U.S. Department of the Interior
U.S. Geological Survey

Estimating the Likelihood of MTBE Occurrence in Drinking Water Supplied by Ground-Water Sources in the Northeast and Mid-Atlantic Regions of the United States

By Paul J. Squillace and Michael J. Moran

Open-File Report 00-343

U.S. Department of the Interior
Bruce Babbitt, Secretary

U.S. Geological Survey
Charles G. Groat, Director

Rapid City, South Dakota: 2000

For additional information write to:

District Chief
U.S. Geological Survey
1608 Mt. View Road
Rapid City, SD 57702

Copies of this report can be purchased from:

U.S. Geological Survey
Branch of Information Services
Box 25286
Denver, CO 80225-0286

FOREWORD

The mission of the U.S. Geological Survey (USGS) is to assess the quantity and quality of the earth resources of the Nation and to provide information that will assist resource managers and policymakers at Federal, State, and local levels in making sound decisions. Assessment of water-quality conditions and trends is an important part of this overall mission.

One of the greatest challenges faced by water-resources scientists is acquiring reliable information that will guide the use and protection of the Nation's water resources. That challenge is being addressed by Federal, State, interstate, and local water-resource agencies and by many academic institutions. These organizations are collecting water-quality data for a host of purposes that include: compliance with permits and water-supply standards; development of remediation plans for a specific contamination problem; operational decisions on industrial, wastewater, or water-supply facilities; and research on factors that affect water quality. An additional need for water-quality information is to provide a basis on which regional- and national-level policy decisions can be based. Wise decisions must be based on sound information. As a society we need to know whether certain types of water-quality problems are isolated or ubiquitous, whether there are significant differences in conditions among regions, whether the conditions are changing over time, and why these conditions change from place to place and over time. The information can be used to help determine the efficacy of existing water-quality policies and to help analysts determine the need for and likely consequences of new policies.

To address these needs, the Congress appropriated funds in 1986 for the USGS to begin a pilot program in seven project areas to develop and refine the National Water-Quality Assessment (NAWQA) Program. In 1991, the USGS began full implementation of the program. The NAWQA Program builds upon an existing base of water-quality studies of the USGS, as well as those of other Federal, State, and local agencies. The objectives of the NAWQA Program are to:

- Describe current water-quality conditions for a large part of the Nation's freshwater streams, rivers, and aquifers.
- Describe how water quality is changing over time.
- Improve understanding of the primary natural and human factors that affect water-quality conditions.

This information will help support the development and evaluation of management, regulatory, and monitoring decisions by other Federal, State, and local agencies to protect, use, and enhance water resources.

The goals of the NAWQA Program are being achieved through ongoing and proposed investigations of 59 of the Nation's most important river basins and aquifer systems, which are referred to as Study Units. These Study Units are distributed throughout the Nation and cover a diversity of hydrogeologic settings. More than two-thirds of the Nation's freshwater use occurs within the 59 Study Units, and more than two-thirds of the people served by public water-supply systems live within their boundaries.

National synthesis of data analysis, made on the basis of an aggregation of comparable information obtained from the Study Units, is a major component of the program. This effort focuses on selected water-quality topics using nationally consistent information. Comparative studies will explain differences and similarities in observed water-quality conditions among Study Units and will identify changes and trends and their causes. The first topics addressed by the national synthesis are pesticides, nutrients, volatile organic compounds, and aquatic biology. Discussions on these and other water-quality topics will be published in periodic summaries of the quality of the Nation's surface and ground water as the information becomes available.

This report is an element of the comprehensive body of information developed as part of the NAWQA Program. The program depends extensively on the advice, cooperation, and information from many Federal, State, interstate, Tribal, and local agencies and the public. The assistance and suggestions of all are greatly appreciated.

Robert M. Hirsch

Robert M. Hirsch
Chief Hydrologist

→ V

CONTENTS

FIGURES

TABLES

Estimating the Likelihood of MTBE Occurrence in Drinking Water Supplied by Ground-Water Sources in the Northeast and Mid-Atlantic Regions of the United States

By Paul J. Squillace *and* Michael J. Moran

ABSTRACT

A multivariate logistic regression model was developed to help explain the presence or absence of methyl *tert*-butyl ether (MTBE) at concentrations equal to or exceeding 0.5 µg/L (micrograms per liter) in ground water from wells in parts of the Northeast and Mid-Atlantic regions of the Nation that are used primarily for drinking water. The model was developed from a calibration data set consisting of information on MTBE in water samples collected from 1,042 wells. MTBE was detected at concentrations equal to or exceeding 0.5 µg/L in 104 samples, but at concentrations equal to or exceeding 5.0 µg/L in only 14 samples. Thus, the model was developed to predict the occurrence of MTBE at greater than or equal to 0.5 µg/L.

A number of factors that describe the conditions in the vicinity of the well were related to the frequency of detection of MTBE but three factors, or variables, most effectively explain MTBE occurrence in a multivariate logistic regression model: MTBE use in gasoline in the study area, the density of above and underground storage tanks, and a soil erodability factor.

A goodness-of-fit test indicated that the logistic regression model fit the calibration data set very well. The model did not, however, provide good estimated probabilities for a validation data set consisting of water samples from 2,787 wells located throughout the entire United States, and thus should not be used to estimate probabilities of detecting MTBE.

