Automatic Adjudication of Symptom-Based Exacerbations in Bronchiectasis Patients Treated With Azithromycin

Joint Statistical Meetings, Seattle, WA, 12th August 2015

Mark C. Wheldon¹,²
Alain C. Vandal¹,³, Anne-Cécile Bourien⁴, Lata Jayaram⁵, Noel Karalus⁶, Cecilia Tong², Hans Hockey⁷, and Conroy Wong³

¹Auckland University of Technology ²Middlemore Clinical Trials ³Counties Manukau Health ⁴ENSAI France ⁵Monash Health Sleep Lab ⁶Waikato District Health Board ⁷Biometrics Matters Ltd.
EMBRACE Trial


Articles

Azithromycin for prevention of exacerbations in non-cystic fibrosis bronchiectasis (EMBRACE): a randomised, double-blind, placebo-controlled trial

Conroy Wong, Lata Jayaram, Noel Karalus, Tam Eaton, Cecilia Tong, Hans Hockey, David Milne, Wendy Fergusson, Christine Tuffery, Paul Sexton, Louanne Storey, Toni Ashton

Summary

**Background** Azithromycin is a macrolide antibiotic with anti-inflammatory and immunomodulatory properties. We tested the hypothesis that azithromycin would decrease the frequency of exacerbations, increase lung function, and improve health-related quality of life in patients with non-cystic fibrosis bronchiectasis.

**Methods** We undertook a randomised, double-blind, placebo-controlled trial at three centres in New Zealand. Between Feb 12, 2008, and Oct 15, 2009, we enrolled patients who were 18 years or older, had had at least one pulmonary exacerbation requiring antibiotic treatment in the past year, and had a diagnosis of bronchiectasis defined by high-resolution CT scan. We randomly assigned patients to receive 500 mg azithromycin or placebo three times a week for 6 months in a 1:1 ratio, with a permuted block size of six and sequential assignment stratified by centre. Participants, research assistants, and investigators were masked to treatment allocation. The coprimary endpoints were rate of event-based exacerbations in the 6-month treatment period, change in forced expiratory volume in 1 s (FEV₁) before bronchodilation, and change in total score on St George’s respiratory questionnaire (SGRQ). Analyses were by intention to treat. This study is registered with the Australian New Zealand Clinical Trials Registry, number ACTRN12607000641493.

**Findings** 71 patients were in the azithromycin group and 70 in the placebo group. The rate of event-based exacerbations was 0·59 per patient in the azithromycin group and 1·57 per patient in the placebo group in the 6-month treatment period (rate ratio 0·38, 95% CI 0·26–0·54; p<0·0001). Prebronchodilator FEV₁ did not change from baseline in the
Objectives  Assess the effect of *azithromycin* on health-related quality of life and lung function in patients 18–80 years with bronchiectasis (diagnosed by CT scan).

Design  Multicenter (3), double-blind, placebo-controlled, parallel group (1:1), 141 pts total.

Intervention  500mg azithromycin capsule vs. placebo, 3 days per week, for six months.

1° Endpoints  i) Rate of Event Based Exacerbations (EBEs) over 6 mo. treatment period;
ii) Change in St. George’s Respiratory Questionnaire (tot. score); (+ others).

2° Endpoints  Symptom scores for: sputum purulence, sputum volume, dyspnoea; (+ others).
Exacerbations

- Patients exacerbations and symptom scores recorded prospectively in patient diaries.
- Each patient-day judged *exacerbation* or *no exacerbation*.
- Key symptoms of an exacerbation are
  - Sputum volume
  - Sputum purulence (colour)
  - Dyspnoea (shortness of breath, coughing).
- Two types: Event-based (EBE) and Symptom-based (SBE).
  - *Ascertainment of EBE requires contact with clinician.*
  - SBE is determined from patient diary data.
Figure: Diarized days by Event-based exacerbations status for each EMBRACE location, all patient-days.
Goal

Automatic adjudication of symptom-based exacerbations

New Definition of SBEs

- Adjudication of SBEs originally done by manual review of diaries.
- Automation of SBE adjudication presented opportunity to revisit definition.
- New definition of SBE to be based on a prediction rule validated against the clinically adjudicated EBEs.

