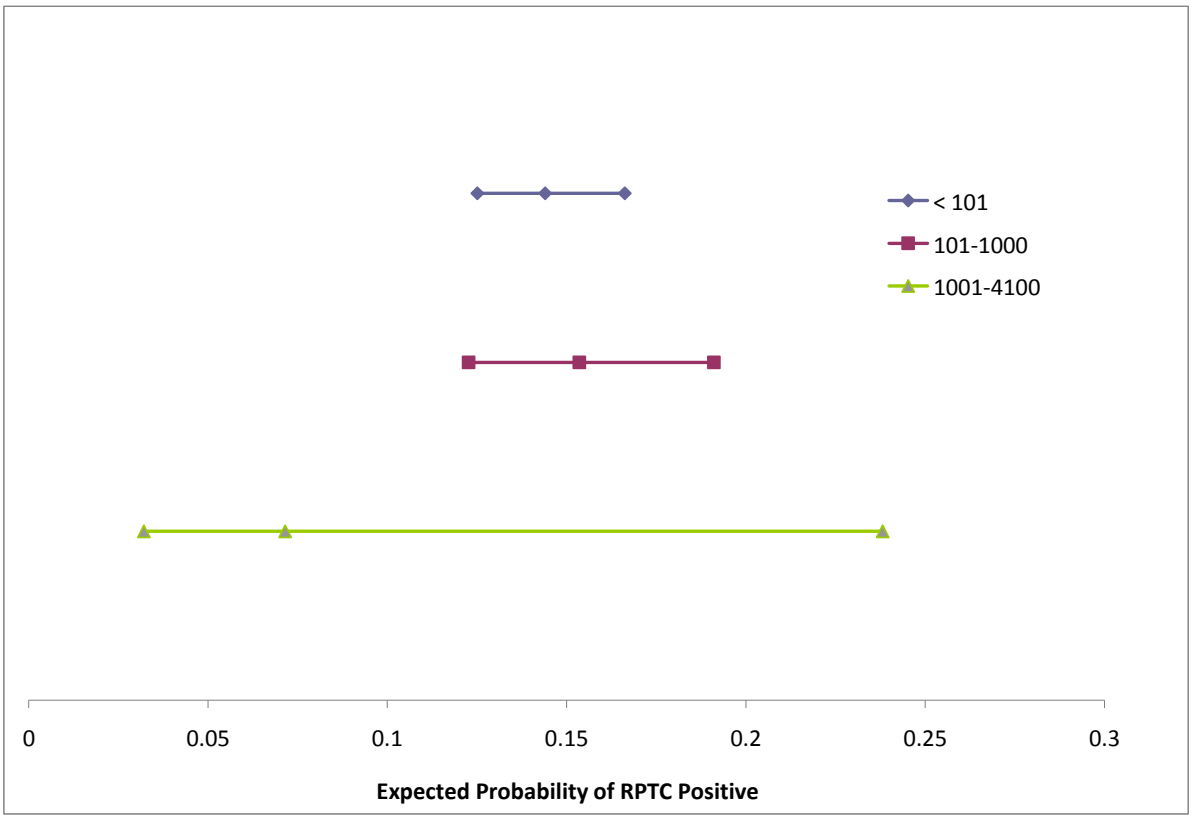
Markov Chain Monte Carlo Sampling was performed using WinBUGS software (Ntzoufras, 2009). This involved encoding the model described above, but with a different parameterization. Based on parameters $\alpha$ and $\beta$, the average positive value is $a = \alpha/(\alpha + \beta)$. A measure of dispersion that is closely associated with standard deviation is $b = (\alpha + \beta)^{-1/2}$. Use of these two parameters $a$ and $b$ (rather than $\alpha$ and $\beta$) is described in Gelman et al. (1995) as useful for this kind of problem. The new parameters are easier to interpret and facilitate the MCMC sampling. Vague, uniform priors were assigned for parameters $a$ and $b$, so as to have little influence on the MCMC sampling.

MCMC results for the key parameter of interest (*a* = average positive probability) were summarized by medians and 95 percent credible intervals. Sample densities for *a* were used to ensure that the MLEs obtained (see 5.5.1) were in the high-density region of the sample.

#### 5.2.4.3 Rationale for Combining Subsets (Pooling Data)

MLEs and MCMC sample results were obtained for all of the basic subsets having sufficient data. Some subsets had insufficient numbers of systems and/or positive samples. For those subsets, professional judgment was used to identify other basic subsets based on perceived general commonalities (e.g., operator skill level, hydraulic characteristics, source water and treatment characteristics, and distribution system complexity) for the purpose of pooling data. For example, Exhibit 5.3 displays MLEs and 95 percent credible intervals for RPTC in three sets of TNCWS systems that disinfect ground water. Having very few systems in the category serving 1001-4100 yielded a very imprecise estimate, while the smaller size categories with larger numbers of systems produced more precise estimates. The figure in Exhibit 5.3 shows that there is no significant size effect, therefore it made sense to combine the these three basic sets. The combined set is "the set of all TNCWS that disinfect ground water and serve at most 4,100 persons." In cases like this, where the data show no significant difference based on system size and combining the data made sense based on professional engineering judgment, the data were pooled to produce improved estimates.

**Exhibit 5.3 Probability of RPTC Positive for TNCWSs that Disinfect Ground Water by System Size**




Beginning with RTTC, EPA completed all of the steps below before following the same steps for RPTC, RTEC, and RPEC:

- For each of the nine combinations of system type and water source, determine whether it makes sense to combine any of the three size categories by using analysis such as shown in Exhibit 5.3.
- Once decisions are made about combining across size categories, produce MLEs and MCMC samples for any new, combined sets and plot the parameter estimates for the sets employing disinfection (DisinfGW and DisinfSW). Combine subsets of systems across water types unless the estimates demonstrate different levels of occurrence.
- After producing MLEs and MCMC samples for any new, combined sets, plot the parameter estimates for NTNCWSs and CWSs. Combine CWSs and NTNCWSs unless the estimates demonstrate different levels of occurrence.

### 5.2.5 Results

Section 5.2.5.1 describes how systems and their data were combined, or pooled, for the purpose of RTTC modeling. Sections 5.2.5.2, 5.2.5.3, and 5.2.5.4 describe the same for modeling RTEC, RPTC, and RPEC, respectively. Section 5.2.5.5 reports the resulting maximum likelihood estimates.

RTTC data were addressed first because a) so much more data were available for RTTC than for the other three measurement types (RTEC, RPTC, and RPEC); and b) summary statistics suggested that systematic differences due to system size and other factors were larger for TC than for EC measurements.

When evaluating data for any one type of measurement, system size was evaluated first because it was expected to be important for a number of reasons, primarily: a) Larger systems have more resources and are more likely to employ full-time professional operators; and b) Having more resources, they may provide better treatment and distribution systems. System type and water type were evaluated after system size, but the order in which these were evaluated does not appear to have been important.

#### 5.2.5.1 Subsets for RTTC

The sections that follow describe how the effects of system size, disinfected water source, and system type (NTNCWS vs. CWS) were considered in deciding how to combine basic subsets for the purpose of modeling RTTC occurrence.

##### 5.2.5.1.1 Size Effects

No significant size effect was observed for TNCWS Disinf GW, for TNCWS Disinf SW, and NTNCWS Disinf GW. For each of these three, all systems serving ≤4,100 were combined into a larger set for RTTC modeling.

The set NTNCWS Disinf SW ≤100 appeared to have significantly greater average pRTTC than the two larger size categories. Therefore, systems serving 101-4,100 were combined into a larger set, but the basic set of systems serving ≤100 was modeled separately.

The set NTNCWS Nondis GW ≤100 also had significantly greater average pRTTC than the two sets for NTNCWS Nondis GW systems serving 101-4100, so systems serving 101-4,100 were combined into a larger set, but the basic set of systems serving ≤100 was modeled separately.

The set TNCWS Nondis GW 1,001-4,100 had significantly lower mean pRTTC than the two smaller size categories of TNCWS Nondis GW. Therefore, the two smaller size categories were combined, but the larger size category was modeled separately.

The sets TNCWS Disinf GW, CWS Disinf SW, CWS Disinf GW, and CWS Nondis GW all showed significantly decreasing average pRTTC with increasing system size. Therefore, the three size categories for these system types were not combined.

#### 5.2.5.1.2 Effect of GW vs. SW (for Disinfecting Systems)

The six basic subsets of disinfecting TNCWS systems did not appear to have a significant effect related to source water type (GW vs. SW). Because there also were no significant size effects, these six basic subsets were combined for the purpose of modeling RTTC.

The six basic subsets of disinfecting NTNCWS systems also showed no effect related to size. Therefore, disinfecting systems serving 101-4100 were combined into one set and those serving ≤100 were combined into another for the modeling RTTC.

The six sets of disinfecting CWS show significant source water effects in addition to their significant size effects, so the six basic CWS sets were retained (not combined) for RTTC modeling.

#### 5.2.5.1.3 Effect of System Type

For disinfecting CWSs, the effects of size were strong. However, size effects were less pronounced for the NTNCWS systems, so these sets of disinfecting systems were not combined across system type.

For nondisinfecting systems, only the smallest size category showed a significant difference between CWSs and NTNCWSs. No significant effect related to system type is evident in the systems serving 101-4,100 people, so those sets were combined. Sets serving ≤100 should be modeled separately.

#### 5.2.5.1.4 Final Subsets for modeling RTTC

The 14 sets used to model RTTC occurrence are shown in Exhibit 5.4 below:

**Exibit 5.4 Final Subsets of Systems Used for RTTC Occurrence Modeling**

| Subset(s)[1] | System Type(s) | Water Type(s) | Population Served |
|---|---|---|---|
| 1,2,3,10,11,12 | TNCWS | Dis GW, SW | ≤4,100 |
| 4,13 | NTNCWS | Dis GW, SW | ≤100 |
| 5,6,14,15 | NTNCWS | Dis GW, SW | 101-4100 |
| 7 | CWS | Dis GW | ≤100 |
| 8 | CWS | Dis GW | 101-1000 |
| 9 | CWS | Dis GW | 1001-4100 |
| 16 | CWS | Dis GW | ≤100 |
| 17 | CWS | Dis GW | 101-1000 |
| 18 | CWS | Dis GW | 1001-4100 |
| 19,20 | TNCWS | Nondis GW | ≤1000 |
| 21 | TNCWS | Nondis GW | 1001-4100 |
| 22 | NTNCWS | Nondis GW | ≤100 |
| 23,24,26,27 | NTNCWS & CWS | Nondis GW | 101-4100 |
| 25 | CWS | Nondis GW | ≤100 |

[1] Subset numbers refer to the numbers given in Exhibit 5.1

#### 5.2.5.2 Subsets for RTEC

Because only routine samples that are TC-positive are assayed for EC, there are far fewer data available for modeling RTEC than RTTC. Overall, only about one in 40 RTTC samples is assayed for RTEC. For this reason, it is more difficult to discern significant differences among the 27 basic sets of systems.

##### 5.2.5.2.1 Size Effects

Over all nine combinations of system type and water type, only the CWSs using SW showed an effect of size. CWSs using SW serving ≤1000 have a significantly greater conditional probability of routine EC positives than systems serving 1001-4100. For the purpose of modeling RTEC, two sets of CWSs using SW are maintained (those serving 1001-4100 and those serving ≤1000).

For each other combinations of system type and water type, the three basic sets (distinguished by population served) can be combined.

##### 5.2.5.2.2 Effect of System Type

Among systems disinfecting GW, CWSs have significantly lower average pRTEC than the TNCWS and NTNCWS systems.

With the exception of the larger (1001-4100) CWSs, there is no significant effect of system type for the surface water systems.

There is no significant effect of system type for the nondisinfected systems.

##### 5.2.5.2.3 Final Subsets for modeling RTEC

The five sets used to model RTEC occurrence are shown in Exhibit 5.5 below.

**Exhibit 5.5 Final Subsets of Systems Used for RTEC Occurrence Modeling**

| Subset(s)[1] | System Type(s) | Water Type(s) | Population Served |
|---|---|---|---|
| 1 - 8 | TNCWS, NTNCWS & CWS | Disinf SW | ≤4100 (except CWS, where Pop Served ≤1000 ) |
| 9 | CWS | Disinf SW | 1001-4100 |
| 10 - 15 | TNCWS & NTNCWS | Disinf GW | ≤4100 |
| 16 - 18 | CWS | Disinf GW | ≤4100 |
| 19 – 27 | TNCWS, NTNCWS & CWS | Nondis GW | ≤4100 |

[1] Subset numbers refer to the numbers given in Exhibit 5.1

#### 5.2.5.3 Subsets for RPTC

Because each RTTC-positive sample triggers repeat sampling, there are considerably more RPTC data than RTEC data. There are about 3 RPTC measurements for every RTEC measurement in the database.

##### 5.2.5.3.1 Size Effects

Among the nine combinations of System Type and Water Type, only four show significant effects of system size. Three of these are sets of CWSs (those disinfecting SW, those disinfecting GW, and those not employing disinfection) and the fourth is TNCWS systems that provide nondisinfected GW.

For the three sets of CWSs, average pRPTC is significantly smaller for systems serving 1001-4100 than for systems serving at most 1000. The two smaller sets (those serving ≤100 and those serving 101-1000) do not have significantly different occurrence rates, so can be combined.

For TNCWS systems providing nondisinfected GW, the three size categories have significantly different average occurrence levels. The average positive rate is greatest for the systems serving ≤100 and is least for those serving 1001-4100. Therefore, EPA kept the three sets separated for the purpose of modeling RPTC.

##### 5.2.5.3.2 Effect of GW vs. SW (for Disinfecting Systems)

No significant effect of water type is observed for TNCWS or NTNCWS systems. CWSs that disinfect SW have significantly smaller average pRPTC than CWSs; this significant effect applies to all three size categories.

##### 5.2.5.3.3 Final Subsets for modeling RPTC

The 12 sets used to model RPTC occurrence are shown in Exhibit 5.6 below.

**Exhibit 5.6 Final Subsets of Systems Used for RPTC Occurrence Modeling**

| Subset(s)[1] | System Type(s) | Water Type(s) | Population Served |
|---|---|---|---|
| 1,2,3,10,11,12 | TNCWS | Disinf GW, SW | ≤ 4100 |
| 4,5,6,13,14,15 | NTNCWS | Disinf GW, SW | ≤ 4100 |
| 7,8 | CWS | Disinf SW | ≤ 1000 |
| 9 | CWS | Disinf SW | 1001-4100 |
| 16,17 | CWS | Disinf GW | ≤ 1000 |
| 18 | CWS | Disinf GW | 1001-4100 |
| 19,20 | TNCWS | Nondis GW | ≤ 100 |
| 20 | TNCWS | Nondis GW | 101-1000 |
| 21 | TNCWS | Nondis GW | 1001-4100 |
| 22,23,24 | NTNCWS | Nondis GW | ≤ 4100 |
| 25,26 | CWS | Nondis GW | ≤ 1000 |
| 27 | CWs | Nondis GW | 1001-4100 |

[1] Subset numbers refer to the numbers given in Exhibit 5.1

### 5.2.5.4 Subsets for RPEC

Because only TC-positive repeat samples were assayed for EC, there is less RPEC data than even the amount of RTEC data in the dataset. On average, there are 56 times as many RTTC as RPEC data in the dataset. Due in part to the limited availability of data, EPA found few significant differences among the 27 basic subsets of systems.

#### 5.2.5.4.1 Size Effects

Among the nine combinations of system type and water type, none shows a significant effect of system size.

#### 5.2.5.4.2 Effect of System Type

After data from different size categories are combined, there is no apparent effect of system type among either the systems disinfecting SW or the systems that provide nondisinfected GW.

Among systems disinfecting GW, the CWSs have significantly smaller average pRPEC than the TNCWS and NTNCWS systems.

#### 5.2.5.4.3 Effect of Water Type

Systems providing nondisinfected GW have significantly smaller average pRPEC than systems that disinfect GW.

