

厚生労働科学研究費補助金（難治性疾患政策研究事業）
（分担）研究報告書

日本語版中枢性感作症調査票Central Sensitization Inventory (CSI-J) における
特異応答項目の検討－頭痛患者と地域住民の対比－

研究分担者 岩田 昇 桐生大学医療保健学部・教授

研究要旨

中枢性感作症候群の評価ツールとして国際的な標準測定となっている Central Sensitization Inventory の日本語版 (CSI-J) への回答における頭痛患者と地域住民の対比により、特異応答項目 DIF (Differential Item Functioning, 同じ潜在特性レベルでも応答傾向が異なっているために見かけ上の得点差を招く状況) という測定バイアスの一種を探索的に検討した。DIF 解析に用いたのは、頭痛外来患者 168 名、地域住民 22,963 名の CSI-J 回答データであった。解析の結果、6 項目 (#1 覚醒時残存疲労、#2 筋骨格痛み、#7 光過敏、#10 頭痛、#11 膀胱不快感・痛み、#21 頻尿) で DIF が検出された。具体的に見ると、#1、#7 および #10 は、外来患者群の方が低い潜在特性での該当回答が認められ、高得点になりやすいものの、#2 および #21 では逆に一般地域住民の方が低い潜在特性での該当回答が認められた。また #11 は DIF を示すものの、高い選択肢の回答頻度が極端に少ないため、実質的な影響は乏しいものと判断された。頭痛項目 (#10) を除くと、DIF は 2 項目ずつで相殺する関係になっているために、結果的に CSI-J 合計では見掛け上の得点差をもたらすものではないことがうかがわれた。今後、他の分担研究の先生方のサイトからの臨床データも加えて、さらなる検討を行いたいと考える。

A. 研究目的

中枢性感作症候群の評価は、これまで主に Central Sensitization Inventory (CSI: Mayerら, 2012) という自己記入式測定尺度に基づいて行われてきた。CSIは25項目・5件法(配点0~4)で構成される、広範なCSS症状を網羅する自己記入式調査票である。2012年に原版(英語)が開発され、日本語版(CSI-J)も2017年に発表され(Tanakaほか, 2017)、信頼性・妥当性に関する報告も公表されている。

CSIのような患者の訴えに基づく健康関連指標(Health-Related Patient-Reported Outcomes: HR-PROs)に関しては、本人の主観であり、応答様式(response style)等も影響するがゆえに、その測定精度を疑問視する声が多かった。その一方で、社会疫学に代表される1990年代以降に隆盛になった学問領域では、広くQOLや幸福感など、心理的側面に関する変数も研究対象にするようになり、何らかの指針が求められていた。

そんな中で推奨されているのが、科学的ツールとしての評価基準COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) (Mokkink et al., 2010)である(図1)。紙面の都合上、本稿ではその説明は省くとして、COSMINチェック4段階の第2段階に、尺度構成項目の検討手続きとして項目反応理論(Item Response Theory: IRT)が位置付けられている。

従来この種の測定評価は、いわゆる項目固定型テストを前提とした古典的テスト理論に基づいていた。しかし、古典的テスト理論には1) 尺度単位(得点)でしか測定評価ができない、2) 信頼性・妥当性も尺度全体でしか保障されないなどの

制約がある。この制約を受けない柔軟な測定評価法がIRTである。IRTでは、測定しようとする潜在特性(θ)に対する個々の項目の測定特性を明らかにし、同時にその潜在特性上での各回答者のレベルを推定することができる。

この第2段階のチェックポイントの一つに特異応答項目(Differential Item Functioning: DIF, Holland & Thayer, 1988)の検討がある。図2に例示するように、DIFとはある外生変数(例、人種/民族)が異なる回答者間で、検討しているある潜在特性(例、抑うつ症状)を同じ潜在特性上に統制しても、項目の該当回答に相違が見られるような状況のことを指す(Clauser & Mazor, 1998)。

DIFは回答(得点)の「真の差」と混同される危険性があるため、ここで両者の違いを明確にしておく必要があると思われる。「真の差」とは、例えば臨床患者と健常者の対比のように、実際に測定対象としている特性のレベル自体が異なっており、そのレベルに応じて該当回答が発現しているため、必然的に得点差が生じているという「真実の差異」のことを指す。

