Personalising Efficacy Scales based on Predominant Symptoms at Baseline

Laverdure-Dupont D[1], Nielsen M[2], Berger AK[1]

[1] Clinical Research – Neuropsychiatry, H. Lundbeck A/S, [2] Biometrics – Data Science, H. Lundbeck A/S

Abstract

Methodological Question Being Addressed

Can clinical outcome assessments be adapted to capture effects on the most severe symptoms for the individual patient and demonstrate a better sensitivity to detect a drug effect in randomized clinical trials?

Introduction

The symptom presentation of patients with neurological or psychiatric disorders is heterogeneous, and to capture the complex manifestation of the conditions, clinical scales typically includes all aspects of the concept (symptom or impact) important to patients (content validity). However, since not all symptoms are expressed in all patients, the use of these assessment tools to support clinical trial endpoints can lead to decreased sensitivity to detect drug-induced changes (bias toward the null). We present here a quantitative approach to individualise broad symptom scales to support efficacy evaluations focused on the predominant symptoms at baseline for each patient. The concept was tested on data obtained with the Montgomery-Asberg Depression Rating Scale (MADRS) in clinical trials with vortioxetine. The MADRS assesses the severity of depressive symptoms and includes 10 items, each rated on a scale ranging from 0-6.

Methods

For a symptom scale which includes several items each associated with a discrete symptom (such as the MADRS), the predominance of a symptom at baseline can be represented by indexing the individual item score at baseline with the total score at baseline for each patient (predominance index for symptom X = item score at baseline/total score at baseline). The predominance index for each item is used to weight the change at end of treatment for this specific item, leading to an increased contribution of the changes observed on the most severe symptoms for an individual, to the overall change from baseline in total score.

Results

Datasets of 5 vortioxetine trials were used, and the changes from baseline in MADRS total score were recalculated after each item change were weighted with its predominance index. For all active treatments in all trials tested, the application of the predominance index increased the mean group difference to placebo in the change from baseline in MADRS total score, with improvements ranging from 0.06 to 1.21 normalised points on the weighted scale. The proportionality between the treatment arms was conserved, and no important impact on the normality of the distribution was observed. However, an increase in variance was noted on weighted outcomes, leading to similar p-values for the standard and weighted scales. The weighted MADRS was also used to recalculate the proportion of responders (defined as those who display an improvement of 50% or greater relative to baseline) in these studies; the results generally show a higher proportion of responders to active treatment with the weighted MADRS, which translated into lower p-values for responder analysis endpoint.

Conclusion

The integration of the predominance of symptoms at baseline may represent an objective method to focus efficacy assessment on symptoms that are important to the individual patient. This quantitative, data-driven approach could be applied to widely accepted symptom scales, so that outcome measures become tailored to each patient, while being maintained in a widely recognized, unified conceptual framework.

The authors report no conflicts of interest for this work.

Background

Patients afflicted with neurological or psychiatric disorders present heterogeneous symptoms, even within a given diagnosis (e.g. major depression). In clinical trials, drug efficacy is often evaluated using clinical scales developed for diagnostic purposes.

Context of Use

- **Clinical Diagnosis**

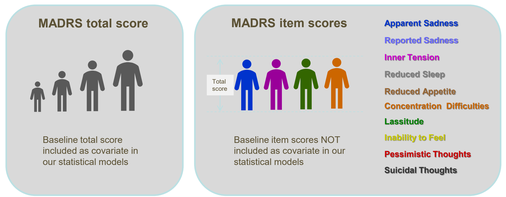
Content Validity: The instrument contains items covering all aspects of the concept important to patients, in order to reflect the complex manifestation of the disease.

- **Clinical Trial Endpoint**

Bias toward the null: Lost of sensitivity to detect drug-induced changes since not all symptoms are expressed in all patients.

Challenge: Identifying the combination of items most relevant for each patient in the trial within a unified clinical evaluation framework

In clinical trials, the ratings at Baseline, of the single items comprised in a symptom scale, can serve to identify the predominant symptoms for each patient.



MADRS properties
- Depressive symptomology
- Clinician-rated symptom severity as reported by the patient
- Simple scoring algorithm (10 items, response scale 0 to 6)

The overall goal is to develop fit-for-purpose clinical outcome assessments to increase patient relevance and scale sensitivity.

Method

Leveraging Item Scores at Baseline

A. **Identifying Symptoms of Primary Importance for each patient** based on the symptom severity at Baseline

B. **Objectively quantify the symptom predominance** by indexing the individual item score with the total score at baseline

$$\text{Predominance Index for Symptom X} = \frac{Item\ Score\ at\ Baseline}{Total\ Score\ at\ Baseline}$$

Data-driven, objective, No arbitrary cut-offs or ranking

C. a) **Weighing item changes at end of treatment** with Symptom Predominance Index
b) **Calculation of the Change from Baseline in Total Score**

→ ↑ contribution of the most severe symptoms on the overall drug effect

→ A drug-induced improvement cannot be expected on a symptom that is not present at Baseline

In sum, the weight or Predominance Index is calculated for each item and adjusts its contribution to the overall improvement at the end of treatment, based on its predominance at Baseline.

