Analyzing diastolic and systolic blood pressure individually or jointly?

Chenglin Ye\textsuperscript{a}, Gary Foster\textsuperscript{a}, Lisa Dolovich\textsuperscript{b}, Lehana Thabane\textsuperscript{a,c}

\textsuperscript{a.} Department of Clinical Epidemiology and Biostatistics, McMaster University
\textsuperscript{b.} Department of Family Medicine, McMaster University
\textsuperscript{c.} Biostatistics Unit, St. Joseph’s Healthcare Hamilton

September 14\textsuperscript{th} 2012
OUTLINE

• Background
• Methods
• Results
• Discussion
• Limitation
GLOBAL DISTRIBUTION OF CVD RELATED DEATHS

Age-standardized deaths due to CVD in a day

Source: WHO, 2009
GLOBAL BURDEN OF HYPERTENSION

- 54% of stroke
- 47% of ischemic heart disease
- 13.5% of all deaths

→ are attributable to high blood pressure (Lawes et al, Lancet, 2008)

Leading risks for deaths in North America
THE DEFINITION OF HYPERTENSION

- Debates on how to define hypertension (HTN)

- Historically, HTN defined solely on DBP

- SBP was related to increased cardiovascular risk
  (The Framingham Heart Study, 1969)

- SBP is a superior predictor to DBP
  (Conen D et al. 2009; Benetos A et al. 2002)
Current classification HTN in 3 subtypes:

<table>
<thead>
<tr>
<th>Classification</th>
<th>SBP</th>
<th>DBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>systolic-diastolic hypertension (SDH)</td>
<td>&gt;= 140</td>
<td>&gt;= 90</td>
</tr>
<tr>
<td>isolated-systolic hypertension (ISH)</td>
<td>&gt;= 140</td>
<td>&lt; 90</td>
</tr>
<tr>
<td>isolated-diastolic hypertension (IDH)</td>
<td>&lt; 140</td>
<td>&gt;= 90</td>
</tr>
</tbody>
</table>

*measured in mmHg

(the 6th JNC report, 1997)
We compare two approaches of analyzing patients’ blood pressure change over a period of time:

1. analyzing SBP and DBP individually

2. analyzing SBP and DBP jointly
A COHORT STUDY

- **Design** = prospective cohort with repeated observations
- **Population** = mainly 65+ years old, living in community
- **Intervention** = Cardiovascular Health Awareness program (CHAP)
- **Outcomes** = SBP and DBP
- **Timeframe of study** = May 2008 to April 2010
A community-wide program (referrals, local media) in 2001

Multiple sessions implemented 22+ communities

Participants get:
- Baseline CVD risk assessment
- BP measure (by BpTRU)
- Education for life style and medication use
- Link to resources to specific modifiable risk factors
- Feedback to family physicians
PARTICIPANTS WITH MULTIPLE VISITS

May 2008 – April 2010:
$n = 13596$

Participants with less than 3 visits = 11050
Participants with at least 3 visits = 2546

Longitudinal analysis
$n = 2546$
METHODS

- Two univariate models: outcome = SBP/DBP
- One bivariate model: outcome = SBP & DBP
- Adjusted for baseline covariates
- Log-transforming the outcomes: SBP & DBP
**The univariate model:**

\[ Y_i(t) = (\beta_0 + \mu_i) + (\beta_1 + b_i)t + (\beta_2 x_1 + \beta_3 x_2 + \cdots + \beta_{k-1} x_k) + \varepsilon_i \]

- A random intercept for patient \( i \): \( \beta_0 + \mu_i \)
- A random slope of time variable \( t \) for patient \( i \): \( \beta_1 + b_i \)
- Baseline covariates with fixed-effects: \( \beta_2 x_1 + \beta_3 x_2 + \cdots + \beta_{k-1} x_k \)
- Usual residual term: \( \varepsilon_i \sim N(0, \Sigma) \)
The bivariate model:

\[
\begin{align*}
SBP: \quad Y_{1i}(t) &= (\beta_{10} + \mu_{1i}) + (\beta_{11} + b_{1i})t + (\beta_{12}x_1 + \beta_{13}x_2 + \cdots + \beta_{1(k-1)}x_k) + \varepsilon_{1i} \\
DBP: \quad Y_{2i}(t) &= (\beta_{20} + \mu_{2i}) + (\beta_{21} + b_{2i})t + (\beta_{22}x_1 + \beta_{23}x_2 + \cdots + \beta_{2(k-1)}x_k) + \varepsilon_{2i}
\end{align*}
\]

