The Frankfurt Complaint Questionnaire for self-assessment of basic symptoms in the early detection of psychosis – factor structure, reliability and predictive validity

Martina Uttinger[1,2], Erich Studerus[1], Sarah Ittig[1], Ulrike Heitz[1], Frauke Schultze-Lutter[3,4], Anita Riecher-Rössler[1]

[1] Center for Gender Research and Early Detection, University of Basel Psychiatric Hospital, University of Basel, Basel, Switzerland

[2] Psychiatric Services Solothurn, Olten, Switzerland

[3] University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland

[4] Department of Psychiatry and Psychotherapy, Medical Faculty, Heinrich-Heine-University, Düsseldorf, Germany

Prof. Dr. med. Anita Riecher-Rössler

Anita.Riecher@upkbs.ch

Tel.: +41 61 325 81 61, Fax: +41 61 325 81 60

Word count of abstract: 198 (max. 200)

Word count of text body 4458 (max. 5000)

Running Head: Self-reported basic symptoms in early psychosis

Abstract

Objectives: Patients with schizophrenia often experience subtle disturbances in several domains of information processing – so-called basic symptoms (BS). BS are already present before onset of frank psychosis and can be assessed by interviews but also by the self-administered Frankfurt Complaint Questionnaire (FCQ). We investigated the factor structure, reliability and predictive validity for transition to psychosis of the FCQ, comparing previously proposed factor solutions containing one, two, four and ten factors.

Methods: Confirmatory factor analysis (CFA) was used in a sample of 117 at-risk mental state (ARMS) and 92 first-episode psychosis (FEP) participants of the Basel FePsy (Early Detection of Psychosis) study.

Results: Although all factor models fitted to the data, ~~depending on the method used,~~ the ~~one~~two- or ~~two~~four-factor solutions ~~with a reduced number of items~~ performed best among the models that used at least half of the FCQ items, suggesting the covariance between FCQ items is best explained by ~~one or two~~two to four underlying factors. No FCQ-~~factor~~ scale predicted transition to psychosis.

Conclusion: We could confirm a ~~one~~two- or ~~two~~four-factor structure of the FCQ in a sample of ARMS and FEP patients using CFA. Contrary to interview-assessed cognitive-perceptive BS, self-assessed BS do not seem to improve prediction of psychosis. This result reinforces reports of poor correspondence between interview- and questionnaire-assessed BS.

Key words: Basic symptoms, self-assessment, Frankfurt complaint questionnaire, confirmatory factor analysis, prediction of psychosis

1. Introduction

Patients with schizophrenia frequently experience not only overt positive and negative symptoms but also various subclinical disturbances in several domains of information processing. These often start already several years before the onset of frank psychosis (Riecher-Rössler et al., 2006; Riecher-Rössler et al., 2009). Following early works of Wilhelm Mayer-Gross (1932), Gerd Huber (1966) first systematically described these subjective disturbances based on in-depth clinical interviews with patients. He coined the term “basic symptoms” (BS), because, based on his pioneer works on dysplastic ventricles in schizophrenia, he assumed them to be the most immediate psychopathological expression of the neurobiological changes underlying the development of psychosis (Huber and Süllwold, 1986). BS are subjectively experienced by patients as deviations from their “normal” self and as a loss of control over different mental functions (Schultze-Lutter, 2009; Schultze-Lutter et al., 2016; Schultze-Lutter et al., 2012b). The 142 BS initially described in the interviewer rated Bonn Scale for the Assessment of Basic Symptoms (BSABS; Gross, 1987) can be present in various numbers and combinations and include self-experienced disturbances in thinking, attention, speech, (body) perception, motor actions, stress tolerance, drive, affective responses as well as central-vegetative processes.

Developed from the BSABS as economic, age-appropriate and severity-sensitive interview-assessments of BS, the Schizophrenia Proneness Instrument, Adult (SPI-A; Schultze-Lutter et al., 2007) and Child and Youth version (SPI-CY; Schultze-Lutter et al., 2010) are the most~~ly~~ used instrument today in the clinic and in research, especially in the early detection of psychosis (Schultze-Lutter et al., 2012b). From interview-assessments of BS, two partially overlapping psychosis-risk criteria were developed, the Cognitive-Perceptive Basic Symptoms (COPER; Klosterkötter et al., 2001) and the Cognitive Disturbances (COGDIS; Schultze-Lutter et al., 2007; Schultze-Lutter et al., 2006). Studies indicated that the combined presence of symptomatic Ultra High Risk Criteria (UHR; Yung and McGorry, 1996) and BS criteria almost triples the risk of developing psychosis when compared to the exclusive presence of either UHR (ultra-high risk) or BS criteria (Michel et

al., 2014; Schultze-Lutter et al., 2014; Schultze-Lutter et al., 2012a). Supported by other findings on a psychosis-risk enhancement by self-experienced cognitive disturbances in UHR samples (Nelson et al., 2012; Parnas et al., 2011; Yung et al., 2005) it was argued that cognitive BS in particular might capture the cognitive core dimension of psychoses that is hardly addressed by UHR criteria (Michel et al., 2014; Schultze-Lutter et al., 2014). All studies so far have been using interviewer rated scales such as the Examination of Anomalous Self-Experience instrument (EASE) or the BSABS. However, as the interview-assessment of BS requires intensive training and is rather time-consuming, and as BS are self-experienced by nature, it is of interest if the same kind of risk-enhancement could be detected in UHR-samples by a self-report questionnaire.

In close collaboration with Huber and parallel to the development of the BSABS, Lilo Süllwold (1991) designed a questionnaire-based assessment of BS, the German Frankfurt Complaint Questionnaire (FCQ; Süllwold, 1991) from example statements of the BSABS. The 98 dichotomous FCQ items partially overlap with each other and refer to identical underlying BS, e.g., a disturbance of receptive speech (BSABS-item C.1.6) is targeted by as much as seven different FCQ items (no. 37, 40, 69, 82, 90, 93 and 94; Michel et al., 2016a). The FCQ has become the main BS-questionnaire, as it was used in several different cultural contexts and was translated into seven languages (Michel et al., 2016b; Moritz et al., 2000b). Furthermore, it is still the most extensive self-report BS questionnaire used. Yet, despite its development based on the BSABS, the FCQ items hardly capture the same phenomena as interview-assessed BS (Mass et al., 1995; Michel et al., 2016a).

Several studies of the factor structure of the FCQ yielded inconclusive results. Apart from the original ten theory-based factors (Süllwold, 1991), various data-based factor solutions including one (Loas et al., 2002; Yon et al., 2008), two (Mass et al., 1997; Mass et al., 1995) and four factors (Süllwold, 1991) of different total numbers of FCQ-items have been suggested based on different samples (Table 1). So far, the model fits of these factor solutions have not been formally tested and compared using confirmatory factor analysis (CFA). Furthermore, in all previous factor analyses, binary FCQ items were treated as

continuous variables, which can severely bias the resulting factor structure (Kubinger, 2003). Specifically, because the exploratory factor analysis (EFA) of Süllwold (1991) was performed on Pearson instead of the more appropriate tetrachoric correlation matrices, items were clustered together not only according to their content but also according to their endorsement probabilities. This may have led to the relatively higher number of factors by the emergence of "pseudo-factors" that are artifacts of item difficulty or extremeness. Thus, this study aimed to close both gaps in knowledge regarding the questionnaire-assessed BS: the model fit and reliability of the different proposed FCQ-factor solutions and the potential property of predicting the risk of transition to psychosis.

First, we analyzed the psychometric properties (factor fit, internal consistency and homogeneity) of the FCQ and its proposed factors using confirmatory factor analysis (CFA) and categorical item methodology in a sample of at-risk mental state (ARMS) and first-episode psychosis (FEP) patients. We hypothesized that the proposed factor models would provide a good fit to our data. Based on the so far strong evidence for a uni- or at most two-dimensionality of the FCQ, we assumed that a one- or two-factor model would fit best. In addition, we expected good internal consistency but not necessarily homogeneity for all proposed factors.

Second, we tested for the first time whether any of the previously proposed FCQ factors was predictive for a later transition to psychosis in ARMS patients. We expected that the total scores and in particular factors mainly consisting of the 34 cognitive and perceptive items most similar to the basic symptoms included in COGDIS and COPER (Table 1) would to be most predictive for psychosis.

