Psicothema

Psicothema 2019, Vol. 31, No. 3, 229-238
doi: 10.7334/psicothema2019.91

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Cortical surface area variations within the dorsolateral prefrontal cortex are better predictors of future cognitive performance than fluid ability and working memory

Francisco J. Román[1], Susanne M. Jaeggi[2], Kenia Martínez,[1] Jesús Privado[3], Lindsay B. Lewis[4,5], Chi-Hua Chen[2], Sergio Escorial[3], William S. Kremen[2], Sherif Karama[4,5], and Roberto Colom[1]

[1] Universidad Autónoma de Madrid, University of California (USA)[2], [3] Universidad Complutense de Madrid, [4] Montreal Neurological Institute (MNI), and [5] McGill University, Montreal (Canada)

Abstract

Background: Are cognitive and biological variables useful for predicting future behavioral outcomes? **Method:** In two independent groups, we measured a set of cognitive (fluid and crystallized intelligence, working memory, and attention control) and biological (cortical thickness and cortical surface area) variables on two occasions separated by six months, to predict behavioral outcomes of interest (performance on an adaptive version of the n-back task) measured twelve and eighteen months later. We followed three stages: discovery, validation, and generalization. In the discovery stage, cognitive/biological variables and the behavioral outcome of interest were assessed in a group of individuals (in-sample). In the validation stage, the cognitive and biological variables were related with a parallel version of the behavioral outcome assessed several months later. In the generalization stage, the validation findings were tested in an independent group of individuals (out-of-sample). **Results:** The key finding revealed that cortical surface area variations within the right dorsolateral prefrontal cortex predict the behavioral outcome of interest in both groups, whereas the cognitive variables failed to show reliable predictive validity. **Conclusions:** Individual differences in biological variables might predict future behavioral outcomes better than cognitive variables concurrently correlated with these behavioral outcomes.

Keywords: Neuroprediction, cognition, neurodiversity.

Resumen

Las variaciones de superficie cortical en la corteza dorsolateral prefrontal predicen mejor el futuro desempeño cognitivo que la inteligencia fluida y la memoria operativa. **Antecedentes:** ¿Predicen las variables cognitivas y biológicas el futuro desempeño cognitivo? **Método:** en dos grupos independientes de participantes se miden variables cognitivas (inteligencia fluida y cristalizada, memoria operativa y control atencional) y biológicas (grosor y superficie cortical) en dos ocasiones separadas por seis meses, para predecir el desempeño en la tarea n-back valorado doce y dieciocho meses después. Se completan tres etapas: descubrimiento, validación y generalización. En la de descubrimiento se valoran en un grupo de individuos las variables cognitivas/biológicas y el desempeño a predecir. En la de validación, se relacionan las mismas variables con una versión paralela de la n-back completada meses después. En la de generalización, los resultados de la validación se replican en un grupo independiente de individuos. **Resultados:** las variaciones de superficie cortical en la corteza dorsolateral prefrontal derecha predicen el desempeño cognitivo en los dos grupos independientes de individuos, mientras que las variables cognitivas no contribuyen a la predicción del desempeño futuro. **Conclusiones:** las diferencias individuales en determinadas variables biológicas predicen el desempeño cognitivo mejor que las variables cognitivas que correlacionan concurrentemente con ese desempeño.

Palabras clave: neuropredicción, cognición, neurodiversidad.

Yarkoni and Westfall (2017) have argued that turning research efforts from explanation to prediction might be of great help for increasing our understanding of human behavior. They underscored the requirement for validating a given model built for a group of individuals (in-sample) in an independent group (out-of-sample). Unfortunately, this validation strategy is rarely followed in current research (Ponsoda et al., 2017). Ideally, models built for the in-sample should be able to predict observations obtained from independent samples (out-of-sample).

In an exhaustive review, Gabrieli, Ghosh, & Whitfield-Gabrieli (2015) highlighted that brain measures (such as gray matter volumes, structural connectivity or resting state connectivity) might outperform cognitive or personality measures (such as fluid reasoning or neuroticism scores) for predicting future behavioral outcomes (such as academic achievement or response to psychological treatment). However, in their review only one out of the seventy-two studies were truly predictive.

Proper predictive studies must follow three stages: discovery, validation, and generalization. The discovery stage involves assessing relationships of cognitive and biological variables with a behavioral outcome of interest. The validation stage separates

Received: January 2, 2019 • Accepted: May 30, 2019
Corresponding author: Francisco J. Román
Facultad de Psicología
Universidad Autónoma de Madrid
28045 Madrid (Spain)
f.javier.roman@uam.es

obtained results into training and test sets, for building a candidate model from the training set, which is assessed on the test set. The observed results allow defining a prediction model. The generalization stage involves applying this latter model to a novel dataset consisting of out-of-sample individuals.

Here we will apply this general framework for studying the relationships between cognitive/biological variables and two behavioral outcomes of interest across three-time points (Time 1, Time 2, and Time 3) in two independent groups of individuals (in-sample –ISG—and out-of-sample –OSG). The measures obtained across times fit the stages proposed by Gabrieli et al. (2015): discovery (Time 1), validation (Time 2 and Time 3) and generalization (Time 2 and Time 3). Time 1 and Time 2 are separated by approx. 6 months, Time 2 and Time 3 are separated by 12 months, and, therefore, Time 1 and Time 3 are separated by 18 months.

The behavioral outcome of interest measured at Time 1 is the performance level achieved on an adaptive version of a visuospatial n-back task (ISG), whereas the behavioral outcome at Time 3 is the performance achieved on a variant of the visuospatial n-back task (faces n-back task) completed by the ISG and the OSG. The biological variables are obtained at Time 1 and Time 2, and they require computing Surface-Based Morphometry (SBM) for assessing individual differences in cortical thickness and cortical surface area in a set of regions of interest (ROIs) (Román et al., 2016). In addition, four cognitive constructs/variables are measured at Time 1 and Time 2 in both groups in order to have reliable cognitive factors related with the behavioral outcome of interest (Martínez et al., 2011; Jaeggi, Buschkuehl, Perrig, & Meier, 2010): fluid reasoning (*Gf*), crystallized ability (*Gc*), working memory capacity (*WMC*), and attention control (*ATT*).

In the discovery phase, the correlation between the cognitive/biological variables and the behavioral outcome of interest at Time 1 (visuospatial n-back) will be analyzed for the ISG. The cognitive and biological variables showing significant correlations at this stage will be considered for predicting the behavioral outcome of interest at Time 3 (faces n-back) again for the ISG. The relevance of these variables will be tested in the validation stage: the cognitive and biological measures assessed at Time 1 in the ISG will be correlated with the behavioral outcome of interest at Time 3 for testing the hypotheses generated in the discovery stage. Finally, in the generalization stage, the relationships between the cognitive/biological variables assessed at Time 1 and the behavioral outcome of interest measured at Time 3 (faces n-back) will be obtained for the OSG. The main goal of this generalization stage is to examine the robustness of the results found for the ISG in the two previous stages (discovery and validation).