INTRODUCTION

Oxygenates, compounds that contain oxygen, are commonly added to gasoline in the United States (U.S.) as an octane enhancer and to promote more complete combustion of gasoline. Octane enhancement began in the late 1970's with the phase-out of tetraethyl lead from gasoline. The use of oxygenates was expanded as a result of the enactment of the Clean Air Act (CAA) Amendments of 1990, which mandate that oxygen be added to gasoline in areas that do not meet certain air quality standards (U.S. Environmental Protection Agency, 1990). Although the CAA Amendments do not specify which oxygenate must be added to gasoline, the oxygenate used most commonly is methyl *tert*-butyl ether (MTBE) because of its low cost, ease of production, and favorable transfer and blending characteristics (Squillace and others, 1995). MTBE is manufactured in large quantities and has been detected frequently in surface and ground water in areas where it is used as a fuel oxygenate (Moran and others, 1999; Moran, Halde, and others, 2000). This has caused concern about water quality in these areas. MTBE in ground water has been associated with taste and odor concerns and potential human-health effects (U.S. Environmental Protection Agency, 1997).

In an attempt to provide information to the U.S. Environmental Protection Agency (USEPA) in support of its efforts to explore and recommend alternatives to the current use of MTBE under section 6 of the Toxic Substances Control Act, the U.S. Geological Survey (USGS) analyzed water-quality data for parts of the Northeast and Mid-Atlantic regions using statistical methods. Specifically, logistic regression analysis was used in an attempt to define relations between the

probability of detecting MTBE in water from wells used primarily for drinking water and selected explanatory variables. The ultimate goals of the work were (1) to estimate the probability that a municipal- or private-supply well will be contaminated by MTBE, and (2) to estimate the total number of these wells in which MTBE will be present at or above concentrations of regulatory concern.

PURPOSE AND SCOPE

The purpose of this report is to provide information to the USEPA in support of the Toxic Substances Control Act ruling that would attempt to limit the use of MTBE. Currently, it is not possible to provide national projections and estimates of the type requested by USEPA using approaches such as deterministic modeling. The desired projections would require development of quantitative relations between the various natural and human-related factors affecting well contamination including hydrogeologic characteristics of the aquifer used as a water-supply source, well characteristics, pumping rate, and MTBE use. Accurate, high-resolution information of this type is not available for many locations on a national scale. Therefore, the USGS assisted by developing a logistic regression model for parts of the Northeast and Mid-Atlantic regions that estimates the probability of detecting MTBE in ground water from wells used primarily for drinking water. The Northeast and Mid-Atlantic regions were selected for study because several large data sets of MTBE analyses from various sources were available for this area. The work for this study entailed the following tasks:

1. Development of a model from a calibration data set that defines the probability of MTBE occurrence at concentrations equal to or exceeding 0.5 µg/L (micrograms per liter), using information on the occurrence of MTBE in ground water from 12 Northeast and Mid-Atlantic States.
2. Attempted validation of the model using an outside data set consisting of data on MTBE in water samples from 2,787 wells located throughout the U.S.

ACKNOWLEDGMENTS

The authors would like to thank Bernard T. Nolan of the USGS National Water-Quality Assessment (NAWQA) Program nutrient synthesis team and Anthony J. Tesoriero of the NAWQA Puget Sound Study Unit for their insight and assistance with the technical aspects of this analysis. In particular, their help was important in developing the final calibration model for estimating the probability of MTBE occurrence in ground water from wells used primarily for drinking water.

STUDY METHODS

The available data set is from 12 Northeast and Mid-Atlantic States and consists of information on MTBE concentrations in ground-water samples from 1,042 wells. One part of the data set consists of compliance monitoring data from 564 municipal water systems where the source could be linked to a specific well location (Grady and Casey, 1999). Data from 47 wells, sampled as part of a joint effort between the USGS, the Metropolitan Water District of Southern California, the Oregon Graduate Institute, and in conjunction with the American Water Works Association Research Foundation, also were included. These wells are all municipal water-supply wells and were sampled prior to treatment. Information from 431 wells, sampled as part of the NAWQA Program, also was included and represented a mix of water uses. Fifty-one of the NAWQA wells were completed as shallow monitoring wells in urban areas and were sampled to determine the effect of land use on ground-water quality (Squillace and Price, 1996). The 51 monitoring wells were included to increase the size of the data set and possibly the number of MTBE detections. The remaining NAWQA wells are used for municipal (28 wells) and private supplies (352 wells).

Logistic regression models were tested to find which independent variables provided the best estimates of the probability of detecting MTBE. The modeling process followed methods outlined by Helsel and Hirsch (1995), Hosmer and Lemeshow (1989) and Kleinbaum (1994). An initial review of the data set revealed only 14 detections of MTBE at or exceeding a concentration of 5.0 µg/L. Figure 1 is a histogram of the number of MTBE detections in the data set at four different censoring concentrations.

