STRUCTURAL BRAIN ATROPHY PREDICT SYMPTOM SEVERITY IN SCHIZOPHRENIA BASED ON GENERALIZED ADDITIVE MODELS

Meng Wang[1,2], Lingzhong Fan[1,2], Bing Liu[3,4]*

[1]Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China.
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China.
[3]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China.
[4]Chinese Institute for Brain Research, Beijing, China.

ABSTRACT

Schizophrenia (SCZ) patients typically vary significantly in symptom severity. Despite numerous studies demonstrate SCZ is linked to brain structure abnormalities, relationships are obscure. In this paper, we establish relationships between structural abnormalities and symptom severity. All analyses are performed in two datasets (discovery: 326 SCZ and 298 normal control (NC); replication: 216 SCZ and 173 NC). We first build normative models in NC group, based on which we calculate atrophy values of cortical thickness, surface area, and gray matter volume in SCZ. Finally, we use atrophy values to predict symptom severity via generalized additive models and further evaluate the marginal effect of each structural feature. We found atrophy values could reliably predict symptom severity across two datasets (discovery: Pearson r = 0.29, $P < 1 \times 10^{-5}$; replication: r = 0.26, $P = 3 \times 10^{-5}$). Our findings could aid in understanding the pathogenesis of symptoms in SCZ.

***Index Terms*—** Schizophrenia, Prediction, Generalized Additive Models, Symptom Severity, Brain Atrophy

1. INTRODUCTION

Schizophrenia (SCZ) is a highly disabling mental disorder along with heterogeneous clinical manifestations. For example, patients with SCZ typically have different levels of symptom severity. Recently, structural magnetic resonance imaging (MRI) has been broadly adopted to examine brain anatomy in SCZ patients. Among various measures of structural MRI, cortical thickness, surface area, and gray matter volume (GMV) are widely used. A large-scale neuroimaging meta-analysis study reported that patients with SCZ have extensive thinner cortex and smaller surface area in both hemispheres, primarily in frontal and temporal lobes regions[1]. Compared to normal controls, patients with SCZ have global GMV reductions, particularly in the subcortical thalamus and across cortical regions, including the ventral and medial prefrontal cortex[2,3]. Although numerous structural MRI studies have concluded that SCZ is associated with widespread abnormalities in brain structure[4], the neurobiological mechanisms inducing these structural alternations are unclear. Furthermore, the relationships between brain anatomical abnormalities and clinical phenotypes such as symptom severity are obscure.

In most existing studies, Pearson correlation is generally used to investigate relationships between structural features and symptom severity. Specifically, two steps are included. First, identify abnormal brain regions in SCZ patients through case-control comparisons on structural features such as cortical thickness, surface area, and GMV. Second, calculate correlation coefficients between mean feature values in each of the detected regions and clinical symptoms. Studies have found that cortical thickness and GMV in frontal and temporal regions are inversely correlated with the positive symptom, whereas surface area in the right frontal regions is inversely correlated with the negative symptom[5,6].

Despite certain regional correlations being found, there are several concerns about the two-step method. First, only statistically significant brain regions from the first step are investigated. Since the significance level is highly dependent on sample size, effects of other potential brain regions are liable to be ignored. Second, the combined effects of structural features are not investigated. The overall contributions are neglected if examine brain regions individually, given these brain regions could be interrelated. Besides, different structural features are not taken together to evaluate integrative effects. Third, Pearson correlation can not identify nonlinear relationships, since it is a measure of linear correlation.

In our study, we worked on the three concerns proposed. Considering anatomical features of cortical thickness, surface area, and GMV mainly show reductions in SCZ patients, we focused on brain regions with

reductions of these features. We first built normative models in the NC group via general linear models (GLMs). Based on normative models, we then defined atrophy values of each structural feature in the SCZ group, indicating global mean feature reductions. We used atrophy values to predict symptom severity through generalized additive models (GAMs). GAM is an additive modeling technique, having advantages of flexibility, interpretability, and regularization[7,8]. Particularly, the impact of features is captured through smooth functions, and thus nonlinear effects are contained. From the view of performance, GAMs are competitive with other popular learning methods. Finally, to achieve interpretability, we evaluated the marginal effect of each feature in predictions.