#### 5.2.5.4.4 Final Subsets for modeling RPEC

The four sets used to model RPEC occurrence are shown below in Exhibit 5.7:

**Exhibit 5.7 Final Subsets of Systems Used for RPEC Occurrence Modeling**

| Subset(s)[1] | System Type(s) | Water Type(s) | Population Served |
|---|---|---|---|
| 1-9 | TNCWS, NTNCWS & CWS | SW | ≤ 4100 |
| 10-15 | TNCWS & NTNCWS | Disinf GW | ≤ 4100 |
| 16-19 | CWS | Disinf GW | ≤ 4100 |
| 19-27 | TNCWS, NTNCWS & CWS | Nondis GW | ≤ 4100 |

[1] Subset numbers refer to the numbers given in Exhibit 5.1

### 5.2.5.5 Maximum Likelihood Estimates

Exhibit 5.8 provides Maximum Likelihood estimates (MLEs) for routine and repeat TC, while Exhibit 5.9 provides MLEs for routine and repeat EC. Note that both of these exhibits present the *a* and *b* parameters as defined in section 5.2.4.2.

## Exhibit 5.8 Maximum Likelihood *a* and *b* Parameter Estimates for RTTC and RPTC

| Disinf? | GW or SW? | Sys Type | Size | No. Systems | aRTTC | bRTTC | aRPTC | bRPTC |
|---|---|---|---|---|---|---|---|---|
| Y | SW | TNCWS | ≤100 | 838 | 0.0231 | 0.2992 | 0.1677 | 0.8864 |
| Y | SW | TNCWS | 101-1000 | 358 | 0.0231 | 0.2992 | 0.1677 | 0.8864 |
| Y | SW | TNCWS | 1001-4100 | 30 | 0.0231 | 0.2992 | 0.1677 | 0.8864 |
| Y | SW | NTNCWS | ≤100 | 114 | 0.0158 | 0.3576 | 0.1064 | 1.0780 |
| Y | SW | NTNCWS | 101-1000 | 172 | 0.0113 | 0.2458 | 0.1064 | 1.0780 |
| Y | SW | NTNCWS | 1001-4100 | 33 | 0.0113 | 0.2458 | 0.1064 | 1.0780 |
| Y | SW | CWS | ≤100 | 484 | 0.0150 | 0.2090 | 0.0646 | 0.6107 |
| Y | SW | CWS | 101-1000 | 2034 | 0.0095 | 0.1490 | 0.0646 | 0.6107 |
| Y | SW | CWS | 1001-4100 | 1699 | 0.0059 | 0.1275 | 0.0343 | 0.2721 |
| Y | GW | TNCWS | ≤100 | 8112 | 0.0231 | 0.2992 | 0.1677 | 0.8864 |
| Y | GW | TNCWS | 101-1000 | 3007 | 0.0231 | 0.2992 | 0.1677 | 0.8864 |
| Y | GW | TNCWS | 1001-4100 | 119 | 0.0231 | 0.2992 | 0.1677 | 0.8864 |
| Y | GW | NTNCWS | ≤100 | 1577 | 0.0158 | 0.3576 | 0.1064 | 1.0780 |
| Y | GW | NTNCWS | 101-1000 | 1780 | 0.0113 | 0.2458 | 0.1064 | 1.0780 |
| Y | GW | NTNCWS | 1001-4100 | 252 | 0.0113 | 0.2458 | 0.1064 | 1.0780 |
| Y | GW | CWS | ≤100 | 3662 | 0.0159 | 0.2243 | 0.0942 | 0.8493 |
| Y | GW | CWS | 101-1000 | 7678 | 0.0124 | 0.2142 | 0.0942 | 0.8493 |
| Y | GW | CWS | 1001-4100 | 2707 | 0.0078 | 0.1841 | 0.0522 | 0.4916 |
| N | GW | TNCWS | ≤100 | 32028 | 0.0476 | 0.3990 | 0.2788 | 1.0680 |
| N | GW | TNCWS | 101-1000 | 10217 | 0.0476 | 0.3990 | 0.2487 | 0.9359 |
| N | GW | TNCWS | 1001-4100 | 217 | 0.0247 | 0.3121 | 0.1694 | 0.4052 |
| N | GW | NTNCWS | ≤100 | 4026 | 0.0366 | 0.4044 | 0.2723 | 0.8917 |
| N | GW | NTNCWS | 101-1000 | 3264 | 0.0276 | 0.3591 | 0.2444 | 0.9913 |
| N | GW | NTNCWS | 1001-4100 | 152 | 0.0276 | 0.3591 | 0.1709 | 0.9322 |
| N | GW | CWS | ≤100 | 3788 | 0.0315 | 0.2680 | 0.1898 | 0.8407 |
| N | GW | CWS | 101-1000 | 3755 | 0.0270 | 0.2498 | 0.1898 | 0.8407 |
| N | GW | CWS | 1001-4100 | 644 | 0.0236 | 0.2697 | 0.1312 | 0.6938 |

Note: Disinfection status ("Disinf?") is from information downloaded from SDWIS/FED. Systems with no indication of disinfection status in SDWIS/FED are assumed to not disinfect.

## Exhibit 5.9 Maximum Likelihood *a* and *b* Parameter Estimates for RTEC and RPEC

| Disinf? | GW or SW? | Sys Type | Size | No. Systems | aRTEC | bRTEC | aRPEC | bRPEC |
|---|---|---|---|---|---|---|---|---|
| Y | SW | TNCWS | ≤100 | 838 | 0.1291 | 0.6287 | 0.0518 | 1.1533 |
| Y | SW | TNCWS | 101-1000 | 358 | 0.1291 | 0.6287 | 0.0518 | 1.1533 |
| Y | SW | TNCWS | 1001-4100 | 30 | 0.1291 | 0.6287 | 0.0518 | 1.1533 |
| Y | SW | NTNCWS | ≤100 | 114 | 0.1291 | 0.6287 | 0.0518 | 1.1533 |
| Y | SW | NTNCWS | 101-1000 | 172 | 0.1291 | 0.6287 | 0.0518 | 1.1533 |
| Y | SW | NTNCWS | 1001-4100 | 33 | 0.1291 | 0.6287 | 0.0518 | 1.1533 |
| Y | SW | CWS | ≤100 | 484 | 0.1291 | 0.6287 | 0.0518 | 1.1533 |
| Y | SW | CWS | 101-1000 | 2034 | 0.1291 | 0.6287 | 0.0518 | 1.1533 |
| Y | SW | CWS | 1001-4100 | 1699 | 0.0751 | 0.8767 | 0.0518 | 1.1533 |
| Y | GW | TNCWS | ≤100 | 8112 | 0.0860 | 0.8796 | 0.0706 | 1.1961 |
| Y | GW | TNCWS | 101-1000 | 3007 | 0.0860 | 0.8796 | 0.0706 | 1.1961 |
| Y | GW | TNCWS | 1001-4100 | 119 | 0.0860 | 0.8796 | 0.0706 | 1.1961 |
| Y | GW | NTNCWS | ≤100 | 1577 | 0.0860 | 0.8796 | 0.0706 | 1.1961 |
| Y | GW | NTNCWS | 101-1000 | 1780 | 0.0860 | 0.8796 | 0.0706 | 1.1961 |
| Y | GW | NTNCWS | 1001-4100 | 252 | 0.0860 | 0.8796 | 0.0706 | 1.1961 |
| Y | GW | CWS | ≤100 | 3662 | 0.0545 | 0.6387 | 0.0395 | 1.1679 |
| Y | GW | CWS | 101-1000 | 7678 | 0.0545 | 0.6387 | 0.0395 | 1.1679 |
| Y | GW | CWS | 1001-4100 | 2707 | 0.0545 | 0.6387 | 0.0395 | 1.1679 |
| N | GW | TNCWS | ≤100 | 32028 | 0.0423 | 0.7886 | 0.0357 | 1.0588 |
| N | GW | TNCWS | 101-1000 | 10217 | 0.0423 | 0.7886 | 0.0357 | 1.0588 |
| N | GW | TNCWS | 1001-4100 | 217 | 0.0423 | 0.7886 | 0.0357 | 1.0588 |
| N | GW | NTNCWS | ≤100 | 4026 | 0.0423 | 0.7886 | 0.0357 | 1.0588 |
| N | GW | NTNCWS | 101-1000 | 3264 | 0.0423 | 0.7886 | 0.0357 | 1.0588 |
| N | GW | NTNCWS | 1001-4100 | 152 | 0.0423 | 0.7886 | 0.0357 | 1.0588 |
| N | GW | CWS | ≤100 | 3788 | 0.0423 | 0.7886 | 0.0357 | 1.0588 |
| N | GW | CWS | 101-1000 | 3755 | 0.0423 | 0.7886 | 0.0357 | 1.0588 |
| N | GW | CWS | 1001-4100 | 644 | 0.0423 | 0.7886 | 0.0357 | 1.0588 |

Note: Disinfection status ("Disinf?") is from information downloaded from SDWIS/FED. Systems with no indication of disinfection status in SDWIS/FED are assumed to not disinfect.

### 5.2.5.6 Derivation of *α, β* from Maximum Likelihood Estimates

Using the estimates of *a* and *b* presented in Exhibits 5.4 and 5.5, EPA derived *α, β* estimates of the variation in PWSs within groups of systems, and incorporated these into the occurrence model as described in section 5.3.

These parameters are derived from *a* and *b* as follows:

$$\alpha = a / b^2$$

$$\beta = (1- a) / b^2$$

Exhibits 5.10 and 5.11 present the estimates of *α, β* corresponding to estimates of a and b in exhibits 5.8 and 5.9.

## Exhibit 5.10 *α* and *β* Parameter Estimates for RTTC and RPTC

| Disinf? | GW or SW? | Sys Type | Size | No. Systems | αRTTC | βRTTC | αRPTC | βRPTC |
|---|---|---|---|---|---|---|---|---|
| Y | SW | TNCWS | ≤100 | 838 | 0.2584 | 10.9160 | 0.2134 | 1.0593 |
| Y | SW | TNCWS | 101-1000 | 358 | 0.2584 | 10.9160 | 0.2134 | 1.0593 |
| Y | SW | TNCWS | 1001-4100 | 30 | 0.2584 | 10.9160 | 0.2134 | 1.0593 |
| Y | SW | NTNCWS | ≤100 | 114 | 0.1239 | 7.6968 | 0.0916 | 0.7690 |
| Y | SW | NTNCWS | 101-1000 | 172 | 0.1866 | 16.3708 | 0.0916 | 0.7690 |
| Y | SW | NTNCWS | 1001-4100 | 33 | 0.1866 | 16.3708 | 0.0916 | 0.7690 |
| Y | SW | CWS | ≤100 | 484 | 0.3438 | 22.5437 | 0.1733 | 2.5080 |
| Y | SW | CWS | 101-1000 | 2034 | 0.4262 | 44.6323 | 0.1733 | 2.5080 |
| Y | SW | CWS | 1001-4100 | 1699 | 0.3608 | 61.1715 | 0.4633 | 13.0432 |
| Y | GW | TNCWS | ≤100 | 8112 | 0.2584 | 10.9160 | 0.2134 | 1.0593 |
| Y | GW | TNCWS | 101-1000 | 3007 | 0.2584 | 10.9160 | 0.2134 | 1.0593 |
| Y | GW | TNCWS | 1001-4100 | 119 | 0.2584 | 10.9160 | 0.2134 | 1.0593 |
| Y | GW | NTNCWS | ≤100 | 1577 | 0.1239 | 7.6968 | 0.0916 | 0.7690 |
| Y | GW | NTNCWS | 101-1000 | 1780 | 0.1866 | 16.3708 | 0.0916 | 0.7690 |
| Y | GW | NTNCWS | 1001-4100 | 252 | 0.1866 | 16.3708 | 0.0916 | 0.7690 |
| Y | GW | CWS | ≤100 | 3662 | 0.3166 | 19.5656 | 0.1305 | 1.2558 |
| Y | GW | CWS | 101-1000 | 7678 | 0.2709 | 21.5260 | 0.1305 | 1.2558 |
| Y | GW | CWS | 1001-4100 | 2707 | 0.2291 | 29.2616 | 0.2160 | 3.9218 |
| N | GW | TNCWS | ≤100 | 32028 | 0.2989 | 5.9827 | 0.2444 | 0.6323 |
| N | GW | TNCWS | 101-1000 | 10217 | 0.2989 | 5.9827 | 0.2839 | 0.8577 |
| N | GW | TNCWS | 1001-4100 | 217 | 0.2532 | 10.0105 | 1.0318 | 5.0589 |
| N | GW | NTNCWS | ≤100 | 4026 | 0.2236 | 5.8897 | 0.3425 | 0.9152 |
| N | GW | NTNCWS | 101-1000 | 3264 | 0.2142 | 7.5386 | 0.2487 | 0.7689 |
| N | GW | NTNCWS | 1001-4100 | 152 | 0.2142 | 7.5386 | 0.1967 | 0.9541 |
| N | GW | CWS | ≤100 | 3788 | 0.4381 | 13.4851 | 0.2685 | 1.1463 |
| N | GW | CWS | 101-1000 | 3755 | 0.4328 | 15.5951 | 0.2685 | 1.1463 |
| N | GW | CWS | 1001-4100 | 644 | 0.3243 | 13.4282 | 0.2726 | 1.8049 |

Note: Disinfection status ("Disinf?") is from information downloaded from SDWIS/FED. Systems with no indication of disinfection status in SDWIS/FED are assumed to not disinfect.

**Exhibit 5.11 *α* and *β* Parameter Estimates for RTEC and RPEC**

| Disinf? | GW or SW? | Sys Type | Size | No. Systems | αRTEC | βRTEC | αRPEC | βRPEC |
|---|---|---|---|---|---|---|---|---|
| Y | SW | TNCWS | ≤100 | 838 | 0.3266 | 2.2033 | 0.0389 | 0.7129 |
| Y | SW | TNCWS | 101-1000 | 358 | 0.3266 | 2.2033 | 0.0389 | 0.7129 |
| Y | SW | TNCWS | 1001-4100 | 30 | 0.3266 | 2.2033 | 0.0389 | 0.7129 |
| Y | SW | NTNCWS | ≤100 | 114 | 0.3266 | 2.2033 | 0.0389 | 0.7129 |
| Y | SW | NTNCWS | 101-1000 | 172 | 0.3266 | 2.2033 | 0.0389 | 0.7129 |
| Y | SW | NTNCWS | 1001-4100 | 33 | 0.3266 | 2.2033 | 0.0389 | 0.7129 |
| Y | SW | CWS | ≤100 | 484 | 0.3266 | 2.2033 | 0.0389 | 0.7129 |
| Y | SW | CWS | 101-1000 | 2034 | 0.3266 | 2.2033 | 0.0389 | 0.7129 |
| Y | SW | CWS | 1001-4100 | 1699 | 0.0978 | 1.2033 | 0.0389 | 0.7129 |
| Y | GW | TNCWS | ≤100 | 8112 | 0.1112 | 1.1815 | 0.0494 | 0.6496 |
| Y | GW | TNCWS | 101-1000 | 3007 | 0.1112 | 1.1815 | 0.0494 | 0.6496 |
| Y | GW | TNCWS | 1001-4100 | 119 | 0.1112 | 1.1815 | 0.0494 | 0.6496 |
| Y | GW | NTNCWS | ≤100 | 1577 | 0.1112 | 1.1815 | 0.0494 | 0.6496 |
| Y | GW | NTNCWS | 101-1000 | 1780 | 0.1112 | 1.1815 | 0.0494 | 0.6496 |
| Y | GW | NTNCWS | 1001-4100 | 252 | 0.1112 | 1.1815 | 0.0494 | 0.6496 |
| Y | GW | CWS | ≤100 | 3662 | 0.1336 | 2.3176 | 0.0290 | 0.7041 |
| Y | GW | CWS | 101-1000 | 7678 | 0.1336 | 2.3176 | 0.0290 | 0.7041 |
| Y | GW | CWS | 1001-4100 | 2707 | 0.1336 | 2.3176 | 0.0290 | 0.7041 |
| N | GW | TNCWS | ≤100 | 32028 | 0.0680 | 1.5399 | 0.0318 | 0.8602 |
| N | GW | TNCWS | 101-1000 | 10217 | 0.0680 | 1.5399 | 0.0318 | 0.8602 |
| N | GW | TNCWS | 1001-4100 | 217 | 0.0680 | 1.5399 | 0.0318 | 0.8602 |
| N | GW | NTNCWS | ≤100 | 4026 | 0.0680 | 1.5399 | 0.0318 | 0.8602 |
| N | GW | NTNCWS | 101-1000 | 3264 | 0.0680 | 1.5399 | 0.0318 | 0.8602 |
| N | GW | NTNCWS | 1001-4100 | 152 | 0.0680 | 1.5399 | 0.0318 | 0.8602 |
| N | GW | CWS | ≤100 | 3788 | 0.0680 | 1.5399 | 0.0318 | 0.8602 |
| N | GW | CWS | 101-1000 | 3755 | 0.0680 | 1.5399 | 0.0318 | 0.8602 |
| N | GW | CWS | 1001-4100 | 644 | 0.0680 | 1.5399 | 0.0318 | 0.8602 |

Note: Disinfection status ("Disinf?") is from information downloaded from SDWIS/FED. Systems with no indication of disinfection status in SDWIS/FED are assumed to not disinfect.

