Construction of a neuro-immune-cognitive pathway-phenotype underpinning the phenome of deficit schizophrenia.

Hussein Kadhem Al-Hakeim[a], Abbas F. Almulla[b], Arafat Hussein Al-Dujaili[c], Michael Maes [d,e,f].

[a] Department of Chemistry, College of Science, University of Kufa, Iraq. E-mail: headm2010@yahoo.com.

[b] Medical Laboratory Technology Department, College of Medical Technology, The Islamic University, Najaf, Iraq. E-mail: abbass.chem.almulla1991@gmail.com.

[c] Senior Clinical Psychiatrist, Faculty of Medicine, Kufa University. E-mail: arafat.aldujaili@uokufa.edu.iq.

[d*] Department of Psychiatry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand;

[e] Department of Psychiatry, Medical University of Plovdiv, Plovdiv, Bulgaria;

[f] IMPACT Strategic Research Centre, Deakin University, PO Box 281, Geelong, VIC, 3220, Australia.

Corresponding author

Prof. Dr. Michael Maes, M.D., Ph.D.

Department of Psychiatry

Faculty of Medicine

Chulalongkorn University

Bangkok

Thailand

 doi:10.20944/preprints201910.0239.v1

https://scholar.google.co.th/citations?user=1wzMZ7UAAAAJ&hl=th&oi=ao

dr.michaelmaes@hotmail.com.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Abstract

In schizophrenia, pathway-genotypes may be constructed by combining interrelated immune biomarkers with changes in specific neurocognitive functions that represent aberrations in brain neuronal circuits. These constructs provide insight on the phenome of schizophrenia and show how pathway-phenotypes mediate the effects of genome X environmentome interactions on the symptomatology/phenomenology of schizophrenia. Nevertheless, there is a lack of knowledge how to construct pathway-phenotypes using Partial Least Squares (PLS) path modeling and Soft Independent Modeling of Class Analogy (SIMCA). This paper aims to provide a step-by-step utilization guide for the construction of pathway-phenotypes that reflect aberrations in the neuroimmune - brain circuit axis (NIBCA) in deficit schizophrenia. This NIBCA index is constructed using immune biomarkers (CCL-2, CCL-11, IL-1β, sIL-1RA, TNF-α, sTNFR1, sTNFR2) and neurocognitive tests (Brief Assessment of Cognition in Schizophrenia) predicting overall severity of schizophrenia (OSOS) in 120 deficit SCZ and 54 healthy participants. Using SmartPLS path analysis, a latent vector is extracted from those biomarkers and cognitive tests, which shows a good construct reliability (Cronbach alpha and composite reliability) and replicability and which is reflectively measured through its NIBCA manifestations. This NIBCA pathway-phenotype explains 75.0% of the variance in PHEMN (psychotic, hostility, excitation, mannerism and negative) symptoms. Using SIMCA, we constructed a NIBCA pathway-class that defines deficit schizophrenia as a qualitatively distinct nosological entity and which allows patients with deficit schizophrenia to be authenticated as belonging to the deficit schizophrenia class. In conclusion, our nomothetic approach to develop a nomological network combining neuro-immune and neurocognitive phenome markers to predict OSOS and cross-validate a diagnostic class generated replicable models reflecting the key phenome of the illness, which may mediate the effects of genome X

 doi:10.20944/preprints201910.0239.v1

environmentome interactions on the final outcome phenome features, namely symptomatology and phenomenology.

Keywords: Deficit schizophrenia, machine learning, cytokines, cognition, inflammation, neuro-immune

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Introduction

Recently, we have demonstrated that the symptomatology and phenomenology of deficit schizophrenia, which comprises negative and PHEM (psychosis, hostility, excitation, and mannerism) symptoms as well as formal thought disorders (FTD) and psychomotor retardation (PMR), is to a large extent explained by neuro-immune pathways [1-7]. The latter include signs of activation of the immune-inflammatory response system (IRS) such as increased interleukin-1β (IL-1β), tumor necrosis factor (TNF)-α and chemokines including CCL-2 or MCP1 and CCL-11 or eotaxin, as well as signs of activation of the compensatory immune-regulatory system (CIRS) such as increased serum levels of the soluble IL-1 receptor antagonist (sIL-1RA), which attenuates IL-1 signaling, and sTNFR2 levels, which lowers TNF-α signaling [3,8].

In schizophrenia, products associated with M1 macrophage (TNF-α, IL-1β, CCL-2), T helper (Th)-1 and Th-2 (CCL-11) functions show multiple neuronal effects including lowered neuroprotection and neurogenesis as well as neurotoxic and excitotoxic effects, which together may cause damage to neuronal circuits in the brain, herein denoted as neuroprogression [1-8]. As such, IRS activation could affect brain tissues via M1 (TNF-α, IL-1β and CCL-2) and Th-2 / eosinophil (CCL-11)-associated products leading to neurocognitive deficits including impairments in episodic and sematic memory and executive functions [1-3]. The latter indicate that brain neural circuits involving the dorsolateral prefrontal cortex, as well as prefronto-striato-thalamic, prefronto-parietal, prefronto-temporal, and amygdala and hippocampal neural networks are affected in deficit schizophrenia [9,10].

The neurocognitive phenomenology of deficit schizophrenia as well as the above-mentioned biological pathways are different manifestations of the phenome of this disorder, whereby “phenome” denotes the collection of all phenotypes including biochemical, behavioral, and cognitive features as well as the phenomenology and symptomatology of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

disorder. It is probable that common causal factors, namely interactions between the genome and the environmentome [11] underpin the biological pathways and associated damage to neuronal circuits [12-14]. **Figure 1** shows the associations between the genome, environmentome and the phenome. As such, both the neuronal brain circuits and IRS/CIRS pathways, which constitute the pathophysiology of deficit schizophrenia, are the measurable psycho-organic manifestations of those genome X environmentome interactions.

While the current state-of-the-art methodology does not allow to measure directly the genome X environmentome interactions, the construction of a pathway-genotype, which combines a series of upstream and downstream biomarkers with the changes in specific behaviors/cognitive functions caused by those pathways, may give insight as to how these genome X environmentome interactions cause specific psycho-organic consequences and thus the phenome of the disorder. These pathway-phenotypes not only reflect the impact of the pathophysiology of the disorder on behavior/cognition but also provide an index of disease risk in a specific individual as well as in the general population and, therefore, the construction of pathway-phenotypes may improve drug discovery processes.

Different machine learning techniques may be used to construct pathway-phenotypes, including Partial Least Squares (PLS) analysis, which is structural equation modeling method that uses pathway modeling on latent constructs extracted from a set of indicator variables [2-6]. PLS allows to validate newly constructed latent vectors (LV) that reflect a pathway-phenotype combining neuro-immune biomarkers and neurocognitive functions in a reflective model. The latter is more adequate than a formative model because a common cause, namely a disorder in the neuroimmune - brain circuit axis (NIBCA), underpins the LV extracted from those data [6]. For example, a pathway-phenotype reflecting a biomarker-staging phenotype was computed in affective disorders by combining a lipid-associated antioxidant enzyme,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

which may affect a cascade of redox-related responses (namely paraoxonase 1), and staging characteristics, including number of episodes and suicidal attempts [15].

Another useful machine learning method to delineate pathway-classes is Soft Independent Modeling of Class Analogy (SIMCA) [16-20]. SIMCA is a supervised machine learning technique which is well-suited to examine whether an existing diagnostic class (e.g. deficit schizophrenia) or a newly developed class (e.g. results of cluster-analysis) is a qualitatively distinct pathway-class which may be discriminated from other classes based on biological and neurocognitive features [1,18-20]. As such, the construction of a pathway-class may delineate a group of patients who belong to a qualitatively distinct nosological entity based on pathway and phenomenological features [1,18-20]. Using SIMCA, we constructed a pathway-class based on the diagnosis of deficit schizophrenia and using IgA levels directed to tryptophan catabolites (TRYCATs), symptoms and neurocognitive impairments as discriminatory variables [18,19]. As such, the clinical diagnosis “deficit schizophrenia” was externally validated as a pathway-class, which reflects the pathophysiology of this condition, namely NIBCA disorders with increased levels of neurotoxic TRYCATs combined with affected brain circuits that mediate episodic and semantic memory and executive functions as well [18,19].

