Convergent and concurrent validity of the Frankfurt Complaint Questionnaire as a screener for psychosis risk

Chantal Michel[†,1], PhD.; Christine Kutschal; Benno G. Schimmelmann[1], Prof., MD; Frauke Schultze-Lutter[1], A./Prof.

[†] corresponding author

[1] University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bolligenstrasse 111 (Haus A), 3000 Bern 60, Switzerland

chantal.michel@kjp.unibe.ch, 0041 (0)31 932 8558

benno.schimmelmann@kjp.unibe.ch, 0041 (0)31 932 8554

frauke.schultze-lutter@kjp.unibe.ch, 0041 (0)31 932 8564

No funding was received in support of the presented study.

Acknowledgement: None.

Corresponding author:

Chantal Michel

University Hospital of Child and Adolescent Psychiatry and Psychotherapy,

Research Department

Bolligenstrasse 111, Haus A

3000 Bern 60, Switzerland

tel. 0041 (0)31 932 8558

fax. 0041 (0)31 932 8569

mail. chantal.michel@kjp.unibe.ch

Acknowledgement:
None.

Conflict of interest:
Drs. Michel and Schultze-Lutter declare that there are no conflicts of interest in relation to the subject of this study.
Prof. Schimmelmann has been a consultant and/or advisor to or has received honoraria from Eli Lilly, Janssen, Novartis, and Shire.

Contributors:
CM, BGS and FSL designed the study; CM and FSL managed the literature searches, undertook the statistical analyses, and wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Abstract

For the broad identification of a risk for a serious mental disorder, valid and reliable screeners are needed to detect those most likely benefitting from a time-consuming and costly in-depth clinical assessment. In the early detection of psychoses, multiple screeners for an ultra-high risk have already been suggested. Yet, no screener explicitly targets an increased risk according to the basic symptom (BS) criteria. We therefore explored the Frankfurt Complaint Questionnaire (FCQ) as a potential screener for BS in comparison to their gold-standard clinical assessment using the Schizophrenia Proneness Instrument (SPI-A/SPI-CY) by examining its convergent (agreement between screener and gold-standard, calculated by the overall percentages of agreement between FCQ and SPI-A/SPI-CY) and concurrent (degree to which a screener can identify individuals with the target condition, examined by diagnostic accuracy measures calculated using thresholds of receiver operating characteristic curves) validity. The sample consisted of 81 patients of a psychosis early detection service (41 with an at-risk mental state or psychosis, and 40 with a nonpsychotic mental disorder). Only two visual perception disturbances reached a beyond chance level of agreement between FCQ and SPI-A/SPI-CY. For the BS-criteria 'cognitive perceptive basic symptoms' and 'cognitive disturbances', only insufficient agreement between assessment techniques were detected with Cohen's kappa being 0.228 and 0.130, respectively, with an overestimation by the FCQ. Diagnostic likelihood ratios indicated only a clinically irrelevant increase in the probability of detecting BS criteria, thus, the concurrent validity of both the total of all and of only criteria-relevant FCQ-items was insufficient. Both concurrent and convergent validity of the FCQ were poor, and the FCQ dramatically overestimated clinician-assessed risk. Our results suggest that the FCQ should not be used as a screener for BS-criteria and that the convergent validity is not guaranteed on the basis of face validity alone, but has to be formally assessed.

Key words: Basic symptoms, convergent validity, concurrent validity, FCQ, SPI-A/SPI-CY

1. Introduction

An indicated prevention of first-episode psychosis has become increasingly important in mental health research for its potential to alter the course of psychotic disorders whose nature, i.e. neurobiological and psychosocial underpinnings, is still largely unknown (Fusar-Poli et al. 2013; Schmidt et al. 2015; Schultze-Lutter et al. 2015). This prevention approach selectively targets patients who already seek help for mental problems including early subclinical signs and symptoms which represent primarily risk indicators or early expressions of the disorder (Mrazek and Haggerty 1994). As the majority of first-episode psychotic disorders, and in particular schizophrenic disorders, are preceded by a prodromal phase of several years in that help might be already sought (Schultze-Lutter et al. 2010; Schaffner et al. 2012), they offer a good starting for an indicated prevention. Thus in the past two decades, efforts have been put into identifying potential sensitive and specific clinical high risk criteria of psychosis to be assessed in clinical interviews by trained mental health professionals (Schmidt et al. 2015; Schultze-Lutter et al. 2015; Fusar-Poli et al. 2015). In this framework, a "clinical high risk state" differs from a "prodromal state" that inevitably leads to the full-blown disorder in that it can remit – temporarily or persistently – or even become chronic (Lin et al. 2015). While different approaches have been studied, two major sets of clinical high risk criteria, are currently used to identify an increased psychosis-risk (Fusar-Poli et al. 2013, 2015; Schultze-Lutter et al. 2015): The ultra-high risk criteria that mainly rely on attenuated or transient psychotic symptoms, and the basic symptom (BS) criteria that rely on subjective disturbances in thought and perception processes (Table 1). Clinical high risk criteria are fulfilled if the defined requirements of onset and frequency are met, with no further differentiation of severity of risk. Of these, the European Psychiatric Association recommends to alternatively use one of the following three clinical high risk criteria for an early detection of psychoses: (1) at least any one attenuated or (2) at least any one transient psychotic symptom and (3) at least any two self-experienced and self-reported cognitive BS with a score of at least '3' (= weekly occurrence within the past 3 months) in the Schizophrenia Proneness Instrument (Schultze-Lutter et al., 2015). It is further recommended that their assessment in clinical interviews should only be carried out by trained specialists with sufficient experience in clinical high risk (Schultze-Lutter et al., 2015). Yet, the application of these semi-structured clinical interviews is time-consuming, requires training,

and is typically carried out in specialised centres to that patients are mainly referred (Fusar-Poli et al. 2013; Schultze-Lutter et al. 2013; Tandon et al. 2012).

Given the waste of resources associated with inadequate referral, the ability to conduct economic, valid and reliable pre-diagnostic screening of individuals with potentially increased psychosis-risk would be of great value. Several, mainly self-report instruments have been proposed to this end (Kline and Schiffman 2014) that are either adaptions of questionnaires originally developed to assess psychotic-like experiences within the community, such as the Community Assessment of Psychotic Experiences (Yung et al. 2009), or developed specifically for ultra-high risk criteria, such as the Prodromal Questionnaire (Ising et al. 2012) or PRIME Screen (Miller et al. 2004). Only two screeners include the assessment of some BS, the PROD-Screen (Heinimaa et al. 2003) and the Eppendorf Schizophrenia Inventory (Maß et al. 2000). Yet both instruments were not constructed with regard to BS criteria, and while the Eppendorf Schizophrenia Instrument only includes questions on four of the 14 criteria-relevant BS – mainly taken from the Frankfurt Complaint Questionnaire (FCQ; Süllwold 1991; Süllwold and Huber 1986), the PROD-Screen includes 6 statements, yet all of them are compound statements tapping into several criteria-relevant as well as criteria-irrelevant BS at once or even into BS and (attenuated) psychotic symptoms at once. Thus, unsurprisingly, both have not been studied for their ability to specifically detect BS criteria (Heinimaa et al. 2003; Maß et al. 2000).

-Table 1-

Furthermore, none of these screeners has yet been examined for its convergent validity, i.e., the degree to which an instrument actually measures the same phenomena as the gold standard instrument (Dawe et al. 2002; Stinson et al. 2006). The gold standard in detection of psychosis-risk is the assessment of clinical high risk criteria within a semi-structured clinical interview by mental health professionals trained in the respective instrument; in case of BS criteria, this is either the adult or the child and youth version of the Schizophrenia Proneness Instrument (Schultze-Lutter et al. 2015). Importantly, convergent validity is not guaranteed on the basis of face validity alone, but has to be formally assessed (Jaeschke et al. 1994).

BS are self-experienced subclinical disturbances in mentation including cognition and perception (Gross 1989; Huber 1966; Huber and Gross 1989). As BS are by definition subjective and

predominately remain private and apparent only to the affected person, their assessment in clinical interviews completely relies on patients' reports regarding the novelty and abnormality of their experiences. BS criteria may therefore be an even better and more suitable target for a self-report screener than the symptomatic ultra-high risk phenomena whose abnormal nature might not be perceived by the patient or might be hard to locate on the proposed psychosis continuum, i.e., might be hard to distinguish from normal experiences (Johns and van Os 2001; Fusar-Poli et al. 2014).

BS were first operationalised by a semi-structured clinical interview in the 142-item Bonn Scale for the Assessment of Basic Symptoms (Gross et al. 1987) with shorter measures, the Schizophrenia Proneness Instrument, Adult (SPI-A; Schultze-Lutter et al. 2007), and Child and Youth version (SPI-CY; Fux et al. 2013; Schultze-Lutter et al. 2012a), subsequently evolving from it (Schultze-Lutter et al. 2012b) and now being used as the gold standard of their assessment (Schultze-Lutter et al. 2015).