2. MATERIAL AND METHOD

2.1. Participants

The discovery dataset (n = 624, SCZ = 326, and NC = 298) was collected from four Chinese hospitals: Peking University Sixth Hospital, Beijing Huilongguan Hospital, Xijing Hospital, and Henan Mental Hospital. All patients with SCZ were diagnosed consensually by two qualified psychiatrists utilizing the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I, Patient Edition). None of the included patients had a history of any other psychiatric disorders, neurological disorders, cognitive disabilities, serious physical diseases, severe head trauma, substance abuse or dependence, and electroconvulsive therapy within the last 6 months. All the engaged healthy subjects were clinically investigated with SCID-I Non-patient Edition, who and their first- and second-degree relatives had no history of any mental disorders. Written informed consent was obtained from all participants or their guardians. The research protocol was approved by the Medical Research Ethics Committees of the local hospitals. SCZ and NC groups were statistically matched in age (two-sample t-test: t = -0.01, P = .98) and gender (chi-squared test: χ^2 = 0.08, P = .77). See Table 1 for the discovery dataset details.

Table 1. Demographics and clinical assessments in the discovery and replication datasets.

Dataset	Category	SCZ	NC	*P*
Discovery	Number	326	298	NA
	Age (y)	27.0 ± 6.1	27.1 ± 5.8	0.98
	Gender (M/F)	168 / 158	158 / 140	0.77
	PANSS Total	81.6 ± 11.1	NA	NA
Replication	Number	216	173	NA
	Age (y)	28.2 ± 7.4	27.7 ± 6.3	0.46
	Gender (M/F)	114 / 102	88 / 85	0.78
	PANSS Total	87.6 ± 12.1	NA	NA

PANSS, Positive and Negative Syndrome Scale; NA, not applicable; y, year; M, male; F, female. Data were presented as mean ± standard deviation.

The replication dataset (n = 389, SCZ = 216, and NC = 173) was recruited from three hospitals in China: Renmin Hospital of Wuhan University, Zhumadian Psychiatric Hospital, and Henan Mental Hospital. The clinical inclusion criteria and evaluation were the same as the discovery dataset. All participants or their guardians provided written informed consent. The study was approved by the Medical Research Ethics Committees of the local hospitals. There were no statistically significant differences in age (two-sample t-test: t = 0.72, P = .46) and gender (chi-squared test: χ^2 = 0.07, P =.78) between SCZ and NC groups. See Table 1 for the replication dataset details.

The symptoms of the patients with SCZ were measured using the Positive and Negative Syndrome Scale (PANSS)[9]. Specifically, all the included patients had a PANSS total score greater than sixty. See Table 1 for the PANSS score details.

2.2. MRI acquisition and processing

In the discovery dataset, T1-weighted (T1w) images were acquired on 3.0T Siemens scanners. The 3D-MPRAGE sequence was performed with parameters: repetition time (TR) = 2530 ms; echo time (TE) = 3.5 or 2.43 ms; flip angle (FA) = 7^o; inversion time (TI) = 1100 ms; voxel size = 1 × 1 × 1 mm^3; matrix size = 256 × 256 × 192. In the replication dataset, T1w images were obtained on 3.0T GE scanners. The 3D-GRE sequence was performed with parameters: TR = 7.8 or 6.8 ms; TE = 3 or 2.5 ms; FA = 7^o; TI = 1100 ms; voxel size = 1 × 1 × 1 mm^3; matrix size = 256 × 256 × 188.

Voxel-based morphometry (VBM) processing was carried out using the VBM8 toolbox[10]. Specifically, it took native T1w images as inputs and generated differently segmented tissue compartments, namely gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) images. non-brain tissues were removed in the procedure. The outputs were registered to standard Montreal Neurological Institute (MNI) space with non-linear deformation using the high dimensional DARTEL algorithm[11]. Quality check was implemented by displaying slices for images and checking for sample homogeneity. All the segmented images were not smoothed and had a voxel size of 1.5mm^3 and a resolution of 121 × 145 × 121.

