Assessment of DPOAE Test-Retest Difference Curves via Hierarchical Gaussian Processes

Tim Hanson

Dept. of Statistics
University of South Carolina

Graduate Colloquium
Northern Illinois University Division of Statistics
March 24, 2017

Cisplatin

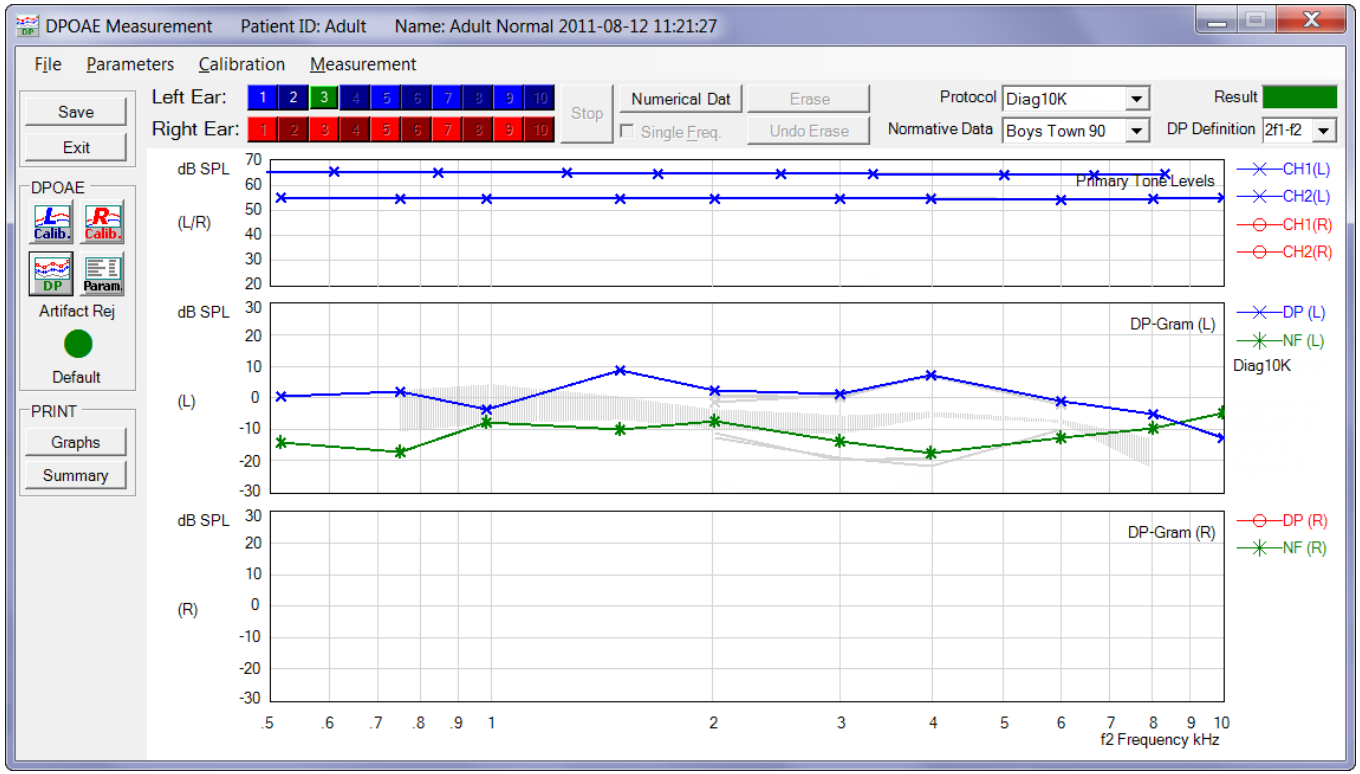
- Cisplatin: chemotherapeudic agent, treats many cancers.
- Can cause ototoxicity: inner ear poisoning & hearing loss.
- Cisplatin chemotherapy causes permanent hearing loss in approximately 70% of children and adolescents.
- Serial monitoring via hearing tests used to assess severe ototoxicity.
- Hearing tests difficult or impossible for very young or very ill cancer patients.

DPOAE

- Distortion production otoacoustic emissions (DPOAE) testing is a promising, non-invasive alternative to behavioral hearing tests.
- OAE elicited by sealing a small speaker & microphone in ear canal and playing tone through speaker.
- Pairs of tones (primary frequency ‘f2’ & secondary frequency) generate ‘distortion product’ OAE, or DPOAE, measured by microphone .
- Most common clinical protocol: play tones at successively increasing f2 and measure DPOAE.
- Generates ‘DP-gram’ that an audiologist can use to evaluate the health of the cochlea.

DPOAE testing on infant

DPOAE test result on healthy adult



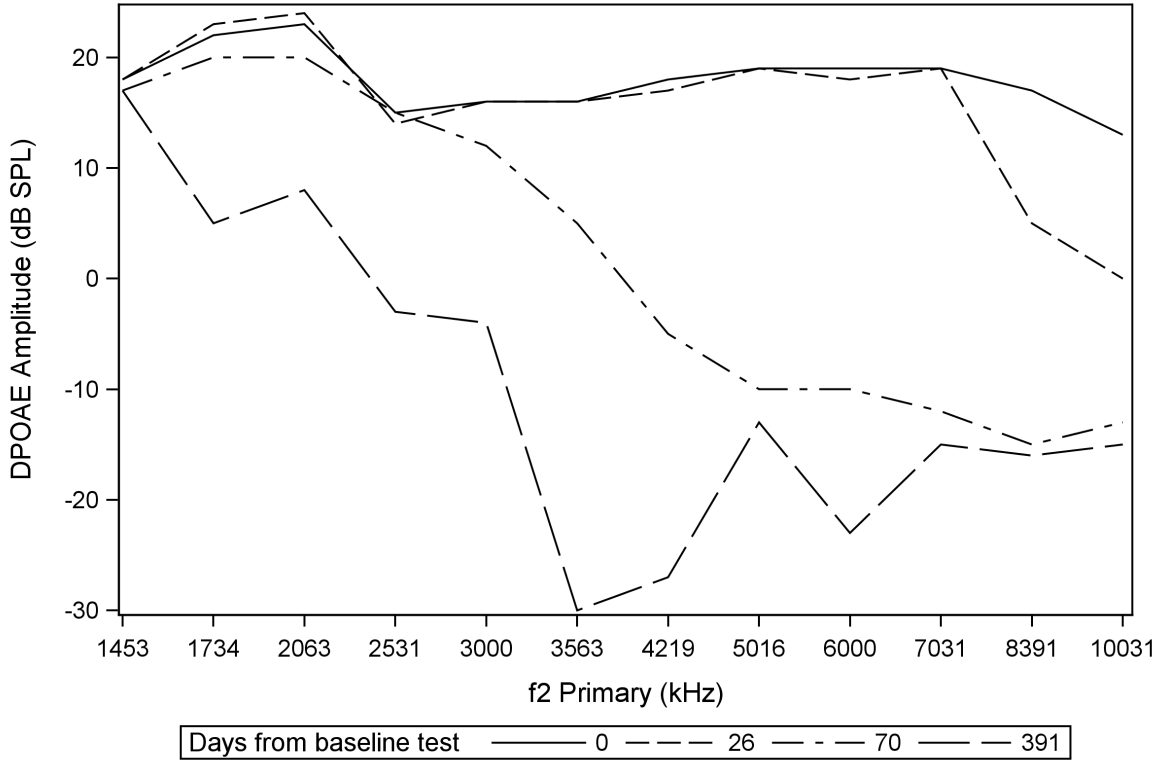
From Mimosa Acoustics webpage.

DP-grams

- DP-grams measured every 3, 4, or 6 weeks show how the cochlea is changing; if significant change observed, course of chemotherapty can be altered.
- Theoretically, each human has smooth DP-gram as a function of f2 at any given time and for a given ear.
- DP-grams change over time and from left to right ear.
- Currently six DPOAE systems in widespread use; typical f2's are 1, 2, 3, 4, 6, and 8 kHz, but others are used depending on system and user.
- Statistical problem: provide normal ranges for test-retest differences, i.e. difference in DP-grams from baseline to followup for normal healthy children.
- Challenge: DP-grams correlated across f2, time, and ear.

