

From Statistical Probability to Machine Learning Prediction Methods in Orofacial Pain and Headache Research: Evolution from Quality Control of Guinness Beer to IBM Innovations

I was asked to write the first editorial of our 2019 volume of the *Journal of Oral & Facial Pain and Headache*—a task that merits some jovial words, first to thank the previous editorial team under the leadership of Barry Sessle, and second, to wish all the best to the new team under the guidance of Rafael Benoliel.

When I was a dental student, a little over four decades ago, we were taught that the Student *t* test was the strongest method to get closer to the truth. Humans always have to search for the route to certainty, a challenge to our species. The *t* test was developed by William Sealy Gosset in 1908, a chemist at the Guinness brewery in Dublin, Ireland. According to the legend, he used the pseudonym Student since his employer was not comfortable with a publication under his name. The *t* test was “the” answer for us as dental students. Years later, it showed some limitations—not a surprise, as all methods have limitations. Caution is mandatory when interpreting results, as I will describe below.

The famous Italian mathematician Carlo Emilio Bonferroni provided more credibility to the *t* test. Adjusting the threshold for repeated testing was a must when rejecting a null hypothesis. Analysis of variance (ANOVA) for multiple variables was developed in the 1920s by an English mathematician in genetics, Ronald A. Fischer, working with large population data.

Correlation and cross-correlation for associations were already among the analytic methods available. In the meantime, the placebo effect was decrypted, and randomized clinical trials (RCT) became the norm—with control groups, blind conditions, study power calculations, representative selection of subjects, etc. We had to get an uncontaminated and clean population to control large variability. Without such a careful and selective process, our data were not possible to explain. And we must not forget that clinical science is based on statistical tests that weigh probability with an arbitrary *P* value of .05 proposed by the same Fischer in 1925; ie, a 1 in 20 chance your significant difference exceeds chance.

It then became clear that to interpret statistical probability, we had to be careful and not assume that we found the answer explaining “the” cause with a single model. Similarly, correlation tests for associations or case series should not be mixed with causality extrapolation. Any analytic method is as good as what we include in the model; the findings of a given

study cannot be generalized to all situations and populations. Statistics help us get closer to a level of certainty, but we must recognize that all our analyses have a percentage of variability or unexplained aspects before we make conclusions.

Then, a “miracle” arrived: systematic analyses, meta-analyses, and meta-analyses of meta-analyses were developed to guide us as scientists and clinicians to extract what makes sense in relation to our hypothesis and clinical context.

The so-called pyramid of evidence gave strength to evidence-based knowledge; it became the principle, with doctor or scientist opinion at the base of the pyramid, case reports above, RCTs in the middle, and meta-analyses on top. Numerous excellent and important methods have been proposed for building a solid meta-analysis, and they have strong merits. I love meta-analyses: They guide and inspire me (although I am sometimes confused when from hundreds of papers selected, a conclusion is made based on two—but hopefully in these situations authors conclude that more studies are needed). As mentioned above, all methods have limitations, and we should remain humble in our interpretation, extrapolation, and application of new findings. We have to remind ourselves that systematic reviews and meta-analyses are as good as the studies included and do not 100% protect us from bias or methodologic limitations even if rigorous methods are used. The results have to make sense; the best outcome(s) for diagnosis or treatment have to be scientifically, socially, and economically relevant.

Interesting point: Can one explain to me why the pyramid of knowledge is always represented as a triangle? In my high school geometry class, we were taught a pyramid has four or five faces. Possibly this bodes for good news—other previously hidden avenues may appear to help us find the best diagnostic method and the best treatment in the era of medicine tailored to the patient’s biology, psychology, and social reality.

A new (or rather, not so new but more accessible) method is currently available in health research: machine learning (ML), although it is not a panacea or a remedy to all our previous limitations or problems. The origin of ML is attributed to Arthur Samuel in 1959 when working at IBM on computer gaming models. ML uses high and fast power calculations integrating mathematical methods and statistics. The

data are analyzed and reanalyzed several times—very rapidly if you have access to a fast computer—to find patterns and/or clusters. The final outcome is that we are able to build a “prediction” model to achieve a more precise and rapid decision in diagnostic and treatment planning.

To have an idea of the power of such methods, just think how fast credit card companies can get back to you if an ongoing misuse event happens with your credit card number. For now, the IBM ML Watson system is a great venture from which the health system still awaits solid results. Consider also the “magic” of automated driving vehicles that albeit need better vision processing to avoid accidents in unexpected conditions.

After decades of limited analytic methods, we can now analyze large sample size populations with several variables. What a change! ML methods include some of the following: multivariate adaptive regression splines; lasso regression; classification and regression trees; bagging; random forests; genetic algorithms; and artificial neural networks.

In my collaborative work with Milton Maluly, Cibele Dal Fabbro, Altay Lino de Sousa, and Sergio Tufik, it is exciting to see that some unknown variables may contribute to explain the occurrence of sleep bruxism at a given age in a specific general population in São Paulo, Brazil. Here again we have to be cautious, as what is found in the São Paulo general population may not fit in your clinical day-to-day reality in Asia, India, Europe, Australia, North America, or Africa. ML is the topic of intensive research in dentistry for clinical diagnostics in radiology, oral medicine, orthodontics, and other disciplines. Run a PubMed query, and you will find several examples!

ML does have some limitations, including:

- Results seem to be valid only within specific samples—use with different populations and/or countries remains to be demonstrated.
- Since ML is reanalyzing data several times, it is learning the data set, getting “familiar” with it due to the fact that it is testing and retesting.

The machine will get better and better, then the “best fit” bias is present. This means that ML tends to “over fit” the model for a specific data set. It is then easy to figure what can happen if you change the data set; ie, results may be different.

- Risk of significance if multiple sites are using algorithms (Bonferroni lives again!).
- Obviously, privacy and bioethical concerns are present with the use of population or patient data. We still have to balance this with the need for high-powered calculations!
- Again, like in all the methods listed above, human bias is always possible. What you select and enter in the ML is your decision and likely to affect the results. It should be a nonbiased process without preconceived ideas or selective omissions.
- ML is the basis of artificial intelligence, a more global application of powerful mathematics to make decisions, as used in speech recognition software, automated driving, etc. More societal concerns are raised with such high-level decision-making.

In conclusion, ML is a powerful tool with superb possibilities to advance our knowledge, but just as we have progressed in other fields, we will surely have more advanced methodology in the pipeline (particularly when you consider the giant steps in the forces underlying ML—computer science and artificial intelligence). The role of the scientist is still irreplaceable, and our critical clinician-scientist judgment must remain sharp and our vigilance levels high. With these limitations in mind, there is still an appropriate setting to consider ML, a predictive method, among other probabilistic methods such as the *t* test and ANOVA. All have their roles and limitations. Welcome to ML, a great tool for clinician-scientists to accelerate the acquisition and application of new knowledge in our practices!

Gilles Lavigne
Associate Editor