Title:
Discrete neurocognitive subgroups in fully or partially remitted bipolar disorder**: associations with functional abilities**

Authors:
Johan Høy Jensen[a], Ulla Knorr[a], Maj Vinberg[a], Lars V. Kessing[a], Kamilla W. Miskowiak[a, 1]

Affiliation:
[a] Psychiatric Centre Copenhagen, Copenhagen University Hospital, Rigshospitalet Dep. 6233, Blegdamsvej 9, DK-2100, Copenhagen, Denmark, Emails: johan.hoey.jensen@regionh.dk, ulla.knorr@regionh.dk, maj.vinberg@regionh.dk, lars.vedel.kessing@regionh.dk, kamilla.miskowiak@regionh.dk

1. Introduction

Bipolar disorder (BD) is a severe and chronic illness characterized by recurrent phases of mood swings found in about 3% of the population (Kessler et al., 2005). Although BD has often been regarded as a purely episodic illness, research within the past two decades highlights persistent cognitive and functional impairment in-between the acute mood episodes (Arts et al., 2008; Harvey et al., 2010). Persistent cognitive deficits in remitted patients with BD are well-documented across several cognitive domains including attention, verbal learning, and executive function (Bora et al., 2009; Reichenberg et al., 2009). There is robust evidence from several studies that patients' persistent cognitive dysfunction is a key contributor to their socio-occupational disability independent of mood symptoms (Depp et al., 2012; Martinez-Arán et al., 2007; Mur et al., 2009).

[1] Corresponding author: Kamilla W. Miskowiak, Psychiatric Centre Copenhagen, Copenhagen University Hospital, Rigshospitalet Dep. 6233, Blegdamsvej 9, DK-2100, Copenhagen, Denmark, Tel.: +45 3864 7087, Fax: +45 3864 7077 Email: kamilla@miskowiak.dk

Functional impairment is prevalent in BD with unemployment rates substantially higher in BD than the general population as studies show between 4- and 10-fold increase in unemployment among BD patients (Huxley and Baldessarini, 2007; Kogan et al., 2004). Further, approximately two thirds of the patients are unable to regain premorbid levels of social and vocational functioning following a single episode (Huxley and Baldessarini, 2007). A recent meta-analysis found that verbal memory and executive function were moderately related to employment outcome (Tse et al., 2014). **Cognitive impairment is among the strongest contributors together with mood symptoms and illness progression to functional disability, lower quality of life, and loss of workforce capacity in BD (Bonnín et al., 2010; Brissos et al., 2008; Torrent et al., 2012). However, the correlation between subjectively experienced and objectively measured cognitive impairment is poor, as shown by our and other research groups (Burdick, Endick and Goldberg, 2005; Jensen et al., 2015; Rosa et al., 2013).** This points to **objectively measured** cognitive dysfunction as a key treatment priority to improve patients' functional recovery **and quality of life after acute mood episodes**.

Meta-analytic findings indicate that the nature of the persistent cognitive dysfunction in BD is non-specific, involving deficits across several domains with moderate to large effect sizes (Bourne et al., 2013). **Several studies have examined cognition in BD using predetermined cut-offs (typically 1-2 standard deviations [SD] from the normal mean [M]) for simple classification of impaired and non-impaired patients (e.g., Jensen et al., 2015; Martino et al., 2014; Reichenberg et al., 2009; Rojo et al., 2010; Volkert et al., 2015). However, there is a scarcity of studies that have used a *data-driven* approach to identify neurocognitive subgroups. Such studies found discrete neurocognitive subgroups in fully or partially remitted** patients with BD **(Bora et al., 2016; Burdick et al., 2014; Lewandowksi et al., 2014)**; a **well performing** "cognitively intact" subgroup with scores equivalent to the performance of healthy control persons (HCs), **one or two** subgroups **of** "selective cognitive impairment" **with lower cognition scores compared to HCs**, and a

subgroup with “global severe impairment” across cognitive domains comparable to cognitive deficits in schizophrenia. **The findings from these studies show some discrepancies with respect to the proportion of patients in each subgroup, possibly due to differences in study samples (e.g., affective symptoms, age, medicine prescribed).** Additional **data-driven subgroup** studies are needed to make any firm conclusions about the pattern of cognitive in BD.

1.1 Aims of the study

The study aims to **identify** discrete neurocognitive subgroups in **a large group of** fully or partially remitted patients with BD **using a data-driven approach**. Building onto the **few studies applying this approach**, the **present** study aims to examine the pattern of the cognitive deficits and the clinical and functional characteristics of the neurocognitive subgroups in terms of perceived stress, cognitive complaints, work **and social adjustment**, quality of life, and medication prescribed.

2. Materials and methods

2.1 Pooling of data

The present study involved pooling of **available *baseline* data from four studies of our research group: two clinical trials targeting cognition (study 1: Miskowiak et al., 2014; study 2: Demant et al., 2015) and from two cross-sectional observational studies of which one has been completed (study 3: Jensen et al., 2015; Ott et al., 2016) and one is ongoing (study 4)** (BD: *N*=201 [study 1=46, study 2=46, study 3=84, study 4=25]; HC: *N*=110 [study 3=86, study 4=24]). Eight patients were excluded due to missing data (BD: *N*=193). We chose to pool data from studies 1-4 because of the largely similar recruitment criteria of BD patients in full or partial remission and large overlap between the applied measures of neurocognitive and functional capacity. Studies 1, 2, and 4 were approved by the Regional Ethics Committee in the Capital Region of Copenhagen. The local ethics committee stated that there was no need for their approval of study 3.

2.2 Participants and screening

A total of 303 individuals aged 18-65 participated in one of the aforementioned studies between September 2009-July 2015, comprising; 193 eligible adult patients with an ICD-10 diagnosis of BD in full or partial remission (defined as Hamilton Depression Rating Scale 17-item [HDRS-17] **scores of ≤7 or 8≤14, respectively, and** Young Mania Rating Scale [YMRS] scores of ≤7 **or** 8≤14, respectively) (Hamilton, 1960; Young et al., 1978), and 110 HCs. **All patients were diagnosed by specialists in psychiatry at the Copenhagen Clinic for Affective Disorders, Psychiatric Centre Copenhagen, and diagnoses of patients from studies 1 and 2 were confirmed using the Schedules for Clinical Assessment in Neuropsychiatry (Wing et al., 1990).** The HCs were recruited from the blood banks at Copenhagen University Hospital, Rigshospitalet (study 3) or Frederiksberg Hospital (study 4).
The patients were referred by psychiatric specialists in Clinic for Affective Disorders, Psychiatric Centre Copenhagen (studies 1-4) or recruited through website advertisements (study 1).
Exclusion criteria **in studies 1-4** included prior history of schizophrenia, current substance abuse, substantial somatic illness or a daily use of benzodiazepines ≥22.5 mg oxazepam. In studies 1 and 2 additional participation criteria for the BD patients included significant subjective cognitive difficulties according to the Massachusetts General Hospital Cognitive and Physical Functioning Questionnaire (CPFQ; scores ≥4 on ≥2 cognitive domains) (Fava et al., 2009). In study 4, patients had HDRS-17 and YMRS scores of ≤7, and baseline data were obtained following 2 weeks of remission. All participants in studies 2 and 4 were aged 18-50. Exclusion criteria for the HCs **(studies 3 and 4)** were dyslexia or any personal or family history of mental illness. Written informed consent was obtained from all participants.

2.3 Neurocognitive tests

Overlapping neuropsychological tests from studies 1-4 include: the Trail Making Test part A (TMT-A) (Army Individual Test Battery, 1944), Rey Auditory Verbal Learning Test (RAVLT) (Corwin and Bylsma, 1993; Rey, 1964), Wechsler's Adult Intelligence Scale, 3rd edition (WAIS-III) Letter-Number-Sequencing (Wechsler, 1997), Repeatable Battery for the Assessment of

Neuropsychological Status (RBANS) Digit Span Forward (Randolph, 1998), and verbal fluency (Borkowski et al., 1967) with the letters S and D.

*2.4 Measures of work **and social adjustment**, subjective cognitive difficulties, perceived stress, and quality of life*

All participants were instructed to complete a set of questionnaires regarding perceived stress (using the 10-item self-report Perceived Stress Scale [PSS]) (Cohen et al., 1983), cognitive and physical functioning (using the 7-item self-report CPFQ) (Fava et al., 2009), work **and social adjustment** (using the 5-item self-rating Work and Social Adjustment Scale [WSAS]) (Mundt et al., 2002), and quality of life (using the 26-item self-report WHOQOL) (WHOQOL Group, 1994). **We chose these questionnaires of functional ability since all measures were applied in studies 1-4.**

2.5 Statistical analyses

2.5.1 Cognitive domain and composite cognition scores

For an easy evaluation of cognitive function, all neurocognitive test scores of the participants were standardized to *z*-scores (***M***=0, ***SD***=1) based on the cognitive performance of the HCs using following formula: (*test score* - *HC test M*) / *HC test SD* (Field, 2013). **Although there is still some controversy regarding the particular cutoff for cognitive impairment (limits range from 1-2 *SDs* below the normal *M*; Bora et al., 2016; Martino et al., 2014), we chose in the present study to set the cut-off score of *z*≥1 below the *M* of the HC sample to define cognitive impairment consistent with Burdick et al. (2014) and our previous approach (Jensen et al., 2015; Ott et al., 2016).** Extreme *z*-scores (≥4 *SD*s below the HCs' *M*) were truncated to *z*=-4.0 **to allow scores with different degree of "severe" cognitive impairment (i.e., scores of >3 *SD*s below normative *M*) (Gualtieri and Morgan, 2008; Royall et al., 2007)**. TMT-A *z*-scores were inversed such that lower scores reflect poorer performance consistent with the direction of the other cognition measures. Using the cognition *z*-scores, we established three cognitive domain *z*-scores for each participant, which again, were standardized (using same formula as above) to the cognitive

domain *z*-scores of HCs: processing speed (*z*-scores for TMT-A), verbal learning (averaged *z*-scores for RAVLT: total recall across five learning trials (I-V), recall following interference (IV), and recall following 30 min. delay), and working memory and executive skills (averaged z-scores for WAIS-III Letter-Number Sequencing, RBANS Digit Span Forward, verbal fluency with letters S and D). **We chose this assignment of neuropsychological tests into cognitive domains given some consistency of this domain grouping in the literature (Lezak et al., 2012) and** to be consistent with previous studies **from our group** (Jensen et al., 2015; Ott et al., 2016). Finally, an overall composite cognition *z*-score was established for each participant by averaging the *z*-scores of the three domains and standardizing this output based on the HCs' composite cognition *z*-scores (same formula as above).

2.5.2 Establishment and validation of neurocognitive subgroups

Initially, independent *t*-tests and χ^2-tests were computed between the complete BD and HC samples to detect differences regarding demographical, clinical, cognitive, functional, and medicine characteristics. Then, a hierarchical cluster analysis (HCA) was conducted with the cognitive domain *z*-scores of the complete BD sample to detect homogeneous neurocognitive subgroups. Ward's linkage was selected as agglomeration procedure, and squared Euclidean distance was chosen to compute similarities between cases. The dendogram was visually inspected to detect the optimal number of clusters explaining the cognitive variance. To validate the clusters retained, a discriminant function analysis (DFA) was conducted since this analysis examines the predictive power of each participant's cognitive domain scores to the neurocognitive subgroup. One-way analyses of variance (ANOVA) and least significance difference (LSD) as post-hoc pairwise comparison were applied to compare the cognition scores of the BD subgroups and HCs. We chose this approach to keep the analyses consistent with the **study** by Burdick et al. (2014) **of comparison reasons**.

2.5.3 Comparisons of neurocognitive subgroups: clinical and functional characteristics

Post-hoc comparisons of the clinical and functional measures of between the HCA-defined neurocognitive subgroups were conducted to gain insight into the implication of cognitive impairment in BD. In line with previous studies from our group (Demant et al., 2015; Jensen et al., 2015; Ott et al., 2016), the CPFQ cognition measure was created by averaging the *z*-scores for item d (focus), e (recall), and a composite measure composed of *z*-scores for items f (word finding) and g (mental acuity) to ensure that each domain weighted equally in the CPFQ cognition measure. All data for the PSS, CPFQ cognition measure, and WHOQOL were standardized to *z*-scores based on the HCs using the same procedure as for the cognitive data (see above). This was not possible for the WSAS since this questionnaire addresses psychiatric patients only.

Clinical, demographical, and prescribed medicine variables include: gender, age, years of education, onset age, illness duration, depressive episodes and (hypo)-manic episodes (categorization: 1, 2 or ≥3 episodes), medication status (lithium, anticonvulsants, antidepressants, antipsychotics, benzodiazepines, melatonin [yes/no]), total number of medications prescribed, and HDRS-17 and YMRS scores. Between-group comparisons of the neurocognitive subgroups were applied with the aforementioned variables and functional data including the WSAS total scores (i.e., work **and social adjustment**), PSS total scores (i.e., perceived stress), CPFQ cognition measure (i.e., subjective cognitive difficulties), and WHOQOL (i.e., quality of life) total and domain scores using ANOVA and LSD post-hoc pairwise comparison or χ^2-tests as appropriate.

All statistical analyses were performed using the Statistical Package for Social Sciences (SPSS; version 22.0; IBM Corporation, Armonk, New York). Statistical significance for all tests were set to an alpha-level of $p<.05$ (two-tailed).

3. Results

3.1 Comparisons of samples

Comparisons between the complete BD and HC samples showed no differences regarding age, gender or years of education ($p \geq .10$) (for details see Table 1). Comparisons between the BD cohorts from the four original studies revealed subtle differences regarding age, age of onset, illness duration, affective symptoms, TMT-A scores, and medication (LSD: p-values$\leq .049$; see Supplementary Table).

3.2 Three neurocognitive subgroups

Visual inspection of the dendogram provided evidence for existence of three neurocognitive subgroups **in** the complete BD sample (data not shown): one subgroup of 89 patients (46.1%) **with "intact" cognitive function**, another subgroup of 63 patients (32.6%) **with "selective" cognitive impairment,** and a third subgroup of 41 **patients (21.2%) with "global" cognitive difficulties** (see Table 2). The DFA revealed presence of two discriminant functions explaining 77% and 23% of the variance (Wilks' λ=.19, χ^2=317.28, p<.01 and Wilks' λ=.60, χ^2=95.55, p<.01, respectively). **Processing speed and verbal memory contributed most to classify BD patients into neurocognitive subgroups (Standardized Coefficients for Function 1: verbal memory =.14, working memory and executive skills=.01, processing speed=.98; Standardized Coefficients for Function 2: Verbal memory =.72, working memory and executive skills=.69, processing speed=-.27).** The 90% of original grouped patients were correctly classified with the DFA, indicating the validity of the three neurocognitive subgroups (see Figure 1 for graphical agglomeration of the neurocognitive subgroups).

The ANOVA analyses showed that cognition measures between subgroups and the HCs were significantly different: composite cognition score (F(3)=99.45, p<.01), verbal learning (F(3)=36.12, p<.01), working memory and executive skills (F(3)=28.52, p<.01), and processing speed (F(3)=140.46, p<.01). Patients in the first (intact) subgroup were cognitively preserved compared to the HCs (range: z=-.11 to .32). LSD *post-hoc* pairwise comparisons between the intact patient subgroup and HCs showed no significant differences in verbal learning, working memory and

executive skills, and the composite cognition score (LSD: $p \geq .19$). Intact patients had significantly higher scores on processing speed than HCs (LSD: $p=.01$). In the second (selectively impaired) subgroup, verbal learning, and working memory and executive skills were within the normal range ($z=-.71$ and $z=-.62$, respectively), but processing speed and, consequently, the overall composite cognition score were severely impaired ($z=-2.53$ and $z=-2.06$, respectively). The selective subgroup scored significantly lower on all cognition measures compared to the intact group (LSD: $p<.01$). The third (globally impaired) group displayed impaired performance on verbal learning ($z=-1.75$), working memory and executive skills ($z=-1.44$), and, consequently, the composite cognition score ($z=-2.02$) while processing speed was **significantly slower compared to HCs (LSD: $p<.01$) but still** within the normal range ($z=-.61$). Post-hoc comparisons between the neurocognitive subgroups showed that all cognition scores of the globally impaired subgroup were significantly lower compared to the intact group (LSD: $p<.01$). Comparison analyses of the global and the selective subgroup revealed greater reduction in processing speed within the selective group (LSD: $p<.01$), and greater impairment of verbal learning and working memory and executive skills in the globally impaired group (LSD: p-values$<.01$), while there was no statistical difference in the composite cognition score (LSD: $p \geq .86$) (see Figure 2 and Table 3).

3.3 Post-hoc validation of the neurocognitive subgroups

As a *post-hoc* validation of the neurocognitive subgroups identified in the HCA, we conducted multiple regression analyses for the complete BD sample with the three cognitive domains, in turn, as dependent variables. Following variables were entered as independent variables **using forced entry**: neurocognitive subgroup (intact/selective/global: 1/2/3), lithium, antipsychotics, and antidepressants (yes/no), benzodiazepines, age, gender, HDRS-17 score, YMRS score, illness duration, number of depressive episodes, manic episodes, number of medications from different classes. The significant predictors for impaired verbal learning ($F(11)=11.52$, $p<.01$, $r^2=.45$, adjusted $r^2=.41$) were categorization in **the selectively or globally impaired subgroup** ($\beta=-.55$,

p<.01) and being male (β=-.32, p<.01) (other p-values ≥.09). The significant predictors of working memory and executive dysfunction (F(11)=9.35, p<.01, r^2=.39, adjusted r^2=.35) were categorization in **the selectively or globally impaired subgroup** (β=-.56, p<.01), more depressive symptoms (β=-.148, p=.04), and a lower number of medications prescribed (β=.21, p=.01) (other p-values ≥.09). Finally, the only significant predictor for slower processing speed (F(11)=4.72, p<.01, r^2=.25, adjusted r^2=.20) was categorization in **the selectively or globally impaired subgroup** (β=-.37, p<.01).

3.4 Comparisons between the BD subgroups: clinical characteristics

Since we found significant age differences between BD subgroups in the complete patient sample (F(3)=4.41, p<.01; LSD: p<.01), we conducted *post-hoc* comparisons of the BD subgroups' clinical characteristics using analyses of covariance (ANCOVA) with adjustment for age and LSD as pairwise comparison.

