A semiparametric Bayesian model for examiner agreement in periodontal research

Elizabeth G. Hill and Elizabeth H. Slate

Division of Biostatistics and Epidemiology, MUSC

September 26, 2011
Dental anatomy

From http://dentalimplants-usa.com/generalinfo/toothnumbering.html
Calibrating examiners

- Larger clinical periodontal studies require multiple examiners
- Accurate measurement of PD requires training
- Calibration studies demonstrate degree of agreement among examiners and with a gold standard examiner
- Agreement varies with examiner, but may also depend on characteristics of the site
Calibrating examiners

- Larger clinical periodontal studies require multiple examiners
- Accurate measurement of PD requires training
- Calibration studies demonstrate degree of agreement among examiners and with a gold standard examiner
- Agreement varies with examiner, but may also depend on characteristics of the site

Goals:
- Quantify agreement
- Determine targets for enhanced training
Pocket and probing depth
True pocket depth

Pocket and probing depth

exposed root
gum attachment

Moderate periodontitis
Pocket and probing depth

*True* pocket depth
PD = 4.6 mm
Motivation

Pocket and probing depth

True pocket depth
PD = 4.6 mm

Observed probed pocket depth

Moderate periodontitis
Pocket and probing depth

True pocket depth
PD = 4.6 mm

Observed probed pocket depth
5 mm < Obs PPD < 6 mm
Pocket and probing depth

**True** pocket depth
PD = 4.6 mm

**Observed** probed pocket depth
5 mm < Obs PPD < 6 mm

**Recorded** PPD

Moderate periodontitis
Pocket and probing depth

**True** pocket depth
PD = 4.6 mm

**Observed** probed pocket depth
5 mm < Obs PPD < 6 mm

**Recorded** PPD
Rec PPD = 5 mm
Examiner calibration pilot study

- Three hygienists (A, B, C) and one standard examiner (S)
- Nine study subjects
- Subjects’ quadrants randomized to examiner pairs
- Pocket depth measured at six sites for all available teeth by two examiners at each site
- As many as 168 sites (336 measurements) per subject
- 1080 duplicate observations (2160 measurements) total
- Inter- (AS, BS, CS, AB, AC, BC) and intra-examiner (AA, BB, CC, SS) measurements collected
## Percent agreement

<table>
<thead>
<tr>
<th>Examiner pair</th>
<th>k</th>
<th>n</th>
<th>Exact</th>
<th>±1 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>%</td>
<td>95% CI</td>
</tr>
<tr>
<td>AS</td>
<td>5</td>
<td>180</td>
<td>62</td>
<td>(36,88)</td>
</tr>
<tr>
<td>BS</td>
<td>5</td>
<td>156</td>
<td>49</td>
<td>(25,73)</td>
</tr>
<tr>
<td>CS</td>
<td>5</td>
<td>180</td>
<td>43</td>
<td>(34,51)</td>
</tr>
<tr>
<td>AB</td>
<td>3</td>
<td>108</td>
<td>45</td>
<td>(28,63)</td>
</tr>
<tr>
<td>AC</td>
<td>3</td>
<td>96</td>
<td>44</td>
<td>(0,89)</td>
</tr>
<tr>
<td>BC</td>
<td>3</td>
<td>120</td>
<td>47</td>
<td>(13,80)</td>
</tr>
<tr>
<td>AA</td>
<td>2</td>
<td>60</td>
<td>73</td>
<td>NA</td>
</tr>
<tr>
<td>BB</td>
<td>2</td>
<td>72</td>
<td>56</td>
<td>(44,67)</td>
</tr>
<tr>
<td>CC</td>
<td>2</td>
<td>78</td>
<td>79</td>
<td>(59,100)</td>
</tr>
<tr>
<td>SS</td>
<td>1</td>
<td>30</td>
<td>80</td>
<td>NA</td>
</tr>
</tbody>
</table>