Furthermore, the relationships between the cognitive/biological variables measured at Time 2 and the behavioral outcome of interest measured at Time 3 (faces n-back) will be analyzed in the ISG (validation 2.0) and in the OSG (generalization 2.0). This is intended to checking the stability and robustness of findings observed in the usual stages described above (discovery, validation, and generalization) but using the cognitive and biological variables registered at Time 2 instead of those registered at Time 1 (Figure 1).

Based on previous results, we expect significant relationships between (1) all cognitive measures and n-back performance (Martínez et al., 2011; Jaeggi et al., 2010); (2) cortical thickness/cortical surface area and the cognitive abilities measured (Román et al., 2014; Vuoksimaa et al., 2014); and (3) cortical thickness/cortical surface area and n-back performance (Barbey et al., 2014).

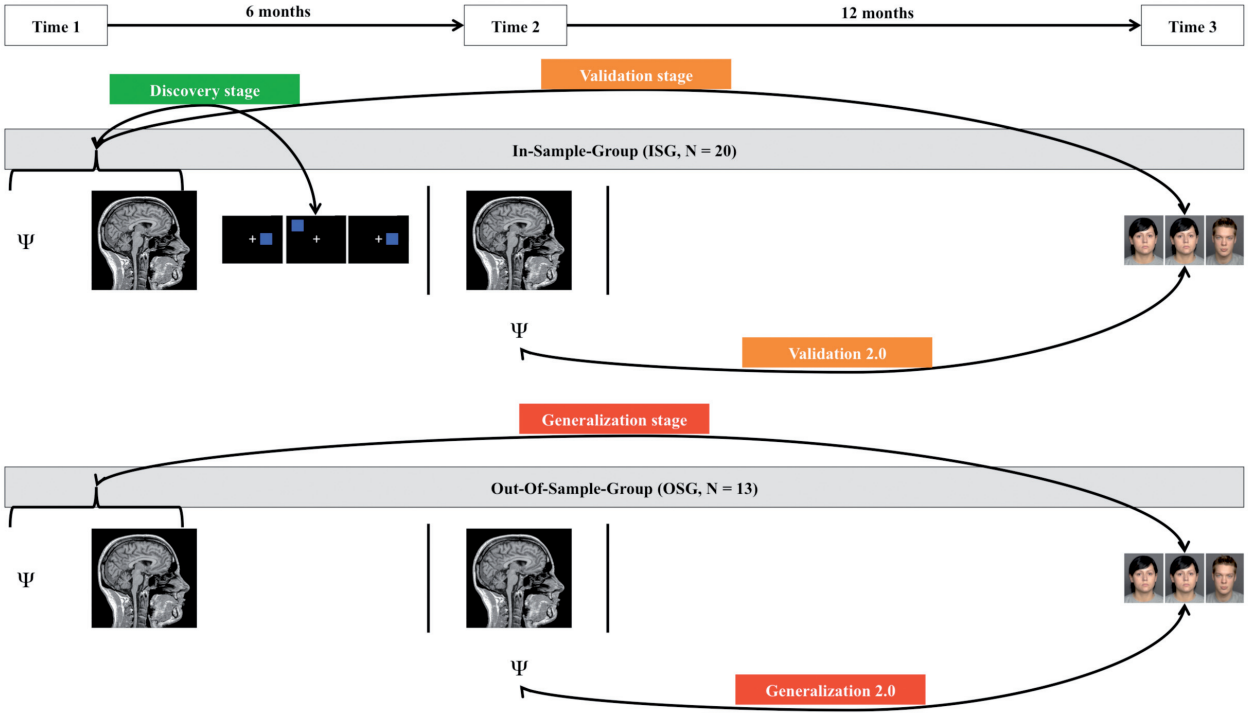


Figure 1. Stages addressed in the present study: discovery, validation (+ validation 2.0), and generalization (+ generalization 2.0). The psychological (fluid reasoning, crystallized ability, working memory, and attention control) and biological (cortical thickness and cortical surface area) variables measured at Time 1 (+ Time 2) predict the behavioral outcome of interest measured at Time 1 (visuospatial n-back) or Time 3 (faces n-back)

Method

Participants

169 psychology undergraduates completed a cognitive assessment battery tapping fluid reasoning (Gf), crystallized ability (Gc), working memory capacity (WMC), and attention control (ATT). After analyzing their performance on the fluid and crystallized tests, 56 right-handed (as evaluated by the Edinburgh Test, Oldfield, 1971) women were selected for representing a range of scores as detailed in Colom et al. (2013). Selection was done following these guidelines: (a) women only for controlling for acknowledged sex differences in brain structure (Escorial et al., 2015; Ruigrok et al., 2014), (b) right-handed, (c) age (range 18-25 years), (d) no tattoos close to the head, (e) no history of psychiatric disorders or addictions, and (f) no chronic use of drugs. Furthermore, participants were selected to represent a wide range of intelligence scores within their reference group.

Half were assigned to the experimental/training group (N = 28) and the other half was assigned to the control group (N = 28). Participants were paid for their participation. Each participant in the training group received 200€, while each control subject received 100€. They signed an informed consent following the Helsinki guidelines (World Medical Association, 2008). Both underwent MRI scanning before and after the time period devoted to the training program (12 weeks). A second psychological assessment took place at the end of the training period. Finally, participants from both groups were invited to a follow-up session devoted to the completion of a parallel version of the adaptive visuospatial n-back task that participants of the training group completed during their first training session sixteen months before. 20 individuals from the training group and 13 from the control group attended the follow-up session.

For the present research, the training participants will serve as the ISG, whereas controls will serve as the OSG. Table 1 and 2 show descriptive data for these groups. There were no significant differences between the ISG and the OSG regarding the cognitive variables measured at baseline (Time 1).

Table 1
In-sample group (ISG, N = 20). Correlations between the cognitive variables assessed at Time 1, performance in the visual n-back task also measured at Time 1, and performance on the faces n-back task completed at Time 3. Means and SDs are also shown in the last two rows. Gf = fluid intelligence, Gc = crystallized intelligence, WMC = working memory capacity, ATT = attention control

	1.	**2.**	**3.**	**4.**	**5.**	**6.**
1. Spatial N-back (time 1)	1	.61**	.46*	.65**	-.25	.61**
2. Gf (time 1)		1	.73**	.70**	.01	.45*
3. Gc (time 1)			1	.63**	-.04	.07
4. WMC (time 1)				1	-.25	.31
5. ATT (time 1)					1	.04
6. Faces N-back (time 3)						1
Mean	3.40	100.40	103.72	248.75	39.75	5.95
Standard Deviation	1.09	13.94	15.71	28.71	50.87	3.77

* $p < .05$
** $p < .01$

Table 2
Out-of-sample group (OSG, N = 13). Correlations between the cognitive variables assessed at Time 1 and performance in the faces n-back task assessed at Time 3. Means and SDs are also shown in the last two rows. Gf = fluid intelligence, Gc = crystallized intelligence, WMC = working memory capacity, ATT = attention control

	1.	**2.**	**3.**	**4.**	**5.**
1. Faces N-back (Time 3)	1	.19	.02	.20	.42
2. Gf (Time 1)		1	.83*	.65*	-.03
3. Gc (Time 1)			1	.43	-.24
4. WMC (Time 1)				1	.04
5. ATT (Time 1)					1
Mean	3.77	100.39	98.29	232.15	50.87
Standard Deviation	.59	15.69	11.64	35.89	20.78

Instruments

Psychological measures

Four cognitive factors were assessed: fluid reasoning (*Gf*), crystallized ability (*Gc*), working memory capacity (WMC), and attention control (ATT). Full details regarding the tests and tasks administered can be found in Colom et al. (2013).