Age-adjusted comparison analyses for cognition scores among the neurocognitive subgroups and HC sample yielded similar results from comparison analyses without adjustment for age (significant in non-adjusted analysis: age-adjusted *p*-values≤.02; not significant in non-adjusted analysis: age-adjusted *p*-values≥.24). No significant differences were found between the three neurocognitive subgroups regarding characteristics, illness load (age of onset, illness duration, mood symptoms, number of previous mood episodes) or BD subtype (p≥.064) (see Table 3). Globally and selectively impaired patients reported poorer work **and social adjustment** (F(2)=3.94, p=.02; LSD: p=.03 and p=.01, respectively), more stress (LSD: p<.01 and p=.01, respectively), and more subjective cognitive difficulties (LSD: p<.01 and p=.02, respectively) than the cognitively intact patients. Selectively impaired patients **were prescribed** a greater number of medications (F(2)=7.25, p=.008; LSD: p=.01 and p<.01, respectively), and were more frequently prescribed

antidepressants ($F(2)=7.13$, $p=.03$; LSD: $p=.04$ and $p=.02$, respectively) and benzodiazepines ($F(2)=7.13$, $p=.03$; LSD: $p=.04$ and $p<.01$, respectively) relative to the other two neurocognitive subgroups. Moreover, patients in the selectively impaired subgroup had lower quality of life due to the environment than the intact patients ($F(3)=3.38$, $p=.04$; LSD: $p=.02$) (see Table 3). **Variance inflation values for all cognition and functional measures were <3, suggesting there should be no concern for multicollinearity in the analyses (Myers, 1990).**

3.5 Post-hoc analyses: characteristics of cognitively impaired patients

For exploratory purposes, we combined the selectively and globally impaired subgroups into one group of 'cognitively impaired' patients who were compared with those who were cognitively intact using ANCOVA adjusted for age. This revealed that **the combined subgroup of selectively and globally impaired patients** had fewer years of education ($F(1)=5.44$, $p=.02$), poorer work **and social adjustment** (higher WSAS scores: $F(1)=7.92$, $p=.01$), more perceived stress (lower PSS scores: $F(1)=10.00$, $p<.01$), greater subjective cognitive difficulties (higher CPFQ cognition scores: $F(1)=8.30$, $p<.01$), lower quality of life (lower WHOQOL scores: $F(1)=6.71$, $p=.01$), and were prescribed a greater number of medications ($F(1)=7.23$, $p=.01$) compared with the cognitively intact group. **These either selectively or globally** impaired patients were also more commonly prescribed antipsychotics ($\chi^2(1)=5.43$, $p=.02$) and benzodiazepines ($\chi^2(1)=9.92$, $p<.01$) compared to those who were cognitively intact.

4. Discussion

The study investigated the presence and clinical characteristics of discrete neurocognitive subgroups in **a large cohort** of fully or partially remitted BD patients. Three neurocognitive subgroups were identified; a cognitively intact subgroup (46.1%), a selectively impaired subgroup

with severe deficits in processing speed (32.6%), and a globally impaired subgroup with substantial difficulties within verbal learning, working memory, and executive skills (21.2%). Compared to the cognitively intact subgroup, patients with global impairment reported greater stress, more subjective cognitive difficulties, poorer work **and social adjustment**, and reduced quality of life in the absence of any differences in other clinical characteristics including age **of onset**, illness load, or medications. The selectively impaired group also displayed higher levels of stress, more subjective cognitive difficulties, poorer work **and social adjustment**, and reduced quality of life compared to the intact group; however, these impairments were generally less pronounced than for globally impaired patients. The selective subgroup was older and was prescribed a greater number of medications, including benzodiazepines and antidepressants relative to the other two neurocognitive subgroups.

The identification of three discrete neurocognitive subgroups in BD using a data-driven approach is in line with findings previous findings (Burdick et al., 2014), although Bora et al. (2016) Lewandowski et al. (2014) found two subgroups of "selective cognitive impairment". However, Bora et al. (2016) examined executive function solely in fully remitted BD patients (HDRS-17 and YMRS scores ≤7), and Lewandowski et al. (2014) included patients with schizophrenia or schizo-affective disorder along with BD patients in their analysis, which limits comparison to the present findings. The finding of this large cognitively intact subgroup of BD patients in full or partial remission is in line with previous evidence for an absence of cognitive impairment in 32-42% of patients (Burdick et al., 2014; Reichenberg et al., 2009; **Volkert et al., 2015**). Selective severe impairment in processing speed in one third of the BD cohort also corroborates with the previously published findings (Burdick et al., 2014) in a similar albeit smaller cohort of 136 BD patients in full or partial remission (HDRS-17 scores of ≤12, and Clinician-

Administered Rating Scale for Mania scores of ≤8) (Altman et al., 1994; Burdick et al., 2014; Hamilton, 1960). Although the present evidence for global impairment in a neurocognitive subgroup of BD patients corresponds to findings of Burdick et al. (Burdick et al., 2014), the proportion of the globally impaired group was substantially smaller (21.2% versus 39.7%) (Burdick et al., 2014). This may be due to differences in the two patient cohorts with our cohort being somewhat younger (*M*[*SD*], 36[10] versus 42[15], respectively) and showing less depressive symptoms (HDRS-17: *M*[*SD*], 6[4] versus 11[9], respectively) **(Samamé, Martino and Strejilevich, 2014; Volkert et al., 2015)**.

The present and previous (Burdick et al., 2014) evidence for an absence of differences between the globally impaired and intact subgroup in age **of onset**, severity of affective symptoms, illness load, BD subtype or medication suggests that the observed deficits in verbal memory and executive dysfunction in the globally impaired subgroup cannot be explained by such clinical characteristics. **This pattern is supported by Bora et al. (2016), although the global subgroup was found to be older, have fewer years of education, longer illness duration, and more commonly given antipsychotics, suggesting that executive function may be sensitive to these factors.** However, we were unable to examine whether cognitive deficits reflect neuroprogression due to lack of detailed information about the exact number of previous affective episodes; 84% and 72% of our BD sample had experienced at least three previous depressive episodes or (hypo-) manic episodes, respectively. **Indeed, Bora et al. (2016) found no statistically significant difference between subgroups with respect to previous depressive, manic, or total number of mood episodes.** Future prospective studies with long follow-up times are thus needed to clarify the nature and developmental trajectory of neurocognitive deficits in BD **(Dias et al., 2012)**.

The selectively impaired subgroup with severe reduction in psychomotor speed was characterized by older age, more polypharmacy**, and more antidepressants and benzodiazepines prescribed**

than both the intact and globally impaired subgroups. **Indeed, age cannot be attributed to differences in cognition between the neurocognitive subgroups since age-adjusted and non-adjusted comparison analyses of cognitive function between subgroups showed similar results.** Considerable evidence indicates that **polypharmacy, benzodiazepines, some antidepressants (mainly tricyclic antidepressants), high plasma levels of lithium** and antipsychotics may have negative effects on psychomotor speed (Bora et al., 2009**; Dias et al., 2012; Goldberg, 200**8), suggesting that the selectively impaired patients' greater use of medications may to some extent explain their slowed psychomotor speed. However, post-hoc multiple regression analysis of the complete BD sample revealed no significant correlations between slowing of processing speed and **medication, polypharmacy or benzodiazepines alone**, suggesting that the relationship between psychomotor slowing and medication is more complex.

The globally impaired subgroup with deficits in verbal learning, working memory, and executive skills showed no differences from the other neurocognitive subgroups with respect to illness load, BD subtype or medication. These findings point to trait-related deficits within verbal memory and executive function that is inherent to BD itself (rather than being explained by medication or external factors). Indeed, this interpretation is consistent with the study by Bora et al. (2016) and meta-analyses of deficits within verbal learning and executive function but not in processing speed in individuals at genetic risk for BD (Bora et al., 2009). The observation that work **and social adjustment**, stress, and quality of life were more negatively affected in of the globally and selectively impaired – but otherwise relatively symptom-free – subgroups compared to patients who were cognitively intact **is somewhat in line with previous reports (Brissos et al., 2008; Burdick et al. 2014; Martino et al., 2014). Burdick et al. (2014) found the global and the selective subgroup to be more occupationally disabled, but found no differences between the subgroups on residential or social functioning. The present results** highlight the clinical relevance of targeting cognitive deficits to improve functional recovery of a

significant proportion of patients with BD. Intensive research efforts have recently been made to find new treatments for cognitive dysfunction in BD. Although several pharmacological and psychological interventions are promising (Miskowiak et al., 2014; Torrent et al., 2013), the evidence is still preliminary. A key methodological problem in cognition trials which may partially explain the many negative findings in the field is the general lack of *objective* neuropsychological screening for cognitive deficits before inclusion of participants in these trials **(Burdick et al., 2015)**. Given our finding that almost half of patients are relatively cognitively intact this is likely to have led to inclusion of a large proportion of patients with little scope for cognitive improvement in these cognition trials. We therefore recommend systematic neuropsychological screening for cognitive dysfunction in BD before commencing a treatment trial targeting cognition to ensure inclusion of an enriched sample of patients with scope for improvement.

A limitation of the study was the cross-sectional study design since this hampers causal inferences regarding the associations between the nature of bipolar disorder itself, cognitive impairment, and functional outcome. Future long-term prospective studies of neurocognitive dysfunction in BD are therefore critically needed. Moderate to severe subjective cognitive difficulties (scores of ≥4 on ≥2 items on the CPFQ) was an inclusion criteria in studies 1 and 2 (i.e., 44% of the present BD sample). This may have introduced an overrepresentation of patients with neurocognitive deficits compared to the general BD population. However, speaking against this, studies by our and other groups have generally found no or only a poor correlation between objective and subjective measures of cognition in remitted BD **(Burdick, Endick and Goldberg, 2005;** Jensen et al., 2015; Rosa et al., 2013). Indeed, we found a larger subgroup of cognitively intact patients **(46.1%)** than previously published **by Burdick et al. (31.6%; 2014)**. **Moreover, post-hoc comparison analyses between studies 1 and 2 (having cognitive complaints as inclusion criterion) versus studies 3 and 4 (not having cognitive complaints as inclusion criterion) showed no significant**

differences on any of the cognitive domains (*p*-values≥.062), suggesting that the present BD sample was representative of BD patients in full or partial remission. Relative to Burdick et al. (2014) the somewhat smaller battery of neuropsychological tests applied in the present study may limit the evaluation of general cognitive deficit. Nevertheless, our findings are highly consistent with the study by Burdick et al. (2014) and the fewer tests should not result in smaller effect size for cognitive deficits since computing of effect sizes relies on the *M* and *SD* of the performance by our HC reference group on each test (which is then averaged to create composite scores for each domain; see details in the methods section). Nevertheless, future investigations of the cognitive heterogeneity in BD should include a more comprehensive battery of neurocognitive tests also including tests of social cognition and visual learning as in Burdick et al. (2014; 2015). It was a limitation of the present study that we had no direct measure of premorbid intelligence since this is strongly correlated with cognitive function. Premorbid intelligence was only assessed in two of the four studies from which the data was pooled and we therefore used educational levels as a proxy of cognitive reserve, which is highly correlated with intelligence levels (Lezak et al., 2012). Another limitation is that we did not formally assess dyslexia in the patients; however, none of the patients reported any difficulties with completing the several self-administered (written) questionnaires or on any of the neuropsychological tasks, indicating that it was unlikely that any patient suffered from dyslexia. Age differences between the neurocognitive subgroups, may have influenced the HCA detection of three neurocognitive subgroups. Nevertheless, no significant age differences were found between the globally impaired and the cognitively intact subgroup, and all analyses comparing the clinical characteristics of the neurocognitive subgroups were adjusted for age. **We used forced entry multiple regression for the post-hoc analyses. The disadvantage of this method to step-wise regression is that all independent variables are entered to the model**

simultaneously, which could possess a problem if multicollinearity was a source of concern (this was not the case in the present study). However, the step-wise methods are limited by selecting explanatory variables of the model based on statistical criteria rather than theoretical groundings (Field, 2013). Another limitation of the study was the lack of precise data on previous of affective episodes beyond >3 or psychotic episodes. However, the comparable proportion of BD type I patients in the present and previous study by Burdick et al. (2014) suggests that differences in history of psychotic symptoms may not explain the larger group of cognitively intact patients in our study. Finally, the use of LSD post-hoc comparison could have increased the likelihood of error due to lack of correction for multiple comparisons. The large sample sizes (BD: *N*=193, HC: *N*=110) with same measures of cognition and functional capacity for all participants were strengths of the study, which provide strong statistical power for the presence of the three neurocognitive subgroups and their clinical and functional characteristics.

We identified three neurocognitive subgroups in fully or partially remitted BD patients. Globally and selectively impaired patients displayed more stress, poorer work **and social adjustment** and reduced quality of life than those who were cognitively intact. These findings highlight a need to screen for cognitive impairment in BD and indicate that novel treatments that target cognition may improve patients' functional recovery.

Acknowledgements
The Lundbeck Foundation and TrygFonden are acknowledged for their contributions to KWM's post-doctorate salary at the Psychiatric Center Copenhagen, Copenhagen University Hospital, Rigshospitalet. The Research Fund of the Mental Health Services in the Capital Region of Denmark

is acknowledged for UK's post-doctorate salary at the Psychiatric Center Copenhagen, Copenhagen University Hospital, Rigshospitalet.

References

1. Altman, E.G., Hedeker, D.R., Janicak, P.G., Peterson, J.L. & Davis, J.M., 1994. The clinician-administered rating scale for mania (CARS-M): development, reliability, and validity. Biol. Psychiat. 36, pp. 124-134.
2. Army Individual Test Battery, 1944). Manual of Directions and Scoring. War Department, Adjutant General's Office, Washington, DC.
3. Arts, B., Jabben, N., Krabbendam, L., Van Os, J., 2008. Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. Psychol. Med. 38, pp. 771-785.
4. Bonnín, C. M., Martínez-Arán, A., Torrent, C., Pacchiarotti, I., Rosa, A. R., Franco, C., Murru, A., Sanchez-Moreno, A., Vieta, E., 2010. Clinical and neurocognitive predictors of functional outcome in bipolar euthymic patients: a long-term, follow-up study. J. Affect. Disord. 121, pp. 156-160.
5. Bora, E., Yucel, M., Pantelis, C., 2009. Cognitive endophenotypes of bipolar disorder: a meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. J. Affect. Disord. 113, pp. 1-20.
6. Bora, E., Hıdıroğlu, C., Özerdem, A., Kaçar, Ö. F., Sarısoy, G., Arslan, F.C., Aydemir, Ö., Tas, Z.C., Vahip, S., Atalay, A., Atasoy, N., Atesci, F., Tümkaya, S., 2016. Executive dysfunction and cognitive subgroups in a large sample of euthymic patients with bipolar disorder. Eur. Neuropsychopharmacol. DOI: http://dx.doi.org/10.1016/j.euroneuro.2016.04.002
7. Borkowski, J.G., Benton, A.L. & Spreen, O., 1967. Word fluency and brain damage. Neuropsychologia 5, pp. 135-140.

8. Bourne, C., Aydemir, Ö, Balanzá-Martéz, V., Bora, E., Brissos, S., Cavanagh, J.T.O., Clark, L., Cubukcuoglu, Z., Dias, V.V., Dittmann, S., 2013. Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. Acta Psychiatr. Scand. 128, pp. 149-162.
9. Brissos, S., Dias, V.V., Carita, A.I., Martinez-Arán, A., 2008. Quality of life in bipolar type I disorder and schizophrenia in remission: clinical and neurocognitive correlates. Psychiatry Res. 160, pp. 55-62.
10. Burdick, K. E., Endick, C. J., Goldberg, J.F., 2005. Assessing cognitive deficits in bipolar disorder: are self-reports valid?. Psychiatry Res. 136, pp. 43-50.
11. Burdick, K.E., Russo, M., Frangou, S., Mahon, K., Braga, R.J., Shanahan, M., Malhotra, A.K., 2014. Empirical evidence for discrete neurocognitive subgroups in bipolar disorder: clinical implications. Psychol. Med. 44, pp. 3083-3096.
12. Burdick, K.E., Ketter, T.A., Goldberg, J.F., Calabrese, J. R., 2015. Assessing cognitive function in bipolar disorder: challenges and recommendations for clinical trial design. J. Clin. Psychiatry, 76.
13. Cohen, S., Kamarck, T., Mermelstein, R., 1983. A global measure of perceived stress. J. Health Soc. Behav. 24, pp. 385-396.
14. Corwin, J., Bylsma, F.W., 1993. Psychological examination of traumatic encephalopathy. Clin. Neuropsychol. 7, pp. 3-21.
15. Demant, K., Vinberg, M., Kessing, L.V., Miskowiak, K.M., 2015. Effects of cognitive remediation on cognitive dysfunction in partially or fully remitted individuals with bipolar disorder: results of a randomised controlled trial. PLoS ONE 10, pp. 1-17.
16. Depp, C.A., Mausbach, B.T., Harmell, A. L., Savla, G.N., Bowie, C.R., Harvey, P.D., Patterson, T.L., 2012. Meta-analysis of the association between cognitive abilities and everyday functioning in bipolar disorder. Bipolar Disord. 14, pp. 217-226.
17. Dias, V.V., Balanzá-Martinez, V., Soeiro-de-Souza, M.G., Moreno, R.A., Figueira, M.L., Machado-Vieira, R., Vieta, E., 2012. Pharmacological approaches in bipolar disorders and the impact on cognition: a critical overview. Acta Psychiatr. Scand. 126, pp. 315-331.

18. Fava, M., Iosifescu, D.V., Pedrelli, P., Baer, L., 2009. Reliability and validity of the Massachusetts general hospital cognitive and physical functioning questionnaire. Psychother. Psychosom. 78, pp. 91-97.
19. Field, A., 2013. Discovering statistics using IBM SPSS statistics. Sage.
20. Gualtieri, C.T., Morgan, D.W., 2008. The frequency of cognitive impairment in patients with anxiety, depression, and bipolar disorder: an unaccounted source of variance in clinical trials. J. Clin. Psychiatry 69, pp. 1122-1130.
21. Goldberg, J.F., 2008. Adverse cognitive effects of psychotropic medications. In: Goldberg, J.F., Burdick, K.E. (Eds.), Cognitive Dysfunction in Bipolar Disorder: A Guide for Clinicians. American Psychiatric Publication, Inc., Washington, DC, pp. 137-158
22. Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatr. 23, pp. 56-62.
23. Harvey, P.D., Wingo, A.P., Burdick, K.E. & Baldessarini, R.J., 2010. Cognition and disability in bipolar disorder: lessons from schizophrenia research. Bipolar Disord. 12, pp. 364-375.
24. Huxley, N., Baldessarini, R.J., 2007. Disability and its treatment in bipolar disorder patients. Bipolar Disord. 9, pp. 183-196.
25. Jensen, J.H., Støttrup, M.M., Nayberg, E., Knorr, U., Ullum, H., Purdon, S.E., Kessing, L.V., Miskowiak, K.W., 2015. Optimising screening for cognitive dysfunction in bipolar disorder: Validation and evaluation of objective and subjective tools. J. Affect. Disord. 187, pp. 10-19.
26. Kessler, R.C., Chiu, W.T., Demler, O., Walters, E.E., 2005. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch. Gen. Psychiatr. 62, pp. 617-627.
27. Kogan, J.N., Otto, M.W., Bauer, M.S., Dennehy, E.B., Miklowitz, D.J., Zhang, H., Ketter, T., Rudorfer, M.V., Wisniewski, S.R., Thase, M.E., 2004. Demographic and diagnostic characteristics of the first 1000 patients enrolled in the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). Bipolar Disord. 6, pp. 460-469.
28. Lezak, M.D., Howieson, D.B., Bigler, E.D., Tranel, D., 2012. Neuropsychological Assessment, 5th edition. Oxford University Press, New York.

29. Lewandowski, K.E., Sperry, S.H., Cohen, B.M., Öngür, D., 2014. Cognitive variability in psychotic disorders: a cross-diagnostic cluster analysis. Psychol. Med. 44, pp. 3239-3248.

