

Factorial structure and long-term stability of the Autonomy Preference Index

Stephane Morandi and Philippe Golay

Service of Community Psychiatry, Department of Psychiatry, University Hospital of Lausanne,
Switzerland

Maria Vazquez-Montes

UK Nuffield Department of Primary Care Health Sciences, University of Oxford, UK
Centre for Statistics in Medicine, University of Oxford, UK

Jorun Rugkåsa

Health Services Research Unit, Akershus University Hospital, Lørenskog; Norway
Social Psychiatry Group, Department of Psychiatry, University of Oxford, UK

Andrew Molodynski, Ksenija Yeeles and Tom Burns

Social Psychiatry Group, Department of Psychiatry, University of Oxford; UK

Author note

Correspondence concerning this article should be addressed to Stephane Morandi, Service of Community Psychiatry, Department of Psychiatry, Place Chauderon 18, 1003 Lausanne, Switzerland. E-mail: Stephane.Morandi@chuv.ch

Funding information

National Institut of Health Research (UK). Award number: RP-PG-0606-1006. Recipient: Tom Burns, DSc

Factorial structure and long-term stability of the Autonomy Preference Index

Abstract

The Autonomy Preference Index scale (API) has been designed to measure patient preference for two dimensions of autonomy: their desire to take part in making medical decisions (Decision Making; DM) and their desire to be informed about their illness and the treatment (Information Seeking; IS). The DM dimension is measured by 6 general items together with 9 items related to three clinical vignettes (3x3 items). The IS dimension is measured by 8 items. While the API is widely used, a review of literature has identified several inconsistencies in the way it is scored. The first aim of this study was to determine the best scoring structure of the API on the basis of validity and reliability evidence. The second aim was to investigate the long-term stability of API scores. 285 patients with a diagnosis of psychosis were assessed as they were about to be discharged from involuntary psychiatric hospitalisation and they were re-assessed after six and twelve months. Confirmatory Factor Analysis (CFA) revealed that a three-factor solution was most adequate and that two distinct DM subscales should be preferred to one total DM score. While internal consistency estimates of the three subscales were good, the long-term stability of API scores was only modest. Multigroup-CFA revealed scalar invariance indicating API scores kept the same meaning longitudinally. In conclusion, a three-factor structure seemed to be most adequate for the API scale. Long-term stability estimates suggested that clinicians should regularly assess patients' preferences for autonomy because API scores fluctuate over time.

Keywords: Autonomy Preference Index; API; patient's autonomy; decision making; information seeking; factor analysis; long-term stability; reliability.

Introduction

Patient-centered care has been a growing concern preoccupation among health providers and politicians for several decades (Scholl, Zill, Haerter, & Dirmaier, 2014). This model of care has been shown to improve patient health and to increase the efficiency of care (Epstein & Street, 2007; Stewart et al., 2000). A prominent dimension of such patient-centeredness is their involvement in care through shared decision making (Barry & Edgman-Levitan, 2012). Engaging patients in the decision making about their health and helping them in making informed choices have been identified as key elements of the shared decision making approach (Scholl et al., 2014). Assessing how the healthcare which is provided matches the patients' preferences for decision making and information seeking is important, as it has been shown to affect both satisfaction and well-being (Kiesler & Auerbach, 2006). The Autonomy Preference Index (API) is one of the most commonly used instruments in measuring patient preferences for autonomy (Chewning et al., 2012; Puschner et al., 2013).

The API was developed by clinicians in the United States for use by general medical patients (Ende, Kazis, Ash, & Moskowitz, 1989). The instrument consists of two scales: a 15 items scale rating preferences in decision making (DM; e.g. "You should feel free to make decisions about everyday medical problems.") and an eight items scale rating information seeking (IS; e.g. "Information about your illness is as important to you as treatment."). The DM scale includes six general items and nine items related to one of three clinical vignettes, generating three specific items each. Each of the general DM and IS items is rated using a five-point Likert scale, with response choices ranging from "strongly disagree" to "strongly agree". DM is also assessed in relation to illness severity using three clinical vignettes of specific diseases, of increasing severity. Patients are asked to consider their preference for who should make decision at these different levels of illness severity. The three items for

each vignette are also rated using a five-point Likert scale. Response options range from “you alone” to “the doctor alone”. DM and IS scales had satisfactory internal consistency and two-week test-retest reliability.

