Schizophrenia Research xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres

A computational algorithm for personalized medicine in schizophrenia

Beom S. Lee [a,*], Roger S. McIntyre [b], James E. Gentle [c], Nan Sook Park [d], David A. Chiriboga [e], Yena Lee [f], Sabrina Singh [a], Marie A. McPherson [a]

[a] Department of Mental Health Law & Policy, Louis de la Parte Florida Mental Health Institute, University of South Florida, Tampa, FL 33612, USA
[b] Department of Psychiatry, University of Toronto, Toronto, Ontario M5T 1R8, Canada
[c] Department of Computational and Data Sciences, George Mason University, Fairfax, VA 22030, USA
[d] School of Social Work, University of South Florida, Tampa, FL 33612, USA
[e] Department of Child & Family Studies, Louis de la Parte Florida Mental Health Institute, University of South Florida, Tampa, FL 33612, USA
[f] Mood Disorders Psychopharmacology Unit, University Health Network, Toronto, Ontario M5T 2S8, Canada

ARTICLE INFO

Article history:
Received 19 January 2017
Received in revised form 14 April 2017
Accepted 1 May 2017
Available online xxxx

Keywords:
Personalized medicine
Computational algorithm
Predictive modeling
GWAS
SNPs
Pharmacogenomics

ABSTRACT

Despite advances in sequencing candidate genes and whole genomes, no method has accurately predicted who will or will not benefit from a specific antipsychotic medication among patients with schizophrenia. We propose a computational algorithm that utilizes a person-centered approach that directly identifies individual patients who will respond to a specific antipsychotic medication. The algorithm was applied to the data obtained from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study. The predictors were either (1) 13 single-nucleotide polymorphisms (SNPs) and 53 baseline variables or (2) 25 SNPs and the same 53 baseline variables, depending on the existing findings and data availability. The outcome variables were either (1) improvement in the Positive and Negative Syndrome Scale (PANSS) (Yes/No) or (2) completion of phase 1/1A (Yes/No). Each of those four predictor-outcome combinations was tried for each of the five antipsychotic medications (Perphenazine, Olanzapine, Quetiapine, Risperidone, and Ziprasidone), leading to 20 prediction experiments. For 18 out of 20 experiments, all three performance measures were greater than 0.50 (sensitivity 0.51–0.79, specificity 0.52–0.79, accuracy 0.52–0.74). Notably, the model provided a promising prediction for Ziprasidone for the case involving completion of phase 1/1A (Yes/No) predicted by 13 SNPs and 53 baseline variables (sensitivity 0.75, specificity 0.74, accuracy 0.74). The proposed algorithm simultaneously used both genetic information and clinical profiles to predict individual patients' response to antipsychotic medications. As the method is not disease-specific but a general algorithm, it can be easily adopted in many other clinical practices for personalized medicine.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The success of personalized medicine requires diagnostic tests that can identify patients who will benefit from targeted treatments (Hamburg and Collins, 2010). However, heterogeneous etiologies of schizophrenia result in high variations in response to a given treatment (Clark et al., 2011; Fanous and Kendler, 2005; Kennedy et al., 2003); schizophrenia is highly heterogeneous in efficacy and liability to side effects of antipsychotics (Pouget et al., 2014). One crucial source of variability in medication response is genetic factors, providing the impetus for applied pharmacogenetics in the selection and treatment regimen for antipsychotic agents in schizophrenia (Adkins et al., 2013; McClay et al., 2011; Need et al., 2009; Pouget et al., 2014; Zhang and Malhotra, 2011). Although existing studies provide clinical evidence of potentially meaningful prediction, the actual prediction for antipsychotic medication efficacy is still suboptimal due to a surfeit of limitations, especially small to modest effect sizes with respect to influencing clinical outcomes (McClay et al., 2011; Motsinger-Reif et al., 2013; Zhang and Malhotra, 2011). Polygenic risk scores (PRSs) were introduced to incorporate multiple genes with small effect sizes and showed potential to explain some variance of the case-control study for the risk of schizophrenia. However, there is still insufficient support for using PRSs in predicting individual predisposition of schizophrenia or personalized effectiveness of antipsychotics (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; The International Schizophrenia Consortium, 2009; Vassos et al., 2017).