In close collaboration with Gerd Huber's work group developing the Bonn Scale for the Assessment of Basic Symptoms, the Frankfurt Complaint Questionnaire (FCQ; Süllwold 1991; Süllwold and Huber 1986) was developed by the work group of Lilo Süllwold as a self-report measure of the same subjective phenomena with its item formulations based on face validity. With 10 of the 14 criteria-relevant BS, the FCQ still offers – in theory – the most comprehensive self-rating assessment of BS criteria, although it has been repeatedly criticized for its lack of diagnostic validity for schizophrenia (e.g., Maß 2005). Furthermore, if the FCQ would indeed measure the same phenomena as the Bonn Scale for the Assessment of Basic Symptoms or its successors SPI-A/SPI-CY, its items should possess excellent convergent validity with the corresponding interview items, and, consequently, the FCQ should have at least satisfactory concurrent validity for a clinical high risk state according to BS criteria. The FCQ has already demonstrated good reliability and internal consistency (Loas et al. 2011; Yon et al. 2008), and therefore may well serve as a useful screener for BS criteria.

This study investigated (i) the convergent validity of BS using the FCQ compared to their clinical assessment using SPI-A/SPI-CY (the gold standard), and (ii) its concurrent validity i.e., the degree to which the FCQ can identify the individuals who currently have a clinician-determined psychosis-risk according to BS criteria.

2. Material and Methods

2.1. Sample

The sample comprised 81 patients aged 8-40 years of the Bern Early Recognition and Intervention Centre for Mental Crises (FETZ-Bern) (Table 2; demographics of the ≥16-year-olds are provided in Supplementary Table 1). Exclusion criteria were somatic and/or psychotropic drug-use related conditions that might account for mental problems, and insufficient reading skills in German. In total, 41 patients (50.6%) were either diagnosed with a psychotic disorder or fulfilled BS and/or ultra-high risk criteria. The remaining 40 patients (49.4%) met criteria for a nonpsychotic mental disorder or complained about only subthreshold nonpsychotic mental problems according to the Mini-International Neuropsychiatric Interview (Sheehan et al. 1998) (Table 2).

-Table 3-

2.2. Instruments

SPI-A/SPI-CY were used for the assessment of clinician-rated BS (henceforth BS-SPI) and BS criteria (Table 1), namely cognitive-perceptive BS (COPER) and cognitive disturbances (COGDIS). With each item assessing exactly one BS, SPI-A/SPI-CY rank BS on a severity scale according to the maximum frequency of their occurrence within the past three months ranging from 0 (absent=BS has not occurred in the past 3 months) to 6 (extreme=BS has occurred daily over some time within the past three months). Symptoms may also be rated as 7 (BS has always been present in same severity; trait), 8 (BS is definitely present, but its frequency of occurrence is unknown), and 9 (presence of BS can neither be unambiguously ruled in nor out). For the purpose of this study, BS-SPI were recoded according to their presence into binary items for better correspondence with the binary rated FCQ-items: 1 (presence) was assigned to scores between 1-8 (i.e., all scores that clearly indicate the presence of the BS) and 0 (absence) was assigned to scores of 0 and 9 (i.e., scores clearly indicating the absence of the BS or only its ambiguous presence).

Presence and absence of self-reported BS (henceforth BS-FCQ) were assessed using the FCQ that comprises of 98 partially overlapping items referring to identical phenomena (Tables 3 and 4). Each item is designed as a statement describing a certain complaint, and the patient is instructed to indicate whether he/she has experienced such a complaint lately – irrespective of its frequency – by ticking

either the Yes or the No box. Complaints only experienced in the past but not lately have to be indicated by ticking the Yes box and adding “formerly” next to it. For the focus of the SPI-A/SPI-CY on the past three months, these FCQ-items were considered as absent in the analyses to avoid a bias towards over-report of BS in the FCQ compared to SPI-A/SPI-CY by different assessment periods. Assignment of BS-SPI and BS-FCQ items followed the originally described item correspondence between the FCQ and Bonn Scale for the Assessment of Basic Symptoms (Süllwold and Huber 1986). Yet, as this description included only a subgroup of items however, item correspondences were complemented by Frauke Schultze-Lutter. (Tables 3 and 4).

2.3. Procedure

In the FETZ-Bern, patients are assessed for BS by well-trained clinical psychologists. All raters were initially trained (concordance rate with expert rating after 10 training sessions 91%) and, to further ensure high data quality over time, are continuously supervised in SPI-A/SPI-CY and SIPS assessments by Frauke Schultze-Lutter, expert in both assessments and main author of SPI-A/SPI-CY. Patients also complete a battery of questionnaires including the FCQ. All patients (and their parents/guardians in the case of minors) provided written informed consent to the scientific use of their anonymous clinical data. The local ethical committee had approved the study.

2.4. Statistical analysis

Using SPSS 21.0, the convergent validity was calculated by the overall percentages of agreement (concordance rate; CR) between the presence of BS-SPI and BS-FCQ. Where several FCQ-items corresponded to one BS-SPI, both single FCQ-items and a compound rating (presence of any one item) were examined. To control for effects of chance, agreement on the presence of symptoms and criteria was additionally calculated using Cohen’s kappa (κ) statistic. A disadvantage generally associated with use of κ is its dependence on the prevalence of an event (Byrt et al. 1993): κ tends to decrease when a response/event is rare, even if the CR is high. In the absence of a satisfactory mathematical solution to this problem, we followed the approach for the appraisal of κ suggested by Burn and Weir (2011) and, additionally, calculated the prevalence index. The prevalence index reports

values between -1 and 1, and is 0 when both responses are equally probable (i.e., their prevalence is 50%). According to Burn et al. (2009), $\kappa \geqslant 0.40$ and CR $\geqslant 75\%$ are considered clinically useful. When information is contradictory, such as when CR exceeds its threshold but κ falls below it, the prevalence index should be considered. With the prevalence index approaching an absolute value of 1, the likelihood of an underestimation of κ increases, and more attention should be paid to CR. Furthermore, concurrent validity (often reported as "discriminant" validity) and diagnostic accuracy of the FCQ for BS criteria were examined. The FCQ provides no information on symptom onset or frequency and so BS criteria had to be redefined to avoid a systematic bias resulting from the unequal consideration of onset and frequency in the assessment modes. To this, psychosis-risk according to SPI was re-defined as at least 1 of the binary BS included in COPER and/or at least 2 of those included in COGDIS. This re-calculation resulted in *n*=49 patients with COPER and *n*=19 with COGDIS.

To examine the concurrent validity of the FCQ-total and the FCQ criteria-relevant BS total (BS-criteria-FCQ), receiver operating characteristic curves were first calculated to detect the optimal threshold for discriminating 'BS-risk' from 'no BS-risk' patients. Using these thresholds, diagnostic accuracy measures (sensitivity, specificity, positive and negative predictive values, positive and negative likelihood ratios (LR)) were then calculated. LRs guided the estimation of concurrent validity for the availability of interpretation guidelines (Jaeschke et al. 1994) that are missing for other accuracy measures that can only be interpreted by less reliable rules-of-thumb (Boyko 1994; Jaeschke et al. 1994).

An age effect was examined by sensitivity analyses in that all analyses were additionally performed on a subsample $\geqslant$16-year-olds (*n*=62) and compared for a systematic bias with the overall results.

3. Results

3.1. Criteria-relevant basic symptoms

Of the 34 BS-FCQ that correspond to criteria-relevant BS-SPI, only two visual perception disturbances (FCQ 24, 29; Table 3) were unequivocally reliable in terms of showing both sufficient CR of $\geq 75\%$ and beyond-chance-agreement of $\kappa > 0.40$. Furthermore, two other visual perception

disturbances (FCQ 51, 84) might be reliably assessable with the FCQ, as they both possessed a sufficient CR and a low κ possibly caused by a high prevalence index. Altogether 15 BS-FCQ had to be regarded as an unreliable assessment despite demonstrating sufficient CRs; in these items, an insufficient κ was unlikely due to a high prevalence index. All remaining 15 BS-FCQ relevant for BS criteria showed no indication of being an adequate estimate of clinician-assessed BS-SPI as both their CRs and κ were below the threshold for clinical usefulness (Table 3).

3.2. Basic symptom criteria

Both BS criteria, COPER and COGDIS, were overestimated by the FCQ by about 30% and 60%. Both CRs and κ fell clearly below the threshold for clinical usefulness therefore (Table 3).

-Table 3-

3.3. Criteria-irrelevant basic symptoms

Of the 37 BS-FCQ not included in the BS criteria, none was unequivocally reliable in terms of showing both a sufficient CR and κ. One body perception disturbance, FCQ 18, might however reliably assess its corresponding BS-SPI (D.8), as it possessed good CR and low κ possibly caused by a high prevalence index (Table 4). The other body perception disturbance included in FCQ and SPI, FCQ 9, had to be regarded as unreliable despite its sufficient CR as its low κ value was unlikely caused by a high PI. All remaining 35 criteria-irrelevant BS-FCQ showed no indication of being an adequate estimate of BS-SPI, as both their CRs and κ fell below the thresholds for clinical usefulness (Table 4).

