Review of Statistical Modeling Methods, Analysis, and Interpretation

University of Michigan Dioxin Exposure Study

March 30, 2009
Introduction

• The UMDES is a very large study with primary objective to

“identify factors that explained variation in serum dioxin concentrations among residents in Midland and Saginaw Counties.”

• Complex sampling and analysis methods
• Confidentiality renders peer review difficult
• The science advisory board (SAB) has not included a PhD statistician since 2006
• As a result MDEQ requested a review by professional statisticians with national experience at large contaminated sediment mega sites
• Desired outcome is a collaborative technical process to develop results applicable to risk management decisions
Objectives of the Review

• Evaluate experimental design and statistical methods to aid MDEQ to:
  – Insure understanding of study conclusions and their strengths and limitations
  – Evaluate the utility and applicability of the UMDES data for risk management decisions
  – If appropriate, determine if modifications to analyses are necessary to improve applicability to risk management decisions
  – Insure that results and interpretations are properly and accurately stated to the public
Presentation Overview

• Summary of primary findings
• Brief discussion of risk assessment components
• Catalog of experimental designs and their strengths and limitations
• Nature of the UMDES design
• Discussion of statistical methods appropriate to UMDES
• Findings
• Recommendations
Primary Findings

• Data are not publicly available beyond UM research team
• Study design is observational which limits the potential to make causal inference
• Statistical modeling—Variable selection by significance tests and stepwise procedures lead to unreliable models (Harrell, 1996)
• Sampling design and selection of subjects may under represent critical target populations
Typical Application of Human Health Risk Assessments for Remedial Decisions

• Michigan DEQ
  – Develop generic cleanup criteria
  – Determine need for and develop site-specific cleanup criteria

• U.S. EPA CERCLA/RCRA Programs
  – Baseline HHRA to evaluate need for remediation/corrective action
  – Use for developing preliminary and final remediation/corrective action goals
Risk Assessment Overview

• Identify concerns = hazard identification
  – What chemicals and what levels?
  – Where are they?

• Determine potential for contact with contamination = exposure assessment

  \[\text{Exposure } \propto \text{ Intensity } \times \text{ Frequency } \times \text{ Duration}\]

• Potential for health effects from contamination = toxicity assessment
  – How much (dose)?

• Potential risk = risk characterization
  – Combine information on exposure and toxicity to determine risk
Exposure Pathway:

• The route a substance takes from its source (where it began) to its end point (where it ends), and how people can come into contact with (or get exposed to) it. An exposure pathway has five parts:
  – a source of contamination (such as an abandoned business);
  – an environmental media and transport mechanism (such with surface water and sediment);
  – a point of exposure (such as a residential property);
  – a route of exposure (eating, drinking, breathing, or touching),
  – a receptor population (people potentially or actually exposed).

• When all five parts are present, the exposure pathway is termed a completed exposure pathway

Definitions provided by ATSDR Glossary of Terms, http://www.atsdr.cdc.gov/glossary.html; last accessed March 26, 2009
Background

- A central goal of the study is to determine which factors explain variation in serum dioxin congener levels, and to quantify how much variation each factor explains.
- In particular, does living on contaminated soil, or living in a house with contaminated house dust lead to increased serum dioxin levels?
- The study is a human exposure pathway study
  - It is not a study of health outcomes.
  - It is not intended to provide information on the geographic distribution of soil contamination with dioxins, furans and PCBs in Midland and Saginaw Counties or elsewhere.
  - The study does not address potential economic consequences of dioxin contamination or exposures.
• **Bottom Up (Mechanistic)**
  - Mechanistic “models”
  - Measurements in soil, sediment and lower trophic levels
  - Models predict receptor exposures

• **Top Down (Empirical)**
  - Receptor and source concentrations are measured
  - Empirical relationships developed
  - Common in ecological studies
  - Biota to sediment or soil accumulation factors (BSAFs)
Hudson River Fish Exposure Model
A Top Down Example

