

# Regional scoring differences observed in clinical trials

Steven D. Targum MD

ISCTM

February 24, 2022

# **Steven D. Targum, M.D.**

## **Disclosures February 2022**

### **Current Professional Affiliations**

Scientific Director, Signant Health

Chief Medical Officer, Functional Neuromodulation Inc (Minneapolis, MN)

Chief Medical Advisor, PAX Neuroscience

### **Equity interests**

Signant Health, Methylation Sciences Inc., EMA Wellness, Functional Neuromodulation Inc., PAX Neuroscience, XR Health

### **Consultation/CDA's within the past three years**

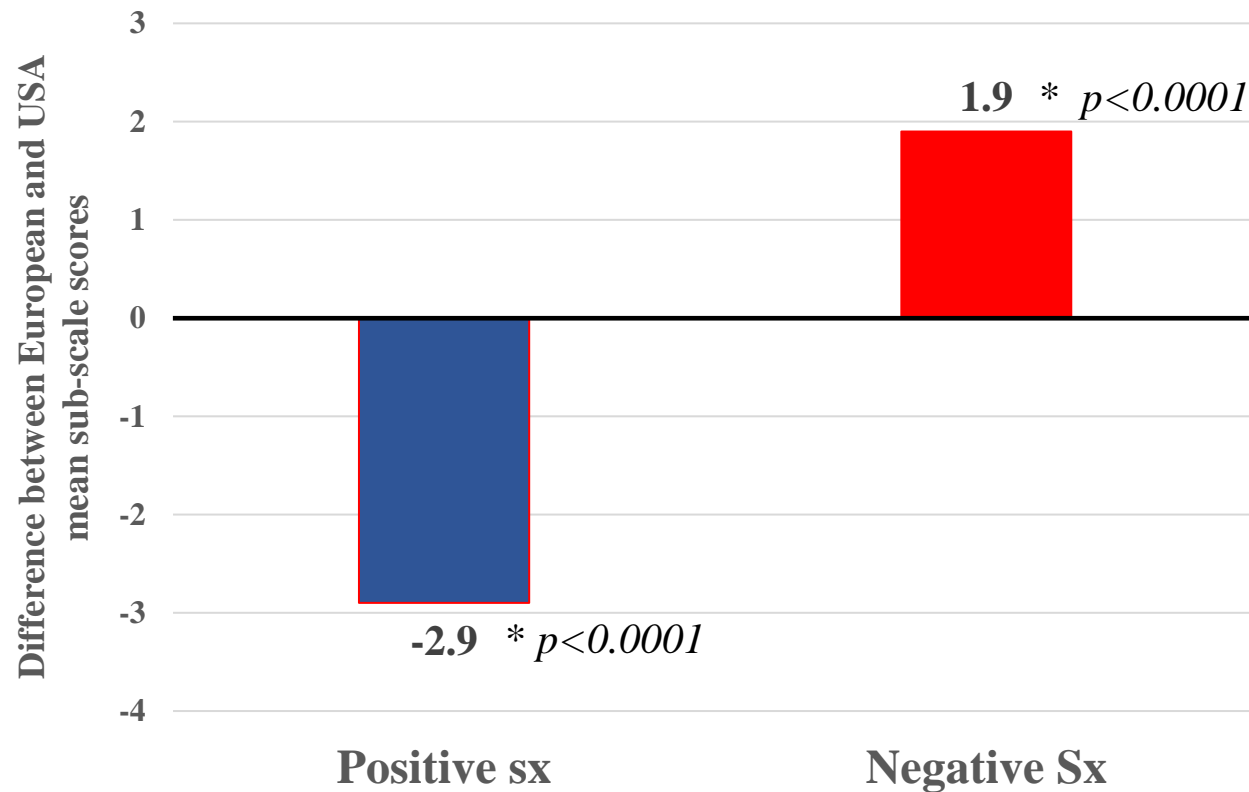
Acadia Pharmaceuticals Inc, AZ Therapies, BioXcel Therapeutics, EMA Wellness, Denovo BioPharma, Epiodyne Inc., Frequency Therapeutics, Functional Neuromodulation Inc, Intracellular Therapeutics Inc., Johnson & Johnson PRD, Karuna Pharmaceuticals, Methylation Sciences Inc., Merck Pharmaceuticals, Navitor Pharmaceuticals, Neurocrine, Pax Neuroscience, Sunovion