-Table 4-

3.4. Concurrent validity

Analyses of the area under the receiver operating characteristic curves of both the FCQ-total and BS criteria-FCQ revealed unsatisfactory areas under the curve of 0.737 (95% *CI*=0.623–0.851, $p<0.001$) and 0.758 (95% *CI*=0.651–0.865, $p<0.001$) with regards to identifying BS-risk patients. When emphasising non-exclusion of potential risk patients (i.e. high sensitivity), the best threshold for

differentiating 'BS-risk' from 'no BS-risk' were a FCQ-total ⩾5 (sensitivity=0.978, specificity=0.161) and BS criteria-FCQ ⩾1 (sensitivity=0.978, specificity=0.194). When trying to optimise both sensitivity and specificity (as reported for most screeners for UHR status), the best threshold were FCQ-total ⩾24 (sensitivity=0.733, specificity=0.645) and BS criteria-FCQ ⩾8 (sensitivity=0.644, specificity=0.677). The diagnostic accuracy measures of FCQ-total and BS criteria-FCQ were unsatisfactory (Table 5, Figure 1).

-Table 5 and Figure 1-

3.5. Age effect

Sensitivity analyses gave no indication of an age effect. In line with the overall results, convergent and concurrent validity were mainly insufficient (Supplementary Tables 2-4).

4. Discussion

The early detection of rare but severe conditions requires easy and broadly applicable screeners to guide the referral of appropriate patients to in-depth diagnostic assessment (Cadman et al. 1984). Psychosis is one such condition. In addition to reliability, an ideal screener would demonstrate excellent concurrent validity by (i) ruling in all patients with the target condition while (ii) ruling out a considerable proportion of those without it. In a practical sense then, a screener for psychosis-risk should therefore possess a sensitivity approaching 100%, a negative diagnostic LR≤0.1 that indicates a 'large and often conclusive' change from pre-screen to post-screen probability of the absence of psychosis-risk (Jaeschke et al. 1994), and a positive predictive value that is greatest in settings in which the prevalence of the condition is highest, i.e., greater in clinical settings than in community settings (Cadman et al. 1984). At the same time, but secondary to the above, a screener should also possess high specificity and a positive diagnostic LR≥5 that indicates at least a moderate increase in the pre-screen to post-screen risk probability (Jaeschke et al. 1994). The screen's discriminant accuracy should also not be largely mediated by confounding conditions, such as comorbid depressive or anxiety disorders (Janes and Pepe 2008), but its items/components should possess strong content and convergent validity (i.e., indeed measure the target condition) (Dawe et al. 2002). When tested

alongside the gold standard of a clinical interview, questionnaire items should highly correlate with their interview counterpart and final screen results should correlate highly with interview results (i.e., have excellent convergent validity) (Dawe et al. 2002). In addition, the screener should assess all aspects of the condition (i.e., possess good content validity). Since the assessment of BS criteria is time-consuming and requires intensive training, and BS assessments completely rely on the patient's self-experience even during a clinical interview, we explored whether the FCQ might be used as a screener for psychosis-risk according to BS criteria.

Surprisingly, the convergent validity of the assessment of BS by the FCQ was generally insufficient despite the subjective nature of BS and their general reliance on self-reporting (Schultze-Lutter 2009). Additionally, its concurrent validity with regard to BS criteria was not satisfactory. Although BS rely less on interviewer appraisal than attenuated positive symptoms, the FCQ's concurrent validity is only as good as that reported among self-report assessments of attenuated positive symptoms when a cut-off is chosen that optimises both sensitivity and specificity (Kline and Schiffman 2014). Thus, in line with reports on differences between self-report and clinical assessments of BS and attenuated positive symptoms (Granö et al. 2011; Maß et al. 1997; Ochoa et al. 2008; Schultze-Lutter et al. 2014; Yung et al. 2009), we found a significant overestimation of clinician-rated BS by self-rated BS.

Only the criteria-relevant BS 'changes in colour vision' and 'micropsia, macropsia' were unequivocally valid in terms of possessing both sufficient CR and κ. Interestingly, it was also a visual perception disturbance, 'blurred vision', that previously demonstrated the highest correlation between the interview (BSABS) and self-rating (FCQ) assessment (r=0.51) (Maß et al. 1997). This symptom however is one of the few criteria-irrelevant perception disturbances (Klosterkötter et al. 2001) and not included in SPI-A/SPI-CY; consequently, it was not analysed in our study. Furthermore, perceptual disturbances have previously shown the highest agreement between self-report and interview versions of the PROD-Screen (Granö et al. 2011). Contrary to our results however, acoustic perception disturbances in the PROD-Screen returned a much higher κ (0.964) than visual ones (0.453). However, PROD-Screen's acoustic perception disturbances also include hallucinations while its visual perception disturbances remain equivalent to perceptual distortions and are therefore comparable to BS (Granö et al. 2011). Consequently, our finding of κ=0.320 for visual perception disturbances is

more comparable to PROD-screen results than the finding of κ=0.249 for acoustic perception disturbances.

For all other BS included in both FCQ and SPI-A/SPI-CY, agreement between assessment modes was at most slight and therefore insufficient. The influence of item wording on convergent validity was sometimes impressive, for example, κ varied between 0.090 and 0.259 for the eight alternative FCQ-items assessing 'disturbances of receptive speech', and even between 0.053 and 0.301 for the three FCQ-items assessing 'disturbances of expressive speech'. The influence of wording on convergent validity has not yet been considered in the construction or translation of early detection screeners that were predominately guided by aspects of face validity. Convergent validity however might also affect the concurrent and predictive validity. The impact of different symptom formulations on convergent validity should therefore be examined for future screeners, their revisions and translations.

With regards to the two BS criteria, COPER and COGDIS, κ and CRs clearly indicated an insufficient clinical usefulness of the FCQ for their assessment; the FCQ overestimated the presence of COPER by 29.6% and that of COGDIS by 58%. Thereby, COGDIS exhibited a much larger discrepancy between the results of self-reports and interviews for a psychosis-risk status than the 29.9% discrepancy reported for the PROD-screen (Granö et al. 2011) which was comparable with the results for COPER. Even when examining a more criteria-independent assessment of the FCQ's concurrent validity, our findings were no more positive. Both negative and positive LRs were clearly insufficient according to their guidelines (Jaeschke et al. 1994) for both FCQ-total and the total of criteria-relevant BS-FCQ. Notably, our sensitivity and specificity values were reasonably comparable to those reported for screeners for ultra-high risk status (Kline and Schiffman 2014). This indicates that LRs as measures of concurrent validity should generally be taken into account to ensure proper validation of screeners.

An advantage of our study is that we not only evaluated the concurrent validity of the FCQ for distinguishing between risk and no-risk, but also examined the FCQ's convergent validity, i.e., the extent to that it actually measures what it is supposed to measure. To date, emphasis in the examination of mainly ultra-high risk screeners has been placed on concurrent validity irrespective of convergent validity. The neglect of convergent validity has arguably resulted in distorted reports however. For example, all three scales of the Community Assessment of Psychotic Experiences

(Stefanis et al. 2002) were considered accurate in discriminating ultra-high risk status (Hanssen et al. 2003; Konings et al. 2006; Mossenheb et al. 2012; Stefanis et al. 2002), but taken in context, only the positive symptoms scale (not the negative and depressive subscales) relates to ultra-high risk criteria. The sample size in this study was comparable to other screening evaluation studies in clinical samples (Kline and Schiffman 2014; Schultze-Lutter et al. 2014) and was also sufficient to detect a clinical useful κ against the null hypothesis of no agreement between assessment modes with a power of 90% (Sim et al. 2005). A clear disadvantage of the current study was however shown in the limited content validity of the FCQ as a result of the incomplete overlap between the BS included in FCQ and SPI-A/SPI-CY. This precluded an examination of the criteria-relevant BS of 'disturbances of abstract thinking', 'unstable ideas of reference', 'thought perseveration', 'derealisation', as well as of two acoustic and five visual perception disturbances.

In conclusion, despite their subjective nature and reliance on self-report, the assessment of BS by the FCQ appears to be no more valid than the self-report screeners of attenuated psychotic symptoms or ultra-high risk criteria and, particularly for lack of content validity, cannot be recommended as a psychosis-risk screener of BS criteria. With respect to the apparent influence of item wording on convergent validity, a future screener should first ensure items' convergent validity before investigating concurrent validity, i.e., the power to discriminate patients with and without BS criteria.

5. References

Boyko, E.J., 1994. Ruling out or ruling in disease with the most sensitive or specific diagnostic test: short cut or wrong turn? Medical Decision Making. 14 (2) 175-179.

Burn, C.C., Pitchard, J.C., Whay, H.R., 2009. Observer reliability for working equine walfare assessment: problems with high prevalences of certain results. Animal Walfare. 18 (2) 177-187.

Burn, C.C., Weir, A.A., 2011. Using prevalence indices to aid interpretation and comparison of agreement ratings between two or more observers. Veterinary Journal 188 (2) 166-170.

