

Defining the minimal clinically important difference (MCID) of the Heinrichs–carpenter quality of life scale (QLS)

BRUNO FALISSARD,¹ CHRISTOPHE SAPIN,² JEAN-YVES LOZE,³ WALLY LANDSBERG⁴ & KARINA HANSEN⁵

1 INSERM U669, Université Paris-Sud and Université Paris-Descartes, APHP, Paris, France

2 Global Analytics, Lundbeck SAS, Issy-les-Moulineaux, France

3 Otsuka Pharmaceutical Europe Ltd, Paris, France

4 Otsuka Pharmaceutical Europe Ltd, London, UK

5 Global Health Economics and Epidemiology Lundbeck SAS, Issy-les-Moulineaux, France

Key words

Minimal Clinically Important Difference, Quality of Life Scale, schizophrenia

Correspondence

Bruno Falissard, INSERM U669, Université Paris-Sud and Université Paris-Descartes, APHP, Paris, France.
Telephone (+33) (0)681827076
Email: bruno.falissard@gmail.com

Received 22 September 2014;

revised 12 February 2015;

accepted 23 April 2015

Abstract

To determine the Minimal Clinically Important Difference (MCID) of the Heinrichs–Carpenter Quality of Life Scale (QLS). Data from the “Schizophrenia Trial of Aripiprazole” (STAR) study were used in this analysis. The MCID value of the QLS total score was estimated using the anchor-based method. These findings were substantiated/validated by comparing the MCID estimate to other measurements collected in the study. Half of the patients (49%) showed improvement in Clinical Global Impressions of Severity (CGI-S) during the trial. The estimated MCID of the QLS total score was 5.30 (standard error: 2.60; 95% confidence interval: [0.16; 10.43]; $p < 0.05$). Patients were divided into two groups: “QLS improvers” (QLS total score increased \geq six points) and “non-improvers”. The QLS improvers had significantly better effectiveness and reported significantly higher levels of preference for their current medications. There was a statistically significant difference between the two groups in the change in two of the four domains of QLS; “Interpersonal relations” and “Intrapsychic foundations” domains during the study. These findings support the value of the estimated MCID for the QLS and may be a useful tool in evaluating antipsychotic treatment effects and improving long-term patient outcomes in schizophrenia. Copyright © 2015 John Wiley & Sons, Ltd.

Introduction

Schizophrenia is often a severe and persistent mental illness, typically accompanied by functional impairment and disability, characterized by poor psychosocial

functioning, difficulties in activities of daily living and interpersonal relationships, low levels of productivity, and high rates of unemployment (Ascher-Svanum *et al.*, 2013, Lenroot *et al.*, 2003, Awad and Voruganti 2008).

Antipsychotic medications are proven effective in treating acute psychosis and reducing the risk of future psychotic episodes. In clinical practice and in randomized controlled trials (RCTs), antipsychotic efficacy is routinely measured according to symptom rating scales e.g. Positive and Negative Syndrome Scale (PANSS).

Since symptoms are medically centred and not necessarily patient oriented, the evolution of quality of life (QoL) score is not always paralleled with evolution of clinical symptoms. This has for example been found in patients with schizophrenia (Wilson-d'Almeida *et al.*, 2013). Meaningful improvement of QoL – interpersonal relations, role performance, and community living skills – may substantially lag behind symptomatic improvement. Both must be considered and their assessment should complement one another. Furthermore, QoL has been shown to be an independent prognostic factor associated with clinical outcome in various chronic diseases, often predicting hospitalization, relapse or survival (Parshall *et al.*, 2008, Sprenkle *et al.*, 2004, Yeo *et al.*, 2006). In the field of psychiatric research QoL has gained increasing acceptance. Boyer *et al.* (2013) demonstrated that QoL at baseline, as assessed by the Short-form 36 (SF-36), is an independent predictor of relapse at a 24-month follow-up in schizophrenia. This was in contrast to symptom severity, measured by PANSS, which failed to significantly predict relapse.

The Heinrichs–Carpenter Quality of Life Scale (QLS) is a clinician-rated scale designed to assess deficit symptoms of schizophrenia and functioning during the preceding four weeks. The QLS comprises 21 items in four subscales: interpersonal relations (household, friends, acquaintances, social activity, social network, social initiative, withdrawal, and sociosexual behaviour), instrumental role (occupational role, work functioning, work level, and work satisfaction) intrapsychic foundations (sense of purpose, motivation, curiosity, anhedonia, aimless inactivity, empathy, and emotional interaction), and commonplace objects and activities. Total and subscale scores are computed as the sum of contributing items, with higher scores indicating better QoL and functioning. The QLS has good inter-rater reliability (Heinrichs *et al.*, 1984), convergent validity (Lehman *et al.*, 1993), and criterion-related validity (Ascher-Svanum *et al.*, 2012).

The QLS is used widely in psychopharmacologic treatment trials for schizophrenia, and has been used in a naturalistic setting trial in order to measure effectiveness e.g. as a primary endpoint in the CUtLASS 1 (Cost-Utility of the Latest Antipsychotic Drugs in Schizophrenia Study; Jones *et al.*, 2006), and more commonly as a secondary endpoint e.g. the CATIE study (Clinical Antipsychotic

Trials of Intervention Effectiveness; Swartz *et al.*, 2007; Philipps *et al.*, 2006; Witte *et al.*, 2012). However, information regarding the interpretation of changes in QLS total and domains scores is limited, despite it being a key element used for evaluation purposes comparing therapeutic options e.g. in RCTs (Fayers and Machin, 2000).

