

A Statistical Analysis of Well Failures in Baltimore County

Xiaoyin Wang¹ and Kevin W. Koepenick²

¹Towson University, Towson and ²Baltimore County Department of Environmental Protection and Resource Management

Abstract: A statistical evaluation of the Baltimore County water well database is performed to gain insight on the sustainability of domestic supply wells in crystalline bedrock aquifers over the last 15 years. Variables potentially related to well yield that are considered included well construction, geology, well depth, and static water level. A variety of statistical methods are utilized to assess correlation and significance from a database of approximately 8,500 wells, and a logistic regression model is developed to predict the probability of well failure by geology type. Results of a two-way analysis of variance technique indicate that the average well depth and yield are statistically different among the established geology groups, and between failed and non-failed wells. The static water level is shown to be statistically different among the geology groups but not among failed and non-failed wells. A logistic regression model results that well yield is the most influential variable for predicting well failure. Static water level and well depth was not found to be significant in predicting well failure.

Key words: ANOVA, Box-Cox transformation, influence analysis, logistic regression, residual, prediction.

1. Introduction

The Baltimore County Master Plan 2010 (Baltimore County Council, 2000) incorporates the designation of two land management areas: the urban area and the rural area. The boundary separating these two land management areas is called the Urban-Rural Demarcation Line (URDL). The urban areas have public water and sewer infrastructure, and the rural areas rely on individual private wells and septic systems. Approximately 80,000 people live in the rural areas where the geology consists of a group of crystalline rock aquifers (metamorphic and igneous) that are commonly referred to as the Piedmont physiographic province. Ground water occurrence (yield) within the crystalline rocks is extremely variable, and there are noted formations where there is relatively poor well productivity (Nutter

and Otten, 1969). The Piedmont aquifers are also unconfined, and therefore, susceptible to contamination from land use practices. Given the nature of the geology, it is important that new development in these rural areas be carefully evaluated to ensure that domestic well water supplies are reasonably protected and sustainable.

The Baltimore County Department of Environmental Protection and Resource Management (DEPRM) is charged with the responsibility of ensuring “safe and adequate” water supplies for proposed development in Baltimore County utilizing wells for their domestic water needs. DEPRM considers the existing setback requirements, well construction regulations, and development regulations to be reasonably adequate to protect existing and proposed water supplies. However, there is continuous concern from residents as to whether or not proposed new development in the rural areas will have adverse impacts on existing land uses. Therefore, gaining a better understanding of well yield sustainability and whether or not well yield failure in the Piedmont can be practically predicted is of great interest to the regulatory, development, and residential communities in Baltimore County. The findings presented in this study may be used to address some of the many questions that have arisen over the years concerning whether existing regulations and practices are sufficient and effective in protecting and preserving domestic water well supplies.

In the sections to follow, we will describe the data set that will be used in this study, analyze some characteristics of the wells, and develop a logistics regression model to predict the well failure probability. We will discuss influence diagnostics to determine the model’s accuracy, and also assess the predict power of the estimated model. Finally, we will discuss the potential ramifications of how the data might be used to change and/or support existing regulations governing rural development.

2. Data Structure

DEPRM manages all the well records for drilling in Baltimore County, which includes information about well locations, well usage, well yield, static water level (distance from the land surface to the depth of water in the well), and total well depth. There are 28 different geologic formations in Baltimore County. However, for the purposes of this study, the Piedmont formations are categorized into eight geologic groups: Gneiss, Granite, Mafic, Marble, Loch Raven Schist, Prettyboy Schist, Other Schists, and Serpentine. Table 1 displays the classification of the geology groups and the corresponding total study area. DEPRM uses a Geographic Information System (GIS) to correlate a geology group with the location of each well that has a known address. Although database maintains records for approximately 21,000 domestic wells as of February 2005, only 8,483 could

be geographically located and matched to a geology group because the database does not have accurate address information for the others. This reflects the approximate number of domestic wells that have been completed since 1990, when address information was added to the well database. However, it does not reflect any known bias toward well geology type, yield, depth, SWL, or well failure.

Table 1: Grouping of geologic formations and well distributions

Group Title (Area)	Formations included in group	# of failed wells / # of wells
Gneiss (43,121 Acre)	Baltimore Gneiss, Franklinville Gneiss, Gunpowder Gneiss, Setters Gneiss, Perry Hall Gneiss, Sykesville Gneiss , Cold Spring Gneiss, Slaughterhouse Gneiss	170/1,685 (10.09%)
Granite (1,181 Acre)	Ellicott City Granite, Woodstock Granite	2/28 (7.14%)
Loch Raven Schist (63,994 Acre)	Loch Raven Schist	443/3,733 (11.63%)
Mafic (10,408 Acre)	Mt. Washington Amphibolite, Hollofield Layered Ultramafite, Bradshaw Layered Amphibolite, James Run-Druid Hill Amphibolite, Raspeburg Amphibolite	14/363 (3.86%)
Marble (23,151 Acre)	Cockeysville Marble, Hydes Marble	29/309 (9.39%)
Prettyboy Schist (63,353 Acre)	Prettyboy Schist	60/1,403 = 4.28%
Other Schists (34,753 Acre)	Pleasant Grove Schist, Sykesville Schist, Oella Formation, Piney Run Formation, Setters Schist	46/933 = 4.93%
Serpentine (3,236 Acre)	Serpentine Ultramafite	3/39 =7.69%

Among the 8,483 wells, there are total 767 (9%) failed wells. Table 1 displays the ratios of the number of failed wells to the number of wells in each geology group, with the resulting percentages being shown in parentheses. The Loch Raven Schist wells have the highest well failure rate of 11.6%, and the Mafic wells have the lowest well failures rate of 3.9%. It should also be noted that the relatively small number of wells in the Granite and Serpentine groups might lessen the significance of statistical inferences for these two geology groups.