Byrt, T., Bishop, J., Carlin, J.B., 1993. Bias, prevalence and kappa. Jorunal of Clincal Epidemiology. 46 (5) 423-429.

Cadman, D., Chambers, L., Feldman, W., Sackett, D., 1984. Assessing the effectiveness of community screening programs. JAMA. 251 (12) 1580-1585.

Dawe, S., Loxton, N.J., Hide, L., Kavanagh, D.J., Mattick, R.P., 2002. Review of diagnostic screening instruments for alcohol and other drug use and other psychiatric disorders. Commonwealth of Australia, AU. http://immunise.health.gov.au/internet/main/publishing.nsf/Content/EB4EDA7A97F80941CA257BF000199DA8/$File/mono48.pdf. Accessed 11 May 2014.

Fusar-Poli, P., Borgwardt, S., Bechdolf, A., Addington, J., Riecher-Rössler, A., Schultze-Lutter, F., Keshavan, M., Wood, S., Ruhrmann, S., Seidman, L.J., et al, 2013. The psychosis high-risk state: a comprehensive state-of-the-art review. JAMA Psychiatry. 70 (1) 107-120.

Fusar-Poli, P., Yung, A.R., McGorry, P., van Os, J., 2014. Lessons learned from the psychosis high-risk state: towards a general staging model of prodromal intervention. Psychological Medicine. 44 (1) 17-24.

Fusar-Poli, P., Cappucciati, M., Rutigliano, G., Schultze-Lutter, F., Bonoldi, I., Borgwardt, S., Riecher-Rössler, A., Addington, J., Perkins, D., Woods, S.W., McGlashan, T.H., Klosterkötter, J., Yung, A.R., McGuire, P., 2015. At risk or not at risk? Meta-analysis of the prognostic accuracy of psychometric interviews for psychosis prediction. World Psychiatry. doi:10.1017/S003329171

Fux, L., Walger, P., Schimmelmann, B.G., Schultze-Lutter, F., 2013. The Schizophrenia Proneness Instrument, Child and Youth version (SPI-CY): practicability and discriminative validity. Schizophrenia Research. 146 (1-3) 69-78.

Granö, N., Karjalainen, M., Itkonen, A., Anto, J., Edlund, V., Heinimaa, M., Roine, M., 2011. Differential results between self-report and interview-based ratings of risk symptoms of psychosis. Early Intervention in Psychiatry. 5 (4) 309-314.

Gross, G., Huber, G., Klosterkötter, J., Linz, M., 1987. Bonner Skala für die Beurteilung von Basisymptomen (BSABS; Bonn Scale for the Assessment of Basic Symptoms). Springer-Verlag, Berlin.

Gross, G., 1989. The 'basic' symptoms of schizophrenia. Br. J. Psychiatry. 155 (Suppl 7) 21-25.

Hanssen, M., Peeters, F., Krabbendam, L., Radstake, S., Verdoux, H., van Os, J., 2003. How psychotic are individuals with non-psychotic disorders? Social Psychiatry and Psychiatric Epidemiology. 38 (3) 149-154.

Heinimaa, M., Salokangas, R.K., Ristkari, T., Plathin, M., Huttunen, J., Ilonen, T., Suomela, T., Korkeila, J., McGlashan, T.H., 2003. PROD-screen - a screen for prodromal symptoms of psychosis. International Journal of Methods in Psychiatric Research. 12 (2) 92-104.

Huber, G., 1966. Reine Defektsyndrome und Basisstadien endogener Psychosen. Fortschritte der Neurologie Psychiatrie. 34 409-426.

Huber, G., Gross, G., 1989. The concept of basic symptoms in schizophrenic and schizoaffective psychoses. Recenti Progressi in Medicina. 80 (12) 646-652.

Ising, H.K., Veling, W., Loewy, R.L., Rietveld, M.W., Rietdijk, J., Dragt, S., Klaassen, R.M., Nieman, D.H., Wunderink, L., Linszen, D.H., et al, M., 2012. The validity of the 16-item version of the Prodromal Questionnaire (PQ-16) to screen for ultra high risk of developing psychosis in the general help-seeking population. Schizophrenia Bulletin. 38 (6) 1288-1296.

Jaeschke, R., Guyatt, G.H., Sackett, D.L., 1994. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? JAMA. 271 (9) 703-707.

Janes, H., Pepe, M.S., 2008. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. American Journal of Epidemiology. 168 (1) 89-97.

Johns, L.C., van Os, J., 2001. The continuity of psychotic experiences in the general population. Clinical Psychological Review. 21 (8)1125-1141.

Kline, E., Schiffman, J., 2014. Psychosis risk screening: A systematic review. Schizophrenia Research. 158 (1-3) 11-18.

Klosterkötter, J., Hellmich, M., Steinmeyer, E.M., Schultze-Lutter, F., 2001. Diagnosing schizophrenia in the initial prodromal phase. Archives of General Psychiatry. 58 (2) 158-164.

Konings, M., Bak, M., Hanssen, M., van Os, J., Krabbendam, L., 2006. Validity and reliability of the CAPE: a selfreport instrument for the measurement of psychotic experiences in the general population. Acta Psychiatrica Scandinavica. 114 (1) 55-61.

Lin, A., Wood, S.J., Nelson, B., Beavan, A., McGorry, P., Yung, A.R., 2015. Outcomes of nontransitioned cases in a sample at ultra-high risk for psychosis. American Journal of Psychiatry. 172 (3) 249-258.

Loas, G., Yon, V., Monestès, J.L., Cuesta, M.J., 2011. Test-retest reliability of the Frankfurt Complaint Questionnaire. Psychological Reports. 108 (2) 503-506.

Maß, R., Hitschfeld, K., Wall, E., Wagner, H.B., 1997. Validität der Erfassung schizophrener Basissymptome. Nervenarzt. 68 (3) 205-211.

Maß, R., Haasen, C., Wolf, K., 2000. Das Eppendorfer Schizophrenie-Inventar (ESI) Entwicklung und Evaluation eines Fragebogens zur Erfassung charakteristischer Selbstwahrnehmungen kognitiver Dysfunktionen schizophren Erkrankter. Nervenarzt. 71 (11) 885-892.

Maß, R., 2005. Eppendorfer Schizophrenie-Inventar (ESI) vs. Frankfurter Beschwerde-Fragebogen (FBF). Nervenarzt. 76 (9) 1109-1116.

Miller, T.J., Cicchetti, D., Markovich, P.J., McGlashan, T.H., Wood, S.W., 2004. The SIPS Screen: a brief self-report screen to detect the schizophrenia prodrome. Schizophrenia Research. 70 (1) 78.

Mrazek, P.J., Haggerty, R.J., 1994. Reducing risk for mental disorders: Frontiers for preventive intervention research. National Academy Press Washington.

Ochoa, S., Haro, J.M., Torres, J.V., Pinto-Meza, A., Palacín, C., Bernal, M., Brugha, T., Prat, B., Usall, J., Alonso, J., Autonell, J., 2008. What is the relative importance of self reported psychotic symptoms in epidemiological studies? Results from the ESEMeD--Catalonia Study. Schizophrenia Research. 102 (1-3) 261-269.

Schaffner, N., Schimmelmann, B.G., Niedersteberg, A., Schultze-Lutter, F., 2012. Versorgungswege von erstmanifesten psychotischen Patienten – eine Übersicht internationaler Studien. Fortschritte der Neurologie, Psychiatrie. 80 (2) 72-78.

Schmidt, S.J., Schultze-Lutter, F., Schimmelmann, B.G., Maric, N.P., Salokangas, R.K.R., Riecher-Rössler, A., van der Gaag, M., Nordentoft, M., Raballo, A., Meneghelli, A., Marshall, M., Morrison, A., Klosterkötter, J., Ruhrmann, S., 2015. EPA guidance on the early intervention in clinical high risk states of psychoses. European Psychiatry. 30 (3) 388-404.

Schultze-Lutter, F., Addington, J., Ruhrmann, S., Klosterkötter, J., 2007. Schizophrenia Proneness Instrument, Adult version (SPI-A). Giovanni Fioriti Editore s.l.r. Rome.

Schultze-Lutter, F., 2009. Subjective symptoms of schizophrenia in research and the clinic: the basic symptom concept. Schizophrenia Bulletin. 35 (1) 5-8.

Schultze-Lutter, F., Ruhrmann, S., Klosterkötter, J., 2009. Early detection of psychosis - establishing a service for persons at risk. European Psychiatry. 24 (1) 1-10.

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. Schizophrenia Bulletin. 36 (1) 182-191.

Schultze-Lutter, F., Marshall, M., Koch, E., 2012a. Schizophrenia Proneness Instrument, Child and Youth version; Extended English Translation (SPI-CY EET). Giovanni Fioriti Editore s.l.r., Rome.

Schultze-Lutter, F., Ruhrmann, S., Fusar-Poli, P., Bechdolf, A., Schimmelmann, B.G., Klosterkötter, J., 2012b. Basic Symptoms and the Prediction of First-Episode Psychosis. Current Pharmaceutical Design. 18 (4) 351-357.