DP-grams: 1.5 year old treated with cisplatin

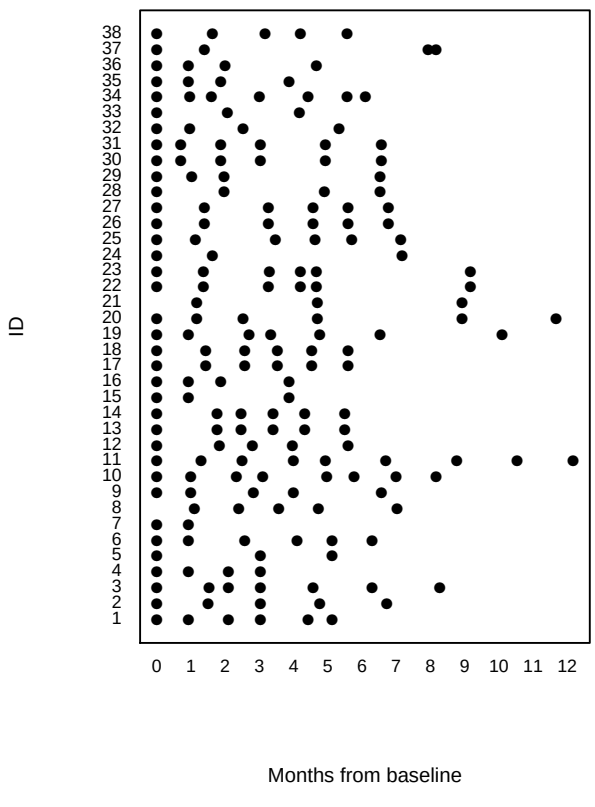
DP-grams for 18 month-old male cancer patient at baseline & about 4, 10, and 56 weeks later.



Data collected

- $n = 38$ healthy children aged 10 years or younger recruited from Oregon Health and Science University Doernbecher Children's Hospital between February 2006 and July 2009
- Subjects have normal hearing sensitivity; measurable DPOAEs; no history of ototoxic treatment, ear pathology, ear surgery, or tympanostomy tubes.
- Test sessions excluded for conductive hearing loss, abnormal tympanometry, or excessive subject noise or non-cooperation.
- DPOAES measured twelve f2 primaries from 1453 to 10031 Hz in half octave steps and using L2/L1 = 65/55 dB SPL and f2/f1 ratio of 1.22.
- Children retested at different times, at different frequencies, and possibly either one or both ears; high degree of unbalance.

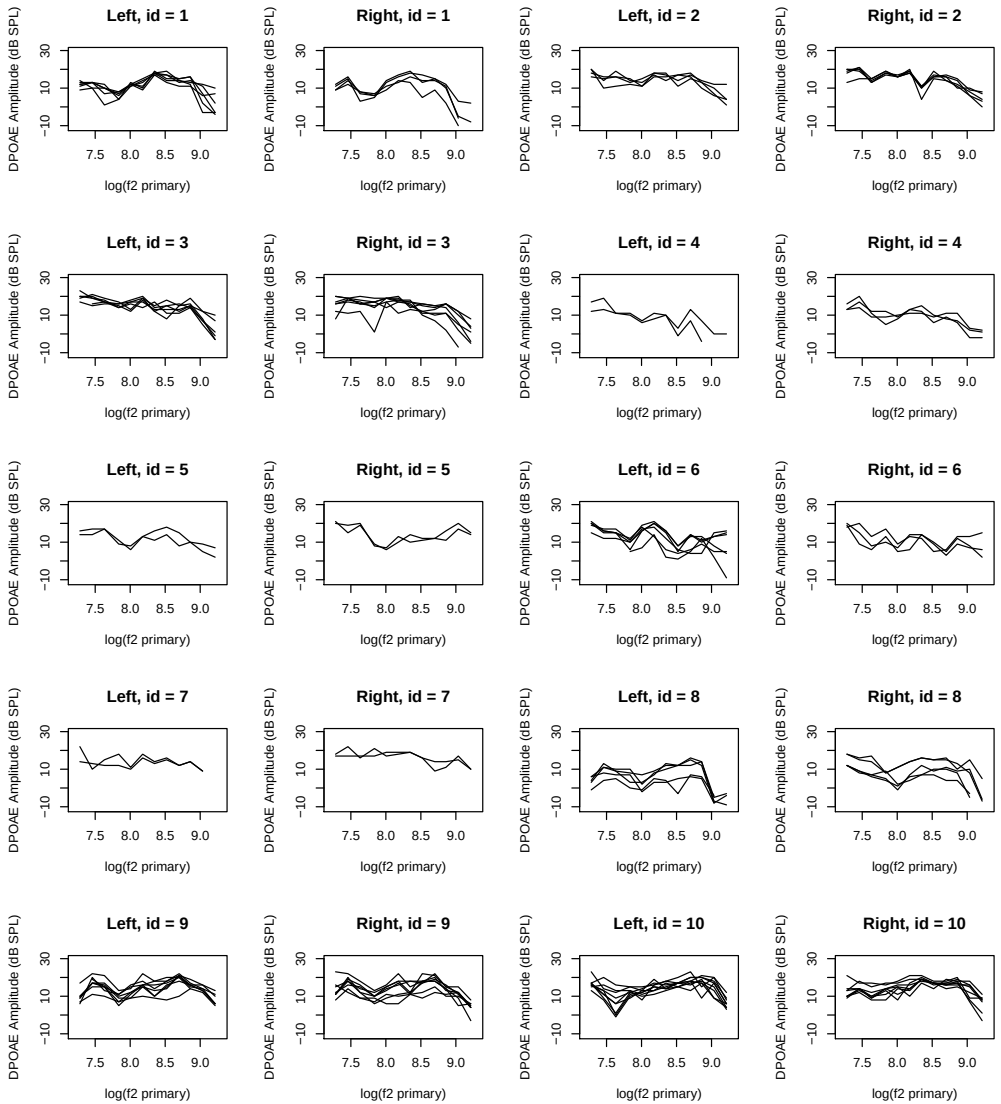
When followup DPOAE were collected



Features of the dataset

1. Two subjects provided no valid baseline data.
2. There is quite a bit of variation in the number of followups and the followup intervals.
3. Most data only cover up to about 7 months of followup.

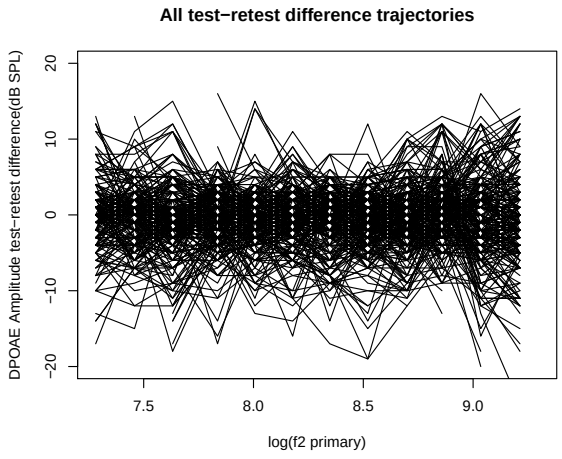
DP-grams for 10 subjects

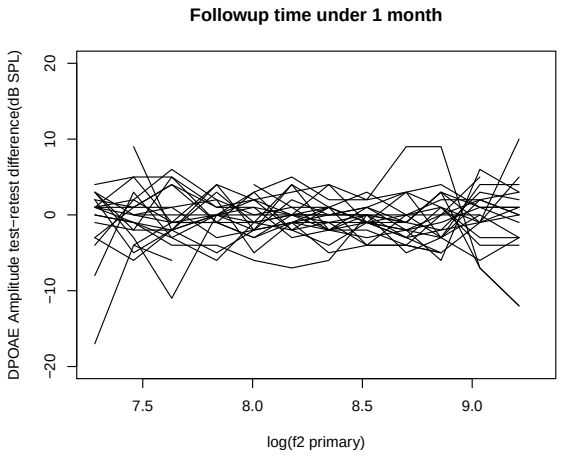


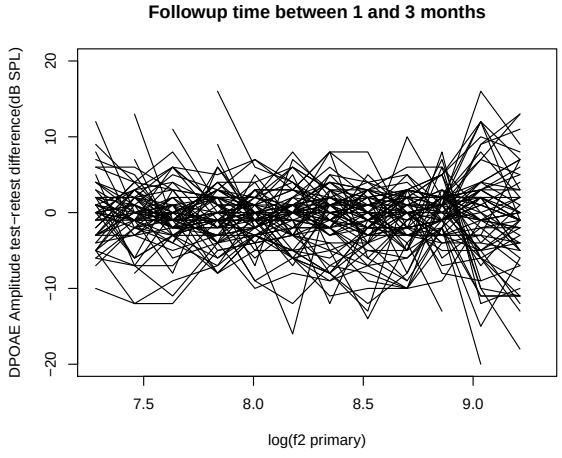
Features of the DP grams

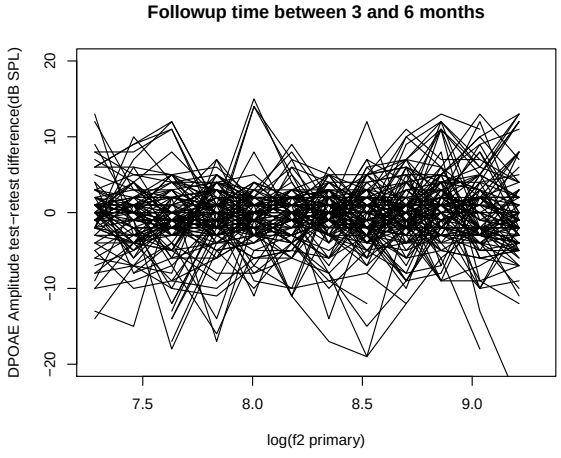
1. 'Intercept' & 'slope' quite different.
2. Overall common shape: decreasing-increasing-decreasing.
3. Strong correlation within subject over time & ear.
4. Variability remarkably constant within subject.