3. Data Analysis

3.1 Correlation analysis

The first question we would like to answer is that if there is a statistically significant correlation exists between geology type and the well failure. A Chi-square test for independency has the results of $\chi^2 = 103.6924$ with $df = 7$ and

p -value < 0.001 . This implies that there is a statistically significant association between geology type and failed/non-failed wells.

3.2 Two-way ANOVA

In the database, there is also information about the characteristics of wells, such as well depth, static water level and well yield. Table 2 displays the summary statistics of these measures of failed and non-failed wells in each geology group. The average well depth, static water level and well yield are different among geology groups, and are different between failed and non-failed wells. Two-way analysis of variance technique is applied to study the difference of these measures among the geology groups, and between the failed and non-failed wells, where the main effects are “geo-group” for the 8 geology groups, and “response” with for failed (“1”) and non-failed (“0”) wells. Due to the missing values in the database, only 8482, 6996 and 8478 wells are used respectively in the analysis of well depth, static water level and well yield.

Table 2: Average (standard deviation) of well depth, static water level and yield

Geo-group		depth (ft)		swl (ft)		yield (1gallon/min)	
Gneiss	Non-Failed	255.07	(115.31)	36.54	(16.43)	9.67	(7.63)
	Failed	313.98	(148.50)	36.86	(15.08)	4.88	(4.61)
Granite	Non-Failed	217.96	(66.66)	29.81	(10.62)	10.43	(9.45)
	Failed	225.00	(106.07)	20.5	(0.7)	8.00	(9.90)
Loch Raven Schist	Non-Failed	316.65	(138.06)	39.62	(16.66)	6.86	(6.68)
	Failed	356.54	(152.77)	40.53	(15.24)	4.01	(4.17)
Mafics	Non-Failed	230.80	(105.54)	29.52	(13.78)	9.29	(7.79)
	Failed	263.00	(129.80)	38.67	(12.26)	6.20	(6.01)
Marble	Non-Failed	267.10	(162.71)	33.30	(18.93)	12.60	(14.33)
	Failed	274.28	(147.21)	30.95	(18.93)	9.04	(6.02)
Pretty boy Schist	Non-Failed	255.30	(102.01)	43.35	(15.80)	8.63	(6.58)
	Failed	292.58	(146.98)	45.39	(16.38)	5.74	(5.30)
Other Schists	Non-Failed	266.86	(117.52)	42.49	(16.24)	9.16	(7.52)
	Failed	292.58	(146.98)	43.04	(16.00)	5.32	(4.01)
Serpentine	Non-Failed	216.67	(108.74)	37.13	(19.61)	10.58	(9.18)
	Failed	246.67	(50.33)	30	(na)	5.03	(3.65)

Starting from a full model with both main and interaction effects, the results reveal that interaction effects are not statistically significant. Therefore, we applied an analysis with the two main effects only. However, the residual analysis shows a certain degree of disagreement with the equal error variance and the normality of error assumptions, see Figure 1(a) the residual plots in part, and (b) normal probability plots. To remedy these departures from the model assumptions, we need to apply transformations on the response variable. It is often

difficult to determine from residual plots which transformation is most appropriate. The Box-Cox procedure automatically identifies a transformation from a family of power functions, and is the most popular used procedure. The Box-Cox transformation with the optimal λ value provided by SAS is applied to the data. The residual plots after Box-Cox transformation in Figure 1(c) show that the error variance appear to be more stable, and the normal probability plots after Box-Cox transformation in Figure 1(d) fall roughly on straight lines. The test results for equal means of well depth, static water level and well yield among geology groups are illustrated in Table 3. It confirms that the average well depth and yield are not the same among the geology groups, nor between the failed and non-failed wells. The data set does not provide enough evidence to conclude that the static water level is different between failed and non-failed wells, but differences among the geology groups are statistically significant.

