Relating Nitrogen Sources and Aquifer Susceptibility to Nitrate in Shallow Ground Waters of the United States

Bernard T. Nolan

Abstract

Characteristics of nitrogen loading and aquifer susceptibility to contamination were evaluated to determine their influence on contamination of shallow ground water by nitrate. A set of 13 explanatory variables was derived from these characteristics, and variables that have a significant influence were identified using logistic regression (LR). Multivariate LR models based on more than 900 sampled wells predicted the probability of exceeding 4 mg/L of nitrate in ground water. The final LR model consists of the following variables: (1) nitrogen fertilizer loading (p-value = 0.012), (2) percent cropland-pasture ($p < 0.001$), (3) natural log of population density ($p < 0.001$), (4) percent well-drained soils ($p = 0.002$), (5) depth to the seasonally high water table ($p = 0.001$), and (6) presence or absence of a fracture zone within an aquifer ($p = 0.002$). Variables 1-3 were compiled within circular, 500-m radius areas surrounding sampled wells, and variables 4-6 were compiled within larger areas representing targeted land use and aquifers of interest. Fitting criteria indicate that the full logistic-regression model is highly significant ($p < 0.001$), compared with an intercept-only model that contains none of the explanatory variables. A goodness-of-fit test indicates that the model fits the data very well, and observed and predicted probabilities of exceeding 4 mg/L nitrate in ground water are strongly correlated ($r^2 = 0.971$). Based on the multivariate LR model, vulnerability of ground water to contamination by nitrate depends not on any single factor but on the combined, simultaneous influence of factors representing nitrogen loading sources and aquifer susceptibility characteristics.

Bernard T. Nolan, U.S. Geological Survey, 413 National Center, Reston, VA 20192
Email: btnolan@usgs.gov; telephone: (703) 648-4000; fax: (703) 648-6693

Introduction

Ground water provides drinking water for more than one-half of the people in the United States, and accounted for 39 percent of water withdrawn to supply cities and towns and 96 percent of water withdrawn by private users in 1990 (Solley and others, 1993). This important national resource is vulnerable to contamination by chemicals, including nitrate, that can pass through soil to the water table. Major sources of nitrate in watersheds of the U.S. include inorganic fertilizer, animal manure, and atmospheric deposition (Puckett, 1994). Nitrate is soluble in water, can easily leach through soil, and can persist in shallow ground water for decades.

Elevated concentrations of nitrate in drinking water are a cause for concern. Ingestion of nitrate by infants can cause low oxygen levels in the blood, a potentially fatal condition (Spalding and Exner, 1993). Other adverse health effects potentially related to ingestion of nitrate in drinking water include spontaneous abortions and non-Hodgkin's lymphoma. Nitrate concentrations of 19-29 milligrams per liter (mg/L) in rural, domestic

wells in Indiana might have caused eight spontaneous abortions among four women during 1991-1994 (Centers for Disease Control and Prevention, 1996). Nitrate concentrations of 4 mg/L or more in water from community wells in Nebraska have been associated with increased risk of non-Hodgkin's lymphoma (Ward and others, 1996). The U.S. Environmental Protection Agency (USEPA) has established a maximum contaminant level (MCL) of 10 mg/L nitrate as nitrogen (N) (U.S. Environmental Protection Agency, 1995). Shallow ground water unaffected by human activities commonly contains less than 2 mg/L of nitrate (Mueller and Helsel, 1996).

Determining where ground water is at risk of nitrate contamination can alert managers of the need to protect water supplies. Approaches to assessing aquifer vulnerability to contamination include overlay and index methods, process-based simulation models, and statistical methods. Advantages of statistical methods are that they are inherently flexible, can readily accommodate differences in spatial scale, and can effectively describe uncertainty (National Research Council, 1993).

Factors that influence contaminant behavior vary widely at large spatial scales, which complicates efforts to model aquifer vulnerability. Additionally, "water quality is a multivariate concept" (Riley and others, 1990) that cannot be summarized by any single constituent or factor. Multiple factors influence nitrate behavior in ground water at international, regional, and local scales. Nitrogen dose, crop type, soil texture, and ground-water level were empirically related to nitrate leaching from topsoil and ground-water nitrate concentration in Europe for the purpose of constructing vulnerability maps (Meinardi and others, 1995). D'Agostino and others (1998) observed that nitrate concentration in ground water beneath the Lucca Plain, Italy, varied substantially depending on local agricultural practices, proximity to the Serchio River, and the presence of a confining layer. At the local scale, overwinter cover cropping and delayed plowing generally decreased nitrate leaching from a calcareous loam soil in Eastern England (Davies and others, 1996).

In previous research in the United States (Nolan and Stoner, 2000), relations between ground-water nitrate concentration and potential explanatory variables were analyzed in simple univariate fashion, resulting in considerable unexplained variation in nitrate concentration. For example, a plot of median nitrate concentration in shallow ground water versus median N load from fertilizer, manure, and atmospheric sources showed generally increasing nitrate response to N loading but exhibited considerable scatter. The variation in nitrate response was attributed in part to areas with well-drained soils or fractured bedrock, which can readily convey even small amounts of applied N to the water table, and to areas with poorly drained soils that impede nitrate movement to ground water even when N load is high. Thus, including a variable such as soil-drainage characteristic in the analysis would explain additional variation.

Multivariate statistical methods can improve on univariate analysis by simultaneously considering the influence of N loading, soil-drainage characteristic, presence/absence of fractured rocks, and other variables on nitrate concentration in ground water. In this paper, logistic regression (LR), a multivariate statistical method, was used to interpret ground-water nitrate data collected as part of the U.S. Geological Survey's National Water-Quality Assessment (NAWQA) Program during 1992-1995. Logistic-regression models containing one (univariate) or more (multivariate) explanatory variables were used to predict the likelihood of exceeding 4 mg/L nitrate in

ground waters of the United States. The objective of this research was to apply multivariate LR to the NAWQA data set to identify variables that in combination significantly influence nitrate concentration in shallow, recently recharged ground water.

Background

Logistic regression has been used extensively in epidemiological studies and is becoming more commonplace in the environmental sciences. Logistic regression differs from classical, linear regression in that the modeled response is the probability of being in a category, rather than the observed quantity of a response variable (Helsel and Hirsch, 1992). The main assumption of LR is that the natural logarithm of the odds ratio (probability of being in a response category) is linearly related to the explanatory variables (Afifi and Clark, 1984). Regression coefficients are estimated using the method of maximum likelihood.

Because a threshold value is specified to define the response categories, LR is well-suited to analysis of non-detects. Nearly 20 percent of shallow ground-water samples collected by NAWQA have nitrate concentration below the detection limit of 0.05 mg/L. Additionally, Tesoriero and others (1998) found that LR yielded significant relations ($p \leq 0.05$) between explanatory variables and nitrate concentration in ground waters of the Puget Sound Basin, northwestern Washington, whereas linear regression did not. Variation in nitrate concentration was such that predicting the probability of elevated nitrate concentration was more feasible than predicting a specific concentration.

