Is Q-Learning a Valid Method of Knowing?

Francisco J. Diaz[1]
[1]Department of Biostatistics, The University of Kansas Medical Center, Mail Stop 1026, 3901 Rainbow Blvd., Kansas City, KS 66160, United States. Email: fdiaz@kumc.edu.

Abstract
A great deal of statistical research on Dynamic Treatment Regimes focuses on Q-learning. However, the mathematical coherence and statistical fundamentation of Q-learning are still very poor. In fact, in Q-learning, it is impossible to distinguish between the model explaining or describing the illness phenomenon and the clinical algorithm for treatment individualization. In addition to this epistemological conundrum, Q-learning is mathematically intractable using standard asymptotic or decision theories. Standard theory cannot be used to test the null hypothesis that a treatment has no effect, or to construct confidence intervals. Incoherent definition of covariates is also common. Researchers have attempted to remedy some of these issues, but questions arise about how should we build models in personalized medicine (PM). We discuss here about these issues. As an alternative, Generalized Linear Mixed Effects Models and Empirical Bayesian Feedback can be used to establish a solid paradigm for the construction of the mathematics and statistics of PM research and practice. In fact, there is a long tradition of mixed modeling for treatment individualization in pharmacological literature.

Key Words: Dynamic treatment regimes, Generalized Linear Mixed Models, Empirical Bayesian Feedback, Q-Learning, Drug Dosage Individualization, Personalized Medicine.

1. The Epistemological Conundrum of Q-learning

A great deal of statistical research concerning dynamic treatment regimes (DTRs) has focused on the implementation of Q-learning and related machine learning 'methods' (Chakraborty et al. 2011, 2013, 2014; Murphy 2003; Robins 2004). The main goal of Q-learning is to build clinical algorithms for treatment individualization in clinical settings. However, the great ebullience with which these 'methods' have been embraced by some research groups in recent years does not commensurate with their mathematical coherence and statistical fundamentation, which are very poor.

Next we describe the problems of Q-learning, although our critiques also apply verbatim to the related proposals by Murphy (2003) and Robins (2004). The most serious problem is that, in Q-learning, it is impossible to distinguish between the model explaining or describing the illness phenomenon and the clinical algorithm for treatment individualization. That is, Q-learning approaches confuse the decision-making process with the mathematical or statistical model of the clinical reality that will be modified because of the decision. This epistemological conundrum, which was also noted by Zajonc (2012), has two undesirable consequences: Q-learning is mathematically intractable using standard decision theory or using standard asymptotic theory.

In particular, standard asymptotic theory cannot be used in Q-learning to test the null hypothesis that a treatment has no effect, or to construct confidence intervals. This issue has been acknowledged by the DTR research community, which labels the issue with the name "nonregularity" (Robins 2004; Chakraborty et al. 2013). Inevitable practical consequences of these problems have been reported. For instance, Rosthøj et al. (2006) report the impossibility of programming some of the formulas from the related method of Murphy (2003), lack of numerical convergence, and instability of the estimates with respect to initial values. Although efforts have been made to solve the problem of nonregularity (see the review of Chakraborty et al. 2014), a convincing solution has not been found yet, if there is such solution.

2. Incoherent Definitions of Covariates

But the inseparability of the decision rule from the model of reality or data is not the only issue of Q-learning. Incoherent definitions of covariates are also common in applications of Q-learning to personalized medicine. Specifically, it is not unusual to combine Q-learning with regression models that have second-stage covariates whose unique values represent non-unique things depending on the patient's history. Besides violating basic properties of mathematical expectations, this clearly prevents making useful interpretations of regression parameters.

Two examples of this issue from the Q-learning literature follow. In a hypothetical example of treatment of alcohol addiction, Chakraborty (2011) describes a covariate A_2 that both symbolizes a second stage treatment and takes on only the values 1 and -1. For responders to the initial treatment, $A_2=1$ represented telephone monitoring. However, for nonresponders, $A_2=1$ represented cognitive behavioral therapy if the initial treatment was naltroxene, but $A_2=1$ represented using naltroxene if the initial treatment was cognitive behavioral therapy.