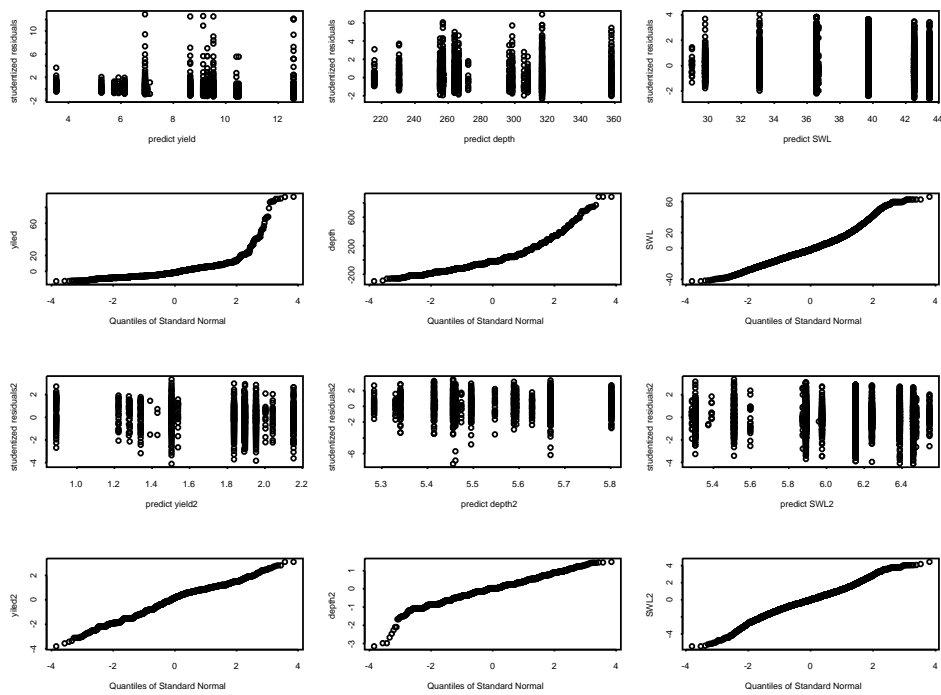


Figure 1: Normal Probability Plots and Studentized Residual Plots
 (a) Studentized Residual Plots Before Box-Cox Transformation
 (b) Normal Probability Plots Before Box-Cox Transformation
 (c) Studentized Residual Plots After Box-Cox Transformation
 (d) Normal Probability Plots After Box-Cox Transformation

Table 3: Two-way analysis of variance of well depth, static water level and yield

	depth		SWL		yield	
	<i>F</i>	<i>p</i> -value	<i>F</i>	<i>p</i> -value	<i>F</i>	<i>p</i> -value
Geo-group	86.07	(< .0001)	43.39	(< .0001)	70.87	(< .0001)
Failed/Non-Failed	64.33	(< .0001)	3.07	0.0798	296.72	(< .0001)

Tukey-Kramer technique is a well-known procedure used to perform pair-wise comparisons simultaneously (Neter, etc. 1996). Tukey's pair-wise multiple comparison shows that the wells in Loch Raven Schist are deeper and have less yield, the wells in the Mafic are shallower, and the wells in the Marble have higher yield than those in other geology groups. The average static water levels of wells are different between most of the geology groups. The majority of test results regarding the geology groups of Granite and Serpentine are not statistically significant due to small data records.

3.3 Logistic regression

Logistic regression models are the most commonly used probabilistic models for a binary (success-failure) response variable such as a "yes/no" question. It has wide applications in biomedical fields, genetics, reliability engineering experiments, social science research, business and environmental studies. A logistic regression model is developed using the well data from DEPRM for the purposes of estimating the well failure probability related to certain variables.

Table 4: Summary of stepwise selection

Step	Effect		DF	Number in	Score Chi-Square	<i>p</i> -value
	Entered	Removed				
1	Yield		1	1	166.7870	< 0.0001
2	Geo-Group		7	2	75.9649	< 0.0001
3	Yield*Geo-Group		7	3	17.1132	0.0167

For this study, we considered four main factors in the model; well depth, static water level, well yield, and geology group, as well as the 11 interaction effects among them. In order to find the most efficient model, a stepwise automatic search procedure with 0.05 level of significance for both entering and removing, is applied to identify the best subset of useful effects to be included in the final model. The summary of model selection results is displayed in Table 4. The outcome model includes two main effects, well yield and geology group. Therefore, the final logistic regression model can be written as

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \alpha_0 + \alpha_i + \beta_0(\text{yield}_{ij}) + \beta_i(\text{yield}_{ij}) + \epsilon_{ij},$$

where i represents the geo-groups with 1 for Gneiss, 2 for Granite, 3 for Mafic, 4 for Marble, 5 for Loch Raven Schist, 6 for Prettyboy Schist, 7 for all other schist, and 8 for Serpentine; j represents each well, and p_{ij} represents the probability of failure for the j th well in the i th geology group. Here α_0 is a baseline or average $\log(odds)$ for all the geo-groups when the yield equals to 0. It is not important that a well having yield equals to 0 be realistic; rather α_0 represents a reference point, and α_i is the deviations from α_0 due to the effect of geo-group i ; β_0 is a baseline or average decrease of $\log(odds)$ for every increase of 1 gallon/min in yield, and β_i is the deviation from β_0 due to the effect of geo-group i . The assumptions for the model are as follows:

- $\sum_{i=1}^8 \alpha_i = 0$;
- $\sum_{i=1}^8 \beta_i = 0$;
- the random error component of the model, ϵ'_{ij} 's, are independent and identical normal distributions with mean 0 and variance 1.

The statistical software SAS (Allison, 2001; Cody& Smith 2005) was used to perform the estimates of these parameters. With the estimated the parameters, we have the equations that can be used to predict the well failure probability, p , based on the initial yield and the geology group, see Table 5. A plot of each equation, shown in Figure 2, reveals that all of the geology groups have an exponential decrease in the probability of well failure with increasing yield. At low yields (1-3 gpm), in particular, the rate of predicted well failure ranges considerably by geology type. It is interesting to note that the Mafic and Prettyboy Schist wells show a significantly lower probability of well failure at the minimum allowable well yield even though the average yield for both of the geology types is lower than nearly all other geology types with the exception of Loch Raven Schist. The Marble and Granite geology groups show a markedly slower decline than other geology groups. In fact, at well yields above 6.33 gpm, the Marble becomes the geology group with the highest probability of well failure. The reason for this difference is not exactly clear, but in the case of the Marble, it may be due to geologic reasons. For instance, the presence of relatively large subsurface solution channels is known to exist in the Marble aquifers and is considered one of the primary reasons for the observed high well yields in this geology group. These solution channels may occasionally collapse or become filled with sediment, thereby reducing what was a high yielding well into a non-productive well. As mentioned earlier, the relatively small data set for the Granite could limit the models reliability for this geology type. The predicted probabilities and odds of

well failure at two specific amount of well yield, 1 gpm (minimum allowable well yield) and 10 gpm, are listed in Table 5.