Biological variables

Specific details regarding MRI acquisition and Surface-Based Morphometry computations are described in detail in Colom et al. (2016a&b) and Román et al. (2016, 2017). Here we provide a brief summary of the key steps.

MRI acquisition: all participants were MRI scanned in a General Electric Signa 3T MR Scanner (whole-body radiofrequency coil for signal excitation and quadrature 8-channel coil for reception) at Time 1 and Time 2. The specific parameters for high-resolution 3D T1-weighted images were: TE = 4.1 ms, TR = 9.1 ms, TI = 450 ms, flip angle = 10°, 170 sagittal slices, acquisition matrix = 256 mm × 256 mm, isotropic voxel size = 1 mm^3.

Surface-Based Morphometry: CIVET pipeline (version 2.0) was employed to compute the estimation of the gray matter indices studied here: Cortical Thickness (CT) and Cortical Surface Area (CSA) following the several steps (see Román et al., 2016). For the analysis, the brain was segmented in several regions of interest (ROIs) following the genetic templates provided by Chen et al., (2012, 2013). Cortical thickness and cortical surface area were computed for each ROI.

Behavioral outcomes at Time 1 and Time 3: adaptive versions of the visuospatial and faces n-back tasks

During 12 weeks, the ISG completed a cognitive training program based on the adaptive n-back task (Jaeggi et al., 2008). The adaptive cognitive training program was carried out between November 14 2011 and February 17 2012 with a break from December 24 2011 to January 9 2012). A full description of the cognitive training program can be found at Colom et al. (2013).

However, for the present research, only their performance on the first session devoted to the adaptive visuospatial n-back task was considered (Time 1). Sixteen months later, 20 participants from the ISG and 13 participants from the OSG completed the adaptive faces n-back task in the planned follow-up session (Time

3). All faces were of neutral valence (see Román et al., 2015, for further details).

Procedure

The timeline is depicted in Figure 1 and can be summarized as follows:

(1) Time 1: the first psychological assessment and the first MRI scanning session took place from September 16 to November 11 2011. The first training session took place in November 14 2011. Here participants from the ISG completed the adaptive visuospatial n-back task for the first time.
(2) Time 2: the second psychological assessment and the second MRI scanning session took place from February 20 to March 30 2012.
(3) Time 3: the follow-up session based on the adaptive faces n-back task (a parallel version of the adaptive visuospatial n-back task) took place in March 2013.

Therefore, on average there was a time gap of approx. (a) 6 months between Time 1 and Time 2, (b) 12 months between Time 2 and Time 3, and (c) 18 months between Time 1 and Time 3.

Data analysis

First, we analyzed the stability of the cognitive and biological variables computing the test-retest correlation between Time 1 and Time 2.

With respect to the predictive analyses, we followed the three stages proposed by Gabrieli et al. (2015) (Figure 1).

Discovery stage: We analyzed the relationships between the cognitive/biological variables (fluid reasoning, crystallized ability, working memory capacity, attention control, cortical thickness, and cortical surface area) assessed at Time 1 and the behavioral outcome of interest at Time 1 (adaptive visuospatial n-back) for the ISG.

Validation stage: We analyzed the relationships between the cognitive/biological variables assessed at Time 1 and the behavioral outcome of interest at Time 3 (faces n-back) for the ISG.

Generalization stage: Here we analyzed the relationships between the cognitive/biological variables assessed at Time 1 and the behavioral outcome of interest at Time 3 (faces n-back) for the OSG.

Afterwards, we moved to the validation 2.0 and generalization 2.0 stages. This next move, not considered by Gabrieli et al. (2015), was aimed at checking the stability and robustness of the findings observed in the standard stages (discovery, validation, and generalization). This is important because our small sample sizes might provide unstable results. With this goal in mind, we considered the relationships between the cognitive/biological variables (fluid reasoning, crystallized ability, working memory, attention control, cortical thickness, and cortical surface area) assessed at Time 2 (instead of those assessed at Time 1) and the behavioral outcome of interest at Time 3 (faces n-back) for the ISG and OSG.

Regarding validation 2.0, the cognitive and biological variables detected in the standard discovery stage were tested using the cognitive and biological variables measured at Time 2 against the behavioral outcome at Time 3 using data from the ISG. With respect to generalization 2.0, the cognitive and biological variables measured at Time 2 were tested against the behavioral outcome at Time 3 using data from the OSG.

Results

Stability of the cognitive and biological variables

First, we computed the test-retest correlation for each cognitive and biological variable. The average gap between Time 1 and Time 2 was approx. 6 months on average for the psychological variables (*Gf, Gc,* WMC, and ATT) and approx. 5 months for the biological variables (cortical thickness and cortical surface area).

Gf, Gc, and WMC showed high test-retest correlations: *Gf* (r = .90, p < .001), *Gc* (r = .78, p < .001), WMC (r = .69, p = .001), while the test-retest correlation for ATT was lower (r = .40, p = .079).

The test-retest correlation for the ROIs of cortical thickness and cortical surface area were substantial ranging from .55 to .88 for cortical thickness and from .91 to .99 for cortical surface area.

Discovery stage (ISG)

Table 1 shows the correlation between the cognitive variables measured at Time 1 and performance on the visuospatial n-back task (Time 1). N-back performance showed a high correlation with *Gf* (r = .61, p = .004) and *WMC* (r = .65, p = .002). *Gc* was moderately correlated with n-back performance (r = .46, p = .039). *ATT* was modestly correlated (r = -.25, p = .289). Therefore, only *Gf* and *WMC* were expected to predict performance on the faces n-back task assessed at Time 3 (validation stage).

