Discussion of some topics in Dr.Zhang's Papers

Dr.Zhang's paper "Chromium Contamination in the city of JinZhou" summarized the whole chromium contamination event. In the other paper ,"Study of the Effect of Environmental Pollution in JinZhou Area on Residents Health", he summarized the mortality studies following the chromium contamination. In this reading report, we try to conduct our own study based on the information offered in these two papers. This effort may help us in identifying the direction of further corporation with Dr.Zhang, and also offers possible topics we can include in the final critic of Dr.Zhang;s study.

I: The relation between the mortality studies and the chromium contamination:

About this relation Dr.Butler has the following point of view. Because the chromium, as an initiator and a promotor of cancer, may need more than 10 years to show its effect in higher cancer rate, the mortality studies conducted only 5-13 years after the chromium contamination may not fully reflect the chromium effect on human health. Given that the alloy company was built in 1940's, and it is reasonable to believe that the environmental pollution by the alloy company began as early as 1940's. The alloy company had regular chromium product after 1965, before that the main pollutants emitted from the alloy company was not chromium. The pollution before 1965 may attribute to the higher death and cancer rate in the mortality studies conducted in 1970-1978. Those pollution would follow the same contamination pattern as the chromium contamination, which is: sewage from alloy plant contaminated underground water first and then due to the movement of the underground water and Nuer River the contamination expanded to the long and narrow region along Nuer River. So that , the mortality study in those villages along Nuer River may not only reflect some effect of chromium contamination, it may reflect some effect of other pollution from the alloy plant before 1965. To validate this argument, we need the level of contamination from alloy company before 1965. We hope we can get some information from Dr.Zhang.

II: The dose responds:

The following paragraph is from the Memo we send to Brent on July 19. The dose response pattern in the chromium contamination area is: the further the location of a village to the alloy plant, the higher lung cancer and stomach cancer death rate. On the surface, this dose pattern does not support the conclusion that the higher cancer death rate was caused by environmental chromium contamination. We thought that one possible plaintiff's argument is that: employees of the alloy company lived around the location of the alloy company. Generally speaking plant employees have higher socio-economic status ,and then are expected to have lower cancer rate than farmers; failure to adjust by occupation can distort the dose-response analysis. Dr.Zhang had the occupation information on the original survey form. When Dr.Zhang collected the information of occupation, Dr.Zhang found that most people (by most, he means more than 95%) in his study were farmers, so he stopped collecting the occupation data. Dr.Zhang agreed with me that in TangHeZi (alloy company location; region I on the map), there might be a slightly higher percentage of the population who were not farmers. Based on the information Dr.Zhang provided us, my conclusion is: farming is the main occupation in most villages in Nuer River Region except TangHeZi.

The question we have now is: if occupation can not be an explanation, then we need an explanation for this pattern. Because if we do not have an reasonable explanation, then our own study may be biased by the same reason. In Dr.Zhang's second paper, on page 7, he wrote:

" ...However the fact that the highest malignant neoplasm death rate appeared in the middle area(not the center) and the fact that stomach cancer death rate increased from the center towards the northern direction revealed that each region located north to the center (down the wind direction) had its specific geological condition."

The question is that: if the geological condition for each village has so much influence on the death rate so that we can not observe the true dose response relation , then statistical study with village as the study unit may not be very informative. Our study based on those villages in the contamination area may show a negative dose response relation. However from Dr.Zhang's map in his second paper, we know that generally the cancer death rate in the contaminated area is higher than non-contaminated area. A reasonable study is to use villages both in contaminated area and uncontaminated area. The death rate information of the uncontaminated area is needed to do such a study. We will discuss this question with Dr.Zhang in the next few phone calls.

III: Measure of Dose and Responds:

In order to investigate the relationship between chromium contamination and human health, the first question is how to measure the effect on human health and how to measure the contamination level.

For the human health problem, we now have the death rate and cancer death rate in Dr.Zhang's second paper. In the phone conversation on July 20, Dr.Zhang told me that he has the death rates for other villages in Nuer River region and in West Suburb. These information is useful for a non-biased statistical model. However, the death rate should be adjusted by age group and/or gender. Because the population information is no longer in existence, we are not likely to get the adjusted death rate except those has been offered in Dr.Zhang's paper. Potential bias caused by unadjustment may be significant. Dr.Zhang also has another measure of the effect of chromium contamination on human heath, which is the average age of cancer incidence. I agree that the average age can reflect part of the effect of chromium on human health. However, it is reasonable to believe that if the chromium contamination can shorten the average age of cancer incidents, it will increase the age adjusted cancer death rate. In another word, if we can not find any evidence that chromium can cause cancer by modeling the adjusted death rate, we are not likely to get a very informative result from modeling the average age of cancer incidents. So I will use the adjusted death rate as outcome for this study (if adjusted death rate is not available, using raw death rate).