Table 1 about here

2. Methods

2.1. Sample and recruitment

Study participants were recruited as part of the prospective "***F***rüherk***e***nnung von ***Psy***chosen" (***FePsy***; English: Early detection of psychosis) project between March 2000 and

July 2016. A detailed description of the study design is provided in Riecher-Rössler et al.,(2007). In brief, patients suspected to have an ARMS for psychosis were referred to our specialized early detection clinic at the Psychiatric University Outpatient Department of the Psychiatric University Hospital Basel, Switzerland, by local psychiatrists, family doctors, or other hospital departments. Some also sought help with us at the advice of family members or through self-referral.

To be eligible for the FePsy study ~~participation~~, patients had to be at least 18 years old. Exclusion criteria were (1) insufficient knowledge of German, (2) indication of an IQ below 70, (3) psychotic symptoms within a clearly diagnosed affective psychosis or borderline personality disorder, (4) symptoms clearly due to organic reasons or substance use, and (5) antipsychotic treatment with a lifetime cumulative chlorpromazine equivalent dose of more than 2500mg. For the present study, patients were also excluded if they had not completed at least 50% of the items of the FCQ.

Written informed consent was obtained from all participating patients. The study was approved by the local ethics committee (EKNZ, Ethikkommission der Nordwest- und Zentralschweiz) and conformed to the Declaration of Helsinki.

2.2. Screening procedure

Patients were screened with the Basel Screening Instrument for Psychosis (BSIP), which has good interrater reliability (Kappa = 0.67) and high psychosis-predictive validity (32%; Riecher-Rössler et al., 2008). Individuals were classified by the BSIP as being in an ARMS for psychosis, having a FEP or not being at risk for psychosis, using criteria corresponding to the UHR-criteria of Yung et al. (1998).

2.3. Assessment of basic symptoms

BS were assessed with the original German paper-pencil version of the FCQ (Süllwold, 1991), presented as part of a larger package of questionnaires at the baseline of the *FePsy*-study. The FCQ contains 98 statements describing particular complaints that are rated dichotomously (yes/no) indicating either presence or absence of the complaint. If

patients had experienced a complaint in the past but not around the time of assessment, they are asked to score “yes” and add the word “formerly” next to it. For the lack of detailed time specifications, however, these items were regarded as absent in the analyses. Earlier, good retest reliability and internal consistency of the FCQ were reported, with most studies reporting Cronbach’s alpha measures of .90 and higher in samples of schizophrenia patients (Süllwold, 1991; Yon et al., 2008).

2.4. Follow-up and transition to psychosis

ARMS patients were followed-up for transition to psychosis for up to five years. During the first year of follow-up, patients were assessed monthly. During the second and third year, assessments took place every three months, thereafter patients were followed up annually.

The Brief Psychiatric Rating Scale (BPRS; Lukoff et al., 1986) items “suspiciousness”, “unusual thought content”, “hallucinations” and “conceptual disorganization”, which are included in the BSIP, were used to determine transition to psychosis in ARMS patients. Patients were considered to have transitioned according to criteria of Yung et al. (1998), i.e. when any one frank psychotic symptom (BPRS score of at least five or four for “hallucinations”) had persisted for more than one week.

2.5. Statistical Analyses

Confirmatory factor analyses (CFA) were used to test the fit of the six different factor solutions described in Table 1. The four-, two- and one-factor models were fitted using the weighted least square mean and variance adjusted estimator (WLSMV) as this is currently considered the best option for conducting a CFA with categorical data (Brown, 2015). Unlike other estimators for categorical data, the WLSMV produces accurate test statistics, parameter estimates, and standard errors even with relatively small sample sizes (Brown, 2015). Furthermore, it provides a large variety of fit indices. We tested goodness of model fit using the model fit X^2, the Comparative Fit Index (CFI), the Tucker–Lewis–Index (TLI), the Root Mean Square Error of Approximation (RMSEA) and the weighted root mean residual

(WRMR). Measures for good, adequate and poor fit respectively for all fit indices used are described in the footnotes below Table 3.

Since the 10-factorial solution was too complex to be fitted using WLSMV, we additionally fitted all models using a Bayesian approach. A comprehensive description of Bayesian Structural Equation Modeling (BSEM) is provided by Muthén and Asparouhov (2012b). Bayesian analysis defines parameters as variables as opposed to constants, using the term “prior” to describe their prior distribution. Priors are typically based on hypotheses deducted from theory or previous analysis (Muthén and Asparouhov, 2012b). Bayesian analyses result in a posterior estimate, which can be considered a compromise between the prior and the likelihood generated using the observed data (Muthén and Asparouhov, 2012b). In case of weakly informative priors, this estimate is comparable to the WLSMV. Goodness of model fit is evaluated using the posterior predictive p-value (ppp), which represents the proportion of model-generated test statistics that exceeds the sample-derived test statistics. Thus, small values indicate poor model fit (Brown, 2015). In addition, the 95% confidence interval (CI) for the difference between the observed and the replicated chi-square values can be used to determine goodness of model fit. Good model fit is indicated if the ppp is close to 0.5, the lower band of the CI is negative and the difference between the observed and replicated chi-square value is close to the middle of the CI (Brown, 2015).

BSEM has several advantages over traditional WLSMV analyses, including allowance for a more reasonable and flexible approach to model testing. Superior performance in smaller samples and computationally less challenging analyses are important benefits of the BSEM approach compared to WLSMV (Muthén and Asparouhov, 2012a). In the case of the current study, it proved to be especially useful to estimate a model for a complex factor solution with ten factors. Furthermore, compared to WLSMV, BSEM analysis allows more flexibility in model specification by relaxing the unrealistic assumption of strictly zero cross-loadings and residual correlations. Therefore, it can be tested whether model fit improves by allowing small variance on cross-loadings, i.e. approximately zero instead of strictly zero (Brown, 2015).

We first performed Bayesian CFA on all six factor solutions (Table 1) with cross-loadings fixed to zero comparable to the previously calculated WLSMV analysis. We used non-informative priors with parameters fixed to exactly zero for freely estimated residual and factor variances, which are the default priors of Mplus (Muthén and Muthén, 1998-2010). Due to initially bad model fit, we additionally relaxed the restraint on cross-loadings for the two-, four- and ten- factorial solutions in order to test whether model fit would improve. We did so by setting sequentially less strict informative priors for cross-loadings starting with variances equal to 0.001, 0.01, progressing to 0.05 and 0.1 as long as model fit would no longer improve. Since the two remaining models contained only one factor, relaxing restraints on cross-loads could not be performed on those models.

The reliabilities of all previously proposed FCQ scales in our sample of ARMS and FEP patients was estimated by the nonlinear SEM reliability coefficient of Green and Yang (2009), which has been termed categorical omega (ω_{cat}) by Kelley et al. (2016). Although Cronbach's alpha is the most popular measure of reliability, we did not report it here because it rests on several assumptions that are rarely met in practice (i.e. a one-factor CFA model perfectly fits the item covariances, the factor loadings of each item are equal, and items are continuous) and thus is frequently biased (cf. Dunn et al., 2014). Categorical omega, on the other hand, does not require the fulfillment of any of these assumptions and therefore has been recommended as the best option for estimating the reliability of composites that are the sum of categorical item scores (Kelley and Pornprasertmanit, 2016). In line with recommendations of Kelley and Pornprasertmanit (2016), we also estimated the confidence intervals for categorical omega by using the bias-corrected-and-accelerated bootstrap. All reliability calculations were performed by using the ci.reliablity function in the R package MBESS version 4.4.0 (Kelley, 2007)

~~Reliability and homogeneity of each solution were based on tetrachoric correlations between the items. We used Cronbach's alpha to determine internal consistency. Homogeneity was determined using Revelle's beta. Revelle's beta is a more conservative estimate that represents the lower half of split-half reliabilities and therefore, is always below or equal to~~

~~Cronbach's alpha. Revelle's beta is considered to apply to a wider range of different conditions as it is not biased under multidimensionality (Revelle, 1979; Soutar, 2009), while Cronbach's alpha relies on the underlying assumption of a common factor.~~

The abilitiesy of the various FCQ measures to predict later transition to psychosis in ARMS patients were tested outside the structural equation modelling framework with univariate Cox proportional hazard models, in which the FCQ totals or factors served as the independent variable and time to transition to psychosis as the outcome measure. For the exploratory nature of this first-time examination, we did not adjust for multiple testing at this step.