The term Minimal Clinically Important Difference (MCID) was first described by Jaeschke *et al.* (1989, p. 408); their operational definition of a MCID was "... The smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management." Early work on MCID was from studies of patient-reported outcomes, in which it represented the smallest improvement that was considered meaningful by the patient (Brozek *et al.*, 2006; Copay *et al.*, 2007).

Establishing superior efficacy/effectiveness relies on the demonstration of significant differences in mean score changes in rating scales between comparator groups. However, interpreting the clinical relevance and potential economic impact of this primary endpoint is not without difficulties. Observed statistically significant changes between two interventions, does not necessarily imply that this difference is clinically important or that changes within patients were clinically relevant. When a primary outcome measure is chosen for the purpose of a clinical trial, it is absolutely key to power the study based on a robust MCID assumption. If this is not done, studies might be powered to detect statistical significant differences that are NOT clinically meaningful, as noted by McGlothlin and Lewis (2014). In disease areas like neurology and psychiatry where symptomatology is evaluated through rating scales and "soft" endpoints, it becomes even more crucial to know by how much a change can be considered as clinically meaningful. The MCID is important when interpreting results of a clinical trial – thus making the results become more useful for treating physicians.

There is a need to more clearly elucidate the clinical relevance of study findings. One such measure is the effect size, often standardized effect size, representing the magnitude of the improvement from baseline to endpoint of the study drug compared with placebo.

The fundamental disadvantage of summarizing a treatment effect as a mean change in score is that modest changes of clinical relevance in individual patients may be lost because they are interpreted as random variation within the error of measurement. Furthermore, it is difficult to estimate the cost effectiveness of outcomes measured according to the mean change from baseline score. As a result, the concept of the MCID is gathering interest

for the comparison of treatment interventions (Burbach *et al.*, 1999; Meyer, 2005).

The objective of this analysis was to determine the MCID of the Heinrichs–Carpenter QLS (Heinrichs *et al.*, 1984). A further objective was to externally validate the MCID threshold value identified in the analysis.

Method

Data source and assessments

In order to estimate the MCID for the QLS total score, data from the “Schizophrenia Trial of Aripiprazole” (STAR) study were used. The design and methodology of the STAR study have been published in detail elsewhere (Kerwin *et al.*, 2007). This study was a multi-centre, randomized, naturalistic, open-label study comparing aripiprazole treatment with standard of care (SOC) treatment for a period of 26 weeks. Patients randomized to the SOC treatment group received one of three selected atypical antipsychotics: olanzapine, quetiapine, or risperidone, based on the investigator’s judgment and patients’ previous response to antipsychotic medication. During the 26-week open-label treatment phase, study visits occurred at weeks 2, 4, 8, 12, 18, and 26 to assess the effectiveness of study treatment.

The QLS was one of the secondary outcomes of the STAR study together with the Clinical Global Impressions of Severity (CGI-S) and Global Improvement (CGI-I) (Guy, 1976), the Preference of Medication (POM) scale assessing treatment satisfaction (Tandon *et al.*, 2006), the Impact of Weight on Quality of Life (IWQoL) scale (Kolotkin *et al.*, 2001), the EuroQol 5 Dimensions (EQ-5D) (EuroQol Group, 1990), healthcare resource use questionnaire (Client Socio-demographic and Service Receipt Inventory, CSSRI) (Chisholm *et al.*, 2000) and the Arizona Sexual Experience Scale (ASEX) (McGahuey *et al.*, 2000). The primary outcome measure was the Investigator Assessment Questionnaire (IAQ), a relative effectiveness measure (Tandon *et al.*, 2005).

Statistical analysis

The analysis consisted of two-steps: first the MCID value of the QLS total score was estimated; then, the findings were substantiated/validated by comparing the MCID estimate to other measurements previously described and collected in the STAR study.

The analysis population was defined as all patients with at least one post-baseline assessment of both QLS and the anchor criterion.

MCID estimation

Although no method is as yet considered the “gold standard” for determining the MCID, the most commonly used are the distribution-based method and the anchor-based method.

The distribution-based method uses the estimation of the standard error of measurement (SEM) of the scale of interest (Wyrwich *et al.*, 1999). The SEM represents the systematic error due to the measurement itself, and previous reports have shown that values between “1 SEM” and “1.96 SEM” are good estimates of the MCID (Wyrwich, 2004; Rejas *et al.*, 2008). The SEM is usually estimated using the following formula:

$$SEM = \sigma_{BL} \cdot \sqrt{1 - r}$$

where σ_{BL} is the baseline value of the standard deviation and r represents the scale reliability. This parameter is usually estimated by the Cronbach’s α coefficient. This method is straightforward to implement, using only data from the scale of interest; the main limitation being the statistical-oriented way of defining a MCID without taking any clinical aspects into consideration.

The anchor-based method allows the computation of the MCID by comparing mean scores of the scale of interest across groups of patients that are known to have small but real differences in their health status (Juniper *et al.*, 1994). Usually, CGI-S or CGI-I are used as criterion (or *anchor*) to define groups of patients with different health status profiles. Then, the difference in scores between two concomitant levels of the anchor (e.g. between CGI-I = 3 and CGI-I = 4) should represent an estimate of the MCID of the score of interest.

One of the main advantages of this method is the use of clinical data in the determination of the MCID (Turner *et al.*, 2010); however, results should be cautiously monitored when this method is applied on placebo-controlled randomized clinical trials in terms of distribution of treated and placebo-patients per anchor level.