- 80 foot spacing for sediment samples
- 300 to 500 fish per species
- Collocated fish and sediment samples at multiple scales
- Biological parameters explain majority of variance
- Adjusted R-squared values are generally low
- Sediment explains less than 10% of variation

\[ \log(C_{fish}) = \beta_0 + \beta_1 \log(Lipid) + \beta_2 \log(Length) \\
+ \beta_3 \log(TOC) + \beta_4 \log(C_{sediment}) \]

Regression model is identical in **form** to the UMDES regression models

![Percent Total PCB Variation in Fish Tissue Explained by Sediment Model](image)
EXPERIMENTAL DESIGN
Specification of Research Questions

• Stepwise variable selection implicitly creates many research questions (thousands of them)
• Important research questions should be specified \textit{a priori} and tested by careful specification of individual models
• Results should be provided in such a way that competing hypotheses can be ranked
Research vs. Risk Management

- Research conducted according to “the scientific method” is an iterative process consisting of:
  - A priori formulation of research questions
  - Study design and sample selection
  - Careful and detailed statistical analyses
  - Formulation of new research questions and insights

- Risk management is a process of integration of diverse sources of information for selection among remedial alternatives
  - unlike academic research findings, remedial selection is often not reversible

- This distinction influences how users of the UMDES must interpret study results
  - Risk managers have fewer iterative cycles with which to refine research questions and to answer them, and false positive (negative) interpretations have costly and, at times, immediate consequences
OUR INTERPRETATION OF UMDES DESIGN AND WHERE IT FITS IN
Types of Study Designs

**Observational**
- Hypothesis generating
- Unbalanced sampling
- Correlated explanatory variables
- Data reduction
- Confirmatory studies needed to verify results
- Arbitrary partitioning of $R^2$

**Designed Experiment**
- Hypothesis testing
- Research questions fully formed
- Independence of variables assured through random assignment of subjects to treatments
- Balanced representation of study
- Unique partitioning of $R^2$

---

**UMDES**

- **Exploratory Observation**
  - Many explanatory variables; data reduction methods are used

- **Confirmatory Observation**
  - Focus is on a set of “primary variables” with *a priori* hypotheses; often a follow-up study

- **Controlled Experiment with Supplemental Variables**
  - Can infer cause and effect; can rank relative importance of explanatory variables
Observational Studies

- In observational studies, treatments are observed, rather than assigned
- It is not reasonable to consider the observed data under different treatments as random samples from a common population
- Systematic differences in populations may exist that affect the response variables
- Designs become unbalanced with respect to treatment combinations
- Controlling for confounding factors is recommended through regression model building
- Model building for causal inference is more difficult than for prediction

Gelman and Hill (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*
Model Building Strategies
Prediction

• Include any variables known a-priori to be important
  – Age, BMI, sex, etc.
• For variables with large effects consider interactions
• Data Reduction:
  – Predictors with interpretable signs can be included regardless of statistical significance
  – Predictors that are non-significant and have the wrong signs should be discarded
  – Predictors that are significant with the wrong signs should be carefully considered and justified with new mechanisms or theories
  – Covariate relationships should be carefully investigated
  – Predictors that are significant with the expected sign are included
• These are recommendations from Gelman and Hill (2007)
• Burnham and Anderson (1998) would follow a similar strategy with the exception that statistical significance would be replaced with information theoretic measures such as the Akaike Information Criterion (Akaike 1974)
• These strategies provide basis for prediction of the response, but not for estimating the effects of manipulating the predictors (i.e. causation)
Three Primary Goals
(stated in the UMDES)

• Evaluate concern that people’s body burdens of dioxins, furans and PCBs are elevated because of environmental contamination
• Determine which factors explain variation in serum congener levels, and to quantify how much variation each factor explains
• Find out whether the elevated levels of dioxins in the soil in the city of Midland, and in the Tittabawassee River flood plain between Midland and Saginaw, have also caused elevated levels of dioxins in residents' bodies
Causal Inference