Ende and colleagues (1989) also reported an exploratory factor analysis using orthogonal rotation that supported the clustering of items on DM as one scale and items on IS a second one. The vignettes were also analyzed as separate subscales. Unfortunately, no further details were provided about this analysis. For investigating the validity of test score interpretation, the association among API scores and other variables were also verified. The DM scale score was correlated with an item related to the patients’ desire to have control over their medical care ($r=0.54$). Finally, the authors confirmed that a diabetes population scored significantly higher on the DM scale than the general population.

The API has been widely used to measure the degree to which patients want to play a role in medical decision-making and their desire to be informed about their treatment (Simon et al., 2010). However, there are several inconsistencies in how the API has been scored. Hamann and colleagues investigated the wishes of 122 patients suffering from schizophrenia to be involved in decisions affecting their treatments but administered only the six general items of the DM subscale (2005). Hill and Laugharne adapted the questions to make them relevant to psychiatric practice using vignettes of different severities of depression (2006). Total scores for DM and IS were obtained by adding the scores of each of the 15 DM items and each of the eight IS items. This adapted version of the API was tested on 205 Community Mental Health Team patients and 160 members of mental health users’ organisations in the UK. Although patients had a wide range of psychiatric diagnoses the DM and IS scales demonstrated adequate internal consistency. The concurrent validity of the DM scale score interpretation was established by correlation with an empirically related global item ($r=0.63$). To explore the preferences of 30 patients with severe mental illness, Adams and colleagues

administered the questionnaire including only the six general items of the DM subscale (2007).

The Factorial structure of the API has also been examined. Simon and colleagues (2010) validated a German version of the API on a sample of 1592 patients comprising those treated for depression in primary care (n=186), hospitalised in surgical and medicine services (n=811) and in an emergency department for minor trauma (n=595). Their confirmatory factor analysis (CFA) suggested that the API could be improved by the removal of several items. Short term reliability coefficients for both the original and the shortened scale were satisfactory. However, the authors encouraged further tests of both the original and adapted versions of the API. More recently, Bonfils and colleagues assessed the factor structure of the API in 293 people with severe mental illness (Bonfils, Adams, Mueser, Wright-Berryman, & Salyers, 2015). The results indicated that the two factors (DM and IS) model did not fit the data well. The authors concluded that two DM items and one IS item needed to be removed to improve the model. No short or long-term reliability evidence of the shortened scales was provided. Furthermore, in both studies, the DM items related to vignettes were not included.

The most recent papers on the factorial structure of the API suggested that the DM items related to vignettes could be omitted because this has been the tendency in past research. It is also suggested that two out of six items of the DM score and one of the IS score should be removed (Bonfils et al., 2015; Simon et al., 2010). These modifications make the measurement of DM and IS quicker and simpler, which is particularly an advantage in community mental health settings. However, they raise concerns about meeting reliability standards in psychological testing especially since the DM score would only be based on 4 items rather than 6 (without vignettes) or 15 (including the vignettes).

Given the lack of clear evidence about the optimal use of the API when applied in severe mental illness, the first aim of this study was to determine the best scoring structure on

the basis of validity and reliability evidence. CFA was used in order to determine whether items should be omitted and whether DM was best accounted by one total score (including both general items and items related to vignettes) or two separate scores. Internal consistency and model-based reliability estimates were also used to select the most adequate scoring method. The results ascertain whether items can be omitted without weakening the reliability of API scores.

The second aim was to investigate whether patients' preferences for autonomy represented a stable and enduring trait. The test retest reliability of the scores of the original scale was estimated with a small sample (N=50) and for a short interval of two weeks (Ende et al., 1989). Neither Hill and Laugharne (2006) nor Bonfils et al (2015) estimated test retest reliability. No data is available concerning measurement invariance over time. To our knowledge, it is the first time that the long-term stability of API scores has been investigated.

Method

Participants

Participants were the 285 patients with available API data at baseline, with a diagnosis of psychosis (85.3% suffering from a schizophrenia), getting ready to be discharged from psychiatric hospital and included in a randomised control trial that tested whether community treatment orders reduced readmissions compared to community voluntary care after a compulsory psychiatric hospitalisation (Burns et al., 2013). Interviews were conducted by research assistants at baseline and repeated at six and twelve months. Recruitment took place from November 2008 to February 2011 in 32 National Health Service Mental Health Trusts predominantly in the Midlands and southern England. Eligible patients were those in adult services (18-65 years of age) with a diagnosis of psychosis, considered

appropriate for community treatment orders and able to give informed consent. 68.1 % (n=194) were male. Mean age was 39.0 years (SD=11.4). Trial procedures, more detailed sociodemographics and clinical characteristics of participants were published by Burns and colleagues (2013). A subgroup analysis from the randomised control trial showed no significant differences between both arms in API scores from baseline to 12 months (Rugkasa et al., 2015). Ethical approval was granted by the Staffordshire National Health Service Research Ethics Committee (reference 08/H1204/131).