Here the primary analysis for the MCID determination was the anchor-based method, using CGI-S as the anchor. The CGI-S was chosen to avoid any memory bias in the CGI-I evaluations as the STAR study was 26 weeks in duration. An analysis of covariance (ANCOVA) model was used, modelling the mean change from baseline to endpoint of the QLS total score, with the difference in CGI-S used as factor, and baseline value of QLS as covariate.

The following sensitivity analyses were conducted to evaluate robustness of the primary analysis: (1) the anchor-based method was applied on other concomitant

health profiles based on difference in CGI-S and (2) the distribution-based approach was applied to the analysis population.

MCID validation phase

In order to substantiate the MCID estimate, two groups of patients were defined according to their QLS total score change from baseline: the patients with a change from baseline greater or equal to the estimated MCID were defined as the “QLS improvers” group, the remaining part of the analysis population, the “QLS non-improvers”.

These two groups were then compared at baseline and endpoint to test whether: (A) any differences at baseline could be identified between these two groups in terms of demographics, medical history or baseline clinical assessments and (B) the clinically relevant changes in QLS total score were also captured through the identification of differences on specific ratings scales, either clinician-, patient- or caregiver-reported.

Baseline characteristics of these groups are described and compared using parametric (*t*-test or Chi-square test) or non-parametric procedures (Mann–Whitney or Fisher’s exact tests) if relevant.

For quantitative measurements (CGI-S, CGI-I, IAQ, POM, IWQoL, EQ-5D), an ANCOVA model was used to estimate mean change from baseline scores per group. Explanatory variables were group (“QLS improvers” versus “QLS non-improvers”) and baseline value of the modelled measure. The alpha risk was set to 0.05, and all analyses were conducted using SAS version 9.2.

Results

Patient disposition and demographics

The patient disposition and demographics of the overall STAR study have been published in detail elsewhere (Kerwin *et al.*, 2007).

The overall STAR study population and group of patients included in this analysis were comparable with respect to the baseline sociodemographic and clinical characteristics (Table 1).

Clinical global impression of severity (CGI-S)

Almost half of the patients (49%) showed improvement in CGI-S during the trial (Figure 1A); one-fifth of the patients (19%) were “improved” with the remainder (29%) “slightly improved” (Figure 1B). Just over one-third (37%) showed no change in CGI-S and 15% had a worsening in CGI-S (Figure 1A).

Estimation of MCID for QLS

Comparing the “slightly improved” group to the “no change” group, the MCID of the QLS total score was estimated to be 5.30 (standard error: 2.60; 95% confidence interval: [0.16; 10.43]; $p < 0.05$) (Table 2).

Sensitivity analyses

Using the anchor-based approach on the subset of patients (“slightly improved” group and “no change” group), the difference of these two subgroups of patients revealed a MCID estimate of 5.30 (standard error: 2.56; 95% confidence interval: [0.23; 10.37], $p < 0.05$) (Table 3).

As sensitivity analyses, the same methodology was applied comparing the difference between the “improved” and “slightly improved” subgroups (Table 4). The MCID of the QLS total score was estimated to be 4.29 (standard error: 2.59; 95% confidence interval: [−0.83; 9.42]; $p = 0.100$).

The results of the distribution-based approach on the overall population revealed a baseline standard deviation (SD) of 14.011 and a Cronbach’s α of 0.942, leading to a SEM of 3.37. Applying the “1 SEM” to “1.96 SEM” rule, gives an estimate for the MCID of the QLS total score ranging between 3.37 and 6.61. As this method is known to provide smaller estimates than the anchor-based method, these results confirm that of the primary analysis.

MCID validation

Patient disposition

Patients were divided into two groups: “QLS improvers” (QLS total score increased \geq six points) and the remaining patients (“QLS non-improvers”). The baseline characteristics of these two groups of patients are presented in Table 5. The CGI-S at baseline was statistically significantly different between the two groups, the “QLS improvers” group being less severe (4.0 versus 4.2). Other demographic and clinical characteristics at baseline were similar between the two groups.

Evaluation at study endpoint

The comparisons of the mean changes from baseline to study endpoint of the clinical-, patient- and caregiver-reported outcomes are presented in Table 6.

There was a statistically significant difference between the two groups for CGI-S. This difference was supported by statistically significantly different ratings for CGI-I, with larger clinical improvements observed in ‘QLS improvers’.

Table 1. Sociodemographic and clinical characteristics at baseline in the overall STAR (Schizophrenia Trial of Aripiprazole) study population and group of patients included in this analysis

| | Analysis population (<i>n</i> = 351) ¹ | Randomized population (<i>n</i> = 555) ² |
|---------------------------------------|--|--|
| Gender [<i>n</i> (%) of male] | 220 (62.7%) | 332 (59.8%) |
| Age (years) | | |
| Mean ± standard deviation (SD) | 38.6 ± 11.4 | 38.5 ± 10.9 |
| Median | 37.0 | 37.0 |
| Baseline weight (kg) | | |
| Mean ± SD | 80.2 ± 17.0 | 80.7 ± 17.2 |
| Median | 78.2 | 79.0 |
| Schizophrenia subtype [<i>n</i> (%)] | | |
| Disorganized | 41 (11.7%) | 56 (10.1%) |
| Catatonic | 6 (1.7%) | 7 (1.3%) |
| Paranoid | 222 (63.3%) | 361 (65.0%) |
| Residual | 44 (12.5%) | 67 (12.1%) |
| Undifferentiated | 38 (10.8%) | 64 (11.5%) |
| Allocated treatment [<i>n</i> (%)] | | |
| Aripiprazole | 179 (51.0%) | 284 (51.2%) |
| Standard of care | 172 (49.0%) | 271 (48.8%) |

¹The analysis population was defined as all patients with at least one post-baseline assessment of both Quality of Life Scale (QLS) and the Clinical Global Impressions of Severity (CGI-S) (anchor criterion).