Nevertheless, no research has examined whether indices of increased IL-1β, TNF-α and CCL-2 and CCL-11 signaling may combine with neurocognitive impairments to shape a pathway-phenotype or pathway-class which reflects the key-phenome of the illness. Moreover, there is a lack of knowledge how to construct pathway-phenotypes and pathway-classes using specific machine learning techniques such as PLS path modeling and SIMCA.

Hence, the present paper aims to explain the adequate methods that should be used to construct pathway-genotypes and pathway-classes. As examples, we show that indices of IL-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

1β, TNF-α and CCL2/11 signaling coupled with neurocognitive deficits may be used towards that purpose.

Methods

Based on our experience with pathway analysis and PLS path modeling [1-4,21] and SIMCA [16,17] we review the steps needed to construct two different types of path-phenotypes, namely latent vectors extracted from a set of indicators using PLS and classes modeled with the same input data in SIMCA models (pathway-class). We will explain how this NIBCA index is constructed using immune biomarkers (CCL-2, CCL-11, IL-1β, sIL-1RA, TNF-α, sTNFR1, sTNFR2) and neurocognitive tests (Brief Assessment of Cognition in Schizophrenia, BRACS) [22] and how the constructed pathway-phenotype may predict a latent vector extracted from the PHEMN symptoms of deficit schizophrenia. Previously we have published that all PHEMN symptoms show highly significant loadings on a LV which is reflectively measured through all PHEMN symptoms and which, therefore, reflects overall severity of schizophrenia (OSOS) [6]). Such findings indicate that a common construct (OSOS) determines negative and PHEM symptoms. Moreover, if there are strong associations between the NIBCA index and the latent trait OSOS it indicates that the NIBCA (early phenome) features may cause at least in part the late phenome features including the phenomenology and symptomatology of the illness.

Towards this end, this study will employ our published data on 120 deficit patients and 54 healthy participants [1]. The socio-demographic data of the patients and controls as well as the measurement in both study groups were presented in our previous paper [1]. In that paper we also defined the inclusion and exclusion criteria for patients and controls as well as the methods used to assay the immune markers and to assess the BACS cognitive tests [22]. The latter comprises “List Learning (probing verbal episodic memory), Digit Sequencing Task (probing working memory), Category Instances (semantic fluency) and Controlled Word

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Association (letter fluency) (both probing verbal fluency and semantic memory), Symbol Coding (probing attention) and Tower of London (probing executive functions, reasoning and problem solving)" [1]. We also assessed the Mini-Mental State Examination (MMSE). In order to assess severity of negative symptoms we used the total score on the SDS [23], the score of the negative subscale of the Positive and Negative Syndrome Scale (PANSS) [24] and the total score on the Scale for the Assessments of Negative Symptoms (SANS) [25]. The severity of PHEM symptoms was estimated using z unit-weighted composite scores based on items of the positive subscale of the Positive and Negative Syndrome Scale (PANSS) [24], the Brief Psychiatric Rating Scale (BPRS) [26] and the Hamilton Depression Rating Scale (HDRS) [27]. The computation of these composites is explained at large in for example by Maes et al. [7]. We have grouped the neuro-immune biomarkers into three relevant profiles, namely z unit-weighted composite scores reflecting IL-1β signaling (computed as zIL-1β + zsIL-1RA), TNF-α signaling (computed as zTNF-α + zsTNFR1 + zsTNFR2) and neurotoxic chemokines signaling (computed as zCCL-2 + zCCL-11) [28].

All controls and patients, as well as the guardians (parents or close family members) of patients, gave written informed consent prior to participation in our study. The study was conducted according to International and Iraq ethics and privacy laws. Approval for the study was obtained from the Institutional Review Board of the University of Kufa (347/2019), which is in compliance with the International Guidelines for Human Research protection as required by the Declaration of Helsinki, The Belmont Report, Council for International Organizations of Medical Sciences (CIOMS) Guideline and International Conference on Harmonization on Good Clinical Practice (ICH-GCP).

Machine learning tests

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

The use of joint principal component analysis (PCA) with a 20-fold cross-validation scheme and standard deviation weighting process, was explained previously [1]. PCA performed on the schizophrenia and control subjects was conducted to display the distribution of both classes in a two-dimensional space whereby both groups are differentiated by marker color and shapes [29]. In addition, the correlation loading plot should be inspected to evaluate the correlation loadings for all variables on PC1 and PC2 with respect to two ellipses whereby the outer ellipse indicates 100% explained variance and the inner ellipse indicates 50% of the explained variance. Al-Hakeim et al. [1,20] also explained the use and merits of Support Vector Machine (SVM), which is useful for classification purposes, with linear kernel (linear SVM) and radial basis function (RBF SVM) [29]. SVM classification results should be cross-validated using a 10-fold cross-validation scheme.

SIMCA is a class modeling technique [29], which constructs PCA models around the patient and control classes using the neuro-immune markers and neurocognitive test scores as modeling variables [1,20]. We divided the participants in two sets, namely a training or calibration set (50% of controls and 50% of schizophrenia and patients) while the remaining patients were used as the validation set. Cross-validation determines the number of PCs used to construct the SIMCA models. Outliers are deleted from this class modeling based on influence, stability and Hotelling's T2 vs samples plots, and thorough inspection of residual values and leverages [1]. Two distances, namely the distance to the class model (Si) and the distance to the class center (Hi or leverage) are used to compute critical limits that define the classes and to decide whether cases belong to a class or are rejected to belong to any class [1]. In this paper, we will use 3 types of SIMCA figures a) the Coomans plot which displays the distances to the models of two classes applying a 5% confidence interval. This plot divides the 2D space into four quadrants, whereby the vertical line indicates the limit of one class while the horizontal line indicates the membership limits of the other class. Samples that fall within

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

the upper right quadrant are considered to be outsiders not belonging to either of the classes, whereas cases that fall in the lower left quatrant could be considered hybrids (that is members that may belong to both classes). b) The Si/Hi plot, which shows the relative distance of the subjects to the class model and the distance to the class centre (leverage or Hi). The critical class limits are, subsequently used to authenticate cases when they are allocated to their target class, to identify aliens when cases belonging to another class intrude into the target class, and to identify outsiders when the cases fall outside the class membership limits [1]. c) The discrimination plot, which shows the discrimination power of all input variables discriminating schizophrenia patients from healthy controls. Moreover, SIMCA also produces two figures of merit a) the confusion matrix with the accuracy of the classification and sensitivity and specificity; and b) the model-to-model distance, which shows the degree of separation of the class models whereby a distance > 3 indicates a significant difference between the groups, values close to zero that there are no differences, and values > 30 that there are highly significant differences.

SmartLPS

Previously we have explained the use of path analysis and Smart PLS path analysis in psychiatric research [1,2,4-6,21]. Here we show how to employ SmartPLS to examine a) whether a LV may be extracted from immune and cognitive data (the pathway-phenotype); b) whether this pathway-phenotype LV predicts a LV extracted from PHEMN symptoms; and c) whether the effects of the immune biomarkers on the PHEMN LV are mediated by neurocognitive deficits [1,2,6]. Complete and consistent bootstrapping using 5000 bootstrap samples is carried out when the model fit data are adequate namely SRMR < 0.080 and the outer model (LVs) show adequate construct reliability with composite reliability > 0.7, Cronbach's alpha > 0.7, rho_A > 0.80 and with an average variance extracted (AVE) > 0.500

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

while all loadings on the LVs should be > 0.6 at p<0.01. Blinfolding is used to examine whether the cross-validated communalities or redundancies are adequate [30]. Confirmatory Tetrad Analysis (CTA) is used to examine whether the models are not misclassified as reflective models. Consequently we computed path coefficients with p-values, and total effects, total indirect effects and specific indirect effects.