Surface-based morphometry (SBM) analysis was performed using Freesurfer version 6.0[12]. The cortical reconstruction processes from T1w images were implemented using a single recon-all command, consisting of several stages including skull stripping, intensity normalization, white matter segmentation, and surface extraction. Cortical thickness was measured by the closest distance between white matter and pial surfaces[13]. The surface area was achieved from triangular surface faces and further converted to vertice representations. For accurate matching of cortical locations among subjects, each individual's reconstructed cortex was finally resampled to an averaged space (fsaverage)[14], containing 163,842

nodes in each hemisphere. Quality control was performed by visual examination for errors of segmentation, intensity normalization, and skull stripping.

2.3. Creation of atrophy features

Three types of brain measurements were used to create atrophy features, i.e. vertex-based cortical thickness and surface area, as well as voxel-based subcortical gray matter volume (GMV). Particularly, GMV in the cortex was not used, considering its interrelationships with another two measurements possibly led to feature redundancy. The subcortex mask applied in GM images was generated based on the Brainnetome atlas[15] by simply zeroing the whole cortex while setting all the subcortical regions to one. Particularly, the cortex-subcortex boundary is consistent with the Freesurfer Desikan–Killiany (DK) atlas[16].

We first built normative models[17,18] from the NC group in each dataset, given different MRI scanner protocols and site numbers. Specifically, we separately performed vertex-wise (or voxel-wise) general linear models (GLMs) for cortical thickness, surface area, and subcortical GMV. Age, gender, and dummied site variables were regarded as covariates (see equation 1).

$$Y = \beta_0+\beta_1\times X_{age}+\beta_2\times X_{gender}+\beta_3\times X_{site\_dummy}+\varepsilon \quad (1)$$

Y represented the three measurements respectively, X were covariates, ε were residual maps. β maps were computed via the least-squares method.

In each dataset, we calculated three different atrophy maps for each patient with SCZ based on the normative models derived from the NC group, following equation 2.

$$W_{amap} = (Y'-\beta_0-\beta_1\times X'_{age}-\beta_2\times X'_{gender}-\beta_3\times X'_{site\_dummy}) / \varepsilon_{std} \quad (2)$$

Y' represented the three measurements in the SCZ group respectively, X' was the corresponding covariates. β_0 to β_3 and ε_{std} values were the results from the normative models, where ε_{std} was the standard deviation of residual maps. W_{amap} was further binarized at a level of -2, indicating 2 standard deviations below the mean of the NC group, controlling for age, gender, and site. We defined the atrophy value as the mean of binarized W_{amap}. Finally, for each patient with SCZ, three atrophy values were computed, representing atrophy degree relative to the NC group in the aspects of cortical thickness, surface area, and subcortical GMV respectively.

2.4. Prediction of symptom severity

Regression analyses were carried out to investigate whether atrophy values were associated with clinical symptoms in SCZ. Specifically, we built regression models to predict PANSS total scores from the three atrophy values.

We built predictive models under the framework of Generalized Additive Models (GAMs)[7,8], which is a powerful yet simple technique combining several advantages, e.g. flexibility, interpretability, and regularization. In our analysis, linear GAMs were selected for simplicity and allowing nonlinear functions for predictor variables (feature functions) while maintaining additivity. Particularly, the feature functions can capture complex relationships but linear between predictor variables (atrophy values) and the response variable (PANSS total scores). We used penalized B splines[19] as feature functions, which have a penalty parameter to control smoothing to prevent overfitting. Besides, we also added a tensor term to create interactions between the three atrophy features.

