

Table 2

Tests of correlation between the abnormal DTI indices and the clinical variables.

	Correlation coefficients (P value)						
	Total disease duration	Duration of current episode	HDRS-17	GAF	MADRS	CGIS	Age
Clusters with significant decrease in FA							
Right frontal white matter (x=12, y=43, z=-18)	-0.296 (0.219)	-0.482 (0.036)*	0.084 (0.732)	-0.012 (0.960)	-0.117 (0.634)	-0.006 (0.980)	0.188 (0.440)
Right frontal white matter (x=37, y=-19, z=59)	0.080 (0.746)	0.318 (0.184)	0.135 (0.583)	-0.462 (0.046)*	0.413 (0.079)	0.412 (0.080)	-0.392 (0.097)
Left frontal white matter	0.247 (0.308)	0.087 (0.723)	0.160 (0.512)	-0.367 (0.122)	0.351 (0.141)	0.377 (0.112)	0.109 (0.657)
Right anterior limb of internal capsule	0.335 (0.160)	0.163 (0.506)	-0.117 (0.634)	-0.479 (0.038)*	0.229 (0.345)	0.363 (0.127)	-0.070 (0.775)
Left anterior limb of internal capsule	0.483 (0.036)*	-0.011 (0.964)	0.078 (0.750)	-0.410 (0.081)	0.384 (0.105)	0.359 (0.131)	-0.308 (0.200)
Left putamen	-0.112 (0.648)	-0.191 (0.433)	-0.026 (0.915)	-0.276 (0.252)	0.120 (0.626)	0.327 (0.171)	-0.397 (0.092)
Mediodorsal nucleus of the right thalamus	0.434 (0.063)	-0.116 (0.635)	0.052 (0.831)	-0.248 (0.307)	0.152 (0.535)	0.326 (0.173)	-0.424 (0.071)
Right cerebellar hemisphere	-0.073 (0.767)	-0.034 (0.891)	0.525 (0.021)*	-0.383 (0.106)	0.272 (0.260)	0.345 (0.148)	0.258 (0.286)
Left cerebellar hemisphere	0.168 (0.492)	0.214 (0.380)	0.434 (0.063)	-0.274 (0.257)	0.252 (0.297)	0.177 (0.470)	0.498 (0.030)*
Peak position of FA histogram	0.576 (0.010)*	0.550 (0.015)*	-0.020 (0.936)	-0.087 (0.723)	0.402 (0.088)	0.040 (0.869)	0.219 (0.368)

Abbreviations: HDRS-17=17-item Hamilton depression rating scale; GAF=Global assessment of functioning; MADRS=Montgomery-Åsberg depression rating scale; CGI-S=Clinical global impression-severity.

* Indicates statistical significance ($P < 0.05$).

cortex, subcortical limbic structures, and monoamine-producing brainstem nuclei (Konarski et al., 2005; Strick et al., 2009). Altered cerebellar levels of glial fibrillary acidic proteins (GFAP) have been observed in the patients with MDD (Fatemi et al., 2004). PET and functional MRI studies have also uncovered distinct variations in cerebellar activity between the patients with MDD and normal subjects (e.g., Dolan et al., 1992; Beauguard et al., 1998; Videbeck et al., 2001). Our finding of decrease in the FA in the bilateral cerebellar hemispheres is also supportive of implication of the cerebellum in the neuropathology of MDD.

Previous DTI studies on non-geriatric patients with MDD who were free of antidepressant medication have reported decreases in the FA in the white matter of the superior longitudinal fasciculus, cingulate, sagittal stratum, posterior thalamic radiation, anterior limb and retrolenticular part of the internal capsule, external capsule, splenium of the corpus callosum, stria terminalis, and the other subcortical and deep white matter of the frontal, parietal and temporal lobes (Li et al., 2007; Ma et al., 2007; Korgaonkar et al., 2011; Ouyang et al., 2011; Wu et al., 2011; Zhu et al., 2011). Added to these are our findings of the decrease in the FA in the left putamen, the mediodorsal nucleus of the right thalamus, and the bilateral cerebellar hemispheres. Taken together, these findings are suggestive of heterogeneity in the areas of the brain that are involved in MDD. It is possible that such heterogeneity arises from a variation in the circuits that are severed among the study population. From our finding of abnormalities in the frontal white matter and mediodorsal nucleus of the thalamus, it is thought that the dorsolateral prefrontal or orbitofrontal circuit is at least involved in our patients (Tekin and Cummings, 2002). Further stratification of analysis based on clinical symptoms may render more consistent results; however, variation in the exact anatomical location of the abnormalities along a single circuit (i.e., in terms of MNI or Talairach coordinates) would remain unsolved (Blood et al., 2010).

In this study, histogram analysis of FA and MD was also performed, as a supplemental analysis to voxel-based analysis. To our knowledge, this is the first study that has applied a

histogram analysis of DTI in psychiatric diseases. This technique involves the creation of a frequency distribution showing the proportion of voxels in an image within a given range of signal intensity (Cercignani, 2011). Two major indices, peak height and location, are usually evaluated in histogram analysis. While the former reflects the change in a part of voxels which contribute to the maximum frequency, the latter is suggestive of rather a global change in voxels which contribute to the maximum frequency (Cercignani et al., 2001). This might suggest that there remained voxels with low FA values left undiscovered by the voxel-based analysis. Additional indices that can be evaluated include skewness (the magnitude of symmetry of the histograms) and kurtosis (peakness of histograms) (Mori et al., 2008). Histogram analysis of DTI has been successfully used in brain tumor classification (Wang et al., 2012), investigation of clinical correlates in leukoariosis (Della Nave et al., 2007) and multiple sclerosis (Cercignani et al., 2001), differentiation between ischemic and non-ischemic patients with moyamoya disease (Mori et al., 2008), evaluation of traumatic brain injury (Marquez de la Plata et al., 2011), and so on. In this study, histogram analysis enabled a discrimination between the patients and normal subjects by its peak position, which is suggestive of its applicability in the evaluation of integrity of brain parenchyma in MDD. When the cut-off value for the peak position of an FA histogram is 0.2, the sensitivity and specificity in distinguishing the patients from normal subjects are 0.89 and 0.74, respectively, as derived from ROC analysis (www.statsdirect.com). The strengths of histogram analysis include its ability to test many thousands of voxels as well as its robust statistical power (Della Nave et al., 2007). Its drawback is limited topographic information.