Missing data in FCQ items was handled using pairwise deletion in models estimated with WLSMV and was automatically adjusted for in Bayesian structural equation models (for details, see (Asparouhov and Muthén, 2010). For reliability and predictive validity analyses, missing data in FCQ items were singly imputed using the missForest algorithm (Stekhoven and Bühlmann, 2012). Although multiple imputation is less biased than single imputation, we opted for a single imputation approach because the fraction of missing data was relatively small and thus unlikely to lead to substantial bias with single imputation. Incomplete follow-up information was handled in predictive validity analysis by applying survival models instead of binary outcome models (e.g. logistic regression). Survival models have the advantage that they can treat patients with incomplete follow-up as censored observations (i.e. patients can still provide information for estimating the probability of transition up to the time they were followed up).

Structural equation models were fitted with Mplus, version 7 (Muthén and Muthén, 1998-2010). All other analyses were performed using the R environment for statistical computing (R Core Team, 2014; Revelle, 2016). Data and scripts of all analyses are available on request. Study method and results are reported according to the guidelines outlined by Brown (2015).

3. Results

3.1. Sample characteristics

Of the 693 patients screened during the recruitment period, 291 were identified as FEP and 289 as ARMS. 140 FEP and 186 ARMS patients consented to participate in the FePsy study, 92 FEP and 117 ARMS additionally completed at least half of the items of the FCQ and were therefore used for the sample of the current study. current study. ~~A sample of 117 ARMS and 92 FEP patients fulfilled the inclusion criteria.~~

ARMS and FEP with and without sufficient FCQ data did not differ statistically with regard to age, sex and years of education. Missing FCQ data in the included patients was relatively small with the majority of items having no more than 2% missingness (see online supplementary material for the proportion of data present in each item and for each pairwise combination)

~~As expected, there were significant differences between the two patient groups with higher mean age, more positive symptoms and a higher BPRS total score in FEP patients (Table 2).~~ Twenty-four (20.5%) ARMS patients later made a transition to psychosis within the follow-up (Table 2). Mean follow-up time was 1.24 years (median 0.80, range 0.03-4.86) for patients with later transition (ARMS-T) and 2.83 years (median 2.53, range 0.05-5.00) for patients without transition (ARMS-NT). The attrition rate in ARMS-NT patients after 1, 2 and 3 years was 26%, 37%, and 55%, respectively. Although incomplete follow up in ARMS-NT patients was mostly due to drop-outs, some patients also had incomplete follow-up because they were recruited less than 5 years before the recruitment period ended.

As expected, there were significant differences between ARMS and FEP patients with higher mean age, more positive symptoms and a higher BPRS total score in FEP patients (Table 2). FEP Patients showed significantly more positive symptoms and a higher overall BPRS score at baseline compared to ~~patients that made a transition to psychosis (~~ARMS-T patients~~)~~. At baseline, there were no statistically significant differences in any of the presented measures between ARMS-T ~~patients~~ and ~~those who did not transition to psychosis (~~ARMS-NT~~)~~ patients, although ARMS-T patients presented more positive symptoms, higher BRPS total scores and higher SANS scores.

Table 2 about here

3.2. Model evaluation

Overall goodness of fit and multiple fit indices for the models estimated with WLSMV are shown in Table 3 and those estimated with Bayesian methods in Table 4. Furthermore, Mplus scripts and standardized parameter estimates for all estimated structural equation models are available in the online supplementary material.

All tested WLSMV models provided an acceptable fit to the data according to the CFI, TLI, and RMSEA (Table 3). Overall, the one-factor solution (24-item short version) suggested by Loas et al. (2002) had the best fitting model with both CFI and TLI being >0.98, although the two- and four-factor models (Mass et al., 1997; Süllwold, 1991) also showed similarly high CFI and TLI indices of ≥0.95. The RMSEA however was best for the four-factor model (0.020) and worst for the short version one-factor model (0.035). Both models based on all 98 FCQ items (Süllwold, 1991; Yon et al., 2008) as well as the two-factor model on 50 items (Mass et al., 1997) provided an equally good fit with similar fit indices. Depending on the fit index, the four-factor model (RMSEA) or the Mass et al. (1997) two-factor-model (CFI, TLI) fit best of all solutions with at least 50 items (Table 3).

Table 3 about here

For all BSEM models, the ppp were >0.04 and fell within the range of the 95%CI for the difference between the observed and the replicated chi-square values, thus indicating adequate model fit (Table 4). In contrast to the WSLMV analyses, the Mass et al. (1995) two-factor model provided the best fit, as its ppp of 0.394 was closest to 0.5 (Table 4), followed by the 24-item short version with a ppp of 0.317. Of the four ≥50-item models, the two-factor model showed the best fitting value with a ppp of 0.143 using cross-loadings fixed to zero.

For the two-factor 50-item model suggested by Mass et al. (1997), an improvement in model fit could be found sequentially with increasing informative priors on the variance of the cross-loadings between the factors (Table 4). The model with the most flexibility allowance

for cross-loads, i.e. 0.1, showed a ppp closest to 0.5 (ppp = 0.193). While the already best fitting second two-factor model improved only slightly in ppp (0.396 for the solution with the most relaxed cross-loads), a considerable improvement in model fit with setting informative priors for variances of cross-loadings at 0.1 was found in the four-factor model, whose ppp improved from 0.081 to 0.261, as well as for the original ten-factor model, whit a ppp improvement from 0.089 to 0.448 (Table 4). Comparing the two 98-item model solutions using increasingly relaxed non-zero cross-loadings, the original theory-based ten-factor model fitted better than the data-based four-, the first two-factor model and both one-factor models but still less well than the best fitting second two-factor solution (Table 4).

Table 4 about here

3.3. ~~Internal consistency and homogeneity~~Scale reliability

Point estimates and confidence intervals of the reliability coefficient omega categorical for all previously proposed FCQ scales in our sample of ARMS and FEP patients are shown in Table 5. Most scales demonstrated good ($0.9 > \omega_{cat} \geq 0.8$) or excellent ($1 \geq \omega_{cat} \geq 0.9$) reliability. The alcohol and schizophrenia specific subscales of Mass et. al (1995) were the only ones with a reliability below 0.8, although they were still in the acceptable range ($0.8 > \omega_{cat} \geq 0.7$). These were also the only scales with a lower bound of the 95% confidence interval below 0.7.

~~For most subscales, Cronbach's alpha was 0.90 or higher, indicating excellent reliability in terms of an internal consistency (Table 5). The FCQ total scale also yielded excellent results with a Cronbach's alpha of 0.96 in ARMS and 0.99 in FEP. Generally, Revelle's beta yielded lower values. For ARMS patients acceptable values were found for only two factors (memory, disturbances of automated responses), while for FEP patients values were frequently still good or at least acceptable (Table 5). Furthermore, several factors showed inacceptable Revelle's beta values of <0.50, mainly in ARMS but also rarely in FEP patients (Table 5). In addition, the homogeneity of the total FCQ was poor in both groups with beta between 0.50 and 0.58.~~

3.4. Survival analysis (prediction of psychosis)

Neither any one of the factors of the six tested models nor the FCQ total score was found to be significantly associated with a later transition to psychosis in ARMS patients (Table 6).