## 5.3 Predictive Modeling of Occurrence for Systems Serving Up to 4,100 People

For the purpose of understanding both the potential to reduce exposure to potential contamination and for estimating incremental costs that may be incurred under the proposed RTCR, EPA developed a model to predict the changes in TC and EC occurrence over the modeled period with respect to requirements of the proposed RTCR regulatory scenarios. This predictive occurrence model takes into account the requirements of existing regulations (TCR and GWR) and develops estimates of changes in occurrence (as well as the frequency of the new Level 1 and Level 2 assessments and corrective actions) based on requirements of the regulatory alternatives considered under the proposed RTCR. The resulting estimates are used to generate cost estimates that can be compared across the three regulatory scenarios[4] – the current TCR, AIP, and Alternative Analysis, as described in Chapter 7 of this EA.

[4] The three regulatory scenarios considered in the RTCR EA (current TCR, AIP, and Alternative Analysis) are described in detail in Chapter 3 of this EA. Briefly, the AIP and Alternative Analysis both differ from the current TCR primarily in the requirements for Level 1 and Level 2 assessments and corrective actions. They also both differ from the current TCR in terms of sampling frequency; in particular, under the Alternative Analysis systems all sample monthly for an initial period, and the reduced sampling schedule does not allow systems to reduce to annual sampling as under the current TCR.

Potential benefits are compared in terms of changes in the level of occurrence associated with each regulatory scenario, as presented in Chapter 6 of this EA.

The occurrence model focuses on a 30-year period beginning with full implementation of the GWR and covering a 25 year period of analysis from the anticipated proposed RTCR effective date of 2015. The predictive model is implemented as a Monte Carlo simulation and currently uses 10,000 iterations (simulated systems) within each of the 27 type, source and size categories to predict the effects of the current rule and the proposed RTCR over the 30 year modeling period.

The PWSs included in this model are those serving ≤4,100 people primarily because these are the systems that will experience the major changes in monitoring frequency and other requirements that will affect their TC and EC occurrence. For systems serving more than 4,100 people, which will not have significant monitoring frequency changes, a separate analysis was performed based on historical violations rates instead of occurrence rates, as explained in section 5.4 of this EA. The characteristics of all systems are described in section 4.3 of this EA.

### 5.3.1 Summary of GWR timing and other relevant factors affecting the TCR and proposed RTCR

The triggered and compliance monitoring requirements of the GWR must be implemented by December 2009, which will be approximately five years prior to the anticipated start of proposed RTCR implementation. Sanitary survey (SS) requirements must be implemented by States by December 2012 for most CWSs, and not until December 2014 for higher performing CWSs and noncommunity CWSs. To account for the impact of the GWR requirements on baseline occurrence for those public water systems using a ground water source (and serving ≤4,100 people), the model performs a number of adjustments to the existing data prior to its application in the model as a baseline.

Using 2007 SDWIS data (as explained in section 4.3.1 of this EA) as a pre-GWR inventory baseline, EPA applies probabilities of $P_{sample}$ and $P_{well}$[5] to determine the number of nondisinfecting systems that are expected to move to a disinfecting status in the five years prior to proposed RTCR implementation. EPA derived $P_{sample}$ and $P_{well}$ [6] in the GWR EA (November 2006); $P_{well}$ is a fixed value of 21.58 percent, and $P_{sample}$ is drawn from a beta distribution with a range of alpha and beta estimates having a median value of 5.8 percent and an expected value of 12.4 percent. Incorporation of these factors into the model ensures that the corrective actions anticipated to be performed under the GWR are reflected in reduced occurrence for systems implementing them. In the predictive model, simulated

[5] The term "$P_{well}$" refers to the probability that a randomly selected well across the United States will test positive for a virus or pathogenic indicator species, such as EC, in its source water; "Psample" is the probability that given a contaminated well, a random sample at the well will eventually test positive for a virus or indicator. The GWR EA (November 2006), and in particular the Baseline Chapter of that document, provide a detailed explanation of the analysis used to generate these pwell and psample estimates.

[6] Under the GWR, GWSs with less than 4 log of treatment for viruses must sample their source for a fecal indicator (e.g., *E. coli*) when they incur a TC+ under the TCR (or RTCR). If sample water is fecal indicator positive, the system must implement remedial actions at the source, one of which achieves disinfection to 4 log for viruses.

nondisinfecting ground water systems having a TC+ take one or more source water samples per the GWR triggered monitoring requirement and test the samples for fecal indicator presence (*E. coli* is assumed to be the fecal indicator used). The probability of observing an EC+ is based on the product of $P_{well}$ and the $P_{sample}$ value drawn for that system. If positive, a second assessment is done to determine if that system implements disinfection or a nondisinfection corrective action based on the estimated proportion choosing those options as defined in the GWR EA for the various types and sizes of systems.

The simulation model keeps track of systems that begin as nondisinfecting ground water systems at the start of the analysis period and elect disinfection at some point during the 30 year period from the start of GWR implementation through the end of the analysis period for the proposed RTCR. The inventory for disinfecting and nondisinfecting ground water systems are then adjusted accordingly for subsequent stages of the analysis.

Exhibit 4.4 in Chapter 4 of this EA presents the baseline inventory of ground water systems by disinfection status at the beginning of proposed RTCR implementation. Those GWSs achieving < 4 log disinfection must sample at their source. Sampling costs were accounted for in the GWR EA (USEPA 2006a).

Throughout the modeling period, GW systems that do not currently disinfect or add disinfection may have EC+ source water samples and implement a nondisinfection corrective action. These corrective actions were assumed to have the same effectiveness as a Level 2 corrective action since it is similar in nature to an acute violation under TCR. As shown in Exhibit 5.12, baseline occurrence probability will be significantly reduced as a function of movement from non-disinfecting to disinfecting status. If the system performs a nondisinfecting corrective action, the model assumes that occurrence is 0% for the remainder of that year and 2 full years after that. The modeling then assumes the P values are only 25 percent of baseline values for an additional five years. The model assumes the system returns to baseline P values after that (approximately) seven year period.

Additionally, the model incorporates the effect of SSs performed under the GWR starting in 2010, which is estimated by EPA to reduce TC and EC occurrence by 10 percent for the remainder of the period of analysis. For transient PWSs, which are on a 5 year SS cycle under the GWR, this reduction is initially applied to 20 percent of qualifying systems for each of the 5 years from 2010 to 2015, the anticipated effective date of the proposed RTCR. For CWSs, which are on a 3 year cycle, this reduction is initially applied to 1/3 of systems annually from 2010 to 2012.

A third GWR requirement that the model incorporates is compliance monitoring by the subset of GWSs that disinfect, for which the model assigns a 10 percent reduction in occurrence throughout the period of analysis. Therefore, the disinfecting GWSs are expected to ultimately have TC and EC occurrence rates that are 81 percent of the values derived from the occurrence distributions (from 90 percent for sanitary surveys x 90 percent for compliance monitoring). These adjustments in the model to account for the GWR are shown in the diagrams of Exhibit 5.13.

A fourth and last GWR requirement that the model incorporates is the use of one of the required four repeat sample for source water testing for GWSs serving ≤1,000 people for EC following a case of a TC positive assay for the current TCR. By comparison, under the

AIP, Alternative Analysis, and the current TCR scenarios for systems serving >1,000 people, only 3 repeat samples are required and in these cases an additional, separate sample is used for the GWR source water EC assay.

Although the actual reductions in occurrence that will result from implementation of GWR requirements may differ from what is predicted based on assumptions used in this model, this would not affect the analysis of how the AIP and Alternative Analysis perform relative to each other since both would be affected similarly. Section 5.3.3.1 describes the various sources of uncertainty considered in this EA, and presents analysis of key variables.

### 5.3.2 Summary of Predictive Model

As the following sections 5.3.2.1 – 5.3.3.3 describe, EPA modeled the expected trend in occurrence of TC and EC positive assays over a pre-RTCR monitoring period of 5 years to capture applicable GWR effects, and a subsequent 25-year period of analysis for the three regulatory scenarios of the proposed RTCR: continuation of the current TCR, AIP, and Alternative Analysis[7]. EPA sought to capture the changes in occurrence from the implementation of the GWR requirements, and in turn, the implementation of requirements of the AIP and the Alternative Analysis. Along with changes in TC and EC occurrence, the model predicts behavioral changes: the number of Level 1 and Level 2 assessments (and associated number of Level 1 or Level 2 corrective actions) to be performed and further resulting adjustments to occurrence, and changes in sampling regimens as systems qualify for reduced monitoring requirements.

#### 5.3.2.1 Approach to Estimating Parameters

As described in section 5.2 of this chapter, EPA derived Beta distributions, characterized by their $\alpha, \beta$ parameters, based on analyses of the variation in TC and EC positive assays that occurred both between and within 27 groupings of systems (as defined in section 5.2.); the resulting $\alpha, \beta$ parameter estimates are presented in Exhibits 5.10 and 5.11. Based on an initial analysis, the 27 groups were condensed into the groups presented in sections 5.2.5.1.4 and 5.2.5.2.3. The final Beta distributions were based on these groupings to make the best use of the available data, combining groups of systems when their occurrence levels were similar to avoid having small sample size in any of the final groups.

#### 5.3.2.2 Description of Predictive Model

For the 30 year analysis period used in this model, all systems are assumed to sample based on the sampling regimen applicable either under the current TCR or the proposed RTCR regulatory option to their system type, size, type of water source, and whether or not the system disinfects. (A list of these categories and the number of systems and associated estimates of population served are provided in Chapter 4 (Baseline Analysis) of this EA.) Assumptions for the percentage of systems on monthly, quarterly, or annual monitoring upon proposed RTCR implementation are shown in Exhibit 5.14a – 5.14c.

[7] Footnote 5 describes the three regulatory scenarios.

The model applies the baseline occurrence associated with each category of PWS at the start of each modeled year (i.e., the level of occurrence in 2005 under the current TCR updated to reflect GWR implementation, as described in section 5.3.1). The TC and EC occurrence (both from routine and repeat samples) are modeled by combining the Beta distributions described previously with the binomial distribution to simulate TC and EC positives (i.e., successes) given the number of samples (i.e., trials) each month.

For each simulated system in the model, a probability (P) is drawn randomly for each of the RTTC, RPTC, RTEC, and RPEC types of samples from their corresponding Beta distributions. These P values are used in the binomial distribution, together with the number of samples taken each month, to predict the number of TC and EC positives. In the absence of any corrective actions being performed under the proposed RTCR, those P values remain the same for that system for the entire modeling period. If TC or EC occurrence is found that results in a specific corrective action under the proposed RTCR, temporary reductions in these P values are made as described further below. In addition, when a system has a TC+ in a particular month that does not result in a specific corrective, it is assumed, (for both the current TCR and the proposed RTCR), that the RTTC rate in the following month will change to reflect either some worsening of the problem or an improvement based on some other actions taken that are not part of the specific corrective actions being considered. This was carried out by selecting at random a different RTTC P value in those months following a TC+ without explicit corrective actions being initiated. EPA determined that this approach provided overall results for the current TCR that approximated what was observed in the 6-Year Review data and SDWIS data as discussed in the section 5.3.3.2 (model validation) below.

Based on the level of sampling and the occurrence estimate for each system, the model determines what percentage of these systems will be required to take an action (i.e., perform a Level 1 or Level 2 assessment in response to exceeding the threshold of multiple TC positive assays or one EC positive assay, and, in some cases, implement a corrective action).

For the percentage of systems predicted to require a Level 1 assessment following the requisite number of TC positives, the model assumes that 10 percent will find and address the exact source of the problem under the AIP and Alternative Analysis. This represents an incremental increase over the number expected to implement such a corrective action under the current TCR. The model assumes that systems that successfully identify and correct the problem will not have a positive assay for the remainder of that year and then for 1 additional full year after the assessment, after which the probability of a TC or EC positive will be 50 percent that of the baseline P value for 3 additional years. After this period of reduced occurrence, the model assumes that occurrence for this subset of systems returns to the baseline level assigned by system from P values initially drawn for that system from the Beta distributions (Exhibit 4.8).[8]

Similarly, for the systems predicted to require a Level 2 assessment based on observing the requisite number of TC and/or EC positives, under the AIP and Alternative Analysis an incremental 10 percent more of those systems than under the current TCR are

[8] These assumptions of efficacy of corrective actions expected to be taken were tested in a sensitivity analysis as described in section 5.3.3.1 of this chapter.

assumed to then implement a Level 2 corrective action. For those systems, the model assumes that occurrence is 0% for the remainder of that year and then for 2 full years after that. The modeling then assumes the P values are only 25 percent of baseline values for an additional five years. The model assumes the system returns to baseline P values after that (approximately) seven year period.

For nondisinfecting GWSs, a percentage of positive samples are estimated (from the Pwell and Psample parameters) to be related to a source water problem. All nondisinfecting GW systems having a sourcewater EC positive[9] will implement either disinfection or a nondisinfecting corrective action. Those adding disinfection essentially get removed from the nondisinfecting category and are reassigned to the disinfecting category for the remainder of the modeling period. For the portion of this subset that are assumed to implement a nondisinfecting corrective action[10], the systems are assumed to experience the same reduced occurrence levels of those addressing Level 2 Assessments (i.e., 0% occurrence for 2 years, then occurrence at 25 percent of baseline for an additional 5 years). As Chapter 7 of this EA explains, no costs are attributed to the proposed RTCR for any corrective action related to source water EC occurrence since such costs have already been considered in the GWR EA (November 2006).