In this paper we propose a computational algorithm to predict the effectiveness of antipsychotic medications for patients with schizophrenia. The algorithm directly predicts health outcomes of a specific antipsychotic for individual patients, by simultaneously utilizing multiple predictors including genotypes of single-nucleotide polymorphisms

* Corresponding author at: Department of Mental Health Law & Policy, University of South Florida, Tampa, FL 33612, USA.
E-mail address: beomlee@usf.edu (B.S. Lee).

http://dx.doi.org/10.1016/j.schres.2017.05.001
0920-9964/© 2017 Elsevier B.V. All rights reserved.

Please cite this article as: Lee, B.S., et al., A computational algorithm for personalized medicine in schizophrenia, Schizophr. Res. (2017), http://dx.doi.org/10.1016/j.schres.2017.05.001

(SNPs) as well as demographic and clinical profiles. Our algorithm aligns with a person-centered approach as opposed to a more typical variable-centered approach.

The variable-centered approach starts from an assumption that the population is homogeneous, so that the persons in the same population are viewed as replaceable data points (von Eye and Bogat, 2006). Many conventional statistical methods belong to the variable-centered approach. For example, regression analysis identifies statistically significant variables that contribute "on average" to the treatment effectiveness for the most representative patient. Although it is useful in finding important predictors (e.g., PRSs) that could affect the response variable, it is less useful in predicting an individual patient's response, especially when the population responses are highly heterogeneous, as is the case for patients with schizophrenia. Another disadvantage of using the variable-centered approach is that it typically estimates only the main effects of the predictor variables while failing to capture interactions and nonlinear relationships in them (Bauer and Shanahan, 2007).

On the other hand, the person-centered approach (e.g., cluster analysis, latent class analysis, finite mixture modeling) in social and behavioral sciences assumes that the members of the population are not homogeneous because development and structure of human behavior are not universal but unique to individual persons or groups of individual persons (von Eye and Bogat, 2006). A person-centered approach therefore has several advantages over a variable-centered approach. First, since persons are classified into empirically meaningful subgroups, confounding or spurious relationships among variables do not pose analytical problems. Second, the classifications of persons derived from analyses can be directly generalized to other groups of people with similar characteristics. Lastly, the person-centered approach is able to include all cases even if they are outliers (Everitt et al., 2011). We applied the person-centered approach to the personalized medicine in schizophrenia to overcome the weakness of the variable-centered approach mentioned above. For example, our proposed algorithm is capable of incorporating the correlations and interactions between predictor variables in predicting individual medication effectiveness. The person-centered approach we used properly models complex interactions and confounding relationships among variables (Bauer and Shanahan, 2007).

2. Methods

2.1. Algorithm

Lee et al. (2014) developed a computational algorithm, called Latent Group Effectiveness Modeling (LGEM), and this predictive modeling successfully identified persons who would truly benefit from the motivational enhancement therapy among patients with substance abuse problems, using patients' sociodemographic and clinical profiles. In the present study we used the analytic approach of LGEM in developing a more generalized and improved algorithm to incorporate patients' genotypes as well as sociodemographic and clinical profiles. Following the main concepts of LGEM, we may identify two subgroups of individuals in the treatment group, the one (G) with relatively *good* outcomes and the other (P) with relatively *poor* outcomes (See Fig. 1-A). Then, we may compare two of the observed groups, G and P, in order to identify the characteristics of individuals who might have derived benefits directly from the treatment. However, by itself this method does not provide robust identification of true beneficiaries from the treatment among the individuals in the treatment group.

As presented in Fig. 1-B, the LGEM approach does allow further classification as follows. In addition to distinguishing between individuals in the treatment group with good and poor outcomes, it allows the group G (good outcome) to be further broken down into two unobserved subgroups: (1) a group GE (good outcome, effective) of the individuals who attained *good* outcomes probably because the treatment was *effective* for them and (2) a group GI (good outcome, ineffective) of the individuals who initially attained *good* outcomes probably because of chance or some other reasons, but whose response to the treatment condition subsequently might have deteriorated. In short, the treatment for group GE is likely to be genuinely *effective* while the treatment for group GI may actually be as *ineffective* as for group P. Then, groups GI and P together make the group I (ineffective) against E (effective), as presented in Fig. 1-B.