30. Martinez-Arán, A., Vieta, E., Torrent, C., Sanchez-Moreno, J., Goikolea, J. M., Salamero, M., Malhi, G.S., Gonzalez-Pinto, A., Daban, C., Alvarez-Grandi, S., Fountoulakis, K., Tabares-Seisdesdos, R., Ayuso-Mateos J.L., 2007. Functional outcome in bipolar disorder: the role of clinical and cognitive factors. Bipolar Disord. 9, pp. 103-113.

31. Martino, D.J., Strejilevich, S.A., Marengo, E., Ibañez, A., Scápola, M., Igoa, A., 2014. Toward the identification of neurocognitive subtypes in euthymic patients with bipolar disorder. J. Affect. Disord. 167, pp. 118-124.

32. Miskowiak, K.W., Ehrenreich, H., Christensen, E.M., Kessing, L.V., Vinberg, M., 2014. Recombinant human erythropoietin to target cognitive dysfunction in bipolar disorder: a double-blind, randomized, placebo-controlled phase 2 trial. J. Clin. Psychiatr. 75, pp. 1347-1355.

33. Mundt, J.C., Marks, I.M., Shear, M.K., Greist, J. M., 2002. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br. J. Psychiatry 180, pp. 461-464.

34. Mur, M., Portella, M.J., Martinez-Arán, A., Pifarre, J., Vieta, E., 2009. Influence of clinical and neuropsychological variables on the psychosocial and occupational outcome of remitted bipolar patients. Psychopathology 42, pp. 148-156.

35. Myers, R., 1990. Classical and modern regression with applications, 2nd edition, Boston, MA: Duxbury

36. Ott, C.V., Bjertrup, A.J., Jensen, J.H., Ullum, H., Sjælland, R., Purdon, S.E., Vieta, E., Kessing, L.V., Miskowiak, K.W., 2016. Screening for cognitive dysfunction in unipolar depression: Validation and evaluation of objective and subjective tools. J. Affect. Disord. 190, pp. 607-615.

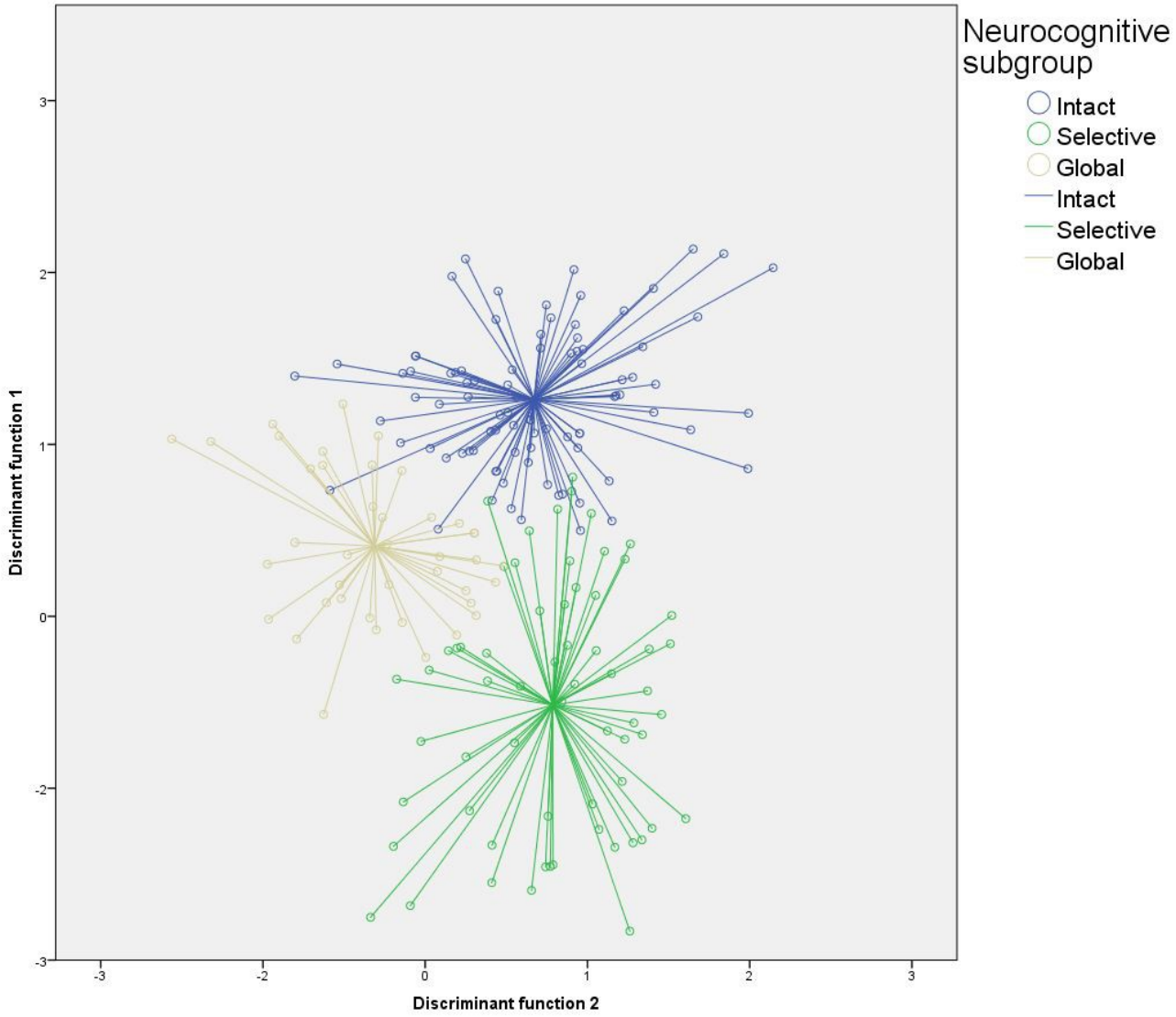
37. Randolph, C., 1998. RBANS manual: Repeatable battery for the assessment of neuropsychological status. The Psychological Corporation, San Antonio, Texas.

38. Reichenberg, A., Harvey, P.D., Bowie, C.R., Mojtabai, R., Rabinowitz, J., Heaton, R.K., Bromet, E., 2009. Neuropsychological function and dysfunction in schizophrenia and psychotic affective disorders. Schizophr. Bull. 35, pp. 1022-1029.
39. Rey, A., 1964. L'examen clinique en psychologie [Clinical psychological examination]. Presses Universitaires de France, Paris.
40. Rojo, E., Pino, O., Guilera, G., Gómez-Benito, J., Purdon, S.E. , Crespo-Facorro, B. , Cuesta, M., Franco, M., Martinez-Arán, A., Segarra, N., Tabarés-Seisdedos, R., Vieta, E., Bernado, M., 2010. Neurocognitive diagnosis and cut-offscores of the Screen for Cognitive Impairment in Psychiatry (SCIP-S). Schizophr. Res. 116, pp. 243-251.
41. Rosa, A.R., Mercade, C., Sanchez-Moreno, J., Sole, B., Bonnin, C.D.M., Torrent, C., Grande, I., Sugranyes, G., Popovic, D., Salamero, M., Kapczinski, F., Vieta, E., Martinez-Arán, A., 2013. Validity and reliability of a rating scale on subjective cognitive deficits in bipolar disorder (COBRA). J. Affect. Disord. 150, pp. 29-36.
42. Royall, D.R., Lauterbach, E.C., Kaufer, D., Malloy, P., Coburn, K.L., Black, K.J., 2007. The cognitive correlates of functional status: a review from the Committee on Research of the American Neuropsychiatric Association. J. Neuropsychiatr. Clin. Neuroci. 19, pp. 249-265.
43. Samamé, C., Martino, D. J., Strejilevich, S. A., 2014. Longitudinal course of cognitive deficits in bipolar disorder: a meta-analytic study. J. Affect. Disord. 164, pp. 130-138.
44. Torrent C., Bonnin C.D.M., Martínez-Arán A., Valle J, Amann B.L., González-Pinto A., Crespo J.M., Ibáñez Á., Garcia-Portilla M.P., Tabarés-Seisdedos R., Arango C., Colom F., Solé B., Pacchiarotti I., Rosa A.R., Ayuso-Mateos J.L., Anaya C., Fernández P., Landín-Romero R., Alonso-Lana S., Ortiz-Gil J., Segura B., Barbeito S., Vega P., Fernández M., Ugarte A., Subirà M., Cerrillo E., Custal N., Menchón J.M., Saiz-Ruiz J., Rodao J.M., Isella S., Alegría A., Al-Halabi S., Bobes J., Galván G., Saiz P.A., Balanzá-Martínez V., Selva G., Fuentes-Durá I., Correa P., Mayoral M., Chiclana G., Merchan-Naranjo J., Rapado-Castro M.,

Salamero M., Vieta E., 2013. Efficacy of functional remediation in bipolar disorder: a multicenter randomized controlled study. Am. J. Psychiatry 170, pp. 852-859.
45. Torrent, C., Martinez-Arán, A., del Mar, B.C., Reinares, M., Daban, C., Sole, B., Rosa, A.R., Tabares-Seisdedos, R., Popovic, D., Salamero, M., Vieta, E., 2012. Long-term outcome of cognitive impairment in bipolar disorder. J. Clin. Psychiatry 73, pp. e899-e905.
46. Tse, S., Chan, S., Ng, K.L., Yatham, L.N., 2014. Meta-analysis of predictors of favorable employment outcomes among individuals with bipolar disorder. Bipolar Disord. 16, pp. 217-229.
47. Wechsler, D., 1997. WAIS-III: Administration and scoring manual: Wechsler adult intelligence scale. Psychological Corporation, San Antonio, Texas.
48. WHOQOL Group, 1994. The development of the World Health Organization quality of life assessment instrument (the WHOQOL). Springer, Berlin, Heidelberg, pp. 41-57.
49. Volkert, J., Kopf, J., Kazmaier, J., Glaser, F., Zierhut, K.C., Schiele, M.A., Reif, A., 2015. Evidence for cognitive subgroups in bipolar disorder and the influence of subclinical depression and sleep disturbances. Eur. Neuropsychopharmacol. 25, pp. 192-202.
50. Wing, J.K., Babor, T., Brugha, T., Burke, J., Cooper, J.E., Giel, R., Jablenski, A., Regier, D., Sartorius, N., 1990. SCAN: Schedules for Clinical Assessment in Neuropsychiatry. Arch. Gen. Psychiatry, 47, pp. 589-593.
51. Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania: reliability, validity and sensitivity. Br. J. Psychiatry 133, pp. 429-435.

Figure 1

Graphical agglomeration of the neurocognitive subgroups.

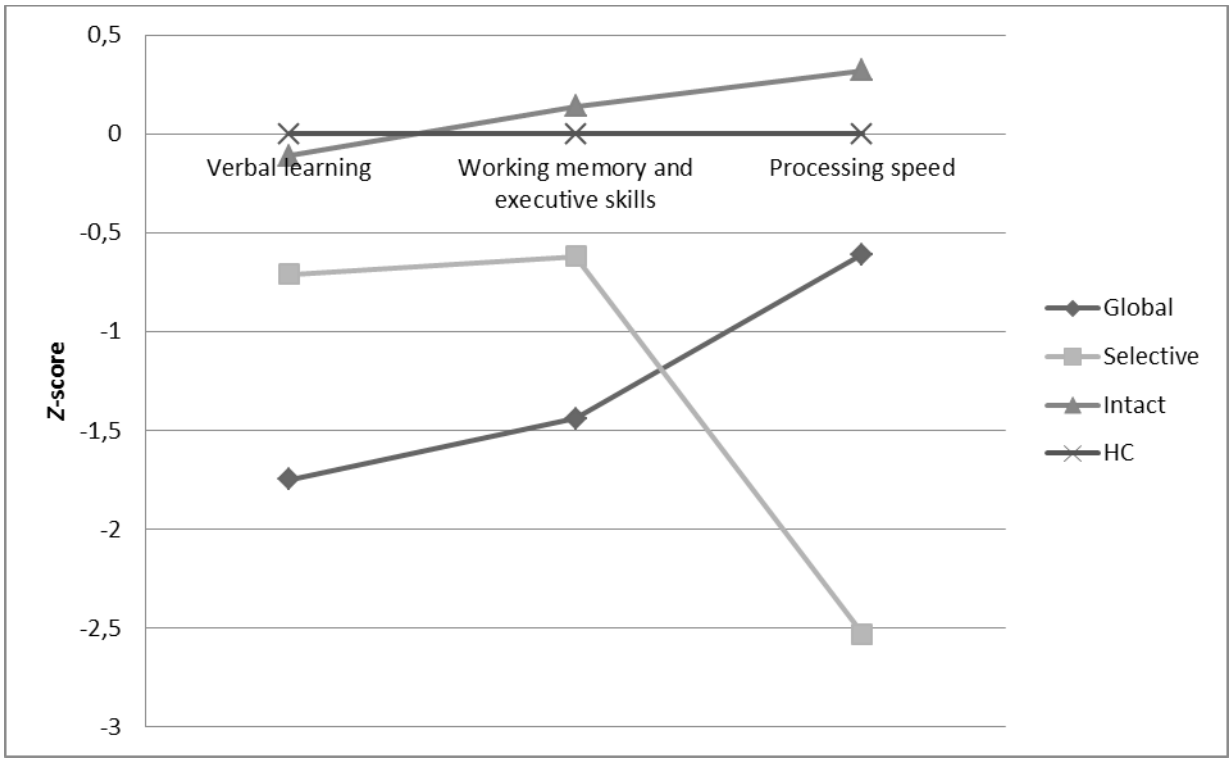


Scatter-plot and centroids of the discriminant values from the discriminant function analysis based

on the neurocognitive subgroups established from the hierarchical cluster analysis.
Abbreviations: Global = globally impaired subgroup, Selective = selectively impaired subgroup,

Intact = cognitively intact subgroup, HC = healthy control

Figure 2
Neurocognitive profiles of the three subgroups and HC sample.



The X-axis indicates the cognitive domains and the Y-axis indicate the mean cognition *z*-score of the neurocognitive subgroups and the HC sample.
Abbreviations: Global = globally impaired subgroup, Selective = selectively impaired subgroup, Intact = cognitively intact subgroup, HC = healthy control

Table 1
Characteristics of the complete BD and HC samples.

			Statistics		
	BD patients	HC	*df*	*t* or χ^2	*p*
N	193	110			-
Age	36 (10)	35 (12)	200.05	.97	.335
Gender, F/M (%)	62/38	57/43	1	.70	.401
Years of education	15 (3)	16 (2)	300	-1.65	.101
Age of onset	22 (10)	-			-

Illness duration, years	15 (11)	-			-
HDRS-17	6 (4)	1 (1)	300	11.07	<.001
Full remission, ≤7, no. (%)	132 (69)	110 (100)			-
Partial remission, 8≤14, no. (%)	60 (31)	0 (0)			-
YMRS	3 (3)	1 (1)	299	6.21	<.001
Full remission, ≤7, no. (%)	177 (93)	110 (100)			-
Partial remission, 8≤14, no. (%)	14 (7)	0 (0)			-
BD type, I/II (%)	58/42	-			-
Depressive episodes (1/2/>3), (%)	2/12/84	-			-
(Hypo-) manic episodes (1/2/>3), (%)	11/17/72	-			-
Composite cognition score	-1.0 (1.5)	0.0 (1.0)	271.03	-6.68	<.001
Verbal learning and memory	-.66 (1.2)	0.0 (1.0)	255.75	-5.17	<.001
Working memory and executive skills	-.44 (1.2)	0.0 (1.0)	300	-3.32	.001
Processing speed	-.80 (1.5)	0.0 (1.0)	294.42	-5.56	<.001
WSAS	16.5 (8.6)	-			-
PSS	-2.3 (1.8)	0.0 (1.0)	298.71	-14.44	<.001
CPFQ cognition measure	-4.5 (3.2)	0.0 (1.0)	220.60	-18.91	<.001
WHOQOL	-3.1 (1.3)	0.0 (1.0)	266.50	-23.50	<.001
Current medication					
Number of medications	2.1 (1.0	-			-
Lithium, no. (%)	110 (58)	-			-
Anticonvulsants, no. (%)	118 (62)	-			-
Antidepressants, no. (%)	40 (21)	-			-
Antipsychotics, no. (%)	89 (47)	-			-
Benzodiazepines, no. (%)	33 (17)	-			-
Melatonin, no. (%)	5 (7)	-			-
No medication, no. (%)	3 (2)	-			-

Independent samples t-test and χ^2 as appropriate. Measures of cognition, the PSS, CPFQ cognition measure, and WHOQOL are given as z-scores, mean (standard deviation), based on the HCs' performance.

Abbreviations: BD = Bipolar disorder, HC = Healthy control, HDRS-17 = Hamilton Depression Rating Scale 17-item, YMRS = Young Mania Rating Scale, WSAS = Work and Social Adjustment Scale, PSS = Perceived Stress Scale, CPFQ = Massachusetts General Hospital Cognitive and Physical Functioning Questionnaire, WHOQOL = World Health Organization's Quality of Life.

Table 2

Comparison between the BD subgroups and HCs across cognitive domain and composite cognition scores.

	Global (*n*=41, 21.2%)	Selective (*n*=63, 32.6%)	Intact (*n*=89, 46.1%)	HC (*n*=110)	*df*	*F*	*p*
Composite cognition score	–2.02 (.66)	–2.06 (1.14)	.20 (.92)	0.0 (1.0)	3	99.45	<.001
							HC *v*. G, *p*<.001
							HC *v*. S, *p*<.001
							HC *v*. I, *p*=.187
							G *v*. S, *p*=.860
							G *v*. I, *p*<.001
							S *v*. I, *p*<.001
Verbal learning	–1.75 (.77)	–.71 (1.11)	-.11 (.97)	0.0 (1.0)	3	36.12	<.001
							HC *v*. G *p*<.001
							HC *v*. S *p*<.001
							HC *v*. I *p*=.438
							G *v*. S *p*<.001
							G *v*. I *p*<.001
							S *v*. I *p*<.001
Working memory and executive skills	–1.44 (.62)	–.62 (1.03)	.14 (1.12)	0.0 (1.0)	3	28.52	<.001
							HC *v*. G *p*<.001
							HC *v*. S *p*<.001
							HC *v*. I *p*=.336
							G *v*. S *p*<.001
							G *v*. I *p*<.001
							S *v*. I *p*<.001
Processing speed	–.61 (.78)	–2.53 (1.12)	.32 (.63)	0.0 (1.0)	3	140.46	<.001
							HC *v*. G *p*<.001

						HC *v*. S *p*<.001
						HC *v*. I *p*=.013
						G *v*. S p<.001
						G *v*. I *p*<.001
						S *v*. I *p*<.001

Analyses of variance with least significance differences as pairwise comparison. Measures of cognition are given as z-scores, mean (standard deviation), based on the HCs' performance.
Abbreviations: BD = bipolar disorder, HC = healthy control, df = degrees of freedom, G = globally impaired subgroup, S = selectively impaired subgroup, I = cognitively intact subgroup.

Table 3
Comparison of the characteristics of the neurocognitive cognitive subgroups.