Because this study is an exploratory study, tests of correlation were performed without correction for multiple comparisons. A number of significant correlations between the FA values and the clinical variables were observed. The finding of positive correlation between the FA values and the clinical variables, especially clinical severity, is unusual. It is possible that the results of tests of correlation are inflated by Type I error associated with multiple comparisons (Gordji and Khamis, 2004). If corrected for

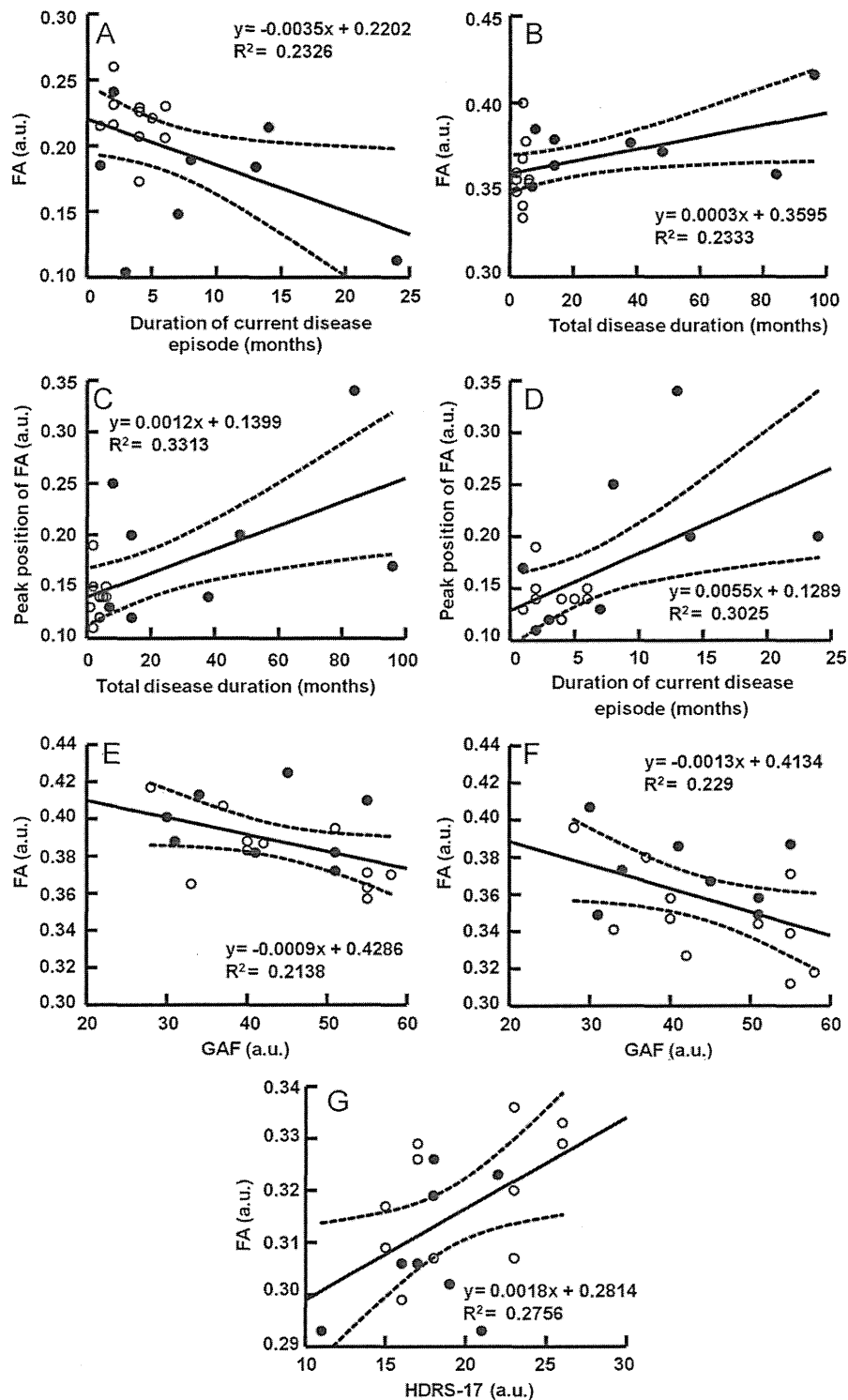


Fig. 4. Scatterplots showing correlation between the abnormal DTI indices and the clinical variables. Correlations between (A) the FA values of right frontal white matter ($x=12, y=43, z=-18$) and the duration of current disease episode, (B) the FA values of the left anterior limb of internal capsule and total disease duration, (C) the peak positions of FA histograms and total disease duration, (D) the peak positions of FA histograms and the duration of current disease episode, (E) the FA values of the right frontal white matter ($x=37, y=-19, z=59$) and the GAF, (F) the FA values of the right anterior limb of internal capsule and the GAF, and (G) the FA values of the right cerebellar hemisphere and the HDRS-17, are shown. White circles represent the medication-naïve patients. Black circles represent the patients who had a previous history of antidepressant medication. The solid lines represent the mean, and the dashed lines represent 95% confidence interval for the mean predicted value. Abbreviations: a.u.=arbitrary unit; FA=fractional anisotropy; GAF=global assessment of functioning; HDRS-17=17-item Hamilton depression rating scale.

multiple comparisons, no pairs with significant correlation would persist. Added to this, the slope of the regression lines might have been modified, and the correlation coefficient values might have

been artificially increased, by the outliers (Hatch and Prihoda, 1992). On the other hand, if these correlations are considered as true correlations, the observation of difference in the pattern of

correlation among the regions would suggest variation in the pattern of injury among the regions. Negative correlation between the duration of current disease episode and the FA value of the right frontal white matter ($x=12$, $y=43$, $z=-18$) would be indicative of impaired frontal white matter integrity with disease duration; and may reflect advancement of axon and/ or myelin injury with time. Positive correlation between disease duration and the regional FA values may be due to alteration in the local microenvironment by the long disease process or previous antidepressant medication. First, patients with long disease duration are exposed to chronic stress. Chronic stress has been demonstrated to contribute to frontostriatal regeneration and reductions in the length and branch numbers of apical dendrites, particularly in the anterior cingulate cortex (Radley et al., 2004). As dendrite formation and increased synaptic density are associated with decrease in the FA, the dendritic changes induced by chronic stress could in turn present with a relative increase in the FA (Baloch et al., 2009; Taylor et al., 2011). However, owing to difference in histological composition between the two structures, the finding of anterior cingulate cortex may not plausibly explain the positive correlation between the FA value of the left anterior limb of the internal capsule and disease duration. Second, in this study, eight patients (those with long disease duration) had a previous history of antidepressant medication. Considering the half-lives of antidepressants, these patients were determined as being free of medication effects. However, the possibility that these patients' tissue microenvironment was modified and/or aggressively treated by previous medication cannot be completely excluded. The significant positive correlation between the regional FA values and clinical severity might be due to astrocytic hypertrophy. Although the evidence is still limited, a recent autopsy report that investigated the white matter underlying anterior cingulate cortex of depressed suicides patients has revealed an increase in the volume of astrocytic processes (representative of astrocytic hypertrophy) in the white matter (Torres-Platas et al., 2011). If it is assumed that astrocytic hypertrophy also occurs in the other white matter regions involved in MDD and that the degree of hypertrophy is severe enough to induce an increase in the regional FA values, the finding of FA increase with disease severity would be explained.

A few limitations of this study need to be addressed. First, correction for eddy current-induced distortion was not performed. This resulted in an imperfect match between the white matter masks constructed from the anatomical images (MPRAGE) and the FA and MD maps. Estimation of the average displacement between the MPRAGE images and the FA maps revealed an average shift of 0.91, 2.18, and 0.30 mm along the x, y, and z directions, respectively. Nevertheless, the magnitude of displacement did not vary significantly between the patients and control subjects, so as to maintain validity of the analytical method for evaluation of global brain parenchymal or white matter abnormalities. Second, the slice thickness used for DTI was 5 mm, and the interslice gap was 1.5 mm. These parameters were chosen so as to allow acquisition of all necessary sequences in clinically acceptable scan time. Ideally, thinner slices without interslice gap are desired. Third, DSM-IV-TR was used to diagnose MDD in the patients, whereas MINI was used to exclude psychiatric diseases in normal subjects. Hence, there might be a question about potential bias arising from the use of different diagnostic assessment methods between the two groups. However, the possibility is considered unlikely as MINI is well-designed to be used to rule out psychiatric diseases, and is compatible with DSM-IV-TR (Sheehan et al., 1998).