²Adapted from Kerwin *et al.* (2007).

These results were substantiated by those observed on IAQ and POM: “QLS improvers” had significantly better effectiveness (through IAQ total score) and reported significantly higher levels of preference for their current medications (through patient- and caregiver-reported POM assessments).

For the health-related QoL assessments there was a statistically significant difference between the two groups in the change in QLS total score, as well as the “Interpersonal relations” and “Intrapsychic foundations” domains. There were no differences between the two groups for the IWQoL scores.

Discussion

Measures of QoL are increasingly used in clinical trials to enhance the delivery of patient-centred care. Suitable estimates of MCID for such measures are important to inform trial design and interpretation of results. To our knowledge, this is the first attempt at estimating the MCID of the QLS, one of the most widely used QoL questionnaires, for patients with schizophrenia, using both anchored- and distribution-based methodologies.

Although commonly used in RCTs in schizophrenia, increasingly, as a primary outcome measure in broad effectiveness trials, little is published on the interpretation of

the QLS scores. In a recent publication, some normative values around QLS total score were defined (Ascher-Svanum *et al.*, 2013). Bushnell *et al.* (2000) found that a six-point in QLS total score was associated with a 20% improvement in clinical symptomatology as assessed by the Brief Psychiatric Rating Scale. Cramer *et al.* (2001) published some elements to interpret the QLS total score changes on a relative scale, a 26% improvement in QLS score being linked to a “improved” clinical status. After six months of treatment, Dunayevitch *et al.* (2006) reported QLS total score changes from baseline of about four points in a subgroup of patients whose health status was not improved. Our estimates of 5.3 points in QLS total score is in line with these findings.

Although of major importance for evaluative purpose, there are no formal gold standard methods to evaluate the MCID of rating scales (Fayers and Machin, 2000). The two most commonly used methodologies, distribution- and anchored-based methods, have been applied leading to quite consistent results that strengthen the overall finding. In the context of randomized placebo-controlled trials, large differences on primary and secondary endpoints are expected on a carefully selected population. Therefore, applying anchored-based methods to estimate the MCID of a scale from this type of study may yield overestimated results; any MCID based

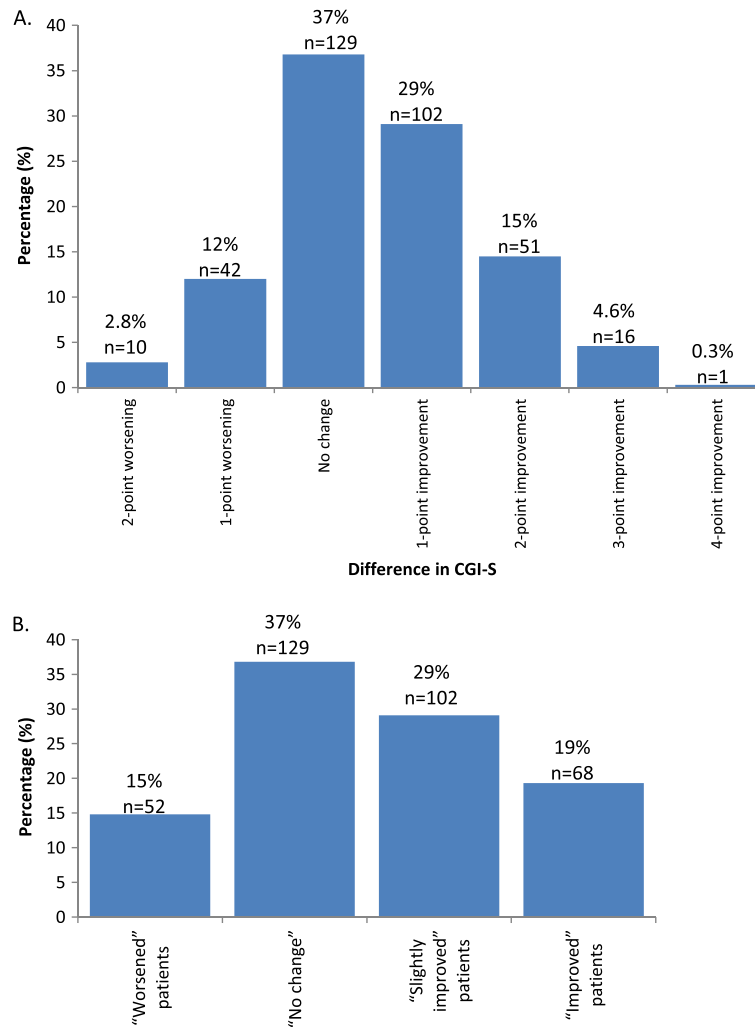


Figure 1. Changes in CGI-S from baseline to end of study (A) and patient subgroups based on CGI-S evolution during the study (B).