For the measure of chromium contamination, we have the following possible candidates: (a) chromium concentration in underground water in each village. The advantage of using this measure is that we have more information on this measure in Dr.Zhang's first paper.Disadvantage is that contamination of the underground water is just a part of the human exposure. In Dr.Zhang's paper:

"At the beginning of 1965, 75 water wells (28.2%) contained Cr^{+6} among 266 water wells in

Nuer village and JinChangBao ....."
"Although the residents began using tap water in 1970's, they contacted a large amount of metal chemicals daily......"

Based on those statements in Dr.Zhang 's paper, we can get the following conclusion:

(1) Not all of the water wells have been contaminated by the same time.

(2) Chromium exposure are not only from drinking water from chromium contaminated water well. After using tap water, chromium exposure are no longer came from drinking water.

(3) Other chromium contraction is a non trivial components of the whole exposure.

(b) Can we use other chromium contamination measure? The answer is unlikely. We do not have enough data of air monitoring and soil monitoring.

(c) Distance as a overall measure. Using the distance from the alloy plant as a surrogate measure of total exposure seems to be reasonable. The advantage and rational of using the distance is that:

(1) In Dr.Zhang's paper: *"Because the main source of contamination was the stack of ore residue, the Cr^{+6} concentration of the water well had a linear correlation with the distance between the water well and the stack..."*. Using distance from the alloy company , where the ore residual was, as a measure of chromium concentration is reasonable.

(2) By using the distance form the alloy company, we have taken into count the range of the contamination. Because the nearer a village is to the alloy company, the higher percentage of water wells is of earlier contamination.

(3) By using the distance from the alloy company, we also surrogate measure the air contamination and food exposure, which is not available otherwise.

The disadvantage of using the distance is that: not all the village with the same distance from the alloy company have the same chromium contamination level, because of the special geological pattern and contamination pattern. For example, the contamination area is a long and narrow region along with Nuer River. A villages located along the Nuer River has a higher chromium concentration compared to a village not located near the river, even if these two villages have the same distance form the alloy company. But because we are just using including the villages near Nuer River now , the distance seems to be a reasonable measure.

Dr.Zhang mentioned he has the contamination data for each village each year, which may be a useful measure for the chromium contamination.

IV: Model of the death rate:

The number of death can be modeled by a Poisson regression. However, we only have death rate in Dr.Zhang's paper, which means we will need total population in each village to get the number of death. Dr.Zhang told us in we first conversation that he thought he can find those data for us. We will use those information after we obtain them from Dr.Zhang. Now we can assume that each village has 15000 residence. The number of 15000 comes from Dr.'s second paper. He stated that TangHeZi is a small town with 15000 residents. The model is :

$$\log(NumberofDeath) = \log(Population) + \beta_0 + \beta_1 exposure$$

The following is an example of doing this Poisson regression using data in Dr.Zhang's paper. I have three models in this data analysis:

Model 1: Using distance from the alloy company as the measure of exposure.
Using Adjusted Neoplasm Death Rate as the measure of the chromium contamination effect on human health.
Using the villages in the contaminated area along Nuer River.

Model2: Same as Model 1 except that we have one more self generated observation in the data set. I generated a new observation assuming that the villages located outside the contaminated area has total population 15000 and with 0 chromium contamination.

Model3: Same as model 2 except assuming population is 50000.

Here is a summary table:

	Coef. of Distance	P-value
Model 1	0.052	0.0321
Model 2	-0.0225	0.0745
Model 3	-0.0380	0.0000

The coefficients change sign form positive in model 1to negative in model 2 and model 3. In fact , the pattern of this contamination is that the cancer death rate in the contaminated area is higher than the non-contaminated area, while in the contaminated area, the furthur a village is from the alloy plant the higher the cancer death rate. If we only use those villages inside the contaminated area we may introduce some bias, such as the higher level catamination can benifit human health. On the other hand, p-value shows more significant if we assume the larger population in non-contaminated area in model 3. In order to get the right information from the Poisson regression model, we need the death rate and population infomation of other villages located outside the contaminated area. Dr.Zhang thought he had those infomation.