Schultze-Lutter, F., Schimmelmann, B.G., Ruhrmann, S., Michel, C., 2013. 'A rose is a rose is a rose', but at-risk criteria differ. Psychopathology. 46 (2) 75-87.

Schultze-Lutter, F., Renner, F., Paruch, J., Julkowski, D., Klosterkötter, J., Ruhrmann, S., 2014. Self-reported psychotic-like experiences are a poor estimate of clinician-rated attenuated and frank delusions and hallucinations. Psychopathology. 47 (3) 194-201.

Schultze-Lutter, F., Michel, C., Schmidt, S.J., Schimmelmann, B.G., Maric, N.P., Salokangas, R.K.R., 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. European Psychiatry. 30 (3) 405-416.

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured psychiatric interview for DSM-IV and ICD-10. Journal of Clinical Psychiatry. 59 (Suppl 20) 22-34.

Sim, J., Wright, C.C., 2005. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Physical Therapy. 85 (3) 257-268.

Stefanis, N.C., Hanssen, M., Smirnis, N.K., Avramopoulos, D.A., Evdokimidis, I.K., Stefanis, C.N., Verdoux, H., Van Os, J., 2002. Evidence that three dimensions of psychosis have a distribution in the general population. Psychological Medicine. 32 (2) 347-358.

Stinson, J.N., Kavanagh, T., Yamada. J., Gill, N., Stevens, B., 2006. Systematic review of the psychometric properties, interpretability and feasibility of self-report pain intensity measures for use in clinical trials in children and adolescents. Pain. 125 (1-2) 143-157.

Süllwold, L., Huber, G., 1986. Schizophrene Basisstörungen. Monographien aus dem Gesamt gebiete der Psychiatrie, Bd.42.: Springer, Berlin Heidelberg New York.

Süllwold, L., 1991. Manual zum Frankfurter Beschwerde-Fragebogen (FBF). Springer, Berlin Heidelberg New York.

Tandon, N., Shah, J., Keshavan, M.S., Tandon, R., 2012. Attenuated psychosis and the schizophrenia prodrome: current status of risk identification and psychosis prevention. Neuropsychiatry (London). 2 (4) 345-353.

Yon, V., Loas, G., Monestès, J.L., 2008. Dimensional structure of the Frankfurt Complaint Questionnaire in a sample of French students. Psychopathology. 41 (2) 85-89.

Yung, A.R., Nelson, B., Baker, K., Buckby, J.A., Baksheev, G., Cosgrave, EM., 2009. Psychotic-like experiences in a community sample of adolescents: implications for the continuum model of psychosis and prediction of schizophrenia. The Australian and New Zealand Journal of Psychiatry. 4 (2) 118-128.

Table 1: Basic symptom criteria

Risk criterion 'Cognitive-Perceptive Basic Symptoms' (COPER)

➲ At least any 1 of the following basic symptoms with a SPI-A score of ≥3 within the last 3 months:

- thought interference
- thought perseveration
- thought pressure
- thought blockages
- disturbance of receptive speech
- decreased ability to discriminate between ideas and perception, fantasy and true memories
- unstable ideas of reference
- derealisation
- visual perception disturbances (excl. hypersensitivity to light or blurred vision)
- acoustic perception disturbances (excl. hypersensitivity to sounds)

➲ First occurrence ≥12 months ago

High-risk criterion 'Cognitive Disturbances' (COGDIS)

➲ At least any 2 of the following basic symptoms with a SPI-A score of ≥3 within the last 3 months:

- inability to divide attention
- thought interference
- thought pressure
- thought blockages
- disturbance of receptive speech
- disturbance of expressive speech
- unstable ideas of reference
- disturbances of abstract thinking
- captivation of attention by details of the visual field

Table 2: Sociodemographic and clinical characteristics

	Patients (N=81)
Age in years: M±SD (range)	20±6(9-40)
Minors (<18 yrs.): n (%)	40 (49%)
Male gender: n (%)	46 (57%)
Current psychosocial functioning (SOFAS): Mdn (range)	65 (35-85)
Main clinical diagnosis according to ICD-10: n (%)	
Mental and behavioral disorders due to psychoactive substance use (F1)	2 (2%)
Schizophrenia, schizotypal and delusional disorders (F2)	8 (10%)
Mood (affective) disorders (F3)	17 (21%)
Neurotic, stress-related and somatoform disorders (F4)	4 (5%)
Disorders of adult personality and behavior (F6)	1 (1%)
Disorders of psychological development (F8)	1 (1%)
Behavioral and emotional disorders with onset usually occurring in childhood and adolescence (F9) and unspecified mental disorder (F9)	15 (19%)
Any 1 at-risk criterion[a]: n (%)	33 (41%)
COPER	17 (21%)
COGDIS	12 (15%)
APS	21 (26%)
BLIPS	2 (2%)

[a] Multiple criteria possible; COPER: cognitive-perceptive basic symptoms; COGDIS: cognitive disturbances; APS: attenuated psychotic symptoms; BLIPS: brief limited intermittent psychotic symptoms

Table 3: Criteria-relevant BS: Agreement on symptom presence of self-reported BS assessed with the 'Frankfurt Complaint Questionnaire' (FCQ) and clinician-rated BS assessed with the 'Schizophrenia Proneness Instrument' (SPI-A / SPI-CY), and on presence of risk criteria "cognitive-perceptive basic symptoms" (COPER) and "cognitive disturbances" (COGDIS)

FCQ items	*SPI-A / SPI-CY items*	*Correspondence rate*	*Kappa (κ)*[a]	*Prevalence index (PI)*[b]	*% present in FCQ*	*% present in SPI-A / SPI-CY*
FCQ 80 (n=81)	Inability to divide attention (B1 / D8)[c]	72.84%	0.314	0.531	37.04%	9.88%
FCQ 31, 66, 71(n=81; any 1)	Disturbance of expressive speech (C5 / D12)[d]	69.14%	0.215	0.519	35.80%	12.35%
FCQ 31 (n=81)		76.54%	0.053	0.716	14.81%	12.35%
FCQ 66 (n=81)		74.07%	0.301	0.568	28.40%	12.35%
FCQ 71[e] (n=81)		76.54%	0.251	0.642	20.99%	12.35%
FCQ 32 (n=81)	Captivation of attention by details of the visual field (O7 / O2)[d]	79.01%	0.120	0.741	20.99%	4.94%
FCQ 13, 54 (any 1; n=81)	Thought interference (C2 / D9)[c,d]	48.15%	0.137	0.259	62.96%	11.11%
FCQ 13 (n=81)		56.79%	0.190	0.346	54.32%	11.11%
FCQ 54 (n=79)		69.62%	0.152	0.570	31.65%	11.39%
FCQ 43, 70 (any 1; n=81)	Thought blockages (C3 / D15) [c,d]	58.02%	0.215	0.062	60.49%	33.33%
FCQ 43 (n=79)		59.49%	0.179	0.185	48.10%	32.91%
FCQ 70 (n=79)		63.29%	0.236	0.228	44.30%	32.91%
FCQ 37, 40, 69, 82, 90, 93, 94 (any 1; n=81)	Disturbance of receptive speech (C4 / D11) [c,d]	34.57%	0.051	0.222	71.60%	6.17%
FCQ 37 (n=80)		57.50%	0.090	0.475	46.25%	6.25%
FCQ 40 (n=80)		78.75%	0.244	0.688	25.00%	6.25%
FCQ 69 (n=79)		65.82%	0.181	0.532	40.51%	6.33%
FCQ 82 (n=79)		62.03%	0.113	0.519	41.77%	6.33%
FCQ 90 (n=79)		68.35%	0.098	0.608	32.91%	6.33%
FCQ 93 (n=79)		72.15%	0.180	0.620	31.65%	6.33%
FCQ 94 (n=79)		79.75%	0.259	0.696	24.05%	6.33%
FCQ 2, 36, 85 (any 1; n=81)	Thought pressure (D3 / D10) [c,d]	43.21%	0.099	0.185	67.90%	13.58%
FCQ 2 (n=81)		59.26%	0.207	0.346	51.85%	13.58%