	Subgroup			Statistics		
	Global	Selective	Intact	*df*	*F* or χ^2	*p*
n	41	63	89			
Age, *min-max*	38 (11), 20-65	39 (11), 18-58	34 (9), 19-61	2	6.70	.002
						G *v.* I, *p*=.020
						S *v.* I, *p*=.001
Gender, F/M (%)	66/34	64/36	60/40	2	.54	.762

Years of education	14 (3)	15 (3)	15 (3)	2	2.80	.064
Age of onset	23 (10)	21 (11)	21 (9)	2	1.10	.334
Illness duration, years	15 (10)	18 (11)	13 (9)	2	1.10	.334
HDRS-17	6 (5)	6 (4)	5 (4)	2	1.74	.178
Full remission, ≤7, no. (%)	**27 (66)**	**38 (60)**	**67 (76)**	**2**	**4.48**	**.106**
Partial remission, 8≤14, no. (%)	**14 (34)**	**25 (40)**	**21 (24)**	**2**	**4.48**	**.106**
YMRS	3 (3)	3 (3)	2 (3)	2	.16	.856
Full remission, ≤7, no. (%)	**39 (95)**	**56 (90)**	**82 (93)**	**2**	**.90**	**.638**
Partial remission, 8≤14, no. (%)	**2 (5)**	**6 (10)**	**6 (7)**	**2**	**.90**	**.638**
BD type, I/II (%)	61/39	61/39	53/47	2	1.17	.557
Depressive episodes (1/2/≥3), (%)	0/13/87	2/12/86	8/11/81	4	5.62	.230
(Hypo-) manic episodes (1/2/≥3), (%)	8/14/78	11/20/69	13/16/71	4	1.39	.847
WSAS	18 (9)	18 (9)	14 (8)	2	3.94	.021
						G *v.* I, *p*=.030
						S *v.* I, *p*=.014
PPS	–2.84 (1.89)	–2.62 (1.59)	–1.89 (1.81)	2	5.16	.007
						G *v.* I, *p*=.005
						S *v.* I, *p*=.013
CPFQ cognition measure	–8.04 (4.95)	–6.86 (4.00)	–5.49 (4.53)	2	5.00	.008
						G *v.* I, *p*=.003
						S *v.* I, *p*=.049
WHOQOL	–3.40 (1.32)	–3.35 (1.14)	–2.89 (1.29)	2	3.38	.038
						G *v.* I, *p*=.042
						S *v.* I, *p*=.027
Physical health	–3.56 (1.55)	–3.38 (1.48)	–2.98 (1.62)	2	2.73	.068
Psychological health	–2.48 (1.35)	–2.76 (1.48)	–2.35 (1.62)	3	2.10	.126
Social relationsships	–1.32 (1.66)	–1.45 (1.31)	–1.03 (1.42)	3	.81	.447
Environment	–2.38 (1.21)	–2.48 (1.36)	–2.01 (1.26)	3	3.38	.036
						S *v.* I, *p*=.015
Current medication						

Number of medications	2.1 (0.9)	2.6 (1.0)	1.9 (1.0)	2	7.25	.001
						G *v.* S, *p*=.009
						S *v.* I, *p*<.001
Lithium, no. (%)	25 (61)	37 (60)	48 (55)	2	.64	.727
Anticonvulsants, no. (%)	23 (56)	41 (66)	54 (61)	2	1.01	.587
Antidepressants, no. (%)	6 (15)	20 (32)	14 (16)	2	7.13	.028
				1	4.06	G *v.* S, *p*=.044
				1	5.55	S *v.* I, *p*=.019
Antipsychotics, no. (%)	23 (56)	33 (53)	33 (38)	2	5.51	.064
Benzodiazepines, no. (%)	6 (15)	20 (32)	7 (8)	2	15.29	<.001
				1	4.06	G *v.* S, *p*=.044
				1	14.56	S *v.* I, *p*<.001
Melatonin, no. (%)	0 (0)	0 (0)	5 (6)	2	6.01	.050
No medication, no. (%)	1 (2)	0 (0)	2 (2)	2	1.47	.480

Analyses of covariance adjusted for age and least significance difference as pairwise comparison and χ^2 as appropriate. Measures of cognition, the PSS, CPFQ cognition measure, and WHOQOL are given as z-scores, mean (standard deviation), based on the HCs' performance.

Abbreviations: HC = healthy controls, df = degrees of freedom, HDRS-17 = Hamilton Depression Rating Scale 17-item, YMRS = Young Mania Rating Scale, BD = bipolar disorder, WSAS = Work and Social Adjustment Scale, PSS = Perceived Stress Scale, CPFQ = Massachusetts General Hospital Cognitive and Physical Functioning Questionnaire, WHOQOL = World Health Organization's Quality of Life, G = globally impaired subgroup, S = selectively impaired subgroup, I = cognitively intact subgroup.

Abstract

Background: Neurocognitive impairment in remitted patients with bipolar disorder contributes to functional disabilities. However, the pattern and impact of these deficits are unclear.

Methods: We pooled data from 193 fully or partially remitted patients with bipolar disorder and 110 healthy controls. Hierarchical cluster analysis was conducted to determine whether there are discrete neurocognitive subgroups in bipolar disorder. The pattern of the cognitive deficits and the characteristics of patients in these neurocognitive subgroups were examined with analyses of covariance and least significance difference pairwise comparison.

Results: Three discrete neurocognitive subgroups were detected: one that was cognitively intact (46.1%), one that was selectively impaired with deficits only in processing speed (32.6%), and one that was globally impaired across verbal learning, working memory, and executive skills (21.2%). The globally and selectively impaired subgroups were characterized by greater perceived stress and cognitive complaints, poorer work **and social adjustment**, and reduced quality of life compared to patients who were cognitively intact.

Limitations: The study design was cross-sectional which limits inferences regarding the causality of the findings.

Conclusion: Globally and selectively impaired bipolar disorder patients displayed more functional disabilities than those who were cognitively intact. The present findings highlight a clinical need to systematically screen for cognitive dysfunction in remitted bipolar disorder and to target residual cognitive dysfunction in future treatment strategies.

Highlights

Three subgroups with unique neurocognitive profiles were found in bipolar disorder.

Intact cognitive function was found in 46% patients.

Subgroups with global (21%) or selective (33%) deficits had most functional problems.

There is need for a more tailored treatment targeting cognition in bipolar disorder.

Comments from the editors and reviewers:

-Reviewer #1

-

The authors present a cluster analysis on data combined from 4 their previous investigations of bipolar disorder I and II (n=193) resulting in three clusters of patients – one cluster with intact verbal learning, working memory, and psychomotor speed scores (relative to their Healthy Control group), a second with verbal learning/memory and working memory within normal limits, but marked impairment of psychomotor speed, and a third with psychomotor speed within normal limits but marked impairment of verbal learning/memory and working memory. The impaired groups differed from the intact group on measures of perceived stress, subjective cognitive complaints, poor work capacity, and reduced quality of life.

This paper is important to both academic and clinical work with Bipolar Disorder, particularly the attention given to cognitive deficits that may underlie disablity in this sample, and in the remarkable proportion of BPD participants with 'intact' cognitive function.

The paper is generally well written and structured, though there were a number of small formatting problems and grammatical (or typographic) errors.

The results are somewhat in accord with an earlier cluster analysis (Burdick et al. 2014) that applied a longer assessment protocol (MCCB) that included tests from a visual memory domain, a reasoning and problem solving (for spatial material) domain, and a social cognition domain that were not included in the present analysis. The processing speed domain in the prior study also that also included BACS Coding in the psychomotor speed domain (in addition to the TMT-A used here), an attention domain that was limited to one test of vigilance and sustained attention (CPT), and a new verbal learning domain (without delayed recall).

Although the battery of tests included in the current analysis was different from that of the Burdick study, the three clusters do have some similarities, and I think this would be of interest to the readership, particularly given that this is apparently the first 'replication' of the earlier work. However, some qualification might be useful given the rather substantial portion of cognitively unimpaired BPD observed, here, particularly given the explicit inclusion of subjects that endorsed subjective deficits on the CPFQ.

In the BD sample of the study there was characterized by a substantial proportion of cognitively 'intact' patients (46.1%) despite having moderate-severe cognitive complaints. The association between cognitive performance and complaints is controversial as our and other research groups have found subjective cognitive impairment to be poorly related to objectively-measured cognitive dysfunction in BD (Rosa et al., 2013; Jensen et al., 2015). Very few studies have investigated neurocognitive subgroups in BD using a data-driven methodology rather than applying a predetermined cut-off score for cognitive impairment. In response to the reviewer's request we have now provided some more qualification of the study and the relation between subjective and objective cognitive impairment:

- Introduction, section 1, p. 2: **Cognitive impairment is among the strongest contributors together with mood symptoms and illness progression to functional disability, lower quality of life, and loss of workforce capacity in BD (Bonnín et al., 2010; Brissos et al., 2008; Torrent et al., 2012). However, the correlation between subjectively experienced and objectively measured cognitive impairment is poor, as shown by our and other research groups (Burdick, Endick and Goldberg, 2005; Jensen et al., 2015; Rosa et al., 2013).**
- Introduction, section 1, p. 2: Replaced "This points to cognitive dysfunction as a key treatment priority to improve patients' functional recovery." with "This points to **objectively measured** cognitive dysfunction as a key treatment priority to improve patients' functional recovery **and quality of life after acute mood episodes**."
- Introduction, section 1, p. 3: **Several studies have examined cognition in BD using predetermined cut-offs (typically 1-2 standard deviations [SD] from the normal mean [M]) for simple classification of impaired and non-impaired patients (e.g., Jensen et al., 2015; Martino et al., 2014; Reichenberg et al., 2009; Rojo et al., 2010; Volkert et al., 2015). However, there is a scarcity of studies that have used a *data-driven* approach to identify neurocognitive subgroups.**
- Introduction, section 1: Replaced "However, a recent study of N=136 BD patients in full or partial remission suggests that cognitive impairment is not a uniform feature of BD; instead discrete neurocognitive subgroups were found; a "cognitively intact" subgroup with scores equivalent to the performance of healthy control persons (HCs) (31.6%), a subgroup with "selective cognitive impairment" (28.7%), and a subgroup with "global severe impairment" across cognitive domains (39.7%) comparable to cognitive deficits in schizophrenia (Burdick et

al., 2014)" with "**Such studies found discrete neurocognitive subgroups in fully or partially remitted** patients with BD **(Bora et al., 2016; Burdick et al., 2014; Lewandowksi et al., 2014)**; a **well performing** "cognitively intact" subgroup with scores equivalent to the performance of healthy control persons (HCs), **one or two** subgroups **of** "selective cognitive impairment" **with lower cognition scores compared to HCs**, and a subgroup with "global severe impairment" across cognitive domains comparable to cognitive deficits in schizophrenia."

- Introduction, section 1, p. 3: **The findings from these studies show some discrepancies with respect to the proportion of patients in each subgroup, possibly due to differences in study samples (e.g., affective symptoms, age, medicine prescribed).**
- Introduction, section 1, p. 3: Replaced "However, as the described study (11) is the only published evidence for the presence of neurocognitive subgroups in BD, additional replication studies are needed to make any firm conclusions about the pattern of cognitive dysfunction in BD." with "Additional **data-driven subgroup** studies are needed to make any firm conclusions about the pattern of cognitive in BD."
- Introduction, section 1.1, p. 3: Replaced "The study aims to 1) determine whether there are discrete neurocognitive subgroups in fully or partially remitted patients with BD as prior demonstrated (Burdick et al., 2014). Building onto the findings of Burdick et al. (Burdick et al., 2014), the study aims to 2) examine the pattern of the cognitive deficits and the clinical and functional characteristics of the neurocognitive subgroups in terms of perceived stress, cognitive complaints, work capacity, quality of life, and medication prescribed." with "The study aims to **identify** discrete neurocognitive subgroups in **a large group of** fully or partially remitted patients with BD **using a data-driven approach**. Building onto the **few studies applying this approach**, the **present** study aims to examine the pattern of the cognitive deficits and the clinical and functional characteristics of the neurocognitive subgroups in terms of perceived stress, cognitive complaints, work **and social adjustment**, quality of life, and medication prescribed."
- References, p. 20: **Bonnín, C. M., Martínez-Arán, A., Torrent, C., Pacchiarotti, I., Rosa, A. R., Franco, C., Murru, A., Sanchez-Moreno, A., Vieta, E., 2010. Clinical and neurocognitive predictors of functional outcome in bipolar euthymic patients: a long-term, follow-up study. J. Affect. Disord. 121, pp. 156-160.**

- References, p. 20: **Bora, E., Hıdıroğlu, C., Özerdem, A., Kaçar, Ö. F., Sarısoy, G., Arslan, F.C., Aydemir, Ö., Tas, Z.C., Vahip, S., Atalay, A., Atasoy, N., Atesci, F., Tümkaya, S., 2016. Executive dysfunction and cognitive subgroups in a large sample of euthymic patients with bipolar disorder. Eur. Neuropsychopharmacol. DOI: http://dx.doi.org/10.1016/j.euroneuro.2016.04.002**
- References, p. 21: **Brissos, S., Dias, V.V., Carita, A.I., Martinez-Arán, A., 2008. Quality of life in bipolar type I disorder and schizophrenia in remission: clinical and neurocognitive correlates. Psychiatry Res. 160, pp. 55-62.**
- References, p. 23: **Lewandowski, K.E., Sperry, S.H., Cohen, B.M., Öngür, D., 2014. Cognitive variability in psychotic disorders: a cross-diagnostic cluster analysis. Psychol. Med. 44, pp. 3239-3248.**
- References, p. 23: **Martino, D.J., Strejilevich, S.A., Marengo, E., Ibañez, A., Scápola, M., Igoa, A., 2014. Toward the identification of neurocognitive subtypes in euthymic patients with bipolar disorder. J. Affect. Disord. 167, pp. 118-124.**
- References, p. 24: **Rojo, E., Pino, O., Guilera, G., Gómez-Benito, J., Purdon, S.E. , Crespo-Facorro, B. , Cuesta, M., Franco, M., Martinez-Arán, A., Segarra, N., Tabarés-Seisdedos, R., Vieta, E., Bernado, M., 2010. Neurocognitive diagnosis and cut-offscores of the Screen for Cognitive Impairment in Psychiatry (SCIP-S). Schizophr. Res. 116, pp. 243-251.**
- References, p. 25: **Torrent, C., Martinez-Aran, A., del Mar, B.C., Reinares, M., Daban, C., Sole, B., Rosa, A.R., Tabares-Seisdedos, R., Popovic, D., Salamero, M., Vieta, E., 2012. Long-term outcome of cognitive impairment in bipolar disorder. J. Clin. Psychiatry 73, pp. e899-e905.**
- References, p. 26: **Volkert, J., Kopf, J., Kazmaier, J., Glaser, F., Zierhut, K.C., Schiele, M.A., Reif, A., 2015. Evidence for cognitive subgroups in bipolar disorder and the influence of subclinical depression and sleep disturbances. Eur. Neuropsychopharmacol. 25, pp. 192-202.**

The Burdick et al. (2014) study also reported a three factor solution to the cluster analysis, similar to the current study, and classified BPD participants into intact (approximately 32%), selective impairment (29%), and global impairment (40%). A replication of the Burdick Clusters would be an important contribution to the literature. The current study resulted in similar clusters comprising 46, 33, and 21% respectively. The authors underscore the similarity in proportions with

'selective' impairment (i.e. primarily psychomotor limitations), and they suggest the discrepancy between studies in the rate of 'intact' or 'global' impairment may relate to a difference in criteria for 'partial remission', or to the somewhat younger sample examined here. It might also relate to the prior study recruitment of only DSM BPD I with psychotic features, and BPD II with psychotic features. The current study applied ICD-10 criteria and did not mention 'with' or 'without' psychotic features.

We thank the reviewer for this valuable observation regarding the psychotic symptoms being an additional explanation for the differences between the findings of our study and the study by Burdick et al (2014). We agree that psychotic features are highly related to cognitive impairment. Indeed, Burdick et al. (2014, p. 3085) included 136 BD patients in total: 105 were diagnosed with BD type I (*n*=60 [57.1%] with psychotic features) and 31 with BD type II (*n*=9 [29%] with psychotic features) according to the Structural Clinical Interview for DSM-IV (SCID). A diagnosis of BD type I is *per se* defined by one or more previous manic episodes, which typically involve psychotic symptoms. In the present BD sample 58% of the patients were diagnosed with BD type I, which is comparable to 51% of patients in the study by Burdick et al. (2014) having history of psychotic features. This suggests that psychotic symptoms cannot explain the different findings between the two studies. However, unfortunately data on previous psychotic episodes was not available for the present study. We have now raised this as a limitation of the present study:

- Discussion, section 4, p. 18: **Another limitation of the study was the lack of precise data on previous of affective episodes beyond >3 or psychotic episodes. However, the comparable proportion of BD type I patients in the present and previous study by Burdick et al. (2014) suggests that differences in history of psychotic symptoms may not explain the larger group of cognitively intact patients in our study.**

Is it possible that a larger battery of tests might be more sensitive to general deficits, particularly given the large effect size observed for visual learning in the Burdick et al. (2014) sample.

We agree with the reviewer on this important point. The study would clearly have benefitted from applying a larger battery of neurocognitive tests for a more exhaustive examination of cognitive function in the patients with bipolar disorder included in the study. This has now been raised as a limitation in the discussion section (see below). Notwithstanding this limitation, the lower number of

tests in the present study compared to the study by Burdick et al (2014) should *not* result in a smaller effect size of cognitive deficits. The reason is that the z-transformation of the individual test scores were based on (and depended on) the normal function in our healthy control group (rather than the number of tests). The computed effect sizes of the cognitive composite scores for the respective domains thus rely on the mean and standard deviation of the reference group. This discussion of the potential reasons for the discrepancy between the present findings and the findings by Burdick et al. (2014) has now been included:

- Discussion, section 4, p. 17: "**Relative to Burdick et al. (2014) the somewhat smaller battery of neuropsychological tests applied in the present study may limit the evaluation of general cognitive deficit. Nevertheless, our findings are highly consistent with the study by Burdick et al. (2014) and the fewer tests should not result in smaller effect size for cognitive deficits since computing of effect sizes relies on the *M* and *SD* of the performance by our HC reference group on each test (which is then averaged to create composite scores for each domain; see details in the methods section). Nevertheless, future investigations of the cognitive heterogeneity in BD should include a more comprehensive battery of neurocognitive tests also including tests of social cognition and visual learning as in Burdick et al. (2014; 2015).**"
- References, p. 20: **Burdick, K.E., Ketter, T.A., Goldberg, J.F., Calabrese, J. R., 2015. Assessing cognitive function in bipolar disorder: challenges and recommendations for clinical trial design. J. Clin. Psychiatry, 76.**

I was a bit confused by the indication that the authors used a z<-1 cut-off score for cognitive impairment to define 'neurocognitive subgroups', similar they indicate, to the method used by Birdick et al. (2014). It is not clear to me how this classification of subgroups was integrated into the analysis. Wouldn't the DFA have classified a given subject into one of the three subgroups. Perhaps I am missing something simple here that could be easily explained. I suspect this classification was applied within the regression analysis of 3.3 in the results section.