Table 6 about here

4. Discussion

The present study investigated for the first time the goodness of fit of all previously ~~various~~ proposed factor structures of the FCQ, ~~regarding their~~ as well as scale reliabilities~~y~~, ~~homogeneity~~ and psychosis-predictive validities~~y~~ of all previously proposed FCQ scales in a ~~.~~ ~~Advancing earlier factor analytical approaches developed for continuous data, we used CFA and methods appropriate for categorical items to compare the model fit of six previously proposed factor solutions~~ in a sample of ARMS and FEP patients. In contrast to previous psychometric investigations of the FCQ, methods appropriate for categorical items were used. Although none of the tested factorial structures was clearly rejected, results indicated that the one factorial solution based on 24 items suggested by Loas et al. (2002) and the two-factorial solution based on 17 items by Mass et al. (1995) provided the best fit to the data. However, if only factorial solutions based on at least 50 items are considered, the two-factorial solution of Mass et al. (1997) and the four factorial solution of Süllwold (1991) performed best. While all tested FCQ scales had at least acceptable reliability, none of these scales could be demonstrated to have psychosis predictive validity. ~~Albeit considerable heterogeneity in findings across methods, the results indicated that the psychometric properties in terms of validation of factor structure, internal consistency, i.e., reliability, and homogeneity of the two factor solutions based on the lowest number of items are most satisfactory. These are the 24-item short version of the FCQ proposed by Loas et al. (2002; WLSMV) and the second two-factor solution found by Mass et al. that both resulted from samples with exclusively and predominately schizophrenia patients.~~

The one factorial solution based on 24 items ~~of the single factor~~ by Loas et al. (2002), which provided the best fit to our data when using WLSMV estimation, ~~were~~ was originally derived by performing a PCA in a sample of schizophrenia patients. Although both parallel analysis and the scree test indicated that the optimal number of factors was two, the authors opted for a one factorial solution based on the finding that all items had their strongest loading on the first factor in the unrotated solution. The one-factorial solution was further refined by retaining only those 24 items that had a factor loading > 0.6 (Loas et al., 2002). The two-factor model of Mass et al. (1995) based on 17 items, which provided the best fit to our data when using Bayesian estimation, was not based on factor analysis but on item-by-item comparisons of the sensitivity to either schizophrenia or alcohol dependency, revealing two factors construed as schizophrenia- and alcohol dependency-specific factors. However, the superior fit of the models suggested by Loas et al. (2002) and Mass et al. (1995) might be explained by the restricted number of items they are based on. To answer the question of how many dimensions are measured by the FCQ, fit indices of models based on a more complete set of items are likely more meaningful. Considering only these models, we found that the two and four factorial solutions of Mass et al. (1997) and Süllwold (1991), respectively, provided the best fit. Overall, both models were approximately equally well fitting. Specifically, while the two factor model had a slightly better CFI, TLI and WRMR with WLSMV estimation and a slightly better ppp with Bayesian estimation, the four factorial model had a slightly better RMSEA with WLSMV estimation.

Thus, ~~O~~our structural results in a sample of ARMS and FEP patients suggest that BS as measured by the FCQ are ~~all~~ indicators of ~~only one or at most two~~two to four underlying dimensions~~ indicative of a general vulnerability for disturbances of mental and central-vegetative processes~~, which is in line with some of the previous research on the factor structure of the FCQ – (Mass et al., 1997; Süllwold, 1991), as well reported results from parallel analyses (Loas et al., 2002; Yon et al., 2008). Contrary to this, dimensional analyses of *interview-assessed* BS (which also excluded central-vegetative complaints) had revealed six clear and replicable dimensions in adult prodromal and manifest psychosis samples

(Schultze-Lutter, 2008b; Schultze-Lutter et al., 2012b), of different specificity for psychosis. This difference between the dimensional structures of self- and interviewer-rated BS is most likely due to the lack of content validity of self-assessed BS when compared to interview-assessed BS (Michel et al., 2016a). Also, potential effects of excluded central-vegetative complaints have not been studied so far. Furthermore, while items of the interview assessments refer to clearly distinct phenomena, items of the FCQ are less markedly defined, and frequently several items relate to the same interview-assessed BS (Michel et al., 2016b). Consequently, the 98 items of the FCQ do not cover all BS described in the BSABS. Moreover, they do not cover four basic symptoms and some acoustic and visual perception disturbances of COPER/COGDIS. These differences in item pools might have additionally affected dimensional analyses.

In the present study, the reliabilities of the FCQ scales were estimated for the first time with a new SEM based reliability coefficient, the so called categorical omega (Green and Yang, 2009; Kelley and Pornprasertmanit, 2016). Unlike Cronbach's alpha, which has been used in all previous psychometric investigations of the FCQ, categorical omega does not rest on the unrealistic assumption of a perfectly fitting unidimensional CFA model with equal factor loadings (i.e., essential tau-equivalence) (Kelley and Pornprasertmanit, 2016). Our results confirm that the FCQ total scales of both the original 98 item (Süllwold, 1991) and shortened 24 item versions (Loas et al., 2002) have excellent reliabilities (i.e. 0.99 and 0.95, respectively). Previously reported alpha values for these scales ranged between 0.95 and 0.97 (Loas et al., 2002; Süllwold, 1991; Yon et al., 2008) and between 0.87 and 0.94 (Loas et al., 2002; Yon et al., 2008), respectively. Hence, our reliability estimates tended to be slightly higher than in previous studies, which might be explained by the fact that our method was not biased downwards under violation of essential tau-equivalence. The reliabilities of FCQ subscales, to our knowledge, have not been investigated in previous studies. We could therefore demonstrate for the first time that these scales have mostly good to excellent reliabilities. The only scales with a reliability < 0.8 were the alcohol and schizophrenia specific subscales of Mass et al. (1995). One possible explanation is that these scales were

not based on factor analytic results or similarity of item content, but on the diagnostic power of single items.

~~Our reliability analyses provided less heterogeneous results and generally confirmed studies on the FCQ reporting excellent to acceptable values for Cronbach's alpha (0.78 by Loas et al., 2002; Schultze-Lutter et al., 2015; 0.97 by Süllwold, 1991; 0.95 by Yon et al., 2008) As Cronbach's alpha might not be accurate if the assumption of unidimensionality is violated, Revelle's beta provides a more accurate evaluation of homogeneity. Supporting a two- rather than one-dimensional structure of the FCQ, lower Revelle's beta values still indicated good homogeneity of the majority of FCQ-factors in FEP patients but not in ARMS patients. This group difference, however, was most likely conveyed by the differences of homogeneity of samples that, from a diagnostic point of view, was high in the schizophrenia-spectrum FEP patients and low in the ARMS patients that ultimately will belong to several diagnostic groups with different response patterns in the FCQ (Mass et al., 1995). As the number of ARMS-T patients was very small, we were not able to run separate analyses for ARMS-T and ARMS-NT. Expecting higher homogeneity in an ARMS-T compared to an ARMS-NT sample or in a combined sample of ARMS-T and FEP, such analyses will be subject of future studies of altogether larger sample size.~~

In line with early studies questioning the psychosis-specificity of the FCQ (Mass et al., 1995), we found no predictive validity with respect to later transition to psychosis in ARMS patients – neither for the FCQ total score nor for any of its subscales. So far, no other self-assessments instruments for BS were tested for predictive validity. This is in contrast to previous studies, in which BS were assessed using semi-structured interviews such as the SPI-A (Schultze-Lutter et al., 2007), that reported a psychosis-predictive value of a subgroup of BS included in COPER and COGDIS of at least comparable quality to UHR criteria (Schultze-Lutter et al., 2014; Schultze-Lutter et al., 2015). This indicates that BS are only predictive of later transition to psychosis if descriptions of patient's experiences are rated based on in-depth clinical interviews by experienced clinicians and are not simply rated by

patients themselves based on a single exemplary statement. This interpretation is in line with previous studies that have generally found poor correlations and convergent validity, respectively, between BS measured with the FCQ and those assessed with the gold-standard of a semi-structured interview (Mass et al., 1995; Michel et al., 2016a). However, due to difficulties in assigning a BSABS or SPI-equivalence to each FCQ-item and vice versa, these previous studies did not include all FCQ items and COPER/COGDIS items, respectively. This further questions the correspondence between FCQ and BSABS or its subsequent scales, SPI-A and SPI-CY. While FCQ-assessed BS seem to predict worse symptomatic outcomes in FEP patients (Prouteau et al., 2004) – likely as a measure of overall symptom severity, the same does not seem to be the case for transition to psychosis for lack of psychosis-specificity of the FCQ. In addition to differences in item content, differences in the item response format could also have contributed to heterogeneous results regarding the predictive validities of self and interview-assessed BS. Specifically, the dichotomous item response format of the FCQ might have precluded the assessment of BS symptoms of low intensity and thus have led to items with low endorsement frequencies. Future studies should therefore investigate whether the use of Likert scales, as implemented in the SPI-A, improves the predictive validities of self-rated BS.