(DBE teatime)
Our approach

- Use model-based approach that naturally incorporates dependence among observations
- Use *ALL* data to borrow strength
- Construct reliability measures ($\kappa_w$, % agreement) from realizations from posterior predictive distribution
1. **TRUE pocket depth**

\[
\log(\theta_{ij}) = \mu + b_i + \varepsilon_{ij}
\]

- \(\theta_{ij}\) is pocket depth for \(j\)th site of \(i\)th subject
- \(b_i \mid \sigma_b^2 \sim \text{Normal}(0, \sigma_b^2)\)
- \(\varepsilon_{ij} \mid \sigma_{\varepsilon}^2 \sim \text{Normal}(0, \sigma_{\varepsilon}^2)\)
- \(b_i\) and \(\varepsilon_{ij}\) independent
2. **OBSERVED** probed pocket depth

\[
\log(O_{ijk}) = \log(\theta_{ij}) + \mathbf{x}'_{ijk} \beta_{ij} + \gamma_{ijk}
\]

- **O** \(_{ijk}\) is depth observed on probe for \(k\)th replicate of \(j\)th site of \(i\)th subject
- **X** \(_{ijk}\) = \((X_{A,ijk}, X_{B,ijk}, X_{C,ijk})'\) - examiner indicators
- **\(\beta\)** \(_{ij}\) = \((\beta_{A,ij}, \beta_{B,ij}, \beta_{C,ij})'\) - associated parameter vector
- **\(\gamma\)** \(_{ijk}\) \(\sim\) Normal\((0, \sigma^2_E)\)
- \(\sigma^2_E\) subscripts are A, B, C or S
3. *RECORDED* probed pocket depth

\[
T_{ijk} = \begin{cases} 
\lfloor O_{ijk} \rfloor & \text{if } 0 \leq O_{ijk} < 15 \\
15 & \text{otherwise}
\end{cases}
\]

- \( T_{ijk} \) is recorded probed pocket depth for \( k \)th replicate of \( j \)th site for \( i \)th subject
- PPD recorded as floor of observed PPD
- Manual probe scored to a maximum depth of 15 mm
3. **RECORDED** probed pocket depth (cont.)

\[
\pi_{t,ijk} = \text{Prob}(T_{ijk} = t | \log(\theta_{ij}), \beta_{ij}, \sigma_E)
\]

\[
= \begin{cases} 
\zeta_{t+1} & \text{if } t = 0 \\
\zeta_{t+1} - \zeta_t & \text{if } t = 1, \ldots, 14 \\
1 - \zeta_t & \text{if } t = 15
\end{cases}
\]

- \( \zeta_t = \Phi \left( \frac{\log(t) - \log(\theta_{ij}) - \mathbf{x}_{ijk}' \beta_{ij}}{\sigma_E} \right) \)
- \( \Phi(\cdot) \) is standard normal CDF
Likelihood

Conditional likelihood proportional to

\[
\prod_{i=1}^{n} \prod_{j=1}^{m_i} \prod_{k=1}^{2} \prod_{t=0}^{15} \pi_{t,ijk} V_{t,ijk}
\]

- \( V_{ijk} = (V_{0,ijk}, V_{1,ijk}, \ldots, V_{15,ijk})' \) is vector of 15 zeros and a single one
- \( T_{ijk} = t \Rightarrow V_{t,ijk} = 1 \)
- \( V_{ijk} | \log(\theta_{ij}), \beta_{ij}, \sigma^2_E \sim \text{Multinomial}(1; \pi_{0,ijk}, \pi_{1,ijk}, \ldots, \pi_{15,ijk}) \)
Prior and hyperprior specifications

- Mean of true distribution ($\mu$) - vague mean zero Normal
- Error terms (five) and random effect - Normal$(0, 1/\sigma^2_\nu)$
- Standard deviations ($\sigma_\nu$) - Uniform$(0, a)$ (Gelman, Bayesian Analysis 2005)

“Non”parametric Component:

- Rater bias parameters ($\beta_{E,ij}$s) - Dirichlet process prior (DPP)
  - DPP naturally gives rise to clusters
  - Reasonable to assume latent class structure
  - Preliminary analysis indicates site-level characteristics cause examiners to be more (less) prone to bias
Dirichlet process prior overview