We thank the reviewer for raising this point. Although there is still some controversy regarding the particular cutoff for cognitive impairment (limits range from 1-2 SD below the normal mean; Bora et al., 2016), we chose in the present study to set the cutoff score of ≥1 SD below the mean of the HC

sample to define cognitive impairment consistent with our previous approach (Jensen et al., 2015; Ott et al., 2016). For an easy evaluation of classification of cognitive impairment, we transformed all participant cognition scores to standardized *z*-scores based on the HCs performance. Thus, cognition *z*-scores of -1 and lower indicate cognitive dysfunction, since the *z*-distribution has a mean of 0 and a SD of 1. The HCA and DFA analyses were conducted using these *z*-scores. We performed the DFA analysis to validate the clusters retained by the HCA analysis; if the DFA analysis would classify the BD patients in the same neurocognitive subgroups as the HCA analysis, the HCA analysis is valid. In response to the reviewer's request we have now increased clarity by adding the following explanations:

- Materials and methods, section 2.5.1, p. 6: "**For an easy evaluation of cognitive function, all neurocognitive test scores of the** participants were standardized to *z*-scores (***M***=0, ***SD***=1) based on the cognitive performance of the HCs using following formula: (*test score - HC test M*) / *HC test SD (Field, 2013)*. **Although there is still some controversy regarding the particular cutoff for cognitive impairment (limits range from 1-2 *SD*s below the normal *M*; Bora et al., 2016; Martino et al., 2014), we chose in the present study to set the cut-off score of *z*≥1 below the *M* of the HC sample to define cognitive impairment consistent with Burdick et al. (2014) and our previous approach (Jensen et al., 2015; Ott et al., 2016).**"
- References, p. 20: "**Bora, E., Hıdıroğlu, C., Özerdem, A., Kaçar, Ö. F., Sarısoy, G., Arslan, F.C., Aydemir, Ö., Tas, Z.C., Vahip, S., Atalay, A., Atasoy, N., Atesci, F., Tümkaya, S., 2016. Executive dysfunction and cognitive subgroups in a large sample of euthymic patients with bipolar disorder. Eur. Neuropsychopharmacol. DOI: http://dx.doi.org/10.1016/j.euroneuro.2016.04.002**
- References, p. 23: **Martino, D.J., Strejilevich, S.A., Marengo, E., Ibañez, A., Scápola, M., Igoa, A., 2014. Toward the identification of neurocognitive subtypes in euthymic patients with bipolar disorder. J. Affect. Disord. 167, pp. 118-124.**

It might be useful for the authors to include a measure of premorbid intellect in future studies (similar to Burdick et al. 2014).

We thank the reviewer for this valuable suggestion which we will follow in future studies. In the present sample pooled from our previous studies we used years of education as a proxy measure for premorbid intelligence, since these measures are known to be highly correlated and we had obtained this measure from all four studies. In two of these studies we had also assessed premorbid intelligence

directly using the Danish version of the National Adult Reading Test (NART) but could not include this in the present analyses given that these data were not available from the two other studies (the two intervention trials where we had decided to omit the intelligence assessment since participants were already undergoing very extensive examination and we did not want to overload them). We have now raised this as a limitation in the manuscript:

- Discussion, section 4, p. 18: **"It was a limitation of the present study that we had no direct measure of premorbid intelligence since this is strongly correlated with cognitive function. Premorbid intelligence was only assessed in two of the four studies from which the data was pooled and we therefore used educational levels as a proxy of cognitive reserve, which is highly correlated with intelligence levels (Lezak et al., 2012)"**
- References, p. 23: **Lezak, M.D., Howieson, D.B., Bigler, E.D., Tranel, D., 2012. Neuropsychological Assessment, 5th edition. Oxford University Press, New York.**

The selective group exhibited relatively mild cognitive limitations (compared to the HC or the Intact group) but quite severe impairment of psychomotor speed (i.e. TMT-A time). They were not clinically distinct from the 'global' impairment group in proportion of depressive episodes, severity of depressed symptoms, or proportion of BPD II, yet they were receiving more anti-depressant medication and anti-anxiety medication than the other two groups. The authors note this in their results but do not offer comment on this in the discussion.

We thank the reviewer for highlighting this. This is now clarified in the:

- Discussion, section 4, p. 15: "The selectively impaired subgroup with severe reduction in psychomotor speed was characterized by older age, more polypharmacy**, and more antidepressants and benzodiazepines prescribed** than both the intact and globally impaired subgroups."
- Discussion, section 4, p. 15: "Considerable evidence indicates that **polypharmacy, benzodiazepines, some antidepressants (mainly tricyclic antidepressants), high plasma levels of lithium** and antipsychotics may have negative effects on psychomotor speed (Bora et al., 2009; **Dias et al., 2012; Goldberg, 2008**), suggesting that the selectively […] no significant correlations between slowing of processing speed and **medication, polypharmacy or benzodiazepines alone**, suggesting that the relationship between psychomotor slowing and medication is more complex."

- References, p. 22: **Dias, V.V., Balanzá-Martinez, V., Soeiro-de-Souza, M.G., Moreno, R.A., Figueira, M.L., Machado-Vieira, R., Vieta, E. (2012). Pharmacological approaches in bipolar disorders and the impact on cognition: a critical overview. Acta Psychiatr. Scand. 126, pp. 315-331.**
- References, p. 22: **Goldberg, J.F. 2008. Adverse cognitive effects of psychotropic medications. In: Goldberg, J.F., Burdick, K.E. (Eds.), Cognitive Dysfunction in Bipolar Disorder: A Guide for Clinicians. American Psychiatric Publication, Inc., Washington, DC, pp. 137-158**

The authors also note prior work implicating anti-convulsant and anti-psychotic medication in their discussion of 'polypharmacy' contributions but I did not find a similar citation regarding anti-depressants and/or benzodiazepines.

In response to this point we have now included Dias et al. (2012) and Goldberg et al. (2008) as references in the manuscript (see comment above).

The selective and the global impairment groups both differed from the intact group on measures of work capacity, stress, and subjective cognitive difficulties. This is interesting and potentially important (although some comment on expectatiosn regarding their subgroups might have been helpful, and differences in psychsocial outcomes between 'impaired' subgroups would have been particularly interesting). In this regard, the WSAS is a measure of work capacity, but it is also fairly heavily weighted on the home, social, and leisure (e.g. social leisure, private leisure, home management, close relationships). It might be more accurate to represent this in the discussion as a measure of work and social adjustment, rather than just 'work capacity'.

We thank the reviewer for noticing this and have now rephrased this to 'work and social adjustment' throughout the abstract and manuscript, and revised the following:

- Introduction, section 1, p. 2: **Cognitive impairment is among the strongest contributors together with mood symptoms and illness progression to functional disability, lower quality of life, and loss of workforce capacity in BD (Bonnín et al., 2010; Brissos et al., 2008; Torrent et al., 2012). However, the correlation between subjectively experienced and objectively measured cognitive impairment is poor, as shown by our and other research groups (Burdick, Endick and Goldberg, 2005; Jensen et al., 2015; Rosa et al., 2013).**

- Introduction, section 1, p. 2: Replaced “This points to cognitive dysfunction as a key treatment priority to improve patients’ functional recovery.” with “This points to **objectively measured** cognitive dysfunction as a key treatment priority to improve patients’ functional recovery **and quality of life after acute mood episodes**.”
- Introduction, section 1, p. 3: **The findings from these studies show some discrepancies with respect to the proportion of patients in each subgroup, possibly due to differences in study samples (e.g., affective symptoms, age, medicine prescribed).**
- Introduction, section 1.1, p. 3: Replaced “The study aims to 1) determine whether there are discrete neurocognitive subgroups in fully or partially remitted patients with BD as prior demonstrated (Burdick et al., 2014). Building onto the findings of Burdick et al. (Burdick et al., 2014), the study aims to 2) examine the pattern of the cognitive deficits and the clinical and functional characteristics of the neurocognitive subgroups in terms of perceived stress, cognitive complaints, work capacity, quality of life, and medication prescribed.” with “The study aims to **identify** discrete neurocognitive subgroups in **a large group of** fully or partially remitted patients with BD **using a data-driven approach**. Building onto the **few studies applying this approach**, the **present** study aims to examine the pattern of the cognitive deficits and the clinical and functional characteristics of the neurocognitive subgroups in terms of perceived stress, cognitive complaints, work **and social adjustment**, quality of life, and medication prescribed.”
- Discussion, section 4, p. 16: “The observation that work **and social adjustment**, stress, and quality of life were more negatively affected in of the globally and selectively impaired – but otherwise relatively symptom-free – subgroups compared to patients who were cognitively intact **is somewhat in line with previous reports (Brissos et al., 2008; Burdick et al. 2014; Martino et al., 2014). Burdick et al. (2014) found the global and the selective subgroup to be more occupationally disabled, but found no differences between the subgroups on residential or social functioning. The present results** highlight […]”
- References, p. 20: **Bonnín, C. M., Martínez-Arán, A., Torrent, C., Pacchiarotti, I., Rosa, A. R., Franco, C., Murru, A., Sanchez-Moreno, A., Vieta, E., 2010. Clinical and neurocognitive predictors of functional outcome in bipolar euthymic patients: a long-term, follow-up study. J. Affect. Disord. 121, pp. 156-160.**

- References, p. 20: **Bora, E., Hıdıroğlu, C., Özerdem, A., Kaçar, Ö. F., Sarısoy, G., Arslan, F.C., Aydemir, Ö., Tas, Z.C., Vahip, S., Atalay, A., Atasoy, N., Atesci, F., Tümkaya, S., 2016. Executive dysfunction and cognitive subgroups in a large sample of euthymic patients with bipolar disorder. Eur. Neuropsychopharmacol. DOI: http://dx.doi.org/10.1016/j.euroneuro.2016.04.002**
- References, p. 21: **Brissos, S., Dias, V.V., Carita, A.I., Martinez-Arán, A., 2008. Quality of life in bipolar type I disorder and schizophrenia in remission: clinical and neurocognitive correlates. Psychiatry Res. 160, pp. 55-62.**
- References, p. 23: **Lewandowski, K.E., Sperry, S.H., Cohen, B.M., Öngür, D., 2014. Cognitive variability in psychotic disorders: a cross-diagnostic cluster analysis. Psychol. Med. 44, pp. 3239-3248.**
- References, p. 23: **Martino, D.J., Strejilevich, S.A., Marengo, E., Ibañez, A., Scápola, M., Igoa, A., 2014. Toward the identification of neurocognitive subtypes in euthymic patients with bipolar disorder. J. Affect. Disord. 167, pp. 118-124.**
- References, p. 24: **Rojo, E., Pino, O., Guilera, G., Gómez-Benito, J., Purdon, S.E. , Crespo-Facorro, B. , Cuesta, M., Franco, M., Martinez-Arán, A., Segarra, N., Tabarés-Seisdedos, R., Vieta, E., Bernado, M., 2010. Neurocognitive diagnosis and cut-offscores of the Screen for Cognitive Impairment in Psychiatry (SCIP-S). Schizophr. Res. 116, pp. 243-251.**
- References, p. 25: **Torrent, C., Martinez-Aran, A., del Mar, B.C., Reinares, M., Daban, C., Sole, B., Rosa, A.R., Tabares-Seisdedos, R., Popovic, D., Salamero, M., Vieta, E., 2012. Long-term outcome of cognitive impairment in bipolar disorder. J. Clin. Psychiatry 73, pp. e899-e905.**
- References, p. 26: **Volkert, J., Kopf, J., Kazmaier, J., Glaser, F., Zierhut, K.C., Schiele, M.A., Reif, A., 2015. Evidence for cognitive subgroups in bipolar disorder and the influence of subclinical depression and sleep disturbances. Eur. Neuropsychopharmacol. 25, pp. 192-202.**

The differences between groups are entirely expected from the initial cluster analysis (and DFA). A similar issue was raised with Burdick in the 2014 paper and addressed in a paragraph directly commenting on the value of the head-to-head comparisons. The authors may wish to review that paragraph for some suggetions to apply here.

We are not entirely sure what the reviewer refers to by raising this point. Further, we were not able to identify a discussion on "*the value of the head-to-head comparisons*" in Burdick et al. (2014).

-Reviewer #2

-

The study provides interesting results and deserves interest.
However, some important methodological nodes need to be clarified and some limits should be highlighted and discussed.
The difference in education is of statistical significance!! ANOVA 1 way df 1,301, 302 F=9.746; P<0.002*. The methods of recruiting controls could cause a selection bias. We do not know if the low level of education is a characteristic of people with bipolar disorder or is a bias caused by different recruiting methods between cases and controls.*

In fact, there was no difference in years of education between the complete BD and HC sample (p=.10), the three neurocognitive subgroups (p=.064) or between the BD samples of studies 1-4 (p=.37). We were unable to identify what the results mentioned by the reviewer refers to. We apologize for any possible lack of clarity regarding the educational levels.

*Authors say "**additional participation criteria for the BD patients included significant subjective cognitive difficulties according to the Massachusetts General Hospital Cognitive and Physical Functioning Questionnaire (CPFQ; scores ≥4 on ≥2 cognitive domains)"** but they didn't adopted this criteria for controls, thus authors have selected a specific sample, unrepresentative of the universe of people with BD, that has been compared with unmatched healthy controls.*

We thank the reviewer for this valuable input. We agree that differences in inclusion criteria of studies 1 and 2 could have skewed the selection of the complete BD sample. However, this is not the case as there were no significant difference between studies 1 and 2 (having subjective cognitive impairment as inclusion criterion) vs. studies 3 and 4 (not having subjective cognitive impairment as inclusion criterion) on any of the cognitive domains (p-values≥.062). We have now added this as a limitation in the:

- Discussion, section 4, p. 17: "**Moreover, post-hoc comparison analyses between studies 1 and 2 (having cognitive complaints as inclusion criterion) versus studies 3 and 4 (not**

having cognitive complaints as inclusion criterion) showed no significant differences on any of the cognitive domains (*p*-values≥.062), suggesting that the present BD sample was representative of BD patients in full or partial remission."

Even heavier is the fact that exclusion criterion for the HCs were dyslexia but for BD doesn't: most of the tests adopted have a specific use in dyslexia. Around 17% of US schoolchildren have dyslexia (with very high prevalence in conditions as ADHD really comorbid with BD). It is therefore very likely that the sample of cases had a very high proportion of people with dyslexia. The weight of dyslexia in bipolar disorder is poorly studied but should be. (Cederli M, Language and mathematical problems as precursors of psychotic-like experiences and juvenile mania symptoms. Psychol Med. 2014 Apr;44(6):1293-302. doi: 10.1017/S0033291713002018. Epub 2013 Aug 13). Because the clinical symptoms of dyslexia emerge often before the manifestations of bipolar disorder (23/22 years in average in this sample), to know how many people are suffering from dyslexia can bring a bit of order in a situation in which the co occurrence of several factors (drugs, concurrent disorders, educational level, vascular conditions for older [the age range is wider in people with cognitive deficits] etc.) can really cause a chaos difficult to reorder. Is natural to ask how many people among the nearly 200 with bipolar disorder had dyslexia. When dyslexia had onset?
If the sample would be correctly recruited it would be interesting to know how many had dyslexia between BD and how many among the HC. But I think that it is impossible, thus this issue should be addressed in a section on limits

We thank the reviewer for highlighting this point. We agree that the impact of dyslexia in bipolar disorder would impact their performance and deserves to be assessed explicitly in future studies. Although we did not examine if patients had dyslexia explicitly, there were no patients in the studies who reported any difficulties with completing the several self-administered (written) questionnaires or on any of the neuropsychological tasks.
This lack for formal assessment of potential dyslexia has now been raised as a limitation in the:

- Discussion, section 4 p. 18: "**Another limitation is that we did not formally assess dyslexia in the patients; however, none of the patients reported any difficulties with completing the several self-administered (written) questionnaires or on any of the neuropsychological tasks, indicating that it was unlikely that any patient suffered from dyslexia.**"

-Reviewer #3

I read with much interest this study on neuropsychological findings in BD. In general terms the paper is well written and clearly organised. I have a few general comments:
- The neuropsychological evaluation focuses mainly on frontal lobe function with the exception of RAVLT that explores short term verbal memory. RAVLT in particular correlates with the cognitive decline seen for example in pts with Alzheimer disease while all other tests focus mainly on mental speed which may be due to frontal lobe dysfunction but can be affected by medications as well. It is quite clear from their results that the "selective group" differs from the intact and globally impaired ones but further differentiation would be of value.

We thank the reviewer for this valuable comment. In the interpretation of which factors could explain the neurocognitive subgroups we chose to highlight the relatively higher number of medications prescribed for patients in the 'selective group' as we thought this could have had an impact on the substantially slowed psychomotor speed. In response to the reviewer's request, we have now also included a discussion addressing explanations for the 'globally impaired' group in the:

- Discussion, section 4, p. 15f: **"The globally impaired subgroup with deficits in verbal learning, working memory, and executive skills showed no differences from the other neurocognitive subgroups with respect to illness load, BD subtype or medication. These findings point to trait-related deficits within verbal memory and executive function that is inherent to BD itself (rather than being explained by medication or external factors). Indeed, this interpretation is consistent with the study by Bora et al. (2016) and meta-analyses of deficits within verbal learning and executive function but not in processing speed in individuals at genetic risk for BD (Bora et al., 2009)."**

- The association with antidepressants in the selective group is again linked to the previous point. I suspect that these patients are also in partial remission as frontal lobe dysfunction is a typical neurocognitive correlate of depression. The authors presented the mean HDRS score but how many pts were in partial remission?

We agree with the reviewer that depressive symptoms are generally found to be associated with cognitive impairment. In the present study, we found no differences between the frequency of fully/partially remitted patients measured with the HDRS-17 and YMRS between the three neurocognitive subgroups. In response to the reviewer's request, this information is now included in Table 3, p. 33:

	Subgroup			Statistics		
	Global	Selective	Intact	*df*	*F* or χ^2	*p*
…						
HDRS-17	6 (5)	6 (4)	5 (4)	2	1.74	.178
Full remission, ≤7, no. (%)	**27 (66)**	**38 (60)**	**67 (76)**	**2**	**4.48**	**.106**
Partial remission, 8≤14, no. (%)	**14 (34)**	**25 (40)**	**21 (24)**	**2**	**4.48**	**.106**
YMRS	3 (3)	3 (3)	2 (3)	2	.16	.856
Full remission, ≤7, no. (%)	**39 (95)**	**56 (90)**	**82 (93)**	**2**	**.90**	**.638**
Partial remission, 8≤14, no. (%)	**2 (5)**	**6 (10)**	**6 (7)**	**2**	**.90**	**.638**
…						

- The lack of association with lithium is of interest as it has been associated with cognitive deterioration in BD for a long time. Do they have any data on the duration of the treatment also?

The reviewer raises an interesting point. However, we unfortunately have not data from all four studies on the duration of lithium treatment for BD patients.

- The authors did not find a correlation with the episode type but I agree with authors that the total number of affective episodes may be of relevance. In addition the total number of mixed episodes would be even more interesting. Do they have this data?