Test-retest differences by followup time









Data & notation

- Data collected over differing frequencies, at different followup times, and for one or both ears; indexing is a nightmare.
- Had to consider different indexing for different models. Hardest part: data manipulation & bookkeeping.
- $i = 1, \ldots, 38$ subjects.
- Subject i seen at *potentially* $j = 1, \ldots, 12$ different log-f2 $\mathbf{f} = (f_1, \ldots, f_{12})'$ over *all* followup times.
- Subject i observed at T_i times including baseline: $\mathbf{t}_i = t_{i1}, \ldots, t_{iT_i}$.

Gaussian processes

- Gaussian processes becoming very popular for modeling functions nonparametrically. Small number of parameters control smoothness properties.
- Nice video tutorial at

  `http://videolectures.net/gpip06_mackay_gpb/`
- Competitor to splines, neural networks, harmonic expansions; includes many approaches as special case.
- Main problem: computation $O(s^3)$. For us $s \leq 200$; usually much smaller.

Gaussian process in one dimension

- Stochastic process $e(t)$ s.t. the function $e(t)$ observed at $(t_1, \ldots, t_s)'$ is multivariate normal, e.g.

$$(e(t_1), \ldots, e(t_s))' \sim N_s\{\mathbf{0}, \Sigma(t_1, \ldots, t_s)\}.$$

- Only need covariance function $\sigma(s, t) = \text{cov}(e(s), e(t))$.
- Used here: squared exponential $\sigma(s, t) = \sigma^2 \exp(-\theta|s - t|^2)$. Smoothness parameter θ subject-specific later on.
- Generalizes to frequency & ear too: $e(t, f, l)$. Two surfaces in $\mathbb{R}^2$ for each subject, one for each ear.
- Since only a finite number of responses *can ever* be recorded, likelihood is product of multivariate normal kernels; easy to work with.

Hierarchical Gaussian process regression model

Consider mixed model

$$y_{ijkl} = \mu(f_j) + b_{i0} + b_{i1} f_j + e_{ijkl},$$

where

- $i = 1, \ldots, 38$ indexes subject.
- $j = 1, \ldots, 12$ indexes frequency level.
- $k = 1, \ldots, T_i$ indexes the visit time for subject i.
- $l = 1, 2$; $l = 1$ is left ear & $l = 2$ right.
- Overall pop'n curvy $\mu(f)$ plus subject specific line $b_{i0} + b_{i1} f$.
- e_{ijkl} is Gaussian process over f2, time, and ear for i observed at finite number of points.
- $E(b_{i0}) = \beta_0$, $E(b_{i1}) = \beta_1$ and $E(e_{ijkl}) = 0$.

Population mean $\mu(f)$ is penalized B-spline

Easy to work with in mixed model context!

$$\mu(f) = \sum_{s=1}^{S} \gamma_s \phi_s(f).$$

- Knots equispaced over range of log f2 primary levels in the data and $S = 20$ basis functions used.
- Since $\mu(f)$ includes constant or linear functions as special case, mean $\beta_0 + \beta_1 f + \mu(f)$ overspecified unless constraints introduced. Set two of the B-spline coefficients to zero, $\gamma_1 = \gamma_S = 0$ (Gray, 1992).

Population mean $\mu(f)$ is penalized B-spline

The B-spline parameters are $\gamma = (\gamma_2, \ldots, \gamma_{S-1})'$,given a 2nd-order random-walk prior

$$p(\gamma) \propto \lambda^{\frac{S-2}{2}} \exp\{-0.5\lambda \|\mathbf{D}\gamma\|^2\},$$

where $\mathbf{D}$ is a $(S-4) \times (S-2)$ penalty matrix. Following Lang and Brezger (2004), the penalty parameter λ follows

$$\lambda \sim \Gamma(\alpha_1, \alpha_2),$$

with $\alpha_1 = 1$ and $\alpha_2 = 0.005$ or 0.0005.

Building a linear model

Let

$$\mathbf{X}_{ijk} = \mathbf{1}_{L_{ijk}} \otimes (\phi_2(f_j), \ldots, \phi_{S-1}(f_j))$$
$$\mathbf{X}_{ij} = [\mathbf{X}'_{ij1} \cdots \mathbf{X}'_{ijT_i}]'$$
$$\mathbf{Z}_{ijk} = \mathbf{1}_{L_{ijk}} \otimes (1, f_j)$$
$$\mathbf{Z}_{ij} = [\mathbf{Z}'_{ij1} \cdots \mathbf{Z}'_{ijT_i}]'$$

Each child's vector of responses at frequency level f_j follows linear model

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\gamma + \mathbf{Z}_{ij}\boldsymbol{b}_i + \boldsymbol{e}_{ij},$$

for $i = 1, \ldots, 38$ and $j = 1, \ldots, 12$. These vectors are of differing lengths! L_{ijk} is 0, 1, or 2; number of ears looked at for subject i at frequency j & time t_{ik}.

Child-specific deviation from the population trend

- Each child's ear-specific response surface $y_{il}(t, f)$ deviates from the population mean $\beta_0 + \beta_1 f + \mu(f)$ by a smooth mean-zero surface in time and frequency $(b_{i0} - \beta_0) + (b_{i1} - \beta_1)f + e_{il}(t)$.
- Define $\boldsymbol{e}_{ij} = (\boldsymbol{e}'_{ij1}, \ldots, \boldsymbol{e}'_{ijT_i})'$ for child i at f2 level j. The Gaussian process model assumes

$$\boldsymbol{e}_{ij} \overset{ind.}{\sim} N_{n_{ij}}(\mathbf{0}, \Sigma_{ij}),$$

where Σ_{ij} is the covariance matrix of $\boldsymbol{e}_{ij}$ with separable covariance structure

$$\text{cov}(e_{ijkl}, e_{ijk'l'}) = \sigma_i^2 \exp\{-\theta_{ti}|t_{ijk} - t_{ijk'}|^2 - \theta_{ei}|l - l'|^2\}.$$

- If both ears are measured at the same time points at each f2 frequency level, subject-specific covariance is

$$\boldsymbol{e}_{ij} \sim N_{n_{ij}}(\mathbf{0}, \sigma_i^2 \Sigma_{ti} \otimes \Sigma_{ei}).$$

Subject-specific smoothness parameters and lines

- For each subject i, let $\boldsymbol{r}_i = (\boldsymbol{b}_i', \boldsymbol{v}_i')'$ where $\boldsymbol{b}_i = (b_{i0}, b_{i1})'$ and $\mathbf{v}_i = (\log(\sigma_i^2), \log(\theta_{ti}), \log(\theta_{ei}))'$.
- Based on preliminary non-hierarchical individual fits in SAS' `proc mixed`, multivariate normality is reasonable for $\boldsymbol{r}_i$:

$$\mathbf{r}_1, \ldots, \mathbf{r}_n \mid \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r \overset{iid}{\sim} N_5(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r), \tag{1}$$

where

$$\boldsymbol{\mu}_r = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\tau} \end{bmatrix}, \ \boldsymbol{\Sigma}_r = \begin{bmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_{bv} \\ \boldsymbol{\Sigma}_{bv}' & \boldsymbol{\Sigma}_v \end{bmatrix}$$

- Population parameters have prior

$$\boldsymbol{\mu}_r \sim N_5(\boldsymbol{m}_0, \boldsymbol{M}_0), \ \boldsymbol{\Sigma}_r^{-1} \sim \text{Wish}_5(\boldsymbol{Q}, q).$$

Hierarchical linear mixed model

$$\mathbf{y}_{ij}|\mathbf{b}_i, \mathbf{v}_i, \gamma \sim N_{n_{ij}}(\mathbf{X}_{ij}\gamma + \mathbf{Z}_{ij}\mathbf{b}_i, \Sigma_{ij}(\mathbf{v}_i)),$$

$$(\mathbf{b}_i, \mathbf{v}_i)|\mu_r, \Sigma_r \sim N_5(\mu_r, \Sigma_r).$$

Priors placed on μ_r, Σ_r, $\gamma|\lambda$, and λ.

Markov chain Monte Carlo

- Blocks of parameters have conjugate closed-form updates, other blocks updated via adaptive Metropolis-Hastings (Haario, Saksman, and Tamminen, 2001 & 2005). Details in paper.
- Fully 20,000 MCMC iterates were generated with the last 10,000 iterations used for posterior inference. Code written in FORTRAN 90 using IMSL library.
- During the last 10,000 iterations, a child's DP-gram was predicted from the *population*, consisting of responses corresponding to 31 log(f2 primary) levels.
- Based on these samples, both the pointwise and simultaneous 95% credible bands were generated for DP-grams of a randomly selected healthy child.

