Building a Decision Tree with Serial Serology Measurements Improves Classification in a Flavivirus Co-circulation Region

Keith Lombard, MS¹, Fernando J. Arteaga Cebellio, Bs, MSc², Aurelie Gouel-Cheron, MD, PhD³, Francisco Belaunzarán, MD, MSc², Gabriel Nájera-Cancino, MD, MS⁴, Sandra Caballero-Sosa, MD, PhD⁵, Héctor Rincón-León, MD⁵, Emilia Del Carmen Ruis Hernandez, MD⁵, Pilar Ramos Cervantes, MSc⁶, M. Lourdes Guerrero, MD, MS⁶, John Beigel, MD⁷, Karina Trujillo-Murillo, PhD⁸, Gustavo Pedraza, CHEM², Jesús Sepulveda, MD⁹, Kenia Melina Escobedo-Lopez, BS¹, Nora K. Mora-Suarez, BS¹, Monica Reyes-Romero, BS¹, Violeta Ibarra-González, MS², Julia Martínez-Lopez, BS³, Guillermo Ruiz-Palacios, MD, FIDSA⁶ and Sally Hunsberger, PhD⁷.

Introduction

RT-PCR (Reverse Transcriptase Polymerase Chain Reaction) is often considered the “gold standard” for diagnosis of Zika Virus (ZIKV) infection, however it has been shown to have low sensitivity. A possible remedy is to study ZIKV-specific IgG (ZsIgG) and IgM (ZsIgM) antibodies. However, the in-vitro cross-reactivities of Dengue Virus (DENV) and ZIKV-specific antibodies are well known, leading to diagnostic difficulties in an area with co-circulation of the two viruses. Our goal was to use Zika and Dengue serologic assays to build a classification model that improves upon the Positive Predictive Value (PPV) of commercial kits while maintaining sensitivity.

Materials and methods

We conducted a prospective longitudinal study in Southern Mexico where DENV and ZIKV co-circulation occurs. Patients were included in two cohorts: a cohort of subjects presenting with a febrile rash meeting WHO/PAHO Zika case definition and a household cohort. After signed consent, all subjects enrolled were evaluated on Study-visits Days 0, 3 and 7 (for fever rash cohort) and 28. We considered a subject “true positive” for ZIKV or DENV if RT-PCR positive at any time point. The healthy household cohort (with no positive RT-PCR) were considered “true negatives”. We fit a statistical decision tree taking as inputs serial serology measurements and outputting a predicted disease category.

We considered the following variables in our modeling procedure:

- Zika IgM Day 0, Zika IgM Day 28, Change Zika IgM Day 0 to 28
- Zika IgG Day 0, Zika IgG Day 28, Change Zika IgG Day 0 to 28
- Dengue IgM Day 0, Dengue IgM Day 28, Change Dengue IgM Day 0 to 28
- Dengue IgG Day 0, Dengue IgG Day 28, Change Dengue IgG Day 0 to 28
- Zika PCR+ (wherever above the line at any time point indicate a positive diagnosis. This tests lacks PPV and specificity.
- Dengue PCR+ (wherever above the line at any time point indicate a positive diagnosis. This tests lacks sensitivity.
- Household

Results

As of March 2018, we have 32 subjects in the Zika PCR+ group, 32 in the Dengue PCR+ group, and 68 in the household group. The highest Positive Predictive Value achieved by the kit manufacturer recommended cutoffs while maintaining a sensitivity of at least 10% on Zika PCR+ subjects is 30/114 (26%, see Figures 1 and 2), and for Dengue PCR+ subjects is 21/30 (70%).

Our decision tree (Figure 3) achieved a Positive Predictive Value (PPV) of at least 80% on all three disease categories, while maintaining sensitivity above 50% (Table 1).

Conclusions

- A useful diagnostic test will be able to maintain both high sensitivity and PPV. Some of the cutoffs suggested by the serology kit manufacturers enjoyed high sensitivity, but low PPV (such as Zika IgG). Others enjoyed high PPV, but disappointing sensitivity (such as Zika IgM).
- Using serology data in a statistical decision tree improves the PPV exhibited by the kit manufacturer recommended cutoffs, while still maintaining respectable sensitivity.
- Observing the variables present in the decision tree provides insight into the measures that are important for achieving diagnostic accuracy. In general, we note that it is important to consider not only serology at a fixed time point, but also change in serology over time.
- Physicians in regions with co-circulating flaviviruses should be aware of the pitfalls of using only RT-PCR (i.e. low sensitivity) or using only pre-established kit manufacturer recommended cutoffs for diagnosis (i.e. low sensitivity, low PPV, or both).
- This analysis is meant as a proof of concept; the takeaway should be that diagnostic accuracy can be improved by considering serological measures for more than one flavivirus simultaneously.

Acknowledgments

Funded in part by the NCI Contract No. HHSN261200800001E. Funded in part by the Mexican Ministry of Health. All data analysis and statistical modeling were performed in the R programming language. The following R packages were used at some point in the analysis: dplyr, ggplot2, rpart, rpart.plot, stringr.

Contact

Keith Lombard
keith.lombard@nih.gov
(240)-689-5258

References

[1] Clinical Research Directorate/Clinical Monitoring Research Program, Frederick National Laboratory for Cancer Research. 2008, Rockville, MD. ¹Hospital General Tapachula, Tapachula, Mexico. ²Instituto Nacional de Enfermedades Respiratorias, Mexico City, Mexico. ³Instituto Nacional de Cardiología, Mexico City, Mexico. ⁴Instituto Nacional de Artritis, Reumatología y Enfermedades del Envejecimiento, Mexico City, Mexico. ⁵Departamento de Salud y Bienestar, Mexico City, Mexico. ⁶Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado, Tapachula, Mexico, Mexico. ⁷Instituto Mexicano del Seguro Social, Mexico City, Mexico. ⁸Hospital General de Mexico, Mexico City, Mexico. ⁹Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico. ¹⁰Comisión Coordinadora de los Institutos Nacionales de Salud y Hospitales de Alta Especialidad, Mexico City, Mexico.