Construction of the pathway-phenotypes

PCA plot

In order to display the distribution of normal controls and SCZ patients with respect to their biomarker and neurocognitive features we conducted a PCA on the 7 biomarkers (IL-1, IL-1RA, TNF, sTNFR1, sTNFR2, CCL2 and CCL11) and the 7 neurocognitive test scores (MMSE, WLM, WM, VFT, COWA, SC and TOL). **Figure 2** displays the distribution of the patients and controls in a two-dimensional space defined by PC1 (explaining 56% of the variance in the data) and PC2 (explaining 8% of the variance). PC3 (7%), PC4 (6%), PC5 (5%), PC6 (5%) and PC7 (4%) contribute less to the variance in the data set. Controls are shown as blue squares while patients are displayed as red dots. Both categories group together, controls at the right-hand side of the graph and patients at the left-hand side. Moreover, there is wide gap (street) between both classes indicating a strong separation using the biomarker and neurocognitive features. The same pattern of separation can be observed in the sequential plots PC1 versus PC3, PC1 versus PC4, PC1 versus PC5, PC1 versus PC6, and PC1 versus PC7 and as such, the first PC allows to differentiate patients from controls.

The correlation loading plot, which displays the correlation loadings for all 14 features in PC1 and PC2, shows that all neurocognitive features contribute highly significantly to the differentiation of both classes and that also CCL-2 and CCL-11 are significant discriminators.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Other PC combinations also show that TNF-α and sTNFR2 (PC1 and PC4), sTNFR2 (PC1 and PC5), IL-1β (PC1 versus PC6) have some discriminatory power.

Support vector machine

In order to compute an algorithm which allows an adequate separation of patients from controls we conducted RBF SVM classification with ten-fold cross-validation. SVM delineated 33 support vectors namely 14 normal controls and 19 patients. The classification showed a training accuracy of 100% and a validation accuracy of 100%. **Figure 3** shows a plot with the classification results using COWA and CCL-11 as discriminatory features. Controls are again displayed as blue squares and patients as red dots showing that there is a perfect separation of both classes with a wide street between the two classes.

SIMCA

We constructed a SIMCA model using a training set comprising 50% of all patients and 50% of all heathy controls, whereas the other participants served as validation set. One patient was considered to be a statistical outlier and consequently this case was deleted from further modeling. No normal controls were identified as outliers and therefore all controls of the training set were used to compute the SIMCA model. Both controls and patients were modeled using 7 PCs. A Coomans plot was employed to visually evaluate the classification whereby cases are classified according to their distances to both class models developed by SIMCA. **Figure 4** shows the Coomans plot for the comparison of controls and patients applying the 5% confidence interval. Red dots represent patients, while blue dots represent normal controls (both with sample identity numbers). This plot divided the 2D space into four quadrants, whereby the vertical line indicates the limit of normal cotrols indicating that values left of the line show a small distance to the control class and thus should be identified as controls. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

horizontal line indicates the membership limit of the schizophrenia class with values under this line showing a small distance to the schizophrenia class. Cases that fall within the upper right quadrant are considered to be outsiders that do not belong to either the controls or patient class, while cases that fall in the lower left quatrant could be considered hybrids (that is members that may belong to both classes). One normal control and 2 patiens fall within the lower left quadrant but these 3 cases still appear to aggregate with the other class members which are all allocated to the left upper (control class) and lower right quadrant (patient class). The green dots in Figure 4 are the cases in the cross-validation set (with sample identity).

Figure 5 displays the Si/S0 or patient group-membership plot which shows the distances of all cases belonging to the cross-validation set to the schizophrenia model (y-axis) and the schizophrenia model centre or leverage (x-axis). As such, the lower left quadrant represents the group limits of the patient class model as constructed using the training sample. In this figure patients are represented by red dots and controls by blue squares. All patients (except one) were authenticated as belonging to the schizophrenia class and there were no aliens (that is controls intruding in the patient class). The Si/S0 or control group-membership plot showed that all controls (except one) were correctly authenticated and that one patient (the same case as in the schizophrenia group-membership plot) intruded into the control group limits. As such, cross-validation showed that one control and one patient were misclassified yielding a sensitivity of 98.3% and a specificity of 96.3%.

SIMCA analysis showed that the inter-class distance between both models was 38.67 indicating a very good discrimination. **Figure 6** shows the discrimination plot of the 7 biomarkers and the 7 neurocognitive test scores separating the schizophrenia class from the control class. The top-5 features of schizophrenia are in descending order: IL-1β (discriminatory power=9.98), MMSE (9.07), sTNFR2 (8.36), sIL-1RA (7.68), and COWA

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

(7.32). In fact, all biomarkers and all neurocognitive tests have a significant discriminatory power.

Smart PLS analysis

Figure 7 shows the results of PLS-SEM pathway analysis which examined the direct and indirect effects of immune activation (entered as a formative LV extracted from three composite scores, namely ALL-IL-1, All-TNF and CCL2+11) on schizophrenia symptoms (entered as a reflective OSOS LV extracted from 3 negative scale scores, PHEM symptoms, FTD and PMR). This pathway also examines the indirect effects of immune activation mediated by neurocognitive impairments, which was entered as a reflective LV extracted from 5 test scores, namely COWA, TOL, WLM, WM and VFT. The other two cognitive tests did not load highly on this LV and were, therefore, eliminated from the analysis. The overall fit of the path model was adequate with SRMR=0.026, while also the construct reliability of the two reflective LVs was excellent with Cronbach's alpha > 0.944, composite reliability > 0.957, rho_A > 0.953 and average variance extracted > 0.817 and with outer model loadings on the neurocognitive LV > 0.854 and the OSOS LV > 0.876. All three biomarkers contributed significantly (weights between 0.365 and 0.490) to the immune activation LV with All-IL1 showing the highest weight and CCL2+11 the lowest. We found that 75.0% of the variance in the OSOS LV could be explained by the regression on the neurocognitive LV, the biomarker LV, and age, while 69.7% of the variance in the neurocognitive LV was explained by the biomarker score. There were significant specific indirect effects of the biomarker score on OSOS LV, which was mediated by the neurocognitive LV ($t=+8.94$, $p<0.001$). There was also a significant specific indirect effect of education on OSOS LV mediated by neurocognition ($t=-6.34$, $p<0.001$). Blindfolding showed that both the neurocognitive and OSOS LVs has adequate

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

cross-validated reduncancies (>0.529). Confirmatory Tetrad analysis showed that both the OSOS and neurocognitive LVs were not misclassified as reflective models.

Based on this PLS analysis we have then constructed a more parsimonious model whereby elevated IL-1β, TNF-α and chemokine signaling as well as the affected neuroanatomical pathways underpinning neurocognitive impairments were considered to be manifestations of a common cause, i.e. the NIBCA phenome. **Figure 8** shows this model with OSOS LV as the outcome variable and the NIBCA LV extracted from the biomarkers composites and neurocognitive tests as the input variables, while controlling for education, age and sex. The SRMR value was 0.034 indicating a good fit of the model. The construct reliabilities of the NIBCA and OSOS reflective LVs were adequate with Cronbach's alpha > 0.926, composite reliability > 0.941, rho_A > 0.945 and average variances extracted > 0.670. The loadings on both LVs were all > 0.629 (NIBCA LV) and > 0.876 (OSOS LV). We found that 74.4% of the variance in the OSOS index was explained by regression on the NIBCA LV and age (inversely associated). Blindfolding showed that both the OSOS (0.771) and NIBCA (0.561) showed adequate construct cross-validated communalities. CTA showed that both the OSOS and neurocognitive LVs were not misclassified as reflective models.

Discussion

In this study we applied machine learning methods to construct a new pathway-phenotype and pathway-class, which represent the key component of the phenome of deficit schizophrenia. **Figure 9** displays a flow chart of the methods employed to develop such models.