*Does merging scores from study data obtained from multinational studies have limitations?*

- Multi-national studies generally merge data from all geographic regions to achieve the sample size necessary to complete the trial
  - *Merged data does not account for regional scoring differences*
- This presentation will give 3 examples of regional scoring differences pertaining to the PANSS sub-scales, YMRS, and MADRS
- We will ask:
  - *Why are there regional scoring differences?*
  - *Does it matter?*

# Regional scoring differences in a study of negative symptom schizophrenia

*European rater scores had significantly **lower positive** but significantly **higher negative** PANSS sub-scale scores relative to USA raters*



# Regional scoring differences on the PANSS in a study of negative symptom schizophrenia

*Mean European scores had significantly **lower positive** but significantly **higher negative** scores*

Mean PANSS scores	Europe (n= 725)	USA (n=281)	difference	t test; p
	mean score ± SD	mean score ± SD		
<b><u>Total PANSS score</u></b>				
site-based score	73.0 ± 11.6	74.4 ± 9.9	-1.4	t= -1.79; p= 0.07
site-independent score*	73.5 ± 11.5	75.1 ± 10.7	-1.6	t= -2.13; p= 0.033
<b><u>PANSS positive symptom subscale</u></b>				
site-based score	13.1 ± 4.0	16.0 ± 3.3	-2.9	t= -10.54; p <0.0001
site-independent score	13.4 ± 4.0	16.8 ± 2.7	-3.4	t= -12.45; p<0.0001
<b><u>PANSS negative symptom subscale</u></b>				
site-based score	25.6 ± 4.2	24.0 ± 3.6	1.6	t= 5.39; p<0.0001
site-independent score	25.6 ± 4.2	23.2 ± 3.7	2.4	t= 8.35; p<0.0001
<b><u>Marder negative symptom factor score**</u></b>				
site-based score	24.5 ± 4.7	23.4 ± 4.1	1.1	t= 3.42; p= 0.0007
site-independent score	24.5 ± 4.4	23.1 ± 4.1	1.4	t= 4.44; p< 0.0001

\* Site-independent scores were obtained by observing video recording of site-based PANSS interviews

\*\*Marder negative symptom factor includes PANSS items N1, N2, N3, N4, N6, G7, G16

*Sometimes,  
Scrutiny can be  
painfully  
revealing !*



Gaspare Traversi, 1752

# Regional scoring differences: *Why is this happening?*

- Merged data would obscure the regional differences
  - *Are European and U.S. patient populations symptomatically different?*
  - *Are European and U.S. patient populations recruited differently?*
  - *Are European and U.S. raters interpreting symptoms differently?*
  - *Are European and U.S. raters applying different scoring conventions?*
- Regional scoring differences are not uncommon in CNS trials
- Socio-cultural factors may influence the interpretation of symptoms and affect scoring
- *Additional Examples of regional differences observed in training interviews*
  - The YMRS in bipolar disorder (USA, Great Britain, India)
  - The MADRS in the USA and Japan