The predictive value of the FCQ might be further lowered by the lack of controlling for confounders, such as somatic illness or substance use effects, in the assessment of BS with the FCQ. For example, Moritz et al. (2000a) questioned the construct validity of the FCQ, as it found evidence that the FCQ score correlated significantly with the administered dose of antipsychotic medication. The authors found a significant positive correlation of BS measured with the FCQ with neuroleptic dosage instead of the expected opposite treatment effects, even when controlled for other psychopathology. While in interviews the co-occurrence of BS with medication or drug use is controlled for, this is not done in the FCQ. However, a medication effect was controlled for in our sample, because patients with a summarized chlorpromazine equivalent dose of 2500 mg or higher were excluded from study participation. Furthermore, ARMS patients were not treated with antipsychotics after baseline

unless after transition, while FEP patients were treated with antipsychotics only after completion of baseline assessments. Thus, further supported by the lack of differences in medication between patient groups at baseline, we do not assume that such a medication effect significantly biased our results.

A limitation of our study was the rather small sample size of just over 200 participants. Although Bayesian analysis has been shown to perform well even with smaller sample sizes, it was not possible to estimate the factor models for ARMS and FEP patients or ARMS-T and ARMS-NT patients separately. Hence, it was also not possible to estimate scale reliabilities for these groups separately. As most ARMS patients did not transition to psychosis, the two patient groups do not completely pertain to the same spectrum of disorders, and therefore it would be interesting to investigate possible factor invariance across patient groups. A study on BSABS dimensions did not find differences between ARMS-T and FEP patient groups but with a subsample of patients suffering from depressive disorders (Schultze-Lutter, 2008a).

5. Conclusion and directions for future research

Depending on the method used, our results suggest a ~~one-or at most two~~two to four-factorial structure ~~based on a small selection of items t~~o best fit the ~~CFA model~~full 98 item version of the FCQ. Since this is the first study to use CFA, categorical item methodology and a sample of early detection of psychosis patients (ARMS and FEP), more research of this nature is clearly needed. Studies on the factor invariance of the FCQ in separate samples of ARMS and FEP patients as well as ARMS-NT and ARMS-T and a combined sample of ARMS-T and FEP patients would be interesting. Although we could not demonstrate that any FCQ scales have predictive validities, it might still be useful to apply this questionnaire in clinical practice to assess subjectively experienced BS in ARMS patients and thereby to facilitate personalized treatment (e.g. initiating cognitive remediation). Since the FCQ as a self-report represents a more economical possibility to assess BS as

compared to more time- and resource-intensive semi-structured interviews, it seems important to develop scales of better convergent validity of items with interview-assessed BS to fully and broadly exploit the advantages of the BS approach in the early detection of psychosis.

Conflicts of interest

All authors declare not to have any conflicts of interest that might have influenced the content of the manuscript.

The authors of this study were supported by research grants of the Swiss National Science Foundation (SNF) and by the EU FP7 Project «European Network of National Schizophrenia Networks Studying Gene-Environment Interactions» (EU-GEI).

Acknowledgements

We thank all patients who participated in the interviews as well as the referring specialists and colleagues. We also would like to thank the registry office of the Centre for Gender research and early detection– namely Johannes Hapig - for his help with the preparation and submission of the manuscript.

References

Asparouhov T, Muthén B. (2010). Bayesian Analysis Using Mplus: Technical

Implementation.
Brown TA. (2015). *Confirmatory Factor Analysis for Applied Research - Second Edition*. New York

London: The Guilford Press.
Dunn TJ, Baguley T, Brunsden V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *Brit J Psychol* **105**(3): 399-412, DOI: 10.1111/bjop.12046.
Green SB, Yang YY. (2009). Reliability of Summed Item Scores Using Structural Equation Modeling: An Alternative to Coefficient Alpha. *Psychometrika* **74**(1): 155-167, DOI: 10.1007/s11336-008-9099-3.
Gross G. (1987). *BSABS Bonner Skala für die Beurteilung von Basissymptomen Manual, Kommentar, Dokumentationsbogen*. Berlin: Springer.
Huber G. (1966). Reine

Defektsyndrome und Basisstadien endogener Psychosen. *Fortschritte der Neurologie Psychiatrie* **34**: 409-426.
Huber G, Süllwold L. (1986). *Schizophrene Basisstörungen*. Berlin: Springer.
Kelley K. (2007). Methods for the behavioral, educational, and social sciences: An R package. *Behav Res Methods* **39**(4): 979-984, DOI: Doi 10.3758/Bf03192993.
Kelley K, Pornprasertmanit S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychol Methods* **21**(1): 69-92, DOI: 10.1037/a0040086.

Klosterkötter J, Hellmich M, Steinmeyer EM, Schultze-Lutter F. (2001). Diagnosing schizophrenia in the initial prodromal phase. *ArchGenPsychiatry* **58**(2): 158-164.
Kubinger KD. (2003). On artificial results due to using factor analysis for dichotomous variables. *Psychology SciencePsychology Science* **45**(1): 106-110.
Loas G, Yon V, Brien D. (2002). Dimensional structure of the Frankfurt Complaint Questionnaire. *Compr Psychiatry* **43**(5): 397-403.
Lukoff D, Nuechterlein KH, Ventura J. (1986). Manual for the expanded brief psychiatric rating scale. *Schizophr Bull* **12**: 594-602.
Mass R, Haasen C, Krausz M. (1997). Dimensional structure and diagnostic specificity of the Frankfurt Complaint Questionnaire. *European Psychiatry* **12**: 117-123.
Mass R, Krausz M, Gross J. (1995). Basissymptome bei Schizophrenie und Alkoholismus. Eine methodenkritische Vergleichsstudie. Nervenarzt. pp 331-337.
Mass R, Weigel S, Schneider S, Klepsch R. (1998). Schizophrenia-specific basic symptoms. A successful replication. *Psychopathology* **31**(3): 113-119.
Michel C, Kutschal C, Schimmelmann BG, Schultze-Lutter F. (2016a). Convergent and concurrent validity of the Frankfurt Complaint Questionnaire as a screener for psychosis risk. *Journal of Risk Research*: 1-17, DOI: 10.1080/13669877.2016.1179209.
Michel C, Kutschal C, Schimmelmann BG, Schultze-Lutter F. (2016b). Convergentand concurrent validity of the Frankfurt Complaint Questionnaire as a screener for psychosis risk. *Journal of Risk Research*: 1-17, DOI: 10.1080/13669877.2016.1179209.
Michel C, Ruhrmann S, Schimmelmann BG, Klosterkötter J, Schultze-Lutter F. (2014). A stratified model for psychosis prediction in clinical practice. *Schizophr Bull* **40**(6): 1533-1542, DOI: 10.1093/schbul/sbu025.
Moritz S, Krausz M, Gottwalz E, Andresen B. (2000a). The impact of neuroleptic dosage and extrapyramidal side effects on schizophrenic basic symptoms. *Compr Psychiatry* **41**(4): 284-288, DOI: 10.1053/comp.2000.7425.
Moritz S, Krausz M, Gottwalz E, Lambert M, Perro C, Ganzer S, Naber D. (2000b). Cognitive Dysfunction at Baseline

Predicts Symptomatic 1-Year Outcome

in First-Episode Schizophrenics. *Psychopathology* **33**: 48-51.
Muthén B, Asparouhov T. (2012a). Bayesian structural equation modeling: a more flexible representation of substantive theory. *PsycholMethods* **17**(3): 313-335, DOI: 10.1037/a0026802.
Muthén B, Asparouhov T. (2012b). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychol Methods* **17**(3): 313-335, DOI: 10.1037/a0026802.
Muthén LK, Muthén BO. (1998-2010). Mplus User's Guide. Los Angeles, CA.
Nelson B, Thompson A, Yung AR. (2012). Basic self-disturbance predicts psychosis onset in the ultra high risk for psychosis "prodromal" population. *Schizophr Bull* **38**(6): 1277-1287, DOI: 10.1093/schbul/sbs007.
Parnas J, Raballo A, Handest P, Jansson L, Vollmer-Larsen A, Saebye D. (2011). Self-experience in the early phases of schizophrenia: 5-year follow-up of the Copenhagen Prodromal Study. *World Psychiatry* **10**(3): 200-204.
Prouteau A, Verdoux H, Briand C, Lesage A, Lalonde P, Nicole L, Reinharz D, Stip E. (2004). Self-assessed cognitive dysfunction and objective performance in outpatients with schizophrenia participating in a rehabilitation program. *Schizophr Res* **69**(1): 85-91, DOI: 10.1016/j.schres.2003.08.011.
Revelle W. (1979). Hierarchical Cluster Analysis And The Internal Structure Of Tests. *Multivariate Behav Res* **14**(1): 57-74, DOI: 10.1207/s15327906mbr1401_4.
Riecher-Rössler A, Aston J, Ventura J, Merlo M, Borgwardt S, Gschwandtner U, Stieglitz RD. (2008). [The Basel Screening Instrument for Psychosis (BSIP): development, structure, reliability and validity]. *Fortschr Neurol Psychiatr* **76**(4): 207-216, DOI: 10.1055/s-2008-1038155.