As seen in figure 1, as the MTBE concentration used for establishing a binary variable increased, the number of detections decreased substantially. Because of the small number of detections at censoring concentrations of 5.0 and 1.0 µg/L, logistic modeling at these higher concentrations was not possible. According to

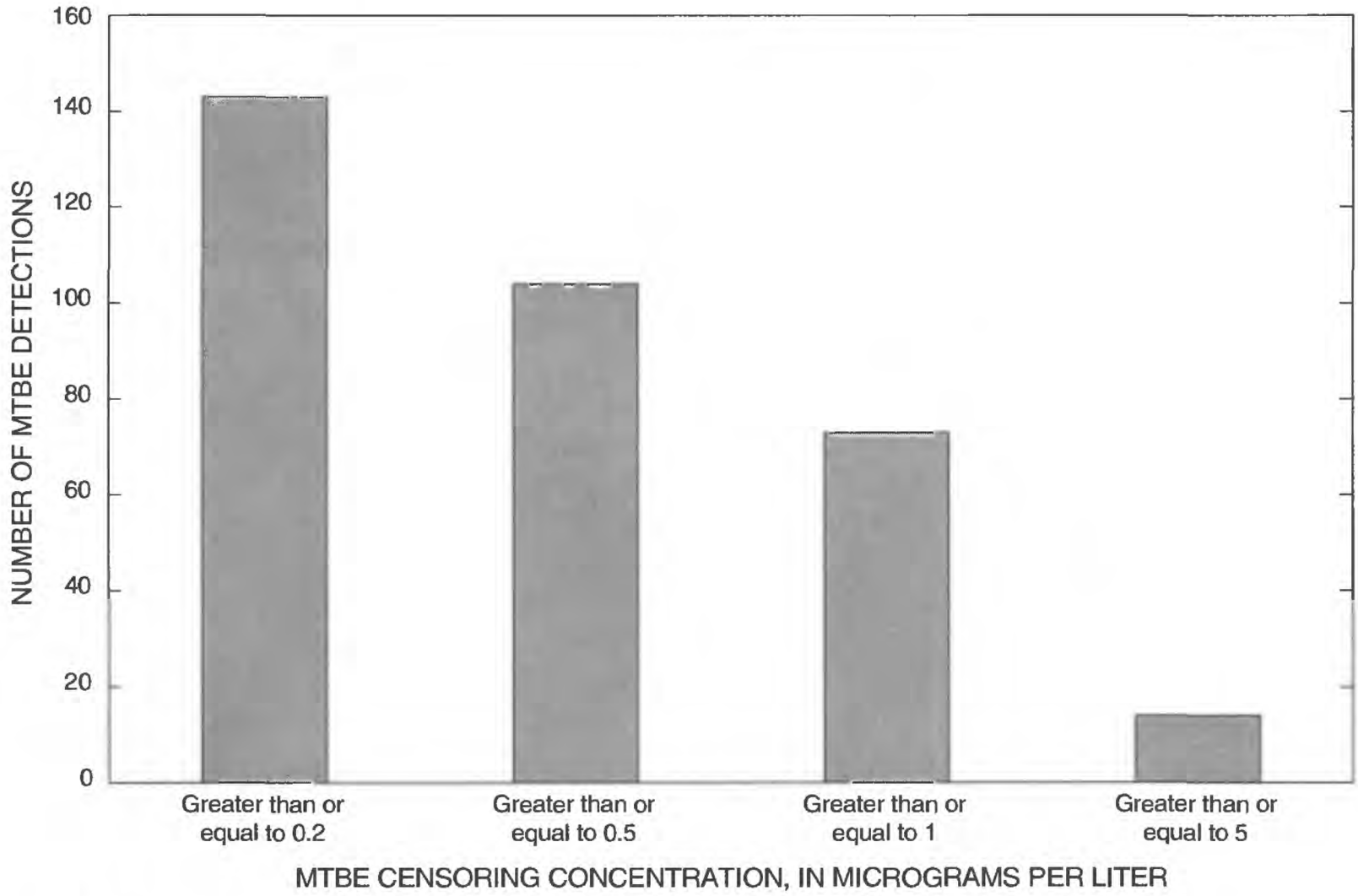

Figure 1. Number of MTBE (methyl *tert*-butyl ether) detections at various censoring concentrations.

Harrell and others (1996), in order to have a model with predictive discrimination to validate, the number of independent variables in the model should not exceed *m*/10, where *m* is the number of detections. This would mean that if the model was developed on the basis of a minimum concentration of 5.0 µg/L, only one independent variable should remain in the model. Having only one independent variable in the model was considered too restrictive; therefore, a lower censoring concentration of 0.5 µg/L was chosen. As shown in figure 1, there were 104 detections of MTBE equal to or exceeding a concentration of 0.5 µg/L.

One of the most important independent variables used in the model is MTBE use in gasoline. Two programs for oxygen use were specified by the CAA Amendments: (1) the Oxygenated Fuels (OXY) Program, in which gasoline must contain 2.7 percent oxygen by weight during the cold season in areas that fail to meet air-quality standards for carbon monoxide, and (2) the Reformulated Gasoline (RFG) Program, in which gasoline must contain 2 percent oxygen by weight year-round in areas having the highest levels of tropospheric ozone (U.S. Environmental Protection Agency, 1990). Using MTBE to meet the oxygen requirements of the CAA Amendments means that gasoline in designated OXY program areas must contain 15 percent MTBE by volume, and gasoline in designated RFG areas must contain 11 percent MTBE by volume. Some areas of the U.S. have chosen to voluntarily meet program requirements (Moran, Clawges, and Zogorski, 2000).