The random-effect components follow a joint distribution:

\[
\begin{bmatrix}
\mu_{1i} \\
\mu_{2i} \\
b_{1i} \\
b_{2i}
\end{bmatrix}
\sim N(0, \mathbf{D}), \text{ where } \mathbf{D} =
\begin{bmatrix}
\sigma_{\mu_1}^2 & \sigma_{\mu_1,\mu_2} & \sigma_{\mu_1,b_1} & \sigma_{\mu_1,b_2} \\
\sigma_{\mu_2} & \sigma_{\mu_2}^2 & \sigma_{\mu_2,b_1} & \sigma_{\mu_2,b_2} \\
\sigma_{b_1} & \sigma_{b_1} & \sigma_{b_1}^2 & \sigma_{b_1,b_2} \\
\sigma_{b_2} & \sigma_{b_2} & \sigma_{b_2} & \sigma_{b_2}^2
\end{bmatrix}
\]
MODELING IRREGULAR VISITS

- Analyzed the actual visit time as continuous variable
- Used a continuous autoregressive (CAR) covariance structure

Continuous time: \[ \Sigma = \begin{bmatrix}
visit 1 & visit 2 & visit 3 & visit 4 \\
\varepsilon^2 & \phi^{h_{1,2}}\varepsilon^2 & \phi^{h_{1,3}}\varepsilon^2 & \phi^{h_{1,4}}\varepsilon^2 \\
\phi^{h_{1,2}}\varepsilon^2 & \varepsilon^2 & \phi^{h_{2,3}}\varepsilon^2 & \phi^{h_{2,4}}\varepsilon^2 \\
\phi^{h_{1,3}}\varepsilon^2 & \phi^{h_{2,3}}\varepsilon^2 & \varepsilon^2 & \phi^{h_{3,4}}\varepsilon^2 \\
\phi^{h_{1,4}}\varepsilon^2 & \phi^{h_{2,4}}\varepsilon^2 & \phi^{h_{3,4}}\varepsilon^2 & \varepsilon^2 \\
\end{bmatrix} \]

\( h_{i,j} = \text{actual time difference between } i \text{ and } j. \)
An alternative the bivariate model in a repeated measure model (without random effect)

By a Kronecker product of the covariance matrix

However, CAR structure is not available in current software
## BASELINE DEMOGRAPHICS

- **Participants’ demographics**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (standard deviation) or % (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>67 (14)</td>
</tr>
<tr>
<td>BMI (kg/m^2)</td>
<td>28 (5)</td>
</tr>
<tr>
<td>Male</td>
<td>37% (13596)</td>
</tr>
<tr>
<td>Pulse rate (BPM)</td>
<td>72 (13)</td>
</tr>
<tr>
<td>SBP (mmHg)</td>
<td>128 (19)</td>
</tr>
<tr>
<td>DBP (mmHg)</td>
<td>74 (11)</td>
</tr>
<tr>
<td>Presented SDH</td>
<td>6% (13571)</td>
</tr>
<tr>
<td>Presented ISH</td>
<td>17% (13571)</td>
</tr>
<tr>
<td>Presented IDH</td>
<td>3% (13571)</td>
</tr>
</tbody>
</table>
## BASELINE DEMOGRAPHICS

- **Participants' medical history**

<table>
<thead>
<tr>
<th>Variable</th>
<th>% (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previously diagnosed HTN</td>
<td>54% (12280)</td>
</tr>
<tr>
<td>On BP medication</td>
<td>50% (12270)</td>
</tr>
<tr>
<td>Heart attack</td>
<td>9% (12304)</td>
</tr>
<tr>
<td>Stroke</td>
<td>3% (12315)</td>
</tr>
<tr>
<td>Transient ischemic attack (TIA)</td>
<td>7% (12310)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>16% (12299)</td>
</tr>
<tr>
<td>Hi cholesterol</td>
<td>41% (12288)</td>
</tr>
</tbody>
</table>
## BASELINE DEMOGRAPHICS

### Participants’ life style

<table>
<thead>
<tr>
<th>Variable</th>
<th>% (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Often felt stressed</td>
<td>17% (12231)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>12% (12308)</td>
</tr>
<tr>
<td>2+ drinks a day</td>
<td>8% (12278)</td>
</tr>
<tr>
<td>3+ times high fat food in a week</td>
<td>10% (12299)</td>
</tr>
<tr>
<td>5+ serving of fruits/vegetables in a day</td>
<td>61% (12219)</td>
</tr>
<tr>
<td>Frequent salt consumption</td>
<td>14% (12228)</td>
</tr>
<tr>
<td>Living alone</td>
<td>29% (12296)</td>
</tr>
<tr>
<td>Moderate exercise in a day</td>
<td>78% (12309)</td>
</tr>
</tbody>
</table>