We thank the reviewer for this point and agree that the total number of mixed episodes would be of particular interest. Unfortunately we do not have data from all four studies on the number of mixed episodes. Also, the number of depressive or manic episodes were categorized 1/2/≥3 given uncertainty in their estimation in the original studies. Therefore, we were unable to conduct analyses of the impact of total number of depressive/manic episodes since the vast majority of BD participants would be placed in the "≥3" category. This has now been raised as a limitation in the:

- Discussion, section 4, p. 18: **Another limitation of the study was the lack of precise data on previous of affective episodes beyond >3 or psychotic episodes. However, the comparable proportion of BD type I patients in the present and previous study by Burdick et al. (2014) suggests that differences in history of psychotic symptoms may not explain the larger group of cognitively intact patients in our study.**

-***Reviewer #4***

-

Comment: Based on recent findings by Burdick et al.'s the authors aim to test whether there are discrete neurocognitive subgroups among BD patients. Furthermore, they want to assess to which extent neurocognition affects functional abilities, perceived stress, subjective cognitive complaints, work capacity and QoL. Pooled data from 4 studies was examined using discriminant analyses, multiple regressions and ANOVAs. Three neurocognitive groups were detected: a cognitively intact, selectively, and globally impaired. The cognitively impaired groups displayed greater stress, cognitive complaints, poorer work capacity, and reduced QoL. I commend the authors for 1. Gathering a large BD sample, 2. Using a range of statistical techniques to thoroughly assess the data, 3. Examining the link between cognition and global functioning. Given these strengths this paper presents novel findings that deserve to be published in this journal. I would, however, like to recommend a few revisions to improve the understanding of the study rationale, methodology, and interpretation of the findings.
Title: I would mention the words "global functioning" or refer to functional abilities.

We thank the reviewer for this suggestion and have now changed the title (p. 1) from: "Discrete neurocognitive subgroups in fully or partially remitted bipolar disorder" to "Discrete neurocognitive subgroups in fully or partially remitted bipolar disorder**: associations with functional abilities**"

Introduction: I would encourage the authors to provide additional information on the link between cognition and global functioning, e.g. perceived stress, QoL. I would also add some mention of potential cognitive differences or similarities between fully remitted/partially remitted BD patients. This would "solidify" the study rationale, and help the readers understand 1.why it is important to study such topic and 2. why studying remitted/partially remitted BD patients is meaningful.

Thank you for this valuable input (as also noted by other reviewers # 1 and 7). We have now revised following in the manuscript:

- Introduction, section 1, p. 2: **Cognitive impairment is among the strongest contributors together with mood symptoms and illness progression to functional disability, lower quality of life, and loss of workforce capacity in BD (Bonnín et al., 2010; Brissos et al., 2008; Torrent et al., 2012). However, the correlation between subjectively experienced and objectively measured cognitive impairment is poor, as shown by our and other research groups (Burdick, Endick and Goldberg, 2005; Jensen et al., 2015; Rosa et al., 2013).**
- Introduction, section 1, p. 2: Replaced "This points to cognitive dysfunction as a key treatment priority to improve patients' functional recovery." with "This points to **objectively measured** cognitive dysfunction as a key treatment priority to improve patients' functional recovery **and quality of life after acute mood episodes**."
- Introduction, section 1, p. 3: **Several studies have examined cognition in BD using predetermined cut-offs (typically 1-2 standard deviations [SD] from the normal mean [M]) for simple classification of impaired and non-impaired patients (e.g., Jensen et al., 2015; Martino et al., 2014; Reichenberg et al., 2009; Rojo et al., 2010; Volkert et al., 2015). However, there is a scarcity of studies that have used a *data-driven* approach to identify neurocognitive subgroups.**
- Introduction, section 1: Replaced "However, a recent study of N=136 BD patients in full or partial remission suggests that cognitive impairment is not a uniform feature of BD; instead discrete neurocognitive subgroups were found; a "cognitively intact" subgroup with scores equivalent to the performance of healthy control persons (HCs) (31.6%), a subgroup with "selective cognitive impairment" (28.7%), and a subgroup with "global severe impairment" across cognitive domains (39.7%) comparable to cognitive deficits in schizophrenia (Burdick et al., 2014)" with "**Such studies found discrete neurocognitive subgroups in fully or partially remitted** patients with BD **(Bora et al., 2016; Burdick et al., 2014; Lewandowksi et al., 2014)**; a **well performing** "cognitively intact" subgroup with scores equivalent to the performance of healthy control persons (HCs), **one or two** subgroups **of** "selective cognitive impairment" **with lower cognition scores compared to HCs**, and a subgroup with "global

severe impairment" across cognitive domains comparable to cognitive deficits in schizophrenia."

- Introduction, section 1, p. 3: **The findings from these studies show some discrepancies with respect to the proportion of patients in each subgroup, possibly due to differences in study samples (e.g., affective symptoms, age, medicine prescribed).**
- Introduction, section 1, p. 3: Replaced "However, as the described study (11) is the only published evidence for the presence of neurocognitive subgroups in BD, additional replication studies are needed to make any firm conclusions about the pattern of cognitive dysfunction in BD." with "Additional **data-driven subgroup** studies are needed to make any firm conclusions about the pattern of cognitive in BD."
- Introduction, section 1.1, p. 3: Replaced "The study aims to 1) determine whether there are discrete neurocognitive subgroups in fully or partially remitted patients with BD as prior demonstrated (Burdick et al., 2014). Building onto the findings of Burdick et al. (Burdick et al., 2014), the study aims to 2) examine the pattern of the cognitive deficits and the clinical and functional characteristics of the neurocognitive subgroups in terms of perceived stress, cognitive complaints, work capacity, quality of life, and medication prescribed." with "The study aims to **identify** discrete neurocognitive subgroups in **a large group of** fully or partially remitted patients with BD **using a data-driven approach**. Building onto the **few studies applying this approach**, the **present** study aims to examine the pattern of the cognitive deficits and the clinical and functional characteristics of the neurocognitive subgroups in terms of perceived stress, cognitive complaints, work **and social adjustment**, quality of life, and medication prescribed."
- References, p. 20: **Bonnín, C. M., Martínez-Arán, A., Torrent, C., Pacchiarotti, I., Rosa, A. R., Franco, C., Murru, A., Sanchez-Moreno, A., Vieta, E., 2010. Clinical and neurocognitive predictors of functional outcome in bipolar euthymic patients: a long-term, follow-up study. J. Affect. Disord. 121, pp. 156-160.**
- References, p. 20: **Bora, E., Hıdıroğlu, C., Özerdem, A., Kaçar, Ö. F., Sarısoy, G., Arslan, F.C., Aydemir, Ö., Tas, Z.C., Vahip, S., Atalay, A., Atasoy, N., Atesci, F., Tümkaya, S., 2016. Executive dysfunction and cognitive subgroups in a large sample of euthymic patients with bipolar disorder. Eur. Neuropsychopharmacol. DOI: http://dx.doi.org/10.1016/j.euroneuro.2016.04.002**

- References, p. 21: **Brissos, S., Dias, V.V., Carita, A.I., Martinez-Arán, A., 2008. Quality of life in bipolar type I disorder and schizophrenia in remission: clinical and neurocognitive correlates. Psychiatry Res. 160, pp. 55-62.**
- References, p. 23: **Lewandowski, K.E., Sperry, S.H., Cohen, B.M., Öngür, D., 2014. Cognitive variability in psychotic disorders: a cross-diagnostic cluster analysis. Psychol. Med. 44, pp. 3239-3248.**
- References, p. 23: **Martino, D.J., Strejilevich, S.A., Marengo, E., Ibañez, A., Scápola, M., Igoa, A., 2014. Toward the identification of neurocognitive subtypes in euthymic patients with bipolar disorder. J. Affect. Disord. 167, pp. 118-124.**
- References, p. 24: **Rojo, E., Pino, O., Guilera, G., Gómez-Benito, J., Purdon, S.E. , Crespo-Facorro, B. , Cuesta, M., Franco, M., Martinez-Arán, A., Segarra, N., Tabarés-Seisdedos, R., Vieta, E., Bernado, M., 2010. Neurocognitive diagnosis and cut-offscores of the Screen for Cognitive Impairment in Psychiatry (SCIP-S). Schizophr. Res. 116, pp. 243-251.**
- References, p. 25: **Torrent, C., Martinez-Aran, A., del Mar, B.C., Reinares, M., Daban, C., Sole, B., Rosa, A.R., Tabares-Seisdedos, R., Popovic, D., Salamero, M., Vieta, E., 2012. Long-term outcome of cognitive impairment in bipolar disorder. J. Clin. Psychiatry 73, pp. e899-e905.**
- References, p. 26: **Volkert, J., Kopf, J., Kazmaier, J., Glaser, F., Zierhut, K.C., Schiele, M.A., Reif, A., 2015. Evidence for cognitive subgroups in bipolar disorder and the influence of subclinical depression and sleep disturbances. Eur. Neuropsychopharmacol. 25, pp. 192-202.**

Methods: please explain why exclusion criteria for HC included dyslexia and not intellectual disability, or learning disabilities for instance? Why does this exclusion criteria apply to HC only?

Thank you for your valuable comment. We estimated dyslexia in the HCs to establish a representative normative sample, but we did not formally assess dyslexia in the patients with BD. None of the patients reported any difficulties with reading or completing the several self-administered (written) questionnaires or on any of the neuropsychological tasks: if a patient with BD had had dyslexia, this would have been expressed at the assessment session. Thus, there are most likely no patients with dyslexia in the BD sample. This information is now added to the manuscript:

- Discussion, section 4, p.18: ”**A limitation was that we did not formally assess dyslexia in the patients with BD; however, none of the patients reported any difficulties with completing**

the several self-administered (written) questionnaires or on any of the neuropsychological tasks. Thus, there were most likely no patients with dyslexia in the complete BD sample."

-did the authors plan to correct results for mood symptoms? Although there are no statistically significant differences across groups subthreshold mood symptoms have been shown to modulate cognition. This issue should be addressed appropriately.

We agree with the reviewer that even sub-syndromal mood symptoms can impact on cognition. We therefore included measures of subthreshold mood symptoms (i.e., HDRS-17 and YMRS scores) as independent variables in the multiple regression model together with neurocognitive subgroup (intact/selective/global: 1/2/3), lithium, antipsychotics, and antidepressants (yes/no), benzodiazepines, age, gender, illness duration, number of depressive episodes, manic episodes, number of medications from different classes to examine what may have contributed to explain cognitive functioning in the BD patients. These regression analyses showed that neither HDRS-17 nor YMRS scores were significant predictors of the any of the cognitive domains, suggesting that they could not explain the cognitive impairments in these subgroups.

-what kind of diagnostic tool did the 4 studies included in this manuscript use to determine the participants' BD diagnosis? E.g.SCID? How many participants were euthymic/depressed/etc.?

We thank the reviewer for raising this point (as also noted by reviewer # 3 and 6). In two of the studies (the intervention trials) from which the data was pooled, patients were diagnosed with the Schedules of Clinical Assessment in Neuropsychiatry (Wing et al., 1990). In the two remaining studies, patients were diagnosed by specialists in psychiatry at the Clinic for Psychiatric Disorder, Psychiatric Centre Copenhagen. This information along with the frequency of patients in full or partial remission is now included in the manuscript:

- Materials and methods, section 2.2, p. 4: **All patients were diagnosed by specialists in psychiatry at the Copenhagen Clinic for Affective Disorders, Psychiatric Centre Copenhagen, and diagnoses of patients from studies 1 and 2 were confirmed using the Schedules for Clinical Assessment in Neuropsychiatry (Wing et al., 1990)**.
- Table 3, p. 29:

	Subgroup	Statistics

	Global	Selective	Intact	*df*	*F* or χ^2	*p*
…						
HDRS-17	6 (5)	6 (4)	5 (4)	2	1.74	.178
Full remission, ≤7, no. (%)	**27 (66)**	**38 (60)**	**67 (76)**	**2**	**4.48**	**.106**
Partial remission, 8≤14, no. (%)	**14 (34)**	**25 (40)**	**21 (24)**	**2**	**4.48**	**.106**
YMRS	3 (3)	3 (3)	2 (3)	2	.16	.856
Full remission, ≤7, no. (%)	**39 (95)**	**56 (90)**	**82 (93)**	**2**	**.90**	**.638**
Partial remission, 8≤14, no. (%)	**2 (5)**	**6 (10)**	**6 (7)**	**2**	**.90**	**.638**
…						

- References, p. 26: **Wing, J.K., Babor, T., Brugha, T., Burke, J., Cooper, J.E., Giel, R., Jablenski, A., Regier, D., Sartorius, N., 1990. SCAN: Schedules for Clinical Assessment in Neuropsychiatry. Arch. Gen. Psychiatry, 47, pp. 589-593.**

-could the authors provide the type of comorbidities in each "cognitive subgroup"?

We agree with the reviewer that this would have been beneficial for the present study. Unfortunately we do not have data on medical or psychiatric comorbidities (although no patient had any other axis 1 diagnosis).

-how did the authors select the questionnaires assessing work capacity, subjective cognitive difficulties etc.?

The questionnaires applied in the present study were chosen based on the large overlap of assessment tools applied in the four studies from which the complete data was pooled. This has now been clarified in:

- Materials and methods, section 2.4, p. 5: "**We chose these questionnaires of functional ability since all measures were applied in studies 1-4.**"

-why didn't the authors select immediate learning for instance, rather than creating a composite score including both immediate learning and recall? The former measure may be even more diagnostically relevant than the latter.

We agree with the reviewer that measures of immediate learning may be of diagnostic relevance. However, the aim of the present study was to examine the pattern of cognitive deficits and functional abilities in discrete neurocognitive subgroups in BD. Thus, we chose to establish a cognitive domain covering verbal learning and memory rather than just immediate learning for a more fulfilling approach in relation to our aim.

-why didn't the authors select a work and social adjustment scale that could be applicable to both BD and HC? If they couldn't standardize WSAS to HC, would it make sense to use this measure anyway?

We thank the reviewer for raising this point. Given that the data was pooled from studies examining neurocognitive function in BD, we had employed the WSAS as a relevant measure for this group. Further, the primary aim of the study was not to investigate the HC group but to examine neurocognitive subgroups in BD and their associations with functional measures.

-did the authors check the data for collinearity? Could they provide the correlational matrix between cognitive and functional scores?

We thank the reviewer for raising this important point. Variance Inflation Factors (VIF) for all cognition and functional measures were <3, which should not be source of concern (<10; Myers, 1990). This indicates that none of the variables share equal variance. Because of space limitations we chose not to include a matrix of cognition and functional variables into the manuscript. However, in response to the reviewer's point we have now added following to the manuscript:

- Results, section 3.4, p. 12: "**Variance inflation values for all cognition and functional measures were <3, suggesting there should be no concern for multicollinearity in the analyses (Myers, 1990).**"
- References, p. 24: **Myers, R., 1990. Classical and modern regression with applications, 2nd edition, Boston, MA: Duxbury**

-could the authors please describe in their statistical analyses section the approach used for the multiple regressions (mentioned in 3.3)? The number of variables included in these regressions is certainly elevated. Could this be a source of concern when interpreting the findings?

We agree with the reviewer that there are a high number of independent explanatory variables in the post-hoc multiple regression analyses. However, these independent variables were included to examine

their potential relevance for the cognition scores and thus neurocognitive subgroups in this exploratory study. Also there were no issues with multicollinearity as mentioned above, which could have led to variables 'cancelling out' one another.

-why was benzodiazepine considered as an individual variable instead of being included in "medication"?

We thank the reviewer for raising this point which is particularly relevant in relation to cognition. We included "benzodiazepines" as a separate independent variable because this type of medicine is associated with cognitive side-effects and we wanted to investigate whether prescription of benzodiazepines differed between the neurocognitive subgroups.

Results: in 3.3 could the authors clarify what they mean by "cognitively impaired subgroup"? Are they referring to globally or selectively cognitively impaired participants?

We agree with the reviewer that this is unclear. To increase clarity for the reader we have now explicitly defined "cognitively impaired subgroup" as both the selective and global subgroup in the manuscript:

- Results, section 3.3, p. 11: "[…] were categorization in **the selectively or globally impaired subgroup** (β=-.55, p<.01) and being male (β=-.32, p<.01) […] were categorization in **the selectively or globally impaired subgroup** (β=-.56, p<.01), […] processing speed (F(11)=4.72, p<.01, r^2=.25, adjusted r^2=.20) was categorization in **the selectively or globally impaired subgroup** (β=-.37, p<.01)."
- Results, section 3.5, p. 12: "This revealed that **the combined subgroup of selectively and globally impaired patients** had fewer […]. **These either selectively or globally** impaired patients were also more […]"

Discussion: could the authors please clarify if on page 11, line 6 and on page 12, line 2, they meant "age" or "age of onset"? I thought they were indeed age differences across BD groups.

We thank the reviewer for noticing these typos which have now been corrected in the manuscript:

- Discussion, section 4, p. 13: "[…] other clinical characteristics including age **of onset**, illness load, or medications."
- Discussion, section 4, p. 14: "[…] evidence for an absence of differences between the globally impaired and intact subgroup in age **of onset**, severity of affective symptoms, illness load […].

Details: In 1.1. please discard 1) and 2) as it is more confusing than helpful.

In response to the reviewer's request, the "1)" and "2)" are now deleted from the Introduction, section 1.1, p. 3.

-Reviewer #5

- The aim of this study was to identify subgroups of cognitive imapirment among cohorts in partially or fully remitted bipolar patients. Thuy have pooled dat from 4 different trials for the purposes of analysis. It would have been helpful for some details of the trials from which that data had been provided, especasilly as to whether there were specific inclusion criteria for the trials (for example patients with cognitive complaints were recruited into study 2), there were some common inclusion criteria although it is not clear what the cut off on the YMRS was (in the text '8<14').

We thank the reviewer for highlighting this relevant point. We included participants who were in partial or full remission according to the HDRS-17 *and* the YMRS. On both these scales scores in the range of 0-7 indicate full remission, and scores in the range of 8-14 indicate partial remission. For clarity reasons we have now included following in the manuscript:

- Materials and methods, section 2.2, p. 4: "193 eligible adult patients with an ICD-10 diagnosis of BD in full or partial remission (defined as Hamilton Depression Rating Scale 17-item [HDRS-17] **scores of ≤7 or 8≤14, respectively, and** Young Mania Rating Scale [YMRS] scores of ≤7 **or** 8≤14, respectively) (Hamilton, 1960; Young et al., 1978), and 110 HCs."
- Materials and methods, section 2.2, p. 4: "**All patients were diagnosed by specialists in psychiatry at the Clinic for Psychiatric Disorder, Psychiatric Centre Copenhagen, and diagnoses of patients from studies 1 and 2 were confirmed using the Schedules for Clinical Assessment in Neuropsychiatry (Wing et al., 1990)**."
- Materials and methods, section 2.2, p. 5:" Exclusion criteria **in studies 1-4** included prior history of schizophrenia, current substance abuse, substantial somatic illness or a daily use of benzodiazepines ≥22.5 mg oxazepam."
- Materials and methods, section 2.2, p. 5: "Exclusion criteria for the HCs **(studies 3 and 4)** were dyslexia or any personal or family history of mental illness."