In conclusion, this study evaluated the integrity of the brain parenchyma in non-geriatric patients with MDD by using two different analytical methods of DTI. The findings suggest an

impaired integrity of the frontal-subcortical circuits and cerebello-cerebellar connections, which are known or thought to mediate human behavior. Consistency in the distribution of abnormalities is lacking among the studies including the current report, which may suggest spatial heterogeneity of the abnormalities in MDD. The peak position of the FA histograms may be used as supplementary to voxel-based analysis, in evaluation of the brain parenchymal integrity in MDD.

Acknowledgement

The authors are greatly indebted to the anonymous reviewers for the invaluable comments. This study was supported by (1) Creation of Innovation Centers for Advanced Interdisciplinary Research Areas Program of Project for Developing Innovation Systems of the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government, and (2) the Grant-in-aid for Scientific Research by the Japan Society for Promotion of Science (Project number 20591430).

References

- Alexopoulos, G.S., Kiosses, D.N., Choi, S.J., Murphy, C.F., Lim, K.O., 2002. Frontal white matter microstructure and treatment response of late-life depression: a preliminary study. *American Journal of Psychiatry* 159, 1929–1932.
- Alexopoulos, G.S., Murphy, C.F., Gunning-Dixon, F.M., Latoussakis, V., Kanellopoulos, D., Klimstra, S., Lim, K.O., Hoptman, M.J., 2008. Microstructural white matter abnormalities and remission of geriatric depression. *American Journal of Psychiatry* 165, 238–244.
- Bae, J.N., MacFall, J.R., Krishnan, K.R., Payne, M.E., Steffens, D.C., Taylor, W.D., 2006. Dorsolateral prefrontal cortex and anterior cingulate cortex white matter alterations in late-life depression. *Biological Psychiatry* 60, 1356–1363.
- Baloch, S., Verma, R., Huang, H., Khurd, P., Clark, S., Yarowsky, P., Abel, T., Mori, S., Davatzikos, C., 2009. Quantification of brain maturation and growth patterns in C57BL/6 J mice via computational neuroanatomy of diffusion tensor images. *Cerebral Cortex* 19, 675–687.
- Basser, P.J., Pierpaoli, C., 1996. Microstructural and physiological features of tissues elucidated by quantitative-diffusion tensor MRI. *Journal of Magnetic Resonance B* 111, 209–219.
- Beaulieu, C., 2011. What makes diffusion anisotropic in the nervous system?. In: Jones, D.K. (Ed.), *Diffusion MRI: Theory, Methods, and Applications*. Oxford University Press, New York, pp. 92–109.
- Beauregard, M., Leroux, J.M., Bergman, S., Arzoumanian, Y., Beaudoin, G., Bourgoin, P., Stip, E., 1998. The functional neuroanatomy of major depression: an fMRI study using an emotional activation paradigm. *Neuroreport* 14, 3253–3258.
- Blood, A.J., Iosifescu, D.V., Makris, N., Perlis, R.H., Kennedy, D.N., Dougherty, D.D., Kim, B.W., Lee, M.J., Wu, S., Lee, S., Calhoun, J., Hodge, S.M., Fava, M., Rosen, B.R., Smoller, J.W., Gasic, G.P., Breiter, H.C., 2010. Microstructural abnormalities in subcortical reward circuitry of subjects with major depressive disorder. *PLoS One* 11, e13945.
- Cercignani, M., Inglese, M., Pagani, E., Comi, G., Filippi, M., 2001. Mean diffusivity and fractional anisotropy histograms of patients with multiple sclerosis. *American Journal of Neuroradiology* 22, 952–958.
- Cercignani, M., 2011. Strategies for patient-control comparison of diffusion MR data. In: Jones, D.K. (Ed.), *Diffusion MRI: Theory, Methods, and Applications*. Oxford University Press, New York, pp. 485–499.
- Della Nave, R., Foresti, S., Pratesi, A., Ginestroni, A., Inzitari, M., Salvadori, E., Giannelli, M., Diciotti, S., Inzitari, D., Mascalchi, M., 2007. Whole-brain histogram and voxel-based analyses of diffusion tensor imaging in patients with leukoariosis: correlation with motor and cognitive impairment. *American Journal of Neuroradiology* 28, 1313–1319.
- Diamond, A., 2000. Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. *Child Development* 71, 44–56.
- Dolan, R.J., Bench, C.J., Brown, R.G., Scott, L.C., Friston, K.J., Frackowiak, R.S., 1992. Regional cerebral blood flow abnormalities in depressed patients with cognitive impairment. *Journal of Neurology, Neurosurgery & Psychiatry* 55, 768–773.
- Egger, K., Schocke, M., Weiss, E., Auffinger, S., Esterhammer, R., Goebel, G., Walch, T., Mechtcheriakov, S., Marksteiner, J., 2008. Pattern of brain atrophy in elderly patients with depression revealed by voxel-based morphometry. *Psychiatry Research* 164, 237–244.
- Fatemi, S.H., Laurence, J.A., Aragi-Niknam, M., Sary, J.M., Schulz, S.C., Lee, S., Gottesman, I.I., 2004. Glial fibrillary acidic protein is reduced in cerebellum of subjects with major depression, but not schizophrenia. *Schizophrenia Research* 69, 317–323.

