guardians of drinking water quality
DRINKING WATER INSPECTORATE

DRINKING WATER INSPECTORATE (DWI)

CHLORINE TASTE & ODOUR CONTACTS AND CHLORINE LEVELS OR VARIABILITY

FINAL REPORT

DWI Ref: 70/2/238
WRc Ref: DWI 8144
February 2010

CHLORINE TASTE & ODOUR CONTACTS AND CHLORINE LEVELS OR VARIABILITY

FINAL REPORT

Report No.: DWI 8144

Date: February 2010

Authors: David Garrow and Joanne Hulance

Contract Manager: Joanne Hulance

Contract No.: 15216-0

RESTRICTION: This report has the following limited distribution:

Drinking Water Inspectorate

Any enquiries relating to this report should be referred to the authors at the following address:

WRc Swindon,
Frankland Road, Blagrove,
Swindon, Wiltshire, SN5 8YF.

Telephone: + 44 (0) 1793 865000
Fax: + 44 (0) 1793 865001
Website: www.wrcplc.co.uk

This document has been produced by WRc plc.

The research was funded by the Department for Environment, Food & Rural Affairs (Defra) under project DWI 70/2/238. The views expressed here are those of the authors and not necessarily those of the Department or DWI.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

I REASONS

Chlorine is used throughout the world as a disinfectant for water supply, and has probably made a greater contribution to the prevention of waterborne disease than any other form of water treatment. However, one disadvantage of the use of chlorine is that it can result in customer complaints of unpleasant taste and odour, either from chlorine itself or from reaction of chlorine with other materials in, or in contact with, the water.

The categorisation of customer contact data was modified in 2006, providing the Drinking Water Inspectorate (DWI) with more detailed information regarding the nature of taste & odour (T&O) contacts, including those relating to chlorine. This provides the DWI with the opportunity to gain a better understanding of any relationships between chlorine concentrations within the distribution network and customer contacts relating to chlorine T&O.

II OBJECTIVES

The aim of this project was to use data held by the DWI over the period 2006 to 2008, to ascertain whether there is any relationship between the chlorine concentrations, or variability in concentration, and chlorine T&O contacts. This data consists of the regulatory sampling results and customer contacts that the companies in England and Wales are required to report to the DWI.

The project had the following specific objectives:

- To extract information from DWI data systems and review the information to identify the most appropriate analysis and statistics to meet the overall objectives of the work;
- To compare the data at zone level on the frequency of customer contacts in relation to chlorine taste and odour (T&O) with the chlorine residual data;
- To compare the data on the frequency of contacts relating to chlorine T&O with the chlorine residual variability measures, and;
- To assess the statistical significance of any correlations identified in the analyses.

III APPROACH

A multiple linear regression analysis technique was used to assess the strength of any relationships between the frequency of customer contacts relating to chlorine T&O and any explanatory factors, such as residual free or residual total chlorine (RDF or RDT respectively). Three measures of chlorine were selected to represent specific characteristics of chlorine concentration within each zone; mean concentration (central tendency), standard deviation (variability) and maximum-minus-mean (sudden peaks in concentration).

Other water quality parameters were included in the regression analysis to represent other factors that may also influence the frequency of chlorine T&O contacts. These parameters were colour to represent chlorine demand, turbidity to represent issues with distribution mains and conductivity to represent the varying hardness of the supply water.

IV CONCLUSIONS

- The statistical analysis of the data held by the DWI identified a weak relationship between mean residual free chlorine concentrations and the frequency of chlorine T&O contacts, at a national level (England and Wales). However, the inclusion of either colour or conductivity suggested that there is a stronger relationship between colour or conductivity

and chlorine T&O contacts than there is between the mean residual free chlorine and chlorine T&O contacts, reducing confidence in the reliability of this relationship.

- No other statistically significant associations were found between other measures of chlorine residual or variability and T&O contacts at the national level, across all three years analysed.
- There were marked differences between water companies which may mask any trends that exist between chlorine concentrations and the frequency of chlorine T&O contacts.
- The analysis of data from three carefully selected water companies suggests that relationships between chlorine concentration and the frequency of chlorine T&O contacts are present for one of these companies. However, these relationships are weak.
- For this one company, the term used to indicate sudden peaks in chlorine concentrations, produced consistently significant models across all three years (2006 to 2008). The relationships observed indicated that the frequency of chlorine T&O contacts increased as the magnitude of the peaks in chlorine concentrations increased. This relationship, although statistically significant, is not strong.
- Including both the mean conductivity and mean colour values, separately within the statistical models, have strengthened the identified relationship between chlorine concentration and the frequency of chlorine T&O contacts.
- The comparison of chloraminated and chlorinated zones in one other company showed a distinct difference in the frequency of chlorine T&O contacts, with chlorinated zones exhibiting a much higher frequency of chlorine T&O contacts than chloraminated zones.
- The relationship observed at the national level (mean residual free chlorine and chlorine T&O contacts) was not found to be statistically significant for any of the companies that were studied in detail, further reducing the confidence in the robustness of the relationship.
- The detailed assessment of a small sample of zones suggests that factors such as income, employment, health and education did not play an important role in determining the frequency of chlorine T&O contacts within these zones.

V RECOMMENDATIONS

It is recommended that further work is undertaken to understand why robust relationships between chlorine concentration or variability and the frequency of chlorine T&O contacts could not be identified at a national level, as follows:

- The weak relationship identified at national-level was not identifiable at individual company level. Further analysis of individual companies might identify the reason for this. This would also provide the opportunity to determine whether relationships do exist, beyond the one company identified in this study.
- Company specific factors, such as source water, operational practices, etc., should be examined further to determine the influence these factors may have on the T&O contacts. In particular, the study into socio-economic factors should be extended.
- There are a number of recommendations relating to the type and format of the data reported to the DWI which would facilitate studies of this nature in the future, as follows:
- A zone reference field should be added to the customer contact dataset in order to facilitate the links between the contacts data, sample results and site data.
- Consistent zone referencing should be used between these three datasets.
- Customer contact data should be reported by the date it was received by the water company. This would allow time series analysis to be conducted.

1. INTRODUCTION

1.1 Overview

Chlorine is used throughout the world as a disinfectant for water supply, and has probably made a greater contribution to the prevention of waterborne disease than any other form of water treatment. However, one disadvantage of the use of chlorine is that it can result in customer complaints of unpleasant taste and odour, either from chlorine itself or from reaction of chlorine with other materials in, or in contact with, the water. The complaints may be associated with high chlorine concentrations, although it is often thought that variability in concentration rather than the concentration itself that causes complaints. Operational conditions for water treatment works or the distribution system can therefore be a significant factor in generating taste and odour complaints from consumers.

The Drinking Water Inspectorate (DWI) regulates public water supplies in England and Wales, and is responsible for assessing the quality of drinking water. This includes the collection of water quality test data provided by individual water companies. Amongst the tests that the water companies regularly perform are measurements of the chlorine concentrations in the water.

Since 2005 DWI has collected data on numbers of drinking water quality contacts received by water companies from their consumers. In 2006, the data requirements were refined such that the number of contacts were broken down in more detail, to reflect the broad range of consumer contacts received and recorded across the industry. This included the categorisation of taste and odour (T&O) contacts to distinguish between chlorine taste and odour and other forms of T&O contacts, such as earthy or musty.

DWI commissioned this study to compare the three years of available customer chlorine taste and odour contact data (2006 to 2008) with the data that the Inspectorate already holds on concentrations of chlorine residuals in the distribution system. The purpose of this study was to provide a better understanding of the relationships at the national level between chlorine concentration, or variability in concentration, and customer contacts relating to T&O, using data taken from DWI data systems.

1.2 Project scope

The aim of this project was to use data held by the DWI over the period 2006 to 2008, to ascertain whether there is any relationship between the chlorine concentrations, or variability in concentration, and chlorine taste and odour contacts. This data consists of the regulatory sampling results and customer contacts that the water companies in England and Wales are required to report to the DWI.

The project had the following specific objectives:

- To extract information for 2006, 2007 and 2008 from DWI data systems on the contact data returns and chlorine sample results in the distribution system;
- To review the information to identify the most appropriate analysis and statistics to meet the overall objectives of the work;

- To compare the data at zone level on the frequency of customer contacts in relation to chlorine taste and odour (T&O) with the chlorine residual data, taking into account factors such as zone size and whether free or total chlorine is reported;

- To compare the data on the frequency of contacts relating to chlorine T&O with the chlorine residual variability measures, taking into account factors such as zone size and whether free or total chlorine is reported; and

- To assess the statistical significance of any correlations identified in the analyses.

1.3 Resume of contents

This report has the following content:

Section 2 describes the data collected and the processing performed in order to put these data into the correct format for the statistical analyses.

Section 3 describes the methodology utilised for the statistical analyses.

Section 4 documents the results of the analyses and the main findings of the project.

Section 5 describes a brief socio-economic analysis of a small number of zones that did not fit the derived models well.

Section 6 details the conclusions derived from this project and Section 7 makes some recommendations for future work.

2. DATA COLLECTION AND PROCESSING

2.1 Data collection

The DWI provided data to WRc from 28 water companies, covering a period of three years (2006, 2007 and 2008). These data consist of the information that the water companies in England and Wales are required to report to the DWI.

The water companies have been anonymised in this report by referring to companies using the code A, B to AA, AB. Each company reports data for specific zones, identified by a unique zone reference for the company. Data were provided on each zone for the following three areas:

- Customer contacts;
- Zone details; and
- Random water quality sampling results.

In total, data were received for about 1,670 zones in each year. The key zone details required for the statistical analyses were:

- Zone reference and site name, to link the different datasets;
- Population, to assess the frequency of contacts per thousand population; and
- Chloramination flag, to determine whether the zone is chloraminated.

Water quality sample data were received for the parameters shown in Table 2.1.

Table 2.1 Water quality parameters in dataset

Parameter name	Parameter code
Colour	A001
Turbidity	A002
Odour	A003
Taste	A004
Hydrogen ion (pH)	A006
Hydrogen ion (pH)	A006A[1]
Hydrogen ion (pH)	A006B[2]
Nitrite	A013A
Total Organic Carbon	A017
Iron	A022
Residual Disinfectant Free	C009
Residual Disinfectant Total	C010
Residual Disinfectant Combined	Combined
Conductivity	D001

[1] A006A was only used to report the hydrogen ion value where it failed the indicator standard. From 01/01/08 onwards, this parameter was no longer reported.

[2] A006B is the parameter code used for A006 before A006A was removed from the reporting for samples taken on or following 01/01/08. Prior to this date, any sample values that failed the indicator standard were duplicated and reported as A006A.

The measures of chlorine concentrations in the dataset are the Residual Disinfectant Free (RDF), Residual Disinfectant Total (RDT) and Residual Disinfectant Combined (RDC, equal to RDT minus RDF).

Customer contacts data were only available as a total number of contacts per zone, per category for each year. It is possible that seasonal differences and trends exist within the customer contact data. However, as the date of contact is not provided to the DWI by the individual water company, it is not possible to perform a time-series analysis. The date of contact should be provided in future to facilitate further work in this area.

2.2 Data processing

The data were collected in a single MS Access database to allow the data to be linked and processed. The data were then linked together to create a single summary table for each year that contained the zone details, contact details and summaries of the water quality parameter results for each zone. These summaries contained only details and test results for zones and not for reservoirs, treatment works, etc. Therefore, only those entries with a zone reference (containing 'Z') were included in the summary tables to ensure that all water quality data were drawn from the distribution samples only.

Summary statistics were calculated for each water quality parameter for each zone in a single year. The summary statistics calculated and recorded in the summary tables were:

- Number of test results for the parameter;
- Mean of test results;
- Standard deviation of test results;
- Minimum value measured;
- Maximum value measured;
- 25-percentile of test results; and
- 75-percentile of test results.

Some sample results have been qualified as "less than" (<) a specific value, when reported to the DWI by the water companies. In these cases the value of the sample result was assumed to be the same as the specified value.

The other key fields in the summary tables were:

- Water Company;
- Site Name;
- Zone Reference;
- Population;

- Chloramination Flag;
- Number of Chlorine T&O Contacts; and
- Chlorine T&O Contacts per Thousand Population.

Data in the customer contacts dataset had no zone reference field and the site name field in this dataset was not always consistent with the site names recorded in the zone details and water quality test results datasets. In addition, different water companies use different formats for the site names reported in this field. This meant that the three datasets (customer contacts, zone details and water quality test results) could not be easily linked using a single reference field. Instead, the contacts dataset had to be matched with the correct zones from the other two datasets using a combination of site names, zone references and truncated zone references. This was a lengthy process and it is recommended that the customer contacts dataset should include a zone reference field and water companies should employ a consistent system of referencing between all the datasets. This will allow for easy linking between datasets.

The number of contacts per zone can vary from year to year for a company, as shown in Figure 2.1 and Figure 2.2.

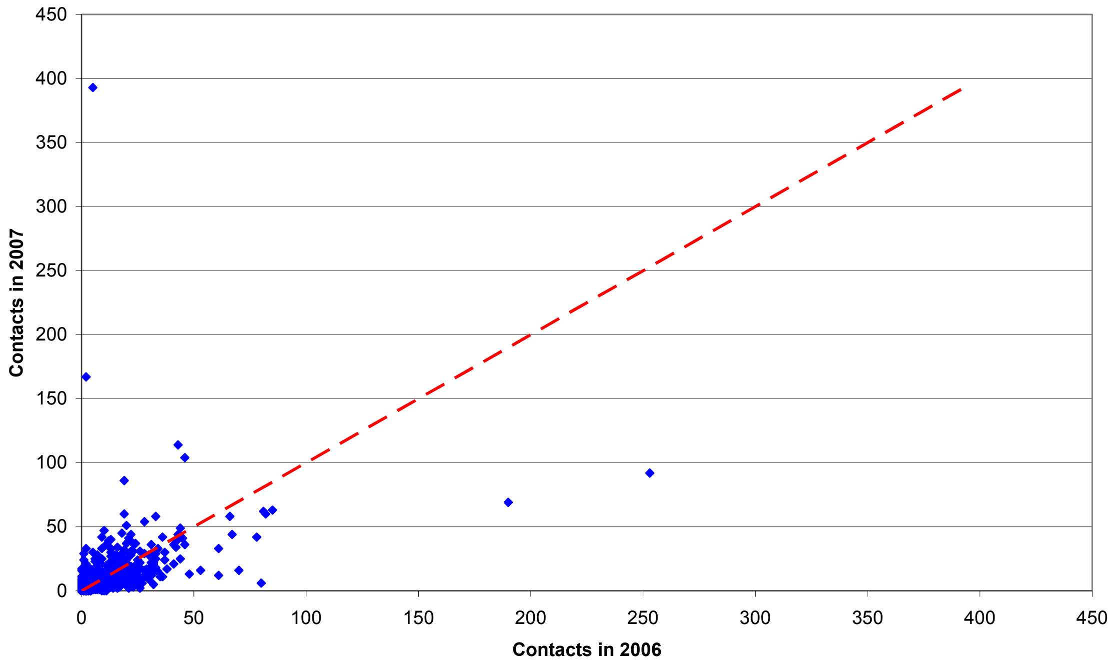

Figure 2.1 Number of chlorine T&O contacts per zone for 2006 versus number of contacts per zone for 2007

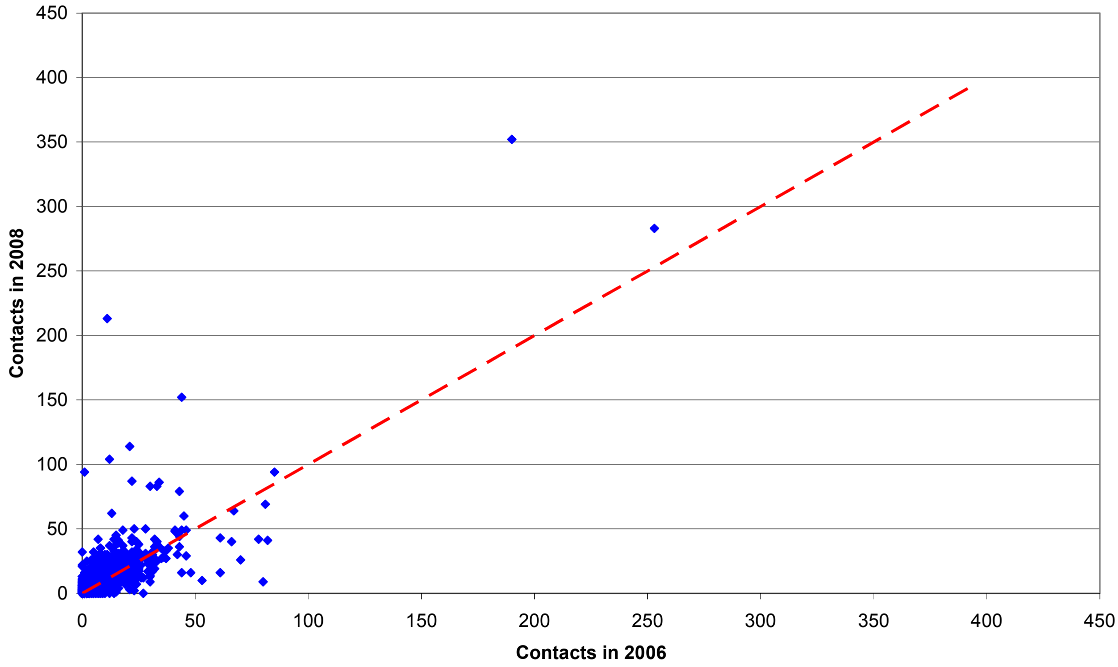

Figure 2.2 Number of chlorine T&O contacts per zone for 2006 versus number of contacts per zone for 2008

These figures show comparisons between the numbers of contacts in the three years for which data was available. It can be seen that, although there can be a large variation in contact frequency between years for some zones, most zones do not vary greatly. Because of the variation between years, each year was modelled separately; however, it is expected that the coefficients of the parameters will be similar between years since the variation in contact frequency is not large.

3. METHODOLOGY

A multiple linear regression analysis technique was used to assess the strength of any relationships between the number of customer contacts relating to chlorine T&O and any explanatory factors, such as RDF or RDT. In order to ensure that the annual statistical summary values calculated for each water quality parameter were generally representative of the levels of the parameter throughout the year, a minimum of 12 samples, for the appropriate parameter, were required to have been undertaken in the year for the zone to be included in the analyses.

Any zones with less than 12 samples in the year were excluded from the analyses. The number of zones excluded depended on the analysis undertaken and the combination of parameters used. At the national level, approximately 11% of the chlorinated zones did not have sufficient RDT sample data. This gives an indication of the proportion of zones that were excluded. The actual numbers of zones included within each analysis are presented in Tables A.3, B.3 and C.3 of the appendices.

Multiple linear regressions produce mathematical models that have the form:

$$y = a.x_1 + b.x_2 + ... + \varepsilon$$

Where:

y is the dependant variable (e.g. number of contacts per 1000 population)

x_i are the independent variables influencing y (e.g. RDT, Colour)

a, b are the coefficients fitted by the regression analysis

ε is the remaining variation that is not explained by the model

RDF and RDT were the parameters selected to represent chlorine concentrations. A large number of possible summary statistics that could represent RDF and RDT were considered, including those identified in Section 2.2. Following initial analyses using these statistics, it was decided to focus on a small number of statistics that would represent the different behaviours of chlorine levels that were considered likely to impact upon contact numbers.

Three different statistics for RDF and RDT were used in the analyses. These are shown in Table 3.1. It must be noted that RDF and RDT levels will vary across a zone. However, only random regulatory reported sample data were included within the analysis, and so should reflect the levels across the zone.

Table 3.1 Statistical measures used to represent chlorine concentrations in regression analyses

Statistical measure	Notes
Mean	Central tendency of chlorine concentrations throughout the year and the zone
Standard deviation	Variability of chlorine concentrations throughout the year and the zone
Maximum minus mean	Represents sudden peaks in chlorine concentrations from normal levels throughout the year and the zone

Figure 3.1 shows how these measures relate to the 233 individual sample results for RDT in a single zone during 2008.