An analogous inconsistency in the definition of a covariate is also manifest in Murphy et al. (2007). In an explanation of how Q-learning can be used for the dynamic treatment of depression, these authors built a dichotomous covariate T_2 representing the switch to a new treatment regardless of which the initial or the new treatment was. Thus, it is unclear what aspect of the clinical phenomenon is being measured by the regression coefficient of T_2.

3. How to build the mathematics of personalized medicine?

Given the apparent epistemological and mathematical problems of Q-learning, we would like to point at a research direction we believe is more promising. The author believes regression models with random effects will revolutionize the mathematical theory and practical applications of pharmacology and personalized medicine (Diaz 2016). In fact, considerable research suggests these models can be used to establish a solid paradigm for the construction of the mathematics and statistics of PM research and practice, especially in the treatment of chronic diseases (Sheiner et al. 1972; Whiting 1986; Diaz et al. 2007; 2012a,b; 2013a,b; 2014; 2016; Zhu and Qu 2016; Cho et al. 2016). The key idea is that regression models with random effects have concepts that allow describing patient populations as a whole (the fixed effects) and, simultaneously, concepts that allow describing patients as individuals (the random effects). Thus, models with random effects

are useful because the variability of a random coefficient is not just a mathematical artifact to control for patients' heterogeneity: it is the result of real variation in the biological and environmental factors that have made humans develop as individuals [Diaz 2016; Diaz et al. 2012, 2013a,b, 2016; Senn 2001, 2016].

There are numerous examples suggesting regression models with random effects have a great potential for DTR development. For instance, the Sheiner School of Pharmacology has advocated for decades the use of drug dosage individualization based on random-effects nonlinear models and empirical Bayesian feedback (EBF) (Sheiner 1972; Whiting et al. 1986; Pillai et al. 2005). Drug dosage individualization based on random-effects linear models has also been investigated, using EBF (Diaz et al. 2007, 2012b) or using conditional inference functions (Zhu and Qu 2016; Wang et al. 2012). Importantly, EBF is firmly anchored to standard decision theory (Diaz et al. 2007, 2012b). In addition, using a combination of random forests and random-effects linear models, Cho et al. (2016) built clinical algorithms to assign patients to the best treatment. As another example, Diaz (2016) has proposed an approach to measuring the individual benefit of a medical or behavioral treatment using generalized linear mixed models.

4. Conclusion

In conclusion, the epistemological and mathematical validities of Q-learning are uncertain. In contrast, both biological and mathematical arguments suggest the potential of regression models with random effects and empirical Bayesian feedback for the development of personalized medicine research and practice, including the development of DTRs. When the normality assumption of random effects is not valid, robust approaches such as those based on conditional inference functions have also a great potential in this area.

References

Chakraborty, B. and Murphy, S.A. (2014), "Dynamic Treatment Regimes," *Annual Review of Statistics and Its Applications*, 1, 447-464.

Chakraborty, B., Laber, E.B. and Zhao, Y. (2013), "Inference for Optimal Dynamic Treatment Regimes Using an Adaptive m-Out-of-n Bootstrap Scheme," *Biometrics*, 69, 714-723.

Chakraborty, B. (2011), "Dynamic Treatment Regimes for Managing Chronic Health Conditions: a Statistical Perspective," *American Journal of Public Health*, 101, 40-45.

Cho, H., Wang, P., and Qu, A. (2016), "Personalized Treatment for Longitudinal Data Using Unspecified Random-Effects Model," *Statistica Sinica* (in press). doi: 10.5705/ ss.202015.0120.

Diaz, F.J. (2016), "Measuring the Individual Benefit of a Medical or Behavioral Treatment Using Generalized Linear Mixed-Effects Models," *Statistics in Medicine*, 35, 4077-4092.