We trained models in the discovery dataset via nested 5-fold cross-validation (CV)[20], an effective design to avoid bias in model selection[21]. The nested 5-fold CV structure contains inner and outer procedures, which are used for parameters tuning (i.e. penalty strength) and testing performance. Altogether, we had 6 independent penalty parameters to optimize, each of which had a wide range of possible values. Considering the high-dimensional search spaces, we applied a randomized search strategy[22] (more than 100,000 times) to estimate these parameters. To further evaluate the generalization performance, we choose the predictive model with the best testing performance and validated it in the replication dataset after retraining on the whole discovery dataset. We calculated the Pearson correlation coefficient and mean absolute error (MAE) between observed and predicted scores to quantify the prediction performance. To achieve the significance (*P* value) of predictions, we performed 100,000 times permutation test by randomly shuffling the response values.

To evaluate the dependencies between the three features (atrophy values) and the target (PANSS total scores), we applied partial dependence plots (PDP) to characterize the marginal effect of atrophy values on the PANSS total scores. Particularly, PDP describes whether the relationship between a feature and the target is linear, nonlinear and monotonic, or more complicated.

3. EXPERIMENT AND RESULT

3.1. Multicollinearity of predictor variables

Multicollinearity is a common problem in regression models when two or more predictor variables are correlated. It has been demonstrated that multicollinearity has an enormous influence on regression analysis since it probably leads to biased results and has unpredictable and inconsistent effects on parameter estimates and significance[23]. A generally accepted method to evaluate the degree of multicollinearity is by examining the variance inflation factor (VIF)[24] for each predictor variable. Specifically, VIF quantifies how much of the variance in the dependent variable explained by each of the predictor variables is inflated. Therefore, before

performing regression analysis, we first detected whether high levels of multicollinearity exist in our predictor variables, i.e. the three types of atrophy values. The VIF values are given in Table 2.

Table 2. VIF values of atrophy variables in the discovery and replication datasets.

Dataset	Cortical thickness	Surface area	Subcortical GMV
Discovery	1.76	1.34	1.70
Replication	1.72	1.55	1.51

As the interpretation by previous studies[23–25], a VIF value of 1 represents there are no correlations between the predictor variable and any others. When the VIF value increases and falls in the range between 1 and 5, moderate correlations exist. If the VIF value is greater than 5, the predictor variable is highly correlated with some others. All our VIF values are slightly greater than 1 and less than 2, indicating that there are minor correlations between atrophy variables, and thus multicollinearity is not a matter of concern.

We built linear GAMs to predict PANSS total scores from atrophy values of cortical thickness, surface area, and subcortical GMV in patients with SCZ. Model training and testing were implemented via nested 5-fold CV in the discovery dataset. The penalty parameters were randomly searched more than 100,000 times from 50 numbers evenly on a log scale from -4 to 2 with a base of 10. To lower the bias introduced by a single run of nested 5-fold CV[26], we repeated this procedure 100 times and calculated median testing performance in terms of Pearson's r and MAE (see Table 3).

Table 3. Testing results of nested 5-fold CV with median performance in the discovery dataset.

	The outer fold of nested 5-fold CV					
	1	2	3	4	5	Mean
Pearson r	0.39	0.25	0.17	0.37	0.28	0.29
MAE	8.26	6.57	8.64	9.17	8.61	8.25

To further evaluate generalization performance, we chose the model with the best testing performance, i.e. the model of outer fold 1 in Table 3. After retraining on the whole discovery dataset, we validated it in the replication dataset. See Fig.1 for model performance in the two datasets.

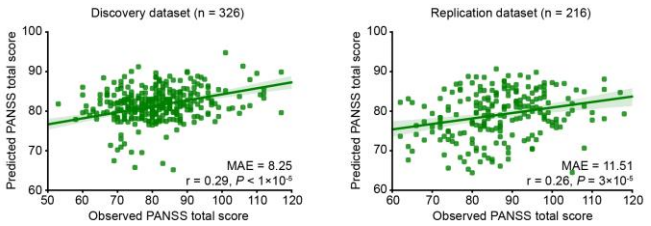

Fig. 1. Atrophy values predicted PANSS total scores. Scatter plots with regression lines show the median testing performance in the discovery dataset (left) and the generalization performance in the replication dataset (right). r was the Pearson correlation between the observed score and predicted score. *P* value was the result of the 100,000 times permutation test. MAE was the mean absolute error.