One observation time, volume tube method

Let

- $\boldsymbol{y}^* = (y_1^*, \ldots, y_{F*}^*)'$ be a vector of correlated responses from a random child drawn from the population at any time across the F^* frequencies $\boldsymbol{f}^* = (f_1^*, \ldots, f_{F*}^*)'$, for either ear
- $\mathbf{Z}_j^* = (1, f_j^*)$ and $\mathbf{Z}^* = [\mathbf{Z}_1^{*\prime} \cdots \mathbf{Z}_{F*}^{*\prime}]'$
- $\mathbf{r}^* = (b_0^*, b_1^*, \log(\sigma^{2*}), \log(\theta_t^*), \log(\theta_e^*))'$
- $\boldsymbol{\Sigma}^* = \sigma^{2*} \boldsymbol{I}_{F*}$

Hierarchical model $\Rightarrow$ random child's response sampled given $(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r, \gamma)$ by first sampling the subject-specific variables

$$\boldsymbol{r}^* \mid \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r \sim N_5(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r),$$

followed by sampling the DP-gram

$$\boldsymbol{y}^* \mid \boldsymbol{r}^*, \gamma \sim N_{F*}(\mathbf{X}^* \gamma + \mathbf{Z}^* \boldsymbol{b}^*, \boldsymbol{\Sigma}^*).$$

One observation time, volume tube method

- Due to linearity, the mean of any $\mathbf{y}^*$ is simply

$$\boldsymbol{\mu}^* = \mathbf{X}^* \bar{\boldsymbol{\gamma}} + \mathbf{Z}^* \bar{\boldsymbol{\beta}}$$

- At each f2 frequency level, the usual equal-tailed pointwise $(1-\alpha)100\%$ credible interval is formed yielding upper and lower pointwise interval endpoints $u_1, \ldots, u_{F^*}, l_1, \ldots, l_{F^*}$, which are well-approximated by

$$u_j = y_j^{*\lceil (1-\alpha/2)M \rceil} \text{ and } l_j = y_j^{*\lceil (\alpha/2)M \rceil}$$

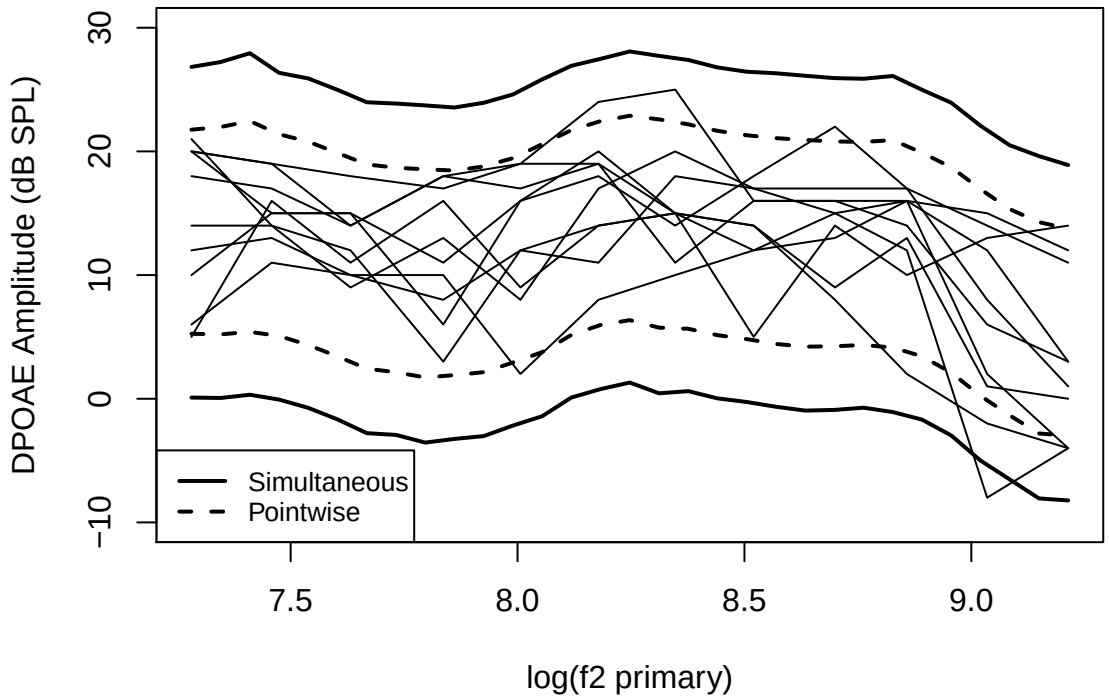
- Each pointwise interval (l_j, u_j) is adjusted by increasing $c > 1$ to

$$(\mu_j^* - c(\mu_j^* - l_j), \mu_j^* + c(u_j - \mu_j^*))$$

until exactly $(1-\alpha)100\%$ of the $\mathbf{y}^{*1}, \ldots, \mathbf{y}^{*M}$ lie between the two adjusted bands.

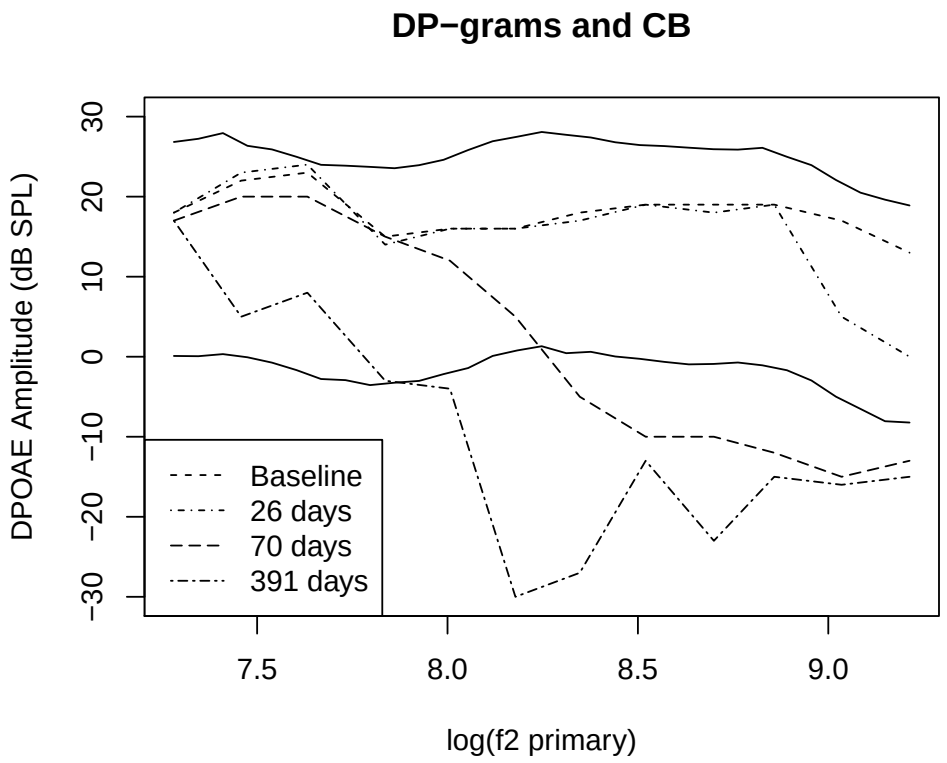
One observation only

95% credible bands (both pointwise and simultaneous) & 10 sample DP-grams from data:



One observation only

Actual cancer patient:



Reference chart for DP-gram test-retest difference

- A 95% reference interval corresponds to the range of DPOAE level shifts that a clinician can reasonably expect to see in a healthy population.
- Let $\mathbf{y}_1^* = (y_{11}^*, \ldots, y_{1F*}^*)'$ and $\mathbf{y}_2^* = (y_{21}^*, \ldots, y_{2F*}^*)'$ be sets of emissions recorded on the same frequencies at times t_1 and t_2, often baseline and then some months later.
- The difference at each frequency is given by the $F^* \times 1$ vector $\boldsymbol{\Delta} = \left[\begin{array}{cc} \boldsymbol{I} & -\boldsymbol{I} \end{array} \right] (\mathbf{y}_1^{*\prime}, \mathbf{y}_2^{*\prime})'$. A short calculation reveals that

$$\boldsymbol{\Delta} \sim N_{F*}\left(\mathbf{0}, 2(1 - \exp\{-\theta_t^* |t_1 - t_2|^2\})\boldsymbol{\Sigma}^*\right)$$

Posterior contour probabilities

- The simultaneous credible band provides a very quick check that a child's response is normal. However, it may miss DP-grams that are unusual in ways different than very high or low responses.
- Also useful to detecting abnormal test-retest differences.
- A contour probability measures how rare or unusual an observation is in a manner similar to a p-value.
- For continuous $\mathbf{y} \sim p(\cdot)$, the contour probability for seeing an observation more unusual than $\mathbf{y}_0$ is

$$P\{p(\mathbf{y}) < p(\mathbf{y}_0)\}$$

Posterior contour probabilities (continued)