Positive = X_1 = 0 = West Suburb
1 = ~~Contaminated Area~~ (Villages Along Nuer River I - IX

Positive X_2 = { 0 West Suburb
Distance From alloy plant = ~~Conta~~ Villages along Nuer River I - IX

→ Negative = X_2 = { 0 West Suburb
Cr^{6+} for either 1965 or 1979 = Villages along Nuer I - IX

- not consistent w/ Cr^{6+} being a cause of cancer

TY-0131

OBS	DISTANCE	EXPOSURE	RATE	ADRATE	NADNUM	POPNUM	LN
1	Tanghezi	.	83.23	71.32	96	15000	9.61581
2	JinChangBao	5.300	86.78	83.62	112	15000	9.61581
3	Nuer River Village	5.300	72.43	71.89	97	15000	9.61581
4	YangXing	2.500	81.27	76.80	103	15000	9.61581
5	Shilitai	0.025	99.30	92.96	125	15000	9.61581
6	WenJiaTun	0.025	96.87	91.12	123	15000	9.61581
7	Bajiazi	0.025	.	.	.	15000	9.61581

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.B
Distribution	POISSON
Link Function	LOG
Dependent Variable	NADNUM
Offset Variable	LN
Observations Used	6
Missing Values	1

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4	2.7852	0.6963
Scaled Deviance	4	2.7852	0.6963

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Pearson Chi-Square	4	2.7919	0.6980
Scaled Pearson X2	4	2.7919	0.6980
Log Likelihood	.	2425.8187	.

The GENMOD Procedure

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	-5.0478	0.0721	4903.5636	0.0000
DISTANCE	1	0.0512	0.0239	4.5918	0.0321
SCALE	0	1.0000	0.0000	.	.

OBS	DISTANCE	EXPOSURE	RATE	ADRATE	NADNUM	POPNUM	LN
1	Tanghezi	.	83.23	71.32	96	15000	9.6158
2	JinChangBao	5.000	86.78	83.62	112	15000	9.6158
3	Nuer River Village	5.000	72.43	71.89	97	15000	9.6158
4	YangXing	2.500	81.27	76.80	103	15000	9.6158
5	Shilitai	0.025	99.30	92.96	125	15000	9.6158
6	WenJiaTun	0.025	96.87	91.12	123	15000	9.6158
7	Bajiazi	0.025	.	.	.	15000	9.6158
8	Not in contamination area	0.000	52.27	54.33	244	50000	10.8198

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.B
Distribution	POISSON
Link Function	LOG
Dependent Variable	NADNUM
Offset Variable	LN
Observations Used	7
Missing Values	1

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	5	18.8723	3.7745
Scaled Deviance	5	18.8723	3.7745

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Pearson Chi-Square	5	19.5817	3.9163
Scaled Pearson X2	5	19.5817	3.9163
Log Likelihood	.	3515.0841	.

The GENMOD Procedure

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	-4.8633	0.0524	8619.2923	0.0000
DISTANCE	1	-0.0380	0.0089	18.3825	0.0000
SCALE	0	1.0000	0.0000	.	.

NOTE: The scale parameter was held fixed.

OBS	DISTANCE	EXPOSURE	RATE	ADRATE	NADNUM	POPNUM	LN
1	Tanghezi	.	83.23	71.32	96	15000	9.61581
2	JinChangBao	5.000	86.78	83.62	112	15000	9.61581
3	Nuer River Village	5.000	72.43	71.89	97	15000	9.61581
4	YangXing	2.500	81.27	76.80	103	15000	9.61581
5	Shilitai	0.025	99.30	92.96	125	15000	9.61581
6	WenJiaTun	0.025	96.87	91.12	123	15000	9.61581
7	Bajiazi	0.025	.	.	.	15000	9.61581
8	Not in contamination area	0.000	52.27	54.33	73	15000	9.61581

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.B
Distribution	POISSON
Link Function	LOG
Dependent Variable	NADNUM
Offset Variable	LN
Observations Used	7
Missing Values	1

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	5	16.0401	3.2080
Scaled Deviance	5	16.0401	3.2080

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Pearson Chi-Square	5	16.2607	3.2521
Scaled Pearson X2	5	16.2607	3.2521
Log Likelihood	.	2659.3947	.

The GENMOD Procedure

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	-4.8940	0.0555	7769.3686	0.0000
DISTANCE	1	-0.0225	0.0126	3.1809	0.0745