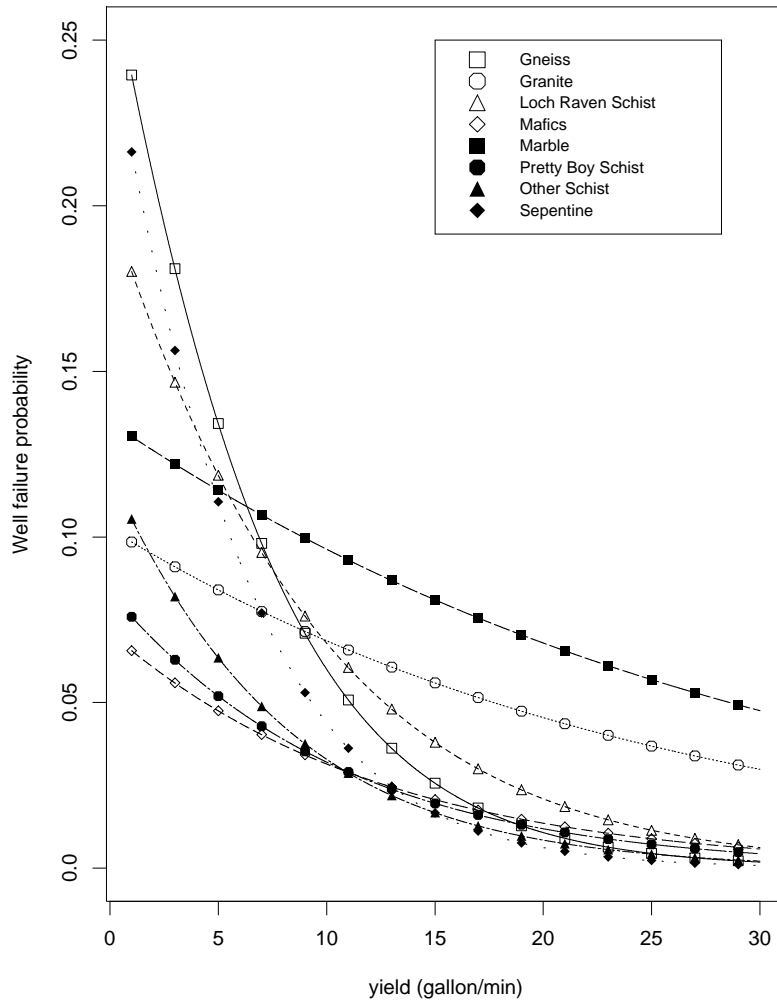


Figure 2: Predicted well failure probabilities

Table 5: Predicted probabilities of well failure probabilities and odds

Geo-group	Equation for failure probability	1 gallon/min		10 gallon/min	
		Probability	Odd	Probability	Odd
Gneiss	$\frac{\exp(0.9783+0.1772 \times \text{yield})}{1+\exp(0.9783+0.1772 \times \text{yield})}$	23.94%	31.48%	6.00%	6.38%
Serpentine	$\frac{\exp(1.0882+0.19935 \times \text{yield})}{1+\exp(1.0882+0.19935 \times \text{yield})}$	21.62%	27.58%	4.39%	4.59%
Loch Raven Schist	$\frac{\exp(1.3928+0.1226 \times \text{yield})}{1+\exp(1.3928+0.1226 \times \text{yield})}$	18.01%	21.97%	6.79%	7.28%
Marble	$\frac{\exp(1.8601+0.0379 \times \text{yield})}{1+\exp(1.8601+0.0379 \times \text{yield})}$	13.03%	14.98%	9.63%	10.66%
Other Schists	$\frac{\exp(2.0004+0.1381 \times \text{yield})}{1+\exp(2.0004+0.1381 \times \text{yield})}$	10.54%	11.78%	3.29%	3.40%
Granite	$\frac{\exp(2.1705+0.0437 \times \text{yield})}{1+\exp(2.1705+0.0437 \times \text{yield})}$	9.85%	10.93%	6.87%	7.38%
Pretty boy Schist	$\frac{\exp(2.3986+0.1010 \times \text{yield})}{1+\exp(2.3986+0.1010 \times \text{yield})}$	7.59%	8.21%	3.20%	3.31%
Mafic	$\frac{\exp(2.5695+0.0858 \times \text{yield})}{1+\exp(2.5695+0.0858 \times \text{yield})}$	6.57%	7.03%	3.15%	3.25%