In brief, we found atrophy values of cortical thickness, surface area, and subcortical GMV could significantly predict PANSS total scores across two independent datasets (discovery dataset: MAE = 8.25, r = 0.29, $P < 1 \times 10^{-5}$; replication dataset: MAE = 11.51, r = 0.26, $P = 3 \times 10^{-5}$).

The partial dependence plot (PDP) was plotted to show the marginal effect of cortical thickness, surface area, and subcortical GMV have on the PANSS total score in predictions. See Fig.2 for the PDP of each feature.

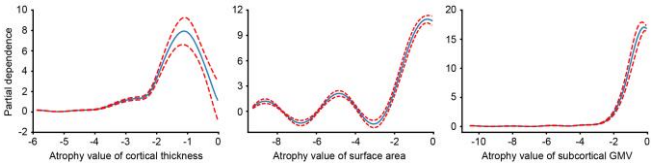

Fig. 2. Partial dependence plot of cortical thickness (left), surface area (middle), subcortical GMV (right) in predictions. The two red lines were 95% confidence intervals.

All of the three features had nonlinear and nonmonotonic effects on PANSS total score, particularly, the effect of surface area was more complex. However, in specific ranges, these effects could be monotonic or nearly linear. For example, the dependence of cortical thickness from -6 to -1 increased monotonically, where the effect was almost linearly increased from -2 to -1, then the effect was decreased linearly from -1 to 0.

4. CONCLUSION

In this paper, we established relationships between structural brain abnormalities of cortical thickness, surface area, and subcortical GMV and symptom severity in patients with SCZ. Instead of performing between-group comparisons, we defined atrophy values for each of the three features to quantify structural alternations in the SCZ group. The atrophy values indicated global mean feature reductions and were calculated based on normative models, derived from the NC group through general linear models. Finally, we used the three atrophy values to predict symptom severity via generalized additive models. Results showed atrophy values could significantly predict PANSS total scores in SCZ patients across two independent datasets. Particularly, the marginal effect each feature had on PANSS total score was nonlinear and nonmonotonic. Our findings could aid in understanding the pathogenesis of symptoms in SCZ.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of each involved hospital.

6. ACKNOWLEDGMENTS

This work was supported by the National Key Basic Research and Development Program (973) (Grant 2011CB707800) and the Natural Science Foundation of China (Grant numbers 81771451).

7. REFERENCES

[1] T. G. M. van Erp *et al.*, "Cortical Brain Abnormalities in 4474 Individuals With Schizophrenia and 5098 Control Subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium," *Biol. Psychiatry*, vol. 84, no. 9, pp. 644–654, Nov. 2018.

[2] N. G. Cascella, S. C. Fieldstone, V. A. Rao, G. D. Pearlson, A. Sawa, and D. J. Schretlen, "Gray-matter abnormalities in deficit schizophrenia," *Schizophr. Res.*, vol. 120, no. 1, pp. 63–70, Jul. 2010.

[3] H. Ananth, I. Popescu, H. D. Critchley, C. D. Good, R. S. J. Frackowiak, and R. J. Dolan, "Cortical and subcortical gray matter abnormalities in schizophrenia determined through structural magnetic resonance imaging with optimized volumetric voxel-based morphometry," *Am. J. Psychiatry*, vol. 159, no. 9, pp. 1497–1505, Sep. 2002.

[4] M. E. Shenton, T. J. Whitford, and M. Kubicki, "Structural neuroimaging in schizophrenia from methods to insights to treatments," *Dialogues Clin. Neurosci.*, vol. 12, no. 3, pp. 317–332, Sep. 2010.

[5] J. L. Padmanabhan *et al.*, "Correlations Between Brain Structure and Symptom Dimensions of Psychosis in Schizophrenia, Schizoaffective, and Psychotic Bipolar I Disorders," *Schizophr. Bull.*, vol. 41, no. 1, pp. 154–162, Jan. 2015.