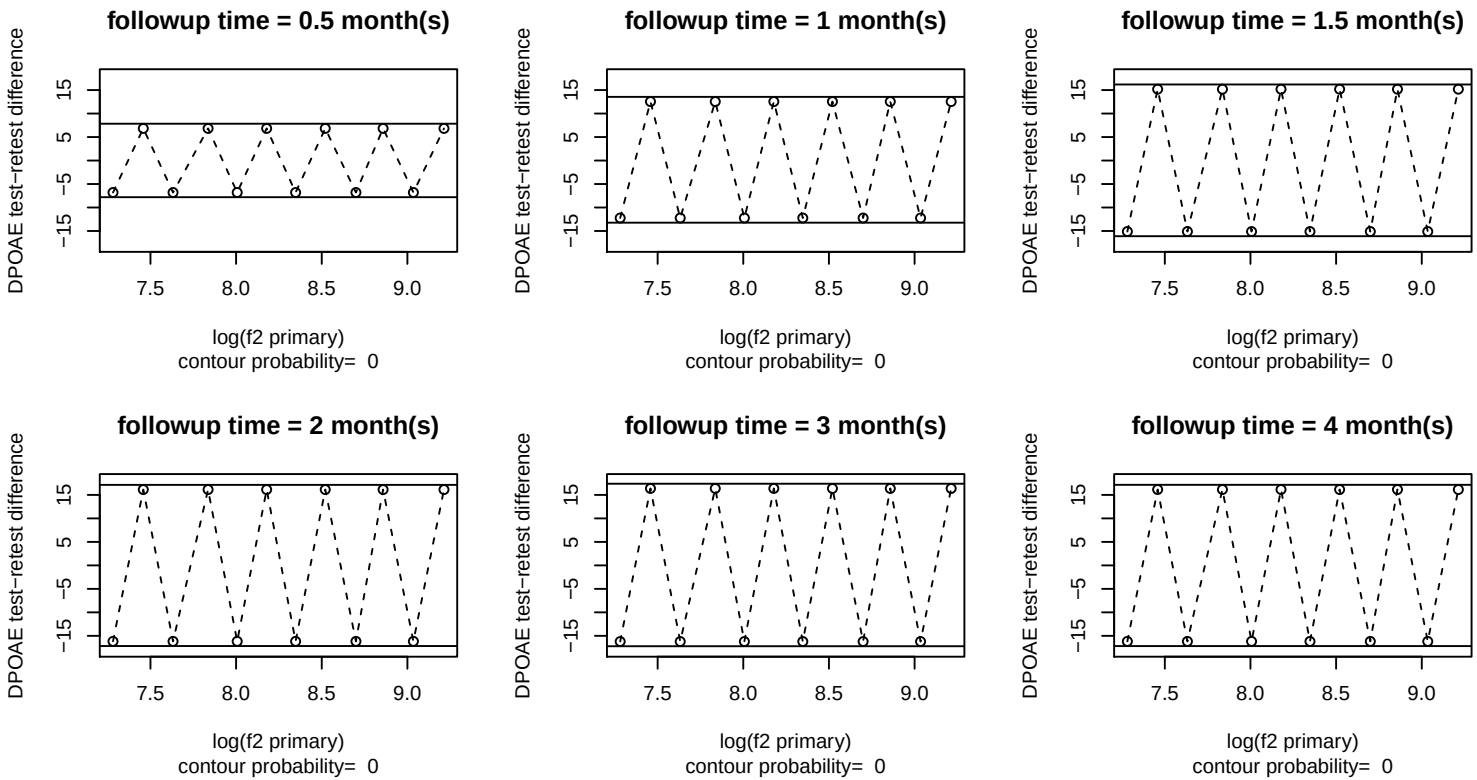
- For one set of measurements contour probability for $\mathbf{y}_0$ is

$$P\{p(\mathbf{y}^*) < p(\mathbf{y}_0)\} = \frac{1}{M}\sum_{m=1}^{M} P\{\chi^2_{F*} > (\mathbf{y}_0 - \mathbf{X}^*\boldsymbol{\gamma}^m - \mathbf{Z}^*\boldsymbol{b}^{*m})'[\boldsymbol{\Sigma}^{*m}]^{-1}(\mathbf{y}_0 - \mathbf{X}^*\boldsymbol{\gamma}^m - \mathbf{Z}^*\boldsymbol{b}^{*m})\}.$$

- Contour probability fordifference of two DP-grams taken at two different visits on the same ear, say $\boldsymbol{\Delta}_0$, is

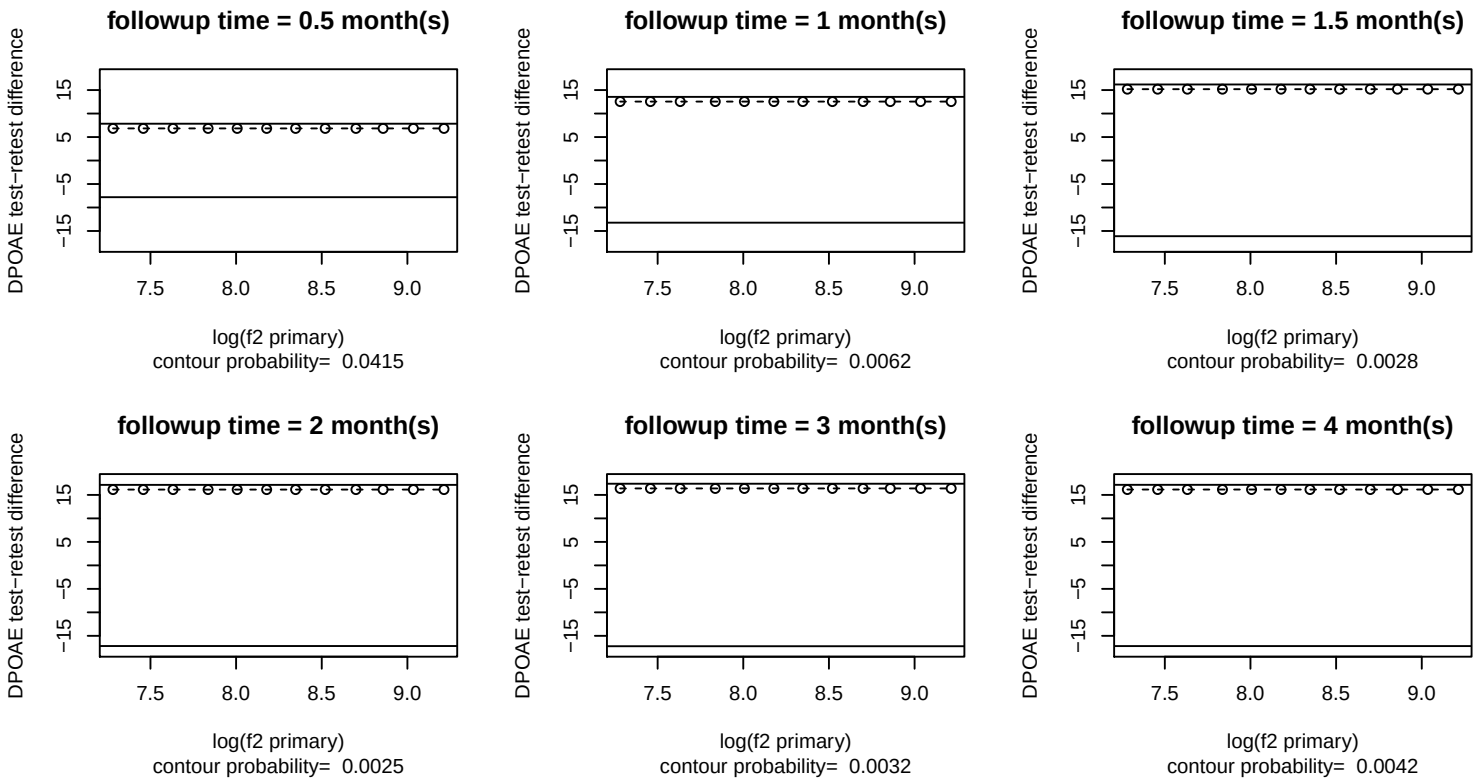
$$P\{p(\boldsymbol{\Delta}^*) < p(\boldsymbol{\Delta}_0)\} = \frac{1}{M}\sum_{m=1}^{M} P\{\chi^2_{F*} > \boldsymbol{\Delta}_0'[2(1 - e^{-\theta_t^{*m}|t_2 - t_1|^2})\boldsymbol{\Sigma}^{*m}]^{-1}\boldsymbol{\Delta}_0'\}$$