Instrument

We used the API adjusted to the mental health setting to assess DM and IS preferences regarding psychiatric care among participants (Hill & Laugharne, 2006). DM general items score and IS score were adjusted to range from 0 to 100. A score of 0 meant a complete lack of a patient's desire to take part in decision making or to be informed. A score of 100 defined the strongest possible patient's desire to take part in decision making or to receive information. The total scores of each vignette were adjusted from 0 to 10. The lowest score (0) corresponded to a patient's desire for the care provider to take complete control over decision making, the mid-range score (5) to a desire for decision making equally shared between the care provider and the patient and the highest score (10) to a patient's desire to take a complete control over decision making.

Statistical analyses

All reversed scored items were re-coded prior to data-analysis. For the CFA, data were treated as categorical ordinal and all the models were estimated using a robust weighted least squares estimator with adjustments for the mean and variance (WLSMV). Three models were estimated. A one-factor model representing a general "autonomy" latent construct was

first tested on the 23 API items. A two-factor model included a DM factor based on the first 6 general items and the 9 items related to vignettes and one IS factor based on the last 8 items was tested next. Finally, a three-factor model including two distinct DM factors (general items and items related to vignettes) and one IS factor was tested. To identify the scale of the latent factors, one factor loading was fixed to one for each latent variable. Each model was compared to a more restrictive alternative including one less factor with a robust chi-square test using the DIFFTEST procedure. Several indicators of model fit were used such as the root mean square error of approximation (RMSEA), the Comparison fit index (CFI), the Tucker–Lewis fit index (TLI) and the Standardized Root Mean Square Residual (SRMR) when available. To the best of our knowledge interpretation of global fit indexes in models with ordered categorical indicators is not as well established as it is with continuous indicators (Hu & Bentler, 1999). With continuous indicators, $RMSEA \leq 0.06$, $SRMR \leq 0.08$ and CFI and TLI ≥ 0.95 are interpreted as good fit while values of $RMSEA \leq 0.08$ and CFI/TLI ≥ 0.90 are often considered as indicating acceptable fit (Hu & Bentler, 1999). While simulation studies by Yu and Muthén suggest that these cut off values works reasonably well with categorical outcomes, the exact cut off scores may still not perfectly apply in the context of our study (Ching-Yun, 2002; B. O. Muthén, 1998-2004). The results from such Monte Carlo simulations have excellent internal validity but can be questionable outside of their reciprocal research situation neither of which corresponds to a 3-factor model with simple structure and 23 categorical ordinal Likert style items. For this reason, model comparisons will be based on robust chi-square testing only. Internal consistency and reliability of the three API subscores that were identified by means of the CFA were estimated by Cronbach's alpha and McDonald's model based Omega. While many publications only report alpha coefficients, numerous criticism of this index have been raised and alternative model based reliability estimates like McDonald's omega have been promoted (Canivez, in press; Nelson,

Canivez, & Watkins, 2013). The stability of API scores over periods of 6 and 12 months was also investigated. The relative test-retest reliability was estimated by both Pearson and intra-class correlation coefficient using a 2-way random-effects model and the absolute agreement definition (ICC (2,1)). Measurement invariance across time was also evaluated with multi-group CFA in order to verify whether the API and items scores kept the same meaning and level over time. Because for some items there was not the same number of categories in the data for all groups (0, 6 & 12 month) categorical methodology required maximum likelihood estimation (for which indicators of fit were not available). Thus invariance analyses were replicated treating data as continuous. With both approaches the same pattern of findings for the comparison between different levels of invariance could be obtained. All statistical analyses were performed with the Mplus statistical package version 7.3 (L. K. Muthén & Muthén, 2012) and IBM SPSS version 22 (IBM Corp., 2013). Because all patients did not necessarily complete all three assessments, analyses were performed using all available data. In order to evaluate the robustness of our findings, all reliability analyses were replicated across multiple imputed (MI) datasets. MI was performed with 20 imputations using the fully conditional specification method and by using previously identified influential demographic variables as predictors. The same pattern of findings could be obtained with either the original data or the pooled results of the 20 imputed datasets.