# **Bipolar disorder:**

## **training interviews reveal regional scoring differences on the YMRS**

*possible socio-cultural differences in interpretation*

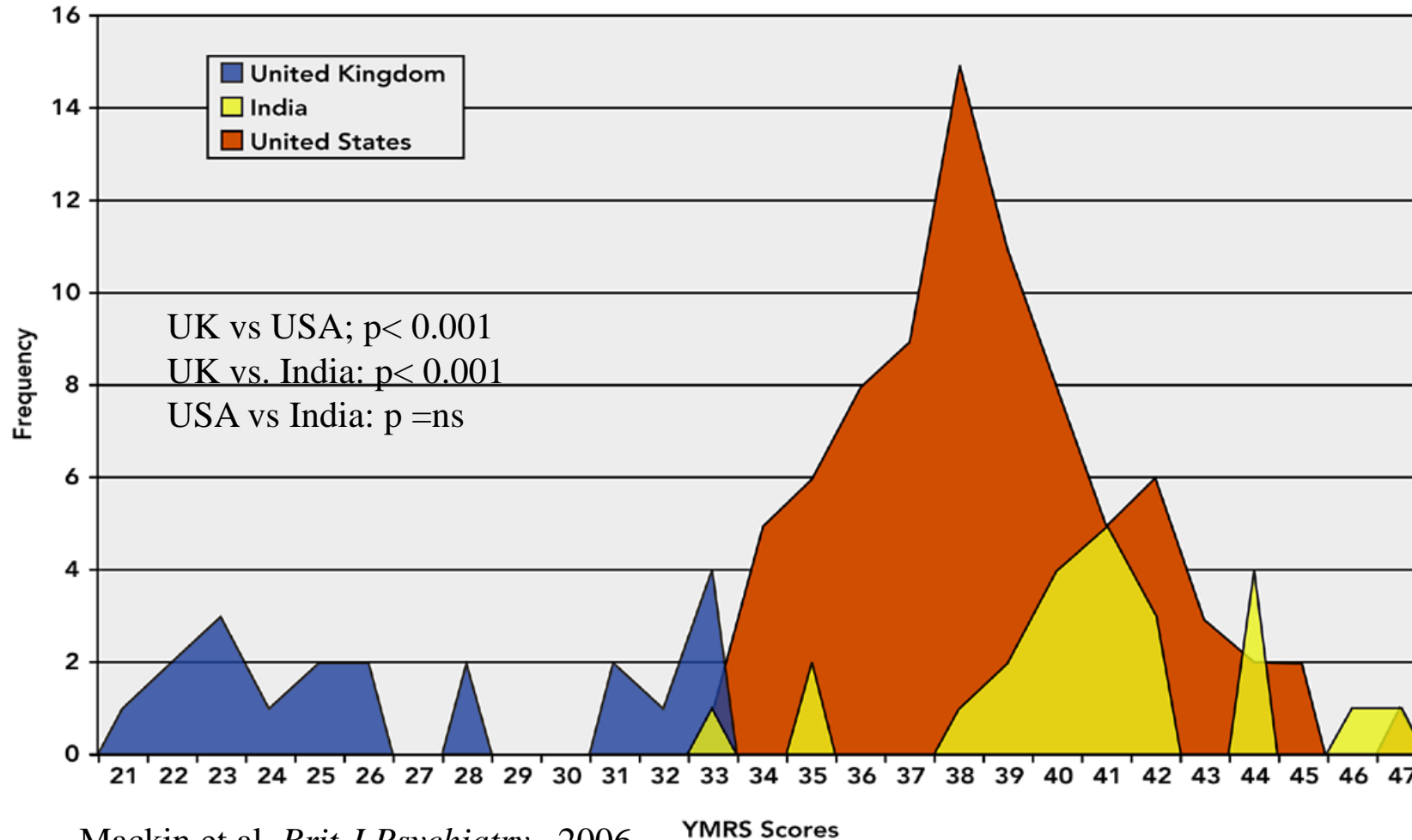
Mackin et al., *Brit J Psychiatr.*, 2006

- Two training YMRS video interviews of English-speaking bipolar subjects conducted in the USA were distributed to trained raters in the USA, Great Britain, and India
- Individual item analyses revealed significant scoring differences between the 3 countries on the total YMRS score and on 10 of the 11 YMRS items on each of the two interviews.
- The most profound individual item differences were:
  - Mood elevation (item 1):  $p < 0.001$
  - Irritability (item 5):  $p < 0.001$
  - Thought content (item 8):  $p < 0.001$
  - Disruptive–aggressive behaviour (item 9):  $p < 0.001$ .



# Distribution of total YMRS scores

## English speaking patient (interview 1)



Mackin et al, *Brit J Psychiatry.*, 2006

# USA vs. Japanese rater comparison of MADRS scores

obtained from rating a **Japanese-speaking** subject during a rater training exercise