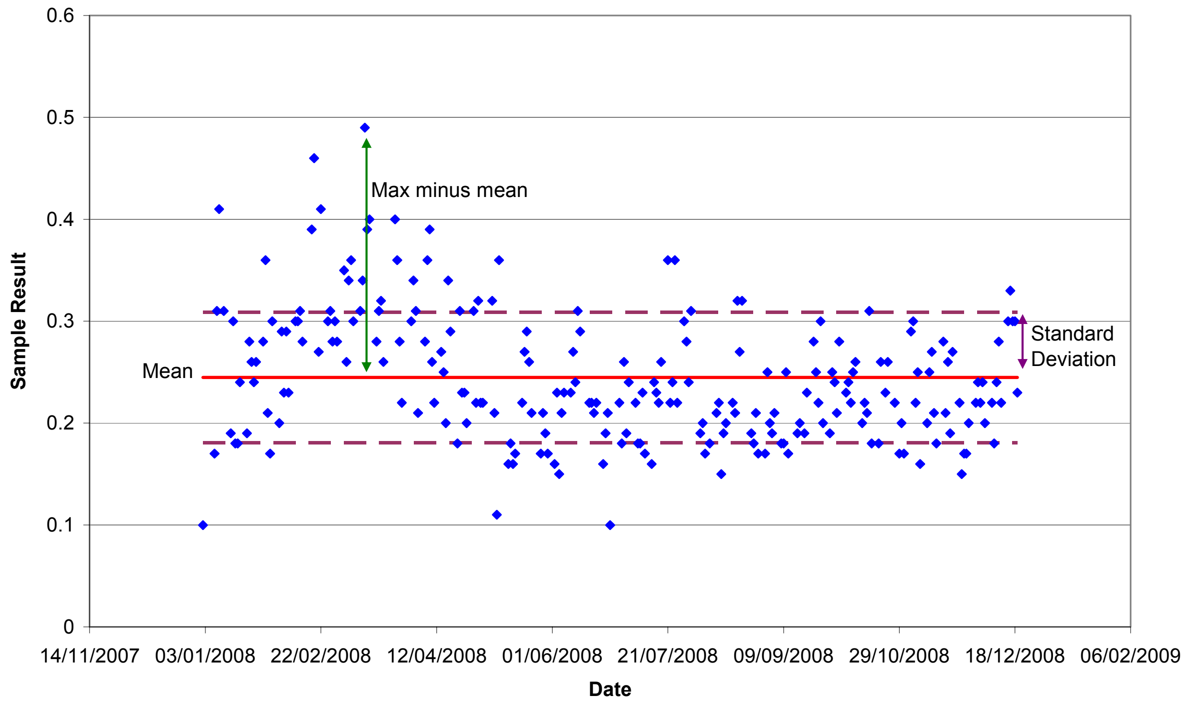

Figure 3.1 Chlorine level measures compared to individual RDT test results for a single zone in 2008

Not all the companies within the dataset report both RDT and RDF levels. There are nine companies (out of 28) which do not report RDF levels in any of the three years for which data were available. Another two companies only report RDF levels in one or two of the three years. There are only two companies that do not report RDT levels in any of the three years. Another two companies only have data on RDT levels in one or two of the three years.

The main aim of this study was to assess the statistical significance of any correlations between chlorine T&O contacts and chlorine concentration or variability. However, it is possible that a number of other factors may also influence customer contacts of this nature.

The possible water quality parameter and the explanatory factors, or water characteristics they represent, are shown in Table 3.2.

Table 3.2 Additional parameters considered for inclusion in regression analysis

Parameters	Characteristic represented
Colour, Total Organic Carbon	Chlorine demand
Turbidity, Iron	Problems with water mains (e.g. bursts, repairs, cleaning)
Conductivity	Hardness/alkalinity
Nitrite	Chloramination

The large number of water quality parameters suggested in this table precluded the inclusion of all the parameters in any modelling as a practical exercise. In addition, it was felt that some parameters might be representing similar characteristics of the water and only one of these parameters need be included. Therefore, careful consideration was given to determine the most appropriate and practical parameter to include in the regression analyses, in addition to the measures of chlorine.

The parameters selected were colour, turbidity and conductivity. Nitrite was not included in the model because there were only about a third of the zones in each year with enough water quality sample results (12 or more) to include this parameter in the models. The mean value of all the sample results in a zone in a single year was used to represent these three parameters in the analyses.

The data cover 28 water companies, each with different policies and practices, and there is a wide spread of zone sizes in the dataset, ranging from a population of 0 or 1 to over 90,000. Thus, a small number of contacts in a zone with a small population could have an excessive weighting in the analysis if this is not accounted for. For example, a single contact in a zone with a population of 250 is equivalent to four contacts per thousand population. This is a high value that might influence the modelling inappropriately.

In order to give appropriate weighting to zones of all sizes, the following approach was taken:

$$\sqrt{n} = A_0 \times \sqrt{p} + A_1 \times \{\sqrt{p} \times x_1\} + A_2 \times \{\sqrt{p} \times x_2\} + ... + \varepsilon$$

Where

n = number of chlorine T&O contacts in a year

p = zone population

x_i = factor i, e.g. mean free chlorine, standard deviation of free chlorine

A_i = coefficients fitted by the regression

ε = unexplained variation

The form of the model can then be transformed into a model for the number of contacts per population:

$$n / p = (A_0 + \sum (A_i \times x_i))^2$$

The basic form of the model involved only a single independent variable - a measure of RDT or RDF – in the regression analyses. Following this, an additional factor (colour, conductivity or turbidity) was added to the model to determine if this altered any relationship identified. Separate models were run for each year: 2006, 2007 and 2008. The independent variables were all assessed at the 95% significance level.

Chloraminated and chlorinated zones were modelled separately. RDF was not considered as an independent variable for chloraminated zones as there should be no, or minimal concentrations of, RDF in these zones. Only a small number of the zones in the dataset were chloraminated. Across the three years, an average of 5.25% of the zones were chloraminated.

An initial assessment of the data suggested that one of the water companies, Company I, had higher numbers of contacts than the other companies in the datasets. This is shown in Figure 3.2, with Company I highlighted. This outlier company was excluded from the analysis in order to prevent it having an excessive weighting upon the modelling. This is discussed further in Section 4.

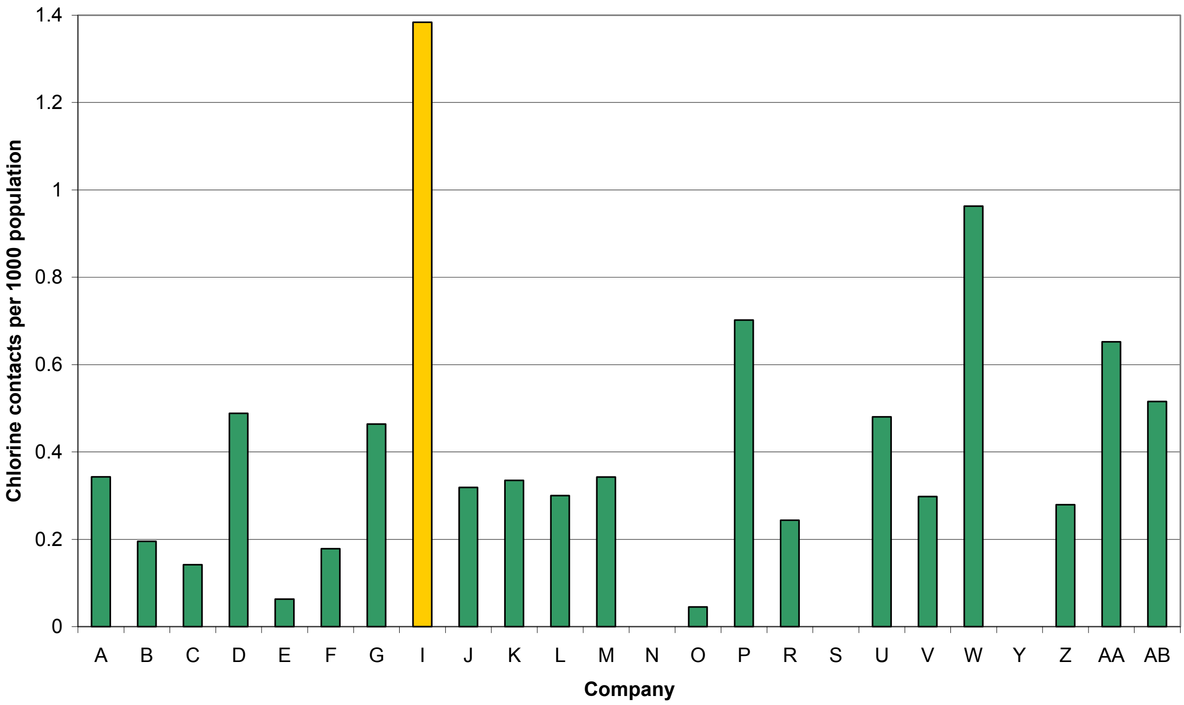

Figure 3.2 Comparison of chlorine contacts per 1000 population by company for 2008

Based on the information presented in Figure 3.2, Company W could also be considered an outlier. This diagram presents the average zonal rate of customer contacts (per 1000 population) for chlorine T&O, in 2008. Company W has three zones, one of which received a higher rate of chlorine T&O contacts. This has resulted in Company W having a high average rate of chlorine T&O contacts. Although the contact rate for this particular zone is high, it is still within the range experienced by the majority of the companies within the dataset. It was therefore felt that Company W should be retained within the analysis dataset.

4. RESULTS AND FINDINGS

4.1 National trends

Regression analyses were performed to determine whether relationships could be identified between the various measures of chlorine and customer contacts arising from chlorine T&O. The analyses were performed at a national level (England and Wales) and included all zones with sufficient sample data (see Section 3).

The results of the regression analyses are shown in Table 4.1. Where all the variables are significant (at the 95% confidence level), then the model exhibits a statistically significant relationship with the number of chlorine contacts per 1000 population. The full results table can be found in Appendix A.

Table 4.1 Summary of results of regression analyses, using chlorine indicators

Model	Chloraminated?	All Variables Significant?		
		2006	2007	2008
RDT_Mean	No	Yes	No	No
RDT_Max minus mean	No	No	No	No
RDT_StDev	No	No	No	No
RDF_Mean	No	No	Yes	No
RDF_Max minus mean	No	No	No	Yes
RDF_StDev	No	No	Yes	Yes
RDT_Mean	Yes	No	No	No
RDT_Max minus mean	Yes	No	No	No
RDT_StDev	Yes	No	No	No

As can been seen from Table 4.1 and the details of the statistical analysis presented in Appendix A, there were no modelled relationships that were found to be significant in all three years.

As mentioned in Section 3, Company I was excluded from the analyses. As an outlier, this water company was excluded because of the excessive weighting that these zones would have on any results. This can be illustrated by considering a single example of a regression analysis. When modelling the relationship between chlorine T&O contacts per 1000 population and the mean zonal RDT in 2008, no statistically significant relationship was found to exist in the national data (see Table 4.1). However, when the zones from Company I were included in the analysis, the relationship was found to be significant, (even at a confidence level of 99.9%). As there were only 23 zones, or 1.6% of the total number in this analysis, from Company I, it is clear that this small number of zones are having a disproportionate impact on the results. It is, therefore, prudent to remove Company I from the analyses.

The initial modelling of the data showed that there were also marked differences in chlorine concentrations and chlorine T&O contacts between the other water companies. The differences between water companies are illustrated in Figure 4.1.

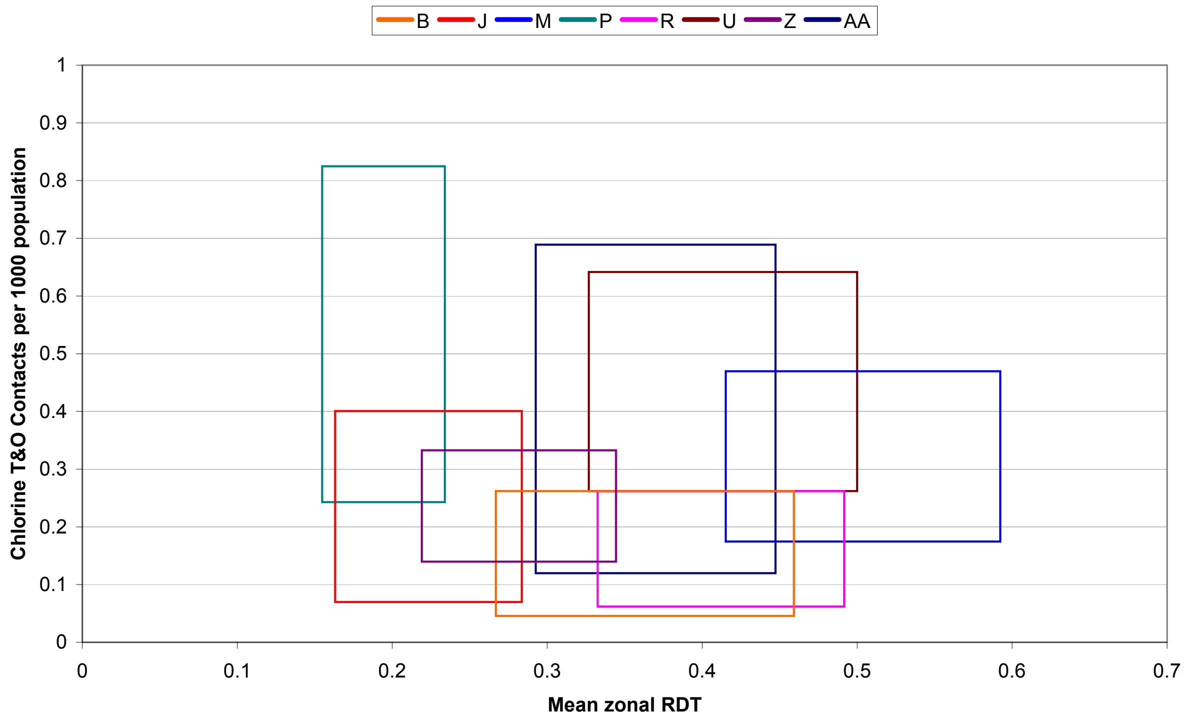

Figure 4.1 Quartile ranges of mean zonal RDT and chlorine contacts for 2008 for the eight companies with the most zones

Each box represents the quartile range of the zonal data (i.e. 50% of the chlorine contacts per 1000 population are to be found between the upper and lower limits of the box and 50% of the zonal mean RDT values are to be found between the left and right limits of the box). The figure shows the data for 2008 for the eight companies with the greatest number of zones, with the company codes shown in the key. Note that similar trends were observed in the data for both 2006 and 2007.

As can be seen from Figure 4.1, the different companies exhibit differing spreads of data. This pattern is fairly consistent and similar over all three years of data.

There are many factors that might contribute to the differences in chlorine T&O contact frequency between water companies. These might include:

- Source water – Different types of source water (e.g. upland surface water, groundwater) might exhibit different relationships between chlorine contacts and chlorine concentrations;

- Socio-economic factors – The socio-economic demographics of a water company's customers may affect how likely they are to report chlorine T&O to the water company;

- Cost of water – Customers may have different expectations based upon the price they pay for their water. This could influence their readiness to contact the water company; and,

- Interactions between water and customers' internal fittings.

Socio-economic factors are discussed in further detail in Section 5. However, data are not available under this current project to enable these other factors to be explored.

In light of the observed differences between water companies, two methods of analysing any underlying relationships between chlorine residuals and chlorine T&O contacts were utilised. In the first, “water company” was included in the model as an independent variable. Within this analysis a value was assigned to each water company to account for the impact of the differences between water companies. This analysis is detailed in the remainder of this section (Section 4.1). Note that this national level analysis did not include Company I, as previously discussed.

If a relationship exists across all (or most) water companies then it should be observable within individual water company data. The second method of analysis therefore considered data from individual companies in isolation. This second approach is detailed in Section 4.2.

When “water company” was included in the national-level models as an independent variable, the chlorine concentration indicator (RDT or RDF) was not found to be significant, with one exception: that of mean RDF in chlorinated zones. This parameter was found to be significant in all three years. The results of these analyses are shown in Table 4.2. Where all the variables are significant, (at the 95% confidence level) then the model exhibits a statistically significant relationship with the number of chlorine contacts per 1000 population[3]. The full results table can be found in Appendix B.

Table 4.2 Summary of results of regression analyses including water company as an independent variable, using chlorine indicators

Model	Chloraminated?	All Variables Significant?		
		2006	2007	2008
RDT_Mean	No	No	No	No
RDT_Max minus mean	No	No	No	No
RDT_StDev	No	No	No	No
RDF_Mean	No	Yes	Yes	Yes
RDF_Max minus mean	No	No	No	No
RDF_StDev	No	No	No	No
RDT_Mean	Yes	No	No	No
RDT_Max minus mean	Yes	No	No	No
RDT_StDev	Yes	No	No	No

This relationship, though statistically significant, does not appear to be strong. This can be seen in Figure 4.2 (all zones) and Figure 4.3 (detail of lower left corner of Figure 4.2, where majority of zones are found). These figures show the number of observed chlorine T&O contacts per 1000 population against the number predicted by the modelled relationship for 2007. The red line indicates the line of perfect match between actual and observed values.

[3] These analyses were also run with Company I included. The results matched those shown in Table 4.2. However, it is still recommended that Company I is excluded from the analyses because of its role as a clear outlier.