Validated Prediction Rule

- Build a regression model for $\text{EBE}_t$ using symptom scores and EBE at times $t \in [t - \delta, t_0)$. 
Data

- Observed ~ 50 000 patient-days observed on 141 patients across 3 centres over 6 months.
- Patients rated severity of
  - Sputum purulence, sputum volume, dyspnoea on a validated 5-point scale, 0 “no symptom” → 4 “very much”.

Table: Example data

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>Pat. Day</th>
<th>EBE</th>
<th>SP</th>
<th>SV</th>
<th>DY</th>
<th>SBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data

- Observed ~ 50,000 patient-days observed on 141 patients across 3 centres over 6 months.
- Patients rated severity of
  - Sputum purulence, sputum volume, dyspnoea
  on a validated 5-point scale, 0 “no symptom” → 4 “very much”.

<table>
<thead>
<tr>
<th>Table: Example data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptoms</strong></td>
</tr>
<tr>
<td>Pat.</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>:</td>
</tr>
</tbody>
</table>
Goal is a Model for Prediction

- Statistical goal is a model to predict a time-ordered, clustered, binary outcome, \( EBE_{i,t} \).
- Selected GLM, logit link, random intercepts for patient.

\[
\text{logit } \Pr(EBE_{i,t} = 1 | \cdot ) = x'_{i,t} \beta + z_i b_i + \epsilon_{i,t}
\]

\( b_i \sim \text{Normal}(0, \tau^2) \perp \epsilon_{i,t} \sim \text{Normal}(0, \sigma^2) \)

(columns of \( X \) are symptom scores and \( EBE_{i,t-\delta} \))

\[
\hat{EBE}_{i,t} = \begin{cases} 1 & \text{if } \hat{Pr}(EBE_{i,t} | \cdot ) > c \\ 0 & \text{if } \hat{Pr}(EBE_{i,t} | \cdot ) \leq c \end{cases}
\]

(\( c \) chosen such that sensitivity = specificity)
- Design parameters are: \( X, \delta, c \).
Method Overview

1. Build a “retrospective” prediction model for $\text{EBE}_t$ using
   - symptom scores
   - observed EBE status at times $t \in [t - \delta, t_0)$.

2. Convert to a “prospective” model for $\text{EBE}_t$ using
   - retrospective design
   - predicted EBE status at times $t \in [t - \delta, t_0)$.

Design

- Fixed effect design matrix, $X$, contains
  - patient-specific symptom scores at a *contemporaneous* time interval, $t \in [-a, t_0]$.
  - patient-specific symptom scores at an *earlier* time interval, $t \in [-c, -b]$, $a < b < c$.
  - EBE status at an earlier time point, $\text{EBE}_{i, t-\delta}$.
- For any given choice of $a, b, c$, symptom scores can be averaged, or not, over the intervals.
72 combinations of $\delta$, $a$, $b$, $c$, and averaging schemes were defined based on existing “by-hand” adjudication methods.

$$2 \times 4 \times 3 \times 3 = 72 \text{ models}$$
Design

72 combinations of $\delta$, $a$, $b$, $c$, and averaging schemes were defined based on existing “by-hand” adjudication methods.

$$\begin{align*}
\text{contemporaneous window} & \times \\
\text{earlier window} & \times \\
\delta & \times \\
\text{avg. schemes} & = 72 \text{ models}
\end{align*}$$

In-sample predictive performance of each compared using AUC.

<table>
<thead>
<tr>
<th>Avg. Scheme</th>
<th>Contemp. Window</th>
<th>Comp. Window</th>
<th>$\delta$</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$[-3, 0]$</td>
<td>$[-11, -7]$</td>
<td>5</td>
<td>0.84</td>
</tr>
<tr>
<td>2</td>
<td>$[-3, 0]$</td>
<td>$[-11, -7]$</td>
<td>10</td>
<td>0.85</td>
</tr>
<tr>
<td>3</td>
<td>$[-3, 0]$</td>
<td>$[-8, -4]$</td>
<td>5</td>
<td><strong>0.97</strong></td>
</tr>
</tbody>
</table>
Selected Model

- Averaging scheme consists in arithmetic average of symptom scores over the interval.
- Includes all two-way interactions between \( \text{EBE}_{i,t-5} \) and averaged symptom scores.