Step 1. Define the biological pathways and associated clinical or phenomenological features which are important to understand the phenome of the disorder. More specifically, activated neuro-immune pathways are key pathways underpinning the pathophysiology of deficit schizophrenia while impairments in episodic and sematic memory as well as deficits in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

executive functions are other important features of the phenome [1,2,18,19]. Indeed, preclinical evidence shows that increased IL-1β, TNF-α and CCL-2/CCL-11 signaling may impact brain circuits through different mechanisms thereby impacting neurocognition (see Introduction). Most importantly, there are also indicants that the neuro-immune pathways and neurocognitive deficits examined here may cause PHEMN symptoms [1,2,18,19] and that a common etiological factor may underpin those pathways and neurocognitive deficits, namely interactions between the genome and the environmentome (see Introduction). As such, there is theoretical evidence to bring immune markers and neurocognitive impairments together in constructs to be modeled by PLS or SIMCA. Of course, it would have been even more interesting if we had measured other neuro-immune biomarkers that play a key role in deficit schizophrenia including deficits in innate immunity such as lowered natural IgM responses and lowered paraoxonase 1 activities, which together increase vulnerability to immune activation, oxidative stress, bacterial load and inflammation [4,31]. Another immune and oxidative stress-associated pathway that is activated in deficit schizophrenia is the tryptophan catabolite (TRYCAT) pathway with increased levels of neurotoxic TRYCATS, including xanthurenic acid, picolinic acid and 3-OH-kunerenine [18,19]. Finally, it would be most interesting to enlarge this set of early phenome markers with brain imaging measurements especially MRI and functional MRI so that pathway-phenotypes could be constructed that directly combine the neurotoxic immune pathways with the tissues (brain) affected and the cognitive symptoms emerging as a consequence of these interactions.

Step 2. Perform PLS and evaluate the reliability of the new pathway-phenotype. In the present paper we have explained how to conduct PLS analysis to construct a pathway-phenotype. Important is that we combined early phenome markers, namely the input indicators including neuro-immune markers and neurocognitive tests, in one construct and examined the validity and unidimesionality of the latent trait extracted from these phenome markers. In

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

addition, we examined the association between these early phenome features and the late phenome indicators namely the PHEMN symptoms, PMR and FTD.

Using bootstrapping (5000 bootstrap samples) to compute path coefficients revealed that the latent phenomenon extracted from the early phenome data predicted an important part of the LV extracted from the PHEMN, PMR and FTD symptoms and that the latter is unidimensional and is reflectively measured through its symptomatic manifestations. This is important because the different symptoms may be conceptualized under an overall single latent vector, reflecting OSOS, which is the cause of the different symptomatic features of the illness [2-4]. Also, the pathway-phenotype showed adequate psychometric properties with good internal consistency reliability, convergent reliability, predictive relevance and construct replicability, indicating that the NIBCA LV is a unidimensional and reliable latent construct. Moreover, also this latent phenomenon is reflectively measured, as assessed with CTA, through biomarker and cognitive features. As such, the NIBCA construct and its etiology (genome X envorimentome) are the common cause of the pathway and cognitive manifestations which are to a large extent determined by NIBCA LV. These findings indicate that the neuro-immune disorders and the organic substrate of the cognitive deficits (e.g. aberrations in long-term potentiation, synaptic sampling and functions, neuroplasticity, and neurogenesis in specific brain regions) are tightly associated phenomena. This is important because the NIBCA index is the measurable part of the phenome that mediates the effects of the genome X environmentome on the late phenome features (the symptoms or phenomenology). As such, our NIBCA index (or a novel NIBCA index enriched with other immune and brain MRI markers) should be used as the new phenome index in genome and environmentome studies and should be considered as a new drug target in schizophrenia research.

STEP 3. Construction of a pathway class using SIMCA. While PLS is an excellent method to construct latent phenomena and examine causal associations, SIMCA is well suited

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

to construct class models using PCA. As explained before [1] one should first examine the distribution of the cases and controls in a two-dimensional PC plot. When clear clusters are observed and both groups are well-separated in this two-dimensional display, it is worthwhile to perform SIMCA on the predefined classes [1,29]. If there are no clear boundaries between the groups, SIMCA will not yield any significant results, but it may be worthwhile to examine whether other clusters of patients may be detected using different unsupervised learning techniques (e.g. cluster analysis), which can then be validated by SIMCA [19]. After cross-validation, SIMCA offers important figures of merit including the model-to-model distance and the confusion matrix. A large model-to-model distance (e.g. > 30) indicates that the multidimensional models constructed around the diagnostic classes are far away from each other in the multidimensional hyperspace and, therefore, that these models are qualitatively different from each other [16-19]. As such, SIMCA can be used to examine whether diagnostic classes are quantitatively or qualitatively different [16-19]. Here we detected a model-to-model distance of > 38 indicating that deficit schizophrenia is a qualitatively different class based on its neuro-immune and neuro-cognitive features. Previously we have shown that based on such biomarkers and cognitive tests, major neurocognitive psychosis (MNP) (largely overlapping with deficit schizophrenia although with more restrictive criteria) is a qualitatively different class as compared with non-deficit schizophrenia [19]. Moreover, the results of the present study show that these features yield a very good classification accuracy for deficit schizophrenia, indicating that the clinical diagnosis of deficit schizophrenia, based on negative symptoms, reflects this pathway-phenotype and, therefore, that deficit schizophrenia may be considered to be a NIBCA pathway-class. It is interesting to note that the combination of NIBCA features in a pathway-phenotype may be used as a reliable index to predict overall severity of schizophrenia (OSOS) as well as the clinical diagnosis of deficit schizophrenia.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

In conclusion, our nomothetic approach and development of a nomological network combining neuro-immune and neurocognitive phenome features to cross-validate diagnostic classes and predict overall severity of illness is an awaited achievement that may radically change the way psychiatry is perceived and practiced. As such, psychiatry will adopt a more nomothetic approach culminating in a biomedical language with pathway-based taxonomies describing the features of pathway-phenotypes and pathway-classes. For example, recently we proposed that, based on results of machine learning, deficit schizophrenia should be renamed as MNP and non-deficit schizophrenia as simple neuro-cognitive psychosis (SNP) [19]. As such, the construction of pathway-classes may sometimes validate older classifications (as we validated the deficit schizophrenia construct in [18]), refute older classifications (e.g. paranoia in [19]), refine older classification (see MNP in [19]) or discover new classifications through unsupervised machine learning [12].

Such discoveries have clinical consequences because, based on SIMCA results, it may be argued that deficit schizophrenia or MNP and non-deficit schizophrenia or SNP should be a introduced as separate diagnostic categories in future classification systems including ICD and DSM [19]. We have shown that the construction of pathway-phenotypes through machine learning including cross-validation techniques generates replicable constructs reflecting the early key phenome of the illness, which mediates the effects of genome X environmentome interactions on the late phenome features, namely symptomatology and phenomenology. A future aim of this research is to integrate the nomothetic approach linking the genome, environmentome and phenome through the construction of pathway-phenotypes with the idiographic approach in order to explain a person's unique phenomenology, which may serve the purpose to develop personalized treatments.

Acknowledgment

We acknowledge the staff of Ibn-Rushd hospital for their help in the collection of samples especially Dr. Mokhlad Swadi Abed. We also acknowledge the work of the high-skilled staff of Asia Clinical Laboratory in Najaf city for their help in the ELISA measurements and Asia lab in the estimation of biomarkers levels.

Funding

There was no specific funding for this specific study.

Conflict of interest

The authors have no conflict of interest with any commercial or other association in connection with the submitted article.

Author's contributions

All the contributing authors have participated in the preparation of the manuscript.

References

1. Sirivichayakul, S.; Kanchanatawan, B.; Thika, S.; Carvalho, A.F.; Maes, M. A new schizophrenia model: immune activation is associated with induction of different neurotoxic products which together determine memory impairments and schizophrenia symptom dimensions. CNS Neurol Disord Drug Targets, 2019, 18(2), 124-140.