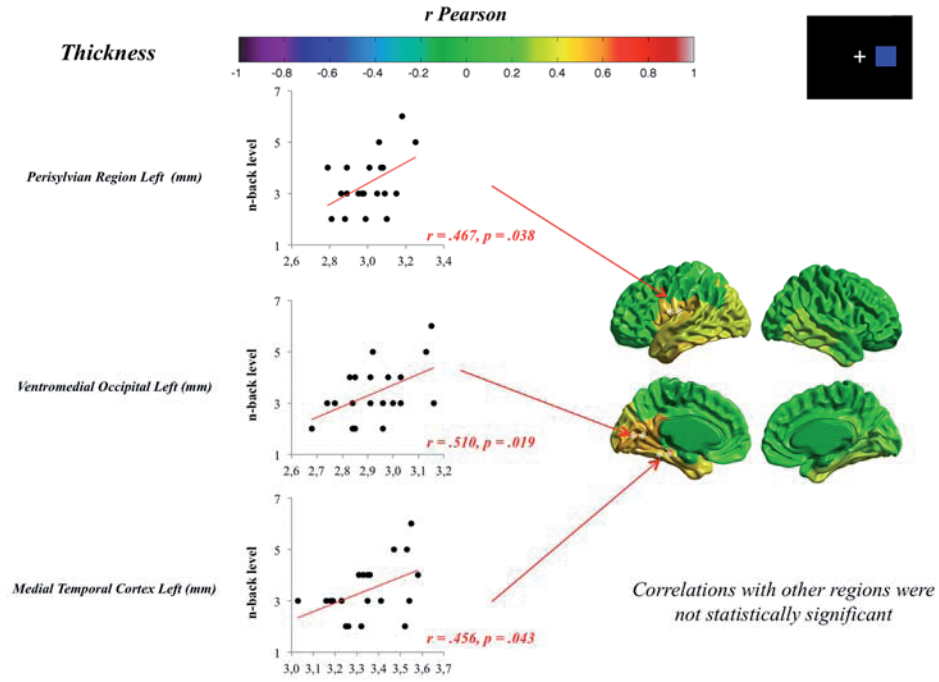
Regarding the biological variables assessed at Time 1, we computed the correlation between cortical thickness and cortical surface area estimated for each region of interest (ROIs) and visuospatial n-back performance assessed at Time 1. Figure 2A shows the significant correlations between cortical thickness and visuospatial n-back performance (Time 1): left ventromedial occipital (r = .51, p = .019), left perisylvian (r = .47, p = .038), and left medial temporal cortices (r = .46, p = .043). Figure 2B shows the significant correlations between cortical surface area and visuospatial n-back performance (Time 1): bilateral dorsolateral prefrontal (r = .63, p = .003 and r = .54, p = .013 for right and left hemisphere, respectively), bilateral superior temporal (r = .57, p = .009 and r = .78, p < .001 for right and left hemisphere, respectively), right anteromedial (r = .50, p = .024), and right posterolateral regions (r = .48, p = .031). Therefore, the identified ROIs were expected to predict performance in the faces n-back task assessed at Time 3 (validation stage).

Validation stage (ISG)

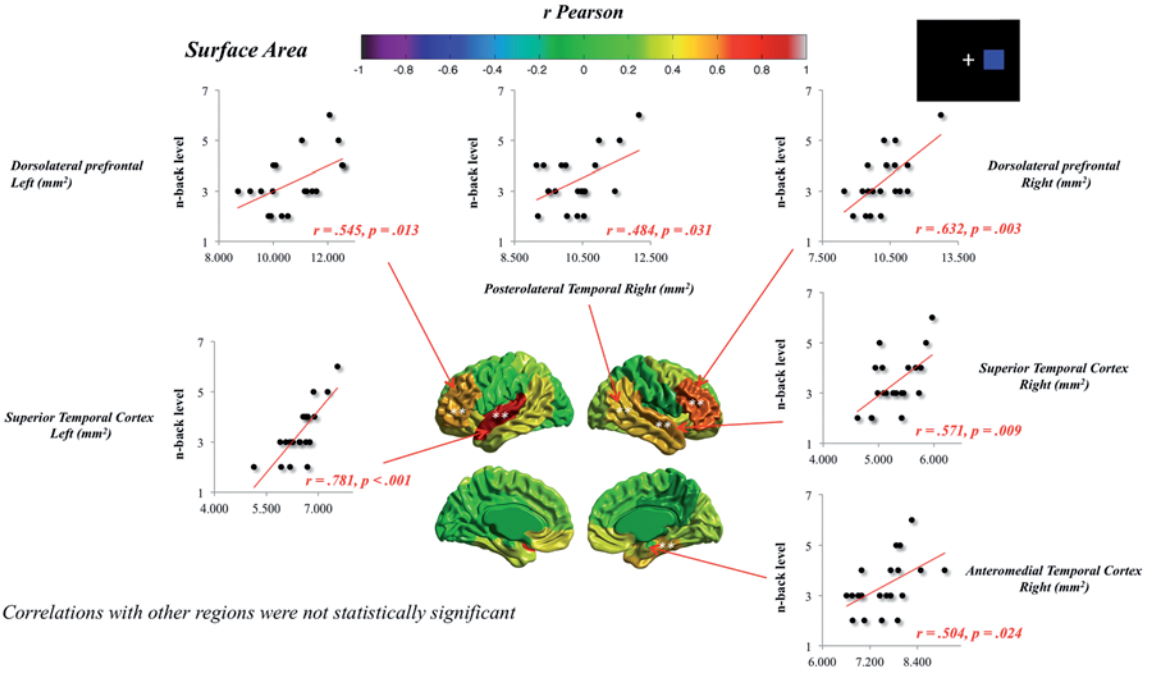
Performance in the faces n-back task completed at Time 3 was considered to validate the results observed in the discovery stage. The correlations between the cognitive variables measured at Time 1 and performance in the faces n-back task assessed at Time 3 for the ISG ranged between .04 and .45 (see Table 1). The hypothesis regarding the predictive validity of *Gf* on n-back performance was supported (r = .45, p = .044). However, the correlation was lower for *WMC* (r = .31, p = .183). Therefore, *Gf* was chosen as the candidate cognitive variable for the generalization stage using data from the OSG.

Regarding the biological variables measured at Time 1; Figure 3A and Figure 3B show the significant relationships for thickness and surface area measured at Time 1 and performance in the faces n-back task measured at Time 3. The only statistically significant correlation between thickness and faces n-back performance was found for the left perisylvian region (r = .45, p = .044). Four

(2A)



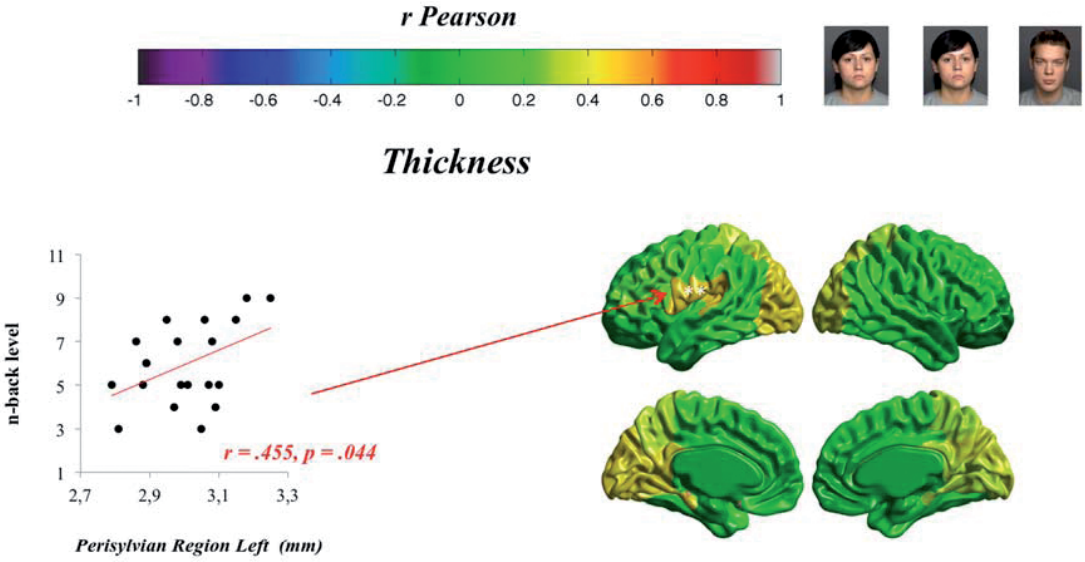
(2B)