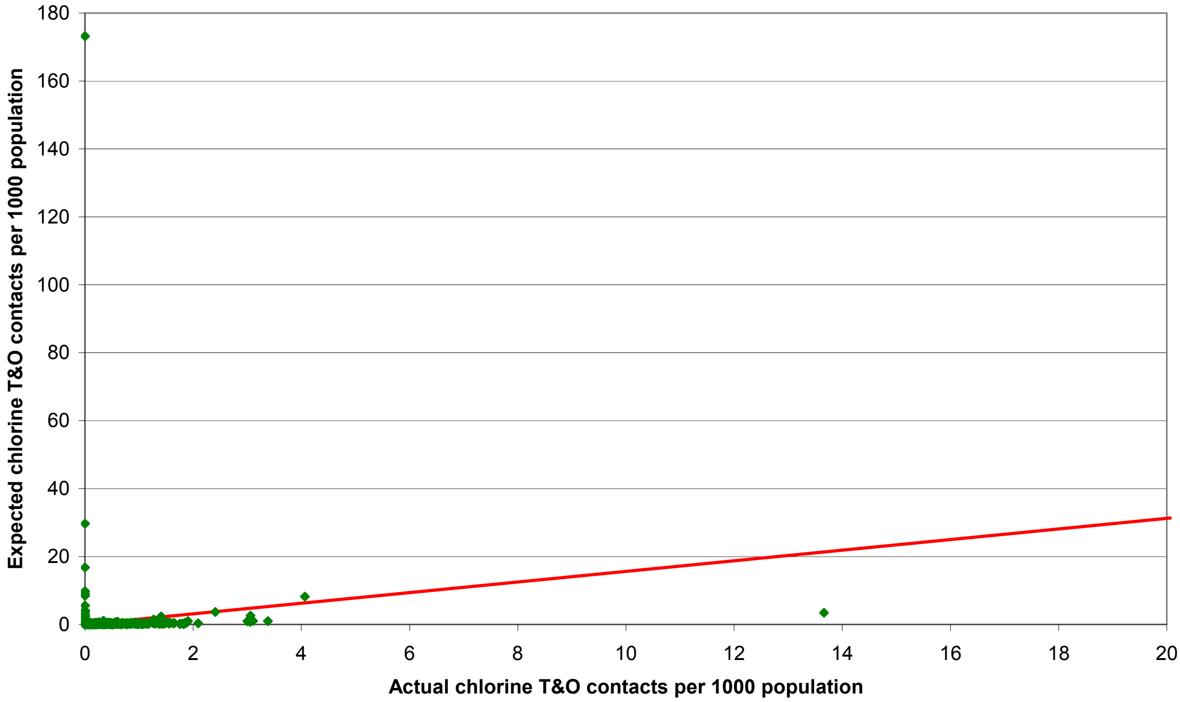

Figure 4.2 **Expected against actual chlorine contacts per 1000 population using RDF mean model for 2007**

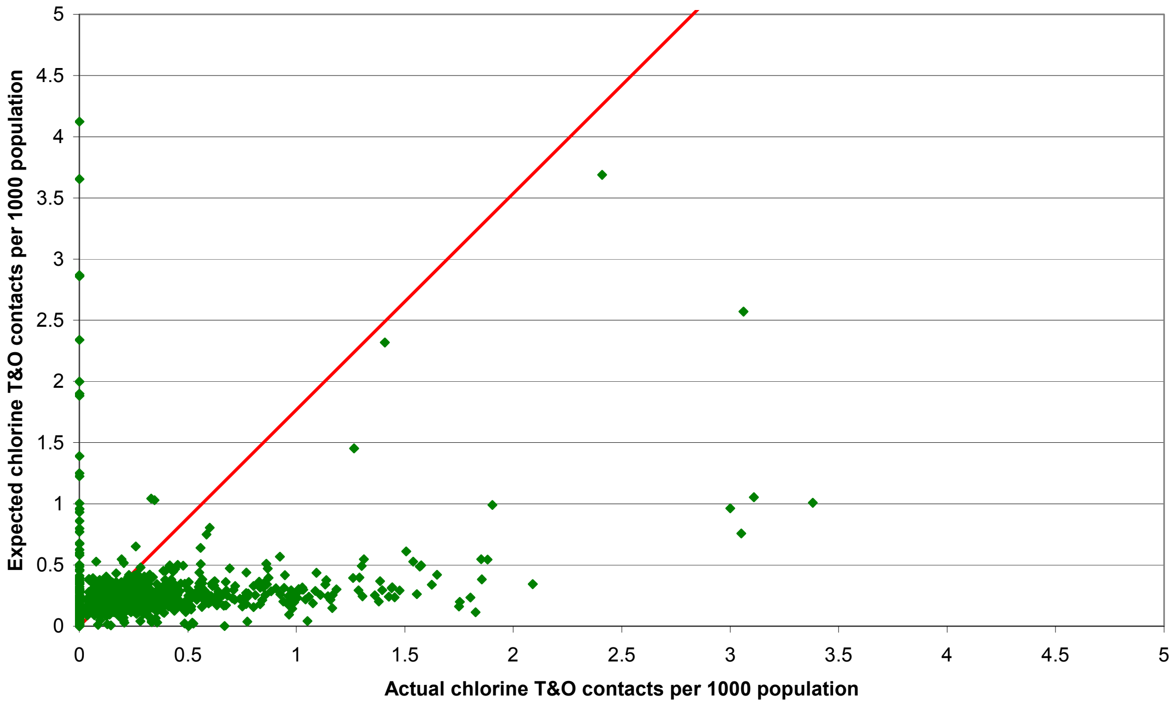

Figure 4.3 **Expected against actual chlorine contacts per 1000 population using RDF mean model for 2007 - detail of majority of zones**

It can be seen from the above figures that the match between the expected and actual values is poor. If the modelled relationship was perfect (i.e. the expected chlorine T&O contacts matched the actual chlorine T&O contacts) then the data points would lie along the red line shown in both figures. The more closely clustered around this line, the stronger the predictive power of the model.

In fact, a large proportion of the actual frequencies of customer contacts were between 0 and 2 contacts per 1000 population. However the model predicts that the same zones should have received between 0 and 0.5 per 1000 population. This clearly demonstrates that the identified relationship, although statistically significant, is weak.

This is further supported by Figure 4.4 which presents the number of chlorine T&O contacts per 1000 population against the mean zonal RDF in 2008, split by water company. There is no obvious relationship discernable from this figure, which again indicates that the correlation between mean RDF and chlorine T&O contacts per 1000 population is weak.

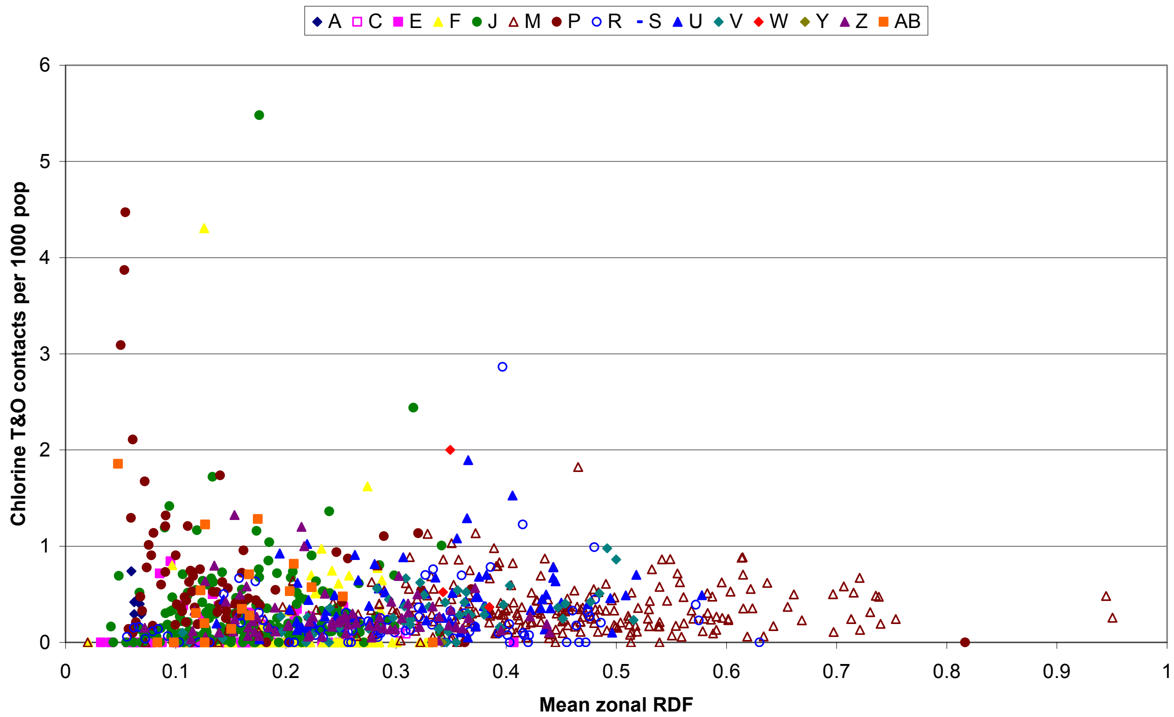

Figure 4.4 Chlorine contacts versus RDF levels for zones in 2008 split by water company

No statistically significant relationship could be identified for mean zonal RDT at national level. It can be confirmed that there is no obvious relationship by looking at Figure 4.5, which shows the number of chlorine T&O contacts per 1000 population against the mean zonal RDT in 2008 for chlorinated zones, split by water company.

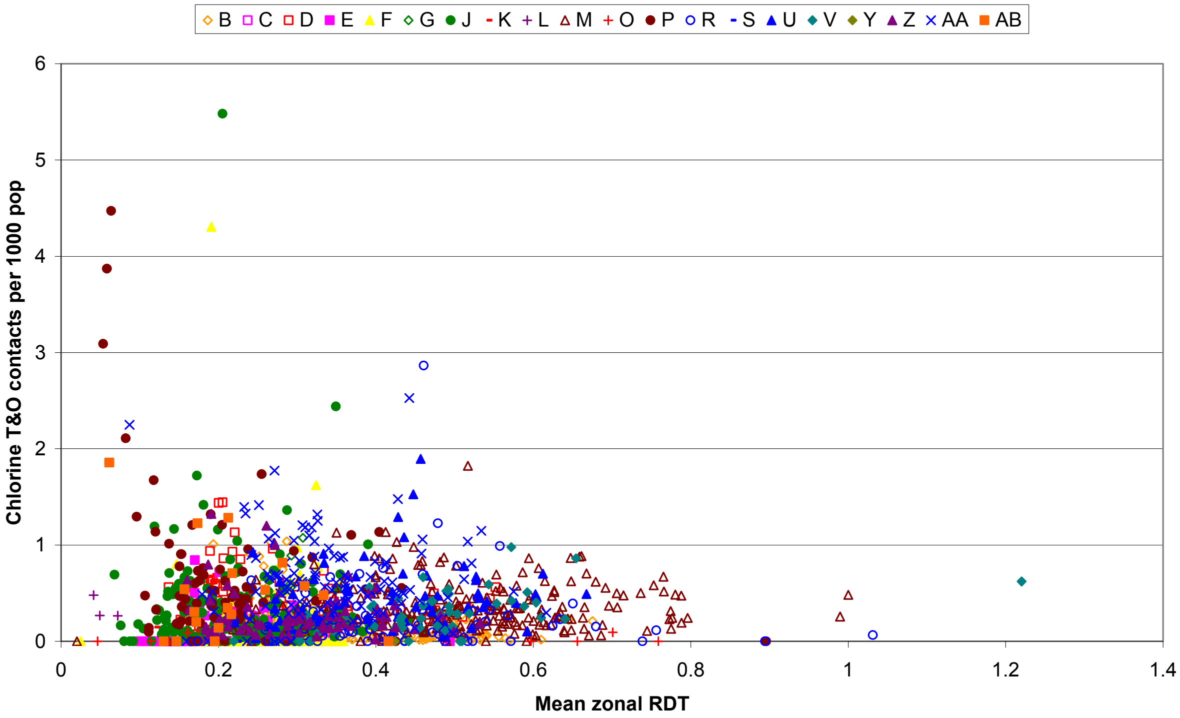

Figure 4.5 Chlorine contacts versus RDT levels for chlorinated zones in 2008 split by water company

Additional parameters were added to the models in order to try and improve the relationships. The results are summarised in Table 4.3, with the full details available in Appendix B.

Table 4.3 Summary of results of regression analyses including water company as an independent variable, including other parameters

Model	Chloraminated?	All Variables Significant?		
		2006	2007	2008
RDT_Mean + Colour_mean	No	No	Yes	No
RDT_Mean + Conductivity_Mean	No	No	No	Yes
RDT_Mean + Turbidity_Mean	No	No	No	No
RDT_Max minus mean + Colour_mean	No	No	No	No
RDT_Max minus mean + Conductivity_Mean	No	Yes	No	No
RDT_Max minus mean + Turbidity_Mean	No	No	No	No
RDT_StDev + Colour_mean	No	No	No	No
RDT_StDev + Conductivity_Mean	No	No	No	No
RDT_StDev + Turbidity_Mean	No	No	No	No

Model	Chloraminated?	All Variables Significant?		
		2006	2007	2008
RDF_Mean + Colour_mean	No	No	Yes	Yes
RDF_Mean + Conductivity_Mean	No	No	Yes	Yes
RDF_Mean + Turbidity_Mean	No	No	No	No
RDF_Max minus mean + Colour_mean	No	No	No	No
RDF_Max minus mean + Conductivity_Mean	No	Yes	No	No
RDF_Max minus mean + Turbidity_Mean	No	No	No	No
RDF_StDev + Colour_mean	No	No	No	No
RDF_StDev + Conductivity_Mean	No	Yes	Yes	No
RDF_StDev + Turbidity_Mean	No	No	No	No
RDT_Mean + Colour_mean	Yes	No	No	No
RDT_Mean + Conductivity_Mean	Yes	No	No	No
RDT_Mean + Turbidity_Mean	Yes	No	No	No
RDT_Max minus mean + Colour_mean	Yes	No	No	No
RDT_Max minus mean + Conductivity_Mean	Yes	No	No	No
RDT_Max minus mean + Turbidity_Mean	Yes	No	No	No
RDT_StDev + Colour_mean	Yes	No	No	No
RDT_StDev + Conductivity_Mean	Yes	No	No	No
RDT_StDev + Turbidity_Mean	Yes	No	No	No

The results presented in this table show there were no modelled relationships that were found to be significant in all three years.

The inclusion of an additional parameter (colour or conductivity) for one of the years (2006) yields a model in which the additional parameter is significant, but RDF mean is not significant. This suggests that there is a stronger relationship between the additional parameter and chlorine T&O contacts than there is between the RDF mean and chlorine T&O contacts. This could indicate that the significant relationship previously identified (when only RDF mean is included in the model) is actually representing a relationship between the additional parameter (colour or conductivity) and the number of chlorine T&O contacts. This could arise because RDF mean is weakly correlated with both colour and conductivity.

With only three available years of data to analyse, it is difficult to ascertain if this finding from 2006 is anomalous or not. Therefore, it is difficult to be confident that the identified national-level relationship is reliable. Further analysis of a greater number of years would indicate whether the impact of additional parameters identified in 2006 is merely an anomaly or whether it is a more common occurrence.

4.2 Individual water company analyses

Although a national level model showing a relationship between chlorine contacts and chlorine concentrations or variability has been found, the data does not fit the model particularly well (see Figure 4.2 and Figure 4.3). If this model is robust, then we would expect to see the same relationship between mean RDF and chlorine T&O contacts within individual water companies. It is, therefore, appropriate to analyse each water company individually to determine if relationships exists between chlorine concentrations and chlorine T&O contacts within a water company.

Although it was not within the scope of this project to investigate in detail any relationships that exist within individual water companies, three companies were selected in order to study this issue briefly (see Table 4.4).

Table 4.4 Water companies selected for individual company analyses and reason for selection

Water Company	Reason for selection
Company M	Large number of zones for analysis. Reasonable range of zonal chlorine levels and contacts. Mostly upland surface water.
Company Z	Reasonable range of zonal chlorine levels and contacts. Mostly groundwater.
Company AA	Mostly lowland surface water. Some chloraminated zones.

The same regression analysis methodology was applied to the datasets of each of the individual companies as was applied to the national dataset with the consideration of the same independent variables (RDT, RDF, Colour, Conductivity and Turbidity).

4.2.1 Results for Company M

Company M had a large number of zones available for analysis with a reasonable range of chlorine levels and chlorine T&O contacts. These factors allow for better determination of any relationships present in the data. Company M had no chloraminated zones and the source water is mainly upland surface water.

The results of the regression analyses are shown in Table 4.5. Where all the variables are significant, (at the 95% confidence level as before) then the model exhibits a statistically significant relationship with the number of chlorine contacts per 1000 population. The full results table can be found in Appendix C.

Table 4.5 Summary of results of regression analyses for Company M, using chlorine indicators

Model	All Variables Significant?		
	2006	2007	2008
RDT_Mean	No	No	No
RDT_Max minus mean	Yes	Yes	Yes
RDT_StDev	Yes	No	No
RDF_Mean	No	No	No
RDF_Max minus mean	Yes	Yes	Yes
RDF_StDev	Yes	No	No

The RDF mean value was not found to be significant in any of the three years for which data was available. This is in contrast to the findings from the national-level analysis and further reduces the confidence in the robustness of the national-level relationship and its applicability to individual companies.

The RDT maximum-minus-mean value, selected to indicate sudden peaks in chlorine concentrations, produced consistently significant models across all three years. This model has been highlighted in yellow in Table 4.5. It can be seen from this table that the RDF maximum-minus-mean also produce models across the three years, but these models were not as significant as the RDT maximum-minus-mean. However, this result does support the selection of the RDT maximum-minus-mean as being the most robust relationship with the frequency of chlorine T&O contacts.

Figure 4.6 shows the number of chlorine T&O contacts per 1000 population against the RDT maximum-minus-mean value for each zone in Company M, in 2008. The same constraints apply to this data as to the national dataset (i.e. only zones with 12 or more sample results in the year are included in the analysis).

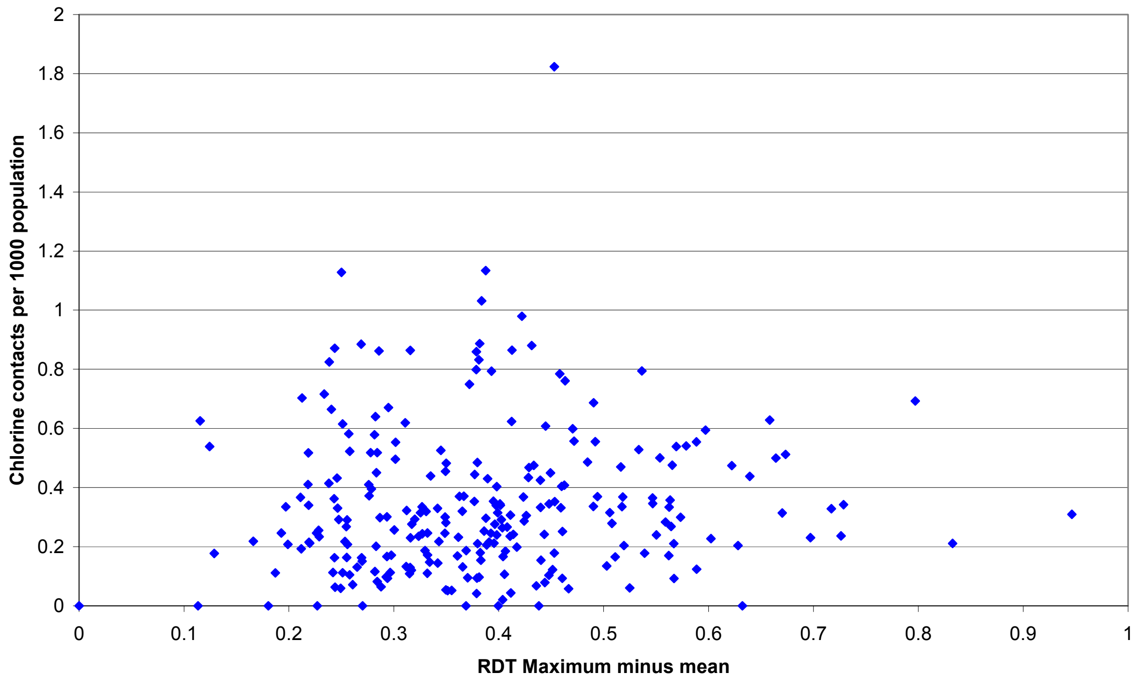

Figure 4.6 Chlorine contacts versus RDT levels for zones in Company M, 2008

No obvious relationship is discernible at first glance in this dataset. However, the regression analyses identified a significant relationship between this measure of chlorine concentrations and the chlorine contacts (Table 4.5). In this relationship, chlorine contacts increase as the magnitude of peaks in chlorine concentrations increases.

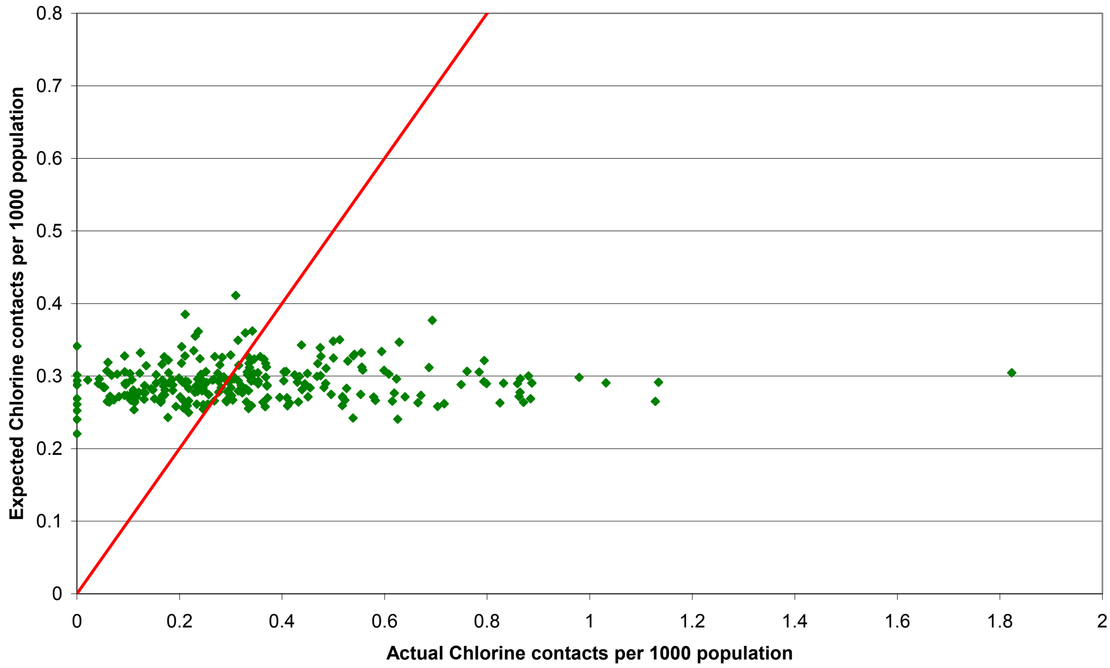

Figure 4.7 Expected against actual chlorine contacts per 1000 population for Company M using RDT maximum-minus-mean model for 2008

This relationship, though significant, is not strong. This can be seen in Figure 4.7, where expected chlorine contacts, as predicted by the identified relationship, are compared to the actual observed chlorine contacts per 1000 population.