2. Sirivichayakul, S.; Kanchanatawan, B.; Thika, S.; Carvalho, A.F.; Maes, M. Eotaxin, an Endogenous Cognitive Deteriorating Chemokine (ECDC), Is a Major Contributor to Cognitive Decline in Normal People and to Executive, Memory, and Sustained Attention Deficits, Formal

Thought Disorders, and Psychopathology in Schizophrenia Patients. Neurotox Res, 2019, 35(1), 122-138.

3. Al-Hakeim, H.K.; Al-Mulla, A.F.; Maes, M. The neuro-immune fingerprint of major neuro-cognitive psychosis or deficit schizophrenia: a supervised machine learning study. Preprint 2019, doi:10.20944/preprints201905.0285.v1. Neurotox Res, in press.

4. Maes, M.; Kanchanatawan, B.; Sirivichayakul, S.; Carvalho, A.F. In Schizophrenia, Deficits in Natural IgM Isotype Antibodies Including those Directed to Malondialdehyde and Azelaic Acid Strongly Predict Negative Symptoms, Neurocognitive Impairments, and the Deficit Syndrome. Mol Neurobiol, 2019, 56, 5122-35.

5. Maes, M.; Kanchanatawan, B.; Sirivichayakul, S.; Carvalho, A.F. In Schizophrenia, Increased Plasma IgM/IgA Responses to Gut Commensal Bacteria Are Associated with Negative Symptoms, Neurocognitive Impairments, and the Deficit Phenotype. Neurotox Res, 2019, 35(3), 684-698.

6. Maes, M.; Sirivichayakul, S.; Kanchanatawan, B.; Carvalho, A.F. In schizophrenia, psychomotor retardation is associated with executive and memory impairments, negative and psychotic symptoms, neurotoxic immune products and lower natural IgM to malondialdehyde. Preprints, 2019, doi: 10.20944/preprints201901.0108.v1.

7. Maes, M.; Sirivichayakul, S.; Kanchanatawan, B.; Vodjani, A. Breakdown of the paracellular tight and adherens junctions in the gut and blood brain barrier and damage to the

vascular barrier in patients with deficit schizophrenia. Neurotox Res, 2019, in press, doi: 10.1007/s12640-019-00054-6.

8. Roomruangwong, C.; Noto, C.; Kanchanatawan, B.; Anderson, G.; Kubera, M.; Carvalho, A.F.; Maes M. The Role of Aberrations in the Immune-inflammatory Reflex System (IRS) and the Compensatory Immune-regulatory Reflex System (CIRS) in Different Phenotypes of Schizophrenia: The IRS-CIRS Theory of Schizophrenia. Mol Neurobiol, 2019, doi: 10.1007/s12035-019-01737-z.

9. Orellana, G.; Alvarado, L.; Muñoz-Neira, C.; Ávila, R.; Méndez, M.F.; Slachevsky, A. Psychosis-related matricide associated with a lesion of the ventromedial prefrontal cortex. J Am Acad Psychiatry Law, 2013, 41(3), 401-106.

10. Orellana, G.; Slachevsky, A. Executive functioning in schizophrenia. Front Psychiatry, 2013, 4, 1–15.

11. Swan, M. Next-Generation Personal Genomic Studies: Extending Social Intelligence Genomics to Cognitive Performance Genomics in Quantified Creativity and Thinking Fast and Slow. Association for the Advancement of Artificial Intelligence (www.aaai.org). Data Driven Wellness: From Self-Tracking to Behavior Change: Papers from the 2013 AAAI Spring Symposium

12. Anderson, G.; Maes, M. Schizophrenia: linking prenatal infection to cytokines, the tryptophan catabolite (TRYCAT) pathway, NMDA receptor hypofunction, neurodevelopment and neuroprogression. Prog Neuropsychopharmacol Biol Psychiatry, 2013, 42, 5-19.

13. Davis, J.; Moylan, S.; Harvey, B.H.; Maes, M.; Berk, M. Neuroprogression in schizophrenia: Pathways underpinning clinical staging and therapeutic corollaries. Aust NZ J Psychiatry, 2014, 48, 512-529.

14. Davis, J.; Eyre, H.; Jacka, F.N.; Dodd, S.; Dean, O.; McEwen, S.; Debnath, M.; McGrath, J.; Maes, M.; Amminger, P.; McGorry, P.D.; Pantelis, C.; Berk, M. A review of vulnerability and risks for schizophrenia: Beyond the two hit hypothesis. Neurosci Biobehav Rev, 2016, 65, 185-194.

15. Maes, M.; Moraes, J.B., Congio, A., Bonifacio, K.L., Barbosa, D.S., Vargas, H.O.; Michelin, A.P., Carvalho, A.F., Nunes, S.O.V. Development of a Novel Staging Model for Affective Disorders Using Partial Least Squares Bootstrapping: Effects of Lipid-Associated Antioxidant Defenses and Neuro-Oxidative Stress. Mol Neurobiol, 2019, 56(9), 6626-6644.

16. Maes, M.; Schotte, C.; Maes, L.; Cosyns, P. Clinical subtypes of unipolar depression: Part II. Quantitative and qualitative clinical differences between the vital and nonvital depression groups. Psychiatry Res, 1990, 34(1), 43-57.

17. Maes, M.; Twisk, FN.; Johnson, C. Myalgic Encephalomyelitis (ME), Chronic Fatigue yndrome (CFS), and Chronic Fatigue (CF) are distinguished accurately: results of supervised learning techniques applied on clinical and inflammatory data. Psychiatry Res, 2012, 200(2-3), 754-60.

18. Kanchanatawan, B.; Sriswasdi, S.; Thika, S.; Sirivichayakul, S.; Carvalho, A.F.; Geffard, M.; Kubera, M.; Maes, M. Deficit schizophrenia is a discrete diagnostic category defined by neuroimmune and neurocognitive features: results of supervised machine learning. Metab Brain Dis, 2018, 33(4), 1053-1067.

19. Kanchanatawan, B.; Sriswasdi, S.; Thika, S.; Stoyanov, D.; Sirivichayakul, S.; Carvalho, A.F.; Geffard, M.; Maes, M. Towards a new classification of stable phase schizophrenia into major and simple neuro-cognitive psychosis: Results of unsupervised machine learning analysis. J Eval Clin Pract 2018, 24(4), 879-891.

20. Al-Hakeim, H,K,; Al-Fadhel, S.Z.; Al-Dujaili, A.H.; Carvalho, A.; Sriswasdi, S.; Maes, M. Development of a Novel Neuro-immune and Opioid-Associated Fingerprint with a Cross-Validated Ability to Identify and Authenticate Unknown Patients with Major Depression: Far Beyond Differentiation, Discrimination, and Classification. Mol Neurobiol, 2019, May 23. doi: 10.1007/s12035-019-01647-0. [Epub ahead of print] PubMed PMID: 31124079.

21. Maes, M.; Vandervorst, C.; Suy, E.; Minner, B.; Raus, J. A multivariate study of simultaneous escape from suppression by dexamethasone of urinary free cortisol, plasma cortisol, adrenocorticotropic hormone and beta-endorphin in melancholic patients. Acta Psychiatr Scand, 1991, 83(6), 480-491.

22. Keefe, R.S.E.; Goldberg, T.E.; Harvey, P.D.; Gold, J.M.; Poe, M.P.; Coughenour, L. The brief assessment of cognition in schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. Schizophr Res, 2004, 68, 283-297.

 doi:10.20944/preprints201910.0239.v1

23. Kirkpatrick, B.; Buchanan, R.W.; McKenney, P.D.; Alphs, L.D.; Carpenter, WT. The schedule for the deficit syndrome: an instrument for research in schizophrenia. Psychiatry Res, 1989, 30, 119-123.

