Bayesian Modeling of Birthweight and Gestational Age

Scott L. Schwartz, Alan E. Gelfand, and Marie Lynn Miranda

Duke University

July 30, 2008
1. Birthweight and Gestational Age

2. A Joint Outcome?

3. Censored Gestational Age

4. The Birth Landscape

5. Results
Birthweight and Gestational Age

Frequently Studied Birth Outcomes:

- Low Birthweight (LBW)  \( b_i < 2500 \) grams
- Very LBW (VLBW)  \( b_i < 1500 \) grams
- Pre Term (PT)  \( g_i < 37 \) weeks
- Very PT (VPT)  \( g_i < 34 \) weeks

\( b_i \): Birthweight and \( g_i \): Gestational Age for Birth \( i \)

- Often studied as adverse birth outcomes in their own right due to strength of relationship.
- Meaning appears to vary by population group.
- Joint Role of Birthweight and Gestational Age is clearly predominant, but not well understood.
### Frequently Studied Birth Outcomes:

- **Low Birthweight (LBW)**: $b_i < 2500$ grams
- **Very LBW (VLBW)**: $b_i < 1500$ grams
- **Pre Term (PT)**: $g_i < 37$ weeks
- **Very PT (VPT)**: $g_i < 34$ weeks

- $b_i$ Birthweight and $g_i$ Gestational Age for Birth $i$

- Often studied as adverse birth outcomes in their own right due to strength of relationship.
- Meaning appears to vary by population group.
- Joint Role of Birthweight and Gestational Age is clearly predominant, but not well understood.
Birthweight and Gestational Age

Frequently Studied Birth Outcomes:

- Low Birthweight (LBW) \( b_i < 2500 \) grams
- Very LBW (VLBW) \( b_i < 1500 \) grams
- Pre Term (PT) \( g_i < 37 \) weeks
- Very PT (VPT) \( g_i < 34 \) weeks

\( b_i \) Birthweight and \( g_i \) Gestational Age for Birth \( i \)

- Often studied as adverse birth outcomes in their own right due to strength of relationship.
- Meaning appears to vary by population group.
- Joint Role of Birthweight and Gestational Age is clearly predominant, but not well understood.
Birthweight and Gestational Age

Frequently Studied Birth Outcomes:

- Low Birthweight (LBW) $b_i < 2500$ grams
- Very LBW (VLBW) $b_i < 1500$ grams
- Pre Term (PT) $g_i < 37$ weeks
- Very PT (VPT) $g_i < 34$ weeks

$ b_i $ Birthweight and $ g_i $ Gestational Age for Birth $ i $

- Often studied as adverse birth outcomes in their own right due to strength of relationship.
- Meaning appears to vary by population group.
- Joint Role of Birthweight and Gestational Age is clearly predominant, but not well understood.
North Carolina Detailed Birth Record 2004--2006 (Restrictions, N=334856)

Restrictions:
- No Congenital Anomalies
- No Alcohol Use
- No Reported Infant Deaths
- Mothers Between 15 and 44 Years of Age (Inclusive)
- Non-Hispanic White, Non-Hispanic Black, and Hispanic Mothers Only
- Single Births

Birthweight (grams)

Gestational Age (weeks)
North Carolina Detailed Birth Record 2004--2006 (Restrictions, N=334856)

**Missing Part of Gestational Age Imputed**

**Restrictions:**
- No Congenital Anomalies
- No Alcohol Use
- No Reported Infant Deaths
- Mothers Between 15 and 44 Years of Age (Inclusive)
- Non-Hispanic White, Non-Hispanic Black, and Hispanic Mothers Only
- Single Births

**Birthweight (grams)**

**Gestational Age (weeks)**
A Joint Outcome?

- Perhaps we can benefit from studying Birthweight and Gestational Age as a Joint Event?