The odds ratio is based on the probability of exceeding a given threshold value:

$$odds\ ratio = \frac{p}{1-p} \tag{1}$$

where

p = probability of exceeding the threshold value

The log of the odds ratio, or logit, transforms a variable constrained between 0 and 1 into a continuous, unbounded variable that is a linear function of the explanatory variables. The resulting LR equation is (Helsel and Hirsch, 1992)

$$\log\left[\frac{p}{1-p}\right] = b_o + \mathbf{bx} \tag{2}$$

where

b_o = constant

bx = vector of slope coefficients and explanatory variables

The logistic transformation converts the predicted values of the response variable back into probability units (Helsel and Hirsch, 1992):

$$p = \frac{e^{(b_o + \boldsymbol{b}\boldsymbol{x})}}{1 + e^{(b_o + \boldsymbol{b}\boldsymbol{x})}} \quad (3)$$

The likelihood ratio test statistic (G) tests the statistical significance of coefficients in the LR model (Hosmer and Lemeshow, 1989; Tesoriero and Voss, 1997):

$$G = -2(L_{\text{int}} - L_{\text{model}}) \quad (4)$$

where

L_{int} = log-likelihood of intercept-only model

L_{model} = log-likelihood of model with one or more explanatory variables

The G statistic is chi-square distributed under the null hypothesis that slope coefficients for the explanatory variables in the model equal zero. For nested models, G is used to determine the significance of adding one or more new explanatory variables to a model. A nested model contains all of the explanatory variables in the simpler model, plus one or more additional variables. The degrees of freedom equals the number of additional variables in the more complex model (Helsel and Hirsch, 1992).

The Wald statistic is the ratio of the maximum likelihood estimate of the slope coefficient to its standard error (Hosmer and Lemeshow, 1989). It is normally distributed and its p-value indicates whether the slope coefficient is significantly different from zero. When one additional variable is introduced into an LR model, the square of the Wald statistic is approximately chi-square distributed with one degree of freedom (Kleinbaum, 1994). The G statistic and the corresponding squared Wald statistic yield approximately the same value in very large samples.

Wald and G statistics were used in this study to build an optimal model, and the Hosmer-Lemeshow (HL) goodness-of-fit statistic was then used to see how well the model fit the data. The G statistic is emphasized in model building because it compares "the observed values of the response variable to predicted values obtained from models with and without the variable in question" (Hosmer and Lemeshow, 1989). In contrast, the HL statistic compares observed values to those fitted under one model.

The HL statistic compares observed and predicted probabilities for data grouped from low to high by values of the predicted probabilities. Ten groups (deciles) commonly are used, with the first group comprising the n/10 observations with the smallest predicted probabilities, and the last group containing the n/10 observations with the largest predicted probabilities (Hosmer and Lemeshow, 1989). Each group yields an average predicted probability and an observed probability based on the number of observed values in the group that are greater than the threshold value. The HL statistic is chi-squared distributed and, because the null hypothesis is that the model fits the data, higher p-values indicate better fit.

Prior studies show that LR can successfully identify variables that significantly affect ground-water quality. Eckhardt and Stackelberg (1995) used logistic regression to predict the probability of exceeding 3 mg/L of nitrate in ground water beneath agricultural, suburban, and undeveloped areas of Long Island, New York. Explanatory variables consisted of population density, percent medium-density residential land use, percent agricultural land use, and depth to the water table. Population and land-use data

were compiled within 0.80-kilometer (km) radius circular areas surrounding each well to reflect the influence of recent (within six years) land use on ground-water quality. Rank correlations between predicted probabilities and observed responses were used to evaluate model performance. Rank correlations range from zero (no fit) to one (perfect fit). The LR models developed for nitrate had rank correlation coefficients of 0.87–0.88 and indicated that nitrate concentration generally increased as population density and percent residential and agricultural land use increased, and as the depth to the water table decreased.

Tesoriero and Voss (1997) used LR to predict the likelihood that nitrate is present in concentrations of 3 mg/L or more in ground waters of the Puget Sound Basin in northwestern Washington. Variables evaluated included well depth, ground-water recharge, soil hydrologic group, surficial geology type, land-use type, and population density. Well depth, surficial geology, and percentages of agricultural and urban lands within 3.2 km of sampled wells best explained elevated nitrate concentration in ground water. Nitrate data from more than 3,000 wells were used to develop and validate the LR model.

Rupert (1998) used separate LR models to predict the probability of detecting atrazine/desethyl-atrazine and the likelihood of nitrate concentration exceeding 2 mg/L in ground waters of the Upper Snake River Basin in southeastern Idaho. The first model indicated that land use, precipitation, soil hydrologic group, and well depth were significantly related to atrazine/desethyl-atrazine detections. The second model indicated that depth to water, land use, and soil drainage were significantly related to elevated nitrate concentration. The best fitting pesticide model resulted in strong correlation (linear correlation coefficient = 0.960) between observed rates of atrazine/desethyl-atrazine detection and probabilities of detection predicted by the model.

Teso and others (1996) used LR to predict the probability of occurrence of DBCP (a nematicide) in ground water beneath land sections of eastern Fresno County, California. Soil particle-size classes (e.g., sandy, loamy) were used as independent variables in the model-building process. The resulting model was statistically significant ($p = 0.017$) and included sandy and fine particle-size classes. The model correctly predicted the contamination status of contaminated sections 89.7 % of the time, but the overall success rate (considering both contaminated and uncontaminated sections) was only 53.2 %. The overall success rate might have been affected by the small number of samples associated with the uncontaminated sections.

Methods

The data set used for this paper comprises 1,230 wells sampled during selected land-use studies conducted in the first 20 NAWQA study areas (called "study units") (Figure 1), which started in 1991 and sampled ground water during 1992-1995. "Land-use studies" evaluate the quality of recently recharged ground water (generally less than 10-years old) in regions that represent the intersection of a targeted land use and an aquifer of interest. Each study unit typically contains one or more land-use studies. The 54 shallow land-use studies used in this research typically comprise 20–30 wells each and range in size from 146 to 62,900 km^2 (median area is 4,370 km^2). Wells sampled in land-use studies usually are installed by NAWQA, and public-supply wells are avoided

because of uncertainties in the location of the recharge area. Springs and agricultural drains were excluded from the analysis because of uncertainties in the source of water and/or contributing land-use area. To preclude undue influence on results by wells that were sampled several times, only one sample per well was used in statistical analyses. The sample selected for analysis represents the date that the land-use study network was sampled by NAWQA for a large suite of water-quality parameters, which occurs once per high-intensity phase of study. Isolated follow-up samples (e.g., for a limited number of parameters to assess seasonal trends) were not considered because these dates vary widely and do not reflect the network sampling date.

All wells were sampled according to procedures described by Koterba and others (1995). Nitrite-plus-nitrate was analyzed by the USGS National Water Quality Laboratory using procedures described in Fishman (1993), and reported concentrations are based on elemental N (e.g., NO_2^- plus NO_3^- as N). Nitrite-plus-nitrate concentration is referred to in this paper as "nitrate" because nitrite contribution to nitrite-plus-nitrate in ground water generally is negligible (Nolan and Stoner, 2000).

Characteristics of N loading and aquifer susceptibility to contamination and their influence on nitrate concentration in shallow ground water were evaluated by developing a set of explanatory variables (Table 1) for use in LR. To develop the N-loading variables, 1991-1993 data representing atmospheric deposition, animal manure (1992 only), and commercial fertilizer sources were compiled by land use within circular, 500-meter (m) radius areas surrounding sampled wells. Fertilizer N was apportioned equally to agricultural and urban land within 500 m of sampled wells to account for residential fertilizer use. Use of residential fertilizer is intensive in heavily populated areas such as Long Island, New York, where estimated annual application rates are as high as 180 kg/hectare (ha) (Porter, 1980). Manure N was apportioned to agricultural land only, and atmospheric N was assumed independent of land use within 500 m of sampled wells. See Nolan and Stoner (2000) for detailed discussion on compilation of N loading variables and other ancillary data. Fertilizer N was considered separately and in combination with manure N and atmospheric N to assess "total N" contribution. In addition to N-loading data, the percent of Anderson Level II cropland-pasture (Anderson and others, 1976) and 1990 population density (U.S. Bureau of the Census, 1991) within 500 m of sampled wells were used as measures of agricultural and urban intensity, respectively. The Anderson land-use data were updated with 1990 Census population data to indicate recent conversion of agricultural land to new residential land (Hitt, 1994). Population density was used as a surrogate for nonagricultural sources of N in urban areas.