- Fujimoto, T., Takeuchi, K., Matsumoto, T., Fujita, S., Honda, K., Higashi, Y., Kato, N., 2008. Metabolic changes in the brain of patients with late-onset major depression. *Psychiatry Research* 164, 48–57.
- Gordji, T., Khamis, H., 2004. Simple solution to a common statistical problem: interpreting multiple tests. *Clinical Therapeutics* 26, 780–786.
- Graff-Guerrero, A., González-Olvera, J., Mendoza-Espinosa, Y., Vaugier, V., García-Reyna, J.C., 2004. Correlation between cerebral blood flow and items of the Hamilton rating scale for depression in antidepressant-naïve patients. *Journal of Affective Disorders* 80, 55–63.
- Greenberg, P.E., Stiglin, L.E., Finkelstein, S.N., Berndt, E.R., 1993. The economic burden of depression in 1990. *Journal of Clinical Psychiatry* 54, 405–418.
- Guy, W., 1976. *ECDEU Assessment Manual for Psychopharmacology*. Department of Health, Education, and Welfare, Rockville, MD: U.S.
- Hall, R.C.W., 1995. Global assessment of functioning: a modified scale. *Psychosomatics* 36, 267–275.
- Hatch, J.P., Pridhoda, T.J., 1992. The effect of influential outliers on parameter estimation in regression analysis. *Biofeedback and Self-Regulation*, 17, 153–156.
- Jones, D.K., Symms, M.R., Cercignani, M.C., Howard, R.J., 2005. The effect of filter size on the outcome of VBM analyses of DT-MRI data. *Neuroimage* 26, 546–554.
- Kessler, R.C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K.R., Rush, A.J., Walters, E.E., Wang, P.S., 2003. National comorbidity survey replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *The Journal of American Medical Association* 289, 3095–3105.
- Konarski, J.Z., McIntyre, R.S., Grupp, L.A., Kennedy, S.H., 2005. Is the cerebellum relevant in the circuitry of neuropsychiatric disorders? *Journal of Psychiatry and Neuroscience* 30, 178–186.
- Korgaonkar, M.S., Grieve, S.M., Koslow, S.H., Gabrieli, J.D., Gordon, E., Williams, L.M., 2011. Loss of white matter integrity in major depressive disorder: evidence using tract-based spatial statistical analysis of diffusion tensor imaging. *Human Brain Mapping* 32, 2161–2171.
- Li, L., Ma, N., Li, Z., Tan, L., Liu, J., Gong, G., Shu, N., He, Z., Jiang, T., Xu, L., 2007. Prefrontal white matter abnormalities in young adult with major depressive disorder: a diffusion tensor imaging study. *Brain Research* 1168, 124–128.
- Ma, N., Li, L., Shu, N., Liu, J., Gong, G., He, Z., Li, Z., Tan, L., Stone, W.S., Zhang, Z., Xu, L., Jiang, T., 2007. White matter abnormalities in first-episode, treatment-naïve young adults with major depressive disorder. *American Journal of Psychiatry* 164, 823–826.
- Malykhin, N., Concha, L., Seres, P., Beaulieu, C., Coupland, N.J., 2008. Diffusion tensor imaging tractography and reliability analysis for limbic and paralimbic white matter tracts. *Psychiatry Research* 164, 132–142.
- Marquez de la Plata, C.D., Yang, F.G., Wang, J.Y., Krishnan, K., Bakhadirov, K., Paliotta, C., Aslan, S., Devous, M.D., Moore, C., Harper, C., McCall, R., Munro Cullum, C., Diaz-Arrastia, R., 2011. Diffusion tensor imaging biomarkers for traumatic axonal injury: analysis of three analytic methods. *Journal of the International Neuropsychological Society* 17, 24–35.
- Montgomery, S.A., Åsberg, M., 1979. A new depression scale designed to be sensitive to change. *British Journal of Psychiatry* 134, 382–389.
- Mori, N., Miki, Y., Fushimi, Y., Kikuta, K., Urayama, S., Okada, T., Fukuyama, H., Hashimoto, N., Togashi, K., 2008. Cerebral infarction associated with moyamoya disease: histogram-based quantitative analysis of diffusion tensor imaging—a preliminary study. *Magnetic Resonance Imaging* 26, 835–840.
- Moseley, M., 2002. Diffusion tensor imaging and aging: a review. *NMR in Biomedicine* 15, 553–560.
- Nagesh, V., Tsien, C.I., Chenevert, T.L., Ross, B.D., Lawrence, T.S., Junick, L., Cao, Y., 2008. Radiation-induced changes in normal-appearing white matter in patients with cerebral tumors: a diffusion tensor imaging study. *International Journal of Radiation Oncology Biology Physics* 70, 1002–1010.
- Nobuhara, K., Okugawa, G., Sugimoto, T., Minami, T., Tamagaki, C., Takase, K., Saito, Y., Sawada, S., Kinoshita, T., 2006. Frontal white matter anisotropy and symptom severity of late-life depression: a magnetic resonance diffusion tensor imaging study. *Journal of Neurology, Neurosurgery and Psychiatry* 77, 120–122.
- Ouyang, X., Tao, H.J., Liu, H.H., Deng, Q.J., Sun, Z.H., Xu, L., Liu, Z.N., Xue, Z.M., 2011. White matter integrity deficit in treatment-naïve adult patients with major depressive disorder. *East Asian Archives of Psychiatry* 21, 5–9.
- Radley, J.J., Sisti, H.M., Hao, J., Rocher, A.B., McCall, T., Hof, P.R., McEwen, B.S., Morrison, J.H., 2004. Chronic behavioral stress induces apical dendritic reorganization in pyramidal neurons of the medial prefrontal cortex. *Neuroscience* 125, 1–6.
- Regenold, W.T., Phatak, P., Marano, C.M., Gearhart, L., Viens, C.H., Hisley, K.C., 2007. Myelin staining of deep white matter in the dorsolateral prefrontal cortex in schizophrenia, bipolar disorder, and unipolar major depression. *Psychiatry Research* 151, 179–188.
- Sala, S., Agosta, F., Pagani, E., Copetti, M., Comi, G., Filippi, M., 2012. Microstructural changes and atrophy in brain white matter tracts with aging. *Neurobiology of Aging* 33, 488–498.
- Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The mini-international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry* 59, 22–33.
- Shimony, J.S., Sheline, Y.I., D'Angelo, G., Epstein, A.A., Benzinger, T.L., Mintun, M.A., McKinstry, R.C., Snyder, A.Z., 2009. Diffuse microstructural abnormalities of normal-appearing white matter in late life depression: a diffusion tensor imaging study. *Biological Psychiatry* 66, 245–252.
- Sijens, P.E., Mostert, J.P., Irwan, R., Potze, J.H., Oudkerk, M., Keyser, J.D., 2008. Impact of fluoxetine on the human brain in multiple sclerosis as quantified by proton magnetic resonance spectroscopy and diffusion tensor imaging. *Psychiatry Research: Neuroimaging* 164, 274–282.
- Strick, P.L., Dum, R.P., Fiez, J.A., 2009. Cerebellum and nonmotor function. *The Annual Review of Neuroscience* 32, 413–434.
- Taoka, T., Kin, T., Nakagawa, H., Hirano, M., Sakamoto, M., Wada, T., Takayama, K., Wuttikul, C., Iwasaki, S., Ueno, S., Kichikawa, K., 2007. Diffusivity and diffusion anisotropy of cerebellar peduncles in cases of spinocerebellar degenerative disease. *Neuroimage* 37, 387–393.
- Taylor, W.D., MacFall, J.R., Boyd, B., Payne, M.E., Sheline, Y.I., Krishnan, K.R.R., Doraiswamy, P.M., 2011. One-year change in anterior cingulate cortex white matter microstructure: relationship with late-life depression outcomes. *American Journal of Geriatric Psychiatry* 19, 43–52.
- Taylor, W.D., MacFall, J.R., Payne, M.E., McQuoid, D.R., Provenzale, J.M., Steffens, D.C., Krishnan, K.R., 2004. Late-life depression and microstructural abnormalities in dorsolateral prefrontal cortex white matter. *American Journal of Psychiatry* 161, 1293–1296.
- Tekin, S., Cummings, J.L., 2002. Frontal-subcortical neuronal circuits and clinical neuropsychiatry: an update. *Journal of Psychosomatic Research* 53, 647–654.
- Tha, K.K., Terae, S., Yabe, I., Miyamoto, T., Soma, H., Zaitsu, Y., Fujima, N., Kudo, K., Sasaki, H., Shirato, H., 2010. Microstructural white matter abnormalities of multiple system atrophy: in vivo topographic illustration by using diffusion-tensor MR imaging. *Radiology* 255, 563–569.
- Torres-Platas, S.G., Hercher, C., Davoli, M.A., Maussion, G., Labonté, B., Turecki, G., Mechawar, N., 2011. Astrocytic hypertrophy in anterior cingulate white matter of depressed suicides. *Neuropsychopharmacology* 36, 2650–2658.
- Tsopoulos, C., Stewart, R., Sawa, G.M., Brayne, C., Ince, P., Thomas, A., Matthews, F.E., Medical Research Council Cognitive Function and Ageing Study, 2011. Neuropathological correlates of late-life depression in older people. *The British Journal of Psychiatry* 198, 109–114.
- Uranova, N.A., Vostrikov, V.M., Orlovskaya, D.D., Rachmanova, V.I., 2004. Oligodendroglial density in the prefrontal cortex in schizophrenia and mood disorders: a study from the Stanley neuropathology consortium. *Schizophrenia Research* 67, 269–275.
- Videbech, P., Ravnkilde, B., Pedersen, A.R., Egander, A., Landbo, B., Rasmussen, N.A., Andersen, F., Stødkilde-Jørgensen, H., Gjedde, A., Rosenberg, R., 2001. The Danish PET/depression project: PET findings in patients with major depression. *Psychological Medicine* 31, 1147–1158.
- Wang, S., Kim, S., Zhang, Y., Wang, L., Lee, E.B., Syre, P., Poptani, H., Melhem, E.R., Lee, J.Y., 2012. Determination of grade and subtypes of meningiomas by using histogram analysis of diffusion-tensor imaging metrics. *Radiology* 262, 584–592.
- Waxman, H.M., Carner, E.A., Dubin, W., Klein, M., 1982. Geriatric psychiatry in the emergency department: characteristics of geriatric and non-geriatric admissions. *Journal of the American Geriatrics Society* 30, 427–432.
- Williams, J.B., 1988. A structured interview guide for the Hamilton depression rating scale. *Archives of General Psychiatry* 45, 742–747.
- Wu, F., Tang, Y., Xu, K., Kong, L., Sun, W., Wang, F., Kong, D., Li, Y., Liu, Y., 2011. White matter abnormalities in medication-naïve subjects with a single short-duration episode of major depressive disorder. *Psychiatry Research* 191, 80–83.
- Yang, Q., Huang, X., Hong, N., Yu, X., 2007. White matter microstructural abnormalities in late-life depression. *International Psychogeriatrics* 19, 757–766.
- Yoo, S.Y., Jang, J.H., Shin, Y.H., Kim, D.J., Park, H.J., Moon, W.J., Chung, E.C., Lee, J.M., Kim, I.Y., Kim, S.I., Kwon, J.S., 2007. White matter abnormalities in drug-naïve patients with obsessive-compulsive disorder: a diffusion tensor study before and after citalopram treatment. *Acta Psychiatrica Scandinavica* 116, 211–219.
- Zhu, X., Wang, X., Xiao, J., Zhong, M., Liao, J., Yao, S., 2011. Altered white matter integrity in first-episode, treatment-naïve young adults with major depressive disorder: a tract-based spatial statistics study. *Brain research* 1369, 223–229.
- Zou, K., Huang, X., Li, T., Gong, Q., Li, Z., Ou-Yang, L., Deng, W., Chen, Q., Li, C., Ding, Y., Sun, X., 2008. Alterations of white matter integrity in adults with major depressive disorder: a magnetic resonance imaging study. *Journal of Psychiatry & Neuroscience* 33, 525–530.