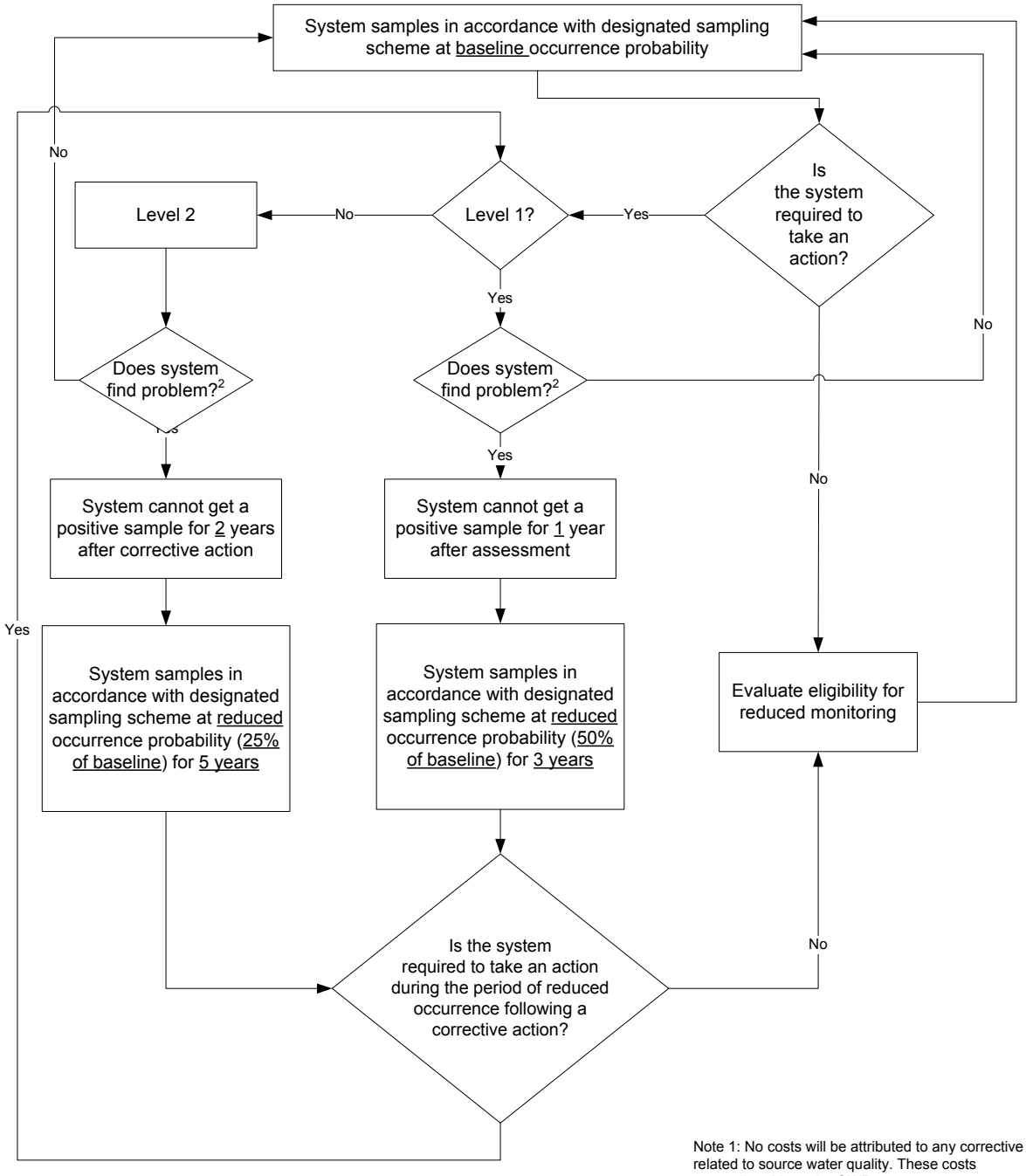
Outputs generated from this predictive simulation model include the following for each modeled year: occurrence of TC and EC positive assays; the number of Level 1 and Level 2 assessments conducted, and the number of Level 1 and Level 2 corrective actions implemented. Based on these model outputs and the criteria for reduced monitoring, the model also provides estimates of the number of systems on monthly, quarterly, and annual sampling regimes (and by implication the number of routine, repeat, and additional routine samples taken annually) under the AIP and Alternative Analysis. These requirements are detailed in Chapter 3 of this EA.

The results of the proposed RTCR predictive occurrence model are presented in the following section, 5.3.3.

[9] The model determines source water occurrence separately based on the subset of nondisinfecting GW systems, as described in section 5.3.1 (2nd paragraph) of this chapter.

[10] Estimates of the number of systems implementing a nondisinfection corrective action are taken from Exhibit 6.21b ("Estimated Distribution of Source Water Contamination Corrective Actions") of the GWR EA (USEPA 2006x). The GWR EA estimated the number of systems choosing a nondisinfection corrective action based on a range; the high end was the percentage of CWS entry points employing disinfection at that time by system size , and the low end was assumed to be 10% based on discussions with State representatives.

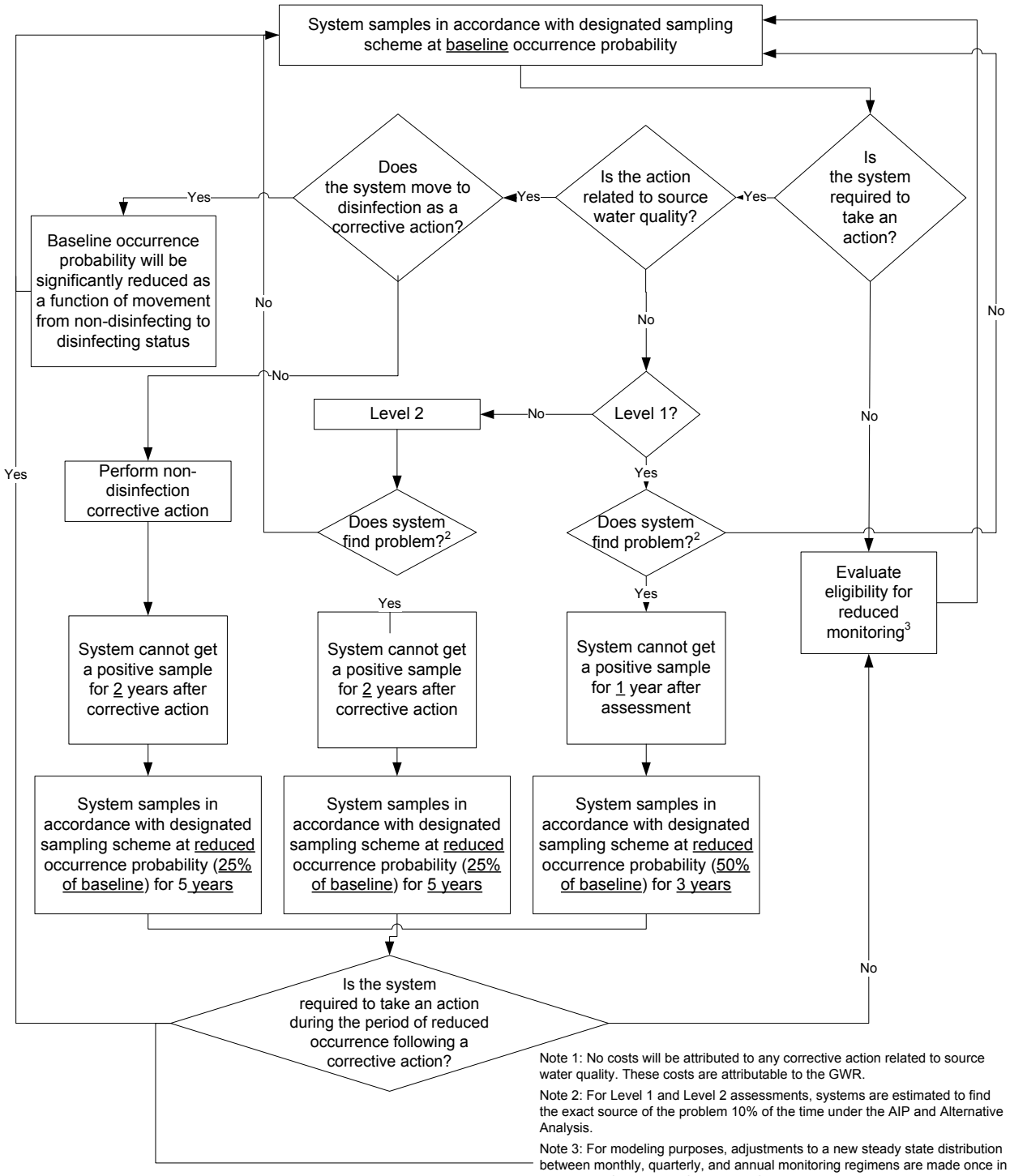
## Exhibit 5.12 Simulated Impact – Surface Water Systems




Note 1: No costs will be attributed to any corrective action related to source water quality. These costs are attributable to the GWR.

Note 2: For Level 1 and Level 2 assessments, systems are estimated to find the exact source of the problem 10% of the time under the AIP and Alternative Analysis.

## Exhibit 5.13 Decision Tree for Occurrence Analysis – Ground Water Systems Serving ≤4,100




Note 1: No costs will be attributed to any corrective action related to source water quality. These costs are attributable to the GWR.

Note 2: For Level 1 and Level 2 assessments, systems are estimated to find the exact source of the problem 10% of the time under the AIP and Alternative Analysis.

Note 3: For modeling purposes, adjustments to a new steady state distribution between monthly, quarterly, and annual monitoring regimens are made once in the period of analysis - after 5 years for noncommunity water systems and 3 years for community water systems. This time frame is meant to correspond with a full sanitary survey cycle for these systems based on the Agency's assumption that States may want to coordinate their evaluation of PWS eligibility for reduced monitoring with the sanitary survey cycle.

## Exhibit 5.14a Percent of GW Systems on M/Q/A Monitoring by System Category under Current TCR (Baseline)

| Disinf? | GW or SW? | Sys Type | Size | No. Systems | Monthly | Quarterly | Annual |
|---|---|---|---|---|---|---|---|
| Y | GW | TNCWS | ≤100 | 13,558 | 4.8% | 62.9% | 32.3% |
| Y | GW | TNCWS | 101-1000 | 6,014 | 7.9% | 66.9% | 25.2% |
| Y | GW | TNCWS | 1001-4100 | 269 | 100.0% | 0.0% | 0.0% |
| Y | GW | NTNCWS | ≤100 | 2,904 | 19.3% | 64.6% | 16.0% |
| Y | GW | NTNCWS | 101-1000 | 3,621 | 18.5% | 66.7% | 14.7% |
| Y | GW | NTNCWS | 1001-4100 | 542 | 100.0% | 0.0% | 0.0% |
| Y | GW | CWS | ≤100 | 6,255 | 88.6% | 13.4% | 0.0% |
| Y | GW | CWS | 101-1000 | 12,762 | 88.5% | 11.5% | 0.0% |
| Y | GW | CWS | 1001-4100 | 5,405 | 100.0% | 0.0% | 0.0% |
| N | GW | TNCWS | ≤100 | 46,642 | 4.8% | 62.9% | 32.3% |
| N | GW | TNCWS | 101-1000 | 15,224 | 7.8% | 66.8% | 25.3% |
| N | GW | TNCWS | 1001-4100 | 348 | 100.0% | 0.0% | 0.0% |
| N | GW | NTNCWS | ≤100 | 5,913 | 19.3% | 64.6% | 16.0% |
| N | GW | NTNCWS | 101-1000 | 4,710 | 18.5% | 66.7% | 14.8% |
| N | GW | NTNCWS | 1001-4100 | 270 | 100.0% | 0.0% | 0.0% |
| N | GW | CWS | ≤100 | 5,806 | 86.6% | 13.4% | 0.0% |
| N | GW | CWS | 101-1000 | 5,597 | 88.6% | 11.4% | 0.0% |
| N | GW | CWS | 1001-4100 | 1,038 | 100.0% | 0.0% | 0.0% |

Notes: 1) Estimates for the size categories presented in this exhibit are produced from weighted averages from the following size categories (and corresponding estimates of proportion sampling monthly, quarterly, or annually): ≤100, 101-500, 501-1000, 1001-2500, 2501-3300, and 3301-4100. Water Type includes disinfection status, as informed by information from SDWIS-FED. 2) Systems with no indication of disinfection ("Disinf?") status in SDWIS-FED were assumed to not disinfect.

## Exhibit 5.14b Percent of GW Systems on M/Q/A Monitoring by System Category under AIP

| Disinf? | GW or SW? | Sys Type | Size | No. Systems | Monthly | Quarterly | Annual |
|---|---|---|---|---|---|---|---|
| Y | GW | TNCWS | ≤100 | 13,558 | 6.4% | 79.8% | 13.8% |
| Y | GW | TNCWS | 101-1000 | 6,014 | 6.5% | 80.2% | 13.2% |
| Y | GW | TNCWS | 1001-4100 | 269 | 100.0% | 0.0% | 0.0% |
| Y | GW | NTNCWS | ≤100 | 2,907 | 3.0% | 89.7% | 7.3% |
| Y | GW | NTNCWS | 101-1000 | 3,621 | 2.5% | 91.2% | 6.3% |
| Y | GW | NTNCWS | 1001-4100 | 542 | 100.0% | 0.0% | 0.0% |
| Y | GW | CWS | ≤100 | 6,255 | 88.6% | 13.4% | 0.0% |
| Y | GW | CWS | 101-1000 | 12,762 | 88.5% | 11.5% | 0.0% |
| Y | GW | CWS | 1001-4100 | 5,405 | 100.0% | 0.0% | 0.0% |
| N | GW | TNCWS | ≤100 | 46,643 | 14.4% | 71.8% | 13.8% |
| N | GW | TNCWS | 101-1000 | 15,225 | 16.1% | 70.5% | 13.4% |
| N | GW | TNCWS | 1001-4100 | 348 | 100.0% | 0.0% | 0.0% |
| N | GW | NTNCWS | ≤100 | 5,919 | 15.1% | 77.6% | 7.3% |
| N | GW | NTNCWS | 101-1000 | 4,711 | 11.4% | 82.0% | 6.5% |
| N | GW | NTNCWS | 1001-4100 | 270 | 100.0% | 0.0% | 0.0% |
| N | GW | CWS | ≤100 | 5,806 | 86.6% | 13.4% | 0.0% |
| N | GW | CWS | 101-1000 | 5,597 | 88.6% | 11.4% | 0.0% |
| N | GW | CWS | 1001-4100 | 1,038 | 100.0% | 0.0% | 0.0% |

Notes: 1) Estimates for the size categories presented in this exhibit are produced from weighted averages from the following size categories (and corresponding estimates of proportion sampling monthly, quarterly, or annually): ≤100, 101-500, 501-1000, 1001-2500, 2501-3300, and 3301-4100. 2) Monthly, quarterly, and annual estimates are for years 11-30. 3) Systems with no indication of disinfection ("Disinf?") status in SDWIS-FED were assumed to not disinfect.

**Exhibit 5.14c Percent of GW Systems on M/Q/A Monitoring by System Category under Alternative Analysis**

| Disinf? | GW or SW? | Sys Type | Size | No. Systems | Monthly | Quarterly | Annual |
|---|---|---|---|---|---|---|---|
| Y | GW | TNCWS | ≤100 | 13,558 | 13.0% | 87.0% | 0.0% |
| Y | GW | TNCWS | 101-1000 | 6,014 | 12.4% | 85.7% | 0.0% |
| Y | GW | TNCWS | 1001-4100 | 269 | 100.0% | 0.0% | 0.0% |
| Y | GW | NTNCWS | ≤100 | 2,907 | 4.9% | 95.1% | 0.0% |
| Y | GW | NTNCWS | 101-1000 | 3,621 | 4.5% | 95.5% | 0.0% |
| Y | GW | NTNCWS | 1001-4100 | 542 | 100.0% | 0.0% | 0.0% |
| Y | GW | CWS | ≤100 | 6,255 | 88.6% | 13.4% | 0.0% |
| Y | GW | CWS | 101-1000 | 12,762 | 88.5% | 11.5% | 0.0% |
| Y | GW | CWS | 1001-4100 | 5,405 | 100.0% | 0.0% | 0.0% |
| N | GW | TNCWS | ≤100 | 46,642 | 25.6% | 74.4% | 0.0% |
| N | GW | TNCWS | 101-1000 | 15,224 | 25.9% | 74.1% | 0.0% |
| N | GW | TNCWS | 1001-4100 | 348 | 100.0% | 0.0% | 0.0% |
| N | GW | NTNCWS | ≤100 | 5,919 | 22.9% | 77.1% | 0.0% |
| N | GW | NTNCWS | 101-1000 | 4,710 | 17.2% | 82.8% | 0.0% |
| N | GW | NTNCWS | 1001-4100 | 270 | 100.0% | 0.0% | 0.0% |
| N | GW | CWS | ≤100 | 5,806 | 86.6% | 13.4% | 0.0% |
| N | GW | CWS | 101-1000 | 5,597 | 88.6% | 11.4% | 0.0% |
| N | GW | CWS | 1001-4100 | 1,038 | 100.0% | 0.0% | 0.0% |

Notes: 1) Estimates for the size categories presented in this exhibit are produced from weighted averages from the following size categories (and corresponding estimates of proportion sampling monthly, quarterly, or annually): ≤100, 101-500, 501-1000, 1001-2500, 2501-3300, and 3301-4100. 2) Monthly, quarterly, and annual estimates are for years 11-30. 3) Systems with no indication of disinfection ("Disinf?") status in SDWIS-FED were assumed to not disinfect.