Results

The API structure

The descriptive statistics for the 23 API items of the 285 participants with data available at the first assessment are detailed in supplementary table 1.

Both the one-factor model ($\chi^2 = 1986.117$, $df = 230$, RMSEA (90% CI) = 0.164 (0.157 – 0.170), CFI = 0.577, TLI = 0.535) and the two factor model ($\chi^2 = 708.617$, $df = 229$,

RMSEA (90% CI) = 0.086 (0.079 – 0.093), CFI = 0.884, TLI = 0.872) showed poor fit to the data. Only the three-factor model showed an acceptable fit to the data ($\chi^2 = 581.639$, $df = 227$, RMSEA (90% CI) = 0.074 (0.067 – 0.081), CFI = 0.915, TLI = 0.905). The results of the robust chi-square difference tests indicated that the three-factor solution should be preferred (1 factor against 2 factors: $\Delta\chi^2 = 166.23$, $\Delta df = 1$, $p < .0001$; 2 factors against 3 factors: $\Delta\chi^2 = 34.98$, $\Delta df = 15$, $p = .003$). Factor loadings of this model are presented in Figure 1. All loadings were statistically significant and of substantial magnitude which indicated that no item should be removed from the API subscales. The highest factor correlation was observed between the two DM factors ($r = 0.68$). However, on the basis of the comparison between the two and the three-factor model, these two factors could not be considered as indistinguishable.

In order to verify the reliability of the API subscales, Cronbach's alpha and McDonald's omega were estimated. Internal consistency for the DM general items score (Cronbach's $\alpha = .77$ & McDonald's $\omega = .83$) was satisfactory to good (Canivez, in press; George & Mallery, 2003; Reise, 2012). Estimates for the DM items related to vignettes were slightly higher (Cronbach's $\alpha = .86$ & McDonald's $\omega = .91$). Finally, internal consistency of the IS items also proved to be good (Cronbach's $\alpha = .80$ & McDonald's $\omega = .89$).

Long-term stability of the API

Long-term stability of the API subscale scores was estimated for 6 and 12 month duration between baseline and follow-up evaluations. Results are reported in Table 1. Pearson and intra-class correlation showed poor long-term stability of the two DM scales scores over 6 and 12 month periods and almost null stability of the IS scale score. While constructed on the basis of only six items, the first DM score was slightly more reliable than the vignette-based nine items score.

Measurement invariance over time

Three multi-group CFA models with increasing level of invariance were compared. The first level was configural invariance. All parameters (factor loadings & thresholds) were freely estimated in the three groups (baseline, 6 month & 12 month), whereas the factor means were fixed at zero in all groups. The second level of longitudinal invariance was metric invariance, in which equivalence constraints on factor loadings across groups were added. Finally scalar invariance had factor loadings and thresholds constrained to be equal across groups and factor means fixed at zero in one group and free in the other groups. Considering configural invariance over time, the fit was acceptable given the ordinal scoring of the items ($\chi^2 = 1624.148$, $df = 681$, RMSEA (90% CI) = 0.079 (0.074 – 0.083), CFI = 0.830, TLI = 0.811, SRMR = 0.075). The fit of the metric and the scalar model were very similar ($\chi^2 = 1671.791$, $df = 721$, RMSEA (90% CI) = 0.077 (0.072 – 0.081), CFI = 0.829, TLI = 0.820, SRMR = 0.081 respectively $\chi^2 = 1722.277$, $df = 761$, RMSEA (90% CI) = 0.075 (0.070 – 0.080), CFI = 0.827, TLI = 0.827, SRMR = 0.084). Comparison between levels of invariance indicated that the most restrictive scalar model should be preferred because additional constraints did not significantly impair model fit (Metric against Configural : $\Delta\chi^2 = 47.643$, $\Delta df = 40$, $p = .190$; Scalar against Configural: $\Delta\chi^2 = 98.129$, $\Delta df = 80$, $p = .124$). This suggested that the meaning and level of API scores and items remained the same and did not suffer from systematic changes.