Aquifer susceptibility variables such as soil hydrologic group (HYDGRP), depth to the seasonally high water table, and percent organic matter were compiled from State Soil Geographic (STATSGO) data (Soil Conservation Service, 1994) as weighted averages within areas representing NAWQA land-use studies (Table 1), using methods described in Nolan and Stoner (2000). These variables were compiled within land-use study areas because the STATSGO mapping units typically are much larger than the 500-m radius circular areas used to compile land-use attributes. STATSGO attributes do not vary appreciably within the small circular areas and the resulting estimate would have resembled a point estimate rather than a weighted average.

Percentages of soil HYDGRPs A and B in land-use study areas were summed to represent "well-drained" soils (Table 1). Soils in HYDGRP A are defined as "deep, well

drained to excessively drained sands or gravels," and HYDGRP B soils are defined as "moderately deep to deep, moderately well drained to well drained soils that have moderately fine to moderately coarse textures" (Soil Conservation Service, 1993). Depth to the seasonally high water table represents the average, minimum unsaturated zone thickness in a land-use study area.

Percent organic matter was used to represent denitrification potential in aquifers (Table 1). It was hoped that percent organic matter in combination with depth to the seasonally high water table in a multivariate model would indicate conditions conducive to denitrification. Denitrification is a microbially assisted process that transforms nitrate to N_2 gas under reducing conditions when organic matter and/or selected reduced minerals are present (Korom, 1992).

A binary variable indicating presence or absence of fractured rock in a land-use study area was coded "1" if the aquifer associated with a given land-use study consists of fractured rocks and "0" if it does not (Table 1). Data indicating the presence or absence of rock fractures were compiled from lithologic descriptions of each land-use study area after consulting with local NAWQA personnel (Dana W. Kolpin, USGS, unpublished data, 2000).

The percent artificially drained soils in land-use study areas (Table 1) was determined by compiling 1992 National Resources Inventory data (Soil Conservation Service, 1994) in a geographic information system (GIS) (Kerie J. Hitt, USGS, unpublished data, 1998). Woodland-to-cropland ratio was calculated by dividing combined percentages of Anderson Level II deciduous forest, evergreen forest, and mixed forest lands within 500 m of sampled wells by combined percentages of cropland-pasture, orchards, groves, vineyards, nurseries, ornamental horticultural areas, and confined feeding operations.

The depth to the top of the open interval (well screen or open borehole) below water level ("sampling depth" in Table 1) was calculated by subtracting depth to ground water from open interval depth, for wells in which open interval depth is greater than depth to water. If open interval depth is less than or equal to depth to ground water, then sampling depth was set to zero. In unconfined aquifers typical of land-use studies, sampling depth represents the distance from the water-table surface to the top of the open interval. Deeper sampling depths generally correspond to older ground water, which is less likely to show effects from recent land use.

Mean annual precipitation within land-use study areas was compiled in a GIS by David M. Wolock (USGS, unpublished data, 1998) for the period 1961-1990 (National Climatic Data Center, 1994) (Table 1). The precipitation data were used to represent the potential for ground-water recharge, which influences nitrate transport through the unsaturated zone.

Descriptive statistics in Table 1 include the median and interquartile range. Medians were used as a measure of central tendency because they are resistant to the effects of outliers typical of skewed data sets. Similarly, the interquartile range (75^{th} percentile minus the 25^{th} percentile) is a resistant measure of spread. It consists of the middle 50 percent of the data; thus it is not influenced by the 25 percent at either end (Helsel and Hirsch, 1992).

Logistic regression was used to predict the likelihood that nitrate concentration in shallow ground water exceeds 4 mg/L. This value indicates effects of human activities

on ground-water quality and also has health significance—the 4 mg/L level has been associated with increased risk of non-Hodgkin's lymphoma in Nebraska (Ward and others, 1996). Model-fitting criteria evaluated in this study consist of the likelihood ratio test statistic (G), the Wald statistic, and the Hosmer-Lemeshow goodness-of-fit test statistic. Explanatory variables in the final model are considered to most influence nitrate concentration in shallow ground water, based on this data set.

Linear regression was used as an additional indicator of goodness-of-fit to compare observed and average predicted probabilities associated with the deciles used to calculate the HL statistic. The coefficient of determination (r^2) was computed for the observed and predicted probabilities, with higher r^2 values indicating better fit. In a related qualitative check, predicted and observed probabilities were plotted and visually compared with a 1:1 line having an intercept of zero. If all of the points fell on the 1:1 line, then the predicted and observed probabilities would agree perfectly.

Results and Discussion

Univariate Logistic-Regression Models

Univariate LR models were developed for the explanatory variables in Table 1 to screen variables for inclusion in subsequent multivariate models. Some of the univariate models are designed to test multiple influences on nitrate contamination of ground water to help identify combinations of variables for use in subsequent multivariate models. For example, models consisting of fertilizer loading in HYDGRP B soils (UV2) and in HYDGRP A and B soils (UV3) investigate the influence of N loading in moderately well-drained to well-drained soils (Table 2).

All of the univariate models except UV6, consisting of population density (G statistic p-value = 0.075), are statistically significant at the 0.05 level (Table 2). All of the univariate models, however, have very low HL p-values, indicating that none fit the data well. The highest HL p-value, only 0.005, is for model UV5 (percent cropland-pasture). (Higher HL p-values indicate better fit because the null hypothesis is that the model fits the data.)

The strength of correlation between observed and predicted probabilities corresponding to deciles of risk was used as an additional indicator of goodness-of-fit to help select variables for inclusion in an initial multivariate LR model. The r^2 values range from 0.086 for model UV15 (precipitation) to 0.760 for model UV8 (percent well-drained soils) (Table 2). Figure 2 shows how linear-regression r^2 indicates degree of logistic-regression model fit. Observed and predicted probabilities associated with model UV8 ($r^2 = 0.760$) generally follow the exact-fit 1:1 line, indicating reasonable agreement between observed and predicted probabilities (Figure 2a). In contrast, observed and predicted probabilities for model UV9 ($r^2 = 0.119$) deviate sharply from the 1:1 line (Figure 2b), indicating poor fit.

Multivariate Logistic-Regression Models

Explanatory variables from the best-fitting univariate models were combined into an initial multivariate LR model (MV1) to evaluate the simultaneous influence of variables that significantly affect nitrate contamination of shallow ground water. The G statistic and strength of correlation between observed and predicted probabilities were used to

select variables for inclusion in the multivariate model. Univariate models consisting of the following variables had statistically significant G statistics (p-value <0.05) and/or r^2 values of about 0.6 or greater, indicating reasonable correlation between observed and predicted probabilities: (1) fertilizer N in HYDGRP B soils, (2) percent cropland-pasture, (3) percent well-drained soils, (4) depth to seasonally high water table, (5) sampling depth, and (6) presence/absence of a rock fracture (Table 2). The r^2 value for sampling depth is only 0.298, but a plot of observed and predicted probabilities appears reasonable except for three points that are somewhat removed from the superimposed 1:1 line indicating exact fit. This variable also was retained to explore a potential interaction with N fertilizer loading. Woodland-to-cropland ratio has a high r^2 value (0.741), but was not retained because it was calculated using percent cropland-pasture, which is already in the model, and because visual inspection of the plot of observed and predicted probabilities revealed a poor fit. Half the points are well removed from the superimposed 1:1 line.

