Statistical Learning Methods to Determine Immune Correlates of Herpes Zoster in Vaccine Efficacy Trials

Peter B. Gilbert and Alexander R. Luedtke

Department of Biostatistics, Bioinformatics, and Epidemiology, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington

Using Super Learner, a machine learning statistical method, we assessed varicella zoster virus-specific glycoprotein-based enzyme-linked immunosorbent assay (gpELISA) antibody titer as an individual-level signature of herpes zoster (HZ) risk in the Zostavax Efficacy and Safety Trial. Gender and pre- and postvaccination gpELISA titers had moderate ability to predict whether a 50–59 year old experienced HZ over 1–2 years of follow-up, with equal classification accuracy (cross-validated area under the receiver operator curve = 0.65) for vaccine and placebo recipients. Previous analyses suggested that fold-rise gpELISA titer is a statistical correlate of protection and supported the hypothesis that it is not a mechanistic correlate of protection. Our results also support this hypothesis.

Keywords. Antibody titer; case-cohort study; correlate of immunity; immune correlate of protection; machine learning; super learner; systems vaccinology; Zostavax Efficacy; safety trial.

For any vaccine that confers protection against an infectious disease outcome, it is important to develop immune response markers that are “correlates of protection.” Within randomized, controlled, preventive vaccine efficacy (VE) trials, correlates can be assessed by measuring pre- and postvaccination immune responses via a case-cohort or nested case-control sampling design, and studying the statistical relationship of these markers with the outcome and with the level of VE to prevent the outcome. For VE trials of preventive herpes zoster (HZ) vaccines, the following questions arise:

1. Correlate of risk (CoR): How well do markers associate with HZ in vaccine recipients and in unvaccinated/placebo recipients?
2. Correlate of protection (CoP): How well do markers in vaccine recipients associate with the level of VE to prevent HZ?
3. Signature of risk: How well do markers classify/predict whether an individual (vaccinated or unvaccinated) will experience HZ?

Qin et al [1] and Plotkin and Gilbert [2] emphasized that these concepts are about statistical associations. A correlate with good properties under criteria for (1) to (3) may or may not be a mechanistic CoP (an immune response that, if manipulated to a set value at the time of pathogen exposure, *determines* that the disease outcome cannot occur). Correlate types (1) and (2) measure population-level associations across subgroups, enabling prediction of

population-level disease rates or VE. Correlate type (3) measures accuracy of individual-level predictions and is a more stringent type of correlate of risk—an accurate signature of risk will also be a strong CoR, but the converse may fail; the reason is that CoRs discriminate average risk whereas signatures discriminate individual risk. The most useful marker correlate should have good properties under criteria for each correlate type, where 2 markers showing similar performance on CoR and CoP criteria may have their utility level ranked/discriminated by signature of risk criteria. Moreover, good mechanistic CoPs typically perform well against signature of risk criteria, meaning that poor signature performance can help establish a marker as a poor mechanistic CoP.

Schmader et al [3] and Gilbert et al [4] showed that the anti-varicella zoster virus (VZV) antibody level fold-rise pre-vaccination to 6 weeks postvaccination was a strong CoR and CoP in the phase III Zostavax Efficacy and Safety Trial (ZEST), and thus useful for defining subgroups with low or high average risk of zoster disease and for defining subgroups with low or high vaccine efficacy. However, antibody level was not assessed as a signature of risk. The current report makes this assessment using state-of-the-art machine learning methodology.

METHODS

ZEST (NCT00534248) randomized 22 439 individuals aged 50–59 years with a history of VZV in 1:1 allocation to receive 1 injection with Zostavax (Merck) or placebo, and followed them for an average of 1.3 years for occurrence of the primary study endpoint, “confirmed HZ disease” [3, 4]. A case-cohort sampling design was used to measure prevaccination (day 1) and week 6 glycoprotein-based enzyme-linked immunosorbent assay (gpELISA) anti-VZV antibody [3, 4]. The gpELISA assay was optimized and validated at Merck Research Laboratories and is based on the quantitative detection of antibodies to VZV

Presented in part: Workshop on Advances and Controversies in our Understanding of Herpes Zoster, The Royal Society of Medicine, London, UK, March 2017.