Discussion

Model comparison revealed that a three-factor structure seemed to be most adequate for the API scale. All item loadings were supported (i.e. no item should be removed from the API subscales). This finding is not in line with those of previous studies that discussed removing two DM items and one IS item (Bonfils et al., 2015; Simon et al., 2010). A likely

reason for the difference in findings is that the categorical ordinal nature of the Likert-type items was not taken into account by either Simon and colleagues (2010) or Bonfils and colleagues (2015) possibly resulting in attenuated correlations and underestimated factor loadings. Another possibility is that the versions of the API were not exactly the same. Simon and colleagues validated a German version of the API (2010), Bonfils and colleagues analysed the factor structure of the original API (2015) and we used the API adjusted to mental health setting (Hill & Laugharne, 2006). In DM items 4 and 6 of this latter version, terms “everyday medical problems”, respectively “check up”, were replaced by “own mental health care” and “mental health team”. The specific wording of the version adapted to psychiatric conditions may have been more relevant to people suffering a severe mental illness making it easier to decide whether they wanted to make decision or not in such situations. Finally, the difference of profile between the samples, in particular that our sample was about to be discharged from psychiatric hospital where they had been treated involuntarily when first assessed, may have influenced the results.

Both DM subscales represented two empirically related but distinct dimensions of a patient’s desire of autonomy. Ende and colleagues (1989) reported that the patient’s desire to make decisions was influenced by the degree of illness presented in the vignette. Patients’ preferences for taking an active part in decision making declined when the vignette was for a more severe disorder. Patients may have identified more with the DM general items than with the specific items in the vignettes, as these did not directly reflect their own experiences. The differences between the two factors could also simply be accounted by method variance. From the standpoint of concurrent validity, the correlation between the two DM factors is reassuringly high. The lower correlation between DM and IS factor confirms that patients’ preferences for making decision and their desire to be informed are separate and distinct elements of shared decision making (Scholl et al., 2014).

On a practical level, the results of this study suggest that vignette items should be handled cautiously in clinical setting and in future research. First, the CFA indicated that incorporating DM general items and vignette-related items to compute a single DM score was not supported by the data. Second, the two DM factors, while heavily correlated, were not identical. In the absence of external criterion measures, it would be unsafe to draw definitive conclusions on whether a DM factor related to vignettes is necessary or not to assess DM in this patient group. We would recommend two separate DM scores be computed to verify that the same conclusions are reached from the general and the vignettes items.

While internal consistency and reliability estimates of the three subscales were shown to be good, the long-term stability of API scores was only modest. Cronbach's alpha and McDonald's omega can be interpreted as estimates of very short-term reliability. In this study, they indicated a 'satisfactory to good' stability for the three subscales scores. The poor long-term stability of the API could be not explained by lack of longitudinal invariance of items over time. Indeed invariance analysis revealed that the same constructs were being measured across time. This pattern of findings leads us to believe that the poor long-term reliability of API scores may be better explained by the inherent variability of the autonomy construct rather than by poor item and scale construction. The latter would very likely have prevented such stringent levels of invariance. Legal status changes over time (coercive vs. voluntary treatment) could also account for the instability of scores during follow-up. Other factors such as the patient's attitude toward medical treatment or employment are associated with a higher desire for decision making and these may change over time and influence the rating at subsequent time points (Hamann et al., 2005; Hill & Laugharne, 2006). Service setting and provider characteristics may also influence the scores. The long-term stability of API scores has not previously been investigated, so an important observation from our study is that individual differences in DM and IS do not appear to be trait-like.

This study has several limitations. The API has often been adapted or modified and psychometric evaluations of the different versions have not been systematically conducted. We decided to use the API adjusted to mental health setting (Hill & Laugharne, 2006), because we considered it the most appropriate to assess DM and IS preferences for psychiatric care among our participants. Although the vignettes were related to depression, we did not modify them for our psychosis population, because we believed that participants could easily understand the scenarios.

Since its construction, the API has been extensively used. However, scoring has been inconsistent. When used to assess desire for autonomy, the DM general items and the vignette items scores should be computed separately. Future studies should be designed to determine whether the clinical vignettes could reasonably be omitted when assessing DM using the API. Clinicians should also regularly assess patients' preferences for autonomy as it may fluctuate over time. Patients may wish to be strongly involved in decision making and information seeking during some phases of their illness but not in others.