Multivariate model MV1, consisting of fertilizer N and variables 2-6 above (Table 3), yielded an HL p-value of 0.067 and thus fits the data somewhat better than the univariate models; but the fit could be much improved. Model MV1 has a highly significant G statistic p-value ($p < 0.001$), compared with the intercept only model that contains none of the explanatory variables.

The presence of both fertilizer N and percent cropland-pasture in model MV1 raises potential multicollinearity concerns. Multicollinearity arises when two or more explanatory variables are closely related, and can result in unrealistic model coefficient sign, unstable slope coefficients, and other problems (Helsel and Hirsch, 1992). If multicollinearity were present, however, the standard error of both fertilizer N and percent cropland-pasture (indicated by the Wald p-values) would be very large because the LR model would not be able to select from among the two variables (Gregory E. Schwarz, USGS, unpublished data, 2000). Wald p-values for model MV1 are highly significant for both fertilizer N (0.041) and percent cropland-pasture (0.001), dispelling multicollinearity concerns. Percent cropland-pasture apparently contains information not embodied in fertilizer N, such as N loading from manure, atmospheric deposition, septic systems, or other sources. Additionally, percent cropland-pasture incorporates information on crop type and tillage practice, which affect N transformations in soil and the efficiency of N uptake by crops.

Exclusion of "total" N (comprising fertilizer, manure, and atmospheric deposition sources) from model MV1 does not mean that manure and atmospheric deposition are insignificant as N sources. Compared with inorganic fertilizer, manure and atmospheric deposition contribute lesser amounts of N annually in the United States (Puckett, 1995) but are important regional sources. Additionally, these variables likely are reflected in the percent cropland-pasture variable. The fertilizer N variable was selected based on statistical performance with the NAWQA national data set.

Nested Multivariate Logistic-Regression Models

Nested LR models were tested to see if addition or removal of a single variable would significantly affect model performance. Models in Table 3 are shown in order of nesting—each successive model within a category (e.g., "initial nested comparison") includes one additional variable. All models shown in Table 3 consist of 987 observations to facilitate unbiased comparisons of model-fitting criteria.

The difference between models MV1 and MV3 is that MV1 contains the percent cropland-pasture variable and MV3 does not, providing an additional check on potential multicollinearity between fertilizer N and percent cropland-pasture. The G statistic for MV1 (p-value = 0.001) indicates that the nested model containing percent cropland-pasture is significantly improved over MV3 at the 0.05 level (Table 3).

An interaction term was tested with model MV1 to see if model performance could be improved. An interaction is present when a covariate (e.g., age) modifies the effect of a risk factor (e.g., gender) on outcome (Hosmer and Lemeshow, 1989). For example, the effect of gender on outcome depends on the age at which the gender comparison is being made. Model MV2 contains the interaction term fertilizer N*sampling depth to test whether the effect of fertilizer loading on nitrate contamination of ground water depends on sampling depth within the aquifer (Table 3). Deeper sampling depths represent older ground water, which is less likely to show effects from recent fertilizer application (Nolan and Stoner, 2000). The G statistic p-value for model MV2, however, is 0.141, indicating that the model is not significantly improved over MV1 at the 0.05 level. Because NAWQA land-use studies are designed to sample shallow, recently recharged ground-water, sampling depth might not vary sufficiently to influence nitrate concentration. The interquartile range of sampling depth (5.3 m) is comparatively low. Additionally, the sign of the sampling depth coefficient used to calculate the interaction term is positive (0.001) in model MV1 and negative (-0.005) in MV2, and the Wald p-values for sampling depth are high for both models (0.660 and 0.417). For these reasons, the interaction term was excluded from subsequent multivariate models.

Reverse selection was attempted to see if removing sampling depth (Wald p-value = 0.660) from model MV1 would improve performance. Reverse-selection results in Table 3 are shown in forward order to allow comparison of G statistics associated with nested models MV4 and MV1. The G statistic p-value for model MV1 is 0.661, indicating that sampling depth is statistically insignificant at the 0.05 level. Wald p-values in the resulting model (MV4) are less than 0.05 for all explanatory variables. Model MV4 is considered better than MV1 based on the highly significant Wald p-values associated with the explanatory variables.

As a final check, the remaining explanatory variables rejected as part of univariate model screening each were introduced into model MV4 to see if they could improve model performance and, in the case of population density, attain statistical significance. With the exception of population density, the remaining variables yielded insignificant Wald p-values ranging from 0.19 (mean annual precipitation) to 0.91 (percent organic matter in soils). Inclusion of population density (model MV5), however, yielded a G statistic p-value of <0.001, indicating that MV5 is significantly improved over model MV4 (Table 3). Additionally, the HL p-value of 0.377 is much improved over that of initial multivariate model MV1 (p = 0.067).

It is important to check the assumption of linearity in the logit after selecting variables for inclusion in a logistic-regression model (Hosmer and Lemeshow, 1989). Natural log (ln) transformation of population density (model MV6 in Table 4) resulted in a more linear response of the logit function and yielded an HL p-value of 0.641, considerably improved over that of model MV5 (p = 0.377).

Evaluating HL p-values is somewhat subjective because little information exists on what range of values constitute acceptable fit. In a study of risk factors associated with

low birth-weight infants, an HL p-value of 0.73 indicated that an LR model fit the data "quite well" (Hosmer and Lemeshow, 1989). In the current paper, the HL p-value of 0.641 is considered evidence of good fit based on visual comparison of plotted observed and predicted probabilities for deciles of risk, and the associated linear-regression r^2 value. The observed and predicted probabilities follow close to the 1:1 exact-fit line (Figure 3), and the r^2 value of 0.971 indicates strong correlation. Model MV6 (Table 4) is considered "best" according to study objectives and fitting criteria emphasized in this study.

Final Multivariate Logistic-Regression Model and Related Explanatory Variables

Of the explanatory variables initially considered, those in MV6 are considered to most significantly influence nitrate contamination of shallow ground water, based on the data set used in this study. Model MV6 consists of variables representing (1) fertilizer N loading (p-value = 0.012), (2) percent cropland-pasture ($p < 0.001$), (3) ln(population density) ($p < 0.001$), (4) percent well-drained soils ($p = 0.002$), (5) depth to the seasonally high water table ($p = 0.001$), and (6) presence or absence of a fracture zone within an aquifer ($p = 0.002$) (Table 4). The Wald p-values associated with these variables are highly significant at the 0.05 level.

The first three variables represent N source terms and the last three variables represent aquifer susceptibility to contamination. Together, they form a conceptual model of aquifer vulnerability consisting of N loading to the land surface followed by transfer of nitrate to ground water through the unsaturated zone. In this model, the degree to which nitrate from aboveground N sources leaches to ground water is influenced by soil type, water-table position, and the presence or absence of fractured rocks.