• The primary goals of the UMDES are best described as causal investigations
• The UMDES is an observational study which limits potential for causal inference
• Careful consideration of balance, overlap, and distribution of the response among covariate combinations is necessary to determine the limits of causal vs. predictive statements
Signs of Trouble

- Nonsensical model results
- Coefficients that change in magnitude and even direction when variables are added or removed from models
- High pairwise correlations among continuous variables
- Significant differences in means of continuous variables among levels of discrete covariates
- Significant multiple regression relationships among predictors
- Large standard errors for regression coefficients
- High variance inflation factors
- Lack of overlap in covariate distributions
- Sample size imbalance among subgroups
- Differences in central tendency and shape of covariate distributions across subgroups
An Example

Objective: Select among two predictors of contaminant concentration in a receptor

• Consider a two variable regression model of the form:

\[ \log_{10}(C_{\text{Receptor}}) = \beta_0 + \beta_1 X_{\text{soil}} + \beta_2 X_{\text{residence}} \]

• \( X_{\text{residence}} \) is a binary indicator

• Forward selection will be used to select the “important” predictor(s)
Soil Only Model (Adjusted $R^2 = 0.68$)

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>660</td>
<td>325</td>
<td>2.03</td>
<td>0.07</td>
<td>-65</td>
</tr>
<tr>
<td>Soil Conc</td>
<td>828</td>
<td>169</td>
<td>4.90</td>
<td>&lt;0.001</td>
<td>451</td>
</tr>
</tbody>
</table>

$y = 828x + 660$

$R^2 = 0.68$
### Forward Stepwise Procedure

**Residence Only Model (Adjusted $R^2 = 0.87$)**

<table>
<thead>
<tr>
<th></th>
<th>Standard Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1527</td>
<td>109</td>
<td>14.01</td>
<td>&lt;0.001</td>
<td>1284</td>
<td>1770</td>
</tr>
<tr>
<td>Residence</td>
<td>1429</td>
<td>169</td>
<td>8.46</td>
<td>&lt;0.001</td>
<td>1053</td>
<td>1805</td>
</tr>
</tbody>
</table>

Start with either of the two variables:

- **Residence**
- **Intercept**

**Analysis of Full Model (Adjusted $R^2 = 0.85$)**

<table>
<thead>
<tr>
<th></th>
<th>Standard Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>Significance Level (P-value)</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1551</td>
<td>334</td>
<td>4.64</td>
<td>0.00</td>
<td>795</td>
<td>2308</td>
</tr>
<tr>
<td>Residence</td>
<td>1457</td>
<td>410</td>
<td>3.55</td>
<td>0.01</td>
<td>530</td>
<td>2384</td>
</tr>
<tr>
<td>Soil Conc</td>
<td>20</td>
<td>265</td>
<td>-0.08</td>
<td><strong>0.94</strong></td>
<td>-619</td>
<td>579</td>
</tr>
</tbody>
</table>

Remove and Try Again

Add variables and test for significance:

- **Residence**
- **Soil Conc**

Negative coefficient
Backward Elimination Procedure: Same results in this instance

**Analysis of Full Model** \( (\text{Adjusted } R^2 = 0.85) \)

<table>
<thead>
<tr>
<th></th>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>Significance Level (P-value)</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1551</td>
<td>334</td>
<td>4.64</td>
<td>0.00</td>
<td>795</td>
<td>2308</td>
</tr>
<tr>
<td>Residence</td>
<td>1457</td>
<td>410</td>
<td>3.55</td>
<td>0.01</td>
<td>530</td>
<td>2384</td>
</tr>
<tr>
<td>Soil Conc</td>
<td>-20</td>
<td>265</td>
<td>-0.08</td>
<td>0.94</td>
<td>-619</td>
<td>579</td>
</tr>
</tbody>
</table>