[6] Y. Xiao *et al.*, "Altered Cortical Thickness Related to Clinical Severity But Not the Untreated Disease Duration in Schizophrenia," *Schizophr. Bull.*, vol. 41, no. 1, pp. 201–210, Jan. 2015.

[7] T. Hastie and R. Tibshirani, "Generalized additive models for medical research," *Stat. Methods Med. Res.*, vol. 4, no. 3, pp. 187–196, Sep. 1995.

[8] T. Hastie and R. Tibshirani, "Generalized Additive Models: Some Applications," *J. Am. Stat. Assoc.*, vol. 82, no. 398, pp. 371–386, Jun. 1987.

[9] S. R. Kay, A. Fiszbein, and L. A. Opler, "The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia," *Schizophr. Bull.*, vol. 13, no. 2, pp. 261–276, Jan. 1987.

[10] H. Matsuda *et al.*, "Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated lie algebra improves the diagnosis of probable Alzheimer Disease," *AJNR Am. J. Neuroradiol.*, vol. 33, no. 6, pp. 1109–1114, Jun. 2012.

[11] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, no. 1, pp. 95–113, Oct. 2007.

[12] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis. I. Segmentation and surface reconstruction," *NeuroImage*, vol. 9, no. 2, pp. 179–194, Feb. 1999.

[13] B. Fischl and A. M. Dale, "Measuring the thickness of the human cerebral cortex from magnetic resonance images," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 20, pp. 11050–11055, Sep. 2000.

[14] B. Fischl, M. I. Sereno, R. B. H. Tootell, and A. M. Dale, "High-resolution intersubject averaging and a coordinate system for the cortical surface," *Hum. Brain Mapp.*, vol. 8, no. 4, pp. 272–284, 1999.

[15] L. Fan *et al.*, "The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture," *Cereb. Cortex N. Y. N 1991*, vol. 26, no. 8, pp. 3508–3526, Aug. 2016.

[16] R. S. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, 2006.

[17] R. Ossenkoppele *et al.*, "Atrophy patterns in early clinical stages across distinct phenotypes of Alzheimer's disease: Origin and Spread of Atrophy in AD Variants," *Hum. Brain Mapp.*, vol. 36, no. 11, pp. 4421–4437, Nov. 2015.

[18] A. M. Tetreault *et al.*, "Network localization of clinical, cognitive, and neuropsychiatric symptoms in Alzheimer's disease," *Brain*, p. awaa058, Mar. 2020.

[19] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with B -splines and penalties," *Stat. Sci.*, vol. 11, no. 2, pp. 89–121, May 1996.

[20] G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Idrobo, Y. Schwartz, and B. Thirion, "Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines," *NeuroImage*, vol. 145, pp. 166–179, Jan. 2017.

[21] G. C. Cawley and N. L. C. Talbot, "On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation," *J. Mach. Learn. Res.*, vol. 11, pp. 2079–2107, Aug. 2010.

[22] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. null, pp. 281–305, Feb. 2012.

[23] M. R. Lavery, P. Acharya, S. A. Sivo, and L. Xu, "Number of predictors and multicollinearity: What are their effects on error and bias in regression?," *Commun. Stat. - Simul. Comput.*, vol. 48, no. 1, pp. 27–38, Jan. 2019.

[24] C. F. Dormann *et al.*, "Collinearity: a review of methods to deal with it and a simulation study evaluating their performance," *Ecography*, vol. 36, no. 1, pp. 27–46, 2013.

[25] J. I. Daoud, "Multicollinearity and Regression Analysis," *J. Phys. Conf. Ser.*, vol. 949, p. 012009, Dec. 2017.

[26] Y. Bengio and Y. Grandvalet, "No Unbiased Estimator of the Variance of K-Fold Cross-Validation," *J. Mach. Learn. Res.*, vol. 5, pp. 1089–1105, Dec. 2004.