Riecher-Rössler A, Gschwandtner U, Aston J, Borgwardt S, Drewe M, Fuhr P, Pflüger M, Radü W, Schindler C, Stieglitz RD. (2007). The Basel early-detection-of-psychosis (FePsy)-study - design and preliminary results. *Acta Psychiatr Scand* **115**(2): 114-125, DOI: ACP854 [pii]

10.1111/j.1600-0447.2006.00854.x.
Riecher-Rössler A, Gschwandtner U, Borgwardt S, Aston J, Pflüger M, Rössler W. (2006). Early detection and treatment of schizophrenia: how early? *Acta Psychiatr Scand Suppl*(429): 73-80, DOI: 10.1111/j.1600-0447.2005.00722.x.
Riecher-Rössler A, Pflueger MO, Aston J, Borgwardt SJ, Brewer WJ, Gschwandtner U, Stieglitz RD. (2009). Efficacy of Using Cognitive Status in Predicting Psychosis: A 7-Year Follow-Up. *Biol Psychiatry* **66**(11): 1023-1030, DOI: 10.1016/j.biopsych.2009.07.020.
Schultze-Lutter F. (2009). Subjective symptoms of schizophrenia in research and the clinic: the basic symptom concept. *Schizophr Bull* **35**(1): 5-8, DOI: 10.1093/schbul/sbn139.
Schultze-Lutter F, Addington J, Ruhrmann S, Klosterkötter J. (2007). *Schizophrenia Proneness Instrument (SPI-A)*. Rome, Italy: Giovanni Fioriti.
Schultze-Lutter F, Debbané M, Theodoridou A, Wood SJ, Raballo A, Michel C, Schmidt SJ, Kindler J, Ruhrmann S, Uhlhaas PJ. (2016). Revisiting the Basic Symptom Concept: Toward Translating Risk Symptoms for Psychosis into Neurobiological Targets. *Front Psychiatry* **7**: 9, DOI: 10.3389/fpsyt.2016.00009.
Schultze-Lutter F, Eckjar. (2008a). The dimensional structure of self-reported "prodromal" disturbances in schizophrenia. *ClinicalNeuropsychiatry* **5**(3): 140-150.
Schultze-Lutter F, Eckjar. (2008b). The dimensional structure of self-reported "prodromal" disturbances in schizophrenia. *Clinical Neuropsychiatry* **5**(3): 140-150.
Schultze-Lutter F, Klosterkötter J, Ruhrmann S. (2014). Improving the clinical prediction of psychosis by combining ultra-high risk criteria and cognitive basic symptoms. *Schizophr Res* **154**(1-3): 100-106, DOI: 10.1016/j.schres.2014.02.010.
Schultze-Lutter F, Michel C, Schmidt SJ, Schimmelmann BG, Maric NP, Salokangas RK, Riecher-Rössler A, van der Gaag M, Nordentoft M, Raballo A, Meneghelli A, Marshall M, Morrison A, Ruhrmann S, Klosterkötter J. (2015). EPA guidance on the early detection of clinical high risk states of psychoses. *Eur Psychiatry* **30**(3): 405-416, DOI: 10.1016/j.eurpsy.2015.01.010.
Schultze-Lutter F, Ruhrmann S, Berning J, Maier W, Klosterkötter J. (2010). Basic symptoms and ultrahigh risk criteria: symptom development in the initial prodromal state. *Schizophr Bull* **36**(1): 182-191, DOI: 10.1093/schbul/sbn072.
Schultze-Lutter F, Ruhrmann S, Fusar-Poli P, Bechdolf A, G. Schimmelmann B, Klosterkotter J. (2012a). Basic Symptoms and the Prediction of First-Episode Psychosis. *Current Pharmaceutical Design* **18**(4): 351-357, DOI: 10.2174/138161212799316064.
Schultze-Lutter F, Ruhrmann S, Fusar-Poli P, Bechdolf A, Schimmelmann BG, Klosterkötter J. (2012b). Basic symptoms and the prediction of first-episode psychosis. *Curr Pharm Des* **18**(4): 351-357.
Schultze-Lutter F, Ruhrmann S, Klosterkötter J. (2006). Can schizophrenia be predicted phenomenologically? In J Johannessen, B Martindale, J Cullberg (eds.) *Evolving Psychosis Different Stages, Different Treatments*. London, New York: Routledge. pp 104-123.
Soutar GN. (2009). Coefficient Beta: A Neglected Statistic. *Australian and New Zealand Marketing Academy*.
Stekhoven DJ, Bühlmann P. (2012). MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**(1): 112-118, DOI: 10.1093/bioinformatics/btr597.
Süllwold L. (1991). *Manual zum Frankfurter Beschwerde-Fragebogen (FBF)*. Berlin: Springer.
Team RC. (2014). R: A language and environment for statistical computing (Version 3.0.3).
Yon V, Loas G, Monestes JL. (2008). Dimensional structure of the Frankfurt Complaint Questionnaire in a sample of French students. *Psychopathology* **41**(2): 85-89, DOI: 10.1159/000111552.
Yung AR, McGorry PD. (1996). The prodromal phase of first-episode psychosis: past and current conceptualizations. *Schizophr Bull* **22**(2): 353-370.

Yung AR, Phillips LJ, McGorry PD, McFarlane CA, Francey S, Harrigan S, Patton GC, Jackson HJ. (1998). Prediction of psychosis - A step towards indicated prevention of schizophrenia. *Brit J Psychiat* **172**: 14-20.
Yung AR, Yuen HP, McGorry PD, Phillips LJ, Kelly D, Dell'Olio M, Francey SM, Cosgrave EM, Killackey E, Stanford C, Godfrey K, Buckby J. (2005). Mapping the onset of psychosis: the Comprehensive Assessment of At-Risk Mental States. *Aust N Z J Psychiatry* **39**(11-12): 964-971, DOI: 10.1080/j.1440-1614.2005.01714.x.

Table 1 The Frankfurt Complaint Questionnaire: Factor solutions and corresponding items.