Negative regression coefficient would go unnoticed in automated procedure

**Residence Only Model** \( (\text{Adjusted } R^2 = 0.87) \)

<table>
<thead>
<tr>
<th></th>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1527</td>
<td>109</td>
<td>14.01</td>
<td>&lt;0.001</td>
<td>1284</td>
<td>1770</td>
</tr>
<tr>
<td>Residence</td>
<td>1429</td>
<td>169</td>
<td>8.46</td>
<td>&lt;0.001</td>
<td>1053</td>
<td>1805</td>
</tr>
</tbody>
</table>
Residence Adjusted Receptor Concentrations vs. Soil Concentration

\[ y = -3.8195x + 2129.2 \]

\[ R^2 = 0.0001 \]
Soil and Place of Residence are Confounded

- The sampling design is unbalanced relative to the predictors
- No overlap in the predictors
- Effects of soil and residence cannot be separated
- Residence may be acting as a surrogate for soil concentrations
- Results should be reported for both variables separately, including confidence intervals and adjusted $R^2$

Receptor vs Soil Concentration

$y = 828x + 660$

$R^2 = 0.68$
Implications of Example

• The effect of soil exposure is conditional on other variables in the model due to confounding
• Variation cannot be partitioned into independent components
• Coefficients cannot be interpreted unconditionally
• The contribution to serum contaminant concentration is a function of other variables in, or out, of the model
• In this example, no conclusion can be drawn regarding importance of soil as opposed to place of residence
• Adjusted $R^2$ is not an indicator of importance of predictors in observational studies because covariates are not independent
• Demond et al. (2008) showed that soil and place of residence are confounded similarly to this example
Selection of Research Subjects - UMDES

• Representation of critical target populations defined by MDEQ (2004) not adequate
  – Critical target populations are those “most likely to have the highest exposures to DLC contamination from Dow”

• Subjects not adequately represented include:
  – Floodplain population
  – High end fish consumers
  – Game consumers
  – Consumers of other animal products associated with the Tittabawassee River, Saginaw River, or Saginaw Bay
  – These critical food chain exposure factors are not necessarily related to the geographically-based study groups identified in the UMDES
Floodplain Example

• Representation of the Floodplain population is not adequate
  – Consists of people who live on or near the 100-year floodplain of the Tittabawassee River

• The portion of the Floodplain population most likely to have elevated body burdens of DLCs live and or use frequently-flooded portions of the Tittabawassee River floodplain (MDEQ 2004)
Definition of the Floodplain Population

1) [Census] blocks in Midland and Saginaw counties which contained any land area in the Federal Emergency Management Administration-defined 100 year flood plain of the Tittabawassee River below the Dow Chemical Company facility in Midland, and above the point where the Tittabawassee and Shiawassee Rivers join and have a mixed flood plain; (Garabrandt 2008a).

Property ownership extends to the river and significant portion of property exceeds 1000ppt
Properties are partially in the floodplain, but residents do not have river access. Most of property is outside the floodplain.

How are these situations differentiated?

Do floodplain exposures represent Reasonable Maximum Exposures?
## UMDES Soil TEQ Summary