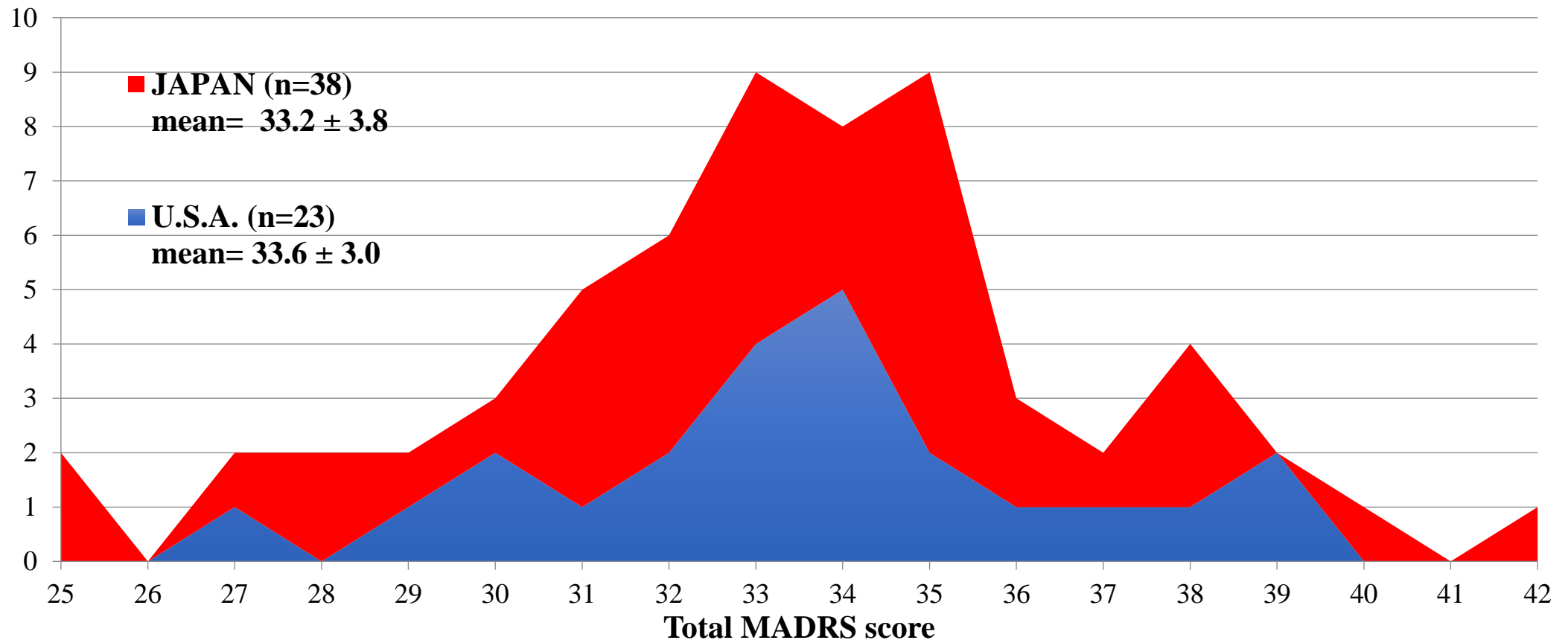
Targum, Nakagawa, and Sato, *Ann Gen Psychiat.*, 2013

- As part of an industry-sponsored rater training program, USA raters observed and scored two English sub-titled video versions of MADRS training interviews conducted in Japan in Japanese
- Sample size: USA raters n=23; Japanese raters=38
- The scoring variability of total MADRS scores between USA and Japanese raters rating the *same* subject was broad (25-42 in one case) but evenly distributed and not significantly different
- Two MADRS items that may be susceptible to cultural interpretation were scored significantly differently by the US and Japanese raters:
  - Lassitude (item 7):  $p= 0.013$  (with bonferroni correction)
  - Inability to feel (item 8):  $p= 0.037$