As with the national level data, if the modelled relationship was perfect (i.e. expected chlorine contacts matched actual chlorine contacts) then the data-points would lie along the red line shown in the figure. The more closely clustered around this line, the stronger the predictive power of the model.

The figure clearly shows that the identified relationship, though statistically significant, is weak.

As described in Section 3, the regression analyses were extended to include other parameters to represent other factors which may influence customer contacts. The results of the regression analyses are shown in Table 4.6. The full results can be found in Appendix C.

Table 4.6 Summary of results of regression analyses for Company M, including other parameters

Model	All Variables Significant?		
	2006	2007	2008
RDT_Mean + Colour_mean	No	No	No
RDT_Mean + Conductivity_Mean	Yes	Yes	Yes
RDT_Mean + Turbidity_Mean	No	No	No
RDT_Max minus mean + Colour_mean	Yes	No	Yes
RDT_Max minus mean + Conductivity_Mean	Yes	Yes	Yes

Model	All Variables Significant?		
	2006	2007	2008
RDT_Max minus mean + Turbidity_Mean	Yes	No	No
RDT_StDev + Colour_mean	Yes	No	Yes
RDT_StDev + Conductivity_Mean	No	Yes	Yes
RDT_StDev + Turbidity_Mean	Yes	No	No
RDF_Mean + Colour_mean	No	No	No
RDF_Mean + Conductivity_Mean	Yes	Yes	Yes
RDF_Mean + Turbidity_Mean	No	No	No
RDF_Max minus mean + Colour_mean	Yes	Yes	Yes
RDF_Max minus mean + Conductivity_Mean	Yes	Yes	Yes
RDF_Max minus mean + Turbidity_Mean	Yes	No	No
RDF_StDev + Colour_mean	Yes	No	Yes
RDF_StDev + Conductivity_Mean	No	Yes	Yes
RDF_StDev + Turbidity_Mean	Yes	No	No

The regression analyses were used to determine that the relationship between RDT maximum-minus-mean and frequency of chlorine T&O contacts was strengthened by the inclusion of either colour or conductivity as an additional independent variable. These relationships are highlighted in yellow in Table 4.6.

This improvement has been demonstrated in Figure 4.8. This figure is the same as Figure 4.7, but for a model including the mean conductivity value.

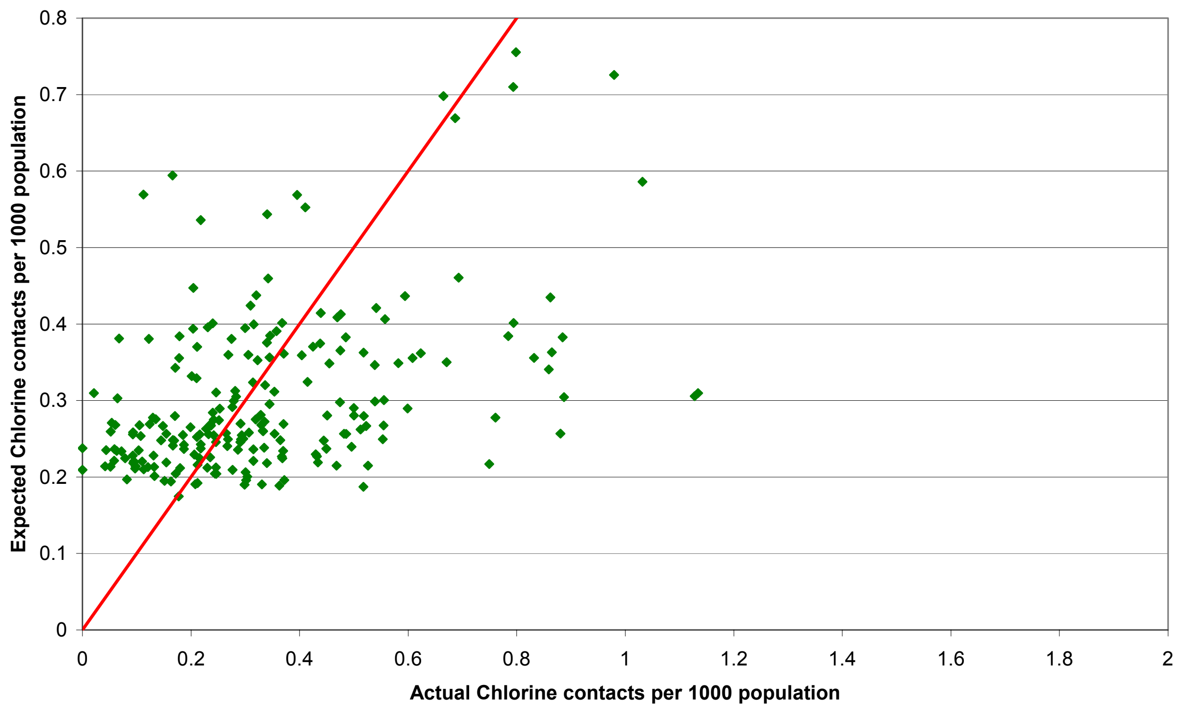

Figure 4.8 Expected against actual chlorine contacts per 1000 population for Company M using RDT maximum-minus-mean and conductivity mean model for 2008

The figure shows an improvement from Figure 4.7, with the data-points more clustered around the red line. The inclusion of conductivity has therefore significantly improved the relationship. In this relationship, chlorine contacts increase with an increasing magnitude in peaks in chlorine (RDT) concentrations and increasing conductivity. Conductivity has been included as an indicator of water hardness, as it was hypothesised that hardness could be linked to customer perception of chlorine T&O, or that customers with hardness problems are more likely to contact their water company with any issues.

Including colour instead of conductivity also improves the model, as shown in Figure 4.9. In this relationship, chlorine contacts increase with an increasing magnitude in peaks in chlorine (RDT) concentrations and decreasing colour. Colour has been included as a surrogate for chlorine demand. If the chlorine demand changes, the customer could experience varying chlorine concentrations. In addition, reactions between organics and chlorine can form T&O compounds more noticeable than chlorine itself but possibly categorised as chlorinous T&O.

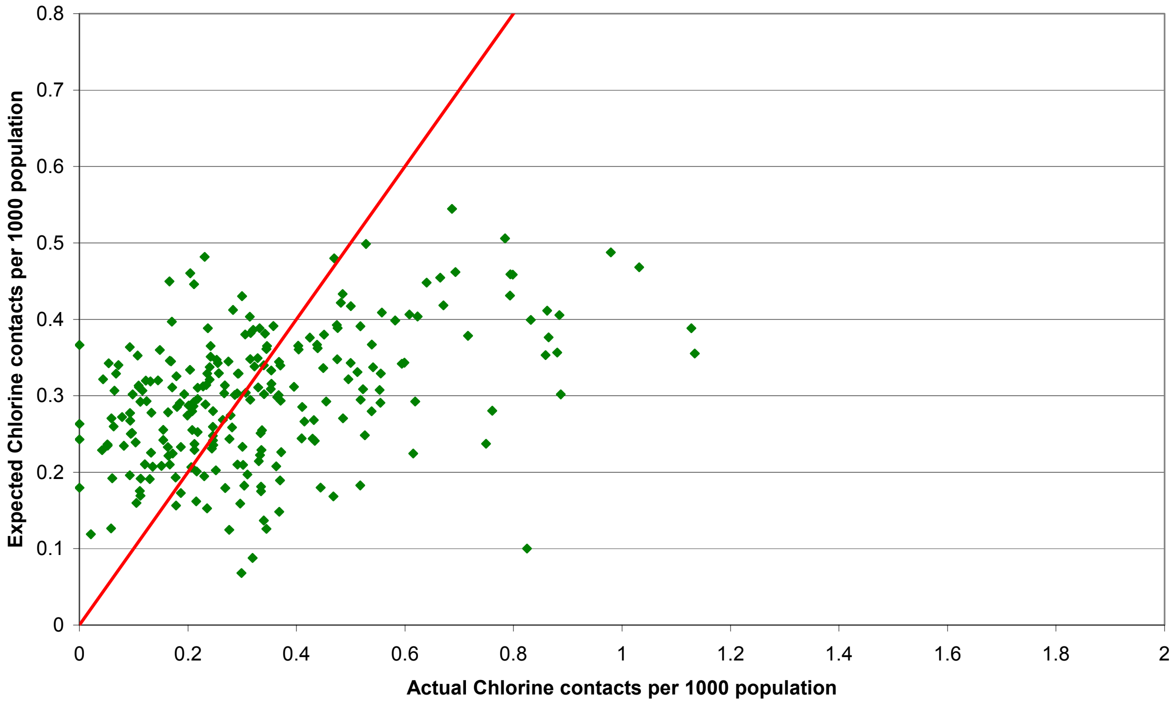

Figure 4.9 Expected against actual chlorine contacts per 1000 population for Company M using RDT maximum-minus-mean and colour mean model for 2008

Similar models exist when using RDF instead of RDT. All the significant models for Company M can be found in Appendix C. Two models, one using RDT and one using RDF, are preferred. These models were selected based upon the average standard errors of the models over the three years of data and the consistency in the models over these years.

The model including RDT has the form:

$$n/p = (A_0 + ((A_1 \times RDT_{\max - mean}) + (A_2 \times Cond_{mean}))^2$$

Where:

n = number of chlorine T&O contacts in a year in zone

p = zone population

$RDT_{max-mean}$ = Residual Disinfectant Total (Chlorine) maximum value minus mean in zone

$Cond_{mean}$ = Mean value of conductivity measured in zone

A_1 = Coefficient for RDT value (specific to year)

A_2 = Coefficient for conductivity value (specific to year)

The model including RDF is of the same form:

$$n/p = (A_0 + ((A_1 \times RDF_{\max - mean}) + (A_2 \times Cond_{mean}))^2$$

Where:

n = number of chlorine T&O contacts in a year

p = zone population

$RDF_{max-mean}$ = Residual Disinfectant Free (Chlorine) maximum value minus mean

$Cond_{mean}$ = Mean value of conductivity measured

A_1 = Coefficient for RDF value (specific to year)

A_2 = Coefficient for conductivity value (specific to year)

The regression analyses for Company M show that a weak relationship between chlorine T&O contacts and chlorine concentrations can be found for individual water companies. However, the mean RDF value was not found to be significant for this company. This is in contrast to the national level analysis.

4.2.2 Results for Company Z

Results derived from Company M's data suggested that intra-company relationships could be identified. Another company, Company Z, was isolated for analysis in order to determine if this was true for other companies as well.

Company Z, like Company M, has no chloraminated zones. It also has a reasonable number of zones for modelling. Company Z was selected for analysis because, in contrast to Company M and Company AA, the source waters are mostly groundwater. Any relationships between chlorine contacts and chlorine concentrations may be different for groundwater than for surface waters. In addition, there are likely to be less chlorination by-products in groundwater, which may affect the frequency of T&O contacts.

The regression analyses performed upon the three years of data could find no significant relationships between the RDT or RDF measures and the chlorine contacts per 1000 population. This was found to be the case with or without the inclusion of the additional independent variables (colour, conductivity and turbidity). The full results table, showing all the potential relationships modelled, can be found in Appendix D.

Of particular interest is the fact that, as with Company M, RDF mean was shown not be a significant variable for the regressions analyses performed on Company Z's data. This is in contrast to the national-level analysis and further reduces the confidence that is placed in the RDF mean/ chlorine T&O contacts model as a robust relationship across all (or the majority of) water companies.

Figure 4.10 shows, as an example, the number of chlorine T&O contacts per 1000 population against the RDT mean value for each zone in Company Z in 2007.

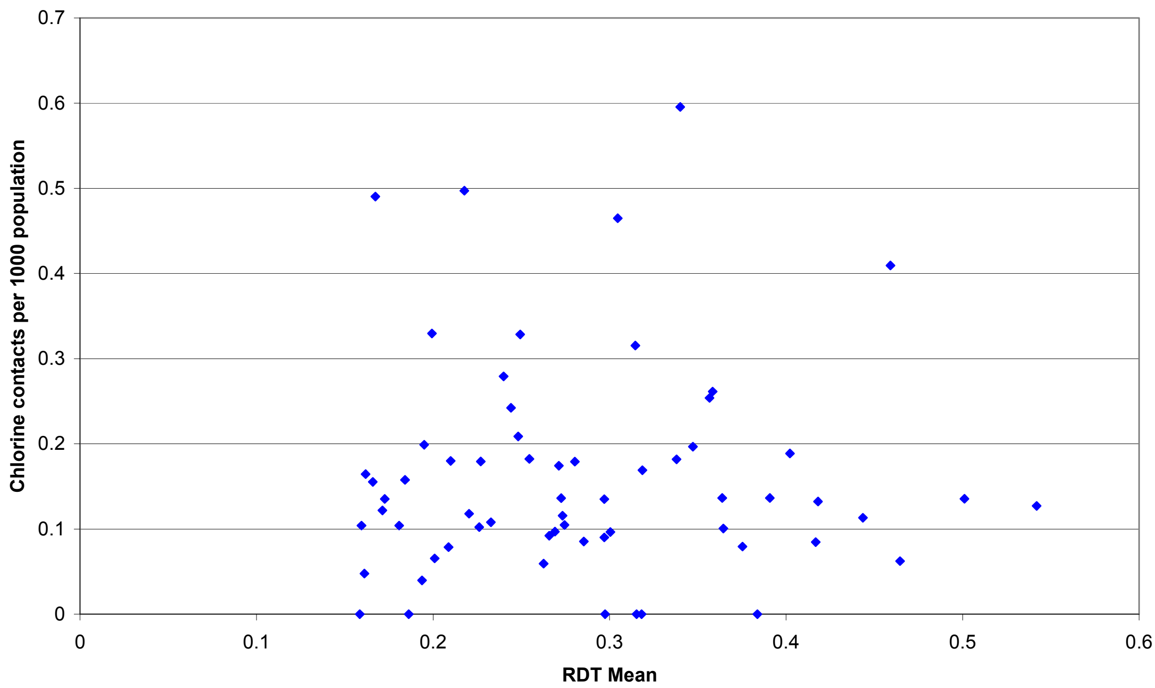

Figure 4.10 Chlorine contacts versus RDT levels for zones in Company Z, 2007

The figure shows that there is no obvious relationship in this data. This was confirmed by the statistical analysis.

The results from Company Z further reduce the confidence that can be placed in the weak national-level relationship that was identified in Section 4.1. It seems likely that this relationship is not appropriate to apply to individual water companies.

4.2.3 Results for Company AA

Company M's data suggested that intra-company relationships could be identified. However, results for Company Z showed no significant relationships. A third company, Company AA, was, therefore, selected for analysis.

This company, unlike Company M and Company Z, has a number of chloraminated zones. It is expected that there will be less formation of chlorination by-products in chloraminated zones. There should also be little or no RDF. This is expected to impact upon chlorine T&O contacts since free chlorine is more noticeable, in odour terms, than combined chlorine. This may reduce the frequency of chlorine T&O contacts in chloraminated zones.

Company AA was selected for analysis because the source waters are mainly lowland surface water. This is in contrast to Company M (upland surface water) and Company Z (groundwater). Company AA also has a reasonable number of zones for modelling. This company does not measure RDF so there are no models including this as a variable.

The regression analyses performed upon the three years of data could find no significant relationships between the RDT measures and the chlorine contacts per 1000 population. This was found to be the case with or without the inclusion of the additional independent variables (colour, conductivity and turbidity) and for both chloraminated and chlorinated zones. The full results table, showing all the potential relationships modelled, can be found in Appendix E.

Figure 4.11 shows, as an example, the number of chlorine T&O contacts per 1000 population against the RDT mean value for each chlorinated zone in Company AA in 2006. Figure 4.12 shows the same for chloraminated zones.

Both figures show that there are no obvious relationships in these datasets. This was confirmed by the statistical analyses.

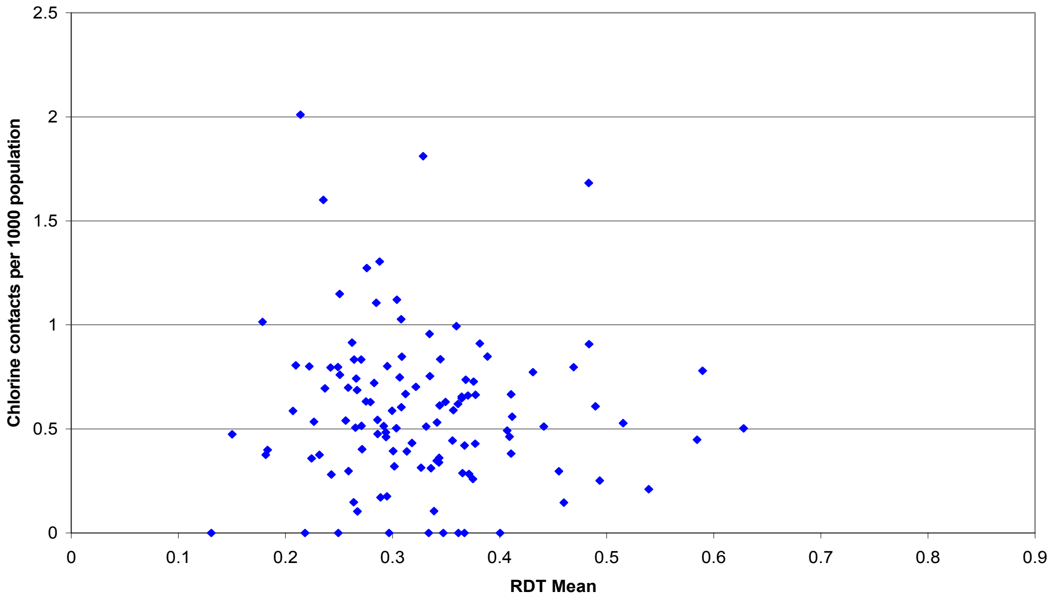

Figure 4.11 Chlorine contacts versus RDT mean concentrations for chlorinated zones in Company AA, 2006

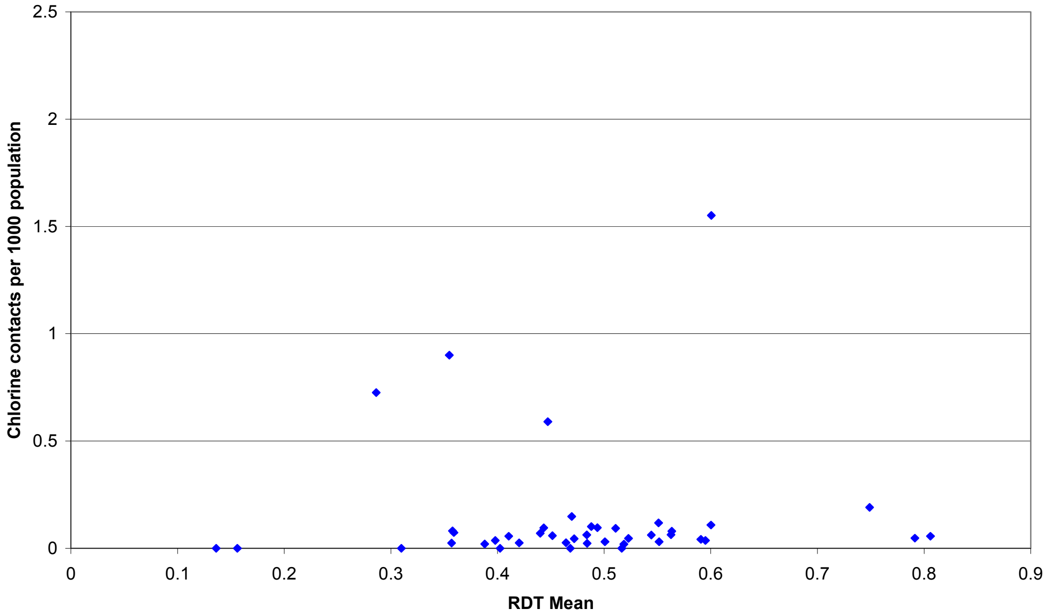

Figure 4.12 Chlorine contacts versus RDT mean concentrations for chloraminated zones in Company AA, 2006

However, a comparison of the two figures clearly shows the impact of chloramination on chlorine T&O contacts. There are more chlorine contacts, on average, in chlorinated zones than there are in chloraminated zones. It is expected that free chlorine provides the majority of the total residual chlorine recorded in chlorinated zones. However, with no measurements of RDF for Company AA, it is not possible to confirm this. The taste threshold for monochloramine is typically much greater than for free chlorine: a factor of 6 to 10 is not unusual. This means that where chloramination is introduced and properly controlled, chlorine T&O contacts can be significantly reduced. There may also be an impact on chlorine T&O contacts relating to the lack of chlorinated by-products in chloraminated zones.