References

- Adams, J. R., Drake, R. E., & Wolford, G. L. (2007). Shared decision-making preferences of people with severe mental illness. *Psychiatric Services*, 58(9), 1219-1221.
- Barry, M. J., & Edgman-Levitan, S. (2012). Shared Decision Making - The Pinnacle of Patient-Centered Care. *New England Journal of Medicine*, 366(9), 780-781.
- Bonfils, K. A., Adams, E. L., Mueser, K. T., Wright-Berryman, J. L., & Salyers, M. P. (2015). Factor structure of the autonomy preference index in people with severe mental illness. *Psychiatry Research*, 228(3), 526-530. doi: 10.1016/j.psychres.2015.06.004
- Burns, T., Rugkasa, J., Molodynski, A., Dawson, J., Yeeles, K., Vazquez-Montes, M., . . . Priebe, S. (2013). Community treatment orders for patients with psychosis (OCTET): a randomised controlled trial. *Lancet*, 381(9878), 1627-1633. doi: 10.1016/s0140-6736(13)60107-5
- Canivez, G. L. (in press). Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation. In K. S. C. DiStefano (Ed.), *Principles and methods of test construction: Standards and recent advancements*. Gottingen: Hogrefe Publishers.
- Chewning, B., Bylund, C. L., Shah, B., Arora, N. K., Gueguen, J. A., & Makoul, G. (2012). Patient preferences for shared decisions: A systematic review. *Patient Education and Counseling*, 86(1), 9-18.
- Ching-Yun, Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. (Doctor of Philosophy in Education), University of California, Los Angeles.

- Ende, J., Kazis, L., Ash, A., & Moskowitz, M. A. (1989). Measuring patients' desire for autonomy: Decision making and information-seeking preferences among medical patients. *Journal of General Internal Medicine*, 4(1), 23-30.
- Epstein, R. M., & Street, R. J. (2007). Patient-centered communication in cancer care: Promoting healing and reducing suffering. Bethesda, MD: National Cancer Institute.
- George, D., & Mallery, P. (2003). *Using SPSS for Windows step by step: a simple guide and reference*. Boston: Allyn and Bacon.
- Hamann, J., Coden, R., Leucht, S., Busch, R., & Kissling, W. (2005). Do patients with schizophrenia wish to be involved in decisions about their medical treatment? *American Journal of Psychiatry*, 162(12), 2382-2384.
- Hill, S. A., & Laugharne, R. (2006). Decision making and information seeking preferences among psychiatric patients. *Journal of Mental Health*, 15(1), 75-84.
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling-a Multidisciplinary Journal*, 6(1), 1-55. doi: 10.1080/10705519909540118
- IBM Corp. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY.
- Kiesler, D. J., & Auerbach, S. M. (2006). Optimal matches of patient preferences for information, decision-making and interpersonal behavior: Evidence, models and interventions. *Patient Education and Counseling*, 61(3), 319-341. doi: 10.1016/j.pec.2005.08.002
- Muthén, B. O. (1998-2004). *Mplus Technical Appendices*. Los Angeles: CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide. Seventh Edition*. Los Angeles: CA: Muthén & Muthén.