<table>
<thead>
<tr>
<th>Soil Composites HP 0-1 Inch</th>
<th>N</th>
<th>Mean</th>
<th>S.E.</th>
<th>Median</th>
<th>75th%ile</th>
<th>95th%ile</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floodplain</td>
<td>203</td>
<td>56.5</td>
<td>9.7</td>
<td>11.4</td>
<td>35.4</td>
<td>223.1</td>
<td>1.1</td>
<td>1881.4</td>
</tr>
<tr>
<td>Near Floodplain</td>
<td>164</td>
<td>52.0</td>
<td>36.7</td>
<td>3.9</td>
<td>10.4</td>
<td>102.9</td>
<td>0.8</td>
<td>2299.8</td>
</tr>
<tr>
<td>Other M/S</td>
<td>168</td>
<td>13.5</td>
<td>2.0</td>
<td>5.3</td>
<td>13.2</td>
<td>59.4</td>
<td>0.8</td>
<td>157.7</td>
</tr>
<tr>
<td>Plume</td>
<td>37</td>
<td>109.2</td>
<td>31.0</td>
<td>58.2</td>
<td>111.9</td>
<td>257.2</td>
<td>6.3</td>
<td>745.5</td>
</tr>
<tr>
<td>Jackson/Calhoun</td>
<td>194</td>
<td>6.9</td>
<td>0.8</td>
<td>3.6</td>
<td>7.6</td>
<td>22.6</td>
<td>0.4</td>
<td>186.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vegetation Composites House Perimeter</th>
<th>N</th>
<th>Mean</th>
<th>S.E.</th>
<th>Median</th>
<th>75th%ile</th>
<th>95th%ile</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floodplain</td>
<td>188</td>
<td>14.2</td>
<td>3.4</td>
<td>3.4</td>
<td>7.4</td>
<td>50.2</td>
<td>0.4</td>
<td>1427.2</td>
</tr>
<tr>
<td>Near Floodplain</td>
<td>69</td>
<td>376.6</td>
<td>354.1</td>
<td>3.3</td>
<td>10.1</td>
<td>152.0</td>
<td>0.6</td>
<td>7994.9</td>
</tr>
<tr>
<td>Other M/S</td>
<td>71</td>
<td>4.2</td>
<td>0.4</td>
<td>3.3</td>
<td>5.1</td>
<td>10.1</td>
<td>1.0</td>
<td>27.5</td>
</tr>
<tr>
<td>Plume</td>
<td>36</td>
<td>37.5</td>
<td>12.7</td>
<td>18.3</td>
<td>31.1</td>
<td>125.4</td>
<td>0.8</td>
<td>268.9</td>
</tr>
<tr>
<td>Jackson/Calhoun</td>
<td>52</td>
<td>4.5</td>
<td>0.6</td>
<td>3.3</td>
<td>6.7</td>
<td>8.7</td>
<td>0.6</td>
<td>25.9</td>
</tr>
</tbody>
</table>
Environmental Health Perspectives Publication (2008):
1. Start with over 100 predictors
2. Variable groupings have many similar variables that are expected to be interrelated
3. Automated selection may obscure confounding


Online 22 December 2008
Variance partitioning results are reported unconditionally in spite of the likely correlations.
Three pages of model coefficients distilled into one primary conclusion

Conclusions: The study provides valuable insights into the relationships between serum dioxins and environmental factors, age, sex, BMI, smoking, and breast feeding. These factors together explain a substantial proportion of the variation in serum dioxin concentrations in the general population. Historic exposures to environmental contamination appeared to be of greater importance than recent exposures for dioxins.

Further Interpretation 2378-TCDD

- Analyses were conducted in $\log_{10}$ so regression coefficients represent ratios of concentration.
- Ratios greater than one indicate positive relationships while those less than one indicate negative relationships

$$C(x) = 10^{\beta_0} \times 10^{\beta_1x}$$

$$\frac{C(1)}{C(0)} = \frac{10^{\beta_0} \times 10^{\beta_1}}{10^{\beta_0}} = 10^{\beta_1}$$

*Percentage Effect* = $\left(10^{\beta_1} - 1\right) \times 100\%$
Estimated Effects Reported in Table 1
(Significant at $\alpha = 0.05$)