Table 2. Least square mean Quality of Life Scale (QLS) total score change from baseline per level of anchor (difference in CGI-S)

| Difference in CGI-S | QLS total score | | | 95% Confidence interval | |
|--|-----------------|----------------|---------|-------------------------|-------|
| | Mean | Standard error | p-Value | Lower | Upper |
| Difference of -1 (n = 102) "slightly improved" | 8.28 | 1.83 | <0.001 | 4.66 | 11.90 |
| Difference of 0 (n = 129) "no change" | 2.98 | 1.85 | 0.109 | -0.67 | 6.63 |
| MCID estimate "slightly improved" minus "no change" | 5.30 | 2.60 | 0.043 | 0.16 | 10.43 |

Table 3. Results of the sensitivity analyses (anchor-based on a subset of patients)

| Difference in CGI-S | QLS total score | | | 95% Confidence interval | |
|--|-----------------|----------------|---------|-------------------------|-------|
| | Mean | Standard error | p-Value | Lower | Upper |
| Difference of -1 (n = 102) "slightly improved" | 8.28 | 1.80 | <0.001 | 4.71 | 11.85 |
| Difference of 0 (n = 129) "no change" | 2.98 | 1.82 | 0.104 | -0.62 | 6.58 |
| MCID estimate "slightly improved" minus "no change" | 5.30 | 2.56 | 0.041 | 0.23 | 10.37 |

Table 4. Results of the sensitivity analyses (anchored-based approach per level of anchor)

| Difference in CGI-S | QLS total score | | | 95% Confidence interval | |
|--|-----------------|----------------|---------|-------------------------|-------|
| | Mean | Standard error | p-Value | Lower | Upper |
| "improved" (n = 68) | 12.57 | 1.89 | <0.001 | 8.81 | 16.32 |
| "slightly improved" (n = 102) | 8.28 | 1.76 | <0.001 | 4.78 | 11.77 |
| MCID estimate "slightly improved" minus "no change" | 4.29 | 2.59 | 0.100 | -0.83 | 9.42 |

Table 5. Patients' baseline characteristics per group

| | "QLS improvers" (n = 126) ¹ | "QLS non-improvers" (n = 225) | p-Value |
|--------------------------|--|-------------------------------|---------|
| Gender (% male) | 66.7% | 60.4% | 0.248 |
| Age | 38.4 ± 11.3 | 38.7 ± 11.5 | 0.828 |
| Body mass index | 27.8 ± 4.8 | 26.8 ± 5.3 | 0.083 |
| Duration of illness | 10.3 ± 9.0 | 10.8 ± 10.4 | 0.654 |
| CGI-S | 4.0 ± 0.8 | 4.2 ± 0.9 | 0.049 |
| EQ-5D index | 0.70 ± 0.26 | 0.71 ± 0.25 | 0.736 |
| IWQoL | | | |
| Public distress | 96.0 ± 10.8 | 94.6 ± 11.3 | 0.264 |
| Physical functioning | 85.9 ± 17.1 | 84.7 ± 20.6 | 0.562 |
| Self-esteem | 81.8 ± 24.2 | 82.5 ± 24.9 | 0.803 |
| Sexual life | 90.2 ± 21.7 | 88.7 ± 24.8 | 0.584 |
| Work | 90.0 ± 18.0 | 91.3 ± 16.8 | 0.496 |
| Total score | 87.5 ± 15.7 | 86.9 ± 17.4 | 0.742 |
| QLS | | | |
| Interpersonal relations | 22.7 ± 8.5 | 21.5 ± 9.3 | 0.257 |
| Instrumental role | 9.9 ± 6.1 | 10.4 ± 5.5 | 0.449 |
| Intrapsychic foundation | 22.1 ± 7.2 | 21.8 ± 7.3 | 0.740 |
| Common object/activities | 7.1 ± 2.5 | 7.1 ± 2.6 | 0.791 |
| Total score | 61.8 ± 20.8 | 60.9 ± 21.2 | 0.713 |

¹QLS total score increased ≥ six points.

Table 6. Comparison of rating scales at study endpoint per group of QLS improvement

| | “QLS improvers” (n = 126) ¹ | “QLS non-improvers” (n = 225) | p-Value |
|-----------------------------------|--|-------------------------------|---------|
| CGI-S | 3.3 ± 1.4 | 3.8 ± 1.2 | <0.001 |
| CGI-I | 2.9 ± 1.6 | 3.5 ± 1.5 | 0.002 |
| IAQ total score | 24.7 ± 6.0 | 26.9 ± 5.4 | <0.001 |
| EQ-5D index | 0.77 ± 0.22 | 0.74 ± 0.24 | 0.167 |
| POM | | | |
| Patient evaluation | 2.0 ± 1.2 | 2.5 ± 1.4 | 0.002 |
| Caregiver evaluation ² | 1.9 ± 0.9 | 2.6 ± 1.4 | 0.007 |
| IWQoL | | | |
| Public distress | 95.0 ± 11.4 | 94.9 ± 11.9 | 0.963 |
| Physical functioning | 87.5 ± 18.2 | 87.3 ± 18.4 | 0.927 |
| Self-esteem | 86.7 ± 21.5 | 85.3 ± 23.0 | 0.599 |
| Sexual life | 89.8 ± 23.8 | 88.7 ± 23.9 | 0.704 |
| Work | 90.4 ± 18.6 | 92.1 ± 16.0 | 0.388 |
| Total score | 89.2 ± 16.1 | 88.8 ± 15.9 | 0.838 |
| QLS | | | |
| Interpersonal relations | 26.9 ± 9.0 | 24.0 ± 9.7 | 0.009 |
| Instrumental role | 11.5 ± 6.7 | 11.4 ± 5.6 | 0.922 |
| Intrapsychic foundation | 25.7 ± 7.6 | 23.7 ± 7.9 | 0.029 |
| Common object/activities | 7.5 ± 2.4 | 7.7 ± 2.5 | 0.539 |
| Total score | 72.7 ± 23.1 | 66.0 ± 22.4 | 0.021 |

¹QLS total score increased \geq six points.