\*

## Distribution of total MADRS scores

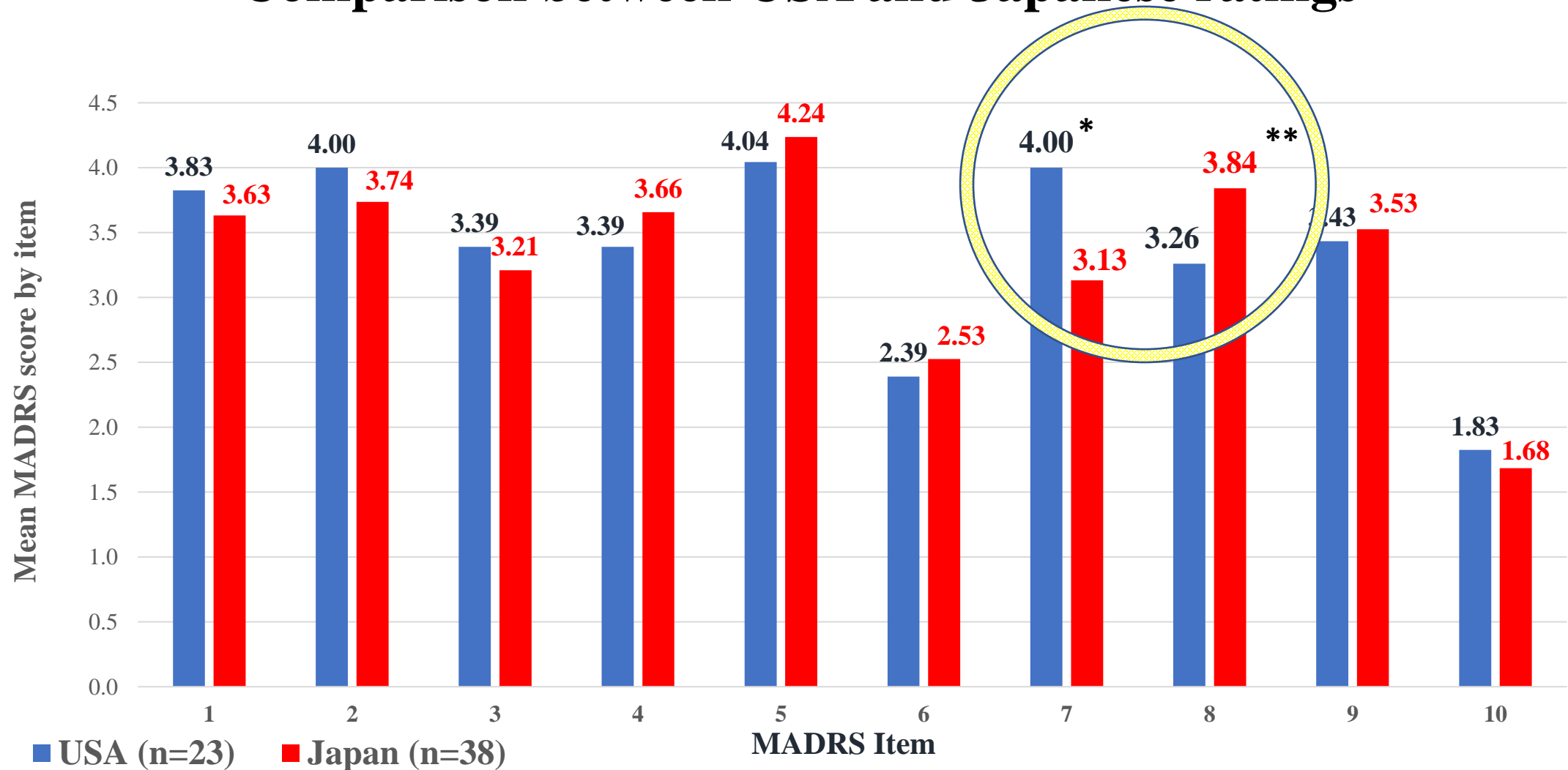
obtained from rating a **Japanese-speaking** subject during a rater training exercise\*



\*USA raters observed English **sub-titled video** version of MADRS interview conducted in Japanese (in Japan)

Targum et al., *Ann Gen Psychiatry* 2013

# Individual MADRS items Comparison between USA and Japanese ratings



\* Item 7: Lassitude:  $p= 0.013$  with bonferroni correction

\*\*Item 8: Inability to feel/decreased interests:  $p= 0.037$  with bonferroni correction

# Regional scoring differences in clinical trials

*not so hidden secrets*

Regional scoring differences  
are not uncommon...

*you just need to look for them*



## **Discussion: Regional scoring differences in clinical trials**

There are regional scoring differences on the PANSS, YMRS, and MADRS.

Merging regional data into a single dataset may adversely affect signal detection in clinical trials

- ***Are the regional differences due to sociocultural differences such that subjects with the same CNS disorders present symptoms differently?***
  - If this is true, do the benefits of merging data between geographic regions outweigh the risks (specifically, the increased variability)?
- ***Are the regional differences due to how trained raters from different regions interpret symptoms or apply scoring conventions?***
  - If true, have we failed to achieve inter-rater reliability between different geographic regions?
  - Are these differences due to the use of ratings scales with different languages?
  - Are all raters reliable?
- ***Are the regional differences due to how subjects are recruited for clinical trials?***
  - If this is true, are raters biased in their assessments?
  - What factors might drive the differences (e.g., subject payments, site structure [for profit vs nonprofit], prior trial experience, etc)?
- ***Do regional scoring differences matter?***
  - *Do we need to incorporate these findings into our statistical analysis plans?*