<table>
<thead>
<tr>
<th>Percentage Change in Serum 2378-TCDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>-100%</td>
</tr>
<tr>
<td>Age-50*</td>
</tr>
<tr>
<td>BMI loss last 12 months</td>
</tr>
<tr>
<td>Months all children breast fed</td>
</tr>
<tr>
<td>Gender (Female:Male)</td>
</tr>
<tr>
<td>Pack-yrs Smoking</td>
</tr>
<tr>
<td>Race (White vs. Other)</td>
</tr>
<tr>
<td>Gender by Age Interaction</td>
</tr>
<tr>
<td>Lived in Midland/Saginaw in 60-79 (Number of Years)*</td>
</tr>
<tr>
<td>Lived on Property where trash or yard waste was burned in 40-59*</td>
</tr>
<tr>
<td>Worked at Dow in 40-59*</td>
</tr>
<tr>
<td>Served as emergency responder in 40-59*</td>
</tr>
<tr>
<td>Served as emergency responder after 1980*</td>
</tr>
<tr>
<td>Did water activities in Tittabawassee R. After 1980 (&gt;=1 per month vs...</td>
</tr>
<tr>
<td>Number of years ate fish from any source after 1980*</td>
</tr>
<tr>
<td>Ate Other Species Saginaw R. or Bay during the last 5 years</td>
</tr>
<tr>
<td>Hunting Tittabawassee Area in 1960-1979 (&gt;=1 per month vs. never)</td>
</tr>
<tr>
<td>Hunting Tittabawassee Area after 1980 (&gt;=1 per month vs. never)-75%</td>
</tr>
</tbody>
</table>

* Effect size for variable applies per year

Nonsensical Results
Path Forward

• Collaborative work at the technical level
• Development of selected multiple regression models that can be used to quantify relationships between serum and critical variables reliably
• Joint development of materials to communicate mutually supportable results
• Development of materials suitable for the MDEQ to review in order to verify that issues identified herein have been addressed and that results can be relied upon for risk management decisions
Summary of Findings

• Conclusions regarding primary factors influencing serum dioxin and furan (D/F) concentrations are based on data that are over processed and under analyzed (interpreted)
• Automated model selection methods used to process data appear to have resulted in overly fitted models that very likely mask important relationships between serum and environmental D/F concentrations
• Models have apparently not been validated and likely have poor out of sample predictive power
• Partial $R^2$ values are incorrectly interpreted as a means to rank importance of variables with regard to D/F exposure
• Reported results fail to recognize the large proportion of variance apparently explained jointly by the collection of environmental variables
Summary of Findings

• Results are stated unconditionally, when they should be qualified as conditional on the other variables included as well as excluded from the final models

• Reported associations between serum and environmental factors are frequently nonsensical and inconsistent with mechanisms known to influence serum dioxin levels

• It appears that subjects included in the floodplain population are likely to not live in areas with elevated soil D/F concentrations

• Soil (D/F) concentrations in the floodplain are in general one to two orders of magnitude higher than those reported in the UMDES study

• Failure to test important hypotheses separately (i.e. food consumption, region of residence, soil concentration and life history) has likely caused confounding amongst critical variables of interest
Recommendations

• Results and findings for critical variables need to be based on individual models with sound theoretical underpinnings based on understood mechanisms of fate and transport and bio-uptake of D/F and PCBs
• All results should include estimates of effect sizes and standard errors and or confidence intervals
• Results that cannot be rectified with the scientific literature should be obviously identified as such and described
• Statements of results should not be released until these modifications have been undertaken and results can be thoroughly peer reviewed
Recommendations

• The science advisory board SAB should recruit and retain one or more PhD statisticians with experience in risk assessment, superfund and remedial decision making, sample survey methodology and linear models theory
  – Candidates for this position should be nominated by Dow, MDEQ, USEPA, ASTDR, NIH and other interested agencies and stakeholders

• Statistical methods need to be revised and published in an applied statistics journal such as Journal of Applied Biological and Environmental Sciences (JABES), Biometrics, Technometrics, or Journal of the American Statistical Association (JASA)
References


Demond, A. et al. 2008. Statistical comparison of residential soil concentrations of PCDDs, PCDFs, and PCBs from Two Communities in Michigan. *Environmental Science and Technology*.