4. Residual Analysis and Influence Diagnostics

It is always very important to examine the outliers and influential observations in the data to refine the model. The estimated model could be quite different if there is an outlier with a large influence. Plots of residuals against explanatory variables and the predicted probabilities are very useful tools to identify outliers. There are two sets of plots in Figure 3. The first set, part (a) consists of scatter plots of two types of residuals, which are the deviance residuals and the Pearson residuals, against well yield, and the predicted well failure probabilities. In each plot, the residuals are clustered into two groups. The upper group of residuals is from the non-failed wells, and the lower group is from failed wells. No obvious outlier is exhibited in the scatter plot of deviance residuals with well yield or the predicted well failure probabilities. However, the scatter plots of the Pearson residuals indicate that one observation with a high value of greater than 12 might be an outlier. In order to identify this potential outlier, scatter plots of the Pearson residuals against well yield of each geology groups were constructed, see part (b) of Figure 3. It can be seen that the potential outlier is referring to a well in Loch Raven Schist. However, it seems to follow the trend line of the other residuals in the upper group. As mentioned by Agresti, when explanatory variables are continuous, there are only one residual for each setting, and a signal residual is often uninformative.

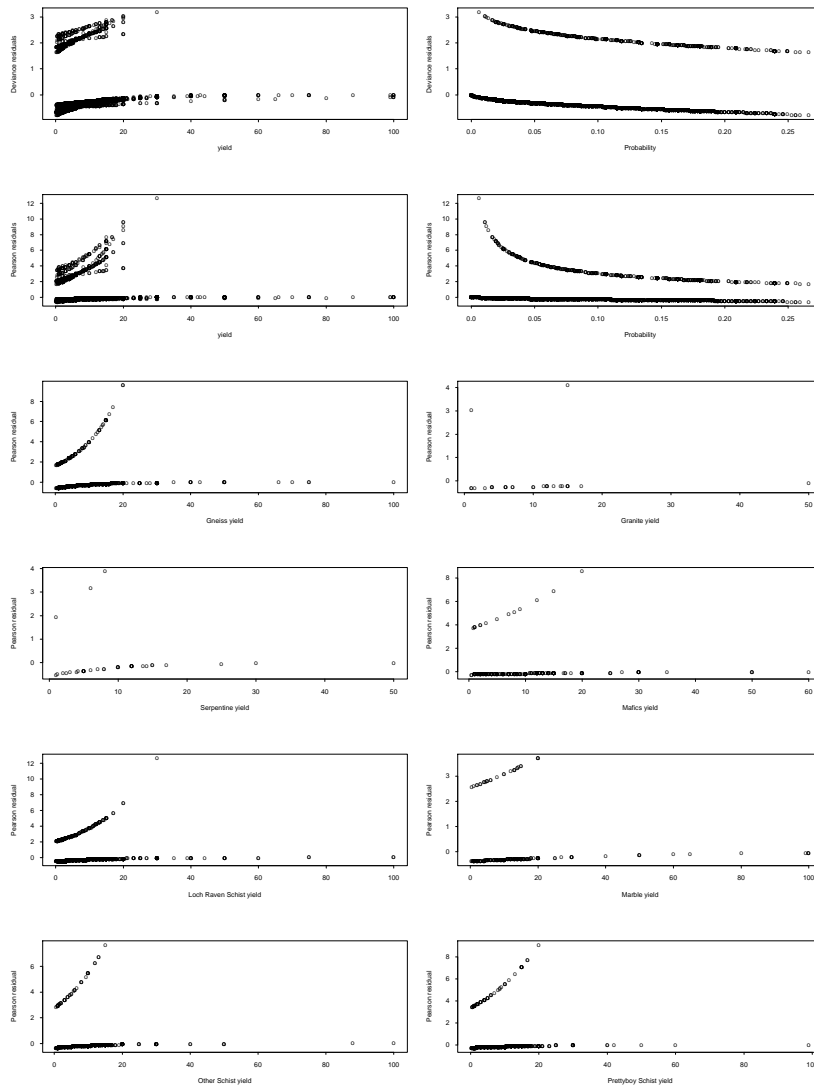


Figure 3: Residual plots: (a) Overall residual plots, (b) Pearson residual plots of each geology group

Other helpful tools used to assess the fitness of a model are diagnostics of an observation's influence on parameter estimates. The greater an observation's leverage, the greater its potential influence. The most commonly used tool to assess the influence of an observation is through the measure of the change in some statistics when the observation is removed from the data. Three standard statistics that serve this purpose are: the joint confidence interval for the parameters, denoted by c ; the chi-square goodness-of-fit statistic, denoted by χ^2 ; and

the deviance goodness-of-fit statistic, denoted by G^2 . The larger the change, the higher influence the observation has on the estimation of the parameter (Agresti 2002).