4.3 Summary

Multi-variable linear regression analysis shows that, at a national level, the relationship between the water company and the number of chlorine contacts per 1000 population was significant, but only a weak relationship between chlorine concentrations and chlorine contacts could be identified. The confidence that can be placed in this relationship is weakened by the individual company analyses, in which this relationship was not observed. It appears that the national-level relationship is not appropriate for application to individual water companies. The differences between water companies are so pronounced that the water company must be taken into account in any analysis of the data.

In addition, analysis of a small number of carefully chosen water companies suggests that relationships with chlorine concentrations are present in one of the three companies assessed. However, the relationships observed in this one company were not strong.

5. SOCIO-ECONOMIC ANALYSIS

Customer contacts reflect customer perceptions of performance rather than being an absolute measure of water quality, such as regulatory sampling. Contacts can be influenced by other factors beyond the quality of water that has been delivered, such as the mix of social/economic groups or a changing population as customers relocate over time. It is therefore possible that socio-economic factors may have an impact upon the frequency of customer contacts. Zones with particularly high or low levels of deprivation may differ from the general trend observed with a water company. In addition to this, socio-economic factors may help explain differences in trends between companies.

Although a full socio-economic analysis was not possible under the scope of this project, a brief study of this issue was undertaken on a small number of zones.

Six zones were selected which exhibited chlorine T&O contacts which differed significantly from the values expected using the relationships established in Section 4.2. As only one of the modelled companies showed a significant relationship between chlorine concentrations and chlorine T&O contacts (Company M), all six zones were drawn from this company.

The zones were selected using the standardised residuals[4] from the fitted model for a single year: 2008. The model selected is that shown in Figure 4.8, where the expected chlorine contacts are calculated using the maximum-minus-mean value for RDT and the mean conductivity. The standardised residuals for 2008 are shown in Figure 5.1, with the selected zones shown in red.

The three zones with the highest residual values (model greatly under-estimates chlorine contacts) and the three zones with the lowest residual values (model greatly over-estimates chlorine contacts) were selected. The six zones each had a large enough population to ensure that low numbers of contacts could not be due to low populations.

[4] Standardised residuals enable zones that exhibit a poor fit with the model to be easily identified. They are calculated from the residuals (observed contacts minus estimated contacts) using the formula:

$$SR_i = \frac{(R_i - \overline{R})}{s.e.}$$

Where: SR_i is the standardised residual for zone i; R_i is the residual for zone i; $\overline{R}$ is the mean of the residuals; and, s.e. is the standard error of the residuals.

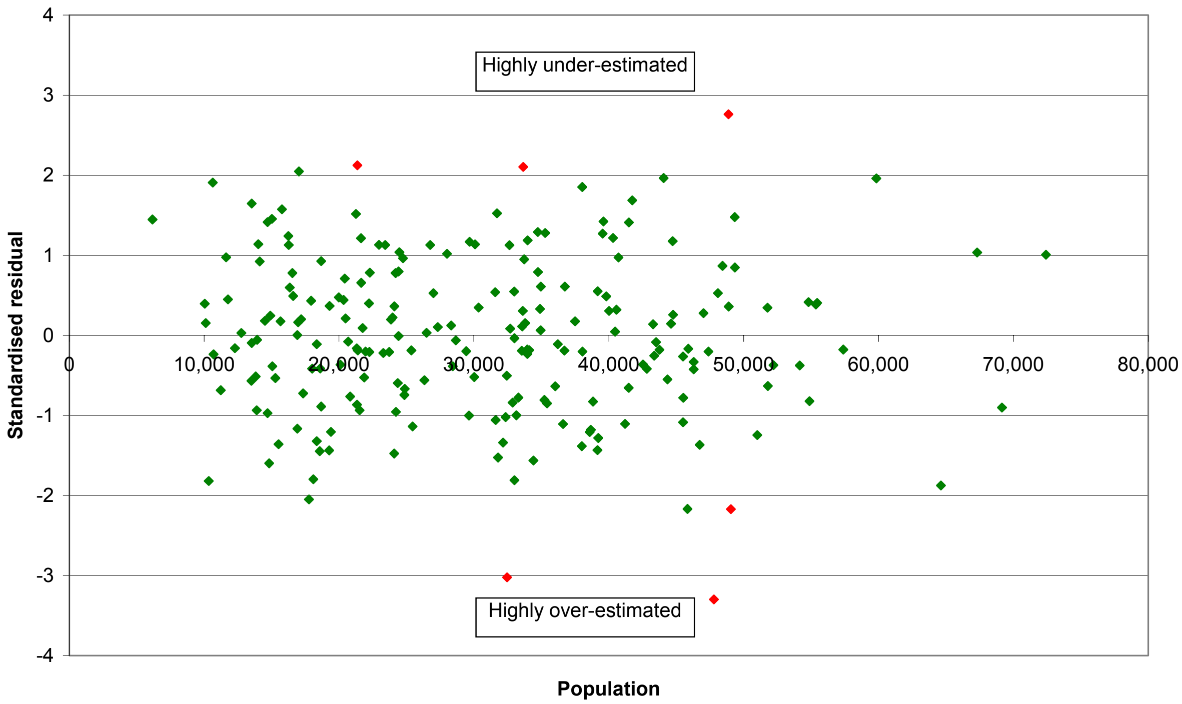

Figure 5.1 Standardised residuals for Company M

It was decided to compare the deprivation scores for these zones to determine if a pattern could be identified. In order to do this, the Index of Multiple Deprivation (IMD) 2007[5] on the Office of National Statistics (ONS) website was used. This assigns an IMD score to each Super Output Area (SOA) in England by combining individual indices for:

- Income;
- Employment;
- Health Deprivation & Disability;
- Education Skills & Training;
- Barriers to Housing & Services;
- Crime; and,
- Living Environment.

A higher IMD score indicates a more deprived SOA. The ONS has ranked all the SOAs in England by their scores, with a rank of 1 as the most deprived area (the area with the highest IMD score) and a rank of 32,482 as the least. These scores and ranks were used to indicate deprivation in the socio-economic analysis.

In order to match the zones in the analysis dataset with the SOAs used by the ONS, location data for the zones was needed.

[5] The Communities and Local Government (CLG) commissioned the Social Disadvantage Research Centre at the University of Oxford to update the Indices of Deprivation 2004 for 2007.

Ideally, postcode data would have been taken from the contacts dataset. This would have enabled a comparison of the number of contacts per population for different areas with different levels of deprivation. However, this dataset was not available.

Instead, postcodes from the water quality parameters dataset were grouped for each of the six zones. These postcodes were used to match the zones to SOAs. The boundaries for a zone and the boundaries for the SOAs are unlikely to match up exactly. In addition, a SOA that lies within a zone may have been missed if no test was carried out within it. Therefore, the IMD scores for each zone are regarded as a rough estimate of the general level of deprivation for the area.

Once the SOAs for each zone were identified, the IMD scores and ranks were then downloaded from the ONS website. The IMD scores and ranks for each zone were then compared to determine if there was a noticeable difference in these values between those zones that were significantly over-predicted by the model and those that were significantly under-predicted by the model.

Figure 5.2 shows the spread of IMD scores within each of the six zones, labelled *a* to *f*. A higher IMD **score** indicates a more deprived area. Figure 5.3 shows the same for the IMD ranks. A lower IMD **rank** indicates a more deprived area.

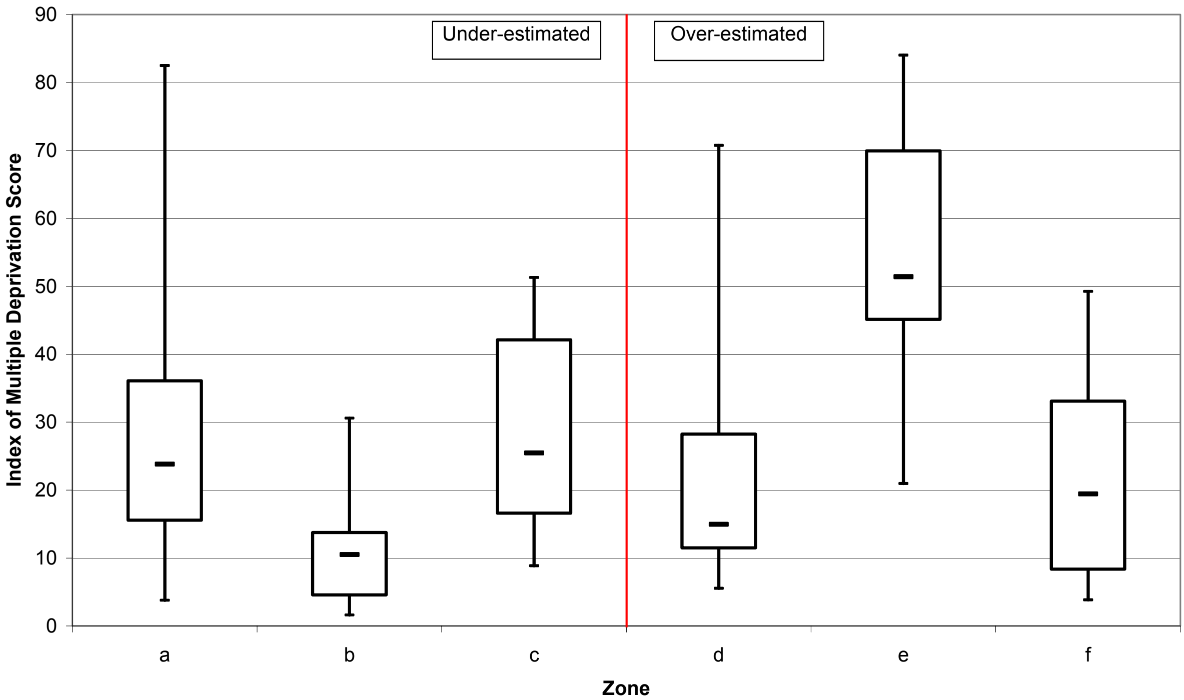

Figure 5.2 Box-plot of SOA IMD scores by zone for under- and over-predicted zones using 2008 Company M model

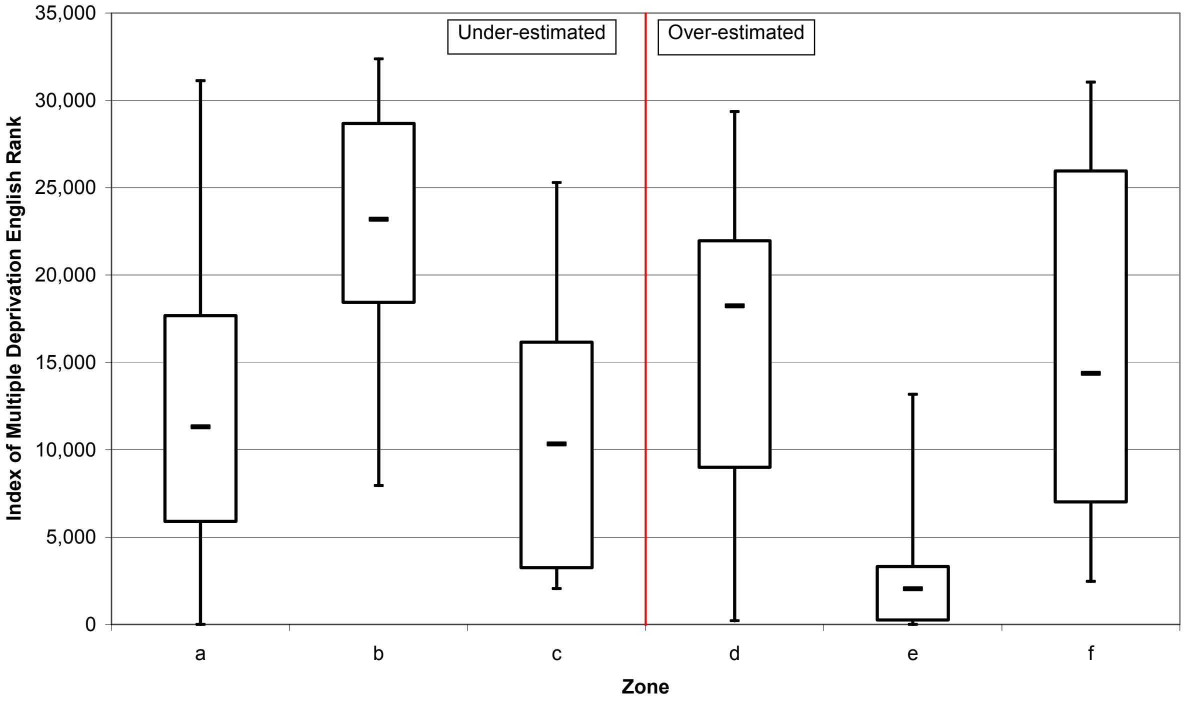

Figure 5.3 Box-plot of SOA IMD ranks by zone for under- and over-predicted zones using 2008 Company M model

The above figures show no clear difference in the IMD scores or ranks between those zones which were under-estimated by the model and those that were over-estimated. If socio-economic factors have any impact on the number of chlorine T&O contacts then it is not noticeable in these six zones.

It can be seen from the figures that Zone *e* contains highly deprived areas and has less chlorine contacts than predicted by the model whilst Zone *b* contains some of the least deprived areas in England and has more chlorine contacts than predicted by the model. Although this pattern is not observed in the other zones studied, further in-depth socio-economic analysis might determine if this is a wide-spread phenomenon.

The contacts recorded in Zone *e* for 2008 were lower than in the previous two years (see Table 5.1). This zone had six contacts for both 2006 and 2007, a number far closer to the expected value of 6.8 from the model. Note that contacts were received from customers in this zone relating to other T&O issues in 2008, which suggests that this is a genuine result for this year, rather than an issue with the data.

Table 5.1 Number of chlorine contacts by year for zones studied in socio-economic analysis

	Number of chlorine contacts		
Zone	**2006**	**2007**	**2008**
a	7	12	42
b	13	6	16
c	7	13	28
d	6	8	6
e	6	6	0
f	15	7	1

In general, it can be seen from Table 5.1 that the three zones which exhibited a greater number of chlorine contacts in 2008 than predicted by the model (*a, b* and *c*) had lower contact numbers in 2006 and 2007, whilst two of the three zones which were over-estimated by the model (*e* and *f*) had less contacts in 2008 than in 2006 and 2007. Therefore, the zones might be considered as having unusual contact numbers in 2008. The large residual values may be due to these zones experiencing an anomalous year instead of any socio-economic factor.

A full socio-economic analysis of this area was not possible within the scope of the project. There was no strong evidence to suggest that deprivation plays an important role in determining the level of chlorine contacts.

6. CONCLUSIONS

A detailed study has been undertaken to compare zonal data on the frequency of chlorine T&O contacts and sample results for chlorine concentrations. The following conclusions can be drawn from this study.

- Multi-variable linear regression identified a weak relationship between mean residual free chlorine concentrations (RDF) and the frequency of chlorine T&O contacts at a national level (England and Wales). However, the inclusion of either colour or conductivity suggested that there is a stronger relationship between colour or conductivity and chlorine T&O contacts than there is between the mean residual free chlorine and chlorine T&O contacts, reducing confidence in the reliability of this relationship. It is therefore difficult to be confident that the national level relationship is reliable.
- No other statistically significant associations were found between other measures of chlorine residual or variability and T&O contacts at the national level, across all three years analysed.
- There were marked differences in chlorine concentrations and chlorine T&O contacts between companies in England and Wales which may mask any relationships that might exist. Company specific factors, such as source water, operational practices, etc., may have a significant influence on the T&O contacts. However, it was not possible to explore these issues within this study.
- An analysis of data from three carefully selected water companies suggests that relationships between chlorine concentrations and T&O contacts for chlorine are present for one of these companies. However, these relationships were weak.
- Only one company out of the three assessed exhibited any relationships between chlorine concentrations and T&O contacts for chlorine.
- The RDT maximum-minus-mean, used to indicate sudden peaks in chlorine concentrations, produced consistently significant models across all three years (2006 to 2008) for this one company. The relationship observed indicated that T&O contacts increased as the magnitude of the peaks in chlorine concentrations increased. This relationship, although significant, is not strong.
- The inclusion of both the mean conductivity and mean colour values separately within the statistical models have strengthened these relationships, as follows:
  - T&O contacts for chlorine increase with increased magnitude of peaks in chlorine concentrations and increasing conductivity;
  - T&O contacts for chlorine increase with increased magnitude of peaks in chlorine concentrations and decreasing colour.
- The results of the regression analysis for this one company suggest that sudden increases in chlorine concentrations have a greater influence on the frequency of chlorine T&O contacts than the mean concentrations of chlorine. Unfortunately it was not possible to explore this phenomenon further, as time series data were not available for customer contacts.

- The comparison of chloraminated and chlorinated zones in one company showed a distinct difference in the frequency of chlorine T&O contacts. Substantially more contacts, on average, were received in chlorinated zones than there were for chloraminated zones. This may be due to either to the greater taste threshold of chloramine, or the fact that it does not react with organic matter, or other materials, to form T&O compounds, or some combination of these effects.

- The relationship (between mean RDF and the frequency of chlorine T&O contacts) observed at the national level was not found to be statistically significant for any of the companies that were studied in detail. This further reduces the confidence in the robustness of the national level relationship and its applicability to individual companies.

- An assessment of the Index of Multiple Deprivation for a small sample of zones suggests that factors such as income, employment, health and education did not play an important role in determining the frequency of chlorine T&O contacts within these zones.

7. RECOMMENDATIONS

The purpose of this study was to provide a better understanding of any relationships between chlorine concentration or variability and customer contacts relating to chlorine T&O. This was done using water quality sample data and annual numbers of contacts reported to the DWI by the water companies in England and Wales.