- $y_i \sim f(y_i | \phi_i)$, $i = 1, \ldots, n$, with $\phi_i$ unknown
- $\phi_i$ is centered around base distribution $G_0$
- Candidate values for $\phi_i$ are drawn from $G_0$ according to concentration parameter $\alpha$
- Traditional Bayesian approach places prior on $\phi_i$
- DPP places prior on $\phi_i$’s distribution

\[
\phi_i | \Gamma \sim \Gamma \\
\Gamma \sim \text{DP}(\alpha G_0)
\]

- Assignment of a given candidate value from $G_0$ to multiple $\phi_i$ may be expected so that $\phi_i$’s cluster based on similarities among the $y_i$’s

Dirichlet process prior overview (cont.)

Practical approach based on finite implementation (Sethuraman 1994)

- Draw $M \leq n$ candidate values, denoted $\phi_m^*$ with $m = 1, \ldots, M$, from $G_0$

- $M^* \leq M$ of these are allocated to one or more of the $\phi_i$

- Assignment of $\phi_i$ to $\phi_m^*$ determined by a multinomial distribution with probability vector $\mathbf{P} \sim \text{Dirichlet}(\alpha G_0(B_1), \ldots, \alpha G_0(B_M))$

- Assignment of multiple $\phi_i$ to the same candidate value $\phi_m^*$ results in the formation of a cluster

- Empty clusters result when none of the $\phi_i$ are assigned to one or more of the $M$ candidate values.
Dirichlet process prior overview (cont.)

- $\phi_i$’s density is discrete and the fineness of the discretization increases with $\alpha$
  - $\alpha$ large $\Rightarrow$ density of $\phi_i$ resembles $G_0$
  - $\alpha$ small $\Rightarrow$ density of $\phi_i$ similar to a finite mixture model

- Dirichlet Process Mixture Model (DPMM)
DPMM for examiner biases

- For our application
  \[ \beta_{E,ij} | \Gamma_E \sim \Gamma_E \]
  \[ \Gamma_E \sim DP(\alpha_E G_{E,0}) \]

- \( G_{E,0} \equiv G_0 \) vague mean zero Normal
- DPMM identifies latent classes of periodontal sites that “cluster” based on examiner-specific biased rating behavior
- \( \alpha_E \equiv 8 \) (tried \( \alpha = 0.5, 1, 2, \ldots, 10, 20 \))
- Total number of classes, \( M_E \equiv 6 \) (tried 2, 3, 4, 5, 6)
- \( M \) limited by number of distinct data values (Congdon 2001)
  - In our data, recorded PPD ranged from 0mm to 8mm
  - Differences in duplicate recorded PPD ranged from -4mm to 4mm
  - \( M_E \leq 9 \)
Posterior clustering inference

- Least-squares clustering identifies examiner-specific classes of biased ratings \( \text{(Dahl 2006)} \)
- \( \mathbf{c}_{E,1}, \ldots, \mathbf{c}_{E,D} \) are \( D \) draws from posterior clustering distribution of \( \beta_{E,ij} \)
- For each \( \mathbf{c}_E \) in \( \mathbf{c}_{E,1}, \ldots, \mathbf{c}_{E,D} \), construct \( \mathcal{L} \times \mathcal{L} \) association matrix, \( \delta(\mathbf{c}_E) \) (\( \mathcal{L} = \text{total number of sites} \))
- \( \delta(\mathbf{c}_E)_{\ell \ell'} = 1 \) when sites \( \ell \) and \( \ell' \) jointly classified, 0 otherwise
- \( \Delta_E = \text{element-wise average of collection of association matrices} \)

Examiner \( E \)'s least-squares cluster

\[
\mathbf{c}_{E}^{\text{LS}} = \arg\min_{\mathbf{c}_E \in \{\mathbf{c}_{E,1}, \ldots, \mathbf{c}_{E,D}\}} \sum_{\ell=1}^{\mathcal{L}} \sum_{\ell'=1}^{\mathcal{L}} \left( \delta(\mathbf{c}_E)_{\ell \ell'} - \Delta_{E,\ell \ell'} \right)^2
\]
Simulation study