FCQ 36 (n=79)		54.43%	0.113	0.342	51.90%	13.92%
FCQ 85[e] (n=79)		62.96%	0.161	0.543	38.27%	13.92%
FCQ 76 (n=81)	Decreased ability to discriminate between ideas and perception, fantasy and true memories (O2 / B1)[c]	75.31%	0.179	0.654	25.93%	8.64%
FCQ 63, 47, 14, 19, 84, 24, 29, 45, 50, 51, 79 (any 1; n=81)	Visual perception disturbances (O4 / B3)[c]	65.82%	0.320	0.177	50.63%	31.65%
FCQ 63 (n=78)	Partial seeing including tubular vision (O4.10 / O.1)	88.46%	0.248	0.833	10.26%	6.41%
FCQ 47 (n=79)	Photopsia (F2 / O3)	81.01%	0.256	0.709	20.25%	8.86%
FCQ 14 (n=81	Changed perception of the face or body of others (D5 / B3.5)	88.89%	0.353	0.815	13.58%	4.94%
FCQ 19 (n=81)	Metamorphopsia (O4.2 / B3.3)	85.19%	0.105	0.827	14.81%	2.47%
FCQ 84 (n=79)		90.12%	-0.035	0.901	5.06%	2.53%
FCQ 24[e] (n=81)	Changes in colour vision (O4.3 / B3.4)	90.12%	0.501	0.778	12.35%	9.88%
FCQ 29 (n=79)	Micropsia, macropsia (F3 / B3.2)	93.67%	0.581	0.835	7.59%	8.86%
FCQ 45 (n=79)	Disturbances of the estimation of distances or sizes (O4.7 / B3.9)	84.81%	0.171	0.797	12.66%	7.59%
FCQ 50[e] (n=78)	Pseudomovements of optic stimuli (O4.5 / B3.7)	84.62%	0.066	0.790	11.54%	6.41%
FCQ 51 (n=78)		91.03%	0.175	0.885	5.13%	6.41%
FCQ 79 (n=76)		90.79%	0.079	0.776	10.53%	6.58%
FCQ 25, 53, 72 (any 1; n=80)	Acoustic perception disturbances (O5 / B5)[c]	72.50%	0.249	0.575	33.75%	8.75%
FCQ 25 (n=80)	Changed intensity/quality of acoustic stimuli (F5 / B5.1)	83.75%	0.299	0.738	17.5%	8.75%
FCQ 53 (n=78)		76.92%	0.257	0.617	26.92%	8.97%
FCQ 72 (n=78)		87.18%	0.305	0.795	11.54%	8.97%
COPER (≥1 BS of 7)[f,g]	**COPER (≥1 BS of 7)[f,g]**	**67.90%**	**0.228**	**0.506**	**90.12%**	**60.49%**
COGDIS (≥2 BS of 7)[f,g]	**COGDIS (≥2 BS of 7)[f,g]**	**41.98%**	**0.130**	**0.049**	**81.48%**	**23.46%**

[a] Evaluation guidelines for κ: 0.00–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1.00 = almost perfect agreement

[b] A high prevalence index represents a low prevalence rate, whereas a low prevalence index represents a high prevalence rate.

[c] Symptom is part of COPER

[d] Symptom is part of COGDIS

[e] Item was not part of the initially described correspondence between FCQ and BSABS by Süllwold & Huber (1986); the assignment was made by F.S.L.

[f] If BS is assessed in the FCQ by several items, at least any 1 of them was rated as presence of BS.

[g] The following BS (SPI-A number / SPI-CY number) that are part of COPER or COGDIS are not assessed in the FCQ: disturbances of abstract thinking (O3 / D7), unstable ideas of reference (D4 / B2), thought perseveration (O1 / D14), derealization (O8 / B7), and some of the acoustic and visual perception disturbances.

Table 4: Unspecific BS: Agreement on symptom presence of self-reported BS assessed with the 'Frankfurt Complaint Questionnaire' (FCQ) and clinician-rated BS assessed with the 'Schizophrenia Proneness Instrument (SPI-A / SPI-CY)

FCQ items	SPI-A / SPI-CY items	Correspondence rate	Kappa (κ)[a]	Prevalence index (PI)[b]	% present in FCQ	% present in SPI-A / SPI-CY
FCQ 16 (n=81)	Change in mood, emotional responsiveness (A2 / A5)	72.84%	0.441	0.210	66.67%	55.56%
FCQ 87 (n=76)	Decreased capacity to discriminate between different kinds of emotions (D1 / D13)	70.37%	0.164	0.605	23.68%	11.84%
FCQ 59, 96 (any 1;n=33)	Disturbance in presenting oneself (- / A8)	63.64%	0.092	0.613	39.39%	3.03%
FCQ 56, 57 (any 1; n=79)	Impaired tolerance to unusual, unexpected or specific novel demands (A1.1 / A4.2)	40.51%	0.024	0.000	73.41%	26.58%
FCQ 61, 89, 97 (any 1; n=79)	Impaired tolerance to certain social everyday situations (A1.2 / A4.3)	59.49%	0.187	0.139	49.37%	36.71%
FCQ 39, 98 (any 1; n=79)	Difficulties concentrating (B3 / A11)	67.09%	0.283	0.291	67.09%	62.03%
FCQ 8, 78, 83[c] (any 1; n=40)	Disturbance in retrieving knowledge from long-term memory (- / D6)	45.00%	0.000	0.450	55.00%	0.0%
FCQ 48[c], 73[c] (any 1; n=79)	Difficulties holding things in mind for less than an hour (B4 / D5)	62.03%	0.275	0.063	58.23%	35.44%
FCQ 27 (n=39)	Disturbance of the comprehension of visual or acoustic stimuli (- / B6)	66.67%	0.090	0.615	35.90%	2.56%
FCQ 10[c], 58, 65[c] (any 1; n=80)	Feeling overly distracted by stimuli (B2 / D16)	62.50%	0.257	0.075	60.00%	47.50%
FCQ 20[c], 22, 33, 81 (any 1; n=80)	Motor interference exceeding simple lack of co-ordination (O9 / D17)	56.25%	-0.025	0.563	42.50%	1.25%
FCQ 7[c], 34, 86 (any 1; n=80)	Motor blockages (O10 / D18)	52.50%	0.050	0.475	50.00%	2.50%
FCQ 6, 11, 17, 38, 44, 46, 77, 95 (any 1; n=80)	Loss of automatic skills (O11 / D19)	32.50%	0.031	0.263	71.25%	3.75%
FCQ 9 (n=81)	Somatopsychic bodily depersonalization (F6 / B8.2)	85.19%	0.173	0.802	12.35%	7.41%
FCQ 18 (n=40)	Bodily sensations of abnormal heaviness, lightness, emptiness, falling, sinking, levitation or elevation (-/B8.6)	87.50%	-0.042	0.875	10.00%	2.50%

[a] Evaluation guidelines for κ: 0.00–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1.00 = almost perfect agreement

[b] A high prevalence index represents a low prevalence rate, whereas a low prevalence index represents a high prevalence rate

[c] Item was not part of the initially described correspondence between FCQ and BSABS by Süllwold & Huber (1986); the assignment was made by F.S.L.

Table 5: Diagnostic accuracy of the cut-off of the FCQ total score and of the sum score of FCQ criteria-relevant BSs

FCQ threshold	*Sensitivity*	*Specificity*	*Positive Predictive Value*	*Negative Predictive Value*	*Positive likelihood ratio(LR^+)*[a]	*Negative likelihood ratio(LR^-)*[a]
FCQ total score ≥24	0.673	0.656	0.750	0.568	1.956	0.498
FCQ criteria-relevant BS ≥8	0.578	0.710	0.743	0.537	1.993	0.594

[a] According to the guidelines (Jaeschke et al. 1994) a LR^+ of <3 and a LR^- of >0.3 generate small and rarely important changes in pre-test probabilities.