Diaz, F.J., Berg, M.J., Krebill, R., Welty, T., Gidal, B.E. , Alloway, R. and Privitera, M. (2013a), "Random-Effects Linear Modeling and Sample Size Tables for Two Special Cross-Over Designs of Average Bioequivalence Studies: The 4-Period, 2-Sequence, 2-Formulation And 6-Period, 3-Sequence, 3-Formulation Designs," *Clinical Pharmacokinetics*, 52, 1033-1043.

Diaz, F.J. and de Leon, J. (2013b), "The Mathematics of Drug Dose Individualization Should Be Built with Random Effects Linear Models". *Therapeutic Drug Monitoring*, 35, 276-277.

Diaz, F.J., Yeh, H-W. and de Leon, J. (2012a), "Role of Statistical Random-Effects Linear Models in Personalized Medicine," *Current Pharmacogenomics and Personalized Medicine*, 10, 22-32.

Diaz, F.J., Cogollo, M., Spina, E., Santoro, V., Rendon, D.M. and de Leon, J. (2012b), "Drug Dosage Individualization Based on a Random-Effects Linear Model," *Journal of Biopharmaceutical Statistics*, 22, 463-484.

Diaz, F.J., Rivera, T.E., Josiassen, R.C. and de Leon, J. (2007), "Individualizing Drug Dosage by Using a Random Intercept Linear Model," *Statistics in Medicine*, 26 2052-2073.

Diaz FJ, Eap CB, Ansermot N, Crettol S, Spina E, and de Leon J. (2014), "Can Valproic Acid Be an Inducer Of Clozapine Metabolism," *Pharmacopsychiatry*, 47:89–96.

Murphy. S.A. (2003), "Optimal Dynamic Treatment Regimes," *Journal of the Royal Statistical Society. Series B*. 65, 331-366.

Murphy, S.A, Oslin D.W., Rush A.J., Zhu J. and MCATS. (2007), "Methodological Challenges in Constructing Effective Treatment Sequences for Chronic Psychiatric Disorders," *Neuropsychopharmacology*, 32,257-62.

Pillai G, Mentré F and Steimer J-L. (2005), "Non-Linear Mixed Effects Modeling - from Methodology and Software Development to Driving Implementation in Drug Development Science," *Journal of Pharmacokinetics and Pharmacodynamics*, 32: 161-183.

Robins, J. (2004), "Optimal Structural Nested Models for Optimal Sequential Decisions," Proceedings of the Seattle Symposium in Biostatistics, 2nd, ed. D Lin, P Heagerty, pp. 189–326. New York: Springer.

Rosthøj, S., Fullwood, C., Henderson, R. and Stewart, S. (2006), "Estimation of Optimal Dynamic Anticoagulation Regimes from Observational Data: A Regret-Based Approach," *Statistics in Medicine*, 25, 4197–215.

Senn S. (2001), "Individual Therapy: New Dawn or False Dawn," *Drug Information Journal*, 35: 1479-1494.

Senn S. (2016), "Mastering Variation: Variance Components and Personalised Medicine," Statistics in Medicine, 35: 966-977.

Sheiner L.B., Rosenberg B., and Melmon K.L. (1972), "Modelling of Individual Pharmacokinetics for Computer-Aided Drug Dosage," *Computers and Biomedical Research*, 5:411–459.

Wang P., Tsai G.F., and Qu A. (2012), "Conditional Inference Functions for Mixed-Effects Models with Unspecified Random-Effects Distribution," *Journal of the American Statistical Association*, 107: 725-736.

Whiting, B., Kelman, A.W. and Grevel, J. (1986), "Population Pharmacokinetics, Theory and Clinical Application." *Clinical Pharmacokinetics*, 11, 387-401.

Zajonc, T. (2012), "Bayesian Inference for Dynamic Treatment Regimes: Mobility, Equity, and Efficiency in Student Tracking," *Journal of the American Statistical Association*, 107, 80-92.

Zhu, X., and Qu, A. (2016), "Individualizing Drug Dosage with Longitudinal Data," *Statistics in Medicine*, in press. doi: 10.1002/sim.7016.