Test-retest differences within band that are unusual



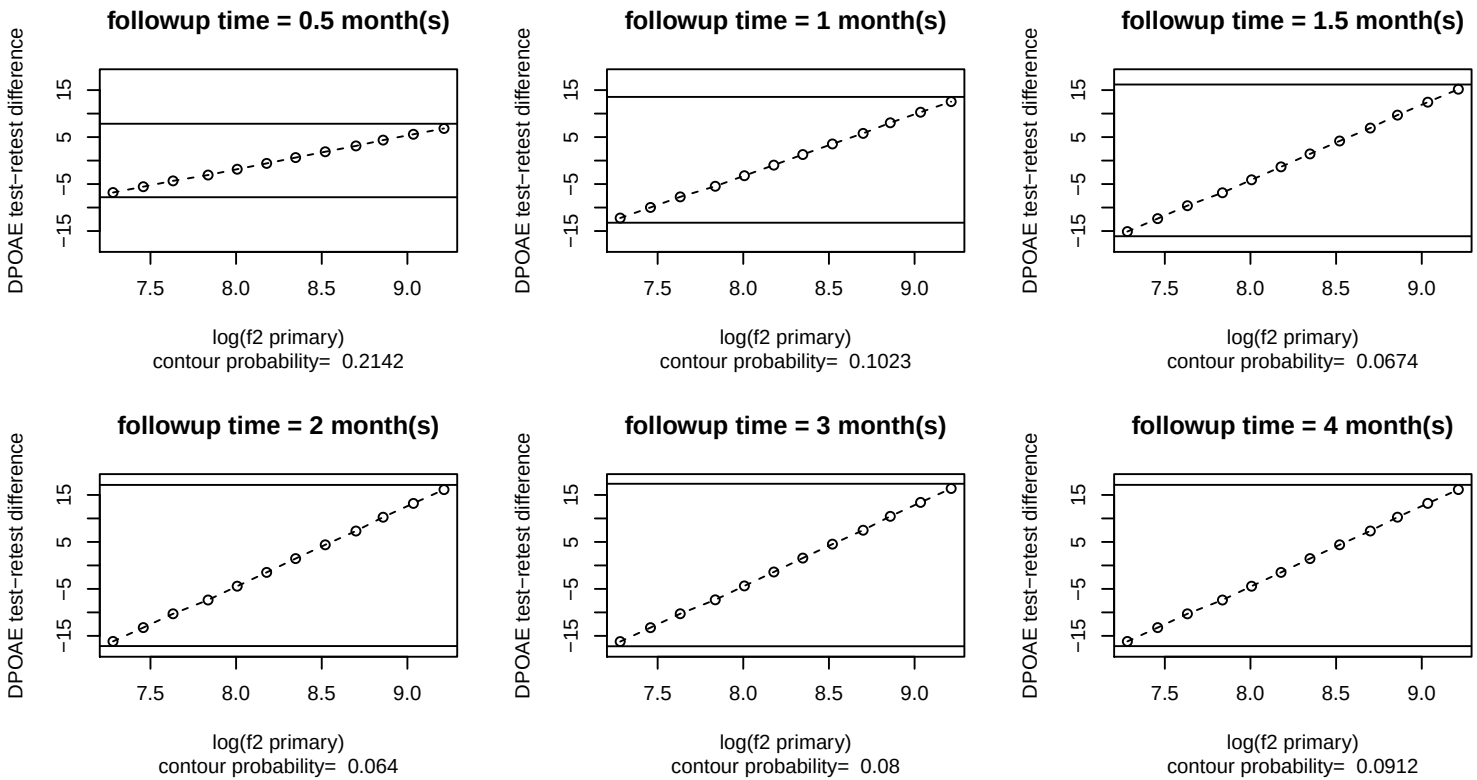
Differences that are too variable.

Test-retest differences within band that are unusual



Vertical shift that falls within band.

Test-retest differences within band that are unusual



DP-grams that cross in the middle.

Bands & contour probabilities for some trajectories

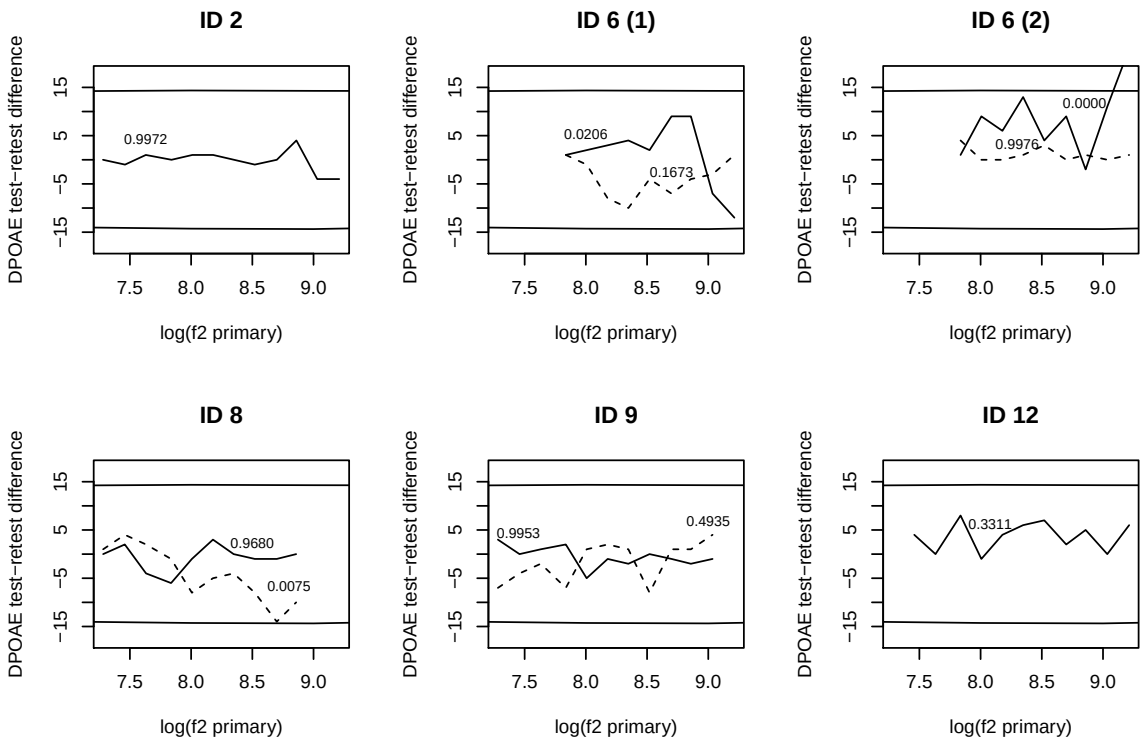
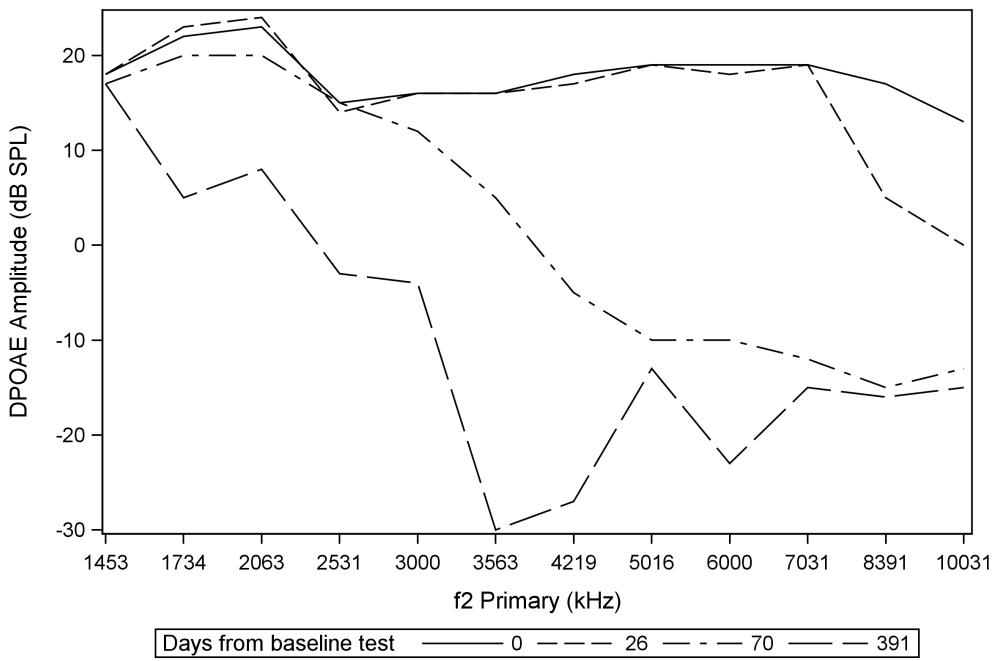


Figure: 10 sample DP-grams of test-retest differences of 5 children and 95% simultaneous credible band; followup time = 1 month.