### 5.3.3 Predictive Model Results

As noted above, the predictive model produces output that includes national estimates of sampling and occurrence in PWSs across the United States, the resulting Level 1 and Level 2 assessments, Level 1 and Level 2 corrective actions, and reductions in sampling and occurrence for these systems in accordance with requirements under the current TCR, the AIP, and the Alternative Analysis. In Exhibits 5.15 through 5.20, output includes model years 1 – 30. Years 1 - 5 are summed for the current TCR only and represent the period prior to commencing monitoring under proposed RTCR and inclusive of GWR implementation. Year 5 in the analysis (also for just the current TCR) is shown separately because it represents the baseline year, reflecting conditions just prior to the start of monitoring under the proposed RTCR. Years 6 through 30 are summed for the current TCR and for each proposed RTCR regulatory alternative because they represent the 25-year period of analysis following promulgation of the proposed RTCR.

For all types of PWSs, the model output is presented for the current TCR (i.e., the baseline regulatory scenario for this EA), the AIP, and the Alternative Analysis. The current TCR without the GWR is provided mainly as a point of comparison with the Six-Year Data

for 2005. The TCR with the GWR serves as the appropriate baseline for comparisons with the proposed RTCR.

EPA's general expectations for TC hit rates for GW systems shown in Exhibits 5.21 – 5.23 are as follows:

1. Years 1-5 show a decline in the TC rates due to GWR implementation effects. These include expected reductions in RTTC rates for all GW systems related to sanitary surveys, for disinfecting GW systems related to the compliance monitoring, and for non-disinfecting GW systems due to corrective actions taken following a source water EC+ sample.

2. Years 6-30 reflect additional reductions in TC rates due to the proposed RTCR (AIP or Alternative Analysis). The proposed RTCR rates are below TCR with the GWR primarily because of the additional corrective actions (Level 1 and Level 2).

3. Differences seen between the AIP and Alternative Analysis are due to different proportions that are performing monthly, quarterly, or annual monitoring. In general, more monitoring (e.g., more doing monthly than quarterly, or more doing quarterly than annual) – *in conjunction with the corrective actions* – will push the TC hit rates down (more samples are expected to be taken and fewer are likely to be positive).

In Exhibit 5.11 (Community GW systems), the graph shows the expected drop in TC rates for years 1-5 as the GWR is implemented prior to proposed RTCR. There are only very small differences between AIP and Alternative Analysis, and these are not observable on this graph. Most of these systems are already on monthly sampling, so there is no significant difference seen between AIP and Alternative Analysis in years 6-8 when those few not on monthly and it is assumed that the monthly/quarterly distribution after this period returns to that under the current TCR. Therefore, the TC hit rates over the 25 year period for these systems show not difference between AIP and Alternative Analysis.

In Exhibit 5.22 (NTNC GW systems), the graph also shows the expected drop in TC rates for years 1-5 as the GWR is implemented prior to the proposed RTCR. For the years 6-10 the hit rates decrease due to the implementation of corrective actions, and the rates for the Alternative Analysis are lower than for AIP since all systems are on monthly sampling under the Alternative Analysis for this period (having the effect of reducing the hit rate as noted above). However, from year 11 to 30, the monthly, quarterly, annual distributions for these two are essentially the same and therefore the TC hit rates are the same over this period.

In Exhibit 5.23 (TNCW GW), the graph also shows the expected drop in TC rates for years 1-5 as the GWR is implemented prior to the proposed RTCR. For the years 6-10, the TC rates drop further as the Level 1 and Level 2 corrective actions are begun for both the AIP and the Alternative Analysis. The hit rate is markedly lower during this period for the Alternative Analysis where all are on monthly, whereas the majority of systems under the AIP are on quarterly. Beginning in Year 11, there is a shift under the AIP toward more systems on monthly than under the TCR, and more systems on quarterly than under the TCR, and hence the overall TC rate drops. For the AIP, beginning in Year 11, a substantial portion move from the all monthly sampling during years 6-10 to quarterly sampling (but none on

annual). Because there are slightly more systems under Alternative Analysis that end up on monthly and quarterly than under the AIP for years 11-30, the TC hit rate for the AIP is slightly higher than for the Alternative Analysis.

EPA's general expectations for TC hit rates for SW systems shown in Exhibits 5.24 – 5.26 are as follows:

1. The current TCR is provided as the baseline for comparison with both the 6-Year data and the proposed RTCR (the GWR does not apply to surface water system and therefore is not included in these graphs).

2. Because all surface water systems of all sizes do monthly monitoring (no systems are on quarterly or annual monitoring), the AIP and Alternative Analysis versions of the proposed RTCR are identical.

3. The reductions in TC rates shown by the proposed RTCR relative to the current TCR therefore reflect the effects of the Level 1 and Level 2 corrective actions only.

In Exhibit 5.24 (Community SW systems), the graph shows little difference in TC hit rates between the current TCR and proposed RTCR. This is primarily because the current TC hit rate is very low for these systems, and therefore there are few Level 1 and Level 2 corrective actions implemented over the 25 year period.

In Exhibit 5.25 (NTNC SW), the graph shows that there is a slightly greater difference between the current TCR and the proposed RTCR for the NTNC SW systems than for the community systems. This is because the baseline TC hit rate is slightly higher and, therefore, more Level 1 and Level 2 corrective actions are performed.

In Exhibit 5.26 (TNCWS SW), the graph indicates that this group exhibits the most significant change between the current TCR and proposed RTCR since the baseline TC hit rate for these systems is approximately twice that of the NTNC and community systems.

Complete results from the model (by year and system size) and additional graphs are shown in Appendices A and B, respectively. The implications of these results in terms of changes in the level of risk associated with contamination of PWS water supplies are discussed in Chapter 6 of this EA. Model results are used to calculate national estimates of incremental costs for PWSs in Chapter 7 of this EA.

## Exhibit 5.15 Ground Water Community Water System Model Output Cumulative Endpoints

| Period | Regulatory Scenario | Reg Rout Samples | Reg Rout Samples if Addls Taken | Repeat Samples | GWR Samples | RTTC+ | RPTC+ | RTEC+ | RPEC+ | GWR EC + | Non-acutes | Acutes | L1 Assmt | L2 Assmt | L1 CA | L2 CA | GWR ND | GWR D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| sum years 6-30 | TCR No GWR | 12,749,629 | 1,046,502 | 812,230 | 0 | 213,979 | 108,508 | 9,905 | 3,992 | 0 | 54,531 | 5,296 | 32,936 | 26,891 | 0 | 0 | 0 | 0 |
| | TCR | 12,795,800 | 861,989 | 586,160 | 92,214 | 175,630 | 65,356 | 7,168 | 2,202 | 1,086 | 38,750 | 3,475 | 25,680 | 16,541 | 0 | 0 | 923 | 162 |
| | AIP | 13,016,022 | 19,017 | 457,547 | 77,257 | 152,516 | 46,637 | 6,072 | 1,459 | 950 | 29,149 | 2,536 | 22,047 | 9,637 | 2,236 | 968 | 802 | 149 |
| | Alt | 13,100,381 | 16,728 | 461,043 | 77,947 | 153,681 | 47,081 | 6,121 | 1,475 | 956 | 29,412 | 2,559 | 22,210 | 9,761 | 2,252 | 980 | 805 | 151 |
| sum years 1-5 | TCR No GWR | 2,550,468 | 250,391 | 194,369 | 0 | 51,210 | 26,038 | 2,408 | 876 | 0 | 13,072 | 1,252 | 7,890 | 6,435 | 0 | 0 | 0 | 0 |
| | TCR | 2,556,095 | 187,274 | 126,884 | 20,449 | 38,110 | 15,380 | 1,671 | 541 | 365 | 8,721 | 867 | 5,472 | 4,116 | 0 | 0 | 256 | 109 |
| Year 5 | TCR No GWR | 510,096 | 41,561 | 32,190 | 0 | 8,487 | 4,221 | 407 | 124 | 0 | 2,174 | 194 | 1,313 | 1,055 | 0 | 0 | 0 | 0 |
| | TCR | 511,727 | 35,310 | 23,910 | 3,817 | 7,179 | 2,730 | 296 | 96 | 49 | 1,613 | 154 | 1,037 | 730 | 0 | 0 | 35 | 14 |

## Exhibit 5.16 Ground Water Nontransient Noncommunity Water System Model Output Cumulative Endpoints

| Period | Regulatory Scenario | Reg Rout Samples | Reg Rout Samples if Addls Taken | Repeat Samples | GWR Samples | RTTC+ | RPTC+ | RTEC+ | RPEC+ | WR EC | Non-acutes | Acutes | L1 Assmt | L2 Assmt | L1 CA | L2 CA | GWR ND | GWR D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| sum years 6-30 | TCR No GWR | 2,660,714 | 325,245 | 262,864 | 0 | 67,750 | 60,013 | 3,617 | 2,450 | 0 | 24,540 | 2,816 | 13,501 | 13,854 | 0 | 0 | 0 | 0 |
| | TCR | 2,668,566 | 265,923 | 175,500 | 41,013 | 54,978 | 33,764 | 2,497 | 1,155 | 521 | 17,592 | 1,616 | 10,675 | 8,532 | 0 | 0 | 435 | 86 |
| | AIP | 2,551,856 | 90,619 | 141,469 | 36,240 | 47,156 | 24,655 | 2,052 | 849 | 474 | 13,482 | 1,238 | 9,731 | 4,989 | 976 | 496 | 394 | 80 |
| | Alt | 3,333,950 | 72,725 | 179,281 | 46,388 | 59,760 | 30,858 | 2,598 | 1,081 | 565 | 16,888 | 1,545 | 11,423 | 7,010 | 1,144 | 713 | 469 | 97 |
| sum years 1-5 | TCR No GWR | 532,245 | 64,803 | 52,280 | 0 | 13,479 | 11,948 | 726 | 446 | 0 | 4,884 | 551 | 2,688 | 2,747 | 0 | 0 | 0 | 0 |
| | TCR | 533,143 | 58,588 | 38,661 | 9,120 | 12,125 | 8,096 | 593 | 301 | 162 | 3,984 | 394 | 2,296 | 2,082 | 0 | 0 | 104 | 58 |
| Year 5 | TCR No GWR | 106,447 | 13,014 | 10,523 | 0 | 2,714 | 2,386 | 156 | 80 | 0 | 967 | 111 | 532 | 545 | 0 | 0 | 0 | 0 |
| | TCR | 106,699 | 11,197 | 7,343 | 1,754 | 2,310 | 1,468 | 101 | 51 | 19 | 754 | 69 | 441 | 382 | 0 | 0 | 14 | 6 |

## Exhibit 5.17 Ground Water Transient Noncommunity Water System Model Output Cumulative Endpoints

| Period | Regulatory Scenario | Reg Rout Samples | Reg Rout Samples if Addls Taken | Repeat Samples | GWR Samples | RTTC+ | RPTC+ | RTEC+ | RPEC+ | WR EC | Non-acutes | Acutes | L1 Assmt | L2 Assmt | L1 CA | L2 CA | GWR ND | GWR D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| sum years 6-30 | TCR No GWR | 7,474,727 | 1,725,880 | 1,456,782 | 0 | 366,435 | 382,461 | 17,296 | 13,456 | 0 | 147,325 | 15,265 | 83,987 | 78,551 | 0 | 0 | 0 | 0 |
| | TCR | 7,520,449 | 1,467,021 | 975,353 | 255,696 | 308,572 | 220,243 | 12,764 | 7,311 | 3,327 | 108,414 | 10,469 | 68,710 | 50,159 | 0 | 0 | 2,816 | 511 |
| | AIP | 9,549,370 | 691,763 | 960,724 | 276,113 | 320,241 | 187,657 | 12,509 | 6,280 | 3,616 | 99,636 | 8,981 | 69,832 | 38,785 | 6,797 | 3,816 | 3,022 | 593 |
| | Alt | 14,719,673 | 532,745 | 1,359,646 | 391,942 | 453,215 | 259,998 | 17,604 | 8,160 | 4,537 | 138,378 | 12,002 | 87,956 | 62,424 | 8,471 | 6,201 | 3,882 | 654 |
| sum years 1-5 | TCR No GWR | 1,501,526 | 353,694 | 298,150 | 0 | 74,985 | 77,933 | 3,593 | 2,650 | 0 | 30,127 | 3,135 | 17,376 | 15,886 | 0 | 0 | 0 | 0 |
| | TCR | 1,503,144 | 323,447 | 215,804 | 56,709 | 68,301 | 53,602 | 3,075 | 1,993 | 1,085 | 24,456 | 2,672 | 15,002 | 12,126 | 0 | 0 | 753 | 333 |
| Year 5 | TCR No GWR | 300,267 | 70,203 | 59,157 | 0 | 14,879 | 15,294 | 736 | 547 | 0 | 5,917 | 658 | 3,436 | 3,139 | 0 | 0 | 0 | 0 |
| | TCR | 300,802 | 60,493 | 40,290 | 10,646 | 12,767 | 9,598 | 536 | 293 | 201 | 4,520 | 444 | 2,756 | 2,207 | 0 | 0 | 147 | 54 |

## Exhibit 5.18 Surface Water Community Water System Model Output Cumulative Endpoints

| Period | Regulatory Scenario | Reg Rout Samples | Reg Rout Samples if Addls Taken | Repeat Samples | GWR Samples | RTTC+ | RPTC+ | RTEC+ | RPEC+ | GWR EC + | Non-acutes | Acutes | L1 Assmt | L2 Assmt | L1 CA | L2 CA | GWR ND | GWR D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| sum years 6-30 | TCR | 3,515,731 | 136,353 | 97,536 | 0 | 27,576 | 5,236 | 2,942 | 230 | 0 | 3,290 | 663 | 2,761 | 1,193 | 0 | 0 | 0 | 0 |
| | AIP | 3,564,342 | 0 | 78,109 | 0 | 26,036 | 3,622 | 2,737 | 211 | 0 | 2,471 | 548 | 2,189 | 830 | 217 | 82 | 0 | 0 |
| | Alt | 3,564,342 | 0 | 78,109 | 0 | 26,036 | 3,622 | 2,737 | 211 | 0 | 2,471 | 548 | 2,189 | 830 | 217 | 82 | 0 | 0 |
| sum years 1-5 | TCR | 703,186 | 27,232 | 19,481 | 0 | 5,511 | 1,052 | 597 | 55 | 0 | 656 | 140 | 541 | 255 | 0 | 0 | 0 | 0 |
| Year 5 | TCR | 140,633 | 5,441 | 3,871 | 0 | 1,095 | 196 | 119 | 11 | 0 | 124 | 27 | 102 | 48 | 0 | 0 | 0 | 0 |

## Exhibit 5.19 Surface Water Nontransient Noncommunity Water System Model Output Cumulative Endpoints

| Period | Regulatory Scenario | Reg Rout Samples | Reg Rout Samples if Addls Taken | Repeat Samples | GWR Samples | RTTC+ | RPTC+ | RTEC+ | RPEC+ | WR EC | Non-acutes | Acutes | L1 Assmt | L2 Assmt | L1 CA | L2 CA | GWR ND | GWR D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| sum years 6-30 | TCR | 222,446 | 13,749 | 10,694 | 0 | 2,801 | 1,073 | 362 | 60 | 0 | 450 | 104 | 276 | 278 | 0 | 0 | 0 | 0 |
| | AIP | 225,823 | 0 | 7,744 | 0 | 2,581 | 650 | 334 | 38 | 0 | 312 | 80 | 222 | 170 | 22 | 17 | 0 | 0 |
| | Alt | 225,823 | 0 | 7,744 | 0 | 2,581 | 650 | 334 | 38 | 0 | 312 | 80 | 222 | 170 | 22 | 17 | 0 | 0 |
| sum years 1-5 | TCR | 44,498 | 2,729 | 2,117 | 0 | 555 | 217 | 72 | 11 | 0 | 89 | 22 | 55 | 56 | 0 | 0 | 0 | 0 |
| Year 5 | TCR | 8,900 | 539 | 415 | 0 | 109 | 41 | 12 | 2 | 0 | 17 | 4 | 11 | 11 | 0 | 0 | 0 | 0 |