Correspondence: P. B. Gilbert, PhD, Department of Biostatistics, Bioinformatics, and Epidemiology, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109 (pgilbert@scharp.org).

The Journal of Infectious Diseases® 2018;218(S2):S99–101

© The Author(s) 2018. Published by Oxford University Press for the Infectious Diseases Society of America. All rights reserved. For permissions, e-mail: journals.permissions@oup.com. DOI: 10.1093/infdis/jiy421

glycoproteins that have been purified from VZV-infected human fibroblasts [5]. This assay provides an alternative to commercial kits, which have poor sensitivity for detection of anti-VZV antibodies. Measurements were made in a 10% random sample of all participants drawn at enrollment (cohort) and in all participants who experienced the HZ study endpoint after week 6 (cases).

Statistical Analysis

Using the ZEST data set, we applied Super Learner [6] to estimate $P(HZ|W, A, S)$, the probability of HZ occurrence during

follow-up conditional on various fixed values of (W, A, S) , where W is baseline variables (age, gender, day 1 gpELISA titer), A is treatment assignment (Zostavax or placebo), and S is week 6 gpELISA data (week 6 titer, fold-rise in titer from day 1 to week 6). Super Learner has three steps: (1) define a large library of statistical algorithms (eg, logistic regression or the LASSO), each of which estimates $P(HZ|W, A, S)$; (2) quantify the classification accuracy of each algorithm via cross-validation; and (3) obtain the Super Learner estimate of $P(HZ|W, A, S)$, corresponding to the optimal linear combination of estimates from the individual

Figure 1. Cross-validated area under the receiver operator characteristic curve (CV-AUC) for 60 statistical algorithms. For ZEST subjects assigned to placebo and Zostavax, each horizontal line shows the point and 95% confidence interval estimate of the CV-AUC for a given statistical algorithm for estimating $P(HZ|W, A, S)$. The algorithms consider inputs W = (age, gender) in the first-row facet (9 algorithms); W = (age, gender, day 1 titer) in the second-row facet (9 algorithms); W = (age, gender) and S = (week 6 titer, fold rise) in the third-row facet (9 algorithms); and W = (age, gender, day 1 titer) and S = (week 6 titer, fold-rise, average of day 1 titer and week 6 titer) in the fourth-row facet (33 algorithms). The Super Learner was fit using the SuperLearner R package, with candidate algorithms Bayes generalized linear model (GLM; SL.bayesglm), GLM (SL.glm), GLM with interactions (SL.glm.interaction), sample mean (SL.mean), generalized additive model (SL.gam), random forest (SL.cforest), and gradient boosting (SL.xgboost). Demo = demographic information (age in years and gender).

algorithms yielding the best cross-validated accuracy. Super Learner was applied to the ZEST data with the SuperLearner R package and 60 statistical algorithms (Figure 1), for each of multiple treatment arm-stratified input variable sets defined based on (W, S), with weighting of subjects by reciprocals of probabilities they were sampled for measurement of day 1 and week 6 gpELISA titers (code provided at <https://github.com/alexluedtkel2/zoster>). Step (2) was implemented using cross-validation with the data set repeatedly divided into 13 of 15 parts (training sets) and the remaining 2 of 15 parts (test sets) and cross-validated area under the receiver operator curve (CV-AUC) as the accuracy criterion. CV-AUC reflects the probability that a future randomly sampled HZ case has a higher estimate of $P(\text{HZ}|\text{W}, \text{S})$ than a future randomly sampled HZ noncase, where 0.5 indicates no classification capacity and 1.0 indicates perfect classification.

RESULTS

Figure 1 shows the Super Learner analysis results. For the placebo group, the best model, considering age and gender has CV-AUC = 0.6 (top left panel), which is driven by higher HZ risk in women than men [4]. When adding titer data CV-AUC increases to 0.65, where any titer information (day 1, week 6, fold-rise, or average of day 1 and week 6) provides the same CV-AUC. For the vaccine group, age, gender, and day 1 titer have no capacity to predict HZ (CV-AUCs ≤ 0.5 , top right panel), whereas the Gradient Boosting models considering week 6, fold-rise, or average titer between day 1 and week 6 all achieve maximum CV-AUC at about 0.65 (bottom right panel), similar to the placebo group. Thus, gender plus any titer information for the placebo group and week 6 titer information for the vaccine group (regardless of gender) provide similar ability to classify HZ outcome moderately well. This assessment of “moderate” performance is based on normative standards for reasonably diagnostic biomarkers.