- Comparable in size to real calibration data
  - 9 subjects, 3 examiners, 1 standard
- Examiners S, A no bias
- Examiner B biased on deep pockets; C biased, severely on DLMM

**True pocket depth model**

\[
\log(T_{ijk}) = \log(\theta_{ij}) + \beta_{B,ij} \cdot I(\theta_{ij} \geq 4\text{mm}) + \beta_{C_1,ij} + \\
\beta_{C_2,ij} \cdot I(\text{site } j \text{ is distolingual mandibular molar}) + \gamma_{ijk},
\]

- \(\beta_{B,ij} = -0.5\)
- \(\beta_{C_1,ij} = 0.25\)
- \(\beta_{C_2,ij} = -1\)
Simulation results

10K iterations (50,500 burn-in)

<table>
<thead>
<tr>
<th>Description</th>
<th>Parameter</th>
<th>Truth</th>
<th>Median</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean true PD</td>
<td>$\mu$</td>
<td>1</td>
<td>1.03</td>
<td>(0.80, 1.18)</td>
</tr>
<tr>
<td>SD random effect</td>
<td>$\sigma_b$</td>
<td>0.2</td>
<td>0.19</td>
<td>(0.11, 0.40)</td>
</tr>
<tr>
<td>SD true PD</td>
<td>$\sigma_\varepsilon$</td>
<td>0.3</td>
<td>0.29</td>
<td>(0.28, 0.30)</td>
</tr>
<tr>
<td>SD Ex A obs PPD</td>
<td>$\sigma_A$</td>
<td>0.1</td>
<td>0.11</td>
<td>(0.09, 0.13)</td>
</tr>
<tr>
<td>SD Ex B obs PPD</td>
<td>$\sigma_B$</td>
<td>0.25</td>
<td>0.24</td>
<td>(0.22, 0.28)</td>
</tr>
<tr>
<td>SD Ex C obs PPD</td>
<td>$\sigma_C$</td>
<td>0.15</td>
<td>0.15</td>
<td>(0.12, 0.17)</td>
</tr>
<tr>
<td>SD Ex S obs PPD</td>
<td>$\sigma_S$</td>
<td>0.07</td>
<td>0.08</td>
<td>(0.07, 0.10)</td>
</tr>
</tbody>
</table>
## Estimated $\hat{\kappa}_W$

<table>
<thead>
<tr>
<th>Examiner Pair</th>
<th>Truth</th>
<th>Median</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS</td>
<td>0.89</td>
<td>0.85</td>
<td>(0.77, 0.93)</td>
</tr>
<tr>
<td>BS</td>
<td>0.69</td>
<td>0.67</td>
<td>(0.55, 0.82)</td>
</tr>
<tr>
<td>CS</td>
<td>0.66</td>
<td>0.61</td>
<td>(0.50, 0.77)</td>
</tr>
<tr>
<td>AB</td>
<td>0.68</td>
<td>0.63</td>
<td>(0.47, 0.80)</td>
</tr>
<tr>
<td>AC</td>
<td>0.66</td>
<td>0.59</td>
<td>(0.45, 0.76)</td>
</tr>
<tr>
<td>BC</td>
<td>0.55</td>
<td>0.50</td>
<td>(0.33, 0.69)</td>
</tr>
<tr>
<td>AA</td>
<td>0.87</td>
<td>0.84</td>
<td>(0.73, 0.92)</td>
</tr>
<tr>
<td>BB</td>
<td>0.56</td>
<td>0.62</td>
<td>(0.44, 0.79)</td>
</tr>
<tr>
<td>CC</td>
<td>0.72</td>
<td>0.82</td>
<td>(0.71, 0.91)</td>
</tr>
<tr>
<td>SS</td>
<td>0.91</td>
<td>0.87</td>
<td>(0.78, 0.95)</td>
</tr>
</tbody>
</table>
Estimated $\hat{\kappa}_w$ - Corrects for bias!