Although knowing that an area is identified as one required to meet RFG or OXY program requirements can give some insight into the type and amount of oxygenate used, this information alone cannot specifically determine which oxygenate is used in specific metropolitan areas and in what amounts. Data on the volumes of oxygenates and other compounds in gasoline were available from several gasoline surveys. The surveys gathered information on the constituents of gasoline, including the percentage volume of ether oxygenates, used in various metropolitan areas of the country. The National Institute for Petroleum and Energy Research (NIPER) survey contained data for the greatest number of cities and the largest number of samples analyzed (Moran, Clawges, and Zogorski, 2000). Therefore, NIPER data were used to generate information on MTBE use.

Using information on RFG program areas and information from the NIPER survey, water-quality data collected from each well was placed into a high or low MTBE-use category and binary coded for the logistic regression model. Water-quality data from wells in areas that used MTBE in the RFG or OXY programs were classified as high MTBE-use (binary coded as 1) unless data from the survey indicated a long-term average MTBE use of less than 3 percent by volume. Water-quality data from wells in areas that did not use MTBE in the RFG or OXY programs were classified as low MTBE-use (binary coded as 0) unless NIPER data indicated a long-term average MTBE use of greater than or equal to 3 percent by volume. NIPER data were available for only about 70 cities in the U.S.; however, many of the cities using RFG and OXY gasoline are included in the NIPER survey and comprise most of the high-MTBE-use areas. An analysis of MTBE use in larger cities that have gasoline survey information indicates that most cities outside RFG and OXY programs have low average MTBE use (less than 3 percent by volume). Therefore, it was considered safe to assume that most areas outside of RFG and OXY areas use low amounts of MTBE in gasoline and were classified as low MTBE use (binary coded as 0).

Another important variable in the model is the density of storage tanks. This variable represents the number of aboveground and underground storage tanks within each square kilometer in the vicinity of the well. The storage tanks include those regulated for storing any type of material, but most of the tanks store petroleum products. The data for this variable were obtained from StarView Real Estate Version 2.6.1, Vista Information Solutions, Inc., San Diego, California. The density number was obtained by making a geographic grid (in square kilometers) of a point coverage of storage-tank locations.

The third principal variable that is part of the final logistic regression equation is soil erodability. Soil erodability describes the soil's inherent susceptibility to erosion and is closely related to infiltration capacity and structural stability. Soil texture, organic matter content, and permeability are some of the factors inherent to soil type that affect the values of this variable. Data on soil erodability for the country are gathered and compiled by the Soil Conservation Service and are available in digital form as part of a national soil-survey data base (Natural Resources Conservation Service, 2000). Soil erodability values generally range from near zero to almost 0.6 (dimensionless), where low values indicate low erodability and high values indicate high erodability. Soil erodability is negatively correlated with percentage of sand and soil permeability (Spearman's correlation coefficients are both about -0.7). Soil erodability is also negatively correlated with rates of recharge (-0.38); therefore, low values of soil erodability correlate with high ground-water recharge, high percentage of sand, and high soil permeability. To improve statistical performance of the model, soil erodability was converted to a binary variable (high or low) by dividing the data in about equal parts; if soil erodability was equal to or exceeded 0.22, the binary value was set at 0, and if soil erodability was less than 0.22, the value was set at 1. The "1" condition indicates greater aquifer susceptibility to contamination because low values of the soil erodability (less than 0.22) correspond to more permeable soils.

The remainder of the variables considered in the model included a wide variety of other hydrogeologic, soil property, or human-related factors that were believed to be potentially significant in determining MTBE occurrence. Both univariate and multivariate logistic regression models were evaluated for testing the variables. Univariate models were evaluated by testing the variables (listed in table 1) for their ability to predict the detection of MTBE at concentrations equal to or exceeding 0.5 µg/L.

Multivariate models were evaluated by a backward elimination method in which all potential variables were included in the model and then the variables were eliminated, one at a time, until all remaining variables were statistically significant. This procedure allowed variables that were previously eliminated to enter the model if they became significant with a different subset of variables. By using this approach, the large initial model—containing many potential explanatory variables—was reduced to a smaller model. Variables that lack statistical significance in univariate models sometimes attain significance when combined with other variables in a multivariate model. As a final check, each of the eliminated variables was screened one at a time with the smaller multivariate model to ensure that the most significant variables were retained.

Not all variables can be considered together because some are collinear. Collinear variables were considered separately and were tested using backward elimination. Variables were considered significant if the Wald's t-statistic (Helsel and Hirsch, 1995) had a

probability (p) less than or equal to 0.05. The Akaike Information Criterion was used to evaluate the best non-nested models, and the partial likelihood ratio statistic was used to evaluate the best nested models along with the Wald chi-square value (corresponding to the square of the Wald statistic) for the parameter estimate (Helsel and Hirsch, 1995). Many possible interaction terms also were tested for inclusion in the model; however, almost all were not significant.

The usefulness of the model in predicting the occurrence of MTBE in ground water used primarily for drinking water was evaluated by attempting to validate the predictive capability of the regression equation. One way of doing this is to split the calibration data set randomly into two roughly equal parts and test the predictive ability on one part of the model and test validation on the other part (Hosmer and Lemeshow, 1989). The calibration data set, sometimes called the developmental data set (Hosmer and Lemeshow, 1989), is used to establish the parameters and estimated coefficients of the parameters for the logistic regression equation. In this case, the calibration data set was the data set for the Northeast and Mid-Atlantic States. For validation testing, the estimated values of the parameter coefficients are assumed to be fixed constants rather than estimated values. Validating the model is an attempt to test the predictive capability of the regression equation on an outside data set. The outside data set used for validation testing is sometimes referred to as validation data (Hosmer and Lemeshow, 1989).