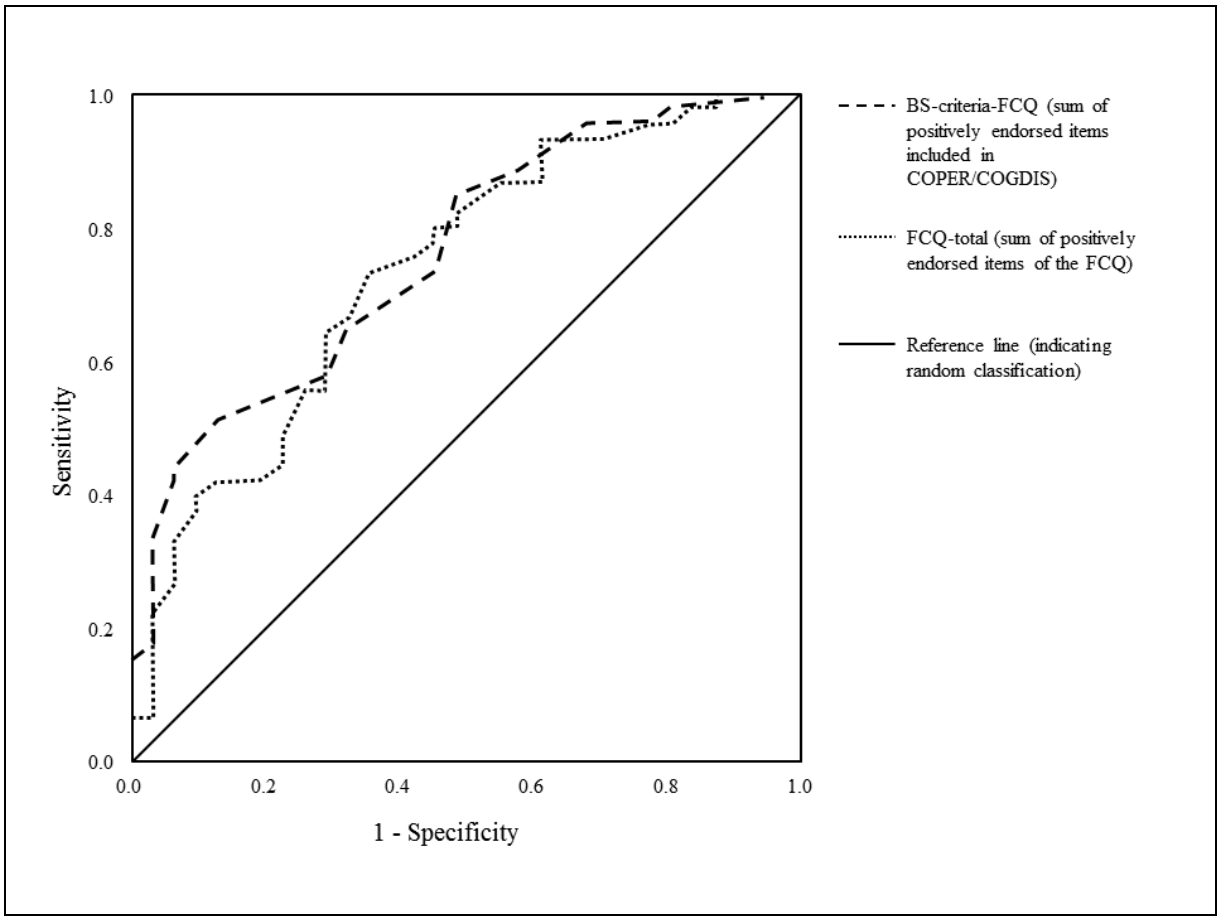

Fig. 1: Overall ability of the FCQ total score and of the total score of FCQ-assessed basic symptoms included in COPER/COGDIS for discriminating patients who had met COPER/COGDIS in the clinical interview from those who did not: Areas under the ROC curve.

Supplementary Table 1: Sociodemographic and clinical characteristics of the subsample of ≥16-year-olds

	Patients (N=62)
Age in years: M±SD (range)	22±5(16-40)
Male gender: n (%)	37 (60%)
Current psychosocial functioning (SOFAS): Mdn (range)	61(35-85)
Main clinical diagnosis according to ICD-10: n (%)	
Mental and behavioral disorders due to psychoactive substance use (F1)	2 (3%)
Schizophrenia, schizotypal and delusional disorders (F2)	7 (11%)
Mood (affective) disorders (F3)	14 (23%)
Neurotic, stress-related and somatoform disorders (F4)	3 (5%)
Disorders of adult personality and behavior (F6)	1 (2%)
Disorders of psychological development (F8)	1 (2%)
Behavioral and emotional disorders with onset usually occurring in childhood and adolescence (F9) and unspecified mental disorder (F9)	4 (6%)
Any 1 at-risk criterion[a]: n (%)	26 (42%)
COPER	16 (26%)
COGDIS	11 (18%)
APS	15 (24%)
BLIPS	2 (3%)

[a] Multiple criteria possible; COPER: cognitive-perceptive basic symptoms; COGDIS: cognitive disturbances; APS: attenuated psychotic symptoms; BLIPS: brief limited intermittent psychotic symptoms

Supplementary Table 2: Criteria-relevant BS: Agreement on symptom presence of self-reported BS assessed with the 'Frankfurt Complaint Questionnaire' (FCQ) and clinician-rated BS assessed with the 'Schizophrenia Proneness Instrument' (SPI-A / SPI-CY), and on presence of risk criteria "cognitive-perceptive basic symptoms" (COPER) and "cognitive disturbances" (COGDIS) for the total sample (n=81) and a subsample of ≥16-year-olds (n=62).

FCQ items Total sample (n=81) Only 16-40 years (n=62)	*SPI-A / SPI-CY items*	*Correspondence rate*	*Kappa (κ)*[a]	*Prevalence index (PI)*[b]	*% present in FCQ*	*% present in SPI-A / SPI-CY*
FCQ 80 (n=81)	Inability to divide attention (B1 / D8)[c]	72.84% 75.81%	0.314 0.402	0.531 0.500	37.04% 37.10%	9.88% 12.90%
FCQ 31, 66, 71(n=81; any 1)	Disturbance of expressive speech (C5 / D12)[d]	69.14% 69.35%	0.215 0.257	0.519 0.467	35.80% 37.10%	12.35% 16.13%
FCQ 31 (n=81)		76.54% 74.19%	0.053 0.046	0.716 0.677	14.81% 16.12%	12.35% 16.12%
FCQ 66 (n=81)		74.07% 77.42%	0.301 0.369	0.568 0.548	28.40% 29.03%	12.35% 16.12%
FCQ 71[e] (n=81)		76.54% 77.42%	0.251 0.281	0.642 0.613	20.99% 22.58%	12.35% 16.12%
FCQ 32 (n=81)	Captivation of attention by details of the visual field (O7 / O2)[d]	79.01% 75.81%	0.120 0.041	0.741 0.726	20.99% 22.58%	4.94% 4.84%
FCQ 13, 54 (any 1; n=81)	Thought interference (C2 / D9)[c,d]	48.15% 48.39%	0.137 0.151	0.259 0.226	62.96% 64.52%	11.11% 12.90%
FCQ 13 (n=81)		56.79% 53.23%	0.190 0.182	0.346 0.274	54.32% 59.68%	11.11% 12.90%
FCQ 54 (n=79)		69.62% 72.58%	0.152 0.231	0.570 0.565	31.65% 30.65%	11.39% 12.90%
FCQ 43, 70 (any 1; n=81)	Thought blockages (C3 / D15) [c,d]	58.02% 58.06%	0.215 0.214	0.062 0.000	60.49% 62.90%	33.33% 37.10%
FCQ 43 (n=79)		59.49% 58.06%	0.179 0.161	0.185 0.129	48.10% 50.00%	32.91% 37.10%
FCQ 70 (n=79)		63.29% 59.68%	0.236 0.173	0.228 0.177	44.30% 45.16%	32.91% 37.10%
FCQ 37, 40, 69, 82, 90, 93, 94 (any 1; n=81)	Disturbance of receptive speech (C4 / D11) [c,d]	34.57% 33.87%	0.051 0.059	0.222 0.177	71.60% 74.19%	6.17% 8.06%

FCQ 37 (n=80)		57.50% 54.84%	0.090 0.097	0.475 0.419	46.25% 50.00%	6.25% 8.06%
FCQ 40 (n=80)		78.75% 79.03%	0.244 0.294	0.688 0.661	25.00% 25.81%	6.25% 8.06%
FCQ 69 (n=79)		65.82% 61.29%	0.181 0.182	0.532 0.452	40.51% 46.77%	6.33% 8.06%
FCQ 82 (n=79)		62.03% 61.29%	0.113 0.132	0.519 0.484	41.77% 43.55%	6.33% 8.06%
FCQ 90 (n=79)		68.35% 62.90%	0.098 0.085	0.608 0.532	32.91% 38.71%	6.33% 8.06%
FCQ 93 (n=79)		72.15% 69.35%	0.180 0.190	0.620 0.565	31.65% 35.48%	6.33% 8.06%
FCQ 94 (n=79)		79.75% 79.03%	0.259 0.294	0.696 0.661	24.05% 25.81%	6.33% 8.06%
FCQ 2, 36, 85 (any 1; n=81)	Thought pressure (D3 / D10) [c,d]	43.21% 43.55%	0.099 0.111	0.185 0.113	67.90% 70.97%	13.58% 17.75%
FCQ 2 (n=81)		59.26% 59.68%	0.207 0.241	0.346 0.274	51.85% 54.84%	13.58% 17.74%
FCQ 36 (n=79)		54.43% 51.61%	0.113 0.107	0.342 0.258	51.90% 56.45%	13.92% 17.74%
FCQ 85[e] (n=79)		62.96% 64.52%	0.161 0.189	0.543 0.419	38.27% 40.32%	13.92% 17.74%
FCQ 76 (n=81)	Decreased ability to discriminate between ideas and perception, fantasy and true memories (O2 / B1)[c]	75.31% 70.97%	0.179 0.171	0.654 0.581	25.93% 30.65%	8.64% 11.29%
FCQ 63, 47, 14, 19, 84, 24, 29, 45, 50, 51, 79 (any 1; n=81)	Visual perception disturbances (O4 / B3)[c]	65.82% 69.35%	0.320 0.401	0.177 0.145	50.63% 53.23%	31.65% 31.75%
FCQ 63 (n=78)	Partial seeing including tubular vision (O4.10 / O.1)	88.46% 90.16%	0.248 0.349	0.833 0.836	10.26% 9.84%	6.41% 6.56%
FCQ 47 (n=79)	Photopsia (F2 / O3)	81.01% 83.87%	0.256 0.371	0.709 0.710	20.25% 20.97%	8.86% 8.06%
FCQ 14 (n=81	Changed perception of the face or body of others (D5 / B3.5)	88.89% 88.71%	0.353 0.315	0.815 0.822	13.58% 12.90%	4.94% 4.84%