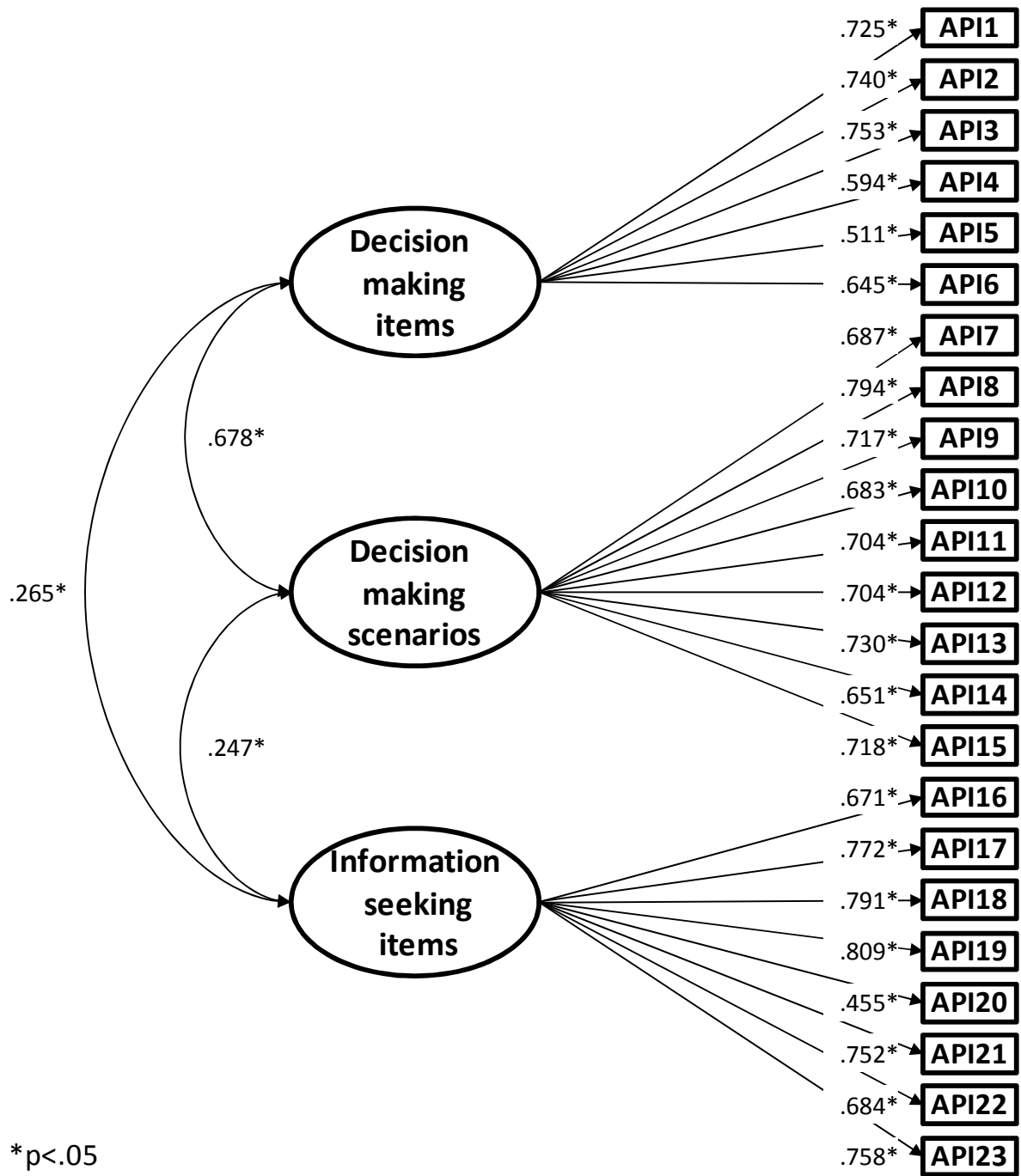
- Nelson, J. M., Canivez, G. L., & Watkins, M. W. (2013). Structural and Incremental Validity of the Wechsler Adult Intelligence Scale-Fourth Edition With a Clinical Sample. *Psychological Assessment*, 25(2), 618-630. doi: 10.1037/a0032086
- Puschner, B., Neumann, P., Jordan, H., Slade, M., Fiorillo, A., Giacco, D., . . . Grp, C. S. (2013). Development and psychometric properties of a five-language multiperspective instrument to assess clinical decision making style in the treatment of people with severe mental illness (CDMS). *BMC Psychiatry*, 13. doi: 10.1186/1471-244x-13-48
- Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, 47(5), 667-696. doi: 10.1080/00273171.2012.715555
- Rugkasa, J., Molodynski, A., Yeeles, K., Montes, M. V., Visser, C., Burns, T., & Grp, O. (2015). Community treatment orders: clinical and social outcomes, and a subgroup analysis from the OCTET RCT. *Acta Psychiatrica Scandinavica*, 131(5), 321-329. doi: 10.1111/acps.12373
- Scholl, I., Zill, J. M., Haerter, M., & Dirmaier, J. (2014). An Integrative Model of Patient-Centeredness - A Systematic Review and Concept Analysis. *Plos One*, 9(9). doi: 10.1371/journal.pone.0107828
- Simon, D., Kriston, L., Loh, A., Spies, C., Scheibler, F., Wills, C., & Härter, M. (2010). Confirmatory factor analysis and recommendations for improvement of the Autonomy-Preference-Index (API). *Health Expectations*, 13(3), 234-243.
- Stewart, M., Brown, J. B., Donner, A., McWhinney, I. R., Oates, J., Weston, W. W., & Jordan, J. (2000). The impact of patient-centered care on outcomes. *Journal of Family Practice*, 49(9), 796-804.

Table 1. Stability of API scores over time: comparisons of baseline and follow-up scores.