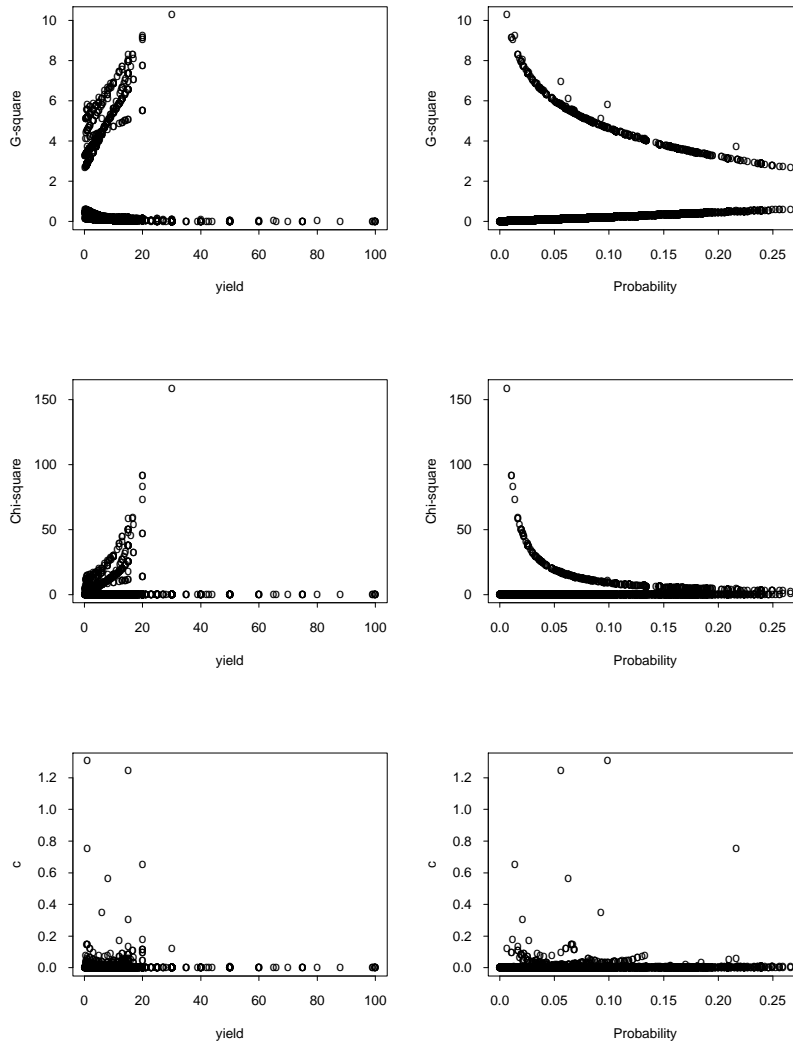


Figure 4: Influence diagnostic

Figure 4 illustrates the scatter plots of the changes of those measures when an observation is deleted against the explanatory variable, well yield, and the predicted well failure probabilities. Similar to Figure 3, there are two clusters in

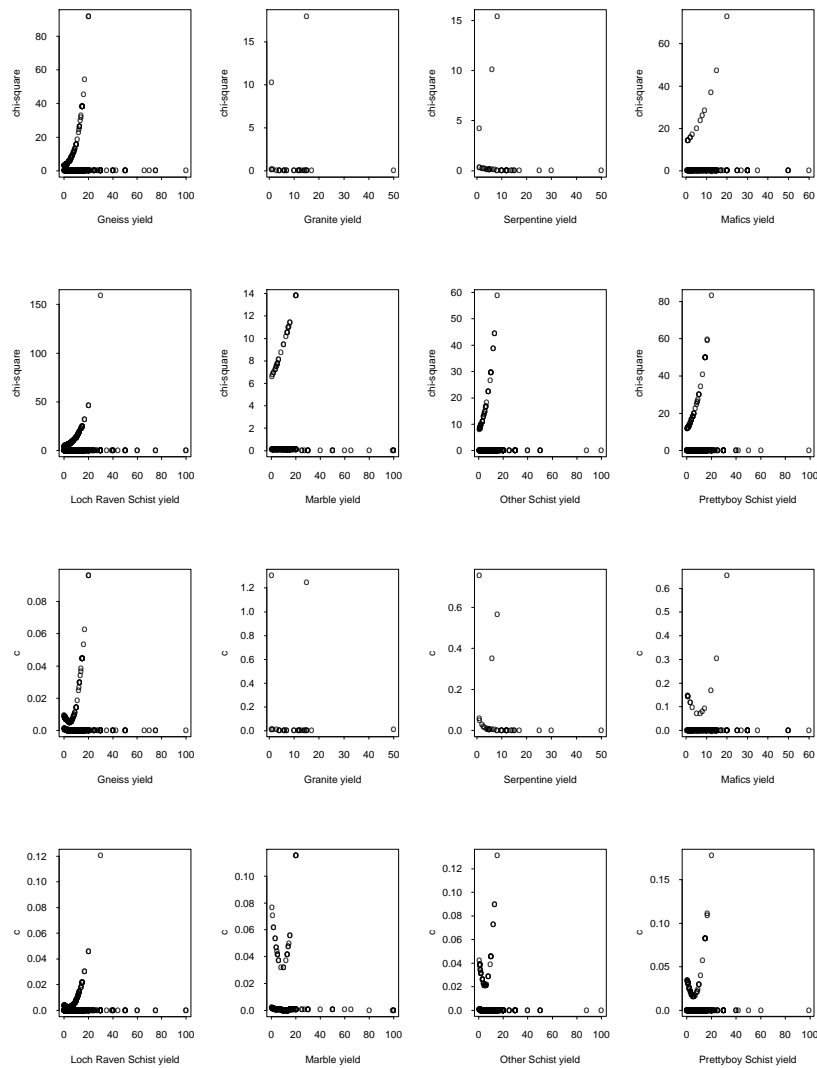


Figure 5: Influence Diagnostic of Each Geology Group: (a) Change on Chi-square Goodness-of-fit Statistic, (b) Change on Confidence Interval.