DISCUSSION

While previous analyses of the ZEST showed that immune response markers based on anti-VZV gpELISA antibody titers yield strong CoRs and CoPs for HZ, the Super Learner classification analyses show that these markers do not yield a strong signature of HZ risk. This shows that a marker can be good at discriminating subgroups with low versus high disease risk or low versus high VE on average—which is very useful for multiple applications—yet have limited ability to determine which specific individuals will experience the disease endpoint. Because the Super Learner methodology optimally selects between different machine learning approaches and their combination for building signatures of risk, this finding may provide compelling evidence that VZV-specific gpELISA antibody titers are not a strong signature of HZ risk (eg, [6]). One potential issue is that the most predictive Super Learner model may be a combination of multiple models, hindering interpretation. In addition, Super Learner

analysis is contingent on the specific metric used to measure classification accuracy and a single metric may not capture all relevant aspects of classification accuracy. These challenges may be addressed by including interpretable models in Super Learner and repeating the analysis using multiple classification accuracy metrics. Another potential issue is that the optimized and validated gpELISA assay still showed intraassay and interassay variabilities exceeding 10% (14% and 17%, respectively), with overall precision (percent relative standard deviation of assay results obtained across different analysts, assay days, VZV glycoprotein lots, and tissue culture cell lots) estimated at 23% [5]. Thus, the gpELISA titer markers that were studied could have had significant portions of their interindividual variability due to technical measurement error, which would diminish their ability to predict HZ.

Our analysis illustrates the value in assessing immune correlates of multiple types/concepts to better understand marker merits and limitations. In addition, the finding that gpELISA titers are not an excellent signature of HZ risk motivates the study of additional markers for improving classification accuracy. Systems vaccinology of high-dimensional markers (eg, cell subpopulation analysis [7]) may also warrant consideration.

Notes

Acknowledgments. The authors thank all the volunteers who participated in the ZEST study, the ZEST protocol 022 study groups, and Lindsay Carpp for editing.

Disclaimer. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Financial support. This work was supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Department of Health and Human Services (grant number R37AI054165).

Supplement sponsorship. This work is part of a supplement sponsored by the Royal Society of Medicine (Royal Charter number RC000525) funded through unrestricted educational grants from Merck, Sanofi Pasteur MSD, The Research Foundation for Microbial Diseases of Osaka University, Seqirus and GlaxoSmithKline.

Potential conflicts of interest. Both authors: No reported conflicts of interest. Both authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Qin L, Gilbert PB, Corey L, McElrath MJ, Self SG. A framework for assessing immunological correlates of protection in vaccine trials. *J Infect Dis* 2007; 196:1304–12.
2. Plotkin SA, Gilbert PB. Nomenclature for immune correlates of protection after vaccination. *Clin Infect Dis* 2012; 54:1615–7.
3. Schmader KE, Levin MJ, Gnann JW Jr, et al. Efficacy, safety, and tolerability of herpes zoster vaccine in persons aged 50–59 years. *Clin Infect Dis* 2012; 54:922–8.
4. Gilbert PB, Gabriel EE, Miao X, et al. Fold rise in antibody titers by measured by glycoprotein-based enzyme-linked immunosorbent assay is an excellent correlate of protection for a herpes zoster vaccine, demonstrated via the vaccine efficacy curve. *J Infect Dis* 2014; 210:1573–81.
5. Hammond O, Wang Y, Green T, et al. The optimization and validation of the glycoprotein ELISA assay for quantitative varicella-zoster virus (VZV) antibody detection. *J Med Virol* 2006; 78:1679–87.
6. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* 2007; 6:Article25.
7. Tsang JS, Schwartzberg PL, Kotliarov Y, et al.; Baylor HIPC Center; CHI Consortium. Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell* 2014; 157:499–513.