Table 1. Variables tested in logistic regression models

[MTBE, methyl *tert*-butyl ether; USEPA, U.S. Environmental Protection Agency; USGS, U.S. Geological Survey; STATSGO, State Soil Geographic Data Base; km^2, square kilometers; --, not applicable; <, less than, ≥, greater than or equal to]

Explanatory variable	Type of variable	Source of variable[1]
More Useful Explanatory Variables (Wald's p-value generally <0.05)		
MTBE use (high =1, low = 0)	Binary	USEPA data
Soil erodability (high = 0, low = 1)	Binary	STATSGO
Total storage-tank density (natural logarithm of aboveground storage tanks + underground storage tanks density at 1-km^2 grid plus 0.00001)	Continuous	StarView
Leaking underground storage tanks (density at 1-km^2 grid)	Continuous	StarView
Natural logarithm of population density	Continuous	1990 Census
Aquifer type (areas were classified by presence or absence of a principal aquifer)	Binary	USGS National Atlas
MTBE use × soil erodability (interaction term)	Binary	--
Poor Explanatory Variables (Wald's p-value generally ≥0.05)		
Aquifer permeability	Categorical	USGS data
Percentage of sand	Continuous	STATSGO
Depth to rock	Continuous	STATSGO
Ground-water use	Continuous	USGS data
Water-table depth	Continuous	STATSGO
Soil permeability	Continuous	STATSGO
Land-surface slope	Continuous	STATSGO
Ground-water recharge	Continuous	USGS data
Well type (drinking/monitoring)	Binary	USGS data

[1]Source of variables: USEPA data (U.S. Environmental Protection Agency, 1999); STATSGO (Natural Resources Conservation Service, 2000); StarView Real Estate Version 2.6.1, Vista Information Solutions, Inc., San Diego, California; 1990 Census (Consortium for International Earth Science Information Network, 1996); USGS National Atlas (U.S. Geological Survey, 1997).

APPLICATION OF LOGISTIC REGRESSION MODEL

Table 1 shows variables that were considered more useful explanatory variables and those that were considered poor explanatory variables. More useful explanatory variables had Wald statistic *p*-values generally less than 0.05 in the multivariate models.

The results obtained here may be unique to the calibration data set used. For example, the density of leaking underground storage tanks (LUST) may become more significant when modeling the probability of occurrence of higher concentrations of MTBE than those used here. A number of variables that were collinear could not be used together in the model. Densities of LUST's, aboveground storage tanks, and underground storage tanks were strongly correlated. Explanatory variables listed as poor had little benefit because the quality of the estimated value of the variable was poor or it simply did not explain the detections of MTBE. Modeling results also showed no difference between water-quality data from monitoring wells and those from the drinking-water wells when a variable representing well type did not yield significant Wald *p*-values during model runs.

The best logistic regression model contained three variables: (1) MTBE use, (2) the natural logarithm of total storage-tank density in the vicinity of the well (here defined as the sum of aboveground tanks and underground tanks plus a constant of 0.00001 to accommodate zero values), and (3) soil erodability. The probability of detecting MTBE at or above 0.5 µg/L in this data set is described by the following equation:

$$P = \frac{e^{-3.37 + 1.47(x) + 0.047\ln(y) + 0.81(z)}}{1 + e^{-3.37 + 1.47(x) + 0.047\ln(y) + 0.81(z)}} \qquad (1)$$

where

x = MTBE use (binary code);
y = natural logarithm of total storage-tank density (continuous); and
z = soil erodability (binary code).

Wald chi-square *p*-values for all variables were about 0.02 or less, which is considered highly significant. Additionally, the model yielded a Hosmer-Lemeshow goodness-of-fit test *p*-value of 0.64, which is also considered very significant (Hosmer and Lemeshow, 1989). High *p*-values for the Hosmer-Lemeshow test are desirable because the null hypothesis is that the model fits the data.

Figures 2 and 3 show how the estimated probabilities of MTBE contamination varied with different values of MTBE use, total storage-tank density, and soil erodability. The greatest probabilities of detecting MTBE, based on equation 1, were about 30 percent. The probability of detecting MTBE is greater for (1) high MTBE use, (2) higher tank density, and (3) low soil erodability. Figures 2 and 3 also show that the 95-percent confidence intervals around the estimated probabilities were large (dashed lines). The overall low estimated probabilities and wide confidence intervals indicate that the model was missing some important variable or that the estimates for the three variables were of poor quality.

Uncertainty analysis was performed on the calibrated model by computing confidence intervals for probabilities estimated with equation 1 and for model parameters (intercept and slope coefficients). Results of this analysis are shown in table 2. The wide confidence intervals, especially in areas of high MTBE use, reflect great uncertainties in the predictive capabilities of the model. These large uncertainties probably arise from a combination of the small data set and model-input and parameter errors.