RESEARCH ARTICLE

Open Access

Utility and limitations of PHQ-9 in a clinic specializing in psychiatric care

Takeshi Inoue*, Teruaki Tanaka, Shin Nakagawa, Yasuya Nakato, Rie Kameyama, Shuken Boku, Hiroyuki Toda, Tsugiko Kurita and Tsukasa Koyama

Abstract

Background: The Patient Health Questionnaire-9 (PHQ-9), despite its excellent reliability and validity in primary care, has not been examined for administration to psychiatric patients. This study assesses the accuracy of PHQ-9 in screening for major depressive episode and in diagnosing major depressive episode in patients of a psychiatric specialty clinic.

Methods: We compared operational characteristics of PHQ-9 as a screening and diagnostic instrument to *DSM-IV-TR* diagnosis by a trained psychiatrist as a reference standard. The reference criteria were "current major depressive episode" or "current major depressive episode with major depressive disorder". PHQ-9 was used with two thresholds: diagnostic algorithm and summary scores (PHQ-9 \geq 10). The optimal cut-off points of PHQ-9 summary scores were analyzed using a receiver operational characteristics (ROC) curve.

Results: For "current major depressive episode", PHQ-9 showed high sensitivity and high negative predictive value at both thresholds, but its specificity and positive predictive value were low. For "current major depressive episode with major depressive disorder", PHQ-9 also showed high sensitivity and high negative predictive value at both thresholds, but the positive predictive value decreased more than that for "current major depressive episode". The ROC analysis showed the optimal cut-off score of 13/14 for "current major depressive episode".

Conclusions: PHQ-9 is useful for screening, but not for diagnosis of "current major depressive episode" in a psychiatric specialty clinic.

Background

The Patient Health Questionnaire (PHQ) was developed in 1999 as a self-report version of the Primary Care Evaluation of Mental Disorders (PRIME-MD), which was designed for criteria-based diagnosis of several mental disorders that are commonly encountered in primary care [1]. The diagnostic validity of the depression module of the PHQ (PHQ-9) has been reported in a study involving 3000 patients in eight primary care clinics. Moreover, PHQ-9 for the diagnosis of major depressive disorder and any mood disorder showed moderate agreement with the diagnosis of the Structured Clinical Interview for *DSM-III-R* (SCID) by mental health professionals [1]. The validity for screening against the *DSM-IV* diagnosis of major depression, reliability, and

feasibility of PHQ-9 is regarded as excellent [2]. PHQ-9 consists of the nine diagnostic criteria items of *DSM-IV* major depressive episode. Therefore, it has good logical validity for the *DSM-IV* diagnosis of major depressive episode.

Two recent meta-analyses have shown good pooled sensitivity (0.80 and 0.77, respectively) and specificity (0.92 and 0.94, respectively) for PHQ-9 against the *DSM-IV* diagnosis of major depressive disorder or major depressive episode in the settings of primary care clinics and clinics other than psychiatric clinics [3,4]. However, the utility and validity of PHQ-9 for the screening for major depressive episode or major depressive disorder has not been examined in a psychiatric specialty clinic. Although one might argue that such a brief self-report screening test is not necessary for specialist psychiatrists, various self-administered questionnaires are often used before a psychiatric interview, even in clinics specializing in psychiatric treatment. Above all, the screening for

* Correspondence: tinoue@med.hokudai.ac.jp
Department of Psychiatry, Hokkaido University Graduate School of Medicine, North 15, West 7, Kita-ku, Sapporo 060-8638, Japan

major depressive episode, which has a high prevalence, using a self-administered questionnaire such as PHQ-9 is useful to prevent underdiagnosis in patients with several other psychiatric disorders that might have comorbidity with major depressive episode. Therefore, the utility and limitation of PHQ-9 for patients with psychiatric disorders in psychiatric specialty clinics must be studied before its application to psychiatric practice.

For this study, a mood disorder specialist psychiatrist administered PHQ-9 to patients of psychiatric specialty clinics of a university hospital, and compared its diagnostic capability with that of *DSM-IV-TR* for diagnosis of major depressive episode and major depressive disorder. In earlier studies in primary care clinics and clinics other than psychiatric clinics, the *DSM-IV* diagnosis was done using SCID by trained interviewers or mental health professionals, but not by psychiatrists [1,5,6], which might be a disadvantage of earlier studies because psychiatric diagnosis, especially of hypomanic episodes, a history of which has a major impact on the accuracy of PHQ-9 screening for major depressive disorder, is extremely difficult, demanding many years of training. Additionally, we evaluated and discussed diagnostic factors contributing to false-positive diagnosis.