Slope coefficients for fertilizer N (0.005), percent cropland-pasture (0.015), and ln(population density) (0.194) are positive (Table 4), indicating that the likelihood of nitrate contamination of ground water increases with increasing levels of N sources. This agrees with prior logistic-regression studies in which the probability of nitrate contamination of ground water increased with increasing percentages of agricultural land near sampled wells (Eckhardt and Stackelberg, 1995; Tesoriero and Voss, 1997). Relations between fertilizer N, extent of agricultural land, and nitrate concentration in ground water are well documented. Hall (1992) analyzed N applications and nitrate concentrations for five wells on a farm in the Conestoga Valley of southeastern Pennsylvania. He observed cause-and-effect relations between changes in rates of applied N in manure and fertilizers and changes in ground-water nitrate concentration, indicating that a significant amount of the applied N is transported with recharge to ground water within four to 19 months after application. Ground-water nitrate concentration generally decreased after implementation of nutrient management plans that saw reductions of N application rates. Time-series data indicated a close relation between applied N and nitrate in ground water, and correlations between applied N and ground-water nitrate were statistically significant at the 90-percent confidence level for all five wells.

Hallberg and Keeney (1993) cited studies that found a direct relation between agricultural land use and nitrate in ground water. As evidence they cite 3- to 60-fold increases of nitrate observed in ground water in agricultural areas, whereas ground water

beneath forestland, grassland, and even pastured areas generally contained less than 2 mg/L nitrate. They describe significant positive correlations between nitrate concentrations in ground water and the percentage area of fertilized crops and with N fertilizer application rates in the vicinity of sampling sites. Additionally, a basin-scale study showed ground-water nitrate increasing in direct proportion to increasing fertilizer use. Nitrogen amounts contributed by rainfall, crop rotation, and soil mineralization were considered in the N budget for the basin.

Nolan and Stoner (2000) showed that nitrate concentration in shallow ground water beneath urban lands generally increased with increasing population density. Median nitrate concentration was as high as 5.4 mg/L in areas of the Willamette Basin near Portland, Oregon. The population of Portland, the state's largest metropolitan area, was about 1.2 million people in 1990, and median population density within 500 m of sampled wells in the area was 2,300 people/km^2. Areas with more than 386 people/km^2 are considered "urban," based on GIS analysis of 1990 Census data and Anderson land-use data (Hitt, 1994).

Domestic sewage and residential fertilizer are major sources of N in some heavily populated areas. Septic systems and cesspools have long been sources of nitrate in ground waters of Long Island, New York, and turf grass is the major crop in Nassau County (Porter, 1980). Annual application rates of N fertilizer to residential lawns in the area ranged from an estimated 80 to 180 kg/ha in the mid 1970s.

Aquifer susceptibility terms in model MV6 comprise percent well-drained soils, depth to the seasonally high water table, and the presence or absence of a rock fracture. The slope coefficient for percent well-drained soils is positive (0.017), indicating that the likelihood of nitrate contamination of ground water increases with better drainage (Table 4). This result agrees with prior research. Tesoriero and Voss (1997) showed that the predicted probability of nitrate contamination of ground water in the Puget Sound Basin is greater for shallow wells in coarse-grained glacial deposits than for shallow wells in fine-grained glacial deposits and in alluvium. Rupert (1998) found that STATSGO soil hydrologic group significantly improved multivariate LR models used to predict contamination of ground water by atrazine/desethyl-atrazine, and that STATSGO soil drainage characteristic significantly improved models for predicting nitrate contamination of ground water. Soil drainage data from STATSGO denote the frequency and duration of periods when soil is free from saturation (Soil Conservation Service, 1994). Poorly drained soils commonly are saturated and can have reducing conditions conducive to denitrification, which lessens the likelihood of nitrate contamination of ground water. Similarly, Burkart and others (1999) observed that nitrate concentration in shallow, unconfined aquifers was negatively correlated with STATSGO HYDGRP C soils, which have moderately fine to fine soil texture and contain a layer that restricts downward movement of water (Soil Conservation Service, 1993). In addition to promoting conditions conducive to denitrification, HYDGRP C soils more likely are artificially drained for improved crop production, which diverts nitrate in infiltrating ground water to nearby streams (Burkart and others, 1999).

The slope coefficient for depth to the seasonally high water table in model MV6 is positive (0.850) (Table 4), indicating that as depth increases, the likelihood of nitrate contamination increases. This result agrees with findings by Burkart and others (1999), who observed a positive correlation between STATSGO seasonally high water table and

nitrate concentration in shallow, unconfined aquifers. These results at first seem counterintuitive because increasing depth to water generally involves greater travel distance and potential to encounter intervening, less permeable layers that inhibit leaching. The NAWQA land-use studies, however, are designed to consistently sample shallow, recently recharged ground water (median depth to water = 4.4 m for this data set). Because depth to ground water is uniformly shallow, travel distance is minimal and the potential for intervening layers is low. Very shallow depth to ground water creates anoxic conditions, which promote denitrification (Böhlke and Denver, 1995; Nolan, 1999; Spruill and others, 1998). Denitrification is fueled by organic matter and selected reduced minerals under anoxic conditions. Increasing depth to the water table reduces the likelihood that soils are saturated, lessening denitrification potential and increasing the likelihood of nitrate contamination of ground water. Korom (1992) provides detailed discussion of denitrification in the saturated zone of aquifers.

An additional explanation for the positive sign of the slope coefficient for depth to the seasonally high water table is that agricultural land is more likely found on well-drained soils than on poorly drained soils. The Spearman correlation between percent cropland-pasture and depth to the seasonally high water table is 0.19, for the data set used in this study. The positive correlation suggests that use of agricultural chemicals is greater in areas with more well-drained soils (i.e., with greater depth to ground water), increasing the likelihood of nitrate contamination of ground water in these areas.

The slope coefficient in model MV6 for presence or absence of a fracture zone is positive (1.033) (Table 4), indicating that the likelihood of nitrate contamination of ground water increases in areas with fractured rocks. Rock fractures can readily convey contaminants to ground water, even in areas where depth to ground water is greater. Water from an aquifer comprising fractured crystalline rocks in southeastern Pennsylvania has a median nitrate concentration of 6.6 mg/L, and the nitrate MCL of 10 mg/L is exceeded in 31 percent of the samples (Nolan and Stoner, 2000). Median depth to ground water in the area is greater (12.8 m) than reported here (4.4 m), but the fractured rocks are susceptible to recharge of water and chemicals from the land surface (Lindsey and others, 1998). Land use in the area consists of mixed forest and agriculture, and ground water is N-rich near hilltops where the agricultural land is most dense.

Data from the Upper Devonian aquifer in northern Iowa indicate that nitrate concentration in ground water is high in karst areas. Karst is eroded limestone that contains fractures and sinkholes that enhance recharge to aquifers. Such features make the rocks extremely porous, especially in areas where overlying deposits are thin or missing. Median nitrate concentration in ground water was 9.6 mg/L in karst material, compared with 6.9 mg/L in very shallow bedrock and <0.1 mg/L in deep bedrock (Hallberg and Keeney, 1993). Surficial recharge delivered nitrate from agricultural areas to the Upper Devonian aquifer in the karst and shallow bedrock areas. Tritium data indicated that modern (post-1953) water has migrated into the aquifer to depths greater than 30 m in these areas. In contrast, the deeper bedrock aquifers contain older water with significantly lower nitrate concentration.