24. Kay, S.R.; Fiszbein, A.; Opler, L.A. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987, 13, 261-276.

25. Andreasen, N.C. The scale for the assessment of negative symptoms (SANS): conceptual and theoretical foundations. Brit J Psychiatry, 1989, Suppl 7, 49-58.

26. Overall, J.E.; Gorham, D.R. The brief psychiatric rating scale. Psycholog Rep, 1962, 10, 799-812.

27. Hamilton, M. A rating scale for depression. J Neurol Neurosurg Psychiatry, 1960, 23, 56-62.

28. Maes, M.; Carvalho, A.F. The Compensatory Immune-Regulatory Reflex System (CIRS) in Depression and Bipolar Disorder. Mol Neurobiol, 2018, 55(12), 8885-8903.

29. CAMO, 2019. The Unscrambler Appendices: Method References, 2019. As assessed 19-3-2019. www.camo.com/helpdocs/The_Unscrambler_Method_References.pdf.

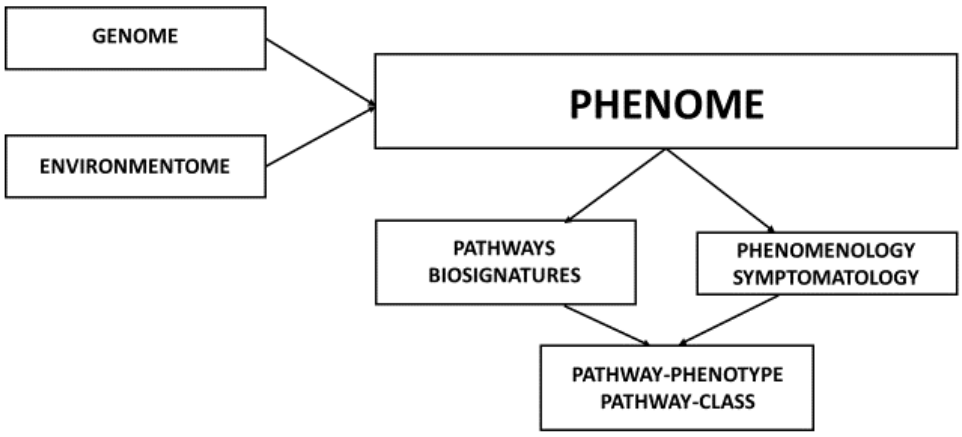
30. Ringle, C.M.; Wende, S.; Becker, J.-M. SmartPLS 3. Bönningstedt: SmartPLS, 2015. Retrieved from http://www.smartpls.com.

 doi:10.20944/preprints201910.0239.v1

31. Matsumoto, A.K.; Maes, M.; Maes, A.; Michelin , A.P..; de Oliveira Semeão, L.; de Lima Pedrão, J.V.. Moreira, E.; Kanchanatawan, B.; Barbosa , D.S.. In Schizophrenia, PON1 Q192R Genotypes and/or Lowered Paraoxonase 1 (PON1) Enzymatic Activity are Significantly Associated with the Deficit Syndrome, Negative Symptoms, Formal Thought Disorders, Psychomotor Retardation, Excitation and Increased IgA Levels to Gram-Negative Microbiota. *Preprints* **2019**, 2019090095 (doi: 10.20944/preprints201909.0095.v1).

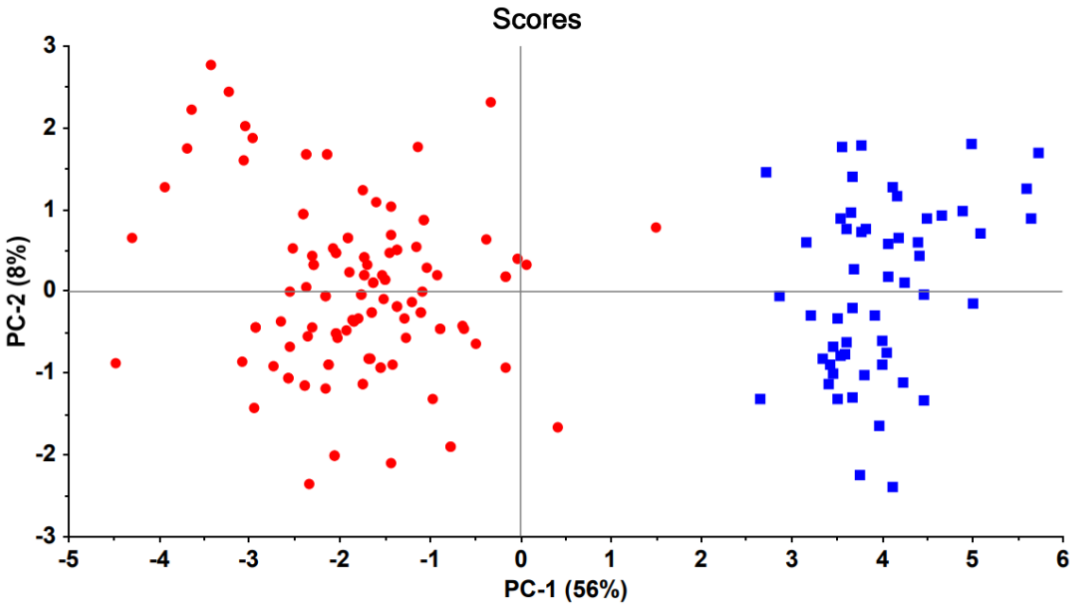
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Figure 1. Associations between the genome, environmentome and the phenome.



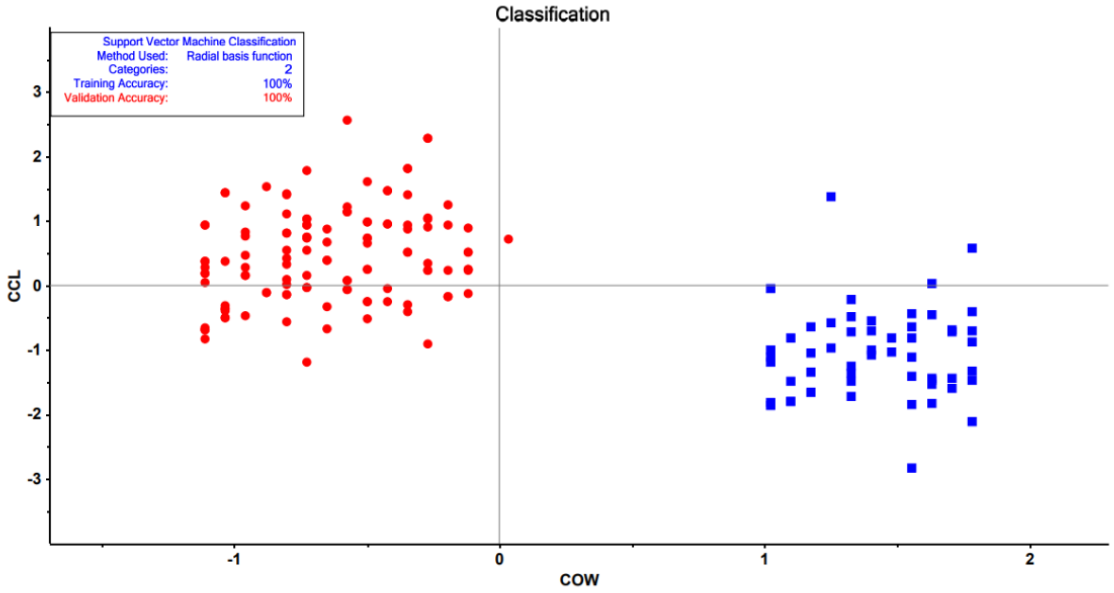
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Figure 2. Principal component (PC) score plot obtained by PC analysis conducted on the 7 neurocognitive scores and the 7 biomarkers and displaying the distribution of patients with deficit schizophrenia (red dots) versus healthy controls (blue squares).