It was not possible to identify strong relationships between chlorine concentration or variability and customer contacts relating to chlorine T&O at a national level, and it is recommended that further work is undertaken to understand why this is the case, as follows:

- The weak relationship between mean residual free chlorine and T&O contacts identified at national-level was not identifiable at individual company level. Further analysis of individual companies might identify why this occurs. This would also provide the opportunity to determine whether relationships between chlorine concentrations and chlorine T&O contacts do exist, beyond the one company identified in this study. Companies should be selected on the following basis:

    - Range of chlorine concentrations across zones;
    - Range of conductivity results across zones;
    - Range of colour values across zones;
    - Sufficient number of zones.

- Company specific factors, such as source water, operational practices, etc., should be examined further to determine the influence these factors may have on the chlorine T&O contacts. In particular, the study into socio-economic factors should be extended.

Studies of this nature can assist the DWI and the UK water industry to better understand network performance and how customers respond to this performance. There are a number of recommendations relating to the type and format of the data collected by the DWI which may facilitate studies of this nature in the future, as follows:

- A zone reference field should be added to the customer contact dataset in order to facilitate the links between the contacts data, sample results and site data.

- Consistent zone referencing should be used between these three datasets.

- Customer contact data should be reported by the date it was received by the water company. This would allow time series analysis to be conducted, which is particularly relevant to the chlorine T&O issue.

- Customer contact data should also be referenced by post-code to enable a more detailed comparison of socio-economic measures, such as the Index of Multiple Deprivation.

APPENDIX A RESULT TABLES FOR NATIONAL-LEVEL REGRESSION ANALYSIS

The results of all the national-level regression analyses for are shown in the following tables. Each line of the table is an individual model.

Table A.1 shows the significance levels for each individual parameter in the regressions. A value of less than 0.05 indicates that the parameter is significant at the 95% confidence level. The significant parameters have been highlighted in the table. A model is considered significant if all of the parameters in the model are significant. The first columns indicate the composition of the model (e.g. Chloraminated zones with RDT_Mean and Colour_mean included as parameters) while the second set of columns (those split by parameter and year) show the significance level for each parameter in each of the three years.

Table A.2 shows the coefficient values for each individual parameter in the regressions. A negative coefficient indicates that the frequency of contacts decreases as the value of the parameter increases. A positive coefficient indicates that the frequency of contacts increases as the value of the parameter increases.

Table A.3 shows the standard errors of the fitted model and the number of data points (zones) included in each model.

Table A.1 Significance levels of national-level regression models

			√(population)			RDT/F * √(population)			Parameter * √(population)		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
N	RDT_Mean	none	<.001	<.001	<.001	0.005	0.091	0.7			
N	RDT_Mean	Colour_mean	<.001	<.001	<.001	<.001	0.351	0.593	<.001	<.001	<.001
N	RDT_Mean	Conductivity_Mean	<.001	<.001	<.001	<.001	0.713	0.051	0.012	<.001	0.022
N	RDT_Mean	Turbidity_Mean	<.001	<.001	<.001	0.01	0.094	0.712	0.545	0.953	0.233
N	RDT_Max minus mean	none	<.001	<.001	<.001	0.39	0.28	0.147			
N	RDT_Max minus mean	Colour_mean	<.001	<.001	<.001	0.481	0.442	0.299	<.001	<.001	<.001
N	RDT_Max minus mean	Conductivity_Mean	<.001	<.001	<.001	<.001	0.291	<.001	0.346	<.001	0.532
N	RDT_Max minus mean	Turbidity_Mean	<.001	<.001	<.001	0.381	0.285	0.162	0.724	0.842	0.216
N	RDT_StDev	none	<.001	<.001	<.001	0.707	0.067	0.053			
N	RDT_StDev	Colour_mean	<.001	<.001	<.001	0.915	0.172	0.091	<.001	<.001	<.001
N	RDT_StDev	Conductivity_Mean	<.001	<.001	<.001	0.781	0.044	0.005	0.107	<.001	0.776
N	RDT_StDev	Turbidity_Mean	<.001	<.001	<.001	0.659	0.074	0.065	0.759	0.864	0.22
N	RDF_Mean	none	<.001	<.001	<.001	0.837	<.001	0.001			
N	RDF_Mean	Colour_mean	<.001	<.001	<.001	0.574	<.001	0.007	0.153	0.055	<.001
N	RDF_Mean	Conductivity_Mean	<.001	<.001	<.001	0.848	<.001	0.017	0.115	0.232	0.553
N	RDF_Mean	Turbidity_Mean	<.001	<.001	<.001	0.955	<.001	0.005	0.857	0.776	0.829
N	RDF_Max minus mean	none	<.001	<.001	<.001	0.106	0.348	0.005			
N	RDF_Max minus mean	Colour_mean	<.001	<.001	<.001	0.166	0.43	0.005	0.218	0.006	<.001
N	RDF_Max minus mean	Conductivity_Mean	<.001	<.001	<.001	0.035	0.011	0.022	0.032	0.77	0.745
N	RDF_Max minus mean	Turbidity_Mean	<.001	<.001	<.001	0.134	0.394	0.011	0.811	0.391	0.69
N	RDF_StDev	none	<.001	<.001	<.001	0.206	0.029	<.001			
N	RDF_StDev	Colour_mean	<.001	<.001	<.001	0.314	0.064	<.001	0.236	0.008	<.001
N	RDF_StDev	Conductivity_Mean	<.001	<.001	<.001	0.65	<.001	0.001	0.1	0.028	0.157
N	RDF_StDev	Turbidity_Mean	<.001	<.001	<.001	0.228	0.048	<.001	0.798	0.433	0.721
Y	RDT_Mean	none	<.001	0.018	<.001	0.452	0.83	0.945			
Y	RDT_Mean	Colour_mean	0.026	0.002	<.001	0.456	0.853	0.396	0.722	0.041	<.001
Y	RDT_Mean	Conductivity_Mean	0.066	0.099	<.001	0.083	0.207	0.345	0.341	0.732	0.096
Y	RDT_Mean	Turbidity_Mean	0.002	0.001	<.001	0.507	0.768	0.376	0.911	0.018	0.008

			√(population)			RDT/F * √(population)			Parameter * √(population)		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
Y	RDT_Max minus mean	none	<.001	0.116	<.001	0.124	0.888	0.934			
Y	RDT_Max minus mean	Colour_mean	0.028	0.014	<.001	0.103	0.635	0.945	0.648	0.036	<.001
Y	RDT_Max minus mean	Conductivity_Mean	0.143	0.013	<.001	0.442	0.077	0.628	0.927	0.244	0.038
Y	RDT_Max minus mean	Turbidity_Mean	0.001	0.012	<.001	0.09	0.252	0.13	0.522	0.009	0.004
Y	RDT_StDev	none	<.001	0.121	<.001	0.199	0.654	0.864			
Y	RDT_StDev	Colour_mean	0.192	0.011	<.001	0.119	0.374	0.526	0.345	0.028	<.001
Y	RDT_StDev	Conductivity_Mean	0.14	0.024	<.001	0.056	0.129	0.254	0.537	0.355	0.073
Y	RDT_StDev	Turbidity_Mean	0.001	0.006	<.001	0.198	0.095	0.048	0.731	0.004	0.001

Table A.2 Coefficients for parameters in national-level regression models

			√(population)			RDT/F * √(population)			Parameter * √(population)		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
N	RDT_Mean	none	0.0147	0.0126	0.0153	-0.0040	0.0023	-0.0006			
N	RDT_Mean	Colour_mean	0.0205	0.0174	0.0203	-0.0063	0.0013	0.0009	-0.0043	-0.0042	-0.0055
N	RDT_Mean	Conductivity_Mean	0.0164	0.0158	0.0173	-0.0069	-0.0006	-0.0040	0.0000	0.0000	0.0000
N	RDT_Mean	Turbidity_Mean	0.0149	0.0124	0.0148	-0.0040	0.0025	-0.0007	-0.0021	0.0002	0.0034
N	RDT_Max minus mean	none	0.0134	0.0133	0.0150	0.0002	0.0001	0.0002			
N	RDT_Max minus mean	Colour_mean	0.0180	0.0178	0.0205	0.0001	0.0001	0.0001	-0.0039	-0.0042	-0.0055
N	RDT_Max minus mean	Conductivity_Mean	0.0119	0.0153	0.0126	0.0045	0.0004	0.0069	0.0000	0.0000	0.0000
N	RDT_Max minus mean	Turbidity_Mean	0.0135	0.0133	0.0145	0.0002	0.0001	0.0002	-0.0012	-0.0007	0.0035
N	RDT_StDev	none	0.0134	0.0131	0.0148	0.0006	0.0018	0.0021			
N	RDT_StDev	Colour_mean	0.0181	0.0177	0.0203	0.0002	0.0013	0.0019	-0.0039	-0.0042	-0.0055
N	RDT_StDev	Conductivity_Mean	0.0137	0.0143	0.0126	0.0014	0.0078	0.0191	0.0000	0.0000	0.0000
N	RDT_StDev	Turbidity_Mean	0.0135	0.0131	0.0143	0.0007	0.0018	0.0022	-0.0010	-0.0006	0.0035
N	RDF_Mean	none	0.0139	0.0121	0.0135	-0.0004	0.0084	0.0063			
N	RDF_Mean	Colour_mean	0.0153	0.0134	0.0164	-0.0012	0.0077	0.0057	-0.0010	-0.0011	-0.0029
N	RDF_Mean	Conductivity_Mean	0.0127	0.0111	0.0131	0.0005	0.0101	0.0061	0.0000	0.0000	0.0000
N	RDF_Mean	Turbidity_Mean	0.0137	0.0123	0.0137	-0.0001	0.0084	0.0059	0.0007	-0.0011	-0.0006

			√(population)			RDT/F * √(population)			Parameter * √(population)		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
N	RDF_Max minus mean	none	0.0137	0.0141	0.0134	0.0002	0.0002	0.0053			
N	RDF_Max minus mean	Colour_mean	0.0147	0.0157	0.0162	0.0002	0.0002	0.0058	-0.0008	-0.0016	-0.0031
N	RDF_Max minus mean	Conductivity_Mean	0.0114	0.0125	0.0130	0.0034	0.0044	0.0055	0.0000	0.0000	0.0000
N	RDF_Max minus mean	Turbidity_Mean	0.0135	0.0145	0.0136	0.0002	0.0002	0.0053	0.0010	-0.0033	-0.0012
N	RDF_StDev	none	0.0136	0.0138	0.0126	0.0019	0.0037	0.0221			
N	RDF_StDev	Colour_mean	0.0146	0.0153	0.0154	0.0017	0.0034	0.0240	-0.0008	-0.0015	-0.0032
N	RDF_StDev	Conductivity_Mean	0.0124	0.0091	0.0110	0.0031	0.0340	0.0271	0.0000	0.0000	0.0000
N	RDF_StDev	Turbidity_Mean	0.0134	0.0142	0.0128	0.0019	0.0036	0.0220	0.0010	-0.0030	-0.0011
Y	RDT_Mean	none	0.0070	0.0103	0.0096	0.0028	-0.0020	-0.0002			
Y	RDT_Mean	Colour_mean	0.0063	0.0193	0.0182	0.0032	0.0018	-0.0027	0.0004	-0.0091	-0.0058
Y	RDT_Mean	Conductivity_Mean	0.0090	0.0177	0.0204	0.0150	-0.0184	-0.0088	0.0000	0.0000	0.0000
Y	RDT_Mean	Turbidity_Mean	0.0073	0.0191	0.0137	0.0026	0.0028	0.0033	-0.0013	-0.0727	-0.0395
Y	RDT_Max minus mean	none	0.0060	0.0087	0.0096	0.0057	0.0020	-0.0004			
Y	RDT_Max minus mean	Colour_mean	0.0051	0.0177	0.0170	0.0064	0.0069	-0.0003	0.0005	-0.0093	-0.0057
Y	RDT_Max minus mean	Conductivity_Mean	0.0074	0.0249	0.0206	0.0048	-0.0186	-0.0049	0.0000	0.0000	0.0000
Y	RDT_Max minus mean	Turbidity_Mean	0.0068	0.0155	0.0133	0.0069	0.0174	0.0087	-0.0079	-0.0840	-0.0466
Y	RDT_StDev	none	0.0064	0.0075	0.0093	0.0123	0.0130	0.0016			
Y	RDT_StDev	Colour_mean	0.0039	0.0167	0.0178	0.0187	0.0266	-0.0055	0.0012	-0.0097	-0.0058
Y	RDT_StDev	Conductivity_Mean	0.0072	0.0227	0.0209	0.0445	-0.0510	-0.0323	0.0000	0.0000	0.0000
Y	RDT_StDev	Turbidity_Mean	0.0069	0.0152	0.0140	0.0130	0.0529	0.0220	-0.0041	-0.0934	-0.0529

Table A.3 Standard errors and number of data points in national-level regression models

			Standard Error				Data Points		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*Average*	*2006*	*2007*	*2008*
N	RDT_Mean	none	1.43	1.37	1.59	1.463	1433	1394	1381
N	RDT_Mean	Colour_mean	1.49	1.37	1.63	1.497	1166	1195	1129
N	RDT_Mean	Conductivity_Mean	1.52	1.46	1.77	1.583	989	975	943
N	RDT_Mean	Turbidity_Mean	1.52	1.44	1.7	1.553	1198	1209	1152
N	RDT_Max minus mean	none	1.43	1.37	1.59	1.463	1433	1394	1381
N	RDT_Max minus mean	Colour_mean	1.5	1.37	1.63	1.500	1166	1195	1129
N	RDT_Max minus mean	Conductivity_Mean	1.52	1.46	1.76	1.580	989	975	943
N	RDT_Max minus mean	Turbidity_Mean	1.53	1.44	1.7	1.557	1198	1209	1152
N	RDT_StDev	none	1.43	1.37	1.59	1.463	1433	1394	1381
N	RDT_StDev	Colour_mean	1.5	1.37	1.63	1.500	1166	1195	1129
N	RDT_StDev	Conductivity_Mean	1.54	1.46	1.76	1.587	989	975	943
N	RDT_StDev	Turbidity_Mean	1.53	1.44	1.7	1.557	1198	1209	1152
N	RDF_Mean	none	1.46	1.35	1.61	1.473	1009	981	1006
N	RDF_Mean	Colour_mean	1.6	1.45	1.73	1.593	797	798	815
N	RDF_Mean	Conductivity_Mean	1.64	1.53	1.83	1.667	717	675	720
N	RDF_Mean	Turbidity_Mean	1.58	1.45	1.74	1.590	823	805	827
N	RDF_Max minus mean	none	1.46	1.37	1.61	1.480	1009	981	1006
N	RDF_Max minus mean	Colour_mean	1.6	1.46	1.73	1.597	797	798	815
N	RDF_Max minus mean	Conductivity_Mean	1.63	1.54	1.83	1.667	717	675	720
N	RDF_Max minus mean	Turbidity_Mean	1.58	1.46	1.74	1.593	823	805	827
N	RDF_StDev	none	1.46	1.36	1.6	1.473	1009	981	1006
N	RDF_StDev	Colour_mean	1.6	1.46	1.72	1.593	797	798	815
N	RDF_StDev	Conductivity_Mean	1.64	1.52	1.83	1.663	717	675	720
N	RDF_StDev	Turbidity_Mean	1.58	1.46	1.73	1.590	823	805	827
Y	RDT_Mean	none	1.03	2.48	1.15	1.553	95	60	101
Y	RDT_Mean	Colour_mean	1.05	2.5	1.03	1.527	86	55	94
Y	RDT_Mean	Conductivity_Mean	1.06	1.01	1.19	1.087	47	16	54
Y	RDT_Mean	Turbidity_Mean	1.06	2.47	1.13	1.553	86	55	94
Y	RDT_Max minus mean	none	1.02	2.48	1.15	1.550	95	60	101

			Standard Error				**Data Points**		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*Average*	*2006*	*2007*	*2008*
Y	RDT_Max minus mean	Colour_mean	1.04	2.5	1.04	1.527	86	55	94
Y	RDT_Max minus mean	Conductivity_Mean	1.09	0.954	1.2	1.081	47	16	54
Y	RDT_Max minus mean	Turbidity_Mean	1.04	2.44	1.12	1.533	86	55	94
Y	RDT_StDev	none	1.03	2.48	1.15	1.553	95	60	101
Y	RDT_StDev	Colour_mean	1.04	2.48	1.03	1.517	86	55	94
Y	RDT_StDev	Conductivity_Mean	1.05	0.986	1.19	1.075	47	16	54
Y	RDT_StDev	Turbidity_Mean	1.05	2.41	1.11	1.523	86	55	94

APPENDIX B RESULT TABLES FOR NATIONAL-LEVEL REGRESSION ANALYSIS INCLUDING WATER COMPANY AS AN INDEPENDENT VARIABLE

The results of all national-level regression analyses with water company included as an independent variable are shown in the following tables. Each line of the table is an individual model. Water company is not shown in the tables, but is always significant.

Table B.1 shows the significance levels for each individual parameter in the regressions. A value of less than 0.05 indicates that the parameter is significant at the 95% confidence level. The significant parameters have been highlighted in the table. A model is considered significant if all of the parameters in the model are significant. The first columns indicate the composition of the model (e.g. Chloraminated zones with RDT_Mean and Colour_mean included as parameters) while the second set of columns (those split by parameter and year) show the significance level for each parameter in each of the three years.

Table B.2 shows the coefficient values for each individual parameter in the regressions. A negative coefficient indicates that the frequency of contacts decreases as the value of the parameter increases. A positive coefficient indicates that the frequency of contacts increases as the value of the parameter increases.

Table B.3 shows the standard errors of the fitted model and the number of data points (zones) included in each model.