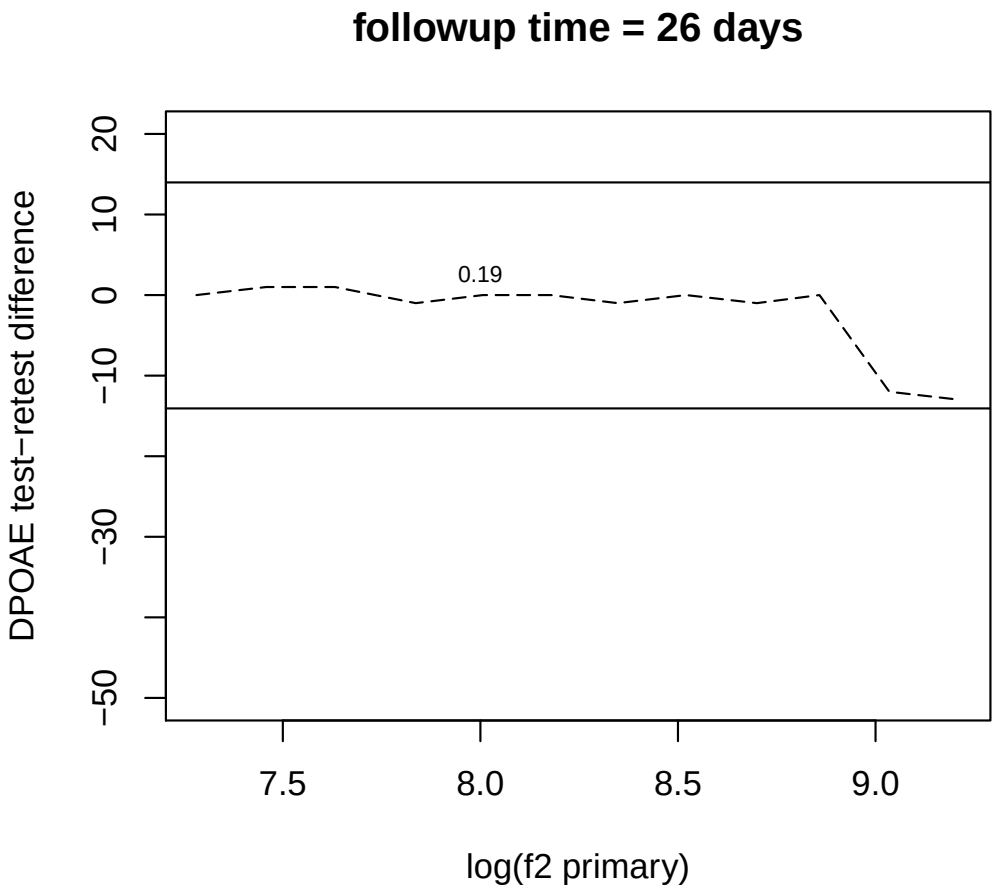
Data analysis: out-of-sample prediction for actual cancer patient

18 month-old male cancer patient's DP-grams from background; posterior mean contour probabilities at 26, 70, and 391 days after baseline are 0.19, 0.00, and 0.00 for the hierarchical model.



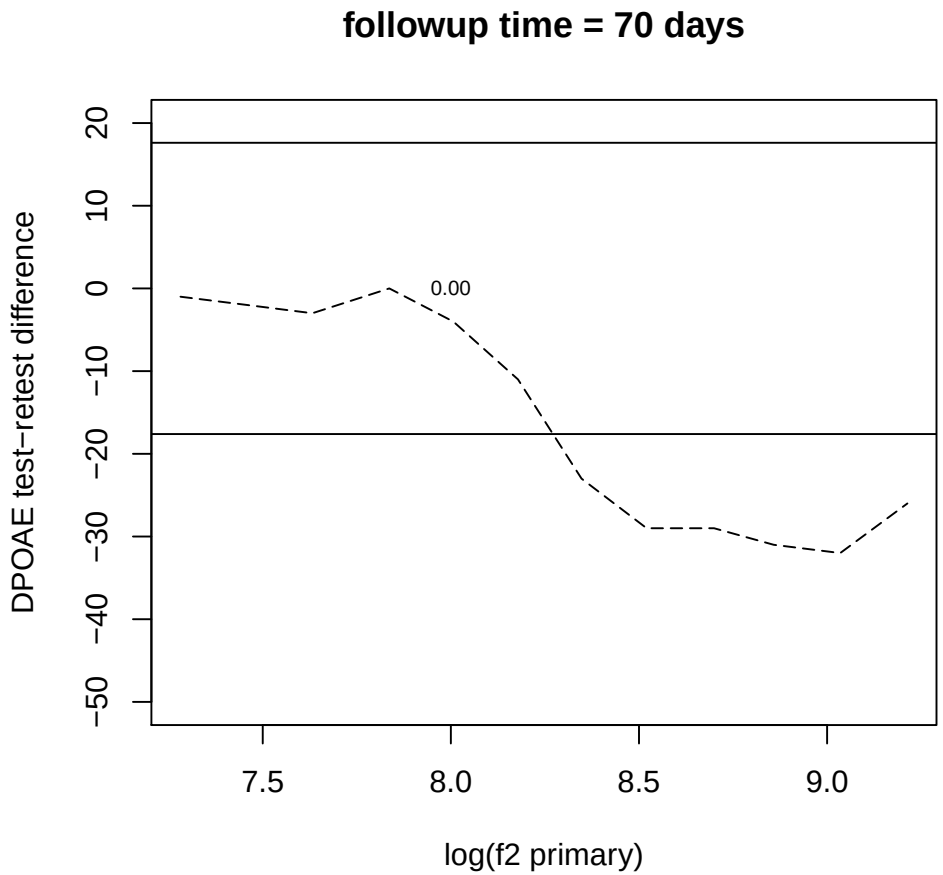
Test-retest difference

Actual cancer patient, first followup time.



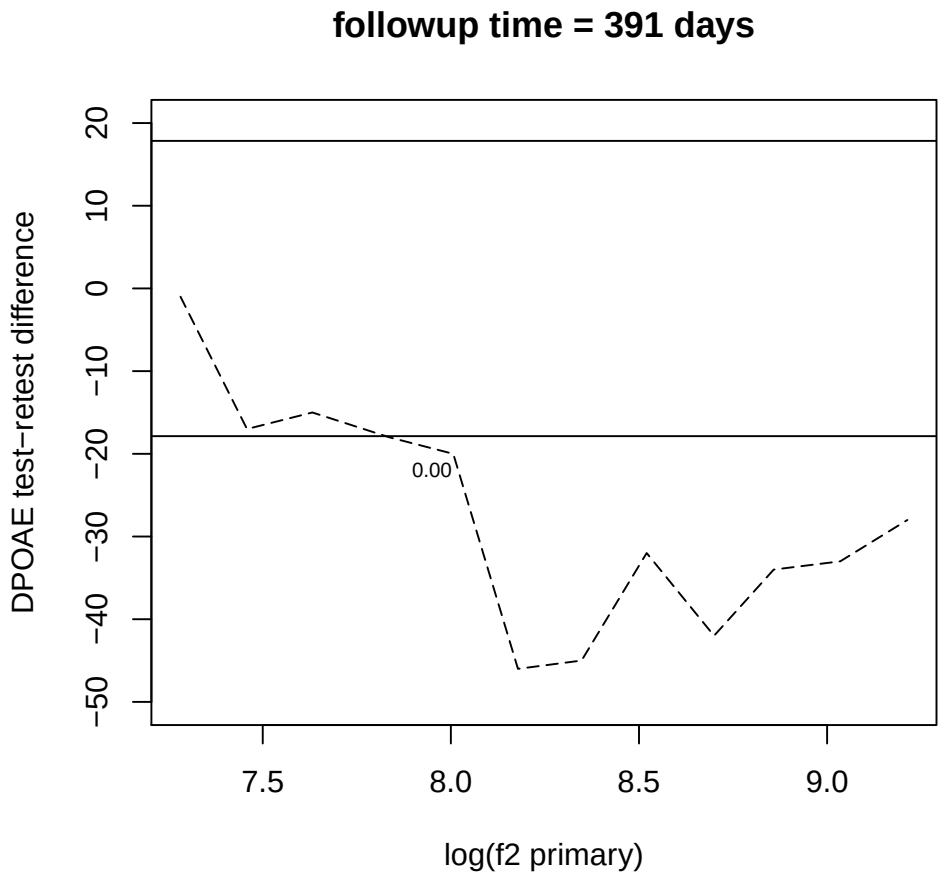
Test-retest difference

Actual cancer patient, second followup time.



Test-retest difference

Actual cancer patient, third followup time.



Age-gender specific model

- There is a well-known physiological basis for an age effect on OAE amplitude: DPOAE amplitude decreases over the first few years of life as the ear canal gets larger and the nervous system matures.
- Since DPOAE levels naturally change with cochlear development, it is desirable to have age-appropriate DPOAE level shift standards as necessary.
- In general, we allow intercepts, slopes, and all three subject-specific variance components to change smoothly with age and gender, yeilding a Gaussian process structural equation model.