- References, p. 26: "**Wing, J.K., Babor, T., Brugha, T., Burke, J., Cooper, J.E., Giel, R., Jablenski, A., Regier, D., Sartorius, N., 1990. SCAN: Schedules for Clinical Assessment in Neuropsychiatry. Arch. Gen. Psychiatry, 47, pp. 589-593.**"

They point out the common exclusion criteria - why was the particluar dose cutoff for benzodiazepine chosen.

We thank the reviewer for raising this valuable point. The cut-off for benzodiazepines of ≥22.5 mg oxazepam was applied in the original studies from which the data were pooled to diminish the risk of substantial impact of benzodiazepines on cognitive function. Due to generalizability of the findings we did not apply benzodiazepines as exclusion from study participation.

A concern is that there may be bias in the individual studies.

Building on the aforementioned point there were no bias in studies 1-4 as these were comparative for inclusion and exclusion criteria and use of benzodiazepines.

The three cognitive domains appear to have been selected on a priori grounds, was using a data reduction technique such as factor analysis considered to identify underlying cognitive domains?

We thank the reviewer for highlighting this important point. The three cognitive domains were not selected on a priori grounds as there is some sort of consensus in the literature of the grouping of cognitive applied in the present study (Lezak et al., 2012). Also, we chose this specific grouping to be consistent with the approach applied in previous studies of our group (e.g., Demant et al., *PLoS ONE*, 2015; Jensen et al., *J Affect Disord*, 2015; Ott et al., *J Affect Disord*, 2016). Of course, it is still debatable whether this is the best approach, and that the neuropsychological tests may be categorized different based on output from data reduction techniques due to test impurity. To clarify this for the reader, we have now added following to the manuscript:

- Materials and methods, section 2.5.1, p. 7: "**We chose this assignment of neuropsychological tests into cognitive domains given some consistency of this domain grouping in the literature (Lezak et al., 2012) and** to be consistent with previous studies **from our group (Jensen et al., 2015; Ott et al., 2016)**."

It would have been helpful to give a brief description of the clusters at the start of section 3.2

In response to the reviewer's request, we have now added following to the manuscript:

- Results, section 3.2, p. 9: "[…] one subgroup of 89 patients (46.1%) **with "intact" cognitive function**, another subgroup of 63 patients (32.6%) **with "selective" cognitive impairment,** and a third subgroup of 41 **patients (21.2%) with "global" cognitive difficulties** (see Table 2)."

How much of the slower processing speed be attributed to a medication effect, rather than an illness effect?

This is a highly relevant question. Among the BD sample there was a weak-moderate correlation between slower processing speed and higher number of medications prescribed (r=.24, p<.01). We could not directly address the question about the effect of "illness" on slowing of processing unless this relates to sub-syndromal mood symptoms. Indeed, we found borderline significant weak correlations between slower processing speed and more sub-syndromal depressive symptoms (HDRS-17: r=.14, p=.051), and slower processing speed and more (hypo-) manic sub-syndromal symptoms (YMRS: r=.13, p=.07) in the BD sample. Given the assumptions described in this paragraph, "medication" contributed significantly more than "illness" to slowing of processing speed.

A problem with theri data is that they have not recorded the number of episodes beyond >3 episodes, the importance of this is that cognitive defects could be the result of the neurotoxic effects of each episode of illness, especailly mania.

We agree with the reviewer on this important point (as also mentioned by reviewers # 1 and 3), and thus we have now added this limitation in the:

- Discussion, section 4, p. 18: **Another limitation of the study was the lack of precise data on previous of affective episodes beyond >3 or psychotic episodes. However, the comparable proportion of BD type I patients in the present and previous study by Burdick et al. (2014) suggests that differences in history of psychotic symptoms may not explain the larger group of cognitively intact patients in our study.**

-Reviewer #6

-

The authors present an analysis of pooled cognitive data from four studies in remitted/partially remitted individuals with bipolar disorder (BD). Through hierarchical cluster analysis, the authors report the emergence of 3 separate subgroups of individuals with BD on the basis of their current cognitive functioning, which they label as: intact, selectively impaired and globally impaired.

This is an important topic given the increasing anecdotal and emerging empirical evidence for substantial variability in the clinical and cognitive presentation of individuals with BD. In this light the main aims of this study are laudable. However, I have some questions that require clarification:

1. *In a footnote in the materials and methods section the authors report that 26 participants in study 3 are overlapping with participants in study 4. Can the authors clarify if these participants have been analyzed twice? I may be reading this wrong but it sounds like all 25 BD participants from study 4 were included in study 3. If this is the case, it is prudent to remove any overlap and only report that data from BD individuals from studies 1-3 were used.*

   We thank the reviewer for pointing out this important point and apologize for the linguistic ambiguity. As the data of the 26 participants were only analyzed once, we have now deleted the footnote to increase clarity for the reader.

2. *The Miskowiak 2014 study is a clinical trial. Can the authors clarify that only the baseline cognitive data from that study was used in this analysis?*

   In response to the reviewer's request, this point is now clarified in the:
   - Materials and methods, section 2.1, p. 4: "The present study involved pooling of **available *baseline* data from four studies of our research group: two clinical trials targeting cognition (study 1: Miskowiak et al., 2014; study 2: Demant et al., 2015) and from two cross-sectional observational studies of which one has been completed (study 3: Jensen et al., 2015; Ott et al., 2016) and one is ongoing (study 4)** […]"

3. *Can the authors clarify whether remission or partial remission is defined on the basis of scores on the YMRS AND the HRSD? Table 2 seems to suggest that individuals with a score of <7 0r 8-14 on either scale were considered remitted/partially remitted. However, typically scores on both scales would be considered in remission criteria.*

We thank the reviewer for highlighting this point (as also noted by reviewer # 3 and 4) We included participants who were in partial or full remission according to the HDRS-17 *and* the YMRS. On both these scales scores in the range of 0-7 indicate full remission, and scores in the range of 8-14 indicate partial remission. To improve clarity for the reader, we added following in the manuscript:

- Materials and methods, section 2.2, p. 4: "[…] ICD-10 diagnosis of BD in full or partial remission (defined as Hamilton Depression Rating Scale 17-item [HDRS-17] **scores of ≤7 or 8≤14, respectively, and** Young Mania Rating Scale [YMRS] scores of ≤7 **or** 8≤14, respectively) […]".
- Table 3, p. 29:

	Subgroup			Statistics		
	Global	Selective	Intact	*df*	*F* or χ^2	*p*
…						
HDRS-17	6 (5)	6 (4)	5 (4)	2	1.74	.178
Full remission, ≤7, no. (%)	**27 (66)**	**38 (60)**	**67 (76)**	**2**	**4.48**	**.106**
Partial remission, 8≤14, no. (%)	**14 (34)**	**25 (40)**	**21 (24)**	**2**	**4.48**	**.106**
YMRS	3 (3)	3 (3)	2 (3)	2	.16	.856
Full remission, ≤7, no. (%)	**39 (95)**	**56 (90)**	**82 (93)**	**2**	**.90**	**.638**
Partial remission, 8≤14, no. (%)	**2 (5)**	**6 (10)**	**6 (7)**	**2**	**.90**	**.638**
…						

4. *The inclusion of the digits forward score in the measure of working memory is questionable, given that digits forwards is considered to be a measure tapping attention span/immediate memory rather than true working memory.*

We thank the reviewer for raising this highly relevant issue (as also noted by reviewer # 5 and 7). We are aware that there is no complete consensus about grouping of cognitive domains in the literature and chose t some groupings on which there is some sort of consensus for, as described by Lezak et al. (2012; for *Digit Span Forward* see p. 649). Of course, it is still debatable whether this is the best approach, and that the neuropsychological tests may be categorized different due to test impurity. To increase clarity for the reader on the rationale for the present groupings, we have now added following description in the manuscript:

- Materials and methods, section 2.5.1, p. 7: "**We chose this assignment of neuropsychological tests into cognitive domains given some consistency of this domain grouping in the literature (Lezak et al., 2012) and** to be consistent with previous studies **from our group (Jensen et al., 2015; Ott et al., 2016)**."
- References, p. 23: **Lezak, M.D., Howieson, D.B., Bigler, E.D., Tranel, D., 2012. Neuropsychological Assessment, 5th edition. Oxford University Press, New York.**

5. *What is the justification for the use of post-hoc LSD tests, other than the fact that Burdick et al., used this method? Post-hoc LSD tests do not really control for multiple comparisons, but rather assume differences between groups on the basis of a significant omnibus test. Thus, the likelihood of error is increased.*

We agree that LSD was used for comparison reasons with the study by Burdick et al. (2014). In the paper by Burdick et al. (2014) it is written that "*Planned contrasts using the HC group as a comparison category was done to detect differences between groups and results were corrected for multiple comparisons with least significant difference (LSD) correction*." (p. 3086). However, we were unable to find another reference confirming that LSD corrects for multiple comparisons. We have now included the following in the manuscript:

- Discussion, section 4, p. 18: "**Finally, the use of LSD post-hoc comparison could have increased the likelihood of error due to lack of correction for multiple comparisons."**
- Materials and methods, section 2.5.2, p. 7: deleted: "adjust for multiple comparison and" in "One-way analyses of variance (ANOVA) and least significance difference (LSD) as post-hoc pairwise comparison were applied **to adjust for multiple comparisons and** compare the cognition scores of the BD subgroups and HCs. We chose this approach to keep the analyses

consistent with the only previously published analyses of neurocognitive subgroups in BD by Burdick et al. (2014)."

6. *Given subtle differences in TMT-A performance, meds, symptoms etc, should site not be used as a covariate in all analyses comparing subgroups on the cognitive measures?*

We thank the reviewer for this point. It is not possible to enter covariates in the HCA. Although covariates could be included in the post-hoc comparisons of the functional variables between the neurocognitive subgroups, the focus of the present analyses was to explore the clinical characteristics of the neurocognitive subgroups and it was thus not particularly meaningful for this purpose to adjust for the study from which the data were pooled.

7. *It would be useful for the authors to explicitly state which variables significantly loaded on each factor that emerged in the DFA so that the reader can determine each cognitive domain's importance in discriminating the groups.*

We thank the reviewer for the suggestion to have included this highly relevant information to the reader. In response to this, we have now added the standardized canonical discriminant function coefficient of the cognitive domains from the DFA to the manuscript:

- Results, section 3.2, p. 9: **"Processing speed and verbal memory contributed most to classify BD patients into neurocognitive subgroups (Standardized Coefficients for Function 1: verbal memory =.14, working memory and executive skills=.01, processing speed=.98; Standardized Coefficients for Function 2: Verbal memory =.72, working memory and executive skills=.69, processing speed=-.27)."**

8. *The authors should specify that although the global group performed within the 'normal' range on measures of processing speed, this group was in fact significantly impaired on this measure in comparison to controls.*

We agree with the reviewer that the global subgroup had statistically significant slower processing speed compared to the healthy controls. However, having the predefined cut-off for cognitive impairment (i.e., $z \leq -1$) in mind, we cannot argue that the global subgroup score of $z=.61$ on processing

speed reflects *clinically* significant cognitive impairment in this domain. We have now clarified this important point by adding the following information:

- Results, section 3.2, p. 10: "[…] while processing speed was **significantly slower compared to HCs (LSD: *p*<.01) but still** within the normal range (*z*=-.61)."

9. *I am not quite sold on the naming convention for the 'global' group. I acknowledge that the authors wised to keep with the methods and naming conventions of Burdick et al – however the measures used by Burdick et al., were more comprehensive and sampled more domains than those used here, and speak more truly to the idea of 'global' impairment than the test/results presented here.*

We agree to some extent with the reviewer on this point. Indeed, in the present study the "globally" impaired subgroup showed cognitive dysfunction in verbal learning, working memory, and executive skills, and presented significant slowing of processing speed compared to the HCs. However, after careful consideration we have chosen to keep the label of the subgroup with "global" cognitive impairment (a) given the more broad cognitive difficulties in this group and (b) to keep the naming of the subgroups consistent with the labels of the three neurocognitive subgroups in the study by Burdick et al. (2014).

10. *The authors may wish to consider the use of step-wise regression to determine whether inclusion within one of the impaired groups was the most important predictor of performance on each of the cognitive domains.*

We thank you for this suggestion. Our reason for choosing multiple regression using forced entry was that this method relies on theoretical groundings rather than statistical criteria. The disadvantage of step-wise regression in this context is that it assesses the fit of one variable based on other variables in the model, which is problematic because of the complex interplay between these factors. For instance, psychotropic medicine is both known to have positive and negative impact on cognition by reducing affective symptoms and the sedative effect itself, respectively. Stepwise regression may have some advantages for exploration, but identification of the most important predictor of performance within a neurocognitive subgroup is beyond the scope of the present study. We have now added following in the manuscript:

- Results, 3.3, p. 11: “Following variables were entered as independent variables **using forced entry**: […]”
- Discussion, section 4, p. 18: “**We used forced entry multiple regression for the post-hoc analyses. The disadvantage of this method to step-wise regression is that all independent variables are entered to the model simultaneously, which could possess a problem if multicollinearity was a source of concern (this was not the case in the present study). However, the step-wise methods are limited by selecting explanatory variables of the model based on statistical criteria rather than theoretical groundings (Field, 2013).**”

11. *Given differences in age between the subgroups and between the subgroups and controls, all cognitive comparison analyses should be re-run with age as a covariate. I can see that this was done for the analyses of clinical variables, but why this was not done for the cognitive comparisons is not clear.*

We thank the reviewer for raising this valuable point. We conducted such post-hoc ANCOVA for differences in the cognition scores between the neurocognitive subgroups and HCs adjusting for age, which gave similar results to the ANOVA analyses of no adjustment for age. This indicates that cognitive differences in neurocognitive subgroups cannot be attributed to differences in age between subgroups. We have now added this to the manuscript:

- Results, section 3.4, p. 11f: “**Age-adjusted comparison analyses for cognition scores among the neurocognitive subgroups and HC sample yielded similar results from comparison analyses without adjustment for age (significant in non-adjusted analysis: age-adjusted *p*-values≤.02; not significant in non-adjusted analysis: age-adjusted *p*-values≥.24).”**
- Discussion, section 4, p. 15: “**Indeed, age cannot be attributed to differences in cognition between the neurocognitive subgroups since age-adjusted and non-adjusted comparison analyses of cognitive function between subgroups showed similar results.”**

12. *One major limitation that the authors have not acknowledged is that their study was not comprehensive in its measurement of the various domains of cognition. It is possible that different results may have been obtained if a more comprehensive battery of measures were used*

We thank the reviewer for raising this this critique which has also been raised by reviewer # 1 and with which we agree. We have now highlighted this as a major limitation:

- Discussion section 4, p. 17: "**Relative to Burdick et al. (2014) the somewhat smaller battery of neuropsychological tests applied in the present study may limit the evaluation of general cognitive deficit. Nevertheless, our findings are highly consistent with the study by Burdick et al. (2014) and the fewer tests should not result in smaller effect size for cognitive deficits since computing of effect sizes relies on the *M* and *SD* of the performance by our HC reference group on each test (which is then averaged to create composite scores for each domain; see details in the methods section). Nevertheless, future investigations of the cognitive heterogeneity in BD should include a more comprehensive battery of neurocognitive tests also including tests of social cognition and visual learning as in Burdick et al. (2014; 2015).**"
- References, p. 20: **"Burdick, K.E., Ketter, T.A., Goldberg, J.F., Calabrese, J. R., 2015. Assessing cognitive function in bipolar disorder: challenges and recommendations for clinical trial design. J. Clin. Psychiatry, 76."**

-Reviewer #7

- The present manuscript reports on a relatively neglected area of research - neurocognitive heterogeneity among individuals with bipolar disorders (BD). The authors found three discrete clusters or subgroups, and replicated some previous findings in the literature. Although being of potential interest, this study presents several problems, as follows.

1. The authors claim that their study is the second one on this topic, building upon the Burdick et al´s report (2014). That is not entirely correct since to my knowledge there have been at least two other similar examinations of the neurocognitive clusters in BD (Martino et al, 2014 with n=100; and Volkert et al, 2015 with n=70). A solid step of any manuscript is a thorough literature review, which is lacking in this case. Therefore, enthusiasm throughout the manuscript should be tempered and the present results should be also discussed with those of the abovementioned studies. Moreover, that could also change the choice of analysis, which was based on the (incorrect) assumption that the Burdick et al´s study was ´the only previously published analysis of neurocognitive subgroups in BD´as mentioned

in the last sentence of section 2.5.2. For instance, Martino et al (2014) used soft and hard criteria to define such subgroups. The authors should consider this in their analyses.

We appreciate the valuable input from the reviewer (as also noted by reviewer # 1 and 4). The aim of the study was to identify neurocognitive subgroups in BD using a *data-driven* approach (Bora et al., 2016; Burdick et al., 2014; Lewandowksi et al., 2014) rather than subgrouping of cognitively impaired BD patients using predetermined cutoffs (e.g., Jensen et al., 2015; Martino et al., 2014; Reichenberg et al., 2009; Rojo et al., 2010; Volkert et al., 2015). This aim and some more qualification of the study are now stated explicitly in the manuscript. We agree with the reviewer that a solid step of any manuscript is a thorough literature review, and in response to the reviewer's request we discussed the present findings with more of previous similar studies. Indeed, we chose to keep our approach consistent with Burdick et al. (2014) of comparison reasons, since Bora et al. (2016) examined executive function solely in fully remitted BD patients (HDRS-17 and YMRS scores ≤7), and Lewandowski et al. (2014) included patients with schizophrenia or schizo-affective disorder along with BD patients in their analysis. We agree to some extent with the reviewer that it may have been beneficial for the present study to apply soft and hard criteria to define cognitive impairment. However, we chose not to use such criteria since the focus of the present study was to identify neurocognitive subgroups using data-driven statistical procedures and not predetermined cutoffs.
Specifically, we have performed the following changes to the manuscript:

- Introduction, section 1, p. 2: **Cognitive impairment is among the strongest contributors together with mood symptoms and illness progression to functional disability, lower quality of life, and loss of workforce capacity in BD (Bonnín et al., 2010; Brissos et al., 2008; Torrent et al., 2012). However, the correlation between subjectively experienced and objectively measured cognitive impairment is poor, as shown by our and other research groups (Burdick, Endick and Goldberg, 2005; Jensen et al., 2015; Rosa et al., 2013).**
- Introduction, section 1, p. 2: Replaced "This points to cognitive dysfunction as a key treatment priority to improve patients' functional recovery." with "This points to **objectively measured** cognitive dysfunction as a key treatment priority to improve patients' functional recovery **and quality of life after acute mood episodes**."
- Introduction, section 1, p. 3: **Several studies have examined cognition in BD using predetermined cut-offs (typically 1-2 standard deviations [SD] from the normal mean [M])**

for simple classification of impaired and non-impaired patients (e.g., Jensen et al., 2015; Martino et al., 2014; Reichenberg et al., 2009; Rojo et al., 2010; Volkert et al., 2015). However, there is a scarcity of studies that have used a *data-driven* approach to identify neurocognitive subgroups.