Figures 2A & 2B. Significant correlations between gray matter indices measured at Time 1 and performance in the visual n-back task assessed at Time 1 (in-sample group, N = 20). (A) Cortical Thickness and (B) Cortical Surface Area

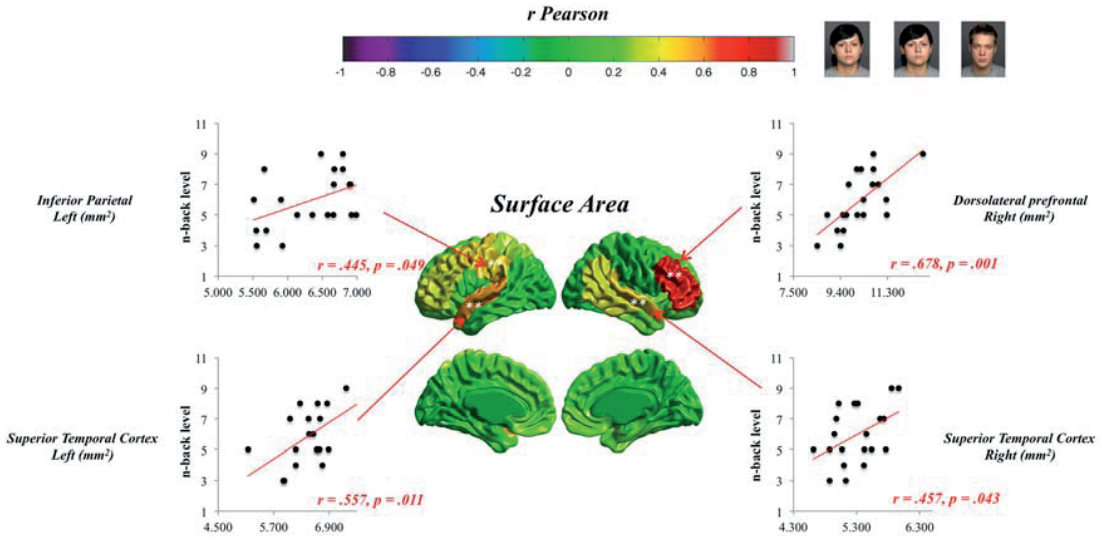
brain regions showed statistically significant correlations between surface area and faces n-back performance: left inferior parietal (r = .44, p = .049), left superior temporal (r = .56, p = .011), right superior temporal (r = .46, p = .043) and right dorsolateral prefrontal regions (r = .68, p = .001). Thus, only these regions emerged as proper predictors of n-back performance in both the discovery and validation stages and these regions were chosen as candidate relevant biological variables for the generalization stage.

(3A)



Correlations with other regions were not statistically significant

(3B)



Correlations with other regions were not statistically significant

Figures 3A & 3B. Significant correlations between gray matter indices assessed at Time 1 and performance differences in the faces n-back task assessed at Time 3 (in-sample group). (A) Cortical Thickness and (B) Cortical Surface Area

Generalization stage (OSG)

The goal here was to check if the relationships between the cognitive/biological variables measured at Time 1 and n-back performance measured at Time 3 observed for the ISG could also be identified on an independent group (OSG). Specifically, *Gf* measured at Time 1 was the only candidate cognitive variable to predict performance in the faces n-back task assessed at Time 3 for the OSG, while the left perisylvian region (thickness), and the right dorsolateral prefrontal (surface area) and bilateral superior temporal

regions (surface area) were the candidate biological variables to predict faces n-back performance at Time 3 for the OSG.

The correlation between *Gf* and faces n-back was non-significant (r = .19, p = .534). Therefore, the cognitive variable considered as relevant according to the ISG was irrelevant for the OSG. Table 2 shows the correlations between the four cognitive variables measured at Time 1 and faces n-back performance measured at Time 3.

With respect to the biological variables, the prediction for thickness was not supported, since the correlation between cortical thickness of the left perisylvian region measured at Time 1 and performance in the faces n-back task measured at Time 3 for the OSG was non-significant (r = .12, p = .694). As shown in Figure 4 (bottom panel) there were three regions where surface area measured at Time 1 was significantly correlated with faces n-back performance measured at Time 3 for the OSG: right dorsolateral prefrontal region (r = .60, p = .030), right precuneus (r = .55, p = .049), and right occipital lobe (r = .69, p = .009). However, only the right dorsolateral prefrontal region showed overlap with the results observed for the ISG.

Validation 2.0 and Generalization 2.0

Next, we analyzed the replicability of findings observed in the regular discovery, validation and generalization stages using the cognitive and biological variables registered at Time 2 for the ISG (Validation 2.0) and for the OSG (Generalization 2.0). Results showed that the cognitive variables fail to predict faces n-back performance in the OSG, as it was observed on the regular generalization stage (see Table 3).

Table 3
Correlations between the psychological factors measured at Time 2 and performance in faces n-back task completed at Time 3 (ISG, N = 20). Gf = fluid intelligence, Gc = crystallized intelligence, WMC = working memory capacity, ATT = attention control

	1.	2.	3.	4.	5.
1. Faces n-back (Time 3)	1	.41°	.02	.27	-.04
2. Gf Time 2		1	.61**	.69**	-.13
3. Gc Time 2			1	.40	-.11
4. WMC Time 2				1	.06
5. ATT Time 2					1

* p = .07
** p < .01

Regarding biological measures in Validation 2.0 stage, no statistically significant correlations were found for thickness. For surface area, two brain regions showed statistically significant correlations: left medial temporal (r = .63, p = .003) and right dorsolateral prefrontal (r =.65, p = .002). Therefore, these regions were included in the generalization stage focused on OSG (see Figure 5).

Finally, we checked if the results observed in validation 2.0 stage can be generalized to the OSG. Table 4 shows the correlations between the psychological measures completed by the OSG at Time 2 and performance in the faces n-back task (Time 3). No statistically significant correlations were found.