²Only a few patients had caregiver evaluation: $n = 27$ and $n = 66$ for the “QLS improvers” and “QLS non-improvers” groups, respectively.

Note: CGI-S, Clinical Global Impressions of Severity; CGI-I, Clinical Global Impressions of Improvement; IAQ, Investigator Assessment Questionnaire; EQ-5D, EuroQoL 5 Dimensions; POM, Preference of Medication questionnaire; IWQoL, Impact of Weight on Quality of Life scale; CSSRI, Client Socio-demographic and Service Receipt Inventory; QLS, Quality of Life Scale.

on distribution-based method would yield underestimated values due to the homogeneity of the selected patient population. In contrast, data from pragmatic trials like STAR can be considered as more appropriate. Indeed, this type of data may ensure that the results are not impacted or driven by any experimental considerations such as therapeutic interventions or study design aspects like schedule of assessment or eligibility criteria. As a consequence, we can expect the MCID estimates from anchored- and distribution-based methods to be more reliable.

MCID is very challenging because the notion of “clinically important” is so subjective and personal. It can be difficult to measure clinical importance for objective endpoints e.g. extending a person’s life by a defined length of time (month/week/day). Further, for subjective endpoints like QoL the task is even more difficult, particularly in disease area of psychiatry where symptomatology is evaluated through rating scales instead of hard endpoints. Nevertheless, when totally different types of estimates (anchor

based or distribution based) converge to the same value, the MCID becomes really tangible. The quality of the anchor here (CGI-S, which is a very well established rating tool used by the practicing clinician) makes it even more concrete. MCID has been used to measure the critical threshold needed to achieve clinically relevant treatment effectiveness pain (Singh *et al.*, 2014; Lauridsen *et al.*, 2009; Ostelo and deVet, 2005). Even with different MCID methods, the results are clinically appropriate and consistent with expectations thus facilitating the presentation and interpretation of results obtained in clinical trials and the transposition of trial results into practice.

When there is no MCID available, the alternative is to dichotomize the score and to estimate the number needed to treat (NNT) (the inverse of the percentage reduction in severity). But in such a situation, there is an important loss of power. Thus MCID determination helps to keep the dimensional score which guarantees the optimal statistical power.

The results in this report must be considered in the context of several limitations.

First, the concept of MCID is subjective and thus difficult to grasp in practice. All experimental approaches thus have obvious limitations. Here, the robustness of the estimates whether from the anchor-based approach or the perspective of error measurement makes the reader more confident in our MCID estimate of 5.3 points in QLS total score. Second, it is difficult to be certain of the generalizability of these results as there is no data on inter-rater reliability. Third, the sample size was too small to assess with sufficient statistical power if the MCID is constant across QoL scores and if the relationship between CGI and QoL scores was really linear.

Since CGI-S is a rater reported outcome, the MCID also represents the smallest meaningful clinical difference that can be detected by the rater. According to the original definition, the MCID should be estimated from the patients' perspective: since the meaningfulness of the difference is important for the patient not the physician.

In conclusion, the concept of the MCID is gaining recognition as a potential metric to describe the effectiveness of current treatments and address the limitations of mean changes in clinical rating scale scores. MCID threshold values for antipsychotics treatments are particularly important in view of the existing high morbidity and mortality in the disorder. However, MCID values are lacking for

commonly used schizophrenia rating scales. Our findings indicate that the MCID for QLS is 5.3 points. Although the MCID value was validated by comparing the MCID estimate to other measurements collected in the study, the results warrant further replication. These initial findings may ultimately be a useful tool in evaluating the efficiency of antipsychotic treatment effects and improving long-term patient outcomes.

Acknowledgements

J.K. Simonsen (employee of H. Lundbeck A/S) provided support in the preparation, revision, and editing of the manuscript.

This research was supported by H. Lundbeck A/S and Otsuka Pharmaceutical Development & Commercialization.

Declaration of interest statement

C. Sapin and K. Hansen are employees of H. Lundbeck A/S. J.Y. Loze and W. Landsberg are employees of Otsuka Pharmaceutical Europe Ltd. B. Falissard has been consultant, expert or has given talks for: E. Lilly, BMS, Servier, SANOFI, GSK, HRA, Roche, Boeringer-Ingelheim, Bayer, Almirall, Allergan, Stallergene, Genzyme, Pierre Fabre, AstraZeneca, Novartis, Janssen, Astellas, Biotronik, Daiichi-Sankyo, Gilead, MSD, Lundbeck.