By advancing one more step as presented in Fig. 1-C, the LGEM revision employed in this paper breaks down group P (poor outcome) into two unobserved subgroups: (1) a group PE (poor outcome, effective) of the individuals who had *poor* outcomes probably because of chance or some other reasons although the treatment was supposed to be *effective* for them and (2) PI (poor outcome, ineffective) of the individuals who attained *poor* outcomes probably because the treatment was *ineffective*. In short, the treatment for group PI is likely to be genuinely *ineffective* while the treatment for group PE may actually be as *effective* as for group G. Then, groups PE and G together make the group E (effective) against I (ineffective), as presented in Fig. 1-C.

The main idea of our modeling is removing possibly uninformative "noise" groups (GI and PE) in order to obtain the genuinely informative reference groups: GE (purely effective) in Fig. 1-B and PI (purely ineffective) in Fig. 1-C. Then, those identified "pure" groups (GE and PI) will be used as reference groups in predicting who will or will not receive benefits from a specific treatment.

The actual decomposition can be done using a variety of techniques available in cluster analysis. Cluster analysis covers a wide range of numerical methods that summarize data with a small number of subgroups or clusters; that is, the clustering procedures generate groups of objects that resemble each other in the same cluster and that are different from the objects in other clusters (Everitt et al., 2011). In the present instance, we used a classical clustering algorithm called the partitioning around medoids (PAM) algorithm (Kaufman and Rousseeuw, 2008) as in Lee et al. (2014). A medoid is the object whose absolute distance is minimal to the other members of the cluster (Everitt et al., 2011), so it is the most representative member in the group. Although the *k*-means algorithm is more frequently used due to its computational simplicity, it is more sensitive to statistical noise and outliers, and in principle it is not suitable for categorical data (Theodoridis and Koutroumbas, 2008). PAM is more robust to noise and outliers, and it can appropriately handle categorical data, which are very common in clinical trials. Also, our algorithm used not only the most representative member of the group (i.e., medoid, called m_1 here) but also possibly the second most representative member of the group (i.e., the object with the second least absolute distance to other members of the cluster, called m_2 here), the third most representative member of the group (i.e., the object with the third least absolute distance to other members of the cluster, called m_3 here), and so on. For the distance measure in our proposed algorithm, we used the dissimilarity measure of Gower (1971) because it can handle both continuous and categorical variables simultaneously and it is capable of calculating the distance measure regardless of the presence of missing values. Each SNP used in computing the Gower dissimilarity measure was regarded as a categorical variable with three possible values: DD, Dd, or dd, where D is the major allele and d is the minor allele.

The algorithm to identify the members of group GE is as follows.

Step 1: Find the r most representative members $[m_1^G, m_2^G, \ldots, m_r^G]$ for group G, then set them to be $[m_1^E, m_2^E, \ldots, m_r^E]$ for group E, as in Fig. 1-B. Likewise, find the r most representative members $[m_1^P, m_2^P, \ldots, m_r^P]$ for group P, then set them to be $[m_1^I, m_2^I, \ldots, m_r^I]$ for group I, as in Fig. 1-B.

Step 2: Relocate the individuals in group G to either group E or I by comparing each individual's mean distances to $[m_1^E, m_2^E, \ldots, m_r^E]$ and $[m_1^I, m_2^I, \ldots, m_r^I]$.

Step 3: Find new sets of $[m_1^E, m_2^E, \ldots, m_r^E]$ and $[m_1^I, m_2^I, \ldots, m_r^I]$ based on the newly assembled E and I.

Please cite this article as: Lee, B.S., et al., A computational algorithm for personalized medicine in schizophrenia, Schizophr. Res. (2017), http://dx.doi.org/10.1016/j.schres.2017.05.001

Download English Version:

https://daneshyari.com/en/article/6821911

Download Persian Version:

https://daneshyari.com/article/6821911

Daneshyari.com