\[
\eta_{i,t} \equiv \logit \Pr(\text{EBE}_{i,t} = 1|\text{EBE}_{i,t-5}, X_{i,t}, b_i) \\
= \beta_0 + \text{EBE}_{i,t-5} \times \\
\left( SV_{i,\text{cont.}} + SP_{i,\text{cont.}} + DY_{i,\text{cont.}} + SV_{i,\text{earl.}} + SP_{i,\text{earl.}} + DY_{i,\text{earl.}} \right) \\
+ b_i + \epsilon_{i,t}
\]

\[
\eta_{i,t} \in (-\infty, \infty), \quad \text{EBE}_{i,t} \in \{0, 1\}, \quad t = 0, 1, \ldots, T, \\
\epsilon_{i,t} \sim \text{Normal}(0, \sigma^2) \quad \perp \quad b_i \sim \text{Normal}(0, \tau^2)
\]
Prospective Prediction

- The selected model is retrospective in that today’s prediction depends on earlier observed EBEs.

\[
\text{logit } \Pr(\text{EBE}_{i,t} = 1 | \text{EBE}_{i,t-5}, \cdot) = x_i'\beta + z_i b_i + \epsilon_{i,t}
\]

- We want a prospective model that uses earlier predictions to make today’s prediction.

\[
\text{logit } \Pr(\text{EBE}_{i,t} = 1 | \hat{\text{EBE}}_{i,t-5}, \cdot) = w_i'\beta + z_i b_i + \epsilon_{i,t}
\]
A Model for Prediction

Goal

- Recall that our goal is a model for prediction.
- ⇒ propose a model (somehow!).
- Verify it has good predictive power.

Two-fold Cross-validation

- Split data into a training set and a hold-out set for validation.
- Randomly select 70 percent of the patients and allocate all their observations to the training set.
- The remainder go into the hold-out set.
Sequential Approach

1. Initialize by generating retrospective predictions, \( \hat{EBE}_{i,1}^{[ret]}, \ldots, \hat{EBE}_{i,5}^{[ret]} \), using \( \hat{\beta} \) and threshold, \( c^{[ret]} \), from the retrospective model.

Using the training set:

2. Sequentially generate prospective predictions \( \hat{EBE}_{i,6}^{[pro]}, \hat{EBE}_{i,7}^{[pro]}, \ldots \).
   - Use “population level” predictions
   - Use \( c^{[ret]} \) to threshold the predicted probabilities (we have to because this is the only \( c \) currently available).

3. Re-estimate the binary threshold, \( c^{[pros]} \), using \( \hat{EBE}_{i,t}^{[pro]} \).

Using the hold-out set:

4. Repeat 2 using \( c^{[pros]} \).
## Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>$c$ (used)</th>
<th>$c$ (Opt.)</th>
<th>Sens. (%)</th>
<th>Spec. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Retro.</td>
<td>0.093</td>
<td>0.093</td>
<td>90</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>Pros.</td>
<td>0.093</td>
<td>0.048</td>
<td>76</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>Pros.</td>
<td>0.048</td>
<td>0.048</td>
<td>83</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>Hold-out Pros.</td>
<td>0.048</td>
<td>—</td>
<td>90</td>
<td>79</td>
<td></td>
</tr>
</tbody>
</table>

Our prospective SBE predictor misses 10 percent of the EBEs (1 in 10) and calls an EBE 21 percent of the time there isn’t one (1 in 5). Relative to using the whole dataset, estimated sensitivity is equal, and specificity is 86 percent (↓13 percentage points).
### Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>$c$ (used)</th>
<th>$c$ (Opt.)</th>
<th>Sens. (%)</th>
<th>Spec. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>Retro.</td>
<td>0.093</td>
<td>0.093</td>
<td>90</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Prosp.</td>
<td>0.093</td>
<td>0.048</td>
<td>76</td>
<td>88</td>
</tr>
</tbody>
</table>

Our prospective SBE predictor misses 10 percent of the EBEs (1 in 10) and calls an EBE 21 percent of the time there isn’t one (1 in 5). Relative to using the whole dataset, estimated sensitivity equal, specificity is 86 percent (↓13 percentage points).
## Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>$c$ (used)</th>
<th>$c$ (Opt.)</th>
<th>Sens. (%)</th>
<th>Spec. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>Retro.</td>
<td>0.093</td>
<td>0.093</td>
<td>90</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Prosp.</td>
<td>0.093</td>
<td>0.048</td>
<td>76</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Prosp.</td>
<td>0.048</td>
<td>0.048</td>
<td>83</td>
<td>83</td>
</tr>
</tbody>
</table>