<table>
<thead>
<tr>
<th>Examiner Pair</th>
<th>Observed</th>
<th>Truth</th>
<th>Median</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS</td>
<td>0.90</td>
<td>0.89</td>
<td>0.85</td>
<td>(0.77, 0.93)</td>
</tr>
<tr>
<td>BS</td>
<td>0.45</td>
<td>0.69</td>
<td>0.67</td>
<td>(0.55, 0.82)</td>
</tr>
<tr>
<td>CS</td>
<td>0.59</td>
<td>0.66</td>
<td>0.61</td>
<td>(0.50, 0.77)</td>
</tr>
<tr>
<td>AB</td>
<td>0.43</td>
<td>0.68</td>
<td>0.63</td>
<td>(0.47, 0.80)</td>
</tr>
<tr>
<td>AC</td>
<td>0.71</td>
<td>0.66</td>
<td>0.59</td>
<td>(0.45, 0.76)</td>
</tr>
<tr>
<td>BC</td>
<td>0.45</td>
<td>0.55</td>
<td>0.50</td>
<td>(0.33, 0.69)</td>
</tr>
<tr>
<td>AA</td>
<td>0.83</td>
<td>0.87</td>
<td>0.84</td>
<td>(0.73, 0.92)</td>
</tr>
<tr>
<td>BB</td>
<td>0.34</td>
<td>0.56</td>
<td>0.62</td>
<td>(0.44, 0.79)</td>
</tr>
<tr>
<td>CC</td>
<td>0.85</td>
<td>0.72</td>
<td>0.82</td>
<td>(0.71, 0.91)</td>
</tr>
<tr>
<td>SS</td>
<td>0.88</td>
<td>0.91</td>
<td>0.87</td>
<td>(0.78, 0.95)</td>
</tr>
</tbody>
</table>

Guggenmoos-Holzmann and Vonk (SIM 1998) noted this bias in $\kappa$. 

(DBE teatime)
Simulation study

Posterior distributions of examiner effects

- Examiner A - 1 class
- Examiners B and C
  - 1 dominant and 1 subordinate class
  - B’s class membership significantly associated with deep pockets ($p < 0.0001$)
  - C’s class membership significantly associated with DLMM sites ($p < 0.0001$)
Application to real calibration data

Compared model fits for 4 variants:

- Model 0: No examiner biases, common examiner variances
- Model 1: No examiner biases, unequal variances
- Model 2: Fixed bias for each examiner ($\beta_{E,ij} = \beta_E$), unequal variances
- Model 3: Site-level biases, unequal variances
Application model selection

- Examined posterior predicted distributions for recorded PPD
- Examined DIC\(_3\) values (Celeux et al. 2006)

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC(_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: No bias, common variance</td>
<td>4560.11</td>
</tr>
<tr>
<td>1: No bias, unequal variance</td>
<td>4402.13</td>
</tr>
<tr>
<td>2: Fixed bias, unequal variance</td>
<td>4129.07</td>
</tr>
<tr>
<td>3: Site-level bias, unequal variance</td>
<td>3381.83</td>
</tr>
</tbody>
</table>
# Agreement measures

<table>
<thead>
<tr>
<th>Examiner Pair</th>
<th>% ±1mm</th>
<th>Model</th>
<th>Observed</th>
<th>$\kappa_w$</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS</td>
<td>94 (83, 100)</td>
<td>95 (89, 99)</td>
<td>.71</td>
<td>.80 (.69, .90)</td>
<td></td>
</tr>
<tr>
<td>BS</td>
<td>88 (71, 100)</td>
<td>89 (78, 95)</td>
<td>.67</td>
<td>.64 (.49, .80)</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>92 (82, 100)</td>
<td>92 (83, 97)</td>
<td>.69</td>
<td>.71 (.59, .84)</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>82 (50, 100)</td>
<td>85 (73, 94)</td>
<td>.63</td>
<td>.60 (.42, .77)</td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>88 (64, 100)</td>
<td>88 (76, 96)</td>
<td>.59</td>
<td>.62 (.44, .79)</td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>81 (59, 100)</td>
<td>88 (77, 96)</td>
<td>.62</td>
<td>.60 (.43, .77)</td>
<td></td>
</tr>
</tbody>
</table>
## Agreement measures