Table B.1 Significance levels of national-level regression models

			√(population)			RDT/F * √(population)			Parameter * √(population)		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
N	RDT_Mean	none	<.001	<.001	<.001	0.424	0.863	0.134			
N	RDT_Mean	Colour_mean	<.001	<.001	<.001	0.746	0.035	0.769	<.001	<.001	<.001
N	RDT_Mean	Conductivity_Mean	<.001	<.001	<.001	0.123	0.923	0.019	0.04	0.001	<.001
N	RDT_Mean	Turbidity_Mean	<.001	<.001	<.001	0.562	0.748	0.09	0.976	0.934	0.529
N	RDT_Max minus mean	none	<.001	<.001	<.001	0.893	0.405	0.163			
N	RDT_Max minus mean	Colour_mean	<.001	<.001	<.001	0.905	0.597	0.242	<.001	<.001	<.001
N	RDT_Max minus mean	Conductivity_Mean	<.001	<.001	<.001	<.001	0.227	0.23	0.036	<.001	<.001
N	RDT_Max minus mean	Turbidity_Mean	<.001	<.001	<.001	0.959	0.427	0.214	0.956	0.927	0.508
N	RDT_StDev	none	<.001	<.001	<.001	0.764	0.465	0.262			
N	RDT_StDev	Colour_mean	<.001	<.001	<.001	0.929	0.476	0.245	<.001	<.001	<.001
N	RDT_StDev	Conductivity_Mean	<.001	<.001	<.001	0.397	0.262	0.352	0.036	<.001	<.001
N	RDT_StDev	Turbidity_Mean	<.001	<.001	<.001	0.687	0.522	0.377	0.905	0.919	0.517
N	RDF_Mean	none	<.001	<.001	<.001	0.01	<.001	0.002			
N	RDF_Mean	Colour_mean	<.001	<.001	<.001	0.536	0.005	0.044	<.001	<.001	<.001
N	RDF_Mean	Conductivity_Mean	<.001	0.007	0.027	0.065	<.001	0.003	<.001	<.001	<.001
N	RDF_Mean	Turbidity_Mean	<.001	<.001	<.001	0.018	<.001	0.005	0.395	0.707	0.732
N	RDF_Max minus mean	none	<.001	<.001	<.001	0.146	0.496	0.461			
N	RDF_Max minus mean	Colour_mean	<.001	<.001	<.001	0.229	0.58	0.949	<.001	<.001	<.001
N	RDF_Max minus mean	Conductivity_Mean	<.001	<.001	<.001	<.001	0.2	0.89	<.001	<.001	<.001
N	RDF_Max minus mean	Turbidity_Mean	<.001	<.001	<.001	0.174	0.545	0.449	0.663	0.362	0.837
N	RDF_StDev	none	<.001	<.001	<.001	0.137	0.245	0.785			
N	RDF_StDev	Colour_mean	<.001	<.001	<.001	0.287	0.371	0.332	<.001	<.001	<.001
N	RDF_StDev	Conductivity_Mean	<.001	0.006	0.003	0.041	0.005	0.343	<.001	<.001	<.001
N	RDF_StDev	Turbidity_Mean	<.001	<.001	<.001	0.178	0.338	0.969	0.653	0.37	0.858
Y	RDT_Mean	none	0.084	0.068	<.001	0.291	0.929	0.84			
Y	RDT_Mean	Colour_mean	0.074	0.005	<.001	0.463	0.588	0.795	0.247	0.032	0.007
Y	RDT_Mean	Conductivity_Mean	0.352	0.087	0.034	0.12	0.62	0.546	0.519	0.162	0.423
Y	RDT_Mean	Turbidity_Mean	0.174	0.006	<.001	0.282	0.794	0.896	0.694	0.03	0.321

			√(population)			RDT/F * √(population)			Parameter * √(population)		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
Y	RDT_Max minus mean	none	0.119	0.222	<.001	0.17	0.664	0.956			
Y	RDT_Max minus mean	Colour_mean	0.075	0.019	<.001	0.137	0.784	0.502	0.175	0.041	0.006
Y	RDT_Max minus mean	Conductivity_Mean	0.782	0.013	0.03	0.926	0.944	0.868	0.709	0.039	0.288
Y	RDT_Max minus mean	Turbidity_Mean	0.135	0.036	<.001	0.08	0.472	0.575	0.342	0.026	0.279
Y	RDT_StDev	none	0.305	0.314	<.001	0.059	0.392	0.716			
Y	RDT_StDev	Colour_mean	0.299	0.018	<.001	0.067	0.11	0.747	0.407	0.015	0.007
Y	RDT_StDev	Conductivity_Mean	0.397	0.412	0.053	0.103	0.275	0.255	0.593	0.657	0.62
Y	RDT_StDev	Turbidity_Mean	0.341	0.046	<.001	0.023	0.147	0.889	0.406	0.016	0.319

Table B.2 Coefficients for parameters in national-level regression models

			√(population)			RDT/F * √(population)			Parameter * √(population)		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
N	RDT_Mean	none	0.0110	0.0105	0.0125	-0.0013	0.0003	-0.0028			
N	RDT_Mean	Colour_mean	0.0164	0.0140	0.0165	0.0006	0.0036	0.0006	-0.0056	-0.0052	-0.0060
N	RDT_Mean	Conductivity_Mean	0.0079	0.0058	0.0077	-0.0031	-0.0002	-0.0055	0.0000	0.0000	0.0000
N	RDT_Mean	Turbidity_Mean	0.0097	0.0091	0.0116	-0.0010	0.0006	-0.0035	0.0001	-0.0003	0.0017
N	RDT_Max minus mean	none	0.0105	0.0105	0.0115	0.0000	0.0001	0.0001			
N	RDT_Max minus mean	Colour_mean	0.0165	0.0150	0.0166	0.0000	0.0001	0.0001	-0.0055	-0.0050	-0.0060
N	RDT_Max minus mean	Conductivity_Mean	0.0053	0.0055	0.0047	0.0039	0.0004	0.0020	0.0000	0.0000	0.0000
N	RDT_Max minus mean	Turbidity_Mean	0.0093	0.0092	0.0102	0.0000	0.0001	0.0001	0.0002	-0.0004	0.0018
N	RDT_StDev	none	0.0106	0.0105	0.0114	-0.0004	0.0006	0.0010			
N	RDT_StDev	Colour_mean	0.0165	0.0150	0.0165	-0.0001	0.0006	0.0011	-0.0055	-0.0050	-0.0060
N	RDT_StDev	Conductivity_Mean	0.0065	0.0051	0.0065	0.0041	0.0041	-0.0059	0.0000	0.0000	0.0000
N	RDT_StDev	Turbidity_Mean	0.0094	0.0092	0.0102	-0.0005	0.0006	0.0009	0.0004	-0.0004	0.0017
N	RDF_Mean	none	0.0095	0.0091	0.0094	0.0059	0.0105	0.0080			
N	RDF_Mean	Colour_mean	0.0154	0.0144	0.0143	0.0016	0.0073	0.0057	-0.0045	-0.0059	-0.0060
N	RDF_Mean	Conductivity_Mean	0.0060	0.0038	0.0034	0.0052	0.0125	0.0091	0.0000	0.0000	0.0000
N	RDF_Mean	Turbidity_Mean	0.0080	0.0080	0.0079	0.0060	0.0107	0.0081	0.0041	-0.0019	0.0010

			√(population)			RDT/F * √(population)			Parameter * √(population)		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
N	RDF_Max minus mean	none	0.0108	0.0115	0.0119	0.0002	0.0001	-0.0013			
N	RDF_Max minus mean	Colour_mean	0.0158	0.0167	0.0160	0.0002	0.0001	0.0001	-0.0046	-0.0064	-0.0063
N	RDF_Max minus mean	Conductivity_Mean	0.0052	0.0060	0.0062	0.0048	0.0021	-0.0003	0.0000	0.0000	0.0000
N	RDF_Max minus mean	Turbidity_Mean	0.0095	0.0109	0.0107	0.0002	0.0001	-0.0015	0.0021	-0.0046	0.0006
N	RDF_StDev	none	0.0106	0.0114	0.0112	0.0020	0.0019	0.0018			
N	RDF_StDev	Colour_mean	0.0157	0.0165	0.0154	0.0015	0.0015	0.0071	-0.0046	-0.0064	-0.0064
N	RDF_StDev	Conductivity_Mean	0.0054	0.0042	0.0050	0.0151	0.0213	0.0078	0.0000	0.0000	0.0000
N	RDF_StDev	Turbidity_Mean	0.0094	0.0108	0.0101	0.0019	0.0017	-0.0003	0.0021	-0.0045	0.0005
Y	RDT_Mean	none	0.0043	0.0140	0.0112	0.0048	-0.0011	-0.0009			
Y	RDT_Mean	Colour_mean	0.0072	0.0328	0.0186	0.0036	0.0069	-0.0011	-0.0025	-0.0162	-0.0045
Y	RDT_Mean	Conductivity_Mean	0.0094	0.0284	0.0202	0.0147	-0.0072	-0.0051	0.0000	-0.0001	0.0000
Y	RDT_Mean	Turbidity_Mean	0.0048	0.0296	0.0148	0.0051	-0.0032	-0.0006	-0.0055	-0.0824	-0.0199
Y	RDT_Max minus mean	none	0.0038	0.0106	0.0108	0.0064	0.0076	0.0003			
Y	RDT_Max minus mean	Colour_mean	0.0063	0.0324	0.0170	0.0070	0.0049	0.0038	-0.0028	-0.0148	-0.0046
Y	RDT_Max minus mean	Conductivity_Mean	0.0026	0.0337	0.0207	0.0008	-0.0007	0.0016	0.0000	-0.0001	0.0000
Y	RDT_Max minus mean	Turbidity_Mean	0.0046	0.0232	0.0136	0.0087	0.0129	0.0034	-0.0138	-0.0845	-0.0221
Y	RDT_StDev	none	0.0026	0.0083	0.0115	0.0285	0.0358	-0.0053			
Y	RDT_StDev	Colour_mean	0.0041	0.0276	0.0177	0.0303	0.0698	0.0047	-0.0018	-0.0179	-0.0045
Y	RDT_StDev	Conductivity_Mean	0.0082	0.0158	0.0185	0.0441	-0.0616	-0.0305	0.0000	0.0000	0.0000
Y	RDT_StDev	Turbidity_Mean	0.0031	0.0202	0.0143	0.0365	0.0624	0.0021	-0.0115	-0.0919	-0.0204

Table B.3 Standard errors and number of data points in national-level regression models

			Standard Error				Data Points		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*Average*	*2006*	*2007*	*2008*
N	RDT_Mean	none	1.18	1.19	1.32	1.230	1432	1393	1380
N	RDT_Mean	Colour_mean	1.18	1.17	1.33	1.227	1165	1194	1128
N	RDT_Mean	Conductivity_Mean	1.26	1.27	1.44	1.323	988	974	942
N	RDT_Mean	Turbidity_Mean	1.22	1.23	1.39	1.280	1197	1208	1151
N	RDT_Max minus mean	none	1.18	1.19	1.32	1.230	1432	1393	1380
N	RDT_Max minus mean	Colour_mean	1.18	1.17	1.33	1.227	1165	1194	1128
N	RDT_Max minus mean	Conductivity_Mean	1.26	1.27	1.44	1.323	988	974	942
N	RDT_Max minus mean	Turbidity_Mean	1.22	1.23	1.39	1.280	1197	1208	1151
N	RDT_StDev	none	1.18	1.19	1.32	1.230	1432	1393	1380
N	RDT_StDev	Colour_mean	1.18	1.17	1.33	1.227	1165	1194	1128
N	RDT_StDev	Conductivity_Mean	1.27	1.27	1.44	1.327	988	974	942
N	RDT_StDev	Turbidity_Mean	1.22	1.23	1.39	1.280	1197	1208	1151
N	RDF_Mean	none	1.25	1.24	1.38	1.290	1008	980	1005
N	RDF_Mean	Colour_mean	1.31	1.28	1.45	1.347	796	797	814
N	RDF_Mean	Conductivity_Mean	1.36	1.35	1.51	1.407	716	674	719
N	RDF_Mean	Turbidity_Mean	1.31	1.31	1.47	1.363	822	804	826
N	RDF_Max minus mean	none	1.25	1.25	1.38	1.293	1008	980	1005
N	RDF_Max minus mean	Colour_mean	1.31	1.29	1.45	1.350	796	797	814
N	RDF_Max minus mean	Conductivity_Mean	1.35	1.37	1.52	1.413	716	674	719
N	RDF_Max minus mean	Turbidity_Mean	1.32	1.33	1.48	1.377	822	804	826
N	RDF_StDev	none	1.25	1.25	1.39	1.297	1008	980	1005
N	RDF_StDev	Colour_mean	1.31	1.29	1.45	1.350	796	797	814
N	RDF_StDev	Conductivity_Mean	1.35	1.37	1.52	1.413	716	674	719
N	RDF_StDev	Turbidity_Mean	1.32	1.33	1.48	1.377	822	804	826
Y	RDT_Mean	none	1.04	2.56	1.08	1.560	94	59	100
Y	RDT_Mean	Colour_mean	1.05	2.53	1.02	1.533	85	54	93
Y	RDT_Mean	Conductivity_Mean	1.03	0.734	1.07	0.945	46	15	53
Y	RDT_Mean	Turbidity_Mean	1.05	2.53	1.06	1.547	85	54	93
Y	RDT_Max minus mean	none	1.03	2.56	1.08	1.557	94	59	100

			Standard Error				Data Points		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*Average*	*2006*	*2007*	*2008*
Y	RDT_Max minus mean	Colour_mean	1.04	2.54	1.02	1.533	85	54	93
Y	RDT_Max minus mean	Conductivity_Mean	1.06	0.744	1.08	0.961	46	15	53
Y	RDT_Max minus mean	Turbidity_Mean	1.04	2.52	1.06	1.540	85	54	93
Y	RDT_StDev	none	1.02	2.54	1.08	1.547	94	59	100
Y	RDT_StDev	Colour_mean	1.03	2.47	1.02	1.507	85	54	93
Y	RDT_StDev	Conductivity_Mean	1.03	0.694	1.06	0.928	46	15	53
Y	RDT_StDev	Turbidity_Mean	1.03	2.47	1.06	1.520	85	54	93

APPENDIX C RESULT TABLES FOR REGRESSION ANALYSIS FOR COMPANY M

The results of all regression analyses for Company M are shown in the following tables. Each line of the table is an individual model.

Table C.1 shows the significance levels for each individual parameter in the regressions. A value of less than 0.05 indicates that the parameter is significant at the 95% confidence level. The significant parameters have been highlighted in the table. A model is considered significant if all of the parameters in the model are significant. The first columns indicate the composition of the model (e.g. RDT_Mean and Colour_mean included as parameters) while the second set of columns (those split by parameter and year) show the significance level for each parameter in each of the three years.

Table C.2 shows the coefficient values for each individual parameter in the regressions. A negative coefficient indicates that the frequency of contacts decreases as the value of the parameter increases. A positive coefficient indicates that the frequency of contacts increases as the value of the parameter increases.

Table C.3 shows the standard errors of the fitted model and the number of data points (zones) included in each model.

Table C.1 Significance levels of regression models for Company M

		√(population)			RDT/F * √(population)			Parameter * √(population)		
Chlorine Value	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
RDT_Mean	none	<.001	<.001	<.001	0.075	0.128	0.419			
RDT_Mean	Colour_mean	<.001	<.001	<.001	0.676	0.243	0.7	<.001	<.001	<.001
RDT_Mean	Conductivity_Mean	0.001	<.001	<.001	<.001	0.002	0.016	<.001	<.001	<.001
RDT_Mean	Turbidity_Mean	<.001	<.001	<.001	0.05	0.184	0.401	0.005	0.587	0.323
RDT_Max minus mean	none	<.001	<.001	<.001	<.001	0.025	0.034			
RDT_Max minus mean	Colour_mean	<.001	<.001	<.001	<.001	0.059	0.003	<.001	<.001	<.001
RDT_Max minus mean	Conductivity_Mean	<.001	<.001	<.001	<.001	0.004	0.038	<.001	<.001	<.001
RDT_Max minus mean	Turbidity_Mean	<.001	<.001	<.001	<.001	0.02	0.043	0.02	0.404	0.491
RDT_StDev	none	<.001	<.001	<.001	<.001	0.702	0.082			
RDT_StDev	Colour_mean	<.001	<.001	<.001	<.001	0.231	0.009	<.001	<.001	<.001
RDT_StDev	Conductivity_Mean	0.284	<.001	<.001	<.001	0.033	0.005	<.001	<.001	<.001
RDT_StDev	Turbidity_Mean	<.001	<.001	<.001	<.001	0.653	0.1	0.009	0.517	0.446
RDF_Mean	none	<.001	<.001	<.001	0.182	0.148	0.444			
RDF_Mean	Colour_mean	<.001	<.001	<.001	0.804	0.327	0.622	<.001	<.001	<.001
RDF_Mean	Conductivity_Mean	<.001	<.001	<.001	<.001	<.001	0.006	<.001	<.001	<.001
RDF_Mean	Turbidity_Mean	<.001	<.001	<.001	0.114	0.223	0.418	0.005	0.612	0.318
RDF_Max minus mean	none	<.001	<.001	<.001	<.001	0.018	0.03			
RDF_Max minus mean	Colour_mean	<.001	<.001	<.001	<.001	0.034	0.012	<.001	<.001	<.001
RDF_Max minus mean	Conductivity_Mean	<.001	<.001	<.001	<.001	0.005	0.011	<.001	<.001	<.001
RDF_Max minus mean	Turbidity_Mean	<.001	<.001	<.001	<.001	0.014	0.036	0.018	0.405	0.444
RDF_StDev	none	<.001	<.001	<.001	<.001	0.399	0.099			
RDF_StDev	Colour_mean	<.001	<.001	<.001	0.001	0.102	0.017	<.001	<.001	<.001
RDF_StDev	Conductivity_Mean	0.245	<.001	<.001	<.001	0.01	0.002	<.001	<.001	<.001
RDF_StDev	Turbidity_Mean	<.001	<.001	<.001	<.001	0.362	0.121	0.005	0.519	0.406

Table C.2 Coefficients for parameters in regression models for Company M

		√(population)			**RDT/F** * √(population)			**Parameter** * √(population)		
Chlorine Value	*Parameter*	2006	2007	2008	2006	2007	2008	2006	2007	2008
RDT_Mean	none	0.0129	0.0145	0.0160	0.0046	0.0050	0.0024			
RDT_Mean	Colour_mean	0.0263	0.0229	0.0274	0.0011	0.0037	-0.0011	-0.0104	-0.0081	-0.0098
RDT_Mean	Conductivity_Mean	0.0049	0.0075	0.0094	0.0114	0.0094	0.0070	0.0000	0.0000	0.0000
RDT_Mean	Turbidity_Mean	0.0107	0.0154	0.0142	0.0052	0.0044	0.0026	0.0152	-0.0058	0.0117
RDT_Max minus mean	none	0.0127	0.0150	0.0149	0.0053	0.0045	0.0057			
RDT_Max minus mean	Colour_mean	0.0248	0.0229	0.0240	0.0055	0.0037	0.0075	-0.0108	-0.0079	-0.0101
RDT_Max minus mean	Conductivity_Mean	0.0091	0.0102	0.0110	0.0048	0.0052	0.0054	0.0000	0.0000	0.0000
RDT_Max minus mean	Turbidity_Mean	0.0113	0.0160	0.0137	0.0050	0.0048	0.0056	0.0121	-0.0089	0.0080
RDT_StDev	none	0.0089	0.0163	0.0144	0.0383	0.0041	0.0175			
RDT_StDev	Colour_mean	0.0209	0.0230	0.0231	0.0352	0.0126	0.0249	-0.0102	-0.0086	-0.0101
RDT_StDev	Conductivity_Mean	0.0018	0.0088	0.0086	0.0544	0.0214	0.0273	0.0000	0.0000	0.0000
RDT_StDev	Turbidity_Mean	0.0073	0.0169	0.0132	0.0376	0.0049	0.0170	0.0136	-0.0069	0.0089
RDF_Mean	none	0.0137	0.0150	0.0162	0.0036	0.0048	0.0022			
RDF_Mean	Colour_mean	0.0267	0.0234	0.0274	0.0006	0.0032	-0.0014	-0.0105	-0.0081	-0.0098
RDF_Mean	Conductivity_Mean	0.0057	0.0075	0.0094	0.0118	0.0108	0.0078	0.0000	0.0000	0.0000
RDF_Mean	Turbidity_Mean	0.0115	0.0158	0.0144	0.0043	0.0042	0.0024	0.0152	-0.0054	0.0119
RDF_Max minus mean	none	0.0127	0.0149	0.0146	0.0058	0.0049	0.0068			
RDF_Max minus mean	Colour_mean	0.0249	0.0227	0.0240	0.0063	0.0043	0.0073	-0.0111	-0.0079	-0.0098
RDF_Max minus mean	Conductivity_Mean	0.0089	0.0102	0.0103	0.0055	0.0054	0.0076	0.0000	0.0000	0.0000
RDF_Max minus mean	Turbidity_Mean	0.0112	0.0158	0.0133	0.0055	0.0053	0.0067	0.0123	-0.0088	0.0089
RDF_StDev	none	0.0099	0.0155	0.0146	0.0337	0.0094	0.0169			
RDF_StDev	Colour_mean	0.0220	0.0223	0.0234	0.0311	0.0179	0.0231	-0.0104	-0.0086	-0.0100
RDF_StDev	Conductivity_Mean	0.0020	0.0079	0.0080	0.0552	0.0271	0.0310	0.0000	0.0000	0.0000
RDF_StDev	Turbidity_Mean	0.0080	0.0161	0.0132	0.0342	0.0104	0.0163	0.0146	-0.0069	0.0097

Table C.3 Standard errors and number of data points in regression models for Company M

		Standard Error				Data Points		
Chlorine Value	*Parameter*	*2006*	*2007*	*2008*	*Average*	*2006*	*2007*	*2008*
RDT_Mean	none	0.968	1.13	0.944	1.014	259	250	252
RDT_Mean	Colour_mean	0.929	1.08	0.881	0.963	238	238	239
RDT_Mean	Conductivity_Mean	0.876	1	0.867	0.914	216	217	215
RDT_Mean	Turbidity_Mean	0.976	1.14	0.96	1.025	238	238	239
RDT_Max minus mean	none	0.942	1.12	0.937	1.000	259	250	252
RDT_Max minus mean	Colour_mean	0.89	1.08	0.865	0.945	238	238	239
RDT_Max minus mean	Conductivity_Mean	0.884	1	0.87	0.918	216	217	215
RDT_Max minus mean	Turbidity_Mean	0.953	1.13	0.953	1.012	238	238	239
RDT_StDev	none	0.944	1.13	0.939	1.004	259	250	252
RDT_StDev	Colour_mean	0.901	1.08	0.868	0.950	238	238	239
RDT_StDev	Conductivity_Mean	0.845	1.01	0.863	0.906	216	217	215
RDT_StDev	Turbidity_Mean	0.953	1.14	0.956	1.016	238	238	239
RDF_Mean	none	0.971	1.12	0.944	1.012	259	259	252
RDF_Mean	Colour_mean	0.93	1.08	0.88	0.963	238	238	239
RDF_Mean	Conductivity_Mean	0.877	0.997	0.864	0.913	216	217	215
RDF_Mean	Turbidity_Mean	0.979	1.14	0.96	1.026	238	238	239
RDF_Max minus mean	none	0.945	1.11	0.936	0.997	259	259	252
RDF_Max minus mean	Colour_mean	0.889	1.08	0.869	0.946	238	238	239
RDF_Max minus mean	Conductivity_Mean	0.883	1	0.866	0.916	216	217	215
RDF_Max minus mean	Turbidity_Mean	0.955	1.13	0.952	1.012	238	238	239
RDF_StDev	none	0.953	1.12	0.94	1.004	259	259	252
RDF_StDev	Colour_mean	0.909	1.08	0.87	0.953	238	238	239
RDF_StDev	Conductivity_Mean	0.85	1.01	0.859	0.906	216	217	215
RDF_StDev	Turbidity_Mean	0.96	1.14	0.956	1.019	238	238	239

APPENDIX D RESULT TABLES FOR REGRESSION ANALYSIS FOR COMPANY Z

The results of all regression analyses for Company Z are shown in the following tables. Each line of the table is an individual model.