FCQ 19 (n=81)	Metamorphopsia (O4.2 / B3.3)	85.19% 85.48%	0.105 0.136	0.827 0.823	14.81% 14.52%	2.47% 3.23%
FCQ 84 (n=79)		90.12% 91.94%	-0.035 -0.040	0.901 0.919	5.06% 4.84%	2.53% 3.23%
FCQ 24[e] (n=81)	Changes in colour vision (O4.3 / B3.4)	90.12% 88.71%	0.501 0.523	0.778 0.726	12.35% 14.52%	9.88% 12.90%
FCQ 29 (n=79)	Micropsia, macropsia (F3 / B3.2)	93.67% 95.16%	0.581 0.703	0.835 0.823	7.59% 6.45%	8.86% 11.29%
FCQ 45 (n=79)	Disturbances of the estimation of distances or sizes (O4.7 / B3.9)	84.81% 83.87%	0.171 0.197	0.797 0.774	12.66% 12.90%	7.59% 9.68%
FCQ 50[e] (n=78)	Pseudomovements of optic stimuli (O4.5 / B3.7)	84.62% 80.33%	0.066 0.042	0.790 0.770	11.54% 14.75%	6.41% 8.20%
FCQ 51 (n=78)		91.03% 88.52%	0.175 0.161	0.885 0.839	5.13% 6.56%	6.41% 8.20%
FCQ 79 (n=76)		90.79% 83.05%	0.079 0.075	0.776 0.758	10.53% 11.86%	6.58% 8.20%
FCQ 25, 53, 72 (any 1; n=80)	Acoustic perception disturbances (O5 / B5)[c]	72.50% 72.58%	0.249 0.237	0.575 0.597	33.75% 33.87%	8.75% 6.45%
FCQ 25 (n=80)	Changed intensity/quality of acoustic stimuli (F5 / B5.1)	83.75% 87.10%	0.299 0.446	0.738 0.742	17.5% 19.35%	8.75% 6.45%
FCQ 53 (n=78)		76.92% 77.42%	0.257 0.219	0.617 0.677	26.92% 25.81%	8.97% 6.45%
FCQ 72 (n=78)		87.18% 87.10%	0.305 0.371	0.795 0.774	11.54% 16.13%	8.97% 6.45%
COPER (≥1 BS of 7)[f,g]	**COPER (≥1 BS of 7)[f,g]**	**67.90%** **70.96%**	**0.228** **0.242**	**0.506** **0.548**	**90.12%** **90.32%**	**60.49%** **64.52%**
COGDIS (≥2 BS of 7)[f,g]	**COGDIS (≥2 BS of 7)[f,g]**	**41.98%** **45.16%**	**0.130** **0.146**	**0.049** **0.129**	**81.48%** **83.87%**	**23.46%** **29.03%**

[a] Evaluation guidelines for κ: 0.00–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1.00 = almost perfect agreement

[b] A high prevalence index represents a low prevalence rate, whereas a low prevalence index represents a high prevalence rate.

[c] Symptom is part of COPER

[d] Symptom is part of COGDIS

[e] Item was not part of the initially described correspondence between FCQ and BSABS by Süllwold & Huber (1986); the assignment was made by F.S.L.

[f] If BS is assessed in the FCQ by several items, at least any 1 of them was rated as presence of BS.

[g] The following BS (SPI-A number / SPI-CY number) that are part of COPER or COGDIS are not assessed in the FCQ: disturbances of abstract thinking (O3 / D7), unstable ideas of reference (D4 / B2), thought perseveration (O1 / D14), derealization (O8 / B7), and some of the acoustic and visual perception disturbances.

Supplementary Table 3: Unspecific BS: Agreement on symptom presence of self-reported BS assessed with the 'Frankfurt Complaint Questionnaire' (FCQ) and clinician-rated BS assessed with the 'Schizophrenia Proneness Instrument (SPI-A / SPI-CY) for the total sample (n=81) and a subsample of ≥16-year-olds (n=62).

FCQ items Total sample Only 16-40 years	*SPI-A / SPI-CY items*	*Correspondence rate*	*Kappa (κ)*[a]	*Prevalence index (PI)*[b]	*% present in FCQ*	*% present in SPI-A / SPI-CY*
FCQ 16 (n=81)	Change in mood, emotional responsiveness (A2 / A5)	72.84% 72.58%	0.441 0.415	0.210 0.274	66.67% 58.06%	55.56% 69.35%
FCQ 87 (n=76)	Decreased capacity to discriminate between different kinds of emotions (D1 / D13)	70.37% 76.27%	0.164 0.100	0.605 0.695	23.68% 20.34%	11.84% 10.17%
FCQ 59, 96 (any 1;n=33)	Disturbance in presenting oneself (- / A8)	63.64% 55.56%	0.092 0.000	0.613 0.556	39.39% 44.44%	3.03% 0.0%
FCQ 56, 57 (any 1; n=79)	Impaired tolerance to unusual, unexpected or specific novel demands (A1.1 / A4.2)	40.51% 40.32%	0.024 -0.003	0.000 0.016	73.41% 72.58%	26.58% 29.03%
FCQ 61, 89, 97 (any 1; n=79)	Impaired tolerance to certain social everyday situations (A1.2 / A4.3)	59.49% 58.06%	0.187 0.168	0.139 0.032	49.37% 53.23%	36.71% 43.55%
FCQ 39, 98 (any 1; n=79)	Difficulties concentrating (B3 / A11)	67.09% 66.13%	0.283 0.237	0.291 0.339	67.09% 69.35%	62.03% 64.52%
FCQ 8, 78, 83[c] (any 1; n=40)	Disturbance in retrieving knowledge from long-term memory (- / D6)	45.00% 38.10%	0.000 0.000	0.450 0.381	55.00% 20.97%	0.0% 0.0%
FCQ 48[c], 73[c] (any 1; n=79)	Difficulties holding things in mind for less than an hour (B4 / D5)	62.03% 62.90%	0.275 0.289	0.063 0.016	58.23% 59.68%	35.44% 38.71%
FCQ 27 (n=39)	Disturbance of the comprehension of visual or acoustic stimuli (- / B6)	66.67% 61.90%	0.090 0.125	0.615 0.524	35.90% 42.86%	2.56% 4.76
FCQ 10[c], 58, 65[c] (any 1; n=80)	Feeling overly distracted by stimuli (B2 / D16)	62.50% 60.56%	0.257 0.234	0.075 0.016	60.00% 59.02%	47.50% 42.62%
FCQ 20[c], 22, 33, 81 (any 1; n=80)	Motor interference exceeding simple lack of co-ordination (O9 / D17)	56.25% 52.50%	-0.025 -0.033	0.563 0.574	42.50% 40.98%	1.25% 1.64%
FCQ 7[c], 34, 86 (any 1; n=80)	Motor blockages (O10 / D18)	52.50% 49.18%	0.050 0.056	0.475 0.426	50.00% 54.10%	2.50% 3.28%
FCQ 6, 11, 17, 38, 44, 46, 77, 95 (any 1; n=80)	Loss of automatic skills (O11 / D19)	32.50% 34.43%	0.031 0.042	0.263 0.246	71.25% 70.49%	3.75% 4.92%
FCQ 9 (n=81)	Somatopsychic bodily depersonalization (F6 / B8.2)	85.19%	0.173	0.802	12.35%	7.41%

		83.87%	0.203	0.774	14.52%	8.06%
FCQ 18 (n=40)	Bodily sensations of abnormal heaviness, lightness, emptiness, falling, sinking, levitation or elevation (-/B8.6)	87.50% 90.48%	-0.042 0.000	0.875 0.905	10.00% 9.52%	2.50% 0.0%

[a] Evaluation guidelines for κ: 0.00–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1.00 = almost perfect agreement

[b] A high prevalence index represents a low prevalence rate, whereas a low prevalence index represents a high prevalence rate

[c] Item was not part of the initially described correspondence between FCQ and BSABS by Süllwold & Huber (1986); the assignment was made by F.S.L.

Supplementary Table 4: Diagnostic accuracy of the cut-off of the FCQ total score and of the sum score of FCQ criteria-relevant BSs for the total sample (n=81) and a subsample of ≥16-year-olds (n=62).

FCQ threshold	*Sensitivity*	*Specificity*	*Positive Predictive Value*	*Negative Predictive Value*	*Positive likelihood ratio(LR^+)*[a]	*Negative likelihood ratio(LR^-)*[a]	*Relative risk*
FCQ total score ≥24	0.673	0.656	0.750	0.568	1.956	0.498	0.577
16-40 years	0.750	0.682	0.811	0.429	2.358	0.367	0.493
FCQ criteria-relevant BS ≥8	0.578	0.710	0.743	0.537	1.993	0.594	0.624
16-40 years	0.632	0.727	0.800	0.533	2.315	0.506	0.583

[a] According to the guidelines (29) a LR^+ of <3 and a LR^- of >0.3 generate small and rarely important changes in pre-test probabilities.