Methods

Subjects

From February 2008 to July 2009, 153 outpatients, who visited the Department of Psychiatry, Hokkaido University Hospital as new patients, were consecutively included in the study. Each had been diagnosed using the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)* [7] by a mood disorder specialist psychiatrist (T.I.) who was blinded to the PHQ-9 results and who had more than 20 years of clinical experience in this field of psychiatry. The Japanese version of PHQ-9 [8] was administered to patients during their waiting time as a routine clinical task. This study was performed in accordance with the Declaration of Helsinki and was approved by the institutional review board of Hokkaido University Hospital.

PHQ-9

The PHQ-9 was self-completed by the patient in written form. Major depressive episodes were diagnosed in two ways using the PHQ-9: diagnostic algorithm and a summary score. The diagnostic algorithmic threshold for diagnosing major depressive episode was regarded as fulfilled if the answer to question #1a or question #1b and five or more of questions #1a–#1i was at least “more than half the days” (question #1i was counted if present at all) [1]. The threshold $PHQ \geq 10$ signified that the summary score of questions #1a–#1i (range = 0–27) must be 10 or higher. This cutoff point was chosen because it

has been reported and recommended the most consistently [3,4,6].

Psychiatric evaluations

DSM-IV-TR diagnoses of various psychiatric disorders including mood disorders (major depressive disorder, minor depressive disorder, bipolar disorder, dysthymic disorder, and cyclothymic disorder) were made by a psychiatrist specializing in mood disorder (T.I.) using the *Quick Reference to Diagnostic Criteria from the DSM-IV-TR* on the same day when the patients answered PHQ-9. The average interview duration was 60 min. In each case, the presence of a current major depressive episode, which can appear in patients with major depressive disorder or bipolar disorder, was identified. In most patients with major depressive disorder, the 17-item Hamilton Depression Rating Scale (HDRS) and the Montgomery–Åsberg Depression Rating Score (MADRS, 10 items) were administered to evaluate the severity of depression in routine clinical work [9,10]. We also used the global assessment of functioning (GAF) scale, which reflects the overall level of psychological, social, and occupational functioning of individuals [7]. Some patients, however, did not receive the HDRS, MADRS, or GAF evaluation because of the limited time of the regular clinical interview.

Data analysis

With respect to criterion validity, we investigated sensitivity, specificity, positive predictive value, negative predictive value, and the overall accuracy for the diagnostic algorithmic threshold and summary score threshold ($PHQ \geq 10$). The *DSM-IV-TR* diagnoses of “current major depressive episode” and “current major depressive episode with major depressive disorder” were the criterion standard.

Receiver Operating Characteristics (ROC) analysis was used to ascertain an optimal cut-off point of PHQ-9 summary scores for screening for major depressive episode in the setting of a psychiatric specialty clinic. This curve shows “sensitivity” versus “1-specificity” for every possible cut-off point. The area under the curve (AUC) was calculated as an indicator of the discriminative property of the scale. A maximal “Youden Index” (sensitivity + specificity – 1) was calculated and suggested as an optimal cut-off point for PHQ-9 [11].

Pearson’s correlation coefficients between PHQ-9 scores and GAF or MADRS or HDRS scores of patients with major depressive disorder were calculated to assess their validity for severity measures of depressive symptoms.

All continuous data are presented as means with standard deviations or 95 % confidence intervals (CIs).

Results

Demographic characteristics and DSM-IV-TR diagnosis of subjects

Demographic characteristics and *DSM-IV-TR* diagnoses of 153 subjects are presented in Table 1. The most common diagnosis was major depressive disorder with subsequent bipolar disorder: 45 % of subjects were diagnosed with mood disorders. Two patients with major depressive disorder had comorbid psychiatric disorders: mild dementia and social anxiety disorder. Among them, only one patient with social anxiety disorder had a current major depressive episode that fulfilled its diagnostic criteria.

Validity of PHQ-9 for the screening for current major depressive episode

Table 2 (left half) presents operating characteristics of PHQ-9 for “current major depressive episode” for the PHQ-9 diagnostic algorithmic threshold and for the cut-

Table 1 Characteristics and *DSM-IV-TR* diagnoses of 153 patients

Characteristic	Value
Sex	
Female, n (%); male, n (%)	96 (63); 57 (37)
Age, mean ± SD (yr)	44.6 ± 18.3
Range	14–82
<i>DSM-IV-TR</i> diagnosis, n (%)	
Major Depressive Disorder	50 (32.7)
Minor Depressive Disorder	2 (1.3)
Dysthymic Disorder	1 (0.7)
Bipolar Disorder	16 (10.5)
Developmental Disorders	2 (1.3)
Dementia	10 (6.5)
Mental Disorders due to a General Medical Condition	7 (4.6)
Substance-Related Disorders	3 (2.0)
Schizophrenia	8 (5.2)
Schizoaffective Disorder	2 (1.3)
Panic Disorder	10 (6.5)
Other Anxiety Disorders	10 (6.5)
Somatiform Disorders	2 (1.3)
Eating Disorders	6 (3.9)
Insomnia	11 (7.2)
Adjustment Disorders	7 (4.6)
Premenstrual Dysphoric Disorder	1 (0.7)
None	5 (3.3)
Comorbidity with Major Depressive Disorder, currently met for a Major Depressive Episode	
Social Anxiety Disorder	1 (0.7)

off point (threshold) of summary score of PHQ-9 ≥ 10 , which has been recommended most consistently as the cut-off point [3,4,6]. Both thresholds had satisfactory sensitivity and negative predictive value: the summary score threshold (PHQ ≥ 10) had slightly higher sensitivity (0.94) and a more negative predictive value (0.94) than the diagnostic algorithmic threshold did, although the former had slightly lower specificity and positive predictive value than the latter did.

Among 73 positive cases for the PHQ-9 diagnostic algorithmic threshold, 34 cases (46.6 %) were falsely positive for having a current major depressive episode, as inferred from low specificity and positive predictive value. The *DSM-IV-TR* diagnoses of false-positive cases were schizophrenia in 4 cases, panic disorder in 4 cases, adjustment disorder in 5 cases, eating disorders in 5 cases, dementia in 2 cases, and insomnia in 5 cases. The false-positive patients showed high PHQ-9 summary scores comparable to those of the true-positive patients (data not shown) and exhibited depressed mood, but their depressive symptoms were not continuous during more than 2 weeks, as defined in *DSM-IV-TR*. Only 2 patients were diagnosed with other depressive disorders (minor depressive disorder and dysthymia).

False-negative cases were far fewer than false-positive cases. This tendency was particularly pronounced for the threshold of the summary score of PHQ-9 ≥ 10

Validity of PHQ-9 for the screening for current major depressive episode with major depressive disorder

Table 2 (right half) presents operating characteristics of PHQ-9 for “current major depressive episode with major depressive disorder” for the PHQ-9 diagnostic algorithmic threshold and for the cut-off point (threshold) of summary score of PHQ-9 ≥ 10 . Both thresholds had satisfactory sensitivity and negative predictive value: the summary score threshold (PHQ ≥ 10) had slightly better sensitivity (0.93) and negative predictive value (0.94) than the diagnostic algorithmic threshold did, although the former had slightly lower specificity and positive predictive value than the latter did.

Compared with false-positive cases for the PHQ-9 diagnostic algorithmic threshold against “current major depressive episode”, the number of those against “current major depressive episode with major depressive disorder” increased by 7 patients. That is to say, the positive predictive value was decreased more than that for “current major depressive episode”. All of these additional false-positive cases were bipolar disorder patients: 1 bipolar I and 6 bipolar II patients. As expected, bipolar depressed patients were positive for PHQ-9 because depressive symptoms of major depressive disorder and bipolar disorder show only subtle differences [7,12].