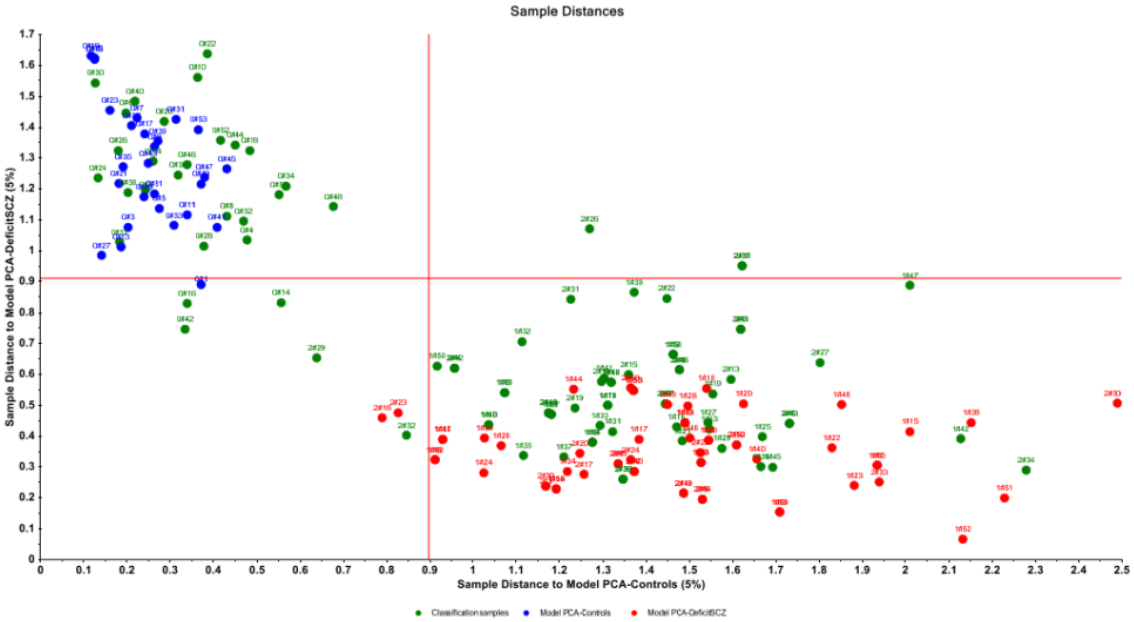
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Figure 3. Results of Support Vector Machine performed on the 7 neurocognitive scores and the 7 biomarkers. This plot was constructed using Controlled Word Association (letter fluency; COW) and CCL (CCL-11 or eotaxin) as discriminatory features.



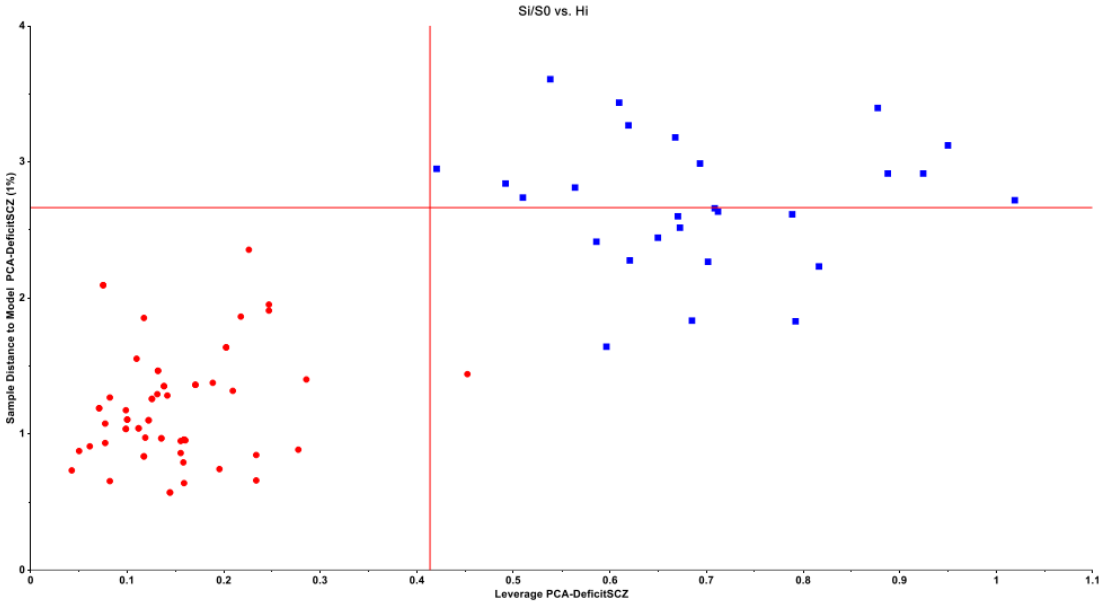
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Figure 4. Results of Soft Independent Modelling of Class Analogy (SIMCA) displaying the Coomans plot for the comparison of patients with deficit schizophrenia and controls applying a 5% confidence interval. Patients are represented by red dots and controls by blue dots (both with sample identity numbers). Green dots are the validation cases.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

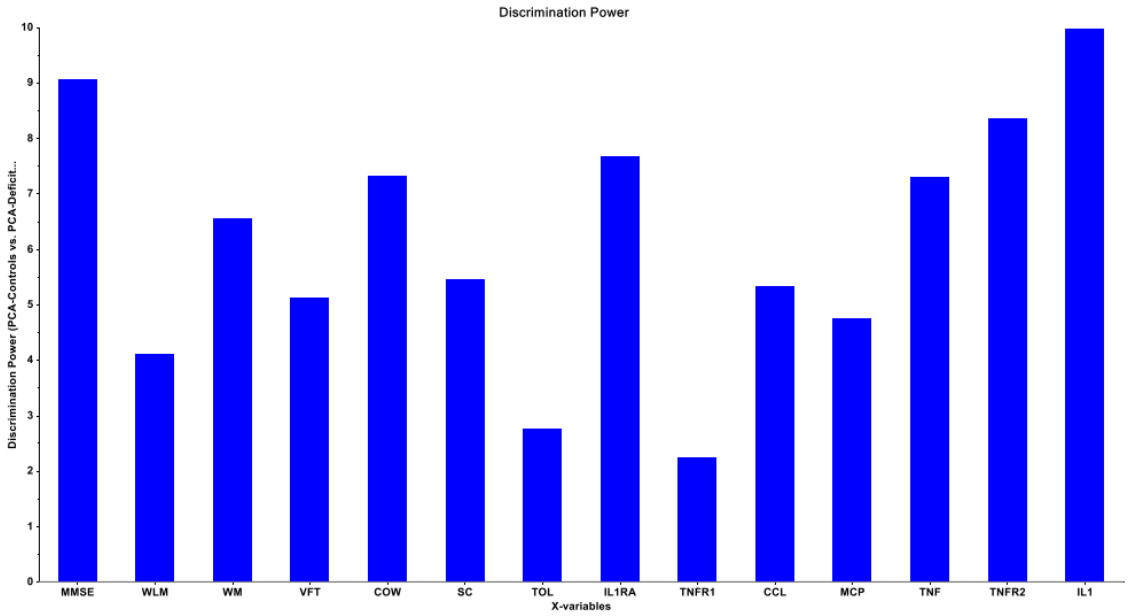
Figure 5. Results of Soft Independent Modelling of Class Analogy (SIMCA) displaying the Si/S0 deficit schizophrenia (DeficitSCZ) group-membership plot, which shows the distances of all validation cases (patients are represented by red dots and controls by blue squares) to the schizophrenia model (y-axis) and the schizophrenia model centre or leverage (x-axis). PCA-DeficitSCZ: denotes the deficit schizophrenia model constructed using principal component analysis (PCA).



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Figure 6 Results of Soft Independent Modelling of Class Analogy (SIMCA) showing the discrimination plot of the 7 biomarkers and the 7 neurocognitive test scores separating the models of controls and deficit schizophrenia.