With respect to the biological variables, we only made predictions for cortical surface area, since results for cortical

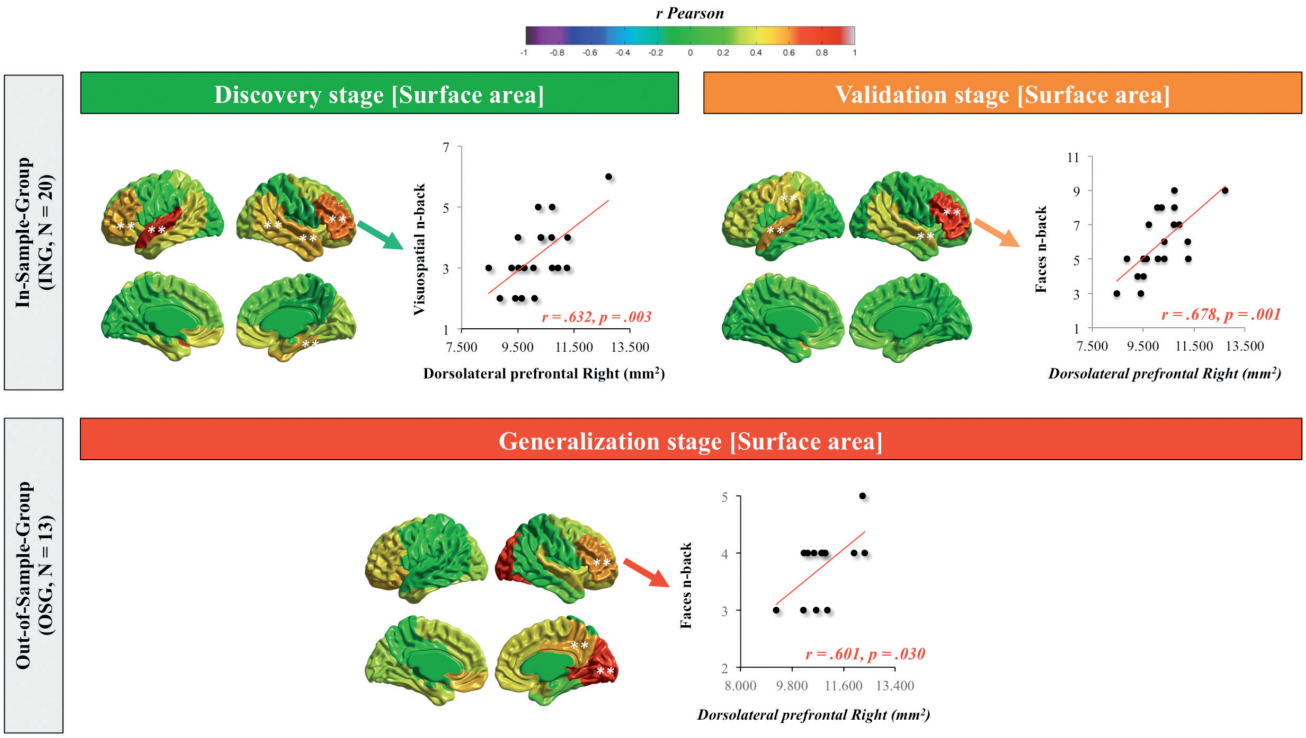


Figure 4. The left top panel shows correlations of the visuospatial n-back task (behavioral outcome measured at Time 1) with cortical surface area measured at Time 1 (discovery stage, in-sample group, ISG). The right top panel shows results of the validation stage: surface area measured at Time 1 and behavioral outcome at Time 3 (faces n-back, ISG). The bottom panel depicts the results of the generalization stage: surface area measured at Time 1 and behavioral outcome at Time 3 (faces n-back task, out-of-sample group, OSG)

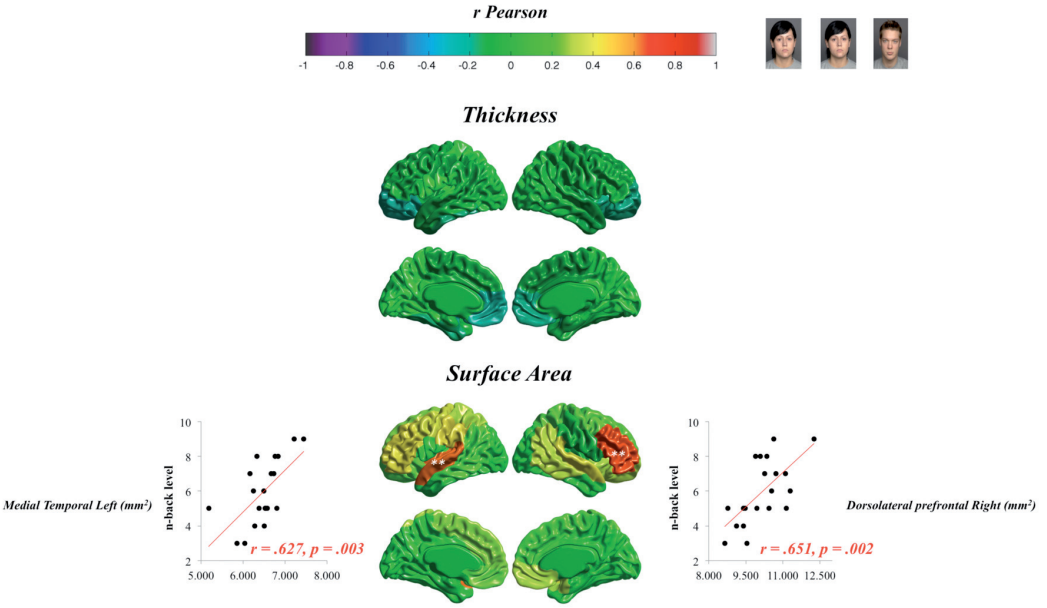


Figure 5. Correlations between thickness (top) and surface area (bottom) measured at Time 2 and performance in the faces n-back task completed at Time 3 (ISG, N = 20)

Table 4
Correlations between the psychological factors measured at Time 2 and performance in the faces n-back task completed at Time 3 (OSG, N = 13). Gf = fluid intelligence, Gc = crystallized intelligence, WMC = working memory capacity, ATT = attention control

	1.	2.	3.	4.	5.
1. Faces n-back (Time 3)	1	.02	.43	.52	-.55
2. Gf Time 2		1	.57*	.15	.08
3. Gc Time 2			1	.13	.24
4. WMC Time 2				1	-.22
5. ATT Time 2					1

* $p < .05$

thickness were not validated in the ISG. Now we test if the identified regions for surface area in the ISG are also relevant for OSG. Figure 6 (top panel) shows the candidate regions for predicting performance in the faces n-back task (left) according to standard discovery stage and the results observed in the validation 2.0 stage (right). Figure 6 (bottom panel) shows the results for the generalization 2.0 stage (OSG).