Validation of the model was attempted by randomly splitting the data from the 1,042 wells into two subsets containing about 520 wells each. This process was repeated a second time to evaluate the effect of the data split on the validation procedure. The variables that were significant for the 1,042 wells (MTBE use, total storage-tank density, and soil erodability) were not consistently significant in the random subsets.

Validation also was attempted by using data from 2,787 NAWQA wells located throughout the U.S. Using the logistic regression equation, probabilities of detecting MTBE were calculated for all the wells. Probabilities were put into deciles and the Hosmer-Lemeshow test statistic was computed. The *p*-value for the Hosmer-Lemeshow test statistic was less than 0.05, indicating that the model did not fit the validation data set (Hosmer and Lemeshow, 1989). Because the model could not be validated with the available data it should not be used for estimation of the probability of MTBE detections either for the Northeast and Mid-Atlantic States or for the rest of the U.S.

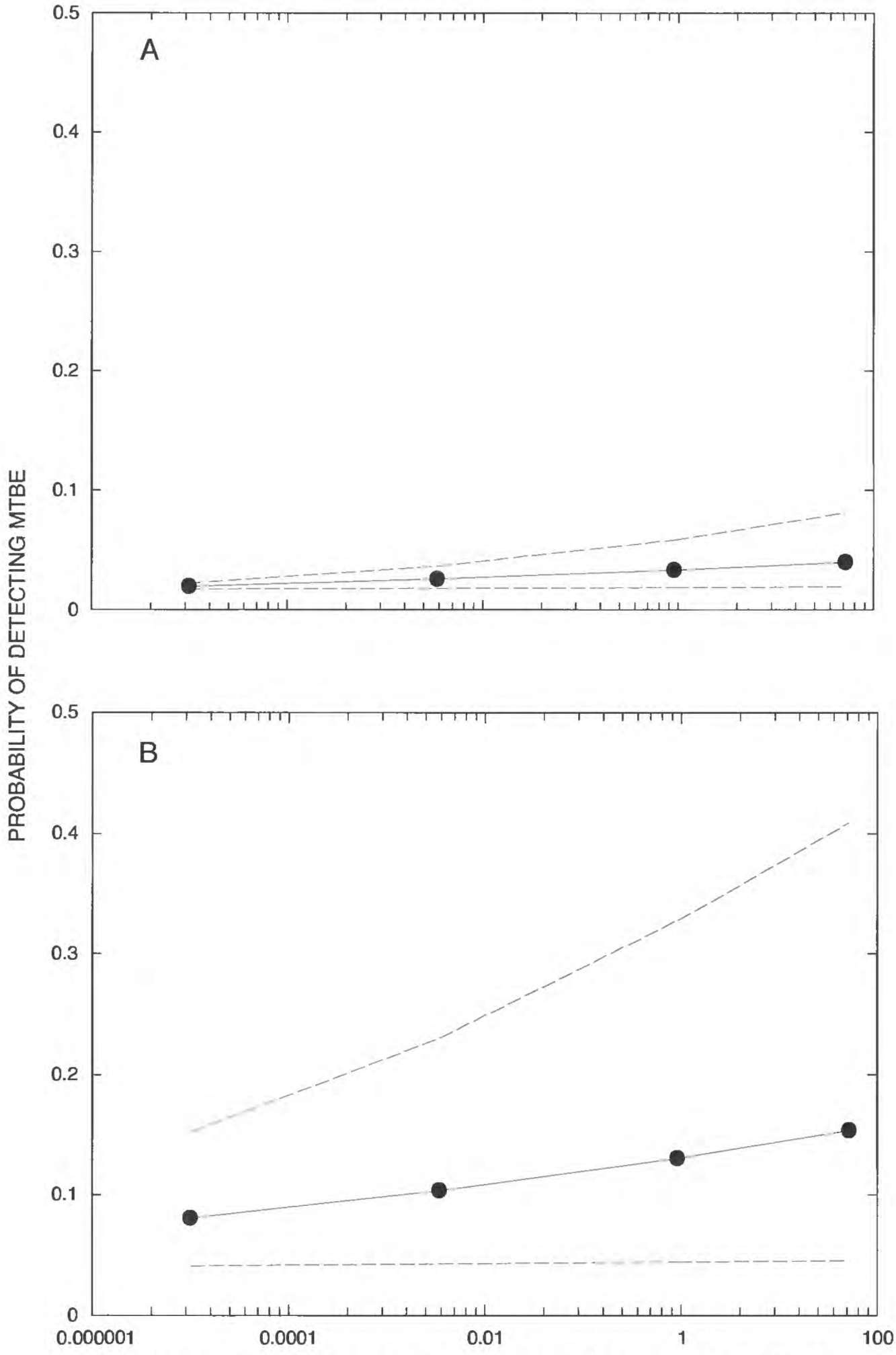

Figure 2. Estimated probabilities of detecting MTBE (methyl *tert*-butyl ether) in (A) areas of little MTBE use and (B) areas where MTBE is used in the Reformulated (RFG) or Oxygenated (OXY) gasoline programs with high soil erodability according to the logistic model described by equation 1.