## Exhibit 5.20 Surface Water Transient Noncommunity Water System Model Output Cumulative Endpoints

| Period | Regulatory Scenario | Reg Rout Samples | Reg Rout Samples if Addls Taken | Repeat Samples | GWR Samples | RTTC+ | RPTC+ | RTEC+ | RPEC+ | WR EC | Non-acutes | Acutes | L1 Assmt | L2 Assmt | L1 CA | L2 CA | GWR ND | GWR D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| sum years 6-30 | TCR | 608,792 | 69,910 | 56,287 | 0 | 14,315 | 9,440 | 1,832 | 506 | 0 | 3,895 | 906 | 2,211 | 2,585 | 0 | 0 | 0 | 0 |
| | AIP | 625,041 | 0 | 36,393 | 0 | 12,131 | 5,089 | 1,612 | 261 | 0 | 2,507 | 581 | 1,737 | 1,351 | 169 | 138 | 0 | 0 |
| | Alt | 625,041 | 0 | 36,393 | 0 | 12,131 | 5,089 | 1,612 | 261 | 0 | 2,507 | 581 | 1,737 | 1,351 | 169 | 138 | 0 | 0 |
| sum years 1-5 | TCR | 121,883 | 14,294 | 11,522 | 0 | 2,931 | 1,923 | 368 | 102 | 0 | 805 | 179 | 448 | 536 | 0 | 0 | 0 | 0 |
| Year 5 | TCR | 24,380 | 2,841 | 2,278 | 0 | 579 | 381 | 73 | 17 | 0 | 159 | 34 | 87 | 106 | 0 | 0 | 0 | 0 |

**Exhibit 5.21 Ground Water Community Water System TC Occurrence**

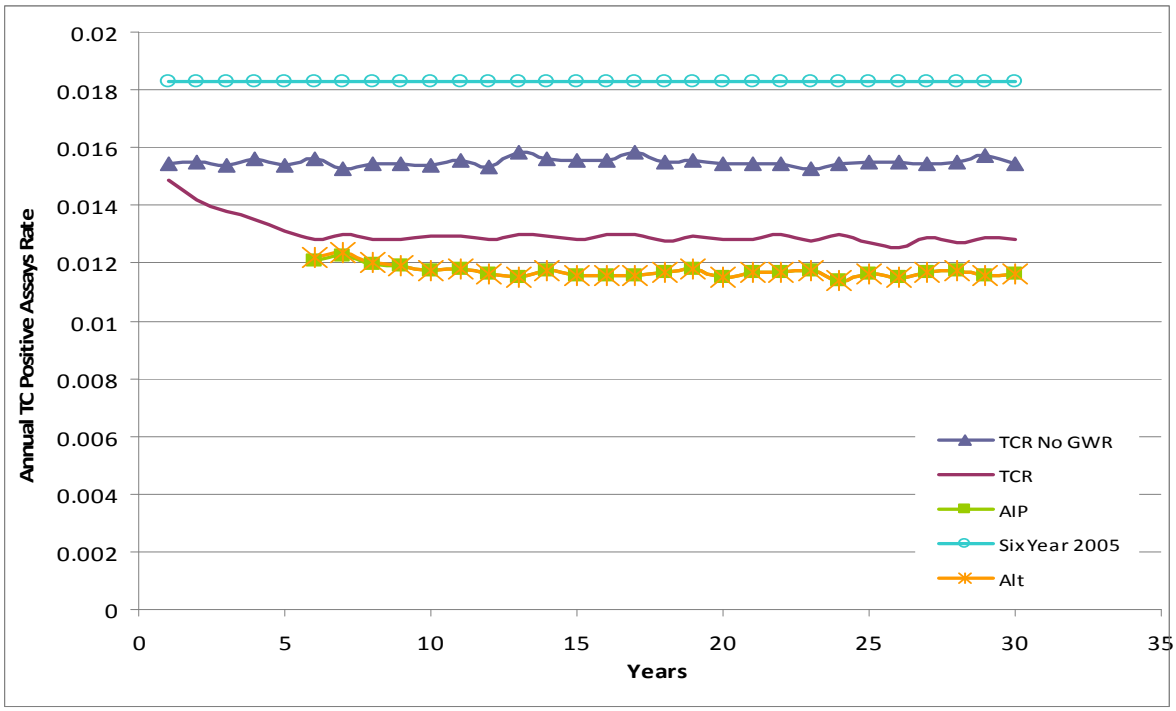
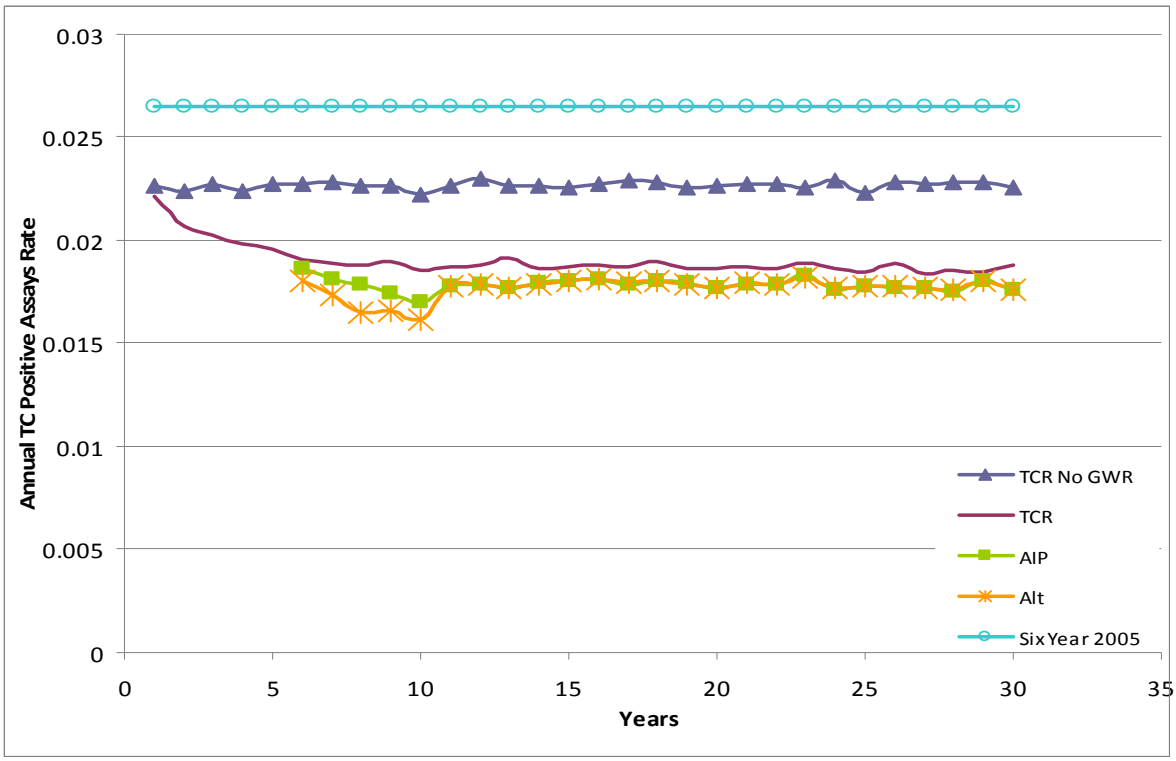
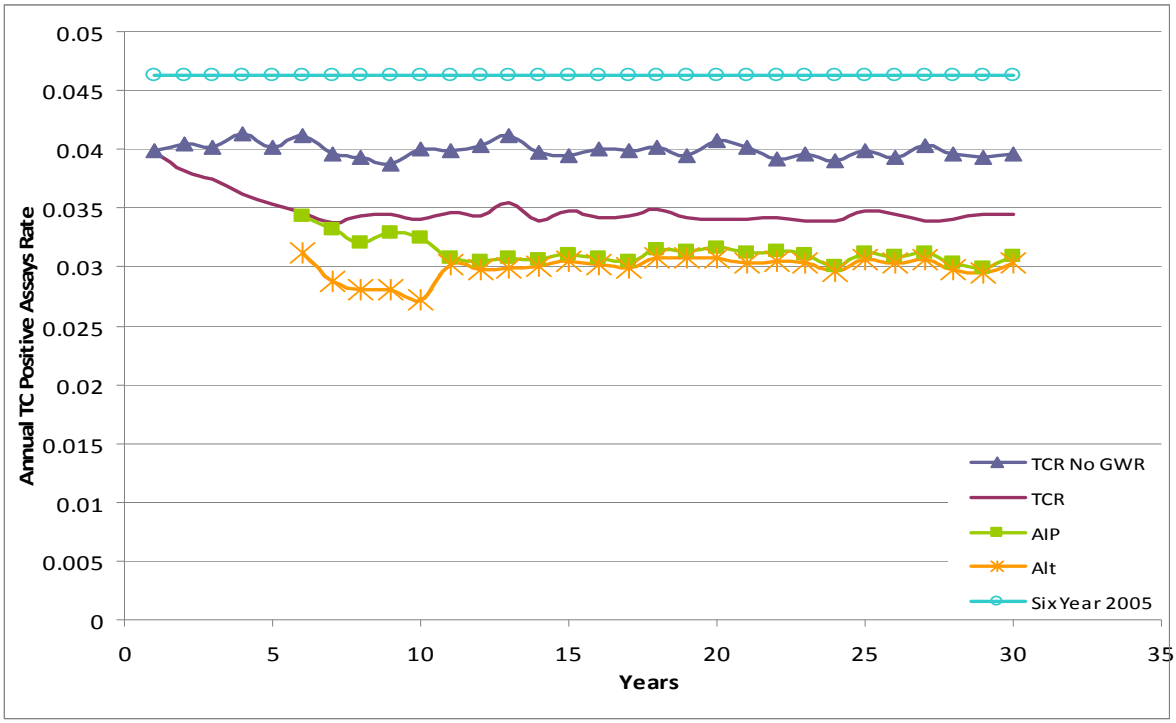



**Exhibit 5.22 Ground Water Nontransient Noncommunity Water System TC Occurrence**



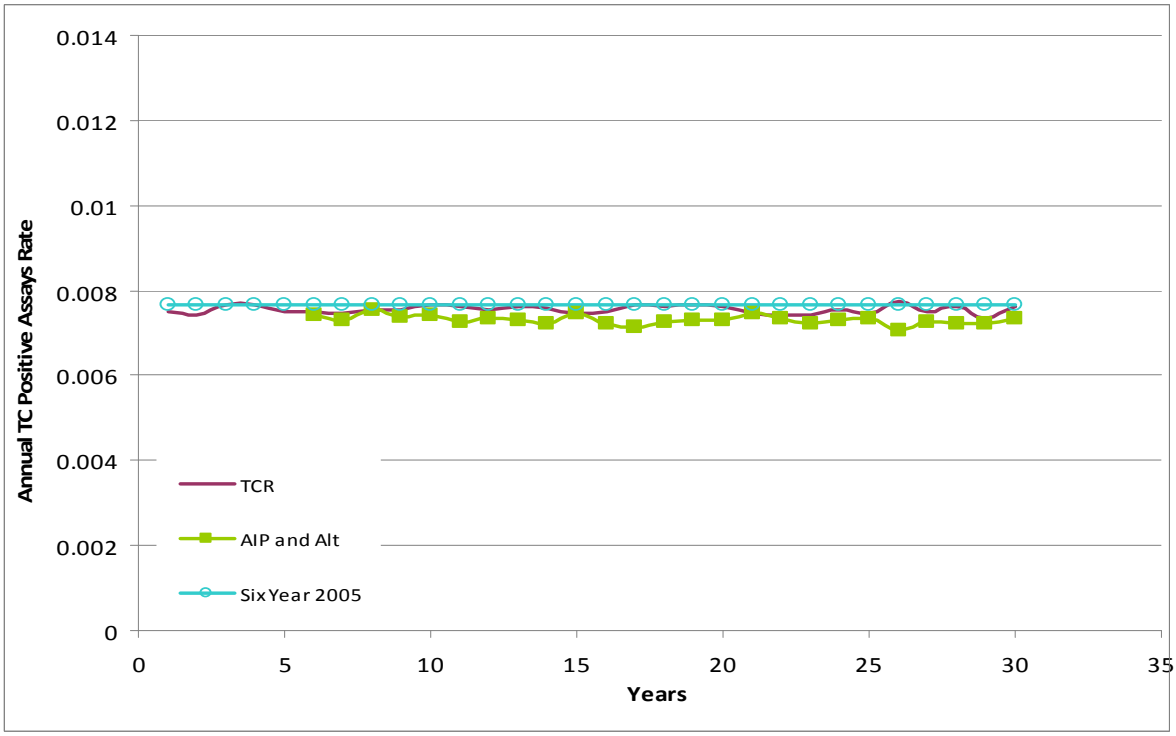

Note: Six Year 2005 TC+ occurrence is representative of all GW NTNCWS. The rate presented may underestimate the occurrence for systems serving 25-4,100 individuals.

## Exhibit 5.23 GW Transient Noncommunity Water System TC Occurrence

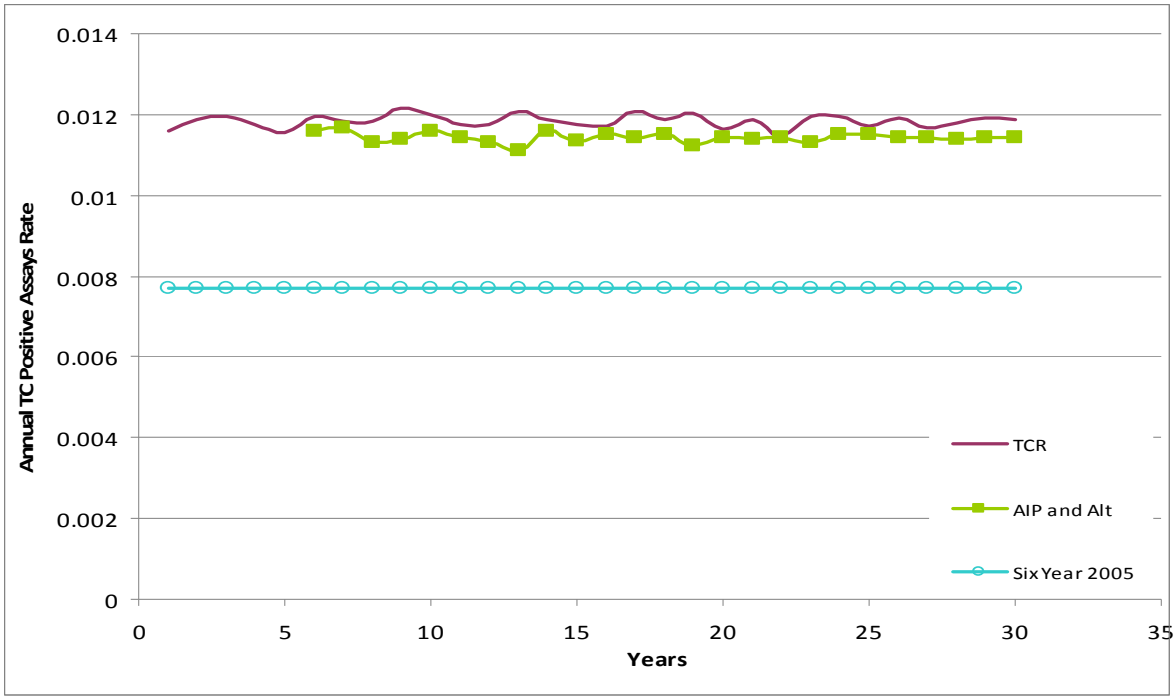



Note: Six Year 2005 TC+ occurrence is representative of all GW TNCWS. The rate presented may underestimate the occurrence for systems serving 25-4,100 individuals.