- Introduction, section 1: Replaced "However, a recent study of N=136 BD patients in full or partial remission suggests that cognitive impairment is not a uniform feature of BD; instead discrete neurocognitive subgroups were found; a "cognitively intact" subgroup with scores equivalent to the performance of healthy control persons (HCs) (31.6%), a subgroup with "selective cognitive impairment" (28.7%), and a subgroup with "global severe impairment" across cognitive domains (39.7%) comparable to cognitive deficits in schizophrenia (Burdick et al., 2014)" with "**Such studies found discrete neurocognitive subgroups in fully or partially remitted** patients with BD **(Bora et al., 2016; Burdick et al., 2014; Lewandowksi et al., 2014)**; a **well performing** "cognitively intact" subgroup with scores equivalent to the performance of healthy control persons (HCs), **one or two** subgroups **of** "selective cognitive impairment" **with lower cognition scores compared to HCs**, and a subgroup with "global severe impairment" across cognitive domains comparable to cognitive deficits in schizophrenia."
- Introduction, section 1, p. 3: **The findings from these studies show some discrepancies with respect to the proportion of patients in each subgroup, possibly due to differences in study samples (e.g., affective symptoms, age, medicine prescribed).**
- Introduction, section 1, p. 3: Replaced "However, as the described study (11) is the only published evidence for the presence of neurocognitive subgroups in BD, additional replication studies are needed to make any firm conclusions about the pattern of cognitive dysfunction in BD." with "Additional **data-driven subgroup** studies are needed to make any firm conclusions about the pattern of cognitive in BD."
- Introduction, section 1.1, p. 3: Replaced "The study aims to 1) determine whether there are discrete neurocognitive subgroups in fully or partially remitted patients with BD as prior demonstrated (Burdick et al., 2014). Building onto the findings of Burdick et al. (Burdick et al., 2014), the study aims to 2) examine the pattern of the cognitive deficits and the clinical and functional characteristics of the neurocognitive subgroups in terms of perceived stress, cognitive complaints, work capacity, quality of life, and medication prescribed." with "The

study aims to **identify** discrete neurocognitive subgroups in **a large group of** fully or partially remitted patients with BD **using a data-driven approach**. Building onto the **few studies applying this approach**, the **present** study aims to examine the pattern of the cognitive deficits and the clinical and functional characteristics of the neurocognitive subgroups in terms of perceived stress, cognitive complaints, work **and social adjustment**, quality of life, and medication prescribed."

- Materials and methods, section 2.5.1, p. 6: "**Although there is still some controversy regarding the particular cutoff for cognitive impairment (limits range from 1-2 *SDs* below the normal *M*; Bora et al., 2016; Martino et al., 2014), we chose in the present study to set the cut-off score of $z \geq 1$ below the *M* of the HC sample to define cognitive impairment consistent with Burdick et al. (2014) and our previous approach (Jensen et al., 2015; Ott et al., 2016).**"
- Discussion, section 4, p. 13: "**The identification of three discrete neurocognitive subgroups in BD using a data-driven approach is in line with findings previous findings (Burdick et al., 2014), although Bora et al. (2016) Lewandowski et al. (2014) found two subgroups of "selective cognitive impairment". However, Bora et al. (2016) examined executive function solely in fully remitted BD patients (HDRS-17 and YMRS scores ≤7), and Lewandowski et al. (2014) included patients with schizophrenia or schizo-affective disorder along with BD patients in their analysis, which limits comparison to the present findings.**"
- Discussion, section 4, p. 14: "**This pattern is supported by Bora et al. (2016), although the global subgroup was found to be older, have fewer years of education, longer illness duration, and more commonly given antipsychotics, suggesting that executive function may be sensitive to these factors.**"
- Discussion, section 4, p. 15: "**Indeed, Bora et al. (2016) found no statistically significant difference between subgroups with respect to previous depressive, manic, or total number of mood episodes.**"
- Discussion, section 4, p. 16: "**Indeed, this interpretation is consistent with the study by Bora et al. (2016) and meta-analyses of deficits within verbal learning and executive function but not in processing speed in individuals at genetic risk for BD (Bora et al., 2009).**"
- References, p. 20: **Bonnín, C. M., Martínez-Arán, A., Torrent, C., Pacchiarotti, I., Rosa, A. R., Franco, C., Murru, A., Sanchez-Moreno, A., Vieta, E., 2010. Clinical and**

neurocognitive predictors of functional outcome in bipolar euthymic patients: a long-term, follow-up study. J. Affect. Disord. 121, pp. 156-160.

- References, p. 20: **Bora, E., Hıdıroğlu, C., Özerdem, A., Kaçar, Ö. F., Sarısoy, G., Arslan, F.C., Aydemir, Ö., Tas, Z.C., Vahip, S., Atalay, A., Atasoy, N., Atesci, F., Tümkaya, S., 2016. Executive dysfunction and cognitive subgroups in a large sample of euthymic patients with bipolar disorder. Eur. Neuropsychopharmacol. DOI: http://dx.doi.org/10.1016/j.euroneuro.2016.04.002**
- References, p. 21: **Brissos, S., Dias, V.V., Carita, A.I., Martinez-Arán, A., 2008. Quality of life in bipolar type I disorder and schizophrenia in remission: clinical and neurocognitive correlates. Psychiatry Res. 160, pp. 55-62.**
- References, p. 23: **Lewandowski, K.E., Sperry, S.H., Cohen, B.M., Öngür, D., 2014. Cognitive variability in psychotic disorders: a cross-diagnostic cluster analysis. Psychol. Med. 44, pp. 3239-3248.**
- References, p. 23: **Martino, D.J., Strejilevich, S.A., Marengo, E., Ibañez, A., Scápola, M., Igoa, A., 2014. Toward the identification of neurocognitive subtypes in euthymic patients with bipolar disorder. J. Affect. Disord. 167, pp. 118-124.**
- References, p. 24: **Rojo, E., Pino, O., Guilera, G., Gómez-Benito, J., Purdon, S.E. , Crespo-Facorro, B. , Cuesta, M., Franco, M., Martinez-Arán, A., Segarra, N., Tabarés-Seisdedos, R., Vieta, E., Bernado, M., 2010. Neurocognitive diagnosis and cut-offscores of the Screen for Cognitive Impairment in Psychiatry (SCIP-S). Schizophr. Res. 116, pp. 243-251.**
- References, p. 25: **Torrent, C., Martinez-Aran, A., del Mar, B.C., Reinares, M., Daban, C., Sole, B., Rosa, A.R., Tabares-Seisdedos, R., Popovic, D., Salamero, M., Vieta, E., 2012. Long-term outcome of cognitive impairment in bipolar disorder. J. Clin. Psychiatry 73, pp. e899-e905.**
- References, p. 26: **Volkert, J., Kopf, J., Kazmaier, J., Glaser, F., Zierhut, K.C., Schiele, M.A., Reif, A., 2015. Evidence for cognitive subgroups in bipolar disorder and the influence of subclinical depression and sleep disturbances. Eur. Neuropsychopharmacol. 25, pp. 192-202.**

2. Regarding the study sample, it is not clear to me whether the 26 overlapping participants from studies 3 and 4 have been counted twice. Therefore, the final sample size was 193 or 167? That should be further clarified. Related with this, at the beginning of section 2.1 it could be mentioned that the four studies belong to the same team.

We agree that this is unclear (as also pointed out by Reviewer # 6) and have now deleted the footnote (Materials and methods, section 2.1) and changed the wording for use of available baseline data to increase clarity for the reader:

- Materials and methods, section 2.1, p. 4: "The present study involved pooling of **available *baseline* data from four studies of our research group** […]"

3. More relevant and concerning is the inclusion of the digits forward, which is a measure of immediate memory and auditory attention, in the working memory/executive domain. Although allocation of tests into domains is always debatable in the field of neuropsychology, there exists established consensus that digits backward tap onto working memory whereas digits forward measure attentional skills. Allocating digits forward with TMT-A in an attention/processing speed domain would be a much better option. In my opinion, the authors´ choice is a conceptual mistake, which may certainly invalidate at least in part the subsequent analyses and therefore the study conclusions.

We thank the reviewer for raising this highly relevant issue, which is also noted by reviewer # 5 and 6. We are aware that there is no complete consensus about grouping of cognitive domains in the literature. We have chosen to apply the groupings on which there is some sort of consensus by Lezak et al. (2012, for *Digit Span Forward* see p. 649). Neuropsychological tests may be categorized different due to test impurity, and as pointed by the reviewer the allocation of such tests to cognitive domains is always debatable in the field of neuropsychology. To make this clear for the reader, we added following to the manuscript:

- Materials and methods, 2.3, p. 7: "**We chose this assignment of neuropsychological tests into cognitive domains given some consistency of this domain grouping in the literature (Lezak et al., 2012) and** to be consistent with previous studies **from our group (Jensen et al., 2015; Ott et al., 2016)**."
- References, p.23: **Lezak, M.D., Howieson, D.B., Bigler, E.D., Tranel, D., 2012. Neuropsychological Assessment, 5th edition. Oxford University Press, New York.**

4. Minor points:
- I agree with truncating extreme z-scores, but please provide a rationale/reference to support the choice of z=-4.0

Based on common consensus in the literature that scores 1-2 SD below the normative mean are considered as mild cognitive impairment, score of 2-3 SD under of the mean as moderate, and below 3 SD as severe cognitive impairment (Gualtieri & Morgan, 2008; Royall et al., 2007). Thus, truncating extreme scores to *z*=-4.0 leaves room for classification of 1 SD range of "various degrees" of severe cognitive impairment. To clarify this for the reader, we added following to the manuscript:

- Materials and methods, section 2.5.1, p.6: Extreme *z*-scores (≥4 *SD*s below the HCs' *M*) were truncated to *z*=-4.0 **to allow scores with different degree of "severe" cognitive impairment (i.e., scores of >3 *SD*s below normative *M*) (Gualtieri and Morgan, 2008; Royall et al., 2007)".**
- References, p. 22: **Gualtieri, C.T., Morgan, D.W., 2008. The frequency of cognitive impairment in patients with anxiety, depression, and bipolar disorder: an unaccounted source of variance in clinical trials. J. Clin. Psychiatry 69, pp. 1122-1130.**
- References, p. 25: **Royall, D.R., Lauterbach, E.C., Kaufer, D., Malloy, P., Coburn, K.L., Black, K.J., 2007. The cognitive correlates of functional status: a review from the Committee on Research of the American Neuropsychiatric Association. J. Neuropsychiatr. Clin. Neuroci. 19, pp. 249-265.**

- Suggested reference for the last sentence in the 1st paragraph of page 12: Samamé, Martino & Strejilevich, 2014.

In response to the reviewer's request this has now been added to the manuscript:

- Discussion, p. 14: "This may be due to differences in the two patient cohorts with our cohort being somewhat younger (*M*[*SD*], 36[10] versus 42[15], respectively) and showing less depressive symptoms (HDRS-17: *M*[*SD*], 6[4] versus 11[9], respectively) **(Samamé, Martino and Strejilevich, 2014; Volkert et al., 2015)**."
- References. p. 25: "**Samamé, C., Martino, D. J., Strejilevich, S. A., 2014. Longitudinal course of cognitive deficits in bipolar disorder: a meta-analytic study. J. Affect. Disord. 164, pp. 130-138."**
- References, p. 26: "**Volkert, J., Kopf, J., Kazmaier, J., Glaser, F., Zierhut, K.C., Schiele, M.A., Reif, A., 2015. Evidence for cognitive subgroups in bipolar disorder and the influence of subclinical depression and sleep disturbances. Eur. Neuropsychopharmacol. 25, pp. 192-202."**

- Same for the last sentence in the 2nd paragraph of page 12: Dias et al, 2012.

We thank the reviewer for suggesting this relevant reference which has now been added to the manuscript:

- Discussion, p. 15: "Future prospective studies with long follow-up times are thus needed to clarify the nature and developmental trajectory of neurocognitive deficits in BD **(Dias et al., 2012)**."
- References, p. 22: "**Dias, V.V., Balanzá-Martinez, V., Soeiro-de-Souza, M.G., Moreno, R.A., Figueira, M.L., Machado-Vieira, R., Vieta, E., 2012. Pharmacological approaches in bipolar disorders and the impact on cognition: a critical overview. Acta Psychiatr. Scand. 126, pp. 315-331."**

Additional minor revisions from the authors

To increase clarity for the reader we performed the following minor revisions to the manuscript:

- Deleted (Introduction, section 1, p. 2): "better" and "positive" in "A recent meta-analysis found that **better** verbal memory and executive function were moderately related to **positive** employment outcome (Tse et al., 2014)."
- Deleted (Materials and methods, section 2.2, p. 2): "**, respectively**" in "In study 4, patients had HDRS-17 and YMRS scores of ≤7, respectively, and baseline data were obtained following 2 weeks of remission.
- Deleted (Materials and methods, section 2.1, p. 4): "**Data for the HC group came from studies 3 and 4 (*N*=110).**", because this information is given in the previous sentence.
- Corrected (Materials and methods, section 2.5.1, p. 6): " […] participants were standardized to *z*-scores (***M***=0, ***SD***=1) based on the cognitive performance of the HCs using following formula: (*test score - HC test M*) / *HC test SD (Field, 2013).*" since abbreviations for *mean* and *standard deviation* are given in the Introduction, section 1.
- Replaced (Results, section 3.2, p. 9): "of" with "in" in "Visual inspection of the dendogram provided evidence for existence of three neurocognitive subgroups **in** the complete BD sample (data not shown).
- Replaced (Results, section 3.4, p. 12): "used" with "were prescribed" in "Selectively impaired patients **were prescribed** a greater number of medications […]"
- Revised (Discussion, 4, p. 13): from "The study investigated the presence and clinical characteristics of discrete neurocognitive subgroups in the, to date, largest cohort of fully or

partially remitted BD patients" to "The study investigated the **existence** and clinical characteristics of discrete neurocognitive subgroups in **a large cohort** of fully or partially remitted BD patients."

- Deleted (Discussion, section 4, p. 13): "were" in "[…] 84% and 72% of our BD sample **were** had experienced at least three previous depressive episodes or (hypo-) manic episodes, respectively."
- Added (Discussion, section 4, p. 15): "(Burdick et al., 2015)" in "A key methodological problem in cognition trials which may partially explain the many negative findings in the field is the general lack of *objective* neuropsychological screening for cognitive deficits before inclusion of participants in these trials **(Burdick et al., 2015)"**
- Corrected: (Discussion, section 4, p. 17): Indeed, we found a larger subgroup of cognitively intact patients **(46.1%)** than previously published **by Burdick et al. (31.6%; 2014)**.

Supplementary Table.

Comparison of the characteristics of the BD samples from the four original studies 1-4.

	Pooled BD samples				**Statistics**		
	Study 1	Study 2	Study 3	Study 4	*df*	*F* or χ^2	*p*
n	46	38	84	25			
Age	34 (8)	41 (12)	36 (10)	33 (10)	3	4.48	.005
							S1 *v.* S2, *p*=.001
							S2 *v.* S3, *p*=.014
							S2 *v.* S4, *p*=.003
Gender, F/M (%)	65/35	66/34	64/36	44/56	3	4.06	.255
Years of education	16 (3)	15 (3)	15 (3)	15 (2)	3	1.05	.370
Age of onset	17 (7)	23 (10)	25 (11)	16 (6)	3	11.76	<.001
							S1 *v.* S2, *p*=.002
							S1 *v.* S3, *p*<.001
							S2 *v.* S4, *p*=.002
							S3 *v.* S4, *p*<.001
Illness duration, years	17 (8)	18 (11)	11 (10)	18 (11)	3	5.68	.001
							S1 *v.* S3, *p*=.002
							S2 *v.* S3, *p*=.001
							S3 *v.* S4, *p*=.008
HDRS-17	6 (4)	9 (4)	5 (4)	3 (2)	3	12.39	<.001
							S1 *v.* S2, *p*=.001
							S1 *v.* S4, *p*=.003
							S2 *v.* S3, *p*<.001
							S2 *v.* S4, *p*<.001
YMRS	2 (3)	3 (3)	3 (4)	1 (1)	3	2.91	.036
							S2 *v.* S4, *p*=.041
							S3 *v.* S4, *p*=.024

BD type, I/II (%)	65/34	42/58	61/39	56/44	3	5.24	.155
Depressive episodes (1/2/≥3), (%)	2/11/87	0/9/91	6/16/78	8/4/88	6	6.74	.345
(Hypo-) manic episodes (1/2/≥3), (%)	9/12/79	11/17/72	16/16/68	0/28/72	6	7.65	.265
Cognitive composite score	-1.08 (1.23)	-1.51 (1.46)	-.79 (1.59)	-.84 (1.29)	3	2.30	.079
Verbal learning and memory	-.78 (1.01)	-.81 (1.05)	-.50 (1.21)	-.74 (1.42)	3	.96	.413
Working memory and executive skills	-.58 (1.02)	-.57 (1.06)	-.30 (1.33)	-.50 (1.03)	3	.81	.489
Processing speed	-.68 (1.30)	-1.46 (1.63)	-.71 (1.58)	-.35 (1.22)	3	3.50	.017
							S1 *v.* S2, *p*=.017
							S2 *v.* S3, *p*=.010
							S2 *v.* S4, *p*=.004
Current medication							
Number of medications	2.4 (.9)	2.1 (.8)	2.2 (1.2)	1.7 (.9)	3	2.89	.037
							S1 *v.* S4, *p*=.005
							S3 *v.* S4, *p*=.029
Lithium, no. (%)	32 (70)	14 (37)	52 (63)	12 (28)	3	11.48	.009
					1	9.00	S1 *v.* S2, *p*=.003
					1	7.41	S2 *v.* S3, *p*=.006
Anticonvulsants, no. (%)	35 (76)	20 (53)	48 (59)	15 (60)	3	5.73	.125
Antidepressants, no. (%)	13 (28)	19 (50)	8 (10)	0 (0)	3	33.69	<.001
					1	4.17	S1 *v.* S2, *p*=.041
					1	7.36	S1 *v.* S3, *p*=.007
					1	8.65	S1 *v.* S4, *p*=.003
					1	24.12	S2 *v.* S3, p<.001
					1	17.90	S2 *v.* S4, *p*<.001
Antipsychotics, no. (%)	24 (52)	8 (21)	46 (56)	11 (44)	3	13.58	.004
					1	8.55	S1 *v.* S2, *p*=.003
					1	12.89	S2 *v.* S3, *p*<.001
					1	3.77	S2 *v.* S4, *p*=.049
Benzodiazepines, no. (%)	6 (13)	14 (37)	11 (13)	2 (8)	3	13.12	.004
					1	6.50	S1 *v.* S2, *p*=.011

					1	8.64	S2 *v.* S3, *p*=.003
					1	6.62	S2 *v.* S4, *p*=.010
Melatonin, no. (%)	1 (2)	2 (5)	1 (1)	1 (4)	3	1.90	.594
No medication, no. (%)	0 (0)	0 (0)	2 (2)	1 (4)	3	2.70	.441

ANOVA with least significance differences as pairwise comparison and χ^2 as appropriate. Measures of cognition are given as z-score mean (standard deviation) relative to the HCs' performance.

Abbreviations: S1 = study 1 (Miskowiak et al., 2014), S2 = study 2 (Demant et al., 2015), S3 = Study 3 (Jensen et al., 2015; Ott et al., 2016), S4 = study 4 (ongoing study), df = degress of freedom, HDRS-17 = Hamilton Depression Rating Scale 17-item, YMRS = Young Mania Rating Scale, BD = bipolar disorder