	N	Pearson's r	95% C.I. Pearson's r	ICC (2,1)	95% C.I. ICC (2,1)
6 months					
<i>Decision making preference general items</i>	150	.49*	.36 – .61	.48*	.35 – .59
<i>Decision making preference related to vignettes</i>	124	.44*	.30 – .58	.44*	.28 – .57
<i>Information seeking preference items</i>	135	.15	-.02 – .28	.14	-.03 – .30
12 months					
<i>Decision making preference general items</i>	181	.44*	.32 – .56	.43*	.30 – .54
<i>Decision making preference related to vignettes</i>	146	.31*	.16 – .46	.31*	.15 – .45
<i>Information seeking preference items</i>	169	.20*	.05 – .32	.19*	.05 – .33

Note. * = $p < .05$. C.I = Confidence Interval. ICC (2,1) = intra-class correlation coefficient using a 2-way random-effects model and the absolute agreement definition.

Figure 1. *Three-factor model for the API scale.*



Supplementary Table 1. *Univariate proportions and counts for the 23 API items at baseline (N = 285).*

	Rating									
	Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
	%	Count	%	Count	%	Count	%	Count	%	Count
<i>Decision making preference general items</i>										
1. The important mental health decisions should be made by the psychiatrist, not the patient	7.1	20	35.3	100	19.1	54	28.3	80	10.2	29
2. Patients should go along with the psychiatrist's advice even if they disagree with it.	2.1	6	25.7	73	18.7	53	40.8	116	12.7	36
3. Hospitalised patients should not be making decisions about their own mental health care.	3.6	10	21.4	60	12.9	36	49.3	138	12.9	36
4. Patients should feel free to make decisions about their own mental health care.	1.8	5	12.4	35	12.7	36	56.9	161	16.3	46
5. If you were unwell, as your illness became worse, you would want the psychiatrist to take greater control.	9.5	27	55.5	157	14.1	40	15.2	43	5.7	16
6. You should decide how frequently you need to be seen by the mental health team	1.8	5	17.7	50	21.6	61	47.0	133	12.0	34
<i>Decision making items related to vignettes</i>										
	Clinician alone		Mostly the clinician		The clinician and you equally		Mostly me		Me alone	
7. Whether you are seen regularly by the mental health team.	8.4	22	11.9	31	28.4	74	18.8	49	32.6	85
8. Whether you are seen regularly by a psychiatrist.	14.3	37	15.4	40	23.6	61	19.3	50	27.4	71
9. Whether you should take an anti-depressant.	18.9	49	14.7	38	23.6	61	15.4	40	27.4	71
10. When you next have an appointment with a psychiatrist.	19.2	50	23.1	60	25.0	65	13.5	35	19.2	50
11. Whether you should have some sick leave from work.	22.0	57	21.2	55	23.2	60	15.4	40	18.1	47
12. Whether you should be treated by talking therapy e.g. counselling, psychotherapy.	19.3	50	14.3	37	30.9	80	15.8	41	19.7	51
13. How often the nurses should check on you to make sure that you are safe.	41.0	107	29.9	78	16.1	42	5.4	14	7.7	20
14. Whether you are well enough to cope with family or friends visiting you.	18.6	48	12.4	32	31.4	81	15.9	41	21.7	56
15. Whether you see a consultant psychiatrist on the day of admission.	36.5	93	21.2	54	19.2	49	10.2	26	12.9	33
<i>Information seeking preference scale</i>										
	Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
16. As you become more unwell you should be told more and more about your illness.	2.6	7	9.6	26	8.1	22	61.0	166	18.8	51
17. You should be kept informed about what is happening inside your body as a result of your illness.	1.1	3	4.1	11	4.1	11	63.0	170	27.8	75
18. Even if the news is bad, you should be well informed.	0.7	2	3.7	10	5.6	15	60.4	162	29.5	79
19. Your psychiatrist should explain the purpose of any investigations, e.g. blood tests.	0.4	1	1.5	4	4.5	12	63.6	171	30.1	81
20. You should be given information only when you ask for it.	4.8	13	21.8	59	10.7	29	43.5	118	19.2	52
21. It is important for you to know all the side effects of your medication.	1.5	4	2.6	7	3.3	9	54.8	148	37.8	102
22. Information about your illness is as important to you as treatment.	0.8	2	5.6	15	7.9	21	60.2	160	25.6	68
23. When there is more than one way to treat a problem, you should be told about all the options.	0.7	2	1.5	4	2.2	6	61.1	165	34.4	93