<table>
<thead>
<tr>
<th>Examiner Pair</th>
<th>Observed</th>
<th>Model</th>
<th>Observed</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS</td>
<td>94 (83, 100)</td>
<td>95 (89, 99)</td>
<td>.71</td>
<td>.80 (.69, .90)</td>
</tr>
<tr>
<td>BS</td>
<td>88 (71, 100)</td>
<td>89 (78, 95)</td>
<td>.67</td>
<td>.64 (.49, .80)</td>
</tr>
<tr>
<td>CS</td>
<td>92 (82, 100)</td>
<td>92 (83, 97)</td>
<td>.69</td>
<td>.71 (.59, .84)</td>
</tr>
<tr>
<td>AB</td>
<td>82 (50, 100)</td>
<td>85 (73, 94)</td>
<td>.63</td>
<td>.60 (.42, .77)</td>
</tr>
<tr>
<td>AC</td>
<td>88 (64, 100)</td>
<td>88 (76, 96)</td>
<td>.59</td>
<td>.62 (.44, .79)</td>
</tr>
<tr>
<td>BC</td>
<td>81 (59, 100)</td>
<td>88 (77, 96)</td>
<td>.62</td>
<td>.60 (.43, .77)</td>
</tr>
<tr>
<td>A·Truth</td>
<td>96 (91, 98)</td>
<td>96 (91, 98)</td>
<td>.81 (.73, .90)</td>
<td></td>
</tr>
<tr>
<td>B·Truth</td>
<td>91 (83, 96)</td>
<td>91 (83, 96)</td>
<td>.69 (.59, .81)</td>
<td></td>
</tr>
<tr>
<td>C·Truth</td>
<td>94 (87, 98)</td>
<td>94 (87, 98)</td>
<td>.74 (.65, .84)</td>
<td></td>
</tr>
<tr>
<td>S·Truth</td>
<td>100 (99,100)</td>
<td>100 (99,100)</td>
<td>.93 (.87, .97)</td>
<td></td>
</tr>
</tbody>
</table>

(Examiner agreement in oral health)
Class 1 - predominantly unbiased

Classes 2 and 4 - one class of significantly associated with
  - Negative bias
  - Significantly associated with mid-tooth ($p = 0.03$) and buccal ($p = 0.016$) sites

Class 3 - positively biased on anterior teeth ($p = 0.028$)
### Examiner A

<table>
<thead>
<tr>
<th>ID</th>
<th>Max</th>
<th>Mand</th>
<th>Posterior</th>
<th>Anterior</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Tooth number**

| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

*Examined by Examiner A on September 26, 2011.*
Examiners B and C posterior clustering inference

- Examiner B - 2 classes
  - Dominant class - mildly negatively biased
  - In subordinate class, all recorded PPD = 0mm on mid-tooth sites
  - Possible association with mandibular sites ($p = 0.052$)
  - Significant association with shallow sites ($p = 0.012$)

- Examiner C - 2 classes
  - Dominant class - mildly negative biased
  - Subordinate class - mild positive bias
  - Possible association with anterior sites ($p = 0.052$)
  - Significant association with deeper sites ($p = 0.0007$)
Summary

- Model-based approach naturally incorporates multiple levels of dependence
- Borrowing strength
  - Corrects bias
  - Improves precision
- Accommodates estimation of agreement with truth
- Interpretation of DPMM classes for $\beta$s
  - Examiner specific
  - Target follow-up training
Acknowledgements

- Carlos Salinas, DMD and COBRE clinical core
- Sara Grossi, DDS, MS
- Keith Kirkwood, DDS, PhD and Steve London, DDS, PhD
- NCRR P20 RR17696, NIDCR U24 DE16508, NIDCR K25 DE016863