each plot. The measures from failed wells are the upper group; the non-failed wells are the lower group of each plot. The two plots in the top panel of Figure 4 illustrate the changes in G^2 when an observation is deleted against well yield and predicted well failure probabilities, respectively. The largest change in G^2 is more than 10. However, there is no clear evidence that this observation has unusually larger influence on G^2 than the others. The two plots in the middle panel of Figure 4 illustrate the changes in χ^2 when an observation is deleted. It

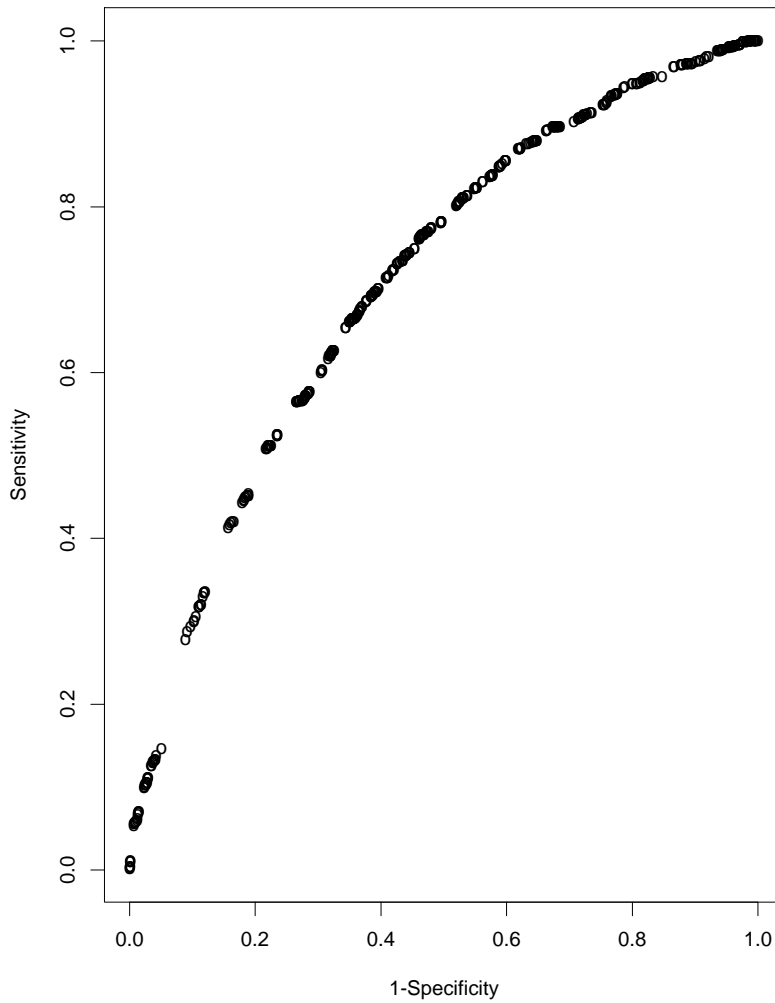


Figure 6: Receiver operating characteristic curve

seems that there is one observation, which has larger influence on the χ^2 than the others, and has the value greater than 150. The bottom panel of Figure 4 illustrates the change in c when an observation is deleted. There are several large values (> 0.4) in the plots.

In order to identify those potential high influence observations, scatter plots of the changes in χ^2 and c against well yield of each geology group are constructed as shown in Figure 5. The top section, part (a), of Figure 5 consists the scatter

plots of the changes in χ^2 against well yield of each geology group. It shows that the potential high influence observation is located in Loch Raven Schist. However, it seems to follow the trend of the line of failed wells. The bottom section, part (b) of Figure 5 presents scatter plots of the changes in c against well yield of each geology. These scatter plots show that only one observation from Mafic with the change in c greater than 0.4 may have high influence on the model.

A logistic regression model was fitted without these potential outliers and high influence observations. The resulting estimated model does not change significantly from the former estimated model. Therefore, we used the former estimation as our final estimated model, and to predict the probability of well failure.

5. Power of the Prediction

The power of the prediction of a logistic model can be summarized by two measures: sensitivity and specificity. For some given cutoff value π_0 , if the predicted probability is greater than π_0 , then the well is predicted to fail, otherwise the well is predicted to not fail. The percentage of correctly predicting well failure is called sensitivity, and the percentage of correctly predicting non-failed well is called specificity. For multiple cutoffs π_0 , a receiver operating characteristic (ROC) curve is a commonly used tool to assess the power of prediction of a logistic model. It is a plot of sensitivity against (1-specificity) for all possible cutoffs π_0 . This curve usually has a concave shape. The larger the area under the curve, the better the prediction. Figure 6 is the ROC curve of our estimated logistic model of predicting well failures. The area under the curve is identical to the value of another measure of predict power, the concordance index, which measures the probability that the predictions and the outcomes are concordant. For our study, the concordance index is 0.708, meaning that overall, we will have a 71% chance of correctly predicting the probability of well failure.