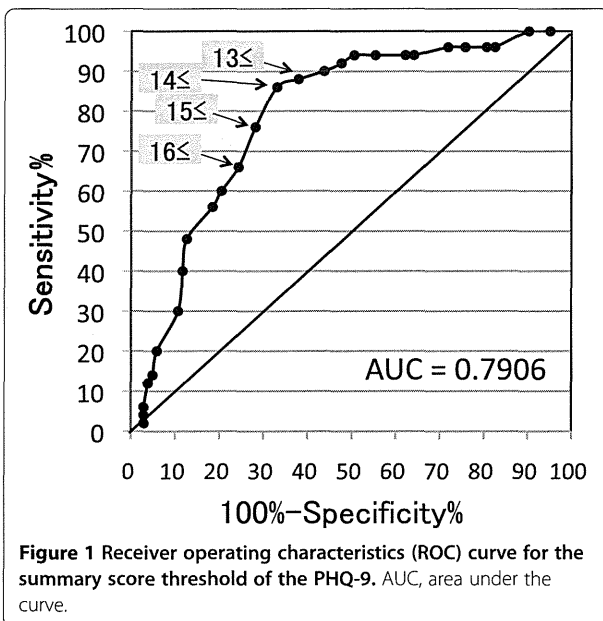
Table 2 Sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios, and overall accuracy for the diagnostic algorithm and the summary score (≥ 10) of the PHQ-9

	Diagnosis of current major depressive episode (MDE)		Diagnosis of MDE with major depressive disorder	
	Diagnostic algorithm of PHQ-9	Summary score (≥ 10) of the PHQ-9	Diagnostic algorithm of PHQ-9	Summary score (≥ 10) of the PHQ-9
True positive, n (%)	39 (25)	47 (31)	32 (21)	38 (25)
False positive, n (%)	34 (22)	52 (34)	41 (27)	61 (40)
False negative, n (%)	11 (7)	3 (2)	9 (6)	3 (2)
True negative, n (%)	69 (45)	51 (33)	71 (46)	51 (33)
Sensitivity (95%CI)	0.78 (0.64–0.88)	0.94 (0.84–0.99)	0.78 (0.62–0.89)	0.93 (0.80–0.98)
Specificity (95%CI)	0.67 (0.57–0.76)	0.50 (0.40–0.60)	0.63 (0.54–0.72)	0.46 (0.36–0.55)
Positive predictive value (95%CI)	0.53 (0.41–0.65)	0.47 (0.37–0.58)	0.44 (0.32–0.56)	0.38 (0.29–0.49)
Negative predictive value (95%CI)	0.86 (0.77–0.93)	0.94 (0.85–0.99)	0.89 (0.80–0.95)	0.94 (0.85–0.99)
Positive likelihood ratio (95%CI)	2.36 (1.76–2.99)	1.86 (1.56–2.02)	2.13 (1.59–2.63)	1.70 (1.41–1.85)
Negative likelihood ratio (95%CI)	0.33 (0.19–0.53)	0.12 (0.04–0.33)	0.35 (0.19–0.59)	0.16 (0.05–0.43)
Overall accuracy (95%CI)	0.71 (0.64–0.76)	0.64 (0.58–0.67)	0.67 (0.61–0.72)	0.58 (0.53–0.61)

95%CI, 95% confidence intervals; MDE, major depressive episode.

Receiver Operating Characteristics (ROC) analysis of the cut-off point (threshold) of summary score of PHQ-9 against “current major depressive episode”

Figure 1 displays the ROC curve of the cut-off point (threshold) of the summary score of PHQ-9 against “current major depressive episode”. The area under the curve (AUC) of 0.79 shows that the PHQ-9 summary scores can discriminate moderately between subjects with and without current major depressive episode [11]. Maximal discrimination between the presence and absence of “current major depressive episode” is achieved at the cut-off point which has the highest sum of sensitivity and specificity, as indicated by the “Youden index”.



The optimal cut-off was 13/14 (sensitivity 0.86, specificity 0.67) in the setting of a psychiatric specialty clinic and higher than that (9/10) recommended in primary care facilities and other specialty clinics [3,4,6]. Screening purposes demand high sensitivity and negative predictive value. Such was the case for a cut-off score of 11/12 or lower (Table 3). For diagnostic purposes, high specificity and positive predictive value are necessary. However, this condition was not reached with any cut-off score because specificity was higher than 90 % with a cut-off score of 21/22 or higher, but positive predictive values of all cut-off scores were lower than 65 %.

Correlation between PHQ-9 summary scores and HDRS, MADRS, or GAF

In patients with major depressive disorder, PHQ-9 summary scores were positively correlated with the severity measure of the depressive symptoms, HDRS and MADRS scores, but were negatively correlated with the overall levels of psychological, social, and occupational functioning, GAF scores [HDRS, $r = 0.55$, 95%CI (0.30, 0.73), $P = 0.0002$, $N = 42$; MADRS, $r = 0.68$, 95%CI (0.45, 0.82), $P < 0.0001$, $N = 36$; GAF, $r = -0.59$, 95%CI (-0.78, -0.30), $P = 0.0005$, $N = 31$].

Discussion

This study revealed that PHQ-9 has high sensitivity and a high negative predictive value in the setting of a clinic specializing in psychiatry as well as in primary care facilities and other specialty clinics [1,3,4,6] so that PHQ-9 is useful for screening purposes for the presence of a current major depressive episode. However, low specificity and a low positive predictive value in the setting of a clinic specializing in psychiatry do not support the use

Table 3 Sensitivity, specificity, positive and negative predictive values (PPV and NPV) at different cut-off scores for PHQ-9 summary scores against "current major depressive episode"

Cut-off	Sensitivity	Specificity	PPV	NPV
7/8	0.94	0.38	0.42	0.93
8/9	0.94	0.45	0.45	0.94
9/10	0.94	0.50	0.47	0.94
10/11	0.92	0.52	0.48	0.93
11/12	0.90	0.56	0.50	0.92
12/13	0.88	0.62	0.53	0.91
13/14 *	0.86	0.67	0.56	0.91
14/15	0.76	0.72	0.57	0.86
15/16	0.66	0.76	0.57	0.82
16/17	0.60	0.80	0.59	0.80
17/18	0.56	0.82	0.60	0.79
18/19	0.48	0.87	0.65	0.78
19/20	0.40	0.88	0.63	0.75
20/21	0.30	0.89	0.58	0.72
21/22	0.20	0.94	0.62	0.71
22/23	0.14	0.95	0.58	0.70

* Denotes the maximum sum of sensitivity and specificity.

of PHQ-9 for diagnostic purposes in contrast to the setting of primary care facilities and other specialty clinics, for which high specificity and a high positive predictive value were reported [1,3,4,6,8]. These findings were shown similarly for both the diagnostic algorithmic threshold and the summary score threshold (PHQ ≥ 10), which were recommended for diagnostic and screening purposes by earlier studies [1,6].

Because of low specificity in both the diagnostic algorithmic threshold and the summary score threshold (PHQ ≥ 10), false-positive cases should be noted in the use of PHQ-9. For a current major depressive episode, schizophrenia, panic disorder, adjustment disorder, eating disorders, dementia, and insomnia were diagnoses of false-positive cases. Compared with patients visiting primary care and other specialty clinics, more patients with various psychiatric disorders showing depressed mood and other various symptoms visit clinics specializing in psychiatry, which engenders more false-positive results.

