**Background and Stats questions**

*Our study aims*

Our study mainly focuses on

1. the carriage of macrolide resistance genes in patients with chronic lung disease and their close contacts
2. the effect of long-term macrolide exposure on the transmission risk of macrolide resistance genes between patients and their healthy close contacts

*Specific clinical questions we aim to address the second study aim*

**Question1. Macrolide resistance gene co-carriage:**

In each treatment group, whether close contacts have/don’t have this gene is dependent/independent of the detection of this gene in their paired patients

**Question2. Macrolide effect on transmission risk:**

As compared with non-macrolide group, whether the macrolide group is more likely to have more transmission cases/less no transmission case?

*Our models in the initial manuscript*: Binary logistic regression

1. Model 1: Question 1
   1. **Independent variable (IV):** detection of the gene in patients (0/1)
      1. 0: this gene is undetected
      2. 1: this gene is detected
   2. **Dependent variable (DV):** detection of the gene in paired close contact (0/1)
      1. 0: this gene is undetected
      2. 1: this gene is detected
2. Model 2: Question 2
3. **Independent variable (IV):** macrolide exposure (Yes/No)
   * 1. Yes: patient received macrolide therapy
     2. No: patient did not receive macrolide therapy
4. **Dependent variable (DV):** Transmission/No transmission case: 0/1
   * 1. 1: transmission case (both patient and close contact have this gene = 1-1 pair)
     2. 0: no transmission case (either patient or close contact have the gene: 1-0/0-1 pair)

*Comments from Statistical reviewer:*

*The primary objective of this prospective study is to estimate and compare macrolide resistance gene detection rates and abundances between MR, MNR, MRCC, and MNRCC cohorts. Study design, data collection, primary endpoints, and statistical methods were clearly described. Logistic regression models were used to associate cohorts with detection and transmission status. Results were presented adequately and clearly. Conclusions were drawn appropriately. I have one minor suggestion (not a concern or question):*

*Seems authors could try logistic GEE model to estimate and compare 4 cohorts (MR, MNR, MRCC, and MNRCC) in one model with respect of resistant genes detection. GEE model would be more efficient than first comparing within treatment pairs then between treatment groups*

Key points of his/her suggestion:

1. Try logistic GEE model
2. In one model
3. More efficient

My simple answer:

1. Did try GEE model:🗸
2. Could address our two clinical questions in one model: ×

What clinical questions does GEE model addressed in one model:

1. Question 1. Macrolide resistance gene co-carriage: 🗸
   1. However I cannot separate into two treatment group as treatment group is the second IV in this model
2. Question 2. Macrolide effect on transmission risk: ×
   1. Macrolide effects on gene detection in close contact: macrolide exposure affect the detection of the gene in close contact?

Explanation:

1. “As compared with non-macrolide group, whether the macrolide group is more likely to have more transmission cases?” ≠ “whether macrolide exposure affect the detection of the gene in close contact?”
2. To address the second question (transmission risk) in GEE model, re-dummy code is needed, in that case, two clinical questions cannot be addressed in one model anymore, which is in the same situation with our original two models.

*Our statistical questions:*

1. Am I interpreting the reviewer’s suggestion (GEE model) in the wrong direction?
2. Does my explanation make sense to you ? Or does that sound logical to you?
3. Did I perform GEE model right?
4. Whether the GEE model I performed is what he/she suggest one?