6. Discussion

In Baltimore County, DEPRM reviews all proposed domestic well locations to ensure adherence to minimum setback distances from domestic wells to other wells, to potential sources of contamination (e.g., septic systems, underground petroleum storage tanks, etc.), to property lines, roads and to buildings. Setback distances and well construction standards were established over 25 year ago to minimize potential influences between wells and to protect well water quality. DEPRM's experience has been that these regulations have generally been effective. However, there are no allowances provided in the regulation for the

potential need to drill replacement water supplies at some point in the future. Unlike the requirements for utilizing an on-site sewage disposal system (OSDS) where a “septic reserve area” must be established prior to issuance of a building permit, there is no requirement in for a “well reserve area.” There have been many instances over the years where replacement water supplies cannot meet the minimum setback requirements, particularly for undersized lots of record, and subdivisions where lots are less than 2 acres in size. Property owners must seek variances to existing setbacks and in some cases have had to acquire easements on neighboring properties to attain adequate well yield and/or water quality. The problem of finding a suitable replacement well location becomes even more problematic when multiple drilling attempts are required to attain a suitable yield. Fortunately, this scenario appears to occur on a relatively small number of cases. Since 1990, when the number of unsuccessful drill attempts (dry holes) per lot were first tracked, over 95% of drilling attempts for replacement wells were successful on the first attempt; 2% had more than 1 drilling attempt; and less than 0.5% had more than 5 drilling attempts.

The statistical analysis provided in this study may be used to argue for regulatory changes that would require “well reserve areas” on all new lots. This would likely increase overall lot size and, therefore, decrease building density. Alternately, one may argue the raising the minimum well yield would provide better protection for property owners. However, this may create a large number of unbuildable areas, and indirectly affect the resale value of existing homes with well yields below the minimum.

Of course, the data presented does not take into account other factors that may impact the well failure rate. In 2002, Maryland experienced arguably one of the worst droughts on record. During that year, there was a 5-fold increase in the number of replacement wells drilled over the previous 10-year average. While the drought caused grave concern for rural residents, the roughly 350 replacement wells drilled in 2002 represent less than 1% of the total number of wells in Baltimore County, and only about 4% of the well population used in this study. The relatively low percentage of wells impacted during the drought seems to indicate that well sustainability in the Piedmont may not be as sensitive to changes in precipitation as generally assumed. The spatial distribution of replacement wells during the drought year indicates that highest percentage of well failure occurred in the Mafic at 2.4%, compared with all other geology groups that had failure rates between 0.9% and 1.5%. This seems contrary to the model presented in this study which indicates that the Mafic wells have the lowest overall failure rate. However, as explained below, the overall well population used to calculate these statistics includes many wells that may be more susceptible to well failure.

In 1980, the state of Maryland adopted regulations requiring more stringent well construction and yield testing practices. In addition, Baltimore County enacted legislation in 1978 requiring that upon transfer of real property, domestic wells must be able to produce a sustained minimum yield of 1 gallon/minute. It is estimated that almost half of the wells currently in use in Baltimore County were drilled prior to 1980 for which there may be little or no well construction information. Since these older wells are generally shallower, and considered more susceptible to drought and yield problems, it is not surprising that DEPRM records show that nearly 70% of the wells replaced due to yield problems from 1989 - 2005 were wells drilled prior to 1980. Clearly, the older wells are slowly being replaced as properties are being transferred and/or residents experience yield problems. The findings in this study should not be strongly influenced strongly by older wells since only wells with complete well information were used (i.e. wells drilled after 1980).

Social trends may also affect the number of well replacements as water consumption in the U.S. has risen over the last few decades. According to the U.S. Environmental Protection Agency, the average household now uses approximately 181 gallons/day, compared with only 164 gallons/day in 1970. The more prevalent use of private swimming pools, landscaping and other outdoor watering needs may add a considerable strain to a domestic well water supply with a low yield.

7. Conclusions

The main goal of this study was to assess whether the well data collected could be used to predict the probability of well failure in the Piedmont. Analysis of the observed data clearly indicates that well failure is correlated strongly with well yield and to a lesser degree with geology type. The relatively high percentage of failure for low yielding wells in certain geology types may be good reason to consider a requirement for well reserve areas during the building/subdivision approval process. This study does not address the possibility that eventually all wells may fail. Certainly, it would require a much longer period of data collection (perhaps 20-40 years) to determine for average well longevity for new and replacement wells.

Acknowledgments

We appreciate the Towson University Applied Mathematics Laboratory directed by Dr. Michael O'Leary who brought a team of students and faculty to work on the earlier phases of this project. The members of the first year team directed by Dr. Andrew Angel are Jennifer Zeigenfuse, Adam Durana, Kristin Seifarth, Alozie Nwoko and Renee Simeon. The members of the second

year team directed by Dr. Xiaoyin Wang are Pete Surgent Christopher DeZago, Adam Warfield, Allyson Rothman, Michael Stephen, and Christopher DeZago.

References

- Agresti, A. (2002). *Categorical data analysis*, 2nd ed. Wiley-Interscience.
- Allison, P. D. (2001). *Logistic Regression Using the SAS System: Theory and Application*. John Wiley & son, Inc. and SAS Institute Inc.
- Cody, R. P. and Smith, J. K. (2005). *Applied Statistics and the SAS Programming Language*, 5th ed. Prentice Hall.
- Neter, J, Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. (1996). *Applied Linear Statistical Models*, 4th ed.. McGraw-Hill/Irwin.
- Nutter, L. T. and Otton, E. G. (1969). *Ground water occurrence in the Maryland Piedmont*, Report of investigations No. 10. Maryland Geological Survey.

Received March 2, 2007; accepted October 4, 2007.

Xiaoyin Wang
Mathematics Department
Towson University
Towson, MD 21252, USA
xwang@towson.edu

Kevin W. Koepenick
Baltimore County Department of Environmental Protection and Resource Management
Towson, MD 21204, USA
kkoepenick@baltimorecountymd.gov