Bivariate Normal Mixture Model:

\[
(b_i, g_i) \sim \sum_{k=1}^{s} p_k 1[v_i, k=1] N(g_i | x_i' \beta_g, k, \sigma_g, k) \\
N(b_i | x_i' \beta_b, k + (g_i - x_i' \beta_g, k) \beta^*_k, \sigma_b | g, k)
\]

- This facilitates interpretation and provides a bias free framework to learn about \(b_i, g_i, \) and \(x_i\).
- “Centering” \(g_i\) “on-the-fly” removes Back-Door bias and provides a non-autocorrelated parameterization.
A Joint Outcome?

- Perhaps we can benefit from studying Birthweight and Gestational Age as a Joint Event?

Bivariate Normal Mixture Model:

\[(b_i, g_i)' \sim \sum_{k=1}^{s} p_k 1[v_i, k=1] N(g_i | x_i' \beta_g, k, \sigma_g, k) \]
\[N(b_i | x_i' \beta_b, k + (g_i - x_i' \beta_g, k) \beta_* k, \sigma_b|g, k)\]

- This facilitates interpretation and provides a bias free framework to learn about \(b_i, g_i, \) and \(x_i.\)
- “Centering” \(g_i\) “on-the-fly” removes Back-Door bias and provides a non-autocorrelated parameterization.
A Joint Outcome?

- Perhaps we can benefit from studying Birthweight and Gestational Age as a Joint Event?

Bivariate Normal Mixture Model:

\[(b_i, g_i)' \sim \sum_{k=1}^{s} p_k 1[v_i,k=1] N(g_i|x_i' \beta_{g,k}, \sigma_{g,k}) \]
\[ N(b_i|x_i' \beta_{b,k} + (g_i - x_i' \beta_{g,k}) \beta_{*k}, \sigma_{b|g,k}) \]

- This facilitates interpretation and provides a bias free framework to learn about \(b_i, g_i,\) and \(x_i\).
- “Centering” \(g_i\) “on-the-fly” removes Back-Door bias and provides a non-autocorrelated parameterization.
Bivariate Normal Mixture Model:

\[(g_i, b_i) \sim \sum_{k=1}^{s} p_k 1[v_{i,k} = 1] N(M_k, S_k)\]

\[M_k = \begin{bmatrix} x_i' \beta_{b,k} \\ x_i' \beta_{g,k} \end{bmatrix}\]

\[S_k = \begin{bmatrix} \sigma_{b|g,k}^2 & \rho_k \sigma_{b|g,k} \sigma_{g,k} \\ \rho_k \sigma_{b|g,k} \sigma_{g,k} & \rho_k \sigma_{g,k}^2 \end{bmatrix}\]

\[
\rho_k = \pm \sqrt{\left(\frac{\beta_{*k} \sigma_{g,k}}{\sigma_{b|g,k}}\right)^2 \left(1 + \left(\frac{\beta_{*k} \sigma_{g,k}}{\sigma_{b|g,k}}\right)^2\right)^{-1}} \\
(\text{where } \pm \text{ matches the sign of } \beta_{*k})
\]

Complete with priors...
North Carolina Detailed Birth Record 2004--2006 (Restrictions, N=334856)

7 Component Mixture Model for Residuals

Restrictions:
- No Congenital Anomalies
- No Alcohol Use
- No Reported Infant Deaths
- Mothers Between 15 and 44 Years of Age (Inclusive)
- Non-Hispanic White
- Non-Hispanic Black, and Hispanic Mothers Only
- Single Births

Birthweight (grams)

Gestational Age (weeks)
North Carolina Detailed Birth Record 2004--2006 (Restrictions, N=334856)

Missing Part of Gestational Age Imputed

Restrictions:
- No Congenital Anomalies
- No Alcohol Use
- No Reported Infant Deaths
- Mothers Between 15 and 44 Years of Age (Inclusive)
- Non-Hispanic White, Non-Hispanic Black, and Hispanic Mothers Only
- Single Births

Birthweight (grams)