- **No significant difference** between participants in:
  - Only 1 visit vs. Multiple visits
**Factors significantly associated with SBP/DBP**

<table>
<thead>
<tr>
<th>Variable</th>
<th>SBP</th>
<th>DBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>BMI</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Male</td>
<td>NS</td>
<td>+</td>
</tr>
<tr>
<td>Pulse rate</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>2+ drinks a day</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Previously diagnosed HTN</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>On BP medication</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Smoking</td>
<td>+</td>
<td>NS</td>
</tr>
<tr>
<td>Diabetes</td>
<td>NS</td>
<td>-</td>
</tr>
<tr>
<td>Feeling stressed</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Heart attack</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hi cholesterol</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Living alone</td>
<td>NS</td>
<td>-</td>
</tr>
</tbody>
</table>

(+) = positively associated  
(-) = negatively associated  
NS = not significant
## RESULTS: MODELING FITTING

<table>
<thead>
<tr>
<th>Goodness of fit</th>
<th>Univariate model</th>
<th>Bivariate model</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 Log likelihood</td>
<td>-51849.3</td>
<td>-52666.8</td>
</tr>
<tr>
<td>AIC</td>
<td>-51837.3</td>
<td>-52652.8</td>
</tr>
<tr>
<td>BIC</td>
<td>-51803.4</td>
<td>-52613.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covariance parameters</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{\mu_1}^2$</td>
<td>0.01125</td>
<td>0.01157</td>
</tr>
<tr>
<td>$\sigma_{\mu_1,\mu_2}$</td>
<td>-</td>
<td>0.008216</td>
</tr>
<tr>
<td>$\sigma_{\mu_2}^2$</td>
<td>0.01187</td>
<td>0.01230</td>
</tr>
<tr>
<td>$\sigma_{\mu_1,b_1}$</td>
<td>-0.00025</td>
<td>-0.00033</td>
</tr>
<tr>
<td>$\sigma_{\mu_2,b_1}$</td>
<td>-</td>
<td>-0.00032</td>
</tr>
<tr>
<td>$\sigma_{b_1}^2$</td>
<td>0.000036</td>
<td>0.000058</td>
</tr>
<tr>
<td>$\sigma_{\mu_1,b_2}$</td>
<td>-</td>
<td>-0.00035</td>
</tr>
<tr>
<td>$\sigma_{\mu_2,b_2}$</td>
<td>-0.00013</td>
<td>-0.00023</td>
</tr>
<tr>
<td>$\sigma_{b_1,b_2}^2$</td>
<td>-</td>
<td>0.000058</td>
</tr>
<tr>
<td>$\sigma_{b_2}^2$</td>
<td>0.000022</td>
<td>0.000046</td>
</tr>
<tr>
<td>$\varepsilon_1$</td>
<td>0.007047</td>
<td>0.006844</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>-7.92E-6</td>
<td>0.000013</td>
</tr>
<tr>
<td>$\varepsilon_2$</td>
<td>0.007147</td>
<td>0.006870</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.000046</td>
<td>4.216E-6</td>
</tr>
</tbody>
</table>
RESULTS: MODELING ESTIMATES

- SBP/DBP reduced over time

<table>
<thead>
<tr>
<th></th>
<th>A bivariate model</th>
<th>Two univariate models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>95% CI</td>
</tr>
<tr>
<td><strong>SBP</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (mmHg)</td>
<td>125</td>
<td>(123, 127)</td>
</tr>
<tr>
<td>Monthly reduction</td>
<td>0.16%</td>
<td>(0.10%, 0.21%)</td>
</tr>
<tr>
<td><strong>DBP</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (mmHg)</td>
<td>71</td>
<td>(70, 73)</td>
</tr>
<tr>
<td>Monthly reduction</td>
<td>0.15%</td>
<td>(0.10%, 0.20%)</td>
</tr>
</tbody>
</table>

*Adjusted for other baseline covariates
** CI = confidence interval
RESULTS: SBP

The graph shows the trend of SBP (systolic blood pressure) over months for different groups: All, SDH, ISH, and IDH. The values for each category are as follows:

- **All**: 149.16, 142.39, 129.06, 126.52
- **SDH**: 142.39, 136.11, 130.06, 126.52
- **ISH**: 149.16, 142.39, 135.41, 130.72
- **IDH**: 149.16, 142.39, 135.41, 130.72

The data indicates a decrease in SBP over the months for all groups.
RESULTS: SBP & DBP

A 95% confidence region for mean SBP/DBP reduction:
• X-axis = SBP reduction
• Y-axis = DBP reduction
• Center = mean reduction, i.e. 5/3 mmHg

Correlation bet. SBP and DBP = 0.69

Vs.

a raw correlation bet. SBP and DBP = 0.5
(without considering within-subject corr.)
For all participants, on average SBP/DBP reduction = 5/3 mmHg (in 18 months)

By subgroups of HTN:
SDH: 23/13 mmHg

ISH: 13/5 mmHg

IDH: 7/8 mmHg
A decrease of 10/5 mmHg (one medication or a change in lifestyle) reduces your risk of developing CVD:

Source: Hypertension Canada
Our conclusion: analyzing SBP and DBP individually or jointly does not seem to change the estimate and its precision.

However, modeling SBP & DBP jointly has a better overall model fitting.

Modeling SBP & DBP jointly produces a better estimate of correlation between SBP and DBP.

Modeling SBP & DBP jointly can be used to derive a time-deponent correlation between SBP and DBP.
SOME LIMITATIONS

- Comparison based on empirical models
- Not a repeated-measure design; participants had irregular visits
- By design, risk factors only collected at baseline (can't look at change of factors over time)
- Non-linear models may be more suitable for the data (e.g. spline model, growth curve model)
Thank you!
Comparison of Marginal and Cluster-Specific Models for analyzing Binary Outcomes in Cluster Randomized Trials with Missing Data: a Simulation Study

Jinhui Ma, Parminder Raina, Joseph Beyene, Lehana Thabane,

September 14th, 2012

Department of Clinical Epidemiology & Biostatistics
McMaster University
Outline

- Design of cluster randomized trials (CRTs)
- Methods for analyzing binary outcomes in CRTs
- Missing data strategies
- Design of the simulation study
- Summary of results
Cluster Randomized Trial

- Cluster randomized trials (CRTs) are increasingly used in health research
Statistical Analysis Methods

- Generalized estimating equations (GEE) method

$$\text{logit}(\text{Pr}(y_{ijl} = 1)) = X_{ijl} \beta_{\text{marginal}}$$

Where

- $y_{ijl}$ denotes the binary outcome of patient $l$ in cluster $j$ in group $i$
- $X_{ijl}$ denotes the vector of cluster or individual level covariates
- $\text{Pr}(y_{ijl} = 1)$ denotes the probability of success

$$\text{logit}(\text{Pr}(y_{ijl} = 1)) = \log \left( \frac{\text{Pr}(y_{ijl} = 1)}{1 - \text{Pr}(y_{ijl} = 1)} \right)$$
Statistical Analysis Methods

- Random-effects logistic regression (RELR)

\[
\log \text{it}(\Pr(y_{ijl} = 1)) = \mathbf{X}_{ijl} \beta_{\text{conditional}} + U_{ij}
\]

Where

\[
U_{ij} \sim N(0, \sigma_B^2)
\]

\(\sigma_B^2\) represents the between – cluster variance
Statistical Analysis Methods

- For a binary outcome

\[ \beta_{\text{marginal}} = \beta_{\text{conditional}} (1 - \rho) \]

Where

\( \rho \) is the intra-cluster correlation coefficient

Comparison of GEE & RELR

- Difference on statistical power is negligible
  - once the total number of clusters is large (≥30)

- Downward biased estimates of standard error (SE) from GEE
  - when the number of clusters per arm is small
  - simple modifications can be used to adjust the SE

- GEE has acceptable accuracy and coverage probability
  - When simple adjustment is made for data with relatively few clusters

Missing Data in CRT

- Missing data is a problem in CRTs
  - Cause potential bias depending on why data are missing
  - Weaken the power of the trial

- Additional concern with missing data in CRTs
  - Entire clusters may be missing
Example of CRT: Assess effectiveness of an educational program in improving the control of blood pressure in elderly patients compared to control.

- Eligible FPs
- Randomization
- Educational Program
- Control
- FP $I_1$ (patients)
- ... (patients)
- FP $I_k$ (patients)
- FP $C_1$ (patients)
- ... (patients)
- FP $C_K$ (patients)

Measure blood pressure for each individual patient (Binary outcome: blood pressure controlled or not).