Results for the generalization 2.0 stage revealed four statistically significant correlations between cortical surface area measured at Time 2 and performance in the faces n-back task (Time 3): bilateral dorsolateral prefrontal region ($r = .63$, $p = .021$; for left and $r = .57$, $p = .042$), right precuneus ($r = .59$, $p = .034$), and right occipital lobe ($r = .64$, $p = .018$). Note that only the right dorsolateral prefrontal region was considered as candidate for

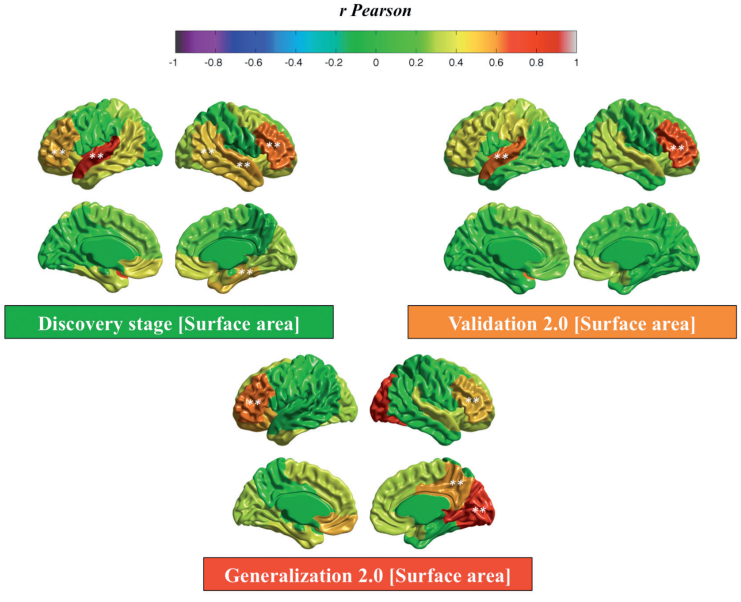


Figure 6. The left top panel shows the correlations between performance on the visual n-back task (Time 1) and cortical surface area estimated at Time 1. These regions were considered as candidates to predict performance in faces n-back task (Time 3). The right top panel shows the results of the predictions (correlations between faces n-back performance and surface area measured at Time 2). Finally, the bottom panel shows the results for OSG: correlations between cortical surface area measured at Time 2 and performance on the faces n-back task (Time 3)

predicting performance in the faces n-back task, and, therefore, only this region did fit the prediction.

Discussion

Here we have investigated if a set of cognitive and biological variables predicts two behavioral outcomes of interest. Two independent samples (ISG = in-sample group—and OSG –out-of-sample group) were considered for pursuing this goal. The cognitive and biological variables were measured in Time 1 and Time 2 (separated on average by approx. 6 months), whereas the behavioral outcomes of interest were measured in Time 1 (adaptive visuospatial n-back performance) and Time 3 (faces n-back performance) (separated on average by approx. 18 months). The cognitive variables were fluid reasoning (Gf), crystallized ability (Gc), working memory capacity (WMC), and attention control (ATT). The biological variables were cortical thickness and cortical surface area.

Three standard stages for proper predictive research were considered: discovery, validation, and generalization.

With respect to the cognitive variables, the findings observed in the discovery stage revealed statistically significant correlations between Gf/WMC and the behavioral outcome (visuospatial n-back) measured at Time 1 in the ISG. Regarding the validation stage, only Gf showed significant correlations with the behavioral outcome (faces n-back) measured at Time 3. Finally, results observed in the generalization stage considering the OSG revealed non-significant correlations between Gf assessed at Time 1 and the behavioral outcome measured at Time 3. Importantly, the same results for both groups (ISG and OSG) emerged when the cognitive variables assessed at Time 2 were considered for predicting the behavioral outcome measured at Time 3. This reinforces the stability of the results observed in the standard predictive analyses (Gabrieli et al., 2015).

Regarding the biological variables, we found significant correlations between cortical thickness measured at Time 1 and the behavioral outcome (visuospatial n-back), also assessed at Time 1, in frontal, temporal and occipital regions (discovery stage). Although the left perisylvian region predicted the behavioral outcome at Time 3 (faces n-back) in the validation stage (ISG), none region was significant in the generalization stage (OSG).

Findings for cortical surface area revealed significant correlations with the behavioral outcome measured at Time 1 (visuospatial n-back) in several frontal and temporal regions (bilateral dorsolateral prefrontal, bilateral superior temporal, right anteromedial, and right posterolateral) (discovery stage). However, only the right DLPFC and the bilateral superior temporal region emerged as significant predictors for the behavioral outcome measured at Time 3 (faces n-back) (validation stage). The single region predicting the behavioral outcome at Time 3 (faces n-back) in the OSG (generalization stage) was the right DLPFC. Therefore, individual variability in cortical surface area within the right DLPFC predicts the behavioral outcome of interest across time and regardless of the sample (Figure 4). Importantly, this key finding was replicated using the biological variables measured at Time 2 (validation 2.0 and generalization 2.0)

Why the DLPFC?

This brain region is known to support cognitive processes such as attention and working memory, as well as general cognitive ability (Barbey et al., 2014; Duncan & Owen, 2000). Also, the DLPFC is typically activated during n-back performance (Owen et al., 2005). Therefore, its salience here is far from surprising. This result was only replicated for the right DLPF. The HAROLD model (Cabeza, 2002) might explain why the results were found in the right region, but not in the left. According to this model, young people (this is the case here) are more lateralized than older people. In addition, we administered two visual working memory tasks for measuring cognitive performance. Reuter-Lorenz et al. (2000) suggested that the right hemisphere might be more involved in spatial than in verbal working memory tasks.

What may be tough to admit is that individual differences in cortical surface area measured within this brain region are better predictors of future cognitive performance (as assessed by the n-back task) than individual differences in a cognitive ability usually correlated with this performance (fluid reasoning). But there is some previous evidence that might help to frame this key finding.

Ullman, Almeida, and Klingberg (2014) have shown that biological variables such as fractional anisotropy, gray matter volume, and brain activation predict cognitive developmental differences. Specifically, investigating the relationships between cognitive variables (n-back, digit span, and fluid reasoning), biological variables and behavioral performance in a visuospatial working memory task measured two years later, the measured psychological variables failed to explain additional variance to the variance already explained by the biological variables. Furthermore, their neuroimaging model was a better predictor of developmental changes observed in the behavioral outcome of interest than their working memory and fluid reasoning measures.

Gabrieli et al. (2015) suggested that biological variables might outperform cognitive variables for predicting future behavioral outcomes. The Ullman, Almeida, and Klingberg's (2014) findings are consistent with this suggestion. Our key finding is also aligned with this perspective: cortical surface area variations within the right DLPFC outperformed fluid reasoning and working memory for *predicting* future n-back performance differences.

Choi et al. (2008) obtained high values ($R = 0.7$) when structural and functional measures were combined for predicting IQ scores. Nevertheless, cortical thickness variations in the left temporal lobe were more related with verbal intelligence, whereas bilateral functional activity differences were more related with fluid ability. These researchers concluded: "brain images can be used to predict complex traits (…) neurometric assessments of intelligence may soon become a useful complement to psychometric testing" (p. 10328).