Gestational Age (weeks)
Uncertainty from Censored Gestational Age

- We don’t actually get to observe $g_i$.
- Instead, we see $g^c_i$, a clinical estimate of the number of completed weeks of Gestation.
- That is, we observe, (Right) Censored Gestational Age, where we lose $u_i$ in $g_i \equiv g^c_i + u_i$, the Unmeasured part of Gestational Age.
- So, everything we know about $u_i$ can be specified by the prior $u_i \sim U[0, 1)$.
- $u_i$ can now be included in posterior inference. This recognizes and incorporates our uncertainty about $g_i$, as well as allow us to learn about it!
- We currently assume $g^c_i$ is measured accurately.
Uncertainty from Censored Gestational Age

- We don’t actually get to observe $g_i$.
- Instead, we see $g_i^c$, a clinical estimate of the number of completed weeks of Gestation.
- That is, we observe, (Right) Censored Gestational Age, where we lose $u_i$ in $g_i \equiv g_i^c + u_i$, the Unmeasured part of Gestational Age.
- So, everything we know about $u_i$ can be specified by the prior $u_i \sim U[0, 1]$.
- $u_i$ can now be included in posterior inference. This recognizes and incorporates our uncertainty about $g_i$, as well as allow us to learn about it!
- We currently assume $g_i^c$ is measured accurately.
Uncertainty from Censored Gestational Age

- We don’t actually get to observe \( g_i \).
- Instead, we see \( g_i^c \), a clinical estimate of the number of *completed* weeks of Gestation.
- That is, we observe, (Right) Censored Gestational Age, where we lose \( u_i \) in \( g_i \equiv g_i^c + u_i \), the Unmeasured part of Gestational Age.
- So, everything we know about \( u_i \) can be specified by the prior \( u_i \sim U[0,1) \).
- \( u_i \) can know be included in posterior inference. This recognizes *and incorporates* our uncertainty about \( g_i \), as well as allow us to learn about it!
- We currently assume \( g_i^c \) is measured accurately.
Uncertainty from Censored Gestational Age

- We don’t actually get to observe $g_i$.
- Instead, we see $g_i^c$, a clinical estimate of the number of *completed* weeks of Gestation.
- That is, we observe, (Right) Censored Gestational Age, where we lose $u_i$ in $g_i \equiv g_i^c + u_i$, the Unmeasured part of Gestational Age.
- So, everything we know about $u_i$ can be specified by the prior $u_i \sim U[0, 1]$.
- $u_i$ can now be included in posterior inference. This recognizes *and incorporates* our uncertainty about $g_i$, as well as allow us to learn about it!
- We currently assume $g_i^c$ is measured accurately.
Uncertainty from Censored Gestational Age

- We don’t actually get to observe $g_i$.
- Instead, we see $g_i^c$, a clinical estimate of the number of completed weeks of Gestation.
- That is, we observe, (Right) Censored Gestational Age, where we lose $u_i$ in $g_i \equiv g_i^c + u_i$, the Unmeasured part of Gestational Age.
- So, everything we know about $u_i$ can be specified by the prior $u_i \sim U[0, 1]$.
- $u_i$ can now be included in posterior inference. This recognizes and incorporates our uncertainty about $g_i$, as well as allow us to learn about it!
- We currently assume $g_i^c$ is measured accurately.
Uncertainty from Censored Gestational Age

- We don’t actually get to observe $g_i$.
- Instead, we see $g_i^c$, a clinical estimate of the number of completed weeks of Gestation.
- That is, we observe, (Right) Censored Gestational Age, where we lose $u_i$ in $g_i = g_i^c + u_i$, the Unmeasured part of Gestational Age.
- So, everything we know about $u_i$ can be specified by the prior $u_i \sim U[0, 1]$.
- $u_i$ can now be included in posterior inference. This recognizes and incorporates our uncertainty about $g_i$, as well as allow us to learn about it!
- We currently assume $g_i^c$ is measured accurately.
Avoiding Misleading Bias

- The roles of Birthweight and Gestational Age are a part of a complicated field of factors.
- Sex and birth order of the baby, age, race, and tobacco use of the mother, are just a few more recognizable factors involved.
- In a regression or stratification context, we run the risk of misunderstanding these variables relationship to birthweight and gestational age:
Avoiding Misleading Bias