Conclusions

Multivariate logistic regression was used in a national-scale analysis to identify variables that significantly influence nitrate contamination of shallow, recently recharged

ground water. The LR model predicts the likelihood that nitrate in ground water exceeds 4 mg/L. The final model consists of variables representing (1) fertilizer N loading (Wald $p = 0.012$), (2) percent cropland-pasture ($p < 0.001$), (3) ln(population density) ($p < 0.001$), (4) percent well-drained soils ($p = 0.002$), (5) depth to the seasonally high water table ($p = 0.001$), and (6) presence or absence of a fracture zone within the aquifer ($p = 0.002$). The Wald p-values are highly significant at the 0.05 level. A goodness-of-fit test indicates that the model fits the data very well, and observed and predicted probabilities of nitrate contamination are strongly correlated ($r^2 = 0.971$). The multivariate LR model fits the data much better than do any of the preliminary univariate models. Based on the model, nitrate contamination of ground water is not caused by any single factor but depends on the combined, simultaneous influence of factors representing N loading sources and aquifer susceptibility characteristics.

Acknowledgments

I thank the many NAWQA personnel who collected and compiled the water-quality data used in this study. I also thank Kerie J. Hitt, Barbara C. Ruddy, and David M. Wolock for compiling and analyzing data in geographic information systems. Lastly, thanks to John Chilton, Paul E. Stackelberg, Anthony J. Tesoriero, and Chester Zenone for providing insightful reviews that substantially improved this paper.

References

Afifi, A.A., and V. Clark. 1984. Logistic Regression. 287-308. *In* Computer-Aided Multivariate Analysis. Lifetime Learning Publ., Belmont, CA.

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer. 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 964.

Böhlke, J.K., and J.M. Denver. 1995. Combined use of groundwater dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds, Atlantic coastal plain, Maryland. Wat. Resources Res. 31:2319-2339.

Burkart, M.R., D.W. Kolpin, R.J. Jaquis, and K.J. Cole. 1999. Agrichemicals in ground water of the midwestern USA: relations to soil characteristics. Ground Water 28:1908-1915.

Centers for Disease Control and Prevention. 1996. Spontaneous abortions possibly related to ingestion of nitrate-contaminated well water—LaGrange County, Indiana, 1991-1994. Morbidity and Mortality Weekly Report 45:569-572.

D'Agostino, V., E.A. Greene, G. Passarella, and M. Vurro. 1998. Spatial and temporal study of nitrate concentration in groundwater by means of coregionalization. Environmental Geology 36:285-295.

Davies, D.B., T.W.D. Garwood, and A.D.H. Rochford. 1996. Factors affecting nitrate leaching from a calcareous loam in East Anglia. Journal of Agricultural Science, Cambridge 126:75-86.

Eckhardt, D.A.V., and P.E. Stackelberg. 1995. Relation of ground-water quality to land use on Long Island, New York. Ground Water 33:1019-1033.

Fishman, M.J. (ed.) 1993. Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—determination of inorganic and organic constituents in water and fluvial sediments. U.S. Geological Survey Open-File Report 93-125.

Hall, D.W. 1992. Effects of nutrient management on nitrate levels in ground water near Ephrata, Pennsylvania. Ground Water 30:720-730.

Hallberg, G.R., and D.R. Keeney. 1993. Nitrate. *In* Alley, W.M., ed., Regional Ground-Water Quality. Van Nostrand Reinhold, New York.

Helsel, D.R., and R.M. Hirsch. 1992. Statistical Methods in Water Resources. Elsevier, New York.

Hitt, K.J. 1994. Refining 1970's land-use data with 1990 population data to indicate new residential development. U.S. Geological Survey Water-Resources Investigations Report 94-4250.

Hosmer, D.W., and S. Lemeshow. 1989. Applied Logistic Regression. John Wiley and Sons, New York.

Kleinbaum, D.G. 1994. Logistic Regression: a Self-Learning Text. Springer-Verlag, New York.

Korom, S.F. 1992. Natural denitrification in the saturated zone: a review. Wat. Resources Res. 28:1657-1668.

Koterba, M.T., F.D. Wilde, and W.W. Lapham. 1995. Ground-water data-collection protocols and procedures for the National Water-Quality Assessment Program: collection and documentation of water-quality samples and related data. U.S. Geological Survey Open-File Report 95-399.

Lindsey, B.D., K.J. Breen, M.D. Bilger, and R.A. Brightbill. 1998. Water quality in the Lower Susquehanna River Basin, Pennsylvania and Maryland, 1992-95. U.S. Geological Survey Circular 1168.

Meinardi, C.R., A.H.W. Beusen, M.J.S. Bollen, O. Klepper, and W.J. Willems. 1995. Vulnerability to diffuse pollution and average nitrate contamination of European soils and groundwater. Wat. Sci. Tech. 31:159-165.

Mueller, D.K., and D.R. Helsel. 1996. Nutrients in the Nation's waters—too much of a good thing? U.S. Geological Survey Circular 1136.

National Climatic Data Center. 1994. National Climatic Data Center, Asheville, North Carolina.

National Research Council. 1993. Ground Water Vulnerability Assessment—Predicting Relative Contamination Potential Under Conditions of Uncertainty. National Academy Press, Washington, D.C.

Nolan, B.T. 1999. Nitrate behavior in ground waters of the southeastern USA. J. Environ. Qual. 28:1518-1527.

Nolan, B.T., and J.D. Stoner. 2000. Nutrients in groundwaters of the conterminous United States, 1992-1995. Environ. Sci. Technol. 34:1156-1165.

Porter, K.S. 1980. An evaluation of sources of nitrogen as causes of ground-water contamination in Nassau County, Long Island. Ground Water 18:617-625.

Puckett, L.J. 1994. Nonpoint and point sources of nitrogen in major watersheds of the United States. U.S. Geological Survey Water-Resources Investigations Report 94-4001.

Puckett, L.J. 1995. Identifying the major sources of nutrient water pollution. Environ. Sci. and Technol. 29:408-414.

Riley, J.A., R.K. Steinhorst, G.V. Winter, and R.E. Williams. 1990. Statistical analysis of the hydrochemistry of ground waters in Columbia River basalts. Journal of Hydrology 119:245-262.

Rupert, M.G. 1998. Probability of detecting atrazine/desethyl-atrazine and elevated concentrations of nitrate (NO_2 + NO_3-N) in ground water in the Idaho part of the Upper Snake River Basin. U.S. Geological Survey Water-Resources Investigations Report 98-4203.

Soil Conservation Service. 1993. National Soils Survey Handbook, title 430-VI. U.S. Department of Agriculture, Soil Conservation Service, Washington, D.C.

Soil Conservation Service. 1994. 1992 National Resources Inventory (CDROM). U.S. Department of Agriculture, Soil Conservation Service, Ft. Worth, Texas.

Soil Conservation Service. 1994. State Soil Geographic (STATSGO) data base for the United States and Puerto Rico (CDROM). U.S. Department of Agriculture, Soil Conservation Service, Ft. Worth, Texas. Also accessible through the internet at <URL: http://www.ftw.nrcs.usda.gov/stat_data.html>.

Solley, W.B., R.R. Pierce, and H.A. Perlman. 1993. Estimated use of water in the United States in 1990. U.S. Geological Survey Circular 1081.

Spalding, R.F., and M.E. Exner. 1993. Occurrence of nitrate in groundwater—a review. J. of Environ. Qual. 22:392-402.

Spruill, T.B., D.A. Harned, P.M. Ruhl, J.L. Eimers, G. McMahon, K.E. Smith, D.R. Galeone, and M.D. Woodside. 1998. Water quality in the Albemarle-Pamlico drainage basin, North Carolina and Virginia, 1992-95. U.S. Geological Survey Circular 1157.