Postulate

As the most dysfunctional domains at Baseline have the most potential for improvement, an increased contribution of these domains is likely to lead to an increased ability to detect a drug response. Excluding or reducing the contribution of absent or mild symptoms will prevent the dilution of the drug effect on relevant symptoms (bias toward the null).

Results

Table 1: Weighting the MADRS using Item Scores at Baseline – Preliminary Results

Study	Treatment	N	Original Δ from placebo in MADRS total score	Weighted Δ from placebo in MADRS total score	Difference Weighted – Original (Δ from placebo in MADRS total score)	Difference Weighted – Original (SEM)	Responders (%) based on original MADRS (p-value)	Responders (%) based on weighted MADRS (p-value)
13267A	PBO	132	NA	NA	NA	NA	36	39
	VOR 15 mg	122	-6.69	-7.76	1.07	0.14	68*** (7.26e-07)	75*** (3.32e-08)
	VOR 20 mg	126	-5.91	-6.83	0.92	0.12	71*** (3.50e-08)	79*** (7.27e-10)
	DUL 60 mg	133	-8.58	-9.79	1.21	0.13	79*** (1.60e-11)	85*** (6.54e-13)
11492A	PBO	87	NA	NA	NA	NA	53	54
	VOR 5 mg	98	-1.6	-2.34	0.74	0.15	72** (0.00605)	77** (0.00137)
	VOR 10 mg	83	-0.66	-1.31	0.65	0.15	77** (0.00113)	84*** (3.89e-05)
	VEN 225 mg	98	-7.51	-8.39	0.88	0.14	84*** (1.21e-05)	87*** (2.71e-06)
11984A	PBO	123	NA	NA	NA	NA	52	60
	VOR 2.5 mg	131	-1.74	-1.8	0.06	0.13	60 (0.18351)	67 (0.24533)
	VOR 5 mg	123	-2.09	-2.33	0.24	0.1	64 (0.06775)	70 (0.12743)
	VOR 10 mg	120	-1.14	-1.3	0.16	0.11	68** (0.00965)	76** (0.00934)
	DUL 60 mg	113	-3.98	-4.62	0.64	0.11	71** (0.00370)	80** (0.00147)
14122A	PBO	187	NA	NA	NA	NA	31	37
	VOR 10 mg	187	-7.27	-7.95	0.68	0.1	48*** (0.000773)	56*** (0.000186)
	VOR 20 mg	196	-7.44	-8.32	0.88	0.11	61*** (1e-08)	71*** (2.32e-11)
T21004-305	PBO	129	NA	NA	NA	NA	26	38
	VOR 1 mg	127	-3.74	-4.4	0.66	0.17	48*** (0.000238)	57** (0.002811)
	VOR 5 mg	124	-7.33	-8.03	0.7	0.12	47*** (0.000441)	59*** (0.000644)
	VOR 10 mg	130	-6.39	-7	0.61	0.12	54*** (8.72e-06)	64*** (4.17e-05)

DUL: duloxetine, VEN: venlafaxine, VOR: vortioxetine

Examples from clinical trials with vortioxetine

Randomised, double-blind, parallel-group, placebo-controlled studies evaluating the efficacy and safety of vortioxetine in adult patients with Major Depressive Disorder

- Primary endpoint: Change from Baseline to end of Treatment in MADRS total score
- Secondary endpoint: Proportion of Responders defined as patients with a ≥ 50% reduction in MADRS total score, at the end of treatment

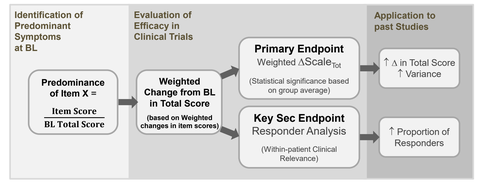
Summary of Results and Discussion

For all active treatments in all 5 trials tested, the change in MADRS total score at the end of treatment and relative to placebo, were larger with the use of the weighted MADRS. This suggests that clinical scales which focuses on improvements on the most severe symptoms at baseline for each patient may be more sensitive to detect drug-induced changes.

The use of the weighted MADRS lead to increased variance (SEM) in the group means. This, in conjunction with larger treatment effect relative to placebo, lead to similar p-values between the original and weighted results on the primary endpoints in the trials tested.

The use of the weighted MADRS to identify the responders to active treatment resulted in a higher proportion of patients meeting the responder criteria in the active treatment arms relative to the placebo arm. As this type of analysis is not based on group means and their associated variance, lower p-values were generally observed on the results of the secondary endpoints obtained with the weighted MADRS in the trials tested.

Potential Applications



The inclusion of the Predominance Index in efficacy scales would promote the development of **data-driven patient-centric efficacy assessments** measuring clinical improvements of domains relevant for each individual patient, while maintaining the statistical power of the standard scale.

As the most impacted domains at Baseline have the most potential for improvement, an increased contribution of these domains is likely to lead to an **increased sensitivity to identify a drug response**.

The identification of clusters of symptoms showing improvement after treatment with a specific drug, could **expand the understanding of the drug's Mode of Action** and help to **identify the appropriate clinical study population**.

Furthermore, drug development for CNS orphan diseases of suspected shared etiology but diverse symptomatology is limited by the availability of a common endpoint to be used in basket trials. By measuring drug effect on the most severe symptoms at baseline for each patient, this approach could capture different phenotypes within a unified measure of efficacy.

The authors report no conflicts of interest for this work.