PHQ-9 is used primarily for screening for the presence of major depressive episode but not major depressive disorder because the *DSM-IV-TR* diagnosis of major depressive disorder demands several exclusion criteria such as the absence of manic or hypomanic episode, but PHQ-9 does not include such exclusion items [4]. A significant number of bipolar disorder patients are invariably misdiagnosed with major depressive disorder by PHQ-9 because a major depressive episode is part of

bipolar disorder if one uses PHQ-9 for the screening for major depressive disorder. Originally, Kroenke et al. noted that before making a final diagnosis, the clinician is expected to rule out physical causes of depression, normal bereavement, and history of a manic episode [6]. Especially for a psychiatric specialty clinic, where bipolar disorder is much more prevalent, Kroenke's notion must be considered. For this reason, we compared the operational characteristics of PHQ-9 against "current major depressive episode" and "current major depressive episode with major depressive disorder". Our analysis of the validity of PHQ-9 for the screening for current major depressive episode with major depressive disorder (Table 2) revealed that the positive predictive value decreased by about 10 % compared with the screening for current major depressive episode. The underdiagnosis of bipolar disorder by PHQ-9 was a main reason for the increased false positives. Further diagnostic workup for past manic or hypomanic episodes or the combination of other screening tools for these episodes can resolve this major disadvantage of PHQ-9.

One might expect that the diagnostic algorithm threshold has better specificity than that of the summary score threshold because the diagnostic algorithm closely mimics the *DSM-IV-TR* diagnosis criteria of major depressive episodes. Nevertheless, against our expectations, the results obtained in this study and previous studies [3,4] showed no marked difference between two thresholds in operational characteristics. The summary score threshold (PHQ ≥ 10) has slightly higher sensitivity and negative predictive value, but slightly lower specificity and positive predictive value than the diagnostic algorithm with no marked difference in this study and a previous study of primary care [4]. The ROC analysis of the cut-off point (threshold) of summary score of PHQ-9 against "current major depressive episode" showed that the optimal cut-off was 13/14, which showed 0.86 of sensitivity and 0.67 of specificity comparable to those of the diagnostic algorithm, in the setting of a psychiatric specialty clinic. Table 3 shows that high specificity (>90 %) was reached with a cut-off score of 21/22 or higher, but a high positive predictive value was not reached with any cut-off score. The salient implication is that PHQ-9 used in a psychiatric specialty clinic might be suitable for screening purposes with the optimal cut-off of 13/14 of the summary scores for major depressive episode, but not for diagnostic purposes. The summary score threshold with different cut-off points for specific purposes might be preferred to the diagnostic algorithm.

Summary scores of PHQ-9 in patients with major depressive disorder were moderately correlated with severity measures of the depressive symptoms, HDRS and MADRS scores, positively, and with the overall levels of psychological, social, and occupational functioning, GAF

scores, negatively. Consistent with our results, an earlier report of primary care described that PHQ-9 scores were correlated linearly with measures of quality of life, self-reported disability days, clinical visits, and self-reported difficulties related to symptoms [6]. In depressive disorder patients of primary care facilities, PHQ-9 scores were correlated moderately with the HDRS (17 items) scores [13]. In this study, the correlation of PHQ-9 scores was highest with the MADRS, which is related to the core concept of depression and which showed about twice the precision in estimating depression as the HDRS (17 items) showed for the average severity of depression [14]. Therefore, PHQ-9 scores might reflect the core symptoms of major depressive disorder, as inferred from items corresponding to the *DSM-IV* criteria items. To date, the correlation of PHQ-9 scores with the standard rating scales of depression, especially the MADRS, has not been reported in psychiatric patients. In addition to screening, PHQ-9 might be useful for measuring the severity of major depressive disorder.

All subjects in this study were psychiatric patients of a university hospital that provides primary and secondary services in Japan. These patients might have more complicated backgrounds than patients in other psychiatric clinics. Accordingly, these findings might not be generally applicable to other populations, which constitutes one limitation of this study.

Conclusions

PHQ-9, a brief questionnaire, reportedly has excellent reliability and validity for screening of major depressive episode and for measuring the severity of depressive symptoms [2]. It has been used in primary care and medical specialty clinics. Results of this study suggest that PHQ-9 is useful also in clinics specializing in psychiatry. This report is the first of a study assessing the validity of the use of PHQ-9 in such clinics. However, caution is necessary for the use of PHQ-9 in psychiatry because PHQ-9 is useful only for screening purposes for "current major depressive episode" as a result of its low positive predictive value. The cut-off point 13/14 of PHQ-9 summary scores, which is higher than that (≥ 10) recommended in primary care, was optimal [6].

Competing interests

The authors report no financial or other relationship that is relevant to the subject of this article. TK has received honoraria from GlaxoSmithKline, Astellas, and Eli Lilly, has received research/grant support from Astellas and GlaxoSmithKline, and is a member of the advisory boards of GlaxoSmithKline and Mitsubishi Tanabe Pharma. The other authors declare that they have no actual or potential conflict of interest.

Authors' contributions

TI designed the study and wrote the protocol. All other authors followed up and evaluated patients, and checked the protocol and discussion. All authors contributed to and approved the final manuscript.

Received: 6 November 2011 Accepted: 3 July 2012
Published: 3 July 2012

References

1. Spitzer RL, Kroenke K, Williams JB, PHQ Primary Care Study Group: Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* 1999, **282**:1737-1744.
2. Furukawa TA: Assessment of mood: guides for clinicians. *J Psychosom Res* 2010, **68**:581-589.
3. Gilbody S, Richards D, Brealey S, Hewitt C: Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007, **22**:1596-1602.
4. Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC: Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *Gen Hosp Psychiatry* 2007, **29**:388-395.
5. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K: Monitoring depression treatment outcomes with the patient health questionnaire-9. *Med Care* 2004, **42**:1194-1201.
6. Kroenke K, Spitzer RL, Williams JB: The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001, **16**:606-613.
7. American Psychiatric Association: *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. 4th edition. Washington, DC: American Psychiatric Association; 2000.
8. Muramatsu K, Miyaoka H, Kamijima K, Muramatsu Y, Yoshida M, Otsubo T, Gejyo F: The patient health questionnaire, Japanese version: validity according to the mini-international neuropsychiatric interview-plus. *Psychol Rep* 2007, **101**:952-960.
9. Montgomery SA, Åsberg M: A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979, **134**:382-389.
10. Williams JB: A structured interview guide for the Hamilton depression rating scale. *Arch Gen Psychiatry* 1988, **45**:742-747.
11. Akobeng AK: Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr* 2007, **96**:644-647.
12. Perlis RH, Brown E, Baker RW, Nierenberg AA: Clinical features of bipolar depression versus major depressive disorder in large multicenter trials. *Am J Psychiatry* 2006, **163**:225-231.
13. Wittkampf K, van Ravesteijn H, Baas K, van de Hoogen H, Schene A, Bindels P, Lucassen P, van de Lisdonk E, van Weert H: The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *Gen Hosp Psychiatry* 2009, **31**:451-459.
14. Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D, Woo A, Trivedi MH: The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *Eur Neuropsychopharmacol* 2006, **16**:601-611.

doi:10.1186/1471-244X-12-73

Cite this article as: Inoue et al.: Utility and limitations of PHQ-9 in a clinic specializing in psychiatric care. *BMC Psychiatry* 2012 **12**:73.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