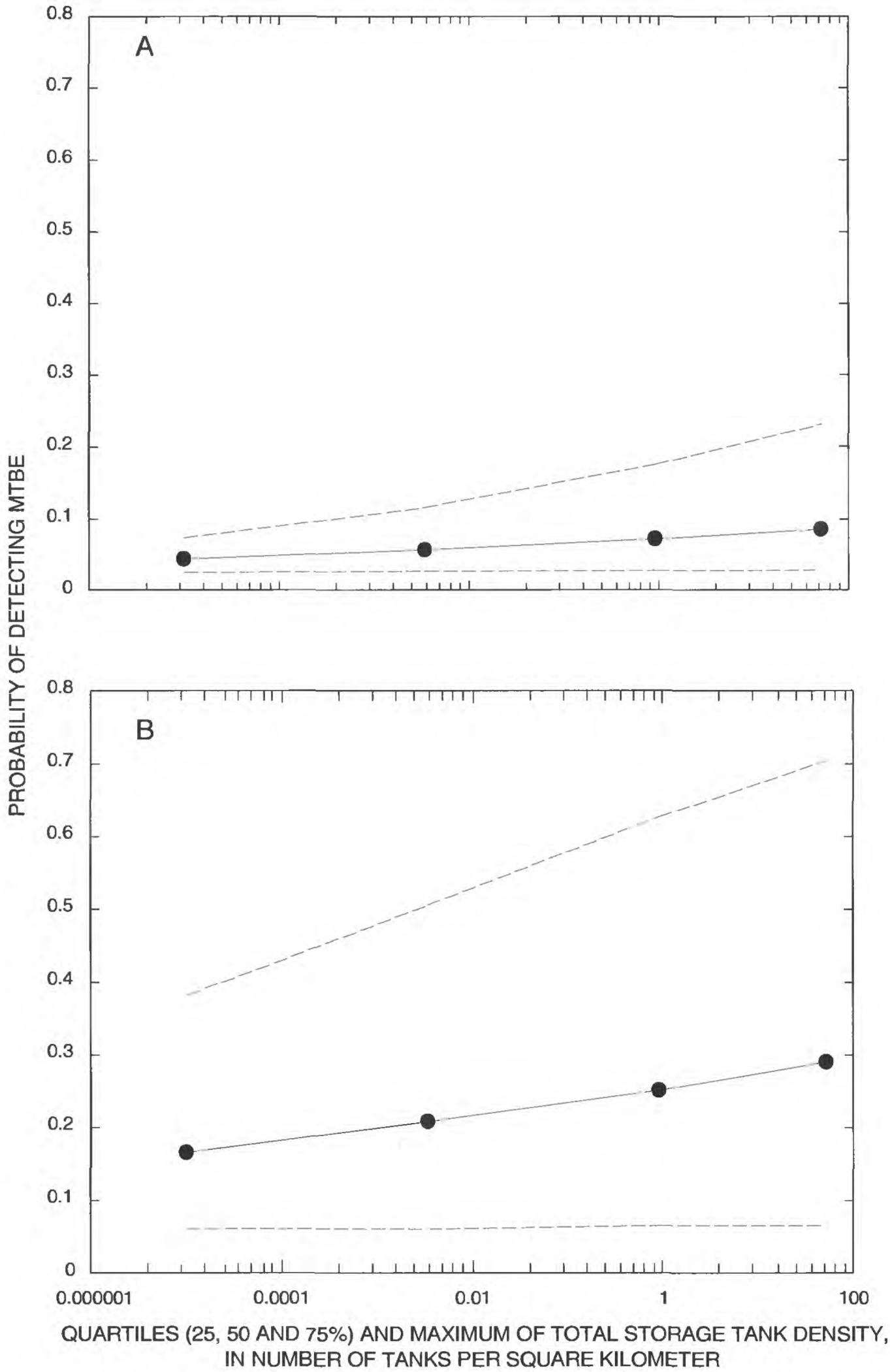

Figure 3. Estimated probabilities of detecting MTBE (methyl *tert*-butyl ether) in (A) areas of little MTBE use and (B) areas where MTBE is used in the Reformulated (RFG) or Oxygenated (OXY) gasoline programs with low soil erodability according to the logistic model described by equation 1.

Table 2. Uncertainty analysis for parameter estimates in final logistic regression model

[MTBE, methyl *tert*-butyl ether]

Explanatory variable	Parameter estimate	Wald 95-percent confidence limits	
		Lower	Upper
Intercept	-3.37	-3.96	-2.77
MTBE-use (high or low)	1.47	.88	2.06
Total storage-tank density in the vicinity of the well	.047	.007	.087
Soil erodability	.81	.38	1.23

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

A number of variables were used in a multivariate logistic regression model to predict the occurrence of MTBE in ground water of the Northeast and Mid-Atlantic States at a concentration equal to or exceeding 0.5 µg/L. The variables most effective in explaining MTBE occurrence were MTBE use in gasoline, density of aboveground and underground storage tanks, and soil erodability. Although these variables help to describe the occurrence of MTBE in ground water in 12 of the Northeast and Mid-Atlantic States, the model could not be validated with data from the entire country and should not be used for estimating the probability of MTBE detections. The model might have been more successful if more ancillary data were available for each well, including factors such as hydrogeologic characteristics of the aquifer used for water supply, well characteristics, pumping rate, MTBE use, and others. Also, if existing model input parameters (for example, total storage tank density, soil erodability, etc.) were more accurately defined, the model would be more accurate and precise in estimating the probability of MTBE occurrence. For example, knowing more accurately the distance between a well and nearby storage tanks and also the direction of ground-water flow in the area would substantially improve the ability of the storage tank variable to predict MTBE occurrence.