Teso, R.R., M.P. Poe, T. Younglove, and P.M. McCool. 1996. Use of logistic regression and GIS modeling to predict groundwater vulnerability to pesticides. Journal of Environ. Qual. 25:425-432.

Tesoriero, A.J., E.L. Inkpen, and F.D. Voss. 1998. Assessing groundwater vulnerability using logistic regression. *In* Proceedings of the Source Water Assessment and Protection Conference, Dallas, TX, April 28-30, 1998.

Tesoriero, A.J., and F.D. Voss. 1997. Predicting the probability of elevated nitrate concentrations in the Puget Sound Basin: implications for aquifer susceptibility and vulnerability. Ground Water 35:1029-1039.

U.S. Bureau of the Census. 1991. 1990 Census of Population and Housing, Public Law 94-171 Data (United States). The Bureau, Washington D.C.

U.S. Environmental Protection Agency. 1995. Drinking water regulations and health advisories. Office of Water, Washington, D.C.

Ward, M.H., S.D. Mark, K.P. Cantor, D.D. Weisenburger, A. Correa-Villaseñor, and S.H. Zahm. 1996. Drinking water nitrate and the risk of non-Hodgkin's lymphoma. Epidemiology 7:465-471.

Table 1. Explanatory variables and descriptive statistics

Variable	*GIS compilation area*	*Minimum*	*Median*	*Maximum*	*Interquartile Range*	*Number of observations*
Nitrogen sources						
N fertilizer loading, kg/ha	500-m well buffer	0.0	27.5	180	64.5	1,230
Total N[a], kg/ha	500-m well buffer	1.1	60.2	224	82.3	1,230
Cropland-pasture, %	500-m well buffer	0.0	84.9	100	96.8	1,230
Population density, people/km^2	500-m well buffer	0.1	19.8	4,135	165	1,230
Aquifer susceptibility						
HYDGRP B soils[b], %	Land-use study area	8.8	41.9	78.4	24.9	1,230
HYDGRP A and B soils, or "well-drained,"[b] %	Land-use study area	16.1	56.0	87.6	22.9	1,230
Organic matter in soils, % by wt.	Land-use study area	0.1	0.6	10.6	1.1	1,230
Depth to seasonally high water table, m	Land-use study area	0.4	1.5	1.8	0.5	1,230
Presence or absence of rock fracture (binary indicator = 0 or 1)	Land-use study area	0	0	1	0	1,199
Artificially drained soils, %	Land-use study area	0.0	0.2	39.0	3.2	1,230
Woodland-to-cropland ratio, %/%	500-m well buffer	0.0	0.0	82.3	0	953
Depth to top of open interval below water, or "sampling depth," m	Well point	0.0	2.0	112	5.3	1,054
Mean annual precipitation, 1961-90, cm	Land-use study area	11.7	97.0	138	84.3	1,230

[a]total N = sum of N loading from fertilizer, manure, and atmospheric deposition
[b]HYDGRP = STATSGO soil hydrologic group

Table 2. Fitting criteria for univariate logistic-regression models

	Model	*Estimated coefficient*	*Likelihood ratio (G) p-value*[a]	*Hosmer-Lemeshow statistic p-value*	*r^2 for obs. and pred. prob.*	*Number of observ.*
	Sources					
UV1	N fertilizer loading, kg/ha	0.0082	<0.001	<0.001	0.368	1,214
UV2	N fertilizer in HYDGRP B soils[b], kg/ha	0.0832	<0.001	<0.001	0.586	191
UV3	N fertilizer in HYDGRP A and B soils[b], kg/ha	0.0064	<0.001	<0.001	0.232	757
UV4	Total N[c] in HYDGRP A and B soils[b], kg/ha	0.0057	<0.001	<0.001	0.253	757
UV5	Cropland-pasture, %	0.0115	<0.001	0.005	0.723	1,214
UV6	Population density, people/km^2	-0.0002	0.075	<0.001	0.150	1,214
	Aquifer susceptibility					
UV7	HYDGRP B soils[b], %	0.0379	<0.001	<0.001	0.472	1,214
UV8	HYDGRP A and B soils, or "well-drained,"[b] %	0.0362	<0.001	<0.001	0.760	1,214
UV9	Organic matter in soils, % by wt.	-0.1070	0.002	<0.001	0.119	1,214
UV10	Depth to seasonally high water table, m	1.5939	<0.001	<0.001	0.631	1,214
UV11	Presence or absence of rock fracture	1.9002	<0.001	--	--	1,183
UV12	Artificially drained soils, %	-0.0385	<0.001	<0.001	0.357	1,214
UV13	Woodland-to-cropland ratio, %/%	-0.2065	0.001	<0.001	0.741	941
UV14	Depth to top of open interval below water, or "sampling depth," m	0.0457	<0.001	<0.001	0.298	1,038
UV15	Mean annual precipitation, 1961-90, cm	-0.0060	<0.001	<0.001	0.086	1,214

[a]for $G = -2(L_o - L)$, where L_o is for intercept-only model
[b]HYDGRP = STATSGO soil hydrologic group
[c]total N = sum of N loading from fertilizer, manure, and atmospheric deposition

Table 3. Fitting criteria for competing multivariate logistic-regression models

	Model[a]	*Likelihood ratio (G) p-value*[b]	*Number of observations*
	Initial nested comparison		
MV3	fert, welldr, wtdep, sampdep, bfract	--	987
MV1	fert, welldr, wtdep, sampdep, bfract, pctcrop	0.001	987
MV2	fert, welldr, wtdep, sampdep, bfract, pctcrop, frtxsd	0.141	987
	Reverse selection		
MV4	fert, pctcrop, welldr, wtdep, bfract	--	987
MV1	fert, pctcrop, welldr, wtdep, bfract, sampdep	0.661	987
	Final nested comparison		
MV4	fert, pctcrop, welldr, wtdep, bfract	--	987
MV5	fert, pctcrop, welldr, wtdep, bfract, popden	<0.001	987