References

- Ascher-Svanum H., Montgomery W.S., McDonnell D.P., Coleman K.A., Feldman P.D. (2012). Treatment-completion rates with olanzapine long-acting injection versus risperidone long-acting injection in a 12-month, open label treatment of schizophrenia: indirect, exploratory comparisons. *International Journal of General Medicine*, **5**, 391–398, DOI: 10.2147/IJGM.S29052
- Ascher-Svanum H., Novick D., Haro J.M., Aguado J., Cui Z. (2013) Empirically driven definitions of 'good,' 'moderate,' and 'poor' levels of functioning in the treatment of schizophrenia. *Quality of Life Research*, **22**, 2085–2094, DOI: 10.1007/s11136-012-0335-z
- Awad A.G., Voruganti L.N. (2008) The burden of schizophrenia on caregivers: a review. *Pharmacoeconomics*, **26**, 149–162.
- Boyer L., Millier A., Perthame E., Aballea S., Auquier P., Toumi M. (2013) Quality of life is predictive of relapse in schizophrenia. *BMC Psychiatry*, **13**, 15, DOI: 10.1186/1471-244X-13-15
- Brozek J.L., Guyatt G.H., Schunemann H.J. (2006) How a well-grounded minimal important difference can enhance transparency of labelling claims and improve interpretation of a patient reported outcome measure. *Health and Quality of Life Outcomes*, **27**, 69, DOI: 10.1186/1477-7525-4-69
- Burback D., Molnar F.J., St John P., Man-Son-Hing M. (1999) Key methodological features of randomized controlled trials of Alzheimer's disease therapy. Minimal clinically important difference, sample size and trial duration. *Dementia and Geriatric Cognitive Disorders*, **10**, 534–540, DOI:10.1159/000017201
- Bushnell D.M., Patrick D.L., Martin M.L., Kody M. C., Buesching D.P., Breier A. (2000) The Quality of Life Scale (QLS) for schizophrenia: assessment of responsiveness to clinical change. *Quality of Life Research*, **9**, 336.
- Chisholm D., Knapp M.R., Knudsen H.C., Amadeo F., Gaité L., van Wijngaarden B. (2000) Client Socio-demographic and Service Receipt Inventory – European version: development of an instrument for international research. EPSILON Study 5. European psychiatric services: inputs linked to outcome domains and needs. *British Journal of Psychiatry*, **177**, s28–s33.
- Copay A.G., Subach B.R., Glassman S.D., Polly D. W. Jr, Schuler T.C. (2007) Understanding the minimum clinically important difference: a review of concepts and methods. *Spine Journal*, **7**, 541–546, DOI: 10.1016/j.spinee.2007.01.008
- Cramer J., Rosenheck R., Xu W., Henderson W., Thomas J., Charney D., Department of Veterans Affairs Cooperative Study Group on Clozapine in Refractory Schizophrenia (2001) Detecting improvement in quality of life and symptomatology in schizophrenia. *Schizophrenia Bulletin*, **27**, 227–324.
- Dunayevitch E., Sethuraman G., Enerson M., Taylor C.C., Lin D. (2006) Characteristics of two alternative schizophrenia remission definitions: relationship to clinical and quality of life outcomes. *Schizophrenia Research*, **86**, 300–308, DOI: 10.1016/j.schres.2006.06.002