Despite the limitations of this initial effort, the approach used in this preliminary modeling holds promise in estimating the number of drinking-water supplies that may be adversely affected by MTBE contamination. The major implication of this initial effort is that a larger calibration data set on MTBE concentrations and ancillary information are needed, and improved ancillary information is needed to reduce the model input and parameter errors. A larger data set could probably be developed because only a 20-percent random sample of drinking-water compliance data was used for this preliminary analysis. However, considerable USGS and USEPA efforts would be required to compile additional data on MTBE usage and gasoline storage tanks as well as other ancillary information. A larger, improved data set for this region would likely provide sufficient MTBE data to validate the 0.5-µg/L MTBE model and potentially to develop a calibrated model at the 5.0-µg/L level.

A second implication of this initial effort is that statistical modeling of the occurrence of MTBE and other volatile organic compounds may be feasible using data from wells sampled by the NAWQA Program throughout the U.S. between 1993 and 1998. There may be sufficient data to develop a model at the 0.2-µg/L level but not at the 5- or 20-µg/L levels because the NAWQA wells primarily sample ambient ground water and generally only low concentrations of MTBE are found. Development of a model from a nationwide data set also may provide better information on explanatory variables than this initial modeling for this region because of the detailed ancillary information collected by NAWQA staff at each sampled well. Additional effort would be needed to improve storage-tank data as part of the initial nationwide modeling effort. A national modeling effort would also be an important prerequisite to developing a national design for assessment of MTBE concentrations in domestic wells and possibly small municipal water-supply systems.

REFERENCES

Consortium for International Earth Science Information Network (CIESIN), 1996, Socioeconomic Data and Applications Center (SEDAC): Archive of Census Related Products, University Center, Michigan.

Grady, S.J., and Casey, G.D., 1999, A plan for assessing the occurrence and distribution of methyl *tert*-butyl ether (MTBE) and other volatile organic compounds in drinking water and ambient ground water of the northeast and mid-Atlantic regions of the United States: U.S. Geological Survey Open-File Report 99-207, 36 p.

Harrell, F.E., Lee, K.L., and Mark, D.B., 1996, Multivariate prognostic models—issues in developing models, evaluating assumptions and adequacy, and measuring and reducing error: Statistics in Medicine, v. 15, p. 361-387.

Helsel, D.R., and Hirsch, R.M., 1995, Statistical methods in water resources: The Netherlands, Elsevier Science, 529 p.

Hosmer, D.W., and Lemeshow, S., 1989, Applied logistic regression: New York, Wiley, 307 p.

Kleinbaum, D.G., 1994, Logistic regression: New York, Springer, 282 p.

Moran, M.J., Clawges, R.M., and Zogorski, J.S., 2000, Identifying the usage patterns of methyl *tert*-butyl ether (MTBE) and other oxygenates in gasoline using gasoline surveys [abs.], *in* American Chemical Society Division of Environmental Chemistry, San Francisco, Calif., March 26-30, 2000: American Chemical Society, v. 40, no. 1, p. 209-212.

Moran, M.J., Halde, M.J., Clawges. R.M., and Zogorski, J.S., 2000, Relations between the detection of methyl *tert*-butyl ether (MTBE) in surface and ground water and its content in gasoline [abs.], *in* American Chemical Society Division of Environmental Chemistry, San Francisco, Calif., March 26-30, 2000: American Chemical Society, v. 40, no. 1, p. 195.

Moran, M.J., Zogorski, J.S., and Squillace. P.J., 1999, MTBE in ground water of the United States—occurrence, potential sources, and long-range transport, *in* Proceedings of the 1999 Water Resources Conference: American Water Works Association, Norfolk, Va., September 26-29, 1999.

Natural Resources Conservation Service, 2000, National STATSGO database, accessed at URL http://www.ftw.nrcs.usda.gov/stat_data.html.

Squillace, P.J., and Price, C.V., 1996, Urban land-use study plans for the National Water-Quality Assessment Program: U.S. Geological Survey Open-File Report 96-217, 19 p.

Squillace, P.J., Pope, D.A., and Price, C.V., 1995, Occurrence of the gasoline additive MTBE in shallow ground water in urban and agricultural areas: U.S. Geological Survey Fact Sheet 114-95, 4 p.

U.S. Environmental Protection Agency, 1999, State winter oxygenated fuels program and list of reformulated gasoline program areas, accessed at URL http://www.epa.gov/oms/regs/fuels.

U.S. Environmental Protection Agency, 1997, Drinking water advisory—consumer acceptability advice and health effects analysis on methyl tertiary-butyl ether: Washington, D.C., Office of Water, EPA-822-F-97-009, 41 p.

U.S. Environmental Protection Agency, 1990, The Clean Air Act Amendments: Washington, D.C., 101st Congress of the United States, 1990, sec. 219, p. S.1630-1938, accessed at URL http://www.epa.gov/oar/caa/caaa.txt.

U.S. Geological Survey, 1997, National Atlas of the United States, accessed at URL http://www.nationalatlas.gov/.

☆ U.S. GOVERNMENT PRINTING OFFICE: 2000 — 673-085 / 50008 Region No. 8