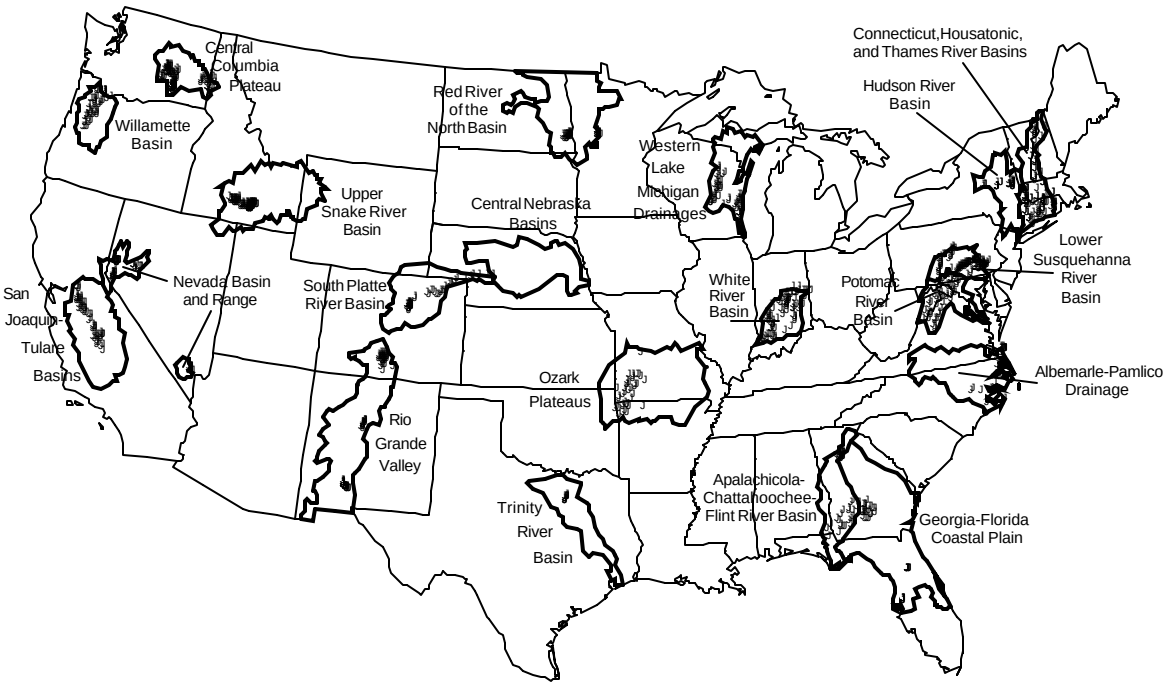
[a]fert = fertilizer N; welldr = percent well-drained soils, or the sum of percentages of soil hydrologic groups A and B; wtdep = depth to seasonally high water table; sampdep = sampling depth, or depth to top of well screen or open borehole below water level; bfract = binary variable indicating presence or absence of fractured rocks; pctcrop = percent cropland-pasture; frtxsd = interaction between fertilizer N and sampling depth; popden = population density

[b]$G = 2(L_c - L_s)$, where L_c is more complex, nested model and L_s is simpler model (Helsel and Hirsch, 1992)

Table 4. Explanatory variables in final multivariate logistic-regression model MV6

Variable	*Estimated coefficient*	*Wald p-value*
Constant	-4.485	<0.001
Fertilizer N, kg/ha	0.005	0.012
Cropland-pasture, %	0.015	<0.001
ln(population density), ln(people/km^2)	0.194	<0.001
Well-drained soils[a], %	0.017	0.002
Depth to season. high water table, m	0.850	0.001
Presence or absence of rock fracture	1.033	0.002

[a]sum of percentages of soil hydrologic groups A and B in area



Wells in NAWQA land-use studies conducted during 1992-1995

Figure 1. Locations of shallow wells sampled as part of NAWQA land-use studies conducted during 1992-1995.

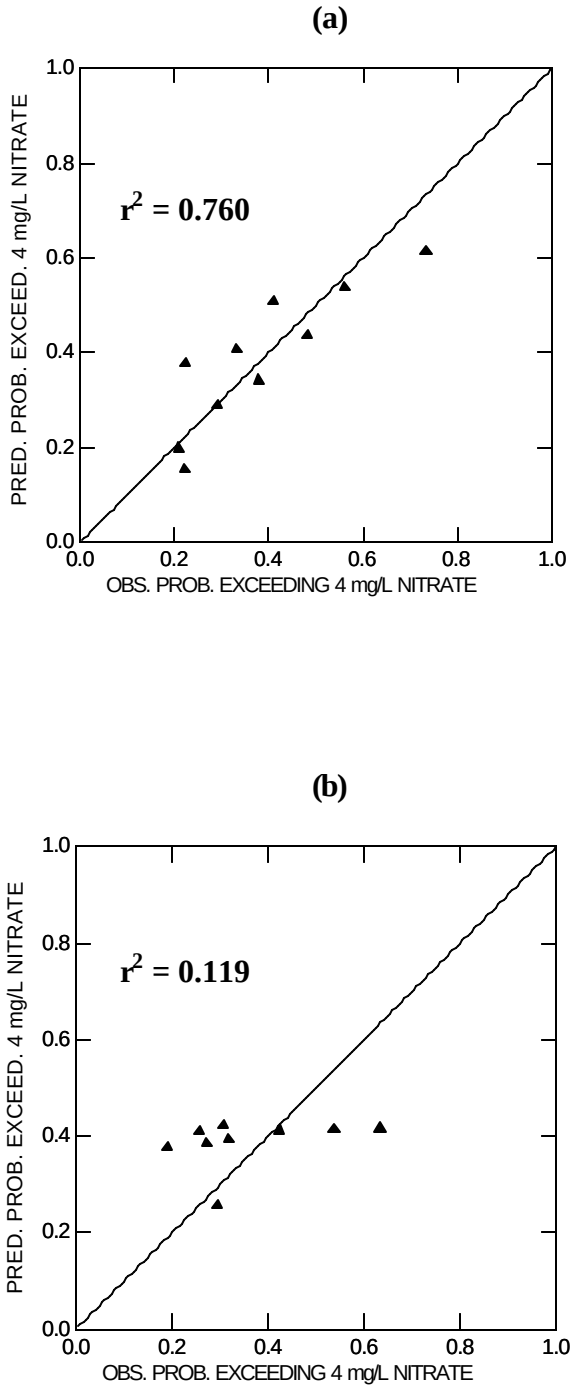


Figure 2. Linear regression fit of observed and predicted probabilities of nitrate exceeding 4 mg/L in shallow ground water, for univariate logistic-regression models representing (a) percent well-drained soils and (b) percent organic matter in land-use study areas.

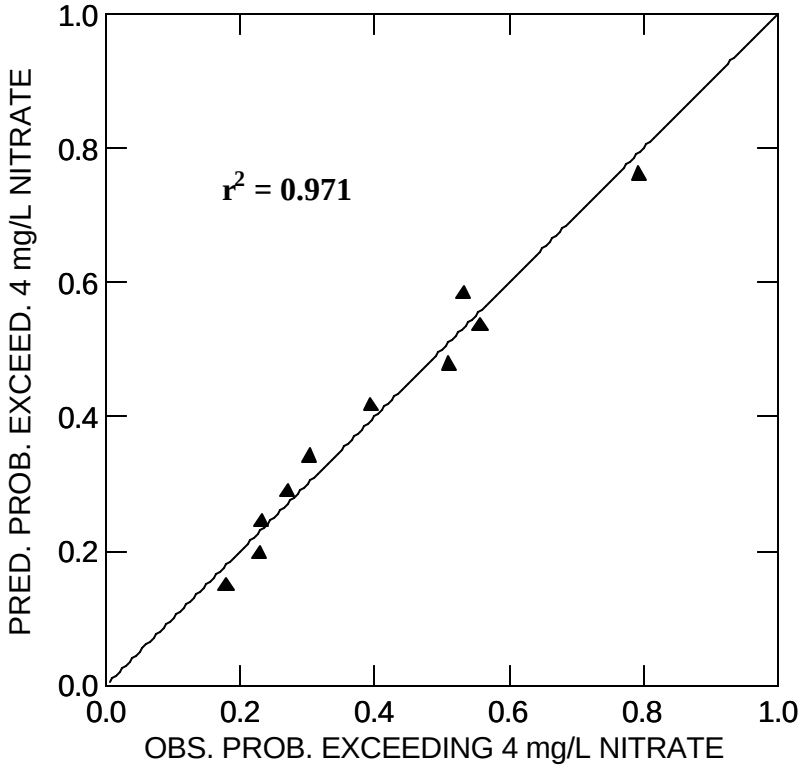


Figure 3. Linear regression fit of observed and predicted probabilities of nitrate exceeding 4 mg/L in shallow ground water, for final multivariate logistic-regression model MV6.