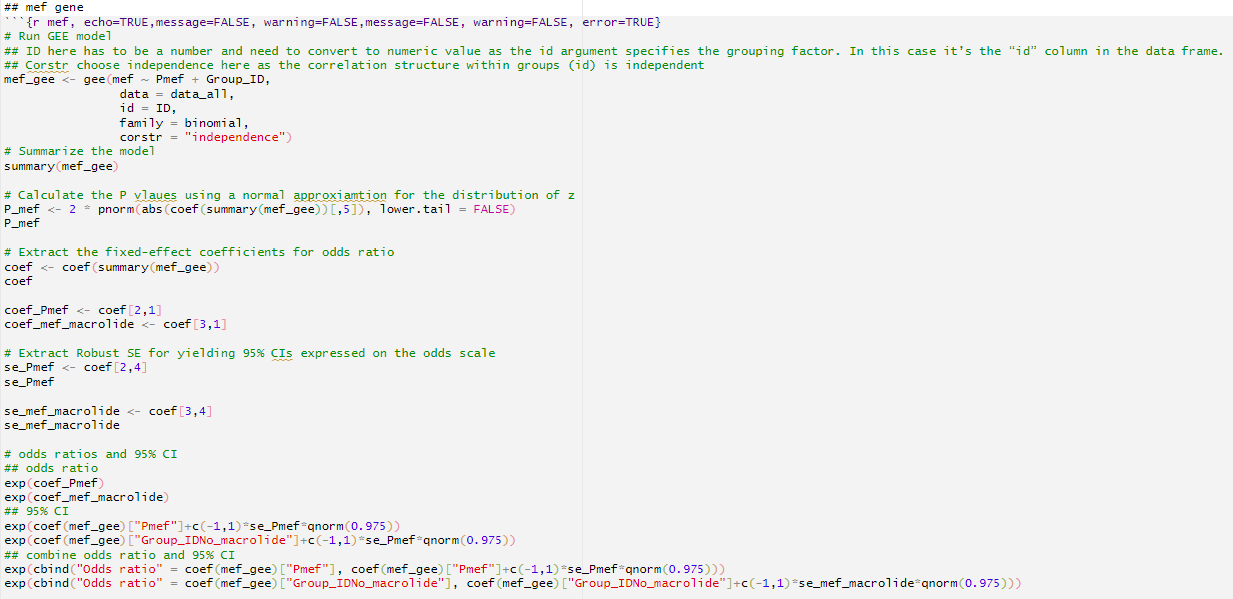
Please feel free to let us know your thoughts!

We really appreciate and value your advice!

The following pages are example of my GEE R script and outcome comparisons

**GEE model scripts**

Take *mef* gene as an example



**Outcome comparisons**

**Question1. Macrolide resistance gene co-carriage:**

Our Binary logistic model:

**In each treatment group, whether close contacts have/don’t have this gene is dependent/independent of the detection of this gene in their paired patients**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Resistance gene** | **MR vs MRCC** | | | **MNR vs MNRCC** | | |
| **Odds ratio**  **(95% CI)** | ***P* value** | ***P* values**  **(post-FDR)** | **Odds ratio**  **(95% CI)** | ***P* value** | ***P* values**  **(post-FDR)** |
| *erm*(B) | 3.4  (0.5-22.9) | 0.21 | 0.38 | 1.3×10-7  (0-Inf) | >0.99 | 0.99 |
| *erm*(C) | 5.1  (0.6-41.9) | 0.13 | 0.29 | 8.5  (0.4-163.9) | 0.16 | 0.99 |
| *erm*(F) | 11.8  (2.3-59.6) | 0.0029† | 0.020 | 1.7  (0.4-7.6) | 0.50 | 0.99 |
| *mef* | 7.3  (1.9-28.4) | 0.0044† | 0.020 | 1.3  (0.3-6.9) | 0.75 | 0.99 |
| *msr*(A) | 1.5  (0.5-4.9) | 0.48 | 0.62 | 1.8  (0.4-8.2) | 0.43 | 0.99 |
| *msr*(E) | 0.8  (0.3-2.7) | 0.74 | 0.83 | 1.1  (0.3-4.5) | 0.87 | 0.99 |
| *tetM* | 2.1×10-7  (0-Inf) | >0.99 | 0.99 | N/A | N/A | N/A |
| *tetO* | 2.7  (0.8-8.5) | 0.099 | 0.29 | 1.7  (0.4-7.6) | 0.50 | 0.99 |
| *tetW* | 2.4  (0.5-12.0) | 0.29 | 0.44 | 2.5×10-7  (0-Inf) | >0.99 | 0.99 |

Analyses could not be performed for *tetM* of the macrolide non-recipient group and for *erm*(A) (both groups) due to no variance between groups.