Our prospective SBE predictor misses 10 percent of the EBEs (1 in 10) and calls an EBE 21 percent of the time there isn’t one (1 in 5). Relative to using the whole dataset, estimated sensitivity is equal, while specificity is 86 percent (↓13 percentage points).
## Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>$c$ (used)</th>
<th>$c$ (Opt.)</th>
<th>Sens. (%)</th>
<th>Spec. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>Retro.</td>
<td>0.093</td>
<td>0.093</td>
<td>90</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Prosp.</td>
<td>0.093</td>
<td>0.048</td>
<td>76</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Prosp.</td>
<td>0.048</td>
<td>0.048</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>Hold-out</td>
<td>Prosp.</td>
<td>0.048</td>
<td>—</td>
<td>90</td>
<td>79</td>
</tr>
</tbody>
</table>

- Our prospective SBE predictor,
  - misses 10 percent of the EBEs (1 in 10)
  - calls an EBE 21 percent of time there isn’t one (1 in 5).

- Relative to using the whole dataset,
  - Estimated sensitivity equal,
  - Specificity is 86 percent (↓ 13 percentage points).
Patient-reported Wellbeing

- Ultimately, we’re interested in patient wellbeing, and a patient-centred measure that is sensitive to changes in physical state.
- Patients also reported wellbeing each day using a 1–5 scale (SGRQ).
- How is our new definition of SBE associated with wellbeing?
- Do we get “closer” to wellbeing with SBE rel. to EBE?
EBE, SBE, and Wellbeing

- Consider EBE and SBE as “tests” for wellbeing,
  - i.e., sens. = Pr(EBE = 1 | WB ≤ 2).
- Wellbeing is dichotomized between 2 and 3.

<table>
<thead>
<tr>
<th>bad</th>
<th>good</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2</td>
<td>3 4 5</td>
</tr>
</tbody>
</table>

(so indicator of “bad” wellbeing corresponds to indicator of “bad” EBE/SBE)
## Wellbeing and EBE

### EBE vs. wellbeing_leq2

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Percent</th>
<th>Col Pct</th>
<th>0</th>
<th>1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>24274</td>
<td>16579</td>
<td>40853</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>52.36</td>
<td>35.76</td>
<td>88.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>83.44</td>
<td>96.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4818</td>
<td>685</td>
<td>5503</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.39</td>
<td>1.48</td>
<td>11.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16.56</td>
<td>3.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29092</td>
<td>17264</td>
<td>46356</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>62.76</td>
<td>37.24</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Spec. = 83%.
- Sens. = 4%.
- Most (96%) episodes of poor wellbeing do not correspond to an EBE.
### Wellbeing and SBE

<table>
<thead>
<tr>
<th></th>
<th>SBE</th>
<th>wellbeing_leq2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Col Pct</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>0</td>
<td>17414</td>
<td>14567</td>
</tr>
<tr>
<td></td>
<td>40.18</td>
<td>33.61</td>
</tr>
<tr>
<td></td>
<td>63.51</td>
<td>91.48</td>
</tr>
<tr>
<td>Total</td>
<td>27420</td>
<td>15923</td>
</tr>
<tr>
<td></td>
<td>63.26</td>
<td>36.74</td>
</tr>
</tbody>
</table>

▶ Well ... we’ve doubled the proportion of bad wellbeing days captured!
▶ Spec. ↓ 20 percentage points.
▶ Work in progress!
Exacerbations are an outcome of interest in the study of bronchiectasis.

Ascertainment of event-based exacerbations (EBEs) requires clinical assessment.

Symptom-based exacerbations (SBEs) are ascertained from patient-reported symptom scores and exacerbation history, coded “by hand”.

We used logistic regression to develop an “automatic” coding scheme; changes in symptoms that are associated with changes in physical state (EBE).

As a classifier of EBE the performance was quite good (sens. 90%, spec. 79%).

Unclear we moved closer to patient-reported wellbeing (SGRQ) . . . to be continued.