Model	Original	Süllwold (1991)	Mass et al. (1997)	Mass et al. (1995)	Loas et al. (2002)	Yon et al. (2008)
Number of factors / items	10 / 98	4 / 98	2 / 50	2 / 18	1 / 24	1 / 98
N		229	692	242	310	399
Sample		Schizophrenia patients	Schizophrenia patients (N=505), Alcohol dependent patients (N=187)	Schizophrenia patients (N=85), Alcohol dependent patients (N=101), both diagnoses (N=222)	Schizophrenia patients	Students
Age (mean)		32.8 years	40.7 years	37.3 years	39.23 years	24.3 years
Gender Male		55.9%	79.1%	67.4%	37.9%	8.3%
Female		44.1%	20.9%	32.6%	26.1%	91.7%
Version	German	German	German	German	French	Spanish
Cronbachs' alpha	n.a.	0.96	n.a.	n.a.	0.78	0.95
Factors	1. Loss of control **2. Simple perception*** 3. Complex perception **4. Language*** **5. Thought*** 6. Memory 7. Motility 8. Lack of automatism 9. Anhedonia and anxiety 10. Sensory overstimulation	**1. Disturbances of automated responses*** **2. Perceptual disturbances*** **3. Depression*** 4. Overinclusion	**1. Dysphoric concomitants of severe illness particularly impairing concentration*** 2. **Subjective experiences of perceptual uncertainties**	**1. Schizophrenia-specific factor** 2. Alcohol-specific factor		
Correspon-ding items*	1. 7, 22, 33, 74, 83, **85**, 86, 96 2. **19**, **24**, **25**, **29**, **45**, **47**, **50**, **51**, **63**, 67, **84**, 92 3. **14**, 23, 26, 27, 97, 30, **76**, **79**, 97 4. **31**, **40**, 42, **66**, **69**, **71**, **82**, **90**, **93**, **94** 5. **2**, 4, 12, 35, **36**, 39, **43**, **54**, **70** 6. 8, **37**, 52, 60, 62, 68, 73, 78, 88, 91 7. 5, 9, 11, 18, 20, 34, 44, 59, 64, **81** 8. 6, **13**, 17, 38, 46, 48, 56, 57, 75, 77 9. 1, 15, 16, 28, 41, 55, **72**, 87, 95, 98 10. 3, 10, 21, **32**, 49, **53**, 58, 61, 65, **80**, 89	1. 33, 35, **36**, 42, 48, 52, **54**, 60, **63**, 67, 68, **69**, **70**, **71**, 73, **85**,88, 89, 91, **94**,95, 96 2. 3, 5, 9, 11, 14, 18, **19**, 20, 21, 23, **24**, **25**, **29**, 30, **32**, 34, **40**, **45**, **47**, **50, 51**, 59, 64, 67, **76**, **79**, **81**, **84**, 87, 92 3. 1, 6, 8, 15, 16, 17, 22, 26, 27, 28, **31**, **37**, 38, 39, **43**, 49, 55, 57, **72**, 75, 77, 78, **82**, **90**, **93**, 97, 98 4. **2**, 4, 7, 10, 12, **13**, 44, 46, **53**, 56, 58, 61, 62, 65, 74, **80**, 83, 86	1. 1, 5-8, 17, 27, **31**, 35-39, 42, 46, 52, **54**, 56, 65, **66**, 68-**71**, 73, 75, 77, 78, **80**, **82**, **90**, **93**, **94**, 95, 96, 98 2. 14, 18, **19**, 23-**25**, **29**, **32**, **45**, **51**, **63**, **76**, **79**, **84**	1. 11, **14**, 15, 63, **81**, **90**, **93**, **94** 2. 2, **13**, 16, 28, 41, **47**, 52, 56, 87	26, 27, 35, **36**, 38, 39, 42, 44, **45**, 46, 55, 60, **66**, **69**, 73, 75, **80**, **85**, 86, 89, **93**, 95, 96, 98	1-98
Method	Grouping according to phenomenological similarities	EFA, varimax rotation	PCA, varimax rotation	Item-by-item comparison	PCA, varimax rotation	PCA, varimax rotation
VE	n.a.	72%	59.6%	n.a.	65.30 %	68.98%

Note: EFA = Exploratory Factor Analysis, PCA = Principal Component Analysis, n.a. = not available, VE= variance explanation
*Factors and items most likely corresponding to COGDIS or COPER in bold (Michel et al., 2016a)

Table 2: Sociodemographic and clinical sample characteristics

		All	ARMS		FEP	p-value			
			ARMS-NT	ARMS-T		ARMS vs FEP	ARMS-NT vs ARMS-T	ARMS-NT vs FEP	ARMS-T vs FEP
N		N=209	N=93	N=24	N=92				
Age mean (SD)*		27.3 (7.99)	25.4 (7.46)	25.6 (5.99)	29.7 (8.37)	<0.001	0.993	<0.001	0.051
Gender**	Women (%)	67 (32.1%)	31 (33.3%)	7 (29.2%)	29 (31.5%)	0.917	1.000	1.000	1.000
	Men (%)	142 (67.9%)	62 (66.7%)	17 (70.8%)	63 (68.5%)				
Years of education mean (SD)*		11.5 (2.87)	11.6 (3.05)	11.4 (2.06)	11.4 (2.89)	0.882	0.931	0.897	0.997
Occupation**	Unemployed (%)	66 (35.3%)	24 (29.3%)	4 (18.2%)	38 (45.8%)	0.099	0.425	0.256	0.151
	Employed (%)	57 (30.5%)	29 (35.4%)	6 (27.3%)	22 (26.5%)				
	In education (%)	49 (26.2%)	23 (28.0%)	9 (40.9%)	17 (20.5%)				
	Other (%)	15 (8.02%)	6 (7.32%)	3 (13.6%)	6 (7.23%)				
~~Years of Follow up mean (SD)*~~			~~2.83 (2.86)~~	~~1.24 (1.30)~~					
BPRS Positive Symptoms mean (SD)*		9.57 (4.50)	6.53 (2.58)	8.13 (2.40)	13.3 (3.86)	<0.001	0.083	<0.001	<0.001
BPRS total score mean (SD)*		43.7 (12.1)	37.3 (8.33)	42.1 (8.86)	50.9 (12.4)	<0.001	0.125	<0.001	0.001
SANS total score mean (SD)*		22.9 (15.7)	22.0 (16.6)	25.8 (16.2)	23.0 (14.7)	0.597	0.568	0.919	0.732

Note: ARMS = at-risk mental state (T = transition, NT = non-transition); FEP = first episode psychosis; BPRS = Brief Psychiatric Rating Scale; SANS = Scale for the Assessment of Negative Symptoms

Continuous variables* were tested using ANOVA, categorical variables** using χ^2 test statistics or Fisher exact test in case of cell counts <5.

Table 3: Fit of the factor models estimated with WLSMV

Model	Number of factors	Para-meters	Chi Square	DF	Chi Square Model fit p-value	CFI	TLI	RMSEA (90% CI)	WRMR
Süllwold (1991)	4	200	4942.4	4553	<0.001	0.961	0.960	0.020 (0.015-0.024)	0.976
Mass et al. (1997)	2	101	1387.1	1174	<0.001	0.965	0.964	0.029 (0.022-0.036)	0.970
Mass et al. (1995)	2	35	160.4	118	0.0057	0.962	0.956	0.041 (0.023-0.057)	0.908
Loas et al. (2002)	1	48	316.3	252	0.0037	0.981	0.980	0.035 (0.021-0.046)	0.859
Yon et al. (2008)	1	196	5168.8	4655	<0.001	0.949	0.948	0.023 (0.019-0.027)	1.027

Note: WLSMV = weighted least square mean and variance adjusted estimator (χ^2<0.05 = good fit, <0.08 = adequate fit, >0.08 = poor fit), CFI = Comparative Fit Index (>0.95 = good fit, >0.9 = adequate fit, <0.9 = poor fit), TLI = Tucker-Lewis-Index (>0.95 = good fit, >0.9 = adequate fit, <0.9 suggests poor fit), RMSEA = Root Mean Square Error of Approximation (<0.05 = good fit, <0.08 = adequate fit, >0.08 = poor fit), WRMR = weighted root mean residual (<1.0 = adequate model fit), CI = Confidence Interval

Table 4: Fit of the factor models estimated with Bayesian method

Model	Number of factors	Cross-loads	Number of parameters	ppp	95% CI for the difference between the observed and replicated chi-square value	
					lower limit	upper limit
Original	10	0	241	0.089	-132.090	556.490
		0.001	1123	0.138	-152.596	528.073
		0.01	1123	0.219	-197.821	458.485
		0.05	1123	0.383	-294.922	383.571
		0.1	1123	0.448	-317.942	356.685
Süllwold (1991)	4	0	200	0.081	-94.850	566.360
		0.001	491	0.082	-99.837	570.740
		0.01	491	0.208	-139.130	540.545
		0.05	491	0.200	-185.160	488.248
		0.1	491	0.261	-223.003	450.478
Mass et al. (1997)	2	0	101	0.143	-75.050	245.540
		0.01	151	0.167	-80.590	243.600
		0.05	151	0.191	-88.520	233.250
		0.1	151	0.193	-88.860	231.240
Mass et al. (1995)	2	0	35	0.394	-45.612	60.755
		0.001	52	0.389	-46.699	61.196
		0.01	52	0.396	-46.801	60.807
		0.05	52	0.405	-47.630	60.197
		0.1	52	0.416	-48.345	59.312
Loas et al. (2002)*	1 / 24 items	0	48	0.317	-60.490	76.970
Yon et al. (2008)*	1 / 98 items	0	196	0.042	-38.680	647.450