- The roles of Birthweight and Gestational Age are a part of a complicated field of factors.
- Sex and birth order of the baby, age, race, and tobacco use of the mother, are just a few more recognizable factors involved.
- In a regression or stratification context, we run the risk of misunderstanding these variables relationship to birthweight and gestational age:
Avoiding Misleading Bias

- The roles of Birthweight and Gestational Age are a part of a complicated field of factors.
- Sex and birth order of the baby, age, race, and tobacco use of the mother, are just a few more recognizable factors involved.
- In a regression or stratification context, we run the risk of misunderstanding these variables relationship to birthweight and gestational age:

  ![Diagram showing the relationship between Smoking, Birthweight, Gestational Age, and residuals with back-door criteria violation and no back-door criteria violation.](attachment:image.png)
Results!!!

- Are in progress...
- We have developed Bayesian specifications in Matlab as well as C using GSL libraries.
- We are currently running posterior simulations with North Carolina Detailed Birth Record (NCDBR) data on CEHI’s private Microsoft network using Cygwin.
- CEHI’s version of the NCDBR consists of all live births in North Carolina from 1990 to 2006 (No Restrictions, N=1,862,405). We currently focus on births from 2004 (Restrictions, N=334,856).
Results!!!

- Are in progress...

- We have developed Bayesian specifications in Matlab as well as C using GSL libraries.

- We are currently running posterior simulations with North Carolina Detailed Birth Record (NCDBR) data on CEHI’s private Microsoft network using Cygwin.

- CEHI’s version of the NCDBR consists of all live births in North Carolina from 1990 to 2006 (No Restrictions, N=1,862,405). We currently focus on births from 2004 (Restrictions, N=334,856).
Results!!!

- Are in progress...

- We have developed Bayesian specifications in Matlab as well as C using GSL libraries.

- We are currently running posterior simulations with North Carolina Detailed Birth Record (NCDBR) data on CEHI’s private Microsoft network using Cygwin.

- CEHI’s version of the NCDBR consists of all live births in North Carolina from 1990 to 2006 (No Restrictions, N=1,862,405). We currently focus on births from 2004 (Restrictions, N=334,856).
Results!!!

- Are in progress...
- We have developed Bayesian specifications in Matlab as well as C using GSL libraries.
- We are currently running posterior simulations with North Carolina Detailed Birth Record (NCDBR) data on CEHI’s private Microsoft network using Cygwin.
- CEHI’s version of the NCDBR consists of all live births in North Carolina from 1990 to 2006 (No Restrictions, N=1,862,405). We currently focus on births from 2004 (Restrictions, N=334,856).
Bayesian Modeling of Birthweight and Gestational Age

Birthweight and Gestational Age: A Joint Outcome?

Censored Gestational Age

The Birth Landscape

Results

JSM 2008, Denver Colorado

Department of
Statistical Science

Children's
Environmental
Health Initiative

Nicholas School
of the Environment

Duke
University

July 30, 2008

Scott L. Schwartz, Alan E. Gelfand, and Marie Lynn Miranda
Bayesian Modeling of Birthweight and Gestational Age

Birthweight and Gestational Age: A Joint Outcome? Censored Gestational Age The Birth Landscape Results

Scott L. Schwartz, Alan E. Gelfand, and Marie Lynn Miranda

July 30, 2008 Slide 14 of 16
Bayesian Modeling of Birthweight and Gestational Age

Birthweight and Gestational Age A Joint Outcome? Censored Gestational Age The Birth Landscape Results

July 30, 2008 Slide 15 of 16

Scott L. Schwartz, Alan E. Gelfand, and Marie Lynn Miranda
Bayesian Modeling of Birthweight and Gestational Age

A Joint Outcome?

Censored Gestational Age

The Birth Landscape

Results

July 30, 2008

Scott L. Schwartz, Alan E. Gelfand, and Marie Lynn Miranda