GEE model:

|  |  |  |  |
| --- | --- | --- | --- |
| **Resistance gene** | **Co-carriage: detection of the gene in patient affect detection of the gene in close contact ?** | | |
| **Odds ratio**  **(95% CI)** | ***P* value** | ***P* values**  **(post-FDR)** |
| *erm*(B) | 2.0  (0.4-11.2) | 0.41 | 0.47 |
| *erm*(C) | 6.1  (1.1-34.5) | 0.042 | 0.11 |
| *erm*(F) | 4.7  (1.8-12.4) | 0.0020 | 0.016 |
| *mef* | 3.8  (1.3-10.5) | 0.012 | 0.048 |
| *msr*(A) | 1.6  (0.6-4.1) | 0.30 | 0.47 |
| *msr*(E) | 0.9  (0.4-2.3) | 0.88 | 0.88 |
| *tet*(O) | 2.2  (0.9-5.6) | 0.084 | 0.17 |
| *tet*(W) | 1.9  (0.4-8.7) | 0.39 | 0.47 |

Analyses could not be performed for ermA and tetM due to variance between groups; GEE model for probability has fitted value very close to 1.

However, I cannot separate into two treatment group as treatment group is the second IV in this model, so in terms of the co-carriage results obtained from GEE is about the overall co-carriage results (including both macrolide group and non-macrolide group).

If I really want to separate them, then the Input IV would be whether patients have the gene, which requires a second GEE model.

**Question2. Macrolide effect on transmission risk:**

Our Binary logistic model: 🗸

|  |  |  |  |
| --- | --- | --- | --- |
| **Resistance**  **gene** | **As compared with non-macrolide group, whether the macrolide group is more likely to have more transmission cases?** | | |
| **Odds ratio**  **(95% CI)** | ***P* value** | ***P* values**  **(post-FDR)** |
| *erm*(A) | 1.0  (0-Inf) | >0.99 | 0.99 |
| *erm*(B) | 1.0  (0.4-2.9) | 0.96 | 0.99 |
| *erm*(C) | 1.0  (0.07-13.9) | >0.99 | 0.99 |
| *erm*(F) | 1.0  (0.4-2.5) | 0.97 | 0.99 |
| *mef* | 1.6  (0.6-3.9) | 0.33 | 0.99 |
| *msr*(A) | 1.3  (0.3-5.0) | 0.73 | 0.99 |
| *msr*(E) | 0.6  (0.2-1.5) | 0.25 | 0.99 |
| *tetM* | 0.5  (0.09-2.7) | 0.43 | 0.99 |
| *tetO* | 0.9  (0.4-2.2) | 0.82 | 0.99 |
| *tetW* | 0.7  (0.3-2.0) | 0.55 | 0.99 |

GEE model:

|  |  |  |  |
| --- | --- | --- | --- |
| **Resistance gene** | **Macrolide effects on gene detection: macrolide exposure affect the detection of the gene in close contact?** | | |
| **Odds ratio**  **(95% CI)** | ***P* value** | **P values**  **(post-FDR)** |
| *erm*(B) | 1.0  (0.3-3.0) | 0.93 | 0.99 |
| *erm*(C) | 0.8  (0.1-4.6) | 0.76 | 0.99 |
| *erm*(F) | 2.1  (0.8-5.2) | 0.11 | 0.88 |
| *mef* | 0.9  (0.4-2.2) | 0.80 | 0.99 |
| *msr*(A) | 0.8  (0.3-2.0) | 0.64 | 0.99 |
| *msr*(E) | 1.6  (0.7-3.8) | 0.26 | 0.99 |
| *tet*(O) | 1.0  (0.4-2.4) | 0.99 | 0.99 |
| *tet*(W) | 1.1  (0.4-3.0) | 0.90 | 0.99 |

Analyses could not be performed for ermA and tetM due to variance between groups; GEE model for probability has fitted value very close to 1