Note: CI = confidence interval; ppp = posterior predictive p-value (ppp close to 0.5 = good model fit)
*Relaxing constraints on cross loads is not possible in models containing only one factor

Table 5: Scale reliabilities

Model	Scale	Number of items	Omega categorical
Original	Loss of control	8	0.84 [0.78; 0.87]
	Simple perception	10	0.86 [0.74; 0.90]
	Complex perception	10	0.86 [0.61; 0.90]
	Language	10	0.91 [0.85; 0.92]
	Thought	10	0.87 [0.82; 0.89]
	Memory	10	0.88 [0.82; 0.91]
	Motility	10	0.84 [0.69; 0.87]
	Loss of automatization	10	0.87 [0.82; 0.89]
	Anxiety / Anhedonia	10	0.80 [0.71; 0.83]
	Sensory overstimulation	10	0.83 [0.76; 0.86]
Süllwold (1991)	Disturbances of automated responses	22	0.95 [0.92; 0.95]
	Perceptual disturbances	30	0.97 [0.93; 0.97]
	Depression	27	0.96 [0.94; 0.96]
	Overinclusion	18	0.90 [0.85; 0.91]
Mass et al. (1998)	Dysphoric concomitants of severe illness particularly impairing concentration	36	0.99 [0.99; 0.99]
	Subjective experiences of perceptual uncertainties	14	0.90 [0.74; 0.92]
Mass et al. (1995)	Alcohol dependency specific subscale	9	0.75 [0.67; 0.80]
	FCQ Schizophrenia specific subscale	8	0.77 [0.67; 0.82]
Loas et al. (2002)	24 item scale	24	0.95 [0.93; 0.95]
Yon et al. (2008)	FCQ total scale	98	0.99 [0.99; 1.00]

Note: FCQ = Frankfurt Complaint Questionnaire; ARMS = At-risk mental state; FEP = First episode psychosis

~~**Table 5:** Scale reliabilities~~

			~~ARMS~~		~~FEP~~	
~~Model~~	~~Scale~~	~~n items~~	~~alpha~~	~~beta~~	~~alpha~~	~~beta~~
~~Original~~	~~Loss of control~~	~~8~~	~~0.80~~	~~0.67~~	~~0.91~~	~~0.82~~
	~~Simple perception~~	~~10~~	~~0.76~~	~~0.48~~	~~0.89~~	~~0.77~~
	~~Complex perception~~	~~10~~	~~0.65~~	~~0.41~~	~~0.88~~	~~0.70~~
	~~Language~~	~~10~~	~~0.84~~	~~0.68~~	~~0.93~~	~~0.84~~
	~~Thought~~	~~10~~	~~0.84~~	~~0.58~~	~~0.92~~	~~0.79~~
	~~Memory~~	~~10~~	~~0.86~~	~~0.72~~	~~0.90~~	~~0.72~~
	~~Motility~~	~~10~~	~~0.70~~	~~0.29~~	~~0.90~~	~~0.81~~
	~~Loss of automatization~~	~~10~~	~~0.86~~	~~0.68~~	~~0.89~~	~~0.35~~
	~~Anxiety/Anhedonia~~	~~10~~	~~0.79~~	~~0.49~~	~~0.83~~	~~0.59~~
	~~Sensory overstimulation~~	~~10~~	~~0.76~~	~~0.42~~	~~0.87~~	~~0.50~~
~~Süllwold (1991)~~	~~Disturbances of automated responses~~	~~22~~	~~0.91~~	~~0.73~~	~~0.96~~	~~0.80~~
	~~Perceptual disturbances~~	~~30~~	~~0.87~~	~~0.51~~	~~0.96~~	~~0.77~~
	~~Depression~~	~~27~~	~~0.92~~	~~0.30~~	~~0.95~~	~~0.58~~

	Overinclusion	18	0.84	0.45	0.93	0.57
Mass et al. (1997)	Dysphoric concomitants of severe illness particularly impairing concentration	36	0.95	0.65	0.97	0.83
	Subjective experiences of perceptual uncertainties	14	0.78	0.63	0.92	0.78
Mass et al. (1995)	Schizophrenia specific factor	8	0.54	0.09	0.81	0.47
	Alcohol dependency specific factor	9	0.72	0.38	0.82	0.54
Loas et al. (2002)	24 item scale	24	0.92	0.67	0.96	0.70
Yon et al. (2008)	FCQ total score	98	0.96	0.50	0.99	0.58

Note: FCQ = Frankfurt Complaint Questionnaire; ARMS = At-risk mental state; FEP = First episode psychosis; Cronbach's alpha (>0.9 = excellent, >0.8 = good, >0.7 = acceptable, >0.6 = questionable, >0.5 = poor and ≤0.5 = inacceptable internal consistency), Revelle's beta (<0.50 = poor, ≥0.70 = acceptable homogenecity)

Table 6: Predictive validities of the FCQ factors

Model	Factor	No transition Mean (SD) N=93	Transition Mean (SD) N=24	*p*-value*
Original	Loss of control	2.08 (1.94)	2.17 (2.12)	0.927
	Simple perception	0.89 (1.40)	0.84 (1.18)	0.843
	Complex perception	1.00 (1.19)	0.71 (0.96)	0.290
	Language	1.73 (2.15)	2.14 (2.27)	0.564
	Thought	4.20 (2.87)	4.36 (2.99)	0.784
	Memory	2.59 (2.63)	2.01 (1.85)	0.161
	Motility	1.26 (1.32)	1.43 (1.67)	0.846
	Loss of automatization	3.14 (2.67)	3.20 (2.87)	0.652
	Anxiety/Anhedonia	3.16 (2.36)	3.57 (2.44)	0.590
	Sensory overstimulation	2.69 (2.15)	3.08 (2.57)	0.391
Süllwold (1991)	Disturbances of automated responses	5.25 (4.78)	5.57 (4.37)	0.970
	Perceptual disturbances	3.17 (3.08)	3.07 (2.94)	0.875
	Depression	7.73 (5.93)	7.97 (6.19)	0.849
	Overinclusion	6.06 (3.91)	6.20 (4.12)	0.987
Mass et al. (1997)	Dysphoric concomitants of severe illness particularly impairing concentration	10.1 (8.40)	10.3 (8.13)	0.841
	Subjective experiences of perceptual uncertainties	1.01 (1.38)	0.79 (1.22)	0.549
Mass et al. (1995)	Schizophrenia specific factor	1.04 (1.07)	1.08 (1.41)	0.970
	Alcohol dependency specific factor	3.45 (2.09)	3.52 (2.28)	0.904
Loas et al. (2002)	24 item scale	6.61 (5.57)	7.04 (5.30)	0.996
Yon et al. (2008)	FCQ total score	22.7 (15.5)	23.5 (15.7)	0.941

Note: *Wald tests from univariate Cox proportional Hazard models

Index of abbreviations

APS:	Attenuated Psychotic Symptoms
ARMS:	At risk mental state
ARMS-T:	At Risk Mental State – Transitioned
ARMS-NT:	At Risk Mental State – Non Transitioned
BSABS:	Bonn Scale for the Assessment of Basic Symptoms
BSIP:	Basler Screening Instrument for early detection of Psychosis
BLIPS:	Brief Limitted Intermittend Psychotic Symptoms
BPRS:	Brief Psychiatric Rating Scale
BS:	Basic symptoms
BSEM	Bayes Structure Equation Modelling
CFA:	Confirmatory Factor Analysis
CFI:	Comparative Fit Index
CI:	Confidence Interval
COPER:	Cognitive Perceptive Basic Symptoms
COGDIS:	Cognitive Disturbances
FCQ:	Frankfurt Complaint Questionnaire
FEP:	First Episode Psychosis
***FePsy**:*	Basel Project of Early detection of Psychosis (German: ***F***rüherk***e***nnung von ***Psy***chosen)
EFA:	Exploratory Factor Analysis
PCA:	Principal Component Analysis
RMSEA:	Root Mean Square Error of Approximation
SANS:	Scale for the Assessment of Negative Symptoms
SPI-A / SPI-CY	Schizophrenia Pronessinstrumen – Adult / Children and Youth version
TLI:	Tucker-Lewis-Index
UHR:	Ultra High Risk
WLSMV:	Weighted Least Square Mean and Variance adjusted estimator
WRMR:	Weighted Root Mean Residual