## Exhibit 5.24 Surface Water Community Water System TC Occurrence

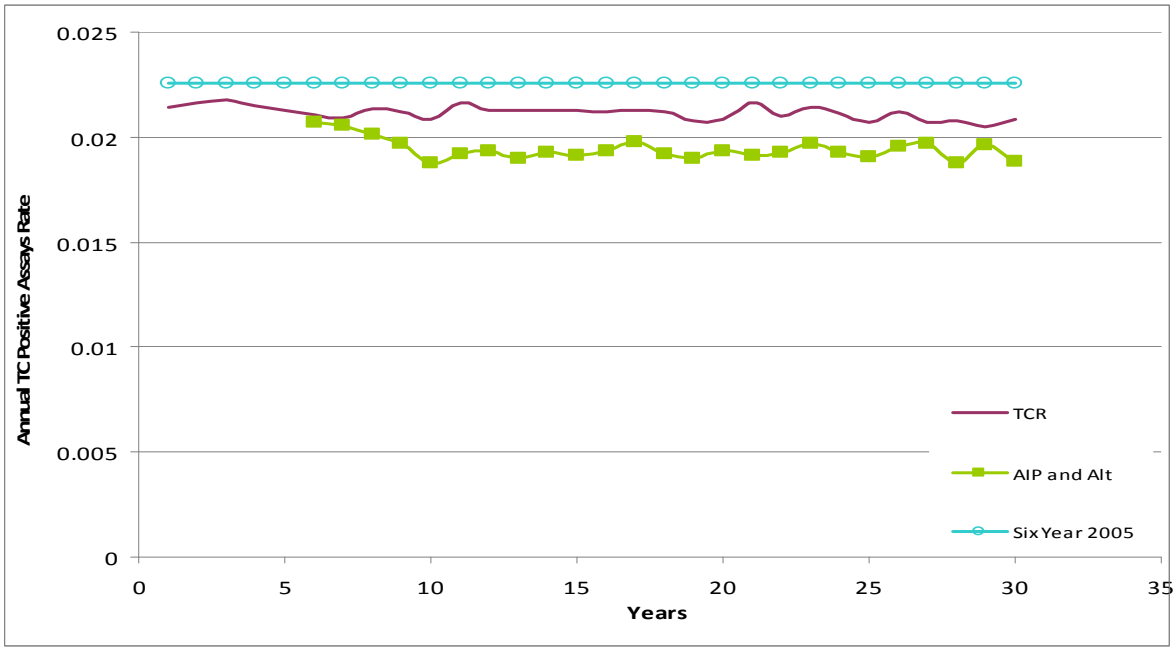


## Exhibit 5.25 SW Nontransient Noncommunity Water System TC Occurrence




Note: Six Year 2005 TC+ occurrence is representative of all SW NTNCWS. The rate presented may underestimate the occurrence for systems serving 25-4,100 individuals.

## Exhibit 5.26 SW Transient Noncommunity Water System TC Occurrence




Note: Six Year 2005 TC+ occurrence is representative of all SW TNCWS. The rate presented may underestimate the occurrence for systems serving 25-4,100 individuals.

#### 5.3.3.1 Uncertainty and Sensitivity Analysis in Occurrence Modeling

There are two primary sources of uncertainty in the proposed RTCR occurrence modeling: uncertainty related to the baseline occurrence model due to the limited data on which the model is based, and uncertainty regarding the frequency and effectiveness of corrective actions on subsequent TC and EC occurrence.

Included in the data limitations is a lack of observed data that reflects implementation of the GWR, which is not available because the GWR requirements will not be implemented until approximately 2010. In the absence of this data, EPA made assumptions regarding the effectiveness of each of the relevant requirements, as described in section 5.3.1 of this chapter. Each of the assumptions related to GWR requirements affects the baseline for this EA, which is the current TCR. The predictive model uses this same baseline (current TCR) to determining the incremental influence of both the AIP and Alternative Analysis requirements on occurrence. Therefore, although the actual reductions in occurrence that will result from implementation of GWR requirements may differ from the assumptions used in this model, this difference would not affect the performance of the AIP and Alternative Analysis relative to each other since any such difference would affect both the AIP and Alternative Analysis similarly. Uncertainties related to other data limitations are discussed in Section 4.2 of this EA.

This section addresses uncertainties in the frequency and effectiveness of corrective actions that the model predicts would be implemented.

As described previously in section 5.3, EPA incorporated the following assumptions (also shown in Exhibits 5.12 and 5.13) for the frequency and effectiveness of the corrective actions that systems will implement following a Level 1 or a Level 2 assessment:

Level 1: Following a Level 1 Assessment in a particular year, 10 percent of the systems performing the assessment will implement corrective actions that will reduce the TC and EC hit rates to 0% for the remainder of that year and for one full year after that. Then, for an additional 3 full years, the TC and EC hit rates for each of these systems will be reduced to 50 percent of their initial values. After that, the TC and EC hit rates return to their initial values.

Level 2: Following a Level 2 Assessment in a particular year, 10 percent of the systems performing the assessment will implement corrective actions that will reduce the TC and EC hit rates to 0% for the remainder of that year and for 2 full years after that. Then, for an additional 5 full years, the TC and EC hit rates for each of these systems will be reduced to 25 percent of their initial values. After that, the TC and EC hit rates return to their initial values.

Also, for the nondisinfecting ground water systems that are found in the simulation to have an EC+ in their source water but do not move to disinfection, the assumptions for the effectiveness of the nondisinfecting corrective actions are identical to those above for Level 2 corrective actions. (However, note that all nondisinfecting systems that discover EC+ source water that do not go to disinfection will implement these corrective actions, not just 10 percent as assumed for systems that already disinfect.)

To assess the influence of these assumptions on the model results, EPA ran the model with alternative assumptions reflecting corrective actions that are less effective (50% of the assumption of efficacy in the primary analysis in this EA) and corrective actions that are more

effective (2x the assumption of efficacy in the primary analysis). Using these alternative assumptions, key model outputs that can serve as proxy indicators of the costs and benefits of the proposed RTCR were compared with those same outputs from the main model assumptions for the AIP.

For the less effective corrective actions, the alternative assumptions for Level 1 and Level 2 are:

Level 1: Following a Level 1 Assessment in a particular year, 5 percent of the systems performing the assessment will implement corrective actions that reduce the TC and EC hit rates to 0% for the remainder of that year only. Then, for an additional 2 full years, the TC and EC hit rates for each of these systems will be reduced to 50 percent of their initial values. After that, the TC and EC hit rates return to their initial values.

Level 2: Following a Level 2 Assessment in a particular year, 5 percent of the systems performing the assessment will implement a corrective action that will reduce the TC and EC hit rates to 0% for the remainder of that year and for 1 full year after that. Then, for an additional 3 full years, the TC and EC hit rates for that system will be reduced to 25 percent of their initial values. After that, the TC and EC hit rates return to their initial values. (These alternative Level 2 durations were also applied to the nondisinfecting GWR corrective actions.)

For the more effective corrective actions, the alternative assumptions for Level 1 and Level 2 are:

Level 1: Following a Level 1 Assessment in a particular year, 20 percent of the systems performing the assessment will implement corrective actions that will reduce the TC and EC hit rates to 0% for the remainder of that year and for two full years after that. Then, for an additional 6 full years, the TC and EC hit rates for that system will be reduced to 50 percent of their initial values. After that, the TC and EC hit rates return to their initial values.

Level 2: Following a Level 2 Assessment in a particular year, 20 percent of the systems performing the assessment will implement corrective actions that will reduce the TC and EC hit rates to 0% for the remainder of that year and for 4 full years after that. Then, for an additional 10 full years, the TC and EC hit rates for that system will be reduced to 25 percent of their initial values. After that, the TC and EC hit rates return to their initial values. (Again, these alternative Level 2 durations were also applied to the nondisinfecting GWR corrective actions.)

A summary of the assumptions used in these alternative analyses is provided in Exhibit 5.27. The analyses were run in the predictive model for the AIP rule for the approximately 60,000 nondisinfecting transient noncommunity groundwater systems serving ≤500 people. The results of the alternative assumptions used in these sensitivity runs are compared with those for the main analysis in Exhibit 5.28.

## Exhibit 5.27 Sensitivity Analysis Assumptions for Frequency and Effectiveness of Corrective Actions following Level 1 or 2 Assessments

| | Type of Assessment | % performing Corrective Action after a Level 1 or Level 2[1] Assessment | Reduced hit rate immediately after Corrective Action | Period of time for reduced hit rate[2] | Additional period of reduced hit rate (yrs) | Reduction from initial hit rate in additional reduced period |
|---|---|---|---|---|---|---|
| Predictive Model | Level 1 | 10% | 0% | Remainder + 1 | 3 | 50% |
| | Level 2 | 10% | 0% | Remainder + 2 | 5 | 25% |
| Sensitivity 1 - Less Effective | Level 1 | 5% | 0% | Remainder only | 2 | 50% |
| | Level 2 | 5% | 0% | Remainder + 1 | 3 | 25% |
| Sensitivity 2 - More effective | Level 1 | 20% | 0% | Remainder + 2 | 6 | 50% |
| | Level 2 | 20% | 0% | Remainder + 4 | 10 | 25% |

Notes: [1]Level 2 assumptions for the predictive model and sensitivity analyses are also applied to 100% of nondisinfecting ground water systems that incur an EC+ in the source water and implement a nondisinfecting corrective action.

[2]"Remainder" refers to the balance of the year following implementation of a corrective action.

**Exhibit 5.28 Effect of Alternative Corrective Action Assumptions on AIP Model Results for Nondisinfecting TNCWS Serving ≤500 People over 25 Years**

| | RTTC+ | RPTC+ | RTEC+ | RPEC+ | GWR EC+ | Non-Acutes | Acutes | L1 Assess-ments | L2 Assess-ments | L1 CAs | L2 CAs |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Less Effective Corrective Actions | 286,433 | 193,446 | 10,302 | 6,582 | 4,188 | 98,705 | 8,681 | 64,084 | 43,302 | 3,223 | 2,165 |
| Main Model for AIP | 270,601 | 168,065 | 9,535 | 5,328 | 3,513 | 88,115 | 7,436 | 60,749 | 34,802 | 5,869 | 3,413 |
| More Effective Corrective Actions | 252,044 | 128,465 | 8,308 | 3,509 | 2,946 | 71,719 | 5,314 | 55,205 | 21,828 | 10,930 | 4,480 |
| **Relative changes factors for alternative corrective action effectiveness assumptions compared with main model assumptions:** | | | | | | | | | | | |
| Less Effective Corrective Actions | 1.06 | 1.15 | 1.08 | 1.24 | 1.19 | 1.12 | 1.17 | 1.05 | 1.24 | 0.55 | 0.63 |
| More Effective Corrective Actions | 0.93 | 0.76 | 0.87 | 0.66 | 0.84 | 0.81 | 0.71 | 0.91 | 0.63 | 1.86 | 1.31 |

The results of these runs indicate that, as would be expected, the less frequent and less effective corrective active assumptions result in finding more TC+ and EC+ samples and having more nonacute and acute events than the main assumptions. In addition, more L1 and L2 assessments are conducted in response to those additional events, but fewer corrective actions are performed.

Conversely, the more frequent and more effective corrective action assumptions lead to fewer TC+, EC+, nonacute, and acute events occurring and more L1 and L2 CAs being performed.

The fewer number of TC+, EC+, nonacute and acute events occurring with the more effective corrective action assumptions relative to the main assumptions can be viewed as events that are prevented from occurring as a result of those more effective actions relative to the main assumptions. Similarly, the increased number of those events occurring with the less effective corrective actions can be viewed as those that would not be prevented relative to the main assumptions.

It is important to note that this sensitivity analysis is not intended as a rigorous, quantitative comparison of the alternatives but rather as a general indicator of the magnitude and direction of change in these outputs relative to the magnitude and direction of the changes in these inputs. Broadly speaking, since the input assumptions reflect changes that are approximately (but not preciscely) factors of one half (0.5) and twice (2.0) the main assumptions,

outputs that have ratios that are substantially less than 0.5 or significantly greater than 2.0 could be considered to be very sensitive to the changes in these input assumptions. Outputs with ratios that fall between 0.5 and 2.0 can be considered to be much less sensitive to the input assumptions (ratios equal to or approaching 1.0 indicate that the output is essentially not sensitive to changes in the input assumptions.

As indicated in Exhibit 5.28, all of the ratios for outputs fall within the 0.5 to 2.0 range, indicating that these outputs do not appear to be highly sensitive to these alternative assumptions for corrective action effectiveness.

The outputs that could serve as indicators of the sensitivity of the benefits on these assumptions are the numbers of TC and EC positives predicted as well as the number of nonacute and acute events predicted. These outputs have factors that generally fall well within the 0.5 to 2.0 range, indicating that these changes in the inputs assumptions appear to result in relatively small changes in these outputs. That is, the less effective corrective action assumptions do not appear to result in "missing" a disproportionately greater number of these events, and the more effective corrective actions appear to result in "finding" a disproportionately greater number of these events.

The outputs that could serve as indicators of the sensitivity of the costs on these assumptions are the Level 1 and Level 2 assessments and the Level 1 and Level 2 corrective actions. The assessments have ratios that, like those for the outputs like TC and EC positives, are generally well within the 0.5 to 2.0 range. The ratios for the corrective actions are also within this range but tend to be close to the 0.5 and 2.0 values. The input assumptions used for the number of corrective actions done is directly proportional to the number of assessments performed, so these changes would be expected to be more directly linear (closer to 0.5 or 2.0). For example, using the more effective corrective action assumptions, the number of TC+ samples predicted is reduced, as are the number of Level 1 assessments that are related to TC+ events. The ratio for Level 1 assessments under the more effective corrective action assumptions is 0.91, indicating that under these assumptions systems will only need to perform about 90% of the Level 1 assessments over the 25 year period (that is, the others are prevented by the more effective actions). However, this set of alternative assumptions includes a factor that 20% of those doing Level 1 assessments go on to do Level 1 corrective actions (which is twice the 10% assumption in the main model). Therefore, we see that almost twice as many of these Level 1 corrective actions are performed as under the main model. More specifically, the ratio for Level 1 corrective actions is 1.86 which is approximately 2x the 0.91 ratio noted for Level 1 assessment predicted.

#### 5.3.3.2 Model Validation

There are a limited number of ways in which outputs from the predictive model for the systems serving ≤4,100 people can be compared to observed data. EPA has identified two types of analyses where comparisons can be made. One of these is to compare the model's predictions of average annual TC and EC hit rates for various types and sizes of systems under the current TCR against the observations from the 2005 Six-Year review data used to parameterize the model. The other is to compare the model's predictions for the current TCR against those reported in SDWIS.

It is necessary to keep in mind that the primary purpose of the modeling effort for the systems serving ≤4,100 people was not to exactly match observed results but to provide a framework for comparing relative changes across regulatory options. Nevertheless, it is important that, to the extent possible, the model outputs for conditions that can be checked against observed data compare reasonably well. This provides assurance that the model is operating in a manner that is a reasonable simulation of how the TCR currently operates and provides some measure of confidence that the relative changes seen for the rule options are meaningful.

Exhibit 5.29 provides a comparison of the model results for average annual TC+ assays under the current TCR with those observed in the 2005 Six-Year Review data.