FP: Family physician practice
Missing Data Strategies

- Complete case analysis
- Standard multiple imputation (MI) using logistic regression method
- Within-cluster MI using logistic regression method
What does the missing apple look like?
What does the missing apple look like?
What does the missing apple look like?
What does the missing apple look like?
Objective of This Project

- To compare the accuracy and efficiency of marginal model (GEE method) and cluster-specific model (RELR) in analysis of binary outcomes in CRTs with missing outcome

- Evaluation criteria
  - Empirical standard error
  - Standardized bias
  - Root mean square error
  - Coverage
Generate complete dataset

**Varied Parameters**
1. Num. of clusters per trial arm  
2. Num. of subjects per cluster  
3. Intra-cluster correlation coefficient

**Fixed Parameters**
1. Treatment effect  
2. Probability of a binary covariate with value of 1

**Distributions**
1. Beta-Binomial distribution to generate clustered binary outcome  
2. Bernoulli distribution to generate binary covariate

Generate dataset with missing data

**Varied Parameters**
1. Percentage of missing data

**Distributions**
1. Bernoulli distribution to generate missing data

---

**Complete case analysis**
**Method:** exclude obs. with missing data

**Multiple imputed datasets**
**Method:** Standard MI

**Multiple imputed datasets**
**Method:** Within-cluster MI

---

**Analyzing**
- Analyzing completed dataset using GEE/RE
- Analyzing reduced dataset using GEE/RE
- Analyzing imputed dataset 1 using GEE/RE
- Analyzing imputed dataset n using GEE/RE
- Analyzing imputed dataset 1 using GEE/RE
- Analyzing imputed dataset n using GEE/RE

**Calculate SB, SE, RMSE, coverage**

---

**Pooling results from multiple analyses**

---

Figure 1: Schematic overview of the simulation study

Abbreviations: MI, multiple imputation; GEE, generalized estimating equations; RE, random-effects logistic regression; SB, standardized bias; SE, standard error; RMSE, root mean square error; obs., observations.
Results
Results

- $m=5, n=500, ICC=0.001$
- $m=5, n=500, ICC=0.01$
- $m=5, n=500, ICC=0.05$
- $m=20, n=50, ICC=0.01$
- $m=20, n=50, ICC=0.05$
- $m=20, n=50, ICC=0.1$
- $m=30, n=30, ICC=0.05$
- $m=30, n=30, ICC=0.1$
- $m=30, n=30, ICC=0.2$

Legend:
- GEE for Complete case analysis
- GEE for standard MI
- GEE for within-cluster MI
- RELR for Complete case analysis
- RELR for standard MI
- RELR for within-cluster MI

% of Missing Data vs. Standardized Bias
Results
Results

- GEE for Complete case analysis
- GEE for standard MI
- GEE for within-cluster MI
- RELR for Complete case analysis
- RELR for standard MI
- RELR for within-cluster MI
## Conclusions

<table>
<thead>
<tr>
<th>Design effect of CRTs(^1)</th>
<th>Percentage of missing data</th>
<th>Missing data strategies</th>
<th>Performance of statistical analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Complete case analysis</td>
<td>GEE</td>
</tr>
<tr>
<td>VIF(^2)&lt;3</td>
<td>&lt;15%</td>
<td>Complete case analysis</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard MI(^3)</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within-cluster MI(^4)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>≥15%</td>
<td>Complete case analysis</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard MI(^3)</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within-cluster MI(^4)</td>
<td>X</td>
</tr>
<tr>
<td>VIF(^2)≥3</td>
<td>&lt;15%</td>
<td>Complete case analysis</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard MI(^3)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within-cluster MI(^4)</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>≥15%</td>
<td>Complete case analysis</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard MI(^3)</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Within-cluster MI(^4)</td>
<td>√</td>
</tr>
</tbody>
</table>

Note: 1 CRTs = Cluster randomized trials  
2 VIF = Variance inflation factor  
3 Standard MI = Standard multiple imputation  
4 Within-cluster MI = Within-cluster multiple imputation, which is not applicable for CRTs with small cluster size
Limitation of This Project

- GEE and RELR are assessed only for
  - CRTs with a completely randomized design
  - CRTs with balanced design

- Conclusion may not be applicable for
  - likelihood based analyses
  - other imputation methods

- Simulation using beta-binomial model gives preference to the GEE model
Thanks for your attention!