MMSE: Mini Mental State Examination; WLM: List Learning; WM: working memory or Digit Sequencing Task; VFT: Category Instances; COW: Controlled Word Association (letter fluency); SC: Symbol Coding; TOL: Tower of London; IL1RA: soluble interleukin-1 receptor antagonist; TNFR1: soluble tumor necrosis factor-α (TNF) receptor 1; CCL: CCL-11; MCP: monocyte chemoattractant protein; IL1: IL-1β.



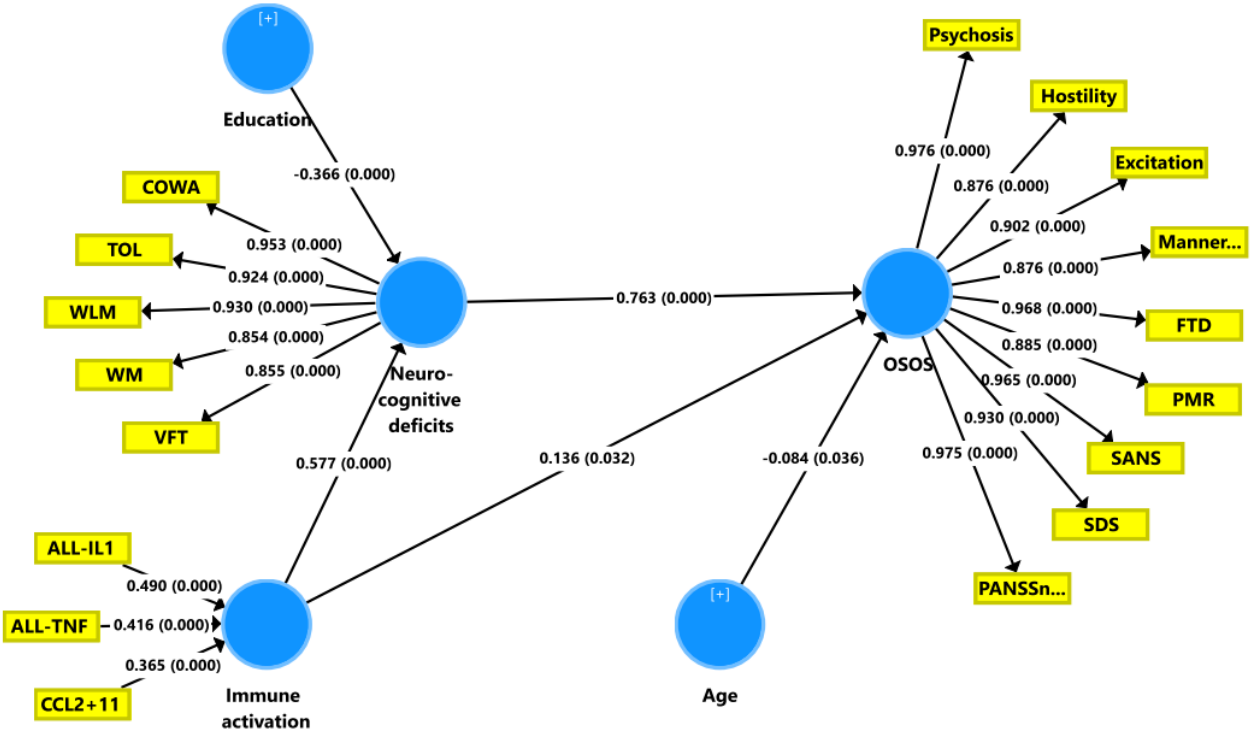
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Figure 7. Results of Partial Least Squares pathway analysis with a latent vector (LV) extracted from 9 symptom domains reflecting overall severity of schizophrenia (OSOS) as final output variable and a LV extracted from 5 cognitive scores (named neuro-cognitive deficit LV) and a LV extracted from 3 biomarker scores (ALL-IL-1, All-TNF and CCL+MCP), named immune activation) as input variables.

All IL1: computed as z score of IL1 (zIL-1; interleukin-1β) + z sIL1RA (soluble IL-1 receptor antagonist); All TNF: computed as z TNF (tumor necrosis factor-α) + z sTNFR1 (soluble TNFR1) + z sTNFR2; CCL2+11: computed as z CCL11 (exotoxin) + z MCP1 (CCL-2); COWA Controlled Word Association (letter fluency); TOL: Tower of London; WLM: List Learning; WM: working memory, Digit Sequencing Task; VFT: Category Instances

Psychosis, hostility, excitation, mannerism; FTD: formal thought disorders; PMR: psycho-motor retardation; SANS: total score on the Scale for the Assessments of Negative Symptoms SDS: total score on Schedule for the Deficit Syndrome; PANNSnegative: the negative subscale of the Positive and Negative Syndrome Scale.

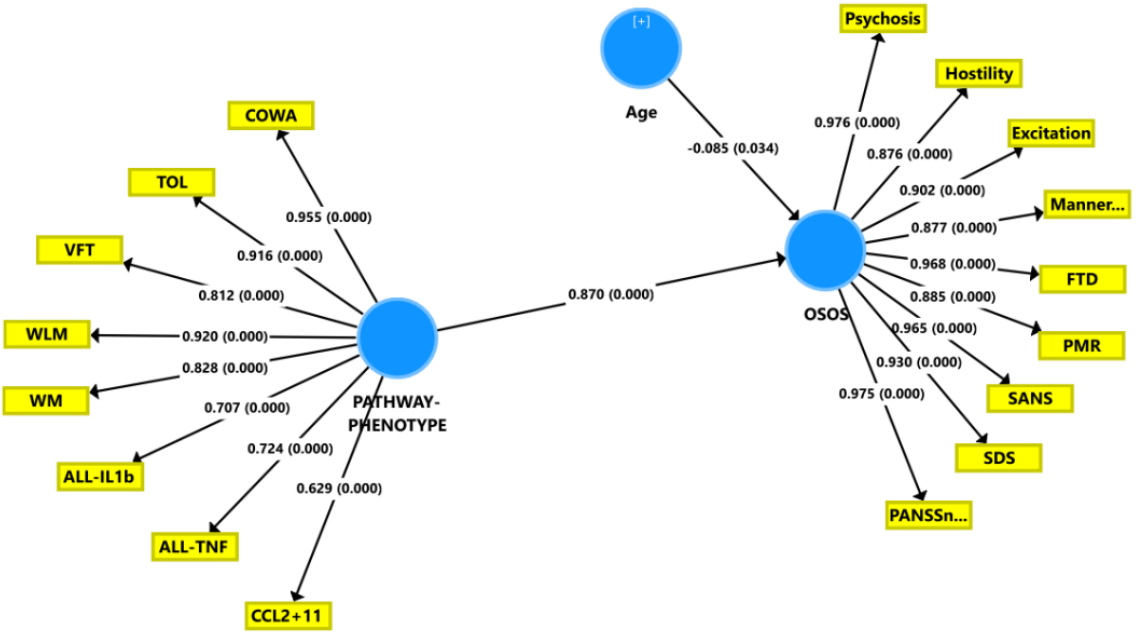
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Figure 8. Partial Least Squares path model with a latent vector extracted from the symptoms of schizophrenia reflecting overall severity of schizophrenia (OSOS) as output variable and a LV extracted from the combined biomarker composites and neurocognitive tests as the input variable, while controlling for education, age and sex.

All IL1b: computed as z score of IL1 (zIL-1; interleukin-1β) + z sIL1RA (soluble IL-1 receptor antagonist); All TNF: computed as z TNF (tumor necrosis factor-α) + z Ln sTNFR1 (soluble TNFR1) + z Ln sTNFR2; CCL2+11: computed as z CCL-11 (exotoxin) + z MCP (CCL-2); COWA Controlled Word Association; TOL: Tower of London; VFT: Category Instances; WLM: List Learning; WM: working memory, Digit Sequencing Task.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0239.v1

Figure 9. Flow chart showing the methods that should be employed to construct pathway-phenotypes using SmartPLS (Partial Least Squares) and SIMCA (Soft Independent Modelling of Class Analogy) using biomarkers and neurocognitive symptoms.