Table D.1 shows the significance levels for each individual parameter in the regressions. A value of less than 0.05 indicates that the parameter is significant at the 95% confidence level. The significant parameters have been highlighted in the table. A model is considered significant if all of the parameters in the model are significant. The first columns indicate the composition of the model (e.g. RDT_Mean and Colour_mean included as parameters) while the second set of columns (those split by parameter and year) show the significance level for each parameter in each of the three years.

Table D.2 shows the coefficient values for each individual parameter in the regressions. A negative coefficient indicates that the frequency of contacts decreases as the value of the parameter increases. A positive coefficient indicates that the frequency of contacts increases as the value of the parameter increases.

Table D.3 shows the standard errors of the fitted model and the number of data points (zones) included in each model.

Table D.1 Significance levels of regression models for Company Z

		√(population)			RDT/F * √(population)			Parameter * √(population)		
Chlorine Value	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
RDT_Mean	none	<.001	<.001	<.001	0.149	0.996	0.21			
RDT_Mean	Colour_mean	0.001	<.001	<.001	0.114	0.919	0.401	0.138	0.614	0.002
RDT_Mean	Conductivity_Mean	0.34	0.007	0.632	0.193	0.936	0.195	0.904	0.489	0.012
RDT_Mean	Turbidity_Mean	0.006	<.001	<.001	0.123	0.943	0.14	0.292	0.79	0.88
RDT_Max minus mean	none	<.001	<.001	<.001	0.333	0.975	0.288			
RDT_Max minus mean	Colour_mean	<.001	<.001	<.001	0.54	0.755	0.822	0.233	0.58	0.001
RDT_Max minus mean	Conductivity_Mean	0.02	0.012	0.468	0.272	0.756	0.924	0.657	0.465	0.012
RDT_Max minus mean	Turbidity_Mean	<.001	<.001	<.001	0.569	0.872	0.568	0.527	0.813	0.881
RDT_StDev	none	<.001	<.001	<.001	0.773	0.636	0.785			
RDT_StDev	Colour_mean	<.001	<.001	<.001	0.521	0.519	0.559	0.125	0.498	<.001
RDT_StDev	Conductivity_Mean	0.083	0.025	0.698	0.938	0.56	0.442	0.833	0.434	0.007
RDT_StDev	Turbidity_Mean	<.001	<.001	<.001	0.463	0.687	0.679	0.22	0.873	0.785
RDF_Mean	none	<.001	<.001	<.001	0.087	0.822	0.15			
RDF_Mean	Colour_mean	0.003	<.001	<.001	0.076	0.77	0.473	0.209	0.637	0.003
RDF_Mean	Conductivity_Mean	0.315	0.006	0.505	0.101	0.783	0.202	0.917	0.506	0.017
RDF_Mean	Turbidity_Mean	0.014	<.001	<.001	0.076	0.718	0.098	0.435	0.765	0.724
RDF_Max minus mean	none	<.001	<.001	<.001	0.785	0.995	0.251			
RDF_Max minus mean	Colour_mean	<.001	<.001	<.001	0.932	0.652	0.564	0.166	0.577	0.001
RDF_Max minus mean	Conductivity_Mean	0.039	0.016	0.312	0.552	0.636	0.697	0.69	0.454	0.011
RDF_Max minus mean	Turbidity_Mean	<.001	<.001	<.001	0.997	0.709	0.501	0.333	0.792	0.849
RDF_StDev	none	<.001	<.001	<.001	0.371	0.687	0.581			
RDF_StDev	Colour_mean	<.001	<.001	<.001	0.291	0.672	0.85	0.122	0.568	0.001
RDF_StDev	Conductivity_Mean	0.133	0.017	0.519	0.509	0.704	0.568	0.917	0.472	0.01
RDF_StDev	Turbidity_Mean	<.001	<.001	<.001	0.255	0.78	0.515	0.206	0.832	0.761

Table D.2 Coefficients for parameters in regression models for Company Z

		√(population)			RDT/F * √(population)			Parameter * √(population)		
Chlorine Value	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
RDT_Mean	none	0.0087	0.0117	0.0114	0.0110	0.0000	0.0065			
RDT_Mean	Colour_mean	0.0125	0.0123	0.0186	0.0133	0.0004	0.0043	-0.0039	-0.0010	-0.0062
RDT_Mean	Conductivity_Mean	0.0072	0.0094	0.0018	0.0139	0.0003	0.0067	0.0000	0.0000	0.0000
RDT_Mean	Turbidity_Mean	0.0118	0.0112	0.0103	0.0131	0.0003	0.0082	-0.0193	0.0019	0.0016
RDT_Max minus mean	none	0.0130	0.0117	0.0149	-0.0043	0.0001	-0.0058			
RDT_Max minus mean	Colour_mean	0.0165	0.0122	0.0206	-0.0032	0.0012	-0.0012	-0.0033	-0.0011	-0.0066
RDT_Max minus mean	Conductivity_Mean	0.0162	0.0090	0.0032	-0.0063	0.0012	0.0006	0.0000	0.0000	0.0000
RDT_Max minus mean	Turbidity_Mean	0.0153	0.0112	0.0139	-0.0033	0.0006	-0.0038	-0.0134	0.0017	0.0018
RDT_StDev	none	0.0112	0.0111	0.0129	0.0054	0.0070	0.0054			
RDT_StDev	Colour_mean	0.0152	0.0119	0.0194	0.0136	0.0103	0.0110	-0.0042	-0.0014	-0.0067
RDT_StDev	Conductivity_Mean	0.0127	0.0084	0.0017	0.0020	0.0090	0.0153	0.0000	0.0000	0.0000
RDT_StDev	Turbidity_Mean	0.0148	0.0109	0.0128	0.0164	0.0063	0.0092	-0.0246	0.0011	-0.0031
RDF_Mean	none	0.0086	0.0115	0.0114	0.0148	0.0011	0.0078			
RDF_Mean	Colour_mean	0.0118	0.0121	0.0188	0.0171	0.0014	0.0039	-0.0033	-0.0009	-0.0062
RDF_Mean	Conductivity_Mean	0.0071	0.0092	0.0025	0.0191	0.0013	0.0070	0.0000	0.0000	0.0000
RDF_Mean	Turbidity_Mean	0.0108	0.0109	0.0100	0.0174	0.0017	0.0099	-0.0143	0.0021	0.0038
RDF_Max minus mean	none	0.0121	0.0118	0.0154	-0.0015	0.0000	-0.0080			
RDF_Max minus mean	Colour_mean	0.0162	0.0120	0.0213	-0.0005	0.0018	-0.0039	-0.0038	-0.0011	-0.0066
RDF_Max minus mean	Conductivity_Mean	0.0152	0.0087	0.0044	-0.0043	0.0019	-0.0028	0.0000	0.0000	0.0000
RDF_Max minus mean	Turbidity_Mean	0.0154	0.0109	0.0143	0.0000	0.0015	-0.0058	-0.0192	0.0018	0.0023
RDF_StDev	none	0.0102	0.0112	0.0125	0.0186	0.0070	0.0104			
RDF_StDev	Colour_mean	0.0144	0.0119	0.0200	0.0244	0.0075	0.0034	-0.0041	-0.0011	-0.0066
RDF_StDev	Conductivity_Mean	0.0107	0.0088	0.0026	0.0190	0.0066	0.0108	0.0000	0.0000	0.0000
RDF_StDev	Turbidity_Mean	0.0140	0.0109	0.0125	0.0270	0.0049	0.0134	-0.0240	0.0015	-0.0033

Table D.3 Standard errors and number of data points in regression models for Company Z

		Standard Error				Data Points		
Chlorine Value	*Parameter*	*2006*	*2007*	*2008*	*Average*	*2006*	*2007*	*2008*
RDT_Mean	none	0.853	0.619	0.742	0.738	83	57	70
RDT_Mean	Colour_mean	0.916	0.63	0.704	0.750	53	52	54
RDT_Mean	Conductivity_Mean	0.974	0.629	0.731	0.778	47	52	53
RDT_Mean	Turbidity_Mean	0.926	0.631	0.771	0.776	53	52	54
RDT_Max minus mean	none	0.859	0.619	0.745	0.741	83	57	70
RDT_Max minus mean	Colour_mean	0.935	0.629	0.708	0.757	53	52	54
RDT_Max minus mean	Conductivity_Mean	0.98	0.628	0.744	0.784	47	52	53
RDT_Max minus mean	Turbidity_Mean	0.945	0.631	0.786	0.787	53	52	54
RDT_StDev	none	0.863	0.618	0.75	0.744	83	57	70
RDT_StDev	Colour_mean	0.935	0.627	0.706	0.756	53	52	54
RDT_StDev	Conductivity_Mean	0.994	0.626	0.74	0.787	47	52	53
RDT_StDev	Turbidity_Mean	0.943	0.63	0.787	0.787	53	52	54
RDF_Mean	none	0.848	0.633	0.739	0.740	83	65	70
RDF_Mean	Colour_mean	0.91	0.63	0.705	0.748	53	52	54
RDF_Mean	Conductivity_Mean	0.963	0.628	0.732	0.774	47	52	53
RDF_Mean	Turbidity_Mean	0.919	0.63	0.767	0.772	53	52	54
RDF_Max minus mean	none	0.863	0.633	0.743	0.746	83	65	70
RDF_Max minus mean	Colour_mean	0.939	0.629	0.706	0.758	53	52	54
RDF_Max minus mean	Conductivity_Mean	0.99	0.627	0.743	0.787	47	52	53
RDF_Max minus mean	Turbidity_Mean	0.948	0.63	0.785	0.788	53	52	54
RDF_StDev	none	0.859	0.632	0.749	0.747	83	65	70
RDF_StDev	Colour_mean	0.928	0.629	0.708	0.755	53	52	54
RDF_StDev	Conductivity_Mean	0.989	0.628	0.742	0.786	47	52	53
RDF_StDev	Turbidity_Mean	0.936	0.631	0.785	0.784	53	52	54

APPENDIX E RESULT TABLES FOR REGRESSION ANALYSIS FOR COMPANY AA

The results of all regression analyses for Company AA are shown in the following tables. Each line of the table is an individual model. There were not enough zones with the required number of test results (12) for conductivity to include this parameter in the models.

Table E.1 shows the significance levels for each individual parameter in the regressions. A value of less than 0.05 indicates that the parameter is significant at the 95% confidence level. The significant parameters have been highlighted in the table. A model is considered significant if all of the parameters in the model are significant. The first columns indicate the composition of the model (e.g. Chloraminated zones with RDT_Mean and Colour_mean included as parameters) while the second set of columns (those split by parameter and year) show the significance level for each parameter in each of the three years.

Table E.2 shows the coefficient values for each individual parameter in the regressions. A negative coefficient indicates that the frequency of contacts decreases as the value of the parameter increases. A positive coefficient indicates that the frequency of contacts increases as the value of the parameter increases.

Table E.3 shows the standard errors of the fitted model and the number of data points (zones) included in each model.

Table E.1 Significance levels of regression models for Company AA

			√(population)			RDT/F * √(population)			Parameter * √(population)		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
N	RDT_Mean	none	<.001	<.001	<.001	0.906	0.554	0.864			
N	RDT_Mean	Colour_mean	<.001	<.001	<.001	0.557	0.972	0.747	0.895	0.174	0.243
N	RDT_Mean	Turbidity_Mean	<.001	<.001	<.001	0.491	0.604	0.849	0.565	0.04	0.37
N	RDT_Max minus mean	none	<.001	<.001	<.001	0.312	0.817	0.414			
N	RDT_Max minus mean	Colour_mean	<.001	<.001	<.001	0.402	0.474	0.32	0.682	0.12	0.327
N	RDT_Max minus mean	Turbidity_Mean	<.001	<.001	<.001	0.59	0.756	0.339	0.771	0.046	0.536
N	RDT_StDev	none	<.001	<.001	<.001	0.693	0.925	0.528			
N	RDT_StDev	Colour_mean	<.001	<.001	<.001	0.781	0.411	0.616	0.833	0.109	0.321
N	RDT_StDev	Turbidity_Mean	<.001	<.001	<.001	0.961	0.974	0.593	0.635	0.048	0.495
Y	RDT_Mean	none	0.054	0.296	0.01	0.786	0.969	0.548			
Y	RDT_Mean	Colour_mean	0.74	0.011	<.001	0.799	0.547	0.671	0.272	0.019	0.01
Y	RDT_Mean	Turbidity_Mean	0.154	0.028	0.069	0.837	0.714	0.667	0.856	0.041	0.622
Y	RDT_Max minus mean	none	0.309	0.978	0.008	0.095	0.409	0.744			
Y	RDT_Max minus mean	Colour_mean	0.339	0.073	<.001	0.03	0.605	0.89	0.109	0.03	0.011
Y	RDT_Max minus mean	Turbidity_Mean	0.332	0.284	0.035	0.063	0.367	0.523	0.693	0.042	0.497
Y	RDT_StDev	none	0.935	0.677	0.216	0.093	0.22	0.363			
Y	RDT_StDev	Colour_mean	0.283	0.105	0.007	0.055	0.093	0.176	0.176	0.011	0.006
Y	RDT_StDev	Turbidity_Mean	0.947	0.579	0.171	0.08	0.11	0.227	0.8	0.023	0.395

Table E.2 Coefficients for parameters in regression models for Company AA

			√(population)			RDT/F * √(population)			Parameter * √(population)		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*	*2006*	*2007*	*2008*
N	RDT_Mean	none	0.0219	0.0217	0.0231	0.0009	-0.0043	0.0013			
N	RDT_Mean	Colour_mean	0.0199	0.0254	0.0212	0.0048	0.0003	-0.0027	0.0011	-0.0119	0.0059
N	RDT_Mean	Turbidity_Mean	0.0212	0.0196	0.0215	0.0056	-0.0043	-0.0016	-0.0102	0.0140	0.0163
N	RDT_Max minus mean	none	0.0234	0.0205	0.0225	-0.0038	-0.0011	0.0038			
N	RDT_Max minus mean	Colour_mean	0.0216	0.0251	0.0194	-0.0037	0.0040	0.0049	0.0035	-0.0137	0.0049
N	RDT_Max minus mean	Turbidity_Mean	0.0233	0.0177	0.0203	-0.0024	0.0017	0.0048	-0.0055	0.0135	0.0112
N	RDT_StDev	none	0.0228	0.0201	0.0225	-0.0058	0.0015	0.0099			
N	RDT_StDev	Colour_mean	0.0216	0.0248	0.0198	-0.0045	0.0160	0.0085	0.0018	-0.0143	0.0050
N	RDT_StDev	Turbidity_Mean	0.0230	0.0183	0.0205	-0.0008	-0.0006	0.0092	-0.0088	0.0136	0.0126
Y	RDT_Mean	none	0.0070	0.0092	0.0082	0.0019	0.0007	0.0034			
Y	RDT_Mean	Colour_mean	0.0020	0.0401	0.0200	0.0017	0.0102	0.0022	0.0055	-0.0303	-0.0099
Y	RDT_Mean	Turbidity_Mean	0.0067	0.0275	0.0106	0.0014	-0.0062	0.0025	0.0029	-0.0922	-0.0122
Y	RDT_Max minus mean	none	0.0031	0.0003	0.0089	0.0111	0.0226	0.0026			
Y	RDT_Max minus mean	Colour_mean	-0.0055	0.0360	0.0207	0.0142	0.0139	0.0010	0.0078	-0.0274	-0.0099
Y	RDT_Max minus mean	Turbidity_Mean	0.0032	0.0141	0.0106	0.0127	0.0242	0.0051	-0.0061	-0.0895	-0.0164
Y	RDT_StDev	none	-0.0004	-0.0050	0.0058	0.0444	0.0801	0.0216			
Y	RDT_StDev	Colour_mean	-0.0074	0.0274	0.0159	0.0492	0.1063	0.0294	0.0065	-0.0316	-0.0103
Y	RDT_StDev	Turbidity_Mean	-0.0003	0.0070	0.0077	0.0467	0.1027	0.0293	-0.0039	-0.0994	-0.0206

Table E.3 Standard errors and number of data-points in regression models for Company AA

			Standard Error				Data Points		
Chloramination	*Chlorine Value*	*Parameter*	*2006*	*2007*	*2008*	*Average*	*2006*	*2007*	*2008*
N	RDT_Mean	none	0.967	0.939	0.939	0.948	117	117	115
N	RDT_Mean	Colour_mean	1.02	1.02	0.952	0.997	83	84	78
N	RDT_Mean	Turbidity_Mean	1.03	1.01	0.96	1.000	84	85	82
N	RDT_Max minus mean	none	0.963	0.94	0.936	0.946	117	117	115
N	RDT_Max minus mean	Colour_mean	1.02	1.02	0.946	0.995	83	84	78
N	RDT_Max minus mean	Turbidity_Mean	1.03	1.01	0.954	0.998	84	85	82
N	RDT_StDev	none	0.966	0.94	0.937	0.948	117	117	115
N	RDT_StDev	Colour_mean	1.03	1.02	0.951	1.000	83	84	78
N	RDT_StDev	Turbidity_Mean	1.03	1.01	0.958	0.999	84	85	82
Y	RDT_Mean	none	1.11	2.92	1.07	1.700	42	42	43
Y	RDT_Mean	Colour_mean	1.08	2.8	0.978	1.619	40	40	41
Y	RDT_Mean	Turbidity_Mean	1.1	2.85	1.07	1.673	40	40	41
Y	RDT_Max minus mean	none	1.08	2.89	1.07	1.680	42	42	43
Y	RDT_Max minus mean	Colour_mean	1.01	2.8	0.98	1.597	40	40	41
Y	RDT_Max minus mean	Turbidity_Mean	1.05	2.82	1.06	1.643	40	40	41
Y	RDT_StDev	none	1.08	2.86	1.06	1.667	42	42	43
Y	RDT_StDev	Colour_mean	1.03	2.71	0.957	1.566	40	40	41
Y	RDT_StDev	Turbidity_Mean	1.05	2.76	1.05	1.620	40	40	41