Age-gender specific model (continued)

Let $\boldsymbol{a}_i$ be a $p \times 1$ vector of baseline covariates associated with child i; the hierarchical model becomes

$$\mathbf{r}_i \mid \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r \overset{ind}{\sim} N_5(\boldsymbol{\mu}_r \boldsymbol{a}_i, \boldsymbol{\Sigma}_r),$$

where

$$\boldsymbol{\mu}_r = \begin{bmatrix} \mathbf{b}' \\ \boldsymbol{\tau}' \end{bmatrix}, \; \boldsymbol{\Sigma}_r = \begin{bmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_{bv} \\ \boldsymbol{\Sigma}'_{bv} & \boldsymbol{\Sigma}_v \end{bmatrix}$$

and

$$\mathbf{b}' = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1p} \\ \beta_{21} & \cdots & \beta_{2p} \end{bmatrix} \text{ and } \boldsymbol{\tau}' = \begin{bmatrix} \tau_{11} & \cdots & \tau_{1p} \\ \tau_{21} & \cdots & \tau_{2p} \\ \tau_{31} & \cdots & \tau_{3p} \end{bmatrix}.$$

Data analysis: age-gender-specific model

- The age-gender-specific model was also fit to the DPOAE data.
- By allowing subject-specific intercept-slope and Gaussian process variance components to be covariate-dependent, the structural equation model may have better predictive power than the hierarchical one, provided that baseline covariate information is available.
- However, in this data analysis, the log-pseudo marginal likelihood (LPML) of the age-gender-specific model is almost the same as that of the hierarchical model.

Data analysis: age-gender-specific model (continued)

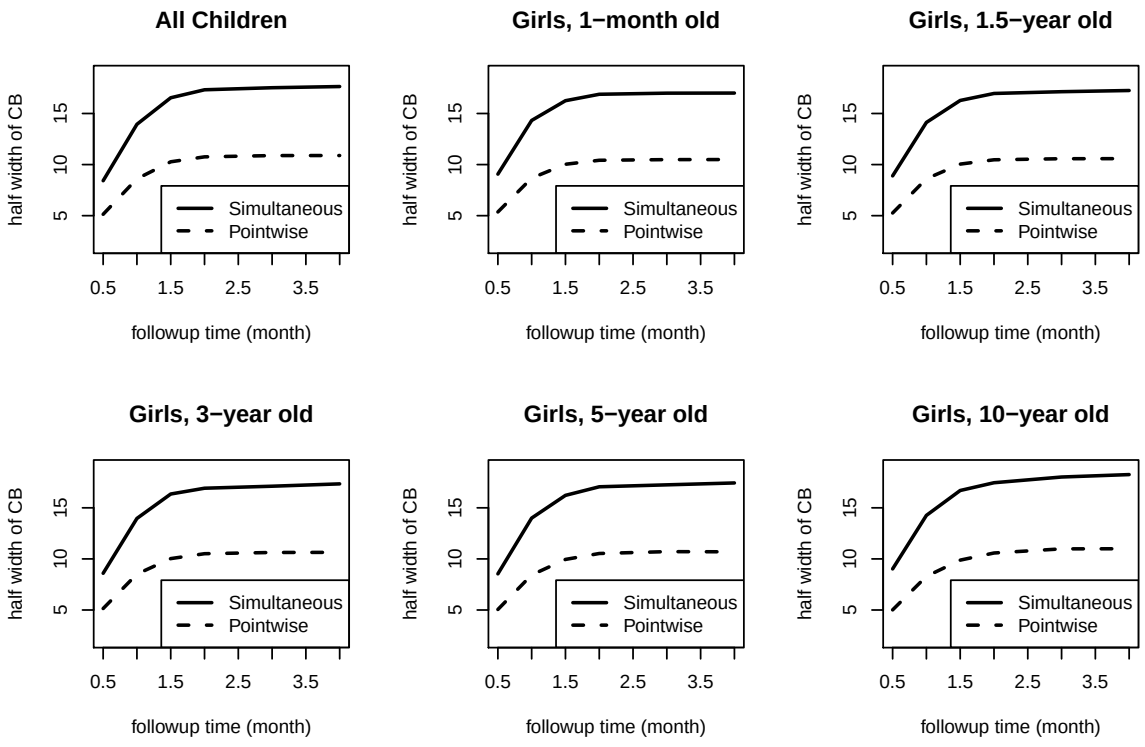


Figure: Half widths of credible bands of test-retest differences for all children and for girls.

Data analysis: age-gender-specific model (continued)

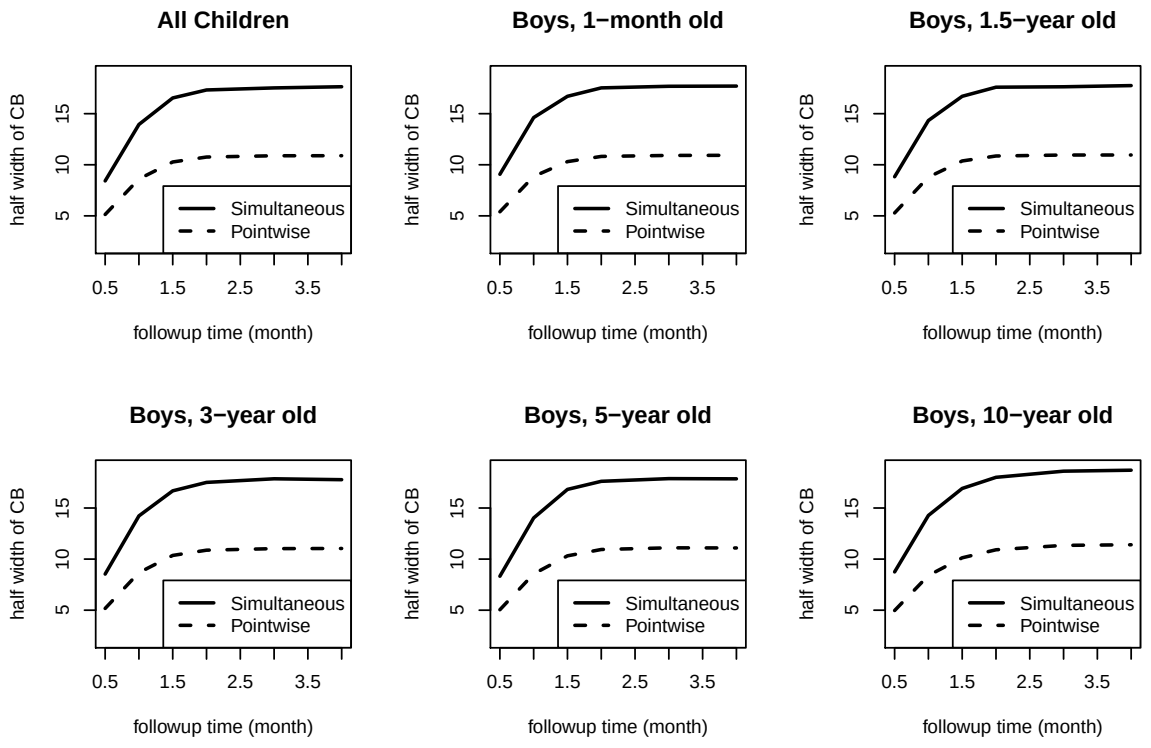


Figure: Half widths of credible bands of test-retest differences for all children and for boys.

Data analysis: age-gender-specific model (continued)

- The previous two figures show that as followup time increases, the credible band tends to be wider.
  - The width of the credible band increases quickly as followup time goes from half a month to two months.
  - After two months, the curve is essentially static, i.e. temporal correlation dies down to almost zero.
- As the children get older, the credible band tends to be wider, reflecting more variability in DPOAE response.
- Boys have wider credible bands than girls at the same age with the same followup time.

Other models

In addition to the two models mentioned previously, we fit four more models:

- Hierarchical model with correlation among f2 frequency levels, i.e. subject-specific surfaces $e_{il}(f, t)$.
- Age-gender-specific model with correlation among f2 primary frequency levels.
- Simple Laird and Ware (1982) linear mixed effects model with individual variances (can fit in `proc mixed` or R).
- Laird and Ware (1982) linear mixed effects model with common variance across all individuals (can fit in `proc mixed` or R).

Model comparison

The LPMLs of the six models:

	LPML
Age-gender-specific	-11785.56
Hierarchical	-11786.93
Age-gender-specific with correlation among f2	-11841.03
Hierarchical with correlation among f2	-11846.62
LMM with individual variances	-14288.11
LMM with common variance	-14723.34

Age-gender model with correlation in time and ear best. Adding correlation in frequency unnecessary and in fact adds noise. Simple mixed models perform very poorly.

Discussion

- Hierarchical & age-gender mixed models $\Rightarrow$ reference charts & contour probabilities for DPOAE test-retest ototoxicity assessment.
- Allows for subject-specific correlation (i.e. smoothness) in frequency, time, and ear coupled with subject-specific linear adjustment to $\mu(f)$.
- Joint work with Junshu Bao (Duquesne); Garnett McMillan and Kristin Knight (National Center for Rehabilitative Auditory Research).