**Exhibit 5.29 Comparison of TC+ Occurrence Predicted as a 25-Year Annual Average under the Current TCR with 2005 Six-Year Review Data**

| System Type | Source Water | Population Served | Six Year review | | Predictive Model – Current TCR | |
|---|---|---|---|---|---|---|
| | | | TC (% Positive) | EC (% Positive) | TC (% Positive) | EC (% Positive) |
| CWS | GW | ≤500 | 2.28% | 0.14% | 1.94% | 0.09% |
| | | 501-1,000 | 1.52% | 0.04% | 1.52% | 0.07% |
| | | 1,001-4,100 | 1.07% | 0.02% | 0.92% | 0.05% |
| | SW | ≤500 | 1.22% | 0.06% | 1.10% | 0.15% |
| | | 501-1,000 | 0.70% | 0.05% | 0.90% | 0.12% |
| | | 1,001-4,100 | 0.57% | 0.03% | 0.57% | 0.04% |
| TNCWS | GW | ≤500 | 4.73% | 0.19% | 4.10% | 0.19% |
| | | 501-1,000 | 4.30% | 0.06% | 3.92% | 0.20% |
| | | >1,000 | 2.02% | 0.03% | 2.14% | 0.13% |
| | SW | ≤500 | 2.35% | 0.31% | 2.13% | 0.27% |
| | | 501-1,000 | 3.63% | 0.00% | 2.19% | 0.28% |
| | | >1,000 | 0.61% | 0.00% | 1.98% | 0.26% |
| NTNCWS | GW | ≤500 | 2.98% | 0.07% | 2.51% | 0.13% |
| | | 501-1,000 | 1.74% | 0.02% | 1.90% | 0.10% |
| | | >1,000 | 1.35% | 0.04% | 1.46% | 0.09% |
| | SW | ≤500 | 1.03% | 0.15% | 1.25% | 0.16% |
| | | 501-1,000 | 0.64% | 0.00% | 1.08% | 0.14% |
| | | >1,000 | 0.08% | 0.00% | 1.03% | 0.13% |

Source: Six Year Review data from Exhibit 4.8. Current TCR data from model output.

The comparison of the predictive model output with the Six-Year data for TC and EC hit rates (based on number of TC routine samples taken) shows a reasonable concordance for all types and sizes of systems. With respect to the differences, the model does not appear to be either systematically overestimating or underestimating the hit rates compared with the Six-Year Review data (except where the Six-Year Review data show 0% observed, the model does show a small positive hit rate).

Exhibit 5.30 provides a comparison of the model results of the average annual nonacute and acute violations for the current TCR with the same metric from SDWIS data for 2007 3Q. (Violations data are downloaded from SDWIS on an annual basis in 3rd quarter only, and are not disaggregated by month.)

**Exhibit 5.30 Comparison of SDWIS Data for Nonacute and Acute Violations with Predictive Model Annual Results for the Current TCR**

| | Non-Acute Violations | | | | Acute Violations | | | |
|---|---|---|---|---|---|---|---|---|
| | GW | | SW | | GW | | SW | |
| | SDWIS | No GWR (Avg) | SDWIS | No GWR (Avg) | SDWIS | No GWR (Avg) | SDWIS | No GWR (Avg) |
| **CWSs** | | | | | | | | |
| ≤500 | 1,787 | 1,670 | 62 | 66 | 124 | 162 | 70 | 15 |
| 501-1,000 | 147 | 192 | 18 | 18 | 7 | 19 | 1 | 4 |
| 1,001-4,100 | 350 | 319 | 88 | 48 | 10 | 31 | 2 | 7 |
| **Totals** | **2,284** | **2,181** | **168** | **132** | **141** | **212** | **73** | **27** |
| **NTNCWSs** | | | | | | | | |
| ≤500 | 881 | 843 | 3 | 13 | 65 | 96 | - | 3 |
| 501-1,000 | 46 | 61 | 2 | 2 | 8 | 7 | - | 0 |
| 1,001-4,100 | 51 | 78 | 9 | 3 | 4 | 9 | - | 1 |
| **Totals** | **978** | **982** | **14** | **18** | **77** | **113** | **-** | **4** |
| **TNCWSs** | | | | | | | | |
| ≤500 | 4,027 | 5,631 | 41 | 139 | 391 | 580 | 13 | 32 |
| 501-1,000 | 118 | 147 | 3 | 7 | 8 | 17 | - | 2 |
| 1,001-4,100 | 44 | 114 | 1 | 10 | 3 | 14 | - | 2 |
| **Totals** | **4,189** | **5,893** | **45** | **156** | **402** | **611** | **13** | **36** |
| **Grand Total** | **7,452** | **9,056** | **227** | **305** | **619** | **935** | **86** | **67** |

Source: SDWIS 2007 Q3 download and predictive model results.

The comparison of the predictive model output with the SDWIS violations data also shows a reasonable concordance for all types and sizes of systems. EPA believes there is some under-counting of actual violations in SDWIS due to monitoring and reporting violations, which the model does not incorporate. Therefore, EPA generally expects that the model would generate results that are higher than the observed SDWIS data. While a comparison reveals that some of the model estimates are lower than the SDWIS statistics, the grand totals for the model are consistently higher than those observed in SDWIS with the exception of acute violations in surface water systems.

## 5.4 *Occurrence Analysis for Systems Serving More Than 4,100*

Systems serving populations greater than 4,100 are similar in many ways. In particular, these systems generally have the resources and are managed in ways that lead to generally high expectations for the integrity of the distribution systems and the ability of the systems to identify and correct problems. Some of these differences include:

- Most of these systems operate with certified operators, are staffed at all operating times, and usually operate continuously.
- Most of these systems have operators for whom producing water is their primary activity, and most are owned by communities rather than run as ancillary businesses.
- As a result of the professional operations, these systems have a much lower incidence of monitoring and reporting violations.
- Many of these systems have their own laboratory for analyzing samples.
- Many more of these systems disinfect their water.
- These systems take at least 5 samples per month. The largest of these systems take hundreds of samples per month. There is no requirement for "additional next month samples" as there is for smaller systems because these larger systems already take at least 5 samples each month.
- The proposed monitoring changes to the TCR mostly do not affect systems serving more than 4,100 people; all are on monthly monitoring, and corrective actions are expected to show only a small incremental increase under the proposed RTCR.

As a result of these shared characteristics, systems serving populations of more than 4,100 are expected to be similar to each other and grouping systems for the purpose of forecasting occurrence is acceptable.

For the largest systems serving >33,000 people, EPA did not quantify changes in occurrence as a result of the proposed RTCR, based on knowledge of the industry and an assumption that their operations will not change significantly under the proposed RTCR.

For systems serving 4,101 to 33,000 people, EPA assumes that occurrence may change based on the extra distribution system awareness created by: a) applying the Level 1 and Level 2 assessments in lieu of prior assessments that may, in some cases, have been less structured, and b) reporting the assessment results. Therefore, EPA developed a simple model to predict the effects of the AIP and Alternative Analysis on PWSs and to compare those data to the baseline data predicted for the current TCR.

### 5.4.1 Model

As input for this model, EPA first considered 2005 TCR data on sampling and positive assays that was compiled under the Six-Year Review, as was incorporated into the model for smaller systems. However, as discussed in section 4.2.2.3, detailed sampling results from the Six Year Review data were not representative of the universe of larger systems. EPA instead found that violations data were adequately representative of the universe of large systems. This input to the model, based on 2007 SDWIS data (see Exhibit 4.8), was used to predict violations that systems serving 4,101 to 33,000 will incur in the 25-year period of proposed RTCR analysis, and

the behavioral responses (Level 1 and Level 2 assessments and Corrective Actions) that would result.

The model uses the same stratification as smaller systems – by system type (CWS, TNCWS, or NTNCWS) and by water source (SW or GW). For each system in a given category the model applies the 2007 SDWIS violations rate for that category. This rate is applied in each year of the analysis, that is, the violations rate is a constant in the model. For estimating the effects on violations under the AIP and Alternative Analysis, the model estimates that 10 percent of the assessments resulting from violations will result in corrective actions, the same value applied in the analysis of the smaller systems. This 10% represents the incremental increase in efficacy of addressing the root cause of system contamination under either regulatory alternative for the proposed RTCR as compared to the current TCR.

Repeat samples are reduced in the model under both regulatory alternatives from the current TCR. They are calculated by applying the ratios of repeat samples to routine samples from the smaller system category (serving 1,001 – 4,100 people) to the systems serving 4,101 to 33,000 people.

The model output, shown in section 5.4.2, includes predictions of the annual number of violations (non-acute = Level 1, acute = Level 2), and the number of Level 1 and Level 2 assessments and corrective actions to be implemented.

### 5.4.2 Model Results

Exhibits 5.31 through 5.33 present model results for systems serving 4,101 – 33,000 people.

### 5.4.3 Model Uncertainty

As explained in section 5.4.1 of this chapter, EPA does not expect systems serving >4,100 people to experience changes in routine monitoring or repeat sample regimes under the regulatory alternatives; furthermore, all systems in this size range are on monthly sampling. Therefore, the larger systems model was relatively simple compared to that for the smaller systems, and EPA did not develop an uncertainty analysis for this group of systems.

## Exhibit 5.31 Results for Systems Serving 4,101 to 33,000 People – Current TCR

| System Size (Population Served) | Number of Routine Monitoring Samples | Number of Routine with Additional Monitoring Samples | Number of Repeat Monitoring Samples | Systems Performing Additional Annual Site Inspections | Number of Non-Acute Violation Assessments (Single Violations) | Number of Corrective Actions (based on Single Non-Acute Violation Assessment) | Number of Acute Violation Assessments | Number of Non-Acute Violation Assessments (Multiple Violations) | Number of Corrective Actions (based on Acute and Multiple Non-Acute Violation Assessments) |
|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I |
| **Community Water Systems (CWSs) - SW** | | | | | | | | | |
| 4,101-33,000 | 10,636,296 | - | 186,781 | - | 2,626 | - | 265 | - | - |
| **Community Water Systems (CWSs) - GW** | | | | | | | | | |
| 4,101-33,000 | 9,145,224 | - | 219,336 | - | 4,164 | - | 138 | - | - |
| **Nontransient Noncommunity Water Systems (NTNCWSs) - SW** | | | | | | | | | |
| 4,101-33,000 | 50,424 | - | 1,632 | - | 51 | - | - | - | - |
| **Nontransient Noncommunity Water Systems (NTNCWSs) - GW** | | | | | | | | | |
| 4,101-33,000 | 153,648 | - | 5,847 | - | 104 | - | 9 | - | - |
| **Transient Noncommunity Water Systems (TNCWSs) - SW** | | | | | | | | | |
| 4,101-33,000 | 40,656 | - | - | - | 4 | - | - | - | - |
| **Transient Noncommunity Water Systems (TNCWSs) - GW** | | | | | | | | | |
| 4,101-33,000 | 156,288 | - | 8,835 | - | 101 | - | 7 | - | - |

Note: Estimates of the number of assessments and corrective actions are incremental increases in activity expected to occur under the AIP and Alternative Analysis relative to the current TCR; therefore, estimates of "zero" under the current TCR do not reflect that no corrective actions occur, but that there is no increase expected in activity from the current baseline.

## Exhibit 5.32 Results for Systems Serving 4,101 to 33,000 People – AIP

| System Size (Population Served) | Number of Routine Monitoring Samples | Number of Routine with Additional Monitoring Samples | Number of Repeat Monitoring Samples | Systems Performing Additional Annual Site Inspections | Number of Level 1 Assessments | Number of Corrective Actions (based on Level 1 Assessments) | Number of Level 2 Assessments (based on Acute Violations) | Number of Level 2 Assessments (based on Non-Acute Violations) | Number of Corrective Actions (based on Level 2 Assessments) |
|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I |
| **Community Water Systems (CWSs) - SW** | | | | | | | | | |
| 4,101-33,000 | 10,636,296 | - | 175,689 | - | 2,626 | 263 | 265 | - | 27 |
| **Community Water Systems (CWSs) - GW** | | | | | | | | | |
| 4,101-33,000 | 9,145,224 | - | 192,934 | - | 4,164 | 416 | 138 | - | 14 |
| **Nontransient Noncommunity Water Systems (NTNCWSs) - SW** | | | | | | | | | |
| 4,101-33,000 | 50,424 | - | 1,395 | - | 51 | 5 | - | - | - |
| **Nontransient Noncommunity Water Systems (NTNCWSs) - GW** | | | | | | | | | |
| 4,101-33,000 | 153,648 | - | 4,811 | - | 104 | 10 | 9 | - | 1 |
| **Transient Noncommunity Water Systems (TNCWSs) - SW** | | | | | | | | | |
| 4,101-33,000 | 40,656 | - | 2,040 | - | 4 | 0 | - | - | - |
| **Transient Noncommunity Water Systems (TNCWSs) - GW** | | | | | | | | | |
| 4,101-33,000 | 156,288 | - | 7,015 | - | 101 | 10 | 7 | - | 1 |

Note: Estimates of the number of assessments and corrective actions are incremental increases in activity expected to occur under the AIP and Alternative Analysis relative to the current TCR; therefore, estimates of "zero" under the current TCR do not reflect that no corrective actions occur, but that there is no increase expected in activity from the current baseline.

## Exhibit 5.33 Results for Systems Serving 4,101 to 33,000 People – Alternative Analysis

| System Size (Population Served) | Number of Routine Monitoring Samples | Number of Routine with Additional Monitoring Samples | Number of Repeat Monitoring Samples | Systems Performing Additional Annual Site Inspections | Number of Level 1 Assessments | Number of Corrective Actions (based on Level 1 Assessments) | Number of Level 2 Assessments (based on Acute Violations) | Number of Level 2 Assessments (based on Non-Acute Violations) | Number of Corrective Actions (based on Level 2 Assessments) |
|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I |
| **Community Water Systems (CWSs) - SW** | | | | | | | | | |
| 4,101-33,000 | 10,636,296 | - | 175,689 | - | 2,626 | 263 | 265 | - | 27 |
| **Community Water Systems (CWSs) - GW** | | | | | | | | | |
| 4,101-33,000 | 9,145,224 | - | 192,934 | - | 4,164 | 416 | 138 | - | 14 |
| **Nontransient Noncommunity Water Systems (NTNCWSs) - SW** | | | | | | | | | |
| 4,101-33,000 | 50,424 | - | 1,395 | - | 51 | 5 | - | - | - |
| **Nontransient Noncommunity Water Systems (NTNCWSs) - GW** | | | | | | | | | |
| 4,101-33,000 | 153,648 | - | 4,811 | - | 104 | 10 | 9 | - | 1 |
| **Transient Noncommunity Water Systems (TNCWSs) - SW** | | | | | | | | | |
| 4,101-33,000 | 40,656 | - | 2,040 | - | 4 | 0 | - | - | - |
| **Transient Noncommunity Water Systems (TNCWSs) - GW** | | | | | | | | | |
| 4,101-33,000 | 156,288 | - | 7,015 | - | 101 | 10 | 7 | - | 1 |

Note: Estimates of the number of assessments and corrective actions are incremental increases in activity expected to occur under the AIP and Alternative Analysis relative to the current TCR; therefore, estimates of "zero" under the current TCR do not reflect that no corrective actions occur, but that there is no increase expected in activity from the current baseline.

## 5.5 Summary of Occurrence Analysis Inputs for Benefit and Cost Analyses

As presented in Exhibits 5.15 – 5.20, the model for small systems (serving ≤4,100 people) produces estimates of the number of samples, the number of Level 1 and Level 2 assessments, and the number of Level 1 and Level 2 corrective actions to be implemented under each regulatory scenario. The incremental change in each of these outputs, in particular sampling regimens and implementations of assessments and corrective actions, relative to the baseline (current TCR) under both alternative regulations considered provides the basis for estimating the benefits in the rule, described in Chapter 6 of this EA. Similarly, the cost model applies unit costs to the activities described by these outputs to produce incremental costs for each regulatory scenario, as described in Chapter 7.

For larger systems (serving >4,100 people), incremental costs are estimated for the additional reporting activity required under the rule, as well as a slight increase in the number of corrective actions implemented for systems serving 4,101 – 33,000 people. As described in section 5.4, however, the sampling regimens are not expected to change for these systems, and they'll experience minimal changes in risk and cost, as described in Chapter 6 and Chapter 7 of this EA, respectively.

## References

Gelman, A., J.B. Carlin, H.S. Stern, and D.R. Rubin. 1995. *Bayesian Data Analysis*. London: Chapman & Hall.

Ntzoufras, Ioannis. *Bayesian Modeling Using WinBUGS*. Hoboken: John Wiley & Sons, Inc.

United States Environmental Protection Agency (USEPA). 2006a. *Ground Water Rule Economic Analysis*. Office of Water. EPA 815-R-06-014. October 2006.

United States Environmental Protection Agency (USEPA). 2007. *The Safe Drinking Water Information System – Federal Version (SDWIS/FED) data 2007 4th Quarter Inventory Data.*