We underscore the adequate predictive nature of the present study. As noted by Gabrieli et al. (2015) these studies are highly unusual. We must admit that the number of individuals analyzed here is low, and, therefore, the general conclusion has to be treated with caution. Further studies using the same framework are strongly required. Nonetheless, the observed key finding was stable when cognitive and biological variables measured at Time 2, instead of those measured at Time 1, were considered for predicting the future behavioral outcome of interest.

In conclusion, individual differences in biological variables such as cortical surface area might predict future behavioral outcomes better than cognitive variables concurrently correlated with these behavioral outcomes. We have shown that this is the case when one

proper predictive design is considered. Although is very difficult and unusual to obtain multiple measures of the variables if interest across time for testing the robustness of predictive models, we think it would provide invaluable information.

Acknowledgements

This project was supported by PSI2017-82218-P (Ministerio de Economía, Industria y Competitividad, Spain).

References

Barbey, A. K., Colom, R., Paul, E. J., & Grafman, J., (2014). Architecture of fluid intelligence and working memory revealed by lesion mapping. *Brain Structure & Function, 219*, 485-494. doi.org/10.1007/s00429-013-0512-z

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: The HAROLD model. *Psychology and Aging*, *17*, 85-100. doi.org/10.1037/0882-7974.17.1.85

Chen, C. H., Fiecas, M., Gutiérrez, E. D., Panizzon, M. S., Eyler, L. T., Vuoksimaa, E., ... & Neale, M. C. (2013). Genetic topography of brain morphology. *Proceedings of the National Academy of Sciences*, *110*, 17089-17094. doi.org/10.1073/pnas.1308091110

Chen, C. H., Gutiérrez, E. D., Thompson, W., Panizzon, M. S., Jernigan, T. L., Eyler, L. T., ... & Lyons, M. J. (2012). Hierarchical genetic organization of human cortical surface area. *Science*, *335*, 1634-1636. doi.org/10.1126/science

Choi, Y. Y., et al. (2008). Multiple bases of human intelligence revealed by cortical thickness and neural activation. *The Journal of Neuroscience, 28*(41), 10323-10329. doi.org/10.1523/JNEUROSCI.3259-08.2008

Colom, R., Martínez, K., Burgaleta, M., Román, F. J., García-García, D., Gunter, J. L., ... & Thompson, P. M. (2016a). Gray matter volumetric changes with a challenging adaptive cognitive training program based on the dual n-back task. *Personality and Individual Differences, 98*, 127-132. doi.org/10.1016/j.paid.2016.03.087

Colom, R., Román, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., ... & Karama, S. (2013). Adaptive n-back training does not improve fluid intelligence at the construct level: Gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence, 41*, 712-727. doi.org/10.1016/j.intell.2013.09.002

Colom, R., Hua, X., Martínez, K., Burgaleta, M., Román, F. J., Gunter, J. L., Carmona, S., Jaeggi, S. M., & Thompson, P. M. (2016b). Brain structural changes following adaptive cognitive training assessed by Tensor-Based Morphometry. *Neuropsychologia, 91*, 77-85. doi.org/10.1016/j.neuropsychologia.2016.07.034

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23,* 475-483.

Escorial, S., Román, F. J., Martínez, K., Burgaleta, M., Karama, S., & Colom, R. (2015). Sex differences in neocortical structure and cognitive performance: A surface-based morphometry study. *Neuroimage, 104,* 355-365. doi.org/10.1016/j.neuroimage.2014.09.035

Gabrieli, J. D., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron, 85*, 11-26. doi.org/10.1016/j.neuron.2014.10.047

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*, 6829-6833. doi.org/10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory, 18*, 394-412. doi.org/10.1080/09658211003702171

Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, M. Á., & Colom, R. (2011). Can fluid intelligence be reduced to 'simple' short-term storage? *Intelligence, 39,* 473-480. doi.org/10.1016/j.intell.2011.09.001

Ponsoda, V., Martínez, K., Pineda-Pardo, J. A., Abad, F. J., Olea, J., … & Colom, R., (2017). Structural brain connectivity and cognitive ability differences: A multivariate distance matrix regression analysis. *Human Brain Mapping*, *38*, 803-816. doi.org/10.1002/hbm.23419

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97-113. doi.org/10.1016/0028-3932(71)90067-4

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*, 46-59. doi.org/10.1002/hbm.20131

Reuter-Lorenz, P. A., Jonides, J., Smith, E. E., Hartley, A., Miller, A., Marshuetz, C., & Koeppe, R. A. (2000). Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET. *Journal of Cognitive Neuroscience*, *12*, 174-187. doi.org/10.1162/089892900561814

Román, F. J., Abad, F. J., Escorial, S., Burgaleta, M., Martínez, K., Álvarez-Linera, J., ... & Colom, R. (2014). Reversed hierarchy in the brain for general and specific cognitive abilities: A morphometric analysis. *Human Brain Mapping, 35,* 3805-3818. doi.org/10.1002/hbm.22438

Román, F. J., García-Rubio, M. J., Privado, J., Kessel, D., López-Martín, S., Martínez, K., ... & Colom, R. (2015). Adaptive working memory training reveals a negligible effect of emotional stimuli over cognitive processing. *Personality and Individual Differences, 74*, 165-170. doi.org/10.1016/j.paid.2014.10.014.doi.org/10.1016/j.paid.2014.10.014

Román, F. J., Iturria-Medina, Y., Martínez, K., Karama, S., Burgaleta, M., Evans, A. C., ... & Colom, R. (2017). Enhanced structural connectivity within a brain sub-network supporting working memory and engagement processes after cognitive training. *Neurobiology of Learning and Memory*, *141*, 33-43. doi.org/10.1016/j.nlm.2017.03.010

Román, F. J., Lewis, L. B., Chen, C. H., Karama, S., Burgaleta, M., Martínez, K., ... & Colom, R. (2016). Gray matter responsiveness to adaptive working memory training: A surface-based morphometry study. *Brain Structure and Function*, *221*, 4369-4382. doi.org/10.1007/s00429-015-1168-7

Ruigrok, A. N., Salimi-Khorshidi, G., Lai, M. C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience & Biobehavioral Reviews*, *39*, 34-50. doi.org/10.1016/j.neubiorev.2013.12.004

Ullman, H., Almeida, R., & Klingberg, T. (2014). Structural maturation and brain activity predict future working memory capacity during childhood development. *The Journal of Neuroscience*, *34*, 1592-1598. doi.org/10.1523/JNEUROSCI.0842-13.2014

Vuoksimaa, E., Panizzon, M. S., Chen, C. H., Fiecas, M., Eyler, L. T., Fennema-Notestine, C., ... & Lyons, M. J. (2014). The genetic association between neocortical volume and general cognitive ability is driven by global surface area rather than thickness. *Cerebral Cortex, bhu018*. doi.org/10.1093/cercor/bhu018

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*, 1100-1122. doi.org/10.1177/1745691617693393

World Medical Association (2008). Declaration of Helsinki - Ethical principles for medical research involving human subjects. *59th WMA General Assembly, Seoul, Korea*.