- EuroQol Group. (1990) EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy*, **16**, 199–208.
- Fayers P.M., Machin D. (2000) *Quality of Life: Assessment, Analysis and Interpretation*, Chichester, John Wiley & Sons, DOI: 10.1002/sim.941
- Guy W. (1976) *ECDEU Assessment Manual for Psychopharmacology Revised*, Rockville, MD, US Department of Health Services.
- Heinrichs D.W., Hanlon T.E., Carpenter W.T. (1984) The Quality of Life Scale: an instrument for rating the schizophrenic deficit syndrome. *Schizophrenia Bulletin*, **10**, 388–398, DOI: 10.1093/schbul/10.3.388
- Jaeschke R., Singer J., Guyatt G.H. (1989) Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clinical Trials*, **10**, 407–415.
- Jones P.B., Barnes T.R., Davies L., Dunn G., Lloyd H., Hayhurst K.P., Murray R.M., Markwick A., Lewis S.W. (2006) Randomized controlled trial of the effect on Quality of Life of second- vs first-generation antipsychotic drugs in schizophrenia: Cost-Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUTLASS 1). *Archives of General Psychiatry*, **63**, 1079–1087, DOI:10.1001/archpsyc.63.10.1079
- Juniper E.F., Guyatt G.H., Willan A., Griffith L.E. (1994) Determining a minimal important change in a disease-specific quality of life questionnaire. *Journal of Clinical Epidemiology*, **47**, 81–87, DOI: 10.1016/0895-4356(94)90036-1
- Kerwin R., Millet B., Herman E., Banki C.M., Lublin H., Pans M., Hanssens L., L'Italien G., McQuade R.D., Beuzen J.N. (2007) A multicentre, randomized, naturalistic, open-label study between aripiprazole and standard of care in the management of community-treated schizophrenic patients Schizophrenia Trial of Aripiprazole: (STAR) study. *European Psychiatry*, **22**, 433–443, DOI: 10.1016/j.eurpsy.2007.03.002
- Kolotkin R.L., Crosby R.D., Kosloski K.D., Williams G.R. (2001) Development of a brief measure to assess quality of life in obesity. *Obesity Research*, **9**, 102–111.
- Lauridsen H.H., Manniche C., Korsholm L., Grunnet-Nilsson N., Hartvigsen J. (2009) What is an acceptable outcome of treatment before it begins? Methodological considerations and implications for patients with chronic low back pain. *European Spine Journal*, **18**, 1858–1866, DOI: 10.1007/s00586-009-1070-1
- Lehman A.F., Postrado L.T., Rachuba L.T. (1993) Convergent validation of quality of life assessments for persons with severe mental illnesses. *Quality of Life Research*, **2**, 327–333.
- Lenroot R., Bustillo J.R., Lauriello J., Keith S.J. (2003) Integrated treatment of schizophrenia. *Psychiatric Services*, **54**, 1499–1507.
- McGahuey C.A., Gelenberg A.J., Laukes C.A., Moreno F.A., Delgado P.L., McKnight K.M., Manber R. (2000) The Arizona Sexual Experience Scale (ASEX): reliability and validity. *Journal of Sex and Marital Therapy*, **26**, 25–40.
- McGlothlin A.E., Lewis R.J. (2014) Minimal clinically important difference: defining what really matters to patients. *JAMA: The Journal of the American Medical Association*, **312**, 1342–1343, DOI: 10.1001/jama.2014.13128
- Meyer R.J. (2005) US regulatory perspective on the minimal clinically important difference in chronic obstructive pulmonary disease. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, **2**, 47–49.
- Ostelo R.W.J.G., deVet H.C.W. (2005) Clinically important outcomes in low back pain. *Best Practice & Research Clinical Rheumatology*, **19**, 593–607.
- Parshall M.B., Mapel D.W., Rice L., Williams A., O'Reilly J. (2008) Predictive validity of short-form health survey [36 items] scales for chronic obstructive pulmonary disease exacerbation. *Heart & Lung*, **37**, 356–365, DOI: 10.1016/j.hrtlung.2007.09.007
- Phillips G.A., Van Brunt D.L., Roychowdhury S.M., Xu W., Naber D. (2006) The relationship between quality of life and clinical efficacy from a randomised trial comparing olanzapine and ziprasidone. *Journal of Clinical Psychiatry*, **67**, 1397–1403.
- Rejas J., Pardo A., Ruiz M.A. (2008) Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patient reported outcomes measures. *Journal of Clinical Epidemiology*, **61**, 350–356, DOI: 10.1016/j.jclinepi.2007.05.011
- Singh J.A., Luo R., Landon G.C., Suarez-Almazor M. (2014) Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. *Journal of Rheumatology*, **41**, 509–515, DOI: 10.3899/jrheum.130609
- Sprengle M.D., Niewoehner D.E., Nelson D.B., Nichol K.L. (2004) The Veterans Short Form 36 questionnaire is predictive of mortality and health-care utilization in a population of veterans with a self-reported diagnosis of asthma or COPD. *Chest*, **126**, 81–89, DOI: 10.1378/chest.126.1.81
- Swartz M.S., Perkins D.O., Stroup T.S., Davis S.M., Capuano G., Rosenheck R.A., Reimherr F., McGee M.F., Keefe R.S., McEvoy J.P., Hsiao J.K., Lieberman J.A., CATIE Investigators. (2007) Effects of antipsychotic medications on psychosocial functioning in patients with chronic schizophrenia: findings from the NIMH CATIE study. *American Journal of Psychiatry*, **164**, 428–436, DOI: 10.1176/appi.ajp.164.3.428
- Tandon R., Devellis R.F., Han J., Li H., Frangou S., Dursun S., Beuzen J.N., Carson W., Corey-Lisle P.K., Falissard B., Jody D.N., Kujawa M. J., L'Italien G., Marcus R.N., McQuade R.D., Ray S., Van Peborgh P., IAQ Validation Study Group. (2005) Validation of the Investigator's Assessment Questionnaire, a new clinical tool for relative assessment of response to antipsychotics in patients with schizophrenia and schizoaffective disorder. *Psychiatry Research*, **136**, 211–221.
- Tandon R., Marcus R.N., Stock E.G., Riera L.C., Kostic D., Pans M., McQuade R.D., Nyilas M., Iwamoto T., Crandall D.T. (2006) A prospective, multicenter, randomized, parallel-group, open-label study of aripiprazole in the management of patients with schizophrenia or schizoaffective disorder in general psychiatric practice: Broad Effectiveness Trial with Aripiprazole (BETA). *Schizophrenia Research*, **84**, 77–89.
- Turner D., Schünemann H.J., Griffith L.E., Beaton D.E., Griffiths A.M., Critch J.N., Guyatt G.H. (2010) The minimal detectable change cannot reliably replace the minimal important difference. *Journal of Clinical Epidemiology*, **63**, 28–36.
- Wilson-d'Almeida K., Karrow A., Bralet M.C., Bazin N., Hardy-Baylé M.C., Falissard B. (2013) In patients with schizophrenia, symptoms improvement can be uncorrelated with quality of life improvement. *European Psychiatry*, **28**, 185–189, DOI: 10.1016/j.eurpsy.2011.10.002
- Witte M.M., Case M.G., Schuh K.J., Ascher-Svanum H. (2012) Effects of olanzapine long-acting injection on levels of functioning among acutely ill patients with schizophrenia. *Current Medical Research and Opinion*, **28**, 315–323, DOI: 10.1185/03007995.2012.657300
- Wyrwich K.W. (2004) Minimal important difference thresholds and the standard error of

- measurement: is there a connection? *Journal of Biopharmaceutical Statistics*, **14**, 97–110, DOI: 10.1081/BIP-120028508
- Wyrwich K.W., Tierney W.M., Wolinsky F.D. (1999) Further evidence supporting a SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *Journal of Clinical Epidemiology*, **52**, 861–873.
- Yeo W., Mo F.K., Koh J., Chan A.T., Leung T., Hui P., Chan L., Tang A., Lee J.J., Mok T.S., Lai P. B., Johnson P.J., Zee B. (2006) Quality of life is predictive of survival in patients with unresectable hepatocellular carcinoma. *Annals of Oncology*, **17**, 1083–1089, DOI: 10.1093/annonc/mdl065