それに対してDIFが存在する場合(図2)には、同じ潜在特性レベルでも該当回答の発現確率が異なる。すなわち、日本人ではかなり低い抑うつレベルでもポジティブ項目の抑制回答が見られるため、結果的にこの種の項目群の得点が高くなってしまふ。図2はIRTの枠組でDIFという測定バイアスが存在する状況が視覚化されている(Iwata, 2002)。

CSIの標準的評価方法は、合計点に基づく5段階評価が知られている(Neblett, 2018)。したがって、もしDIF項目による影響が存在した場合には、その臨床評価の精度が低下することになる。そこ

で2021年度は、一般地域住民と頭痛外来患者のCSI-J回答を材料にしてDIFの検出を試みた。また、男女間のDIFについての検討も行った。

B. 研究方法

1. 調査および参加者

地域住民：地域住民のCSI-J回答データについては、すでに2020年度報告書に記しているため、ここでは詳細は割愛するが、調査票の配布数39,152部、回収数24,189部（回答率61.8%）で、記入不備のない22,963名の回答データを一般集団参照群(reference group)とした。

頭痛患者：本研究班の分担研究機関となっている富永病院頭痛センター(分担研究者：竹島多賀夫、協力者：團野大介)の頭痛外来に来院された患者168名の回答データを臨床群(focal group)とした。なお、各患者の氏名等の個人情報はずべて病院内で削除し、性・年齢・診断名・CSI-J調査票回答のみの電子データの貸与を得て本解析に用いた。

2. 統計解析

いくつかのDIFの検出方法がある中で、ここではIRTによる測定特性に基づく指標K(熊谷, 2012)を用いた。指標Kは次の手続きに沿って実行した。まず、多値型IRT(Samejima, 1969?)による項目反応カテゴリ特性曲線(item response category characteristic curves: IRCCC)を描き、潜在特性(θ)に準じて回答の期待値(expected value curves: EVC)をプロットする。複数の群(対比群)間のEVCで囲まれた部分の面積値を指標Kと呼び、DIF検出の指標とする。シミュレーション検討の結果、この指標Kが(回答選択肢数-1)×0.1(あるいは×0.15)より大きい場合にDIFが存在すると判断される(熊谷, 2012)。CSI-Jは5選択肢なので、先行研究(Iwataら, 2020)にならば、 $K \geq 0.40$ でDIF傾向ありとし、 $K \geq 0.60$ でDIFありと判断した。

(倫理面への配慮)

本研究は厚生労働省科学研究費補助金(難治性疾患など政策研究事業)によるもので、獨協医科大学研究倫理審査委員会、富永病院研究倫理審査委員会および桐生大学研究倫理審査委員会の承認を経て行われた。

C. 研究結果

1. 臨床患者と一般住民のDIFチェック

CSI-J25項目の臨床患者と一般住民との対比による指標Kおよび男女間での対比による指標Kを表1および2に示す。臨床—一般間のDIFが認められた6項目(#1覚醒時残存疲労、#2筋骨格痛み、#7光過敏、#10頭痛、#11膀胱不快感・痛み、#21頻尿)は太字で表記した。

DIF項目の視覚化の前に、参照例としてnon-DIF項目(#18首と肩の筋肉が張った感じがする：指

標 $K=0.14$)のIRCCCおよびEVCを図3に示す。5選択肢(まったくない・まれにある・ときどき・頻繁に・いつも)に対応するIRCCC(左側)の各群の曲線は概ね似たカーブを描いているが、「いつも」回答は一般住民の方が低い θ 値から発現しており、それに呼応してEVC(右側)も θ 値が1を超えたあたりからわずかに高値となっている。

項目1(図4)では、著しく低い θ レベルから患者群の「まれにある」回答は発現し、「ときどき」・「頻繁に」の該当回答も同様に一般住民に先んじて発現している。そのような回答のために、低い θ レベルでも患者群の評定値は高くなるというDIFが存在することがEVCで見取れる。

項目2(図5)では、一般住民のIRCCCに比べて臨床患者のそれは、非常に狭い θ 値の範囲にかたまっている。これは回答選択肢の順との逆転がみられるような場合に生じるIRCCCの特徴で、EVCはむしろ一般住民の方が高得点になりやすい状況にあることを示している。

項目7(図6)は、非常に低い θ 値から臨床患者の該当回答は見られ、しかも高い選択肢も見られていることを示している。そのため、各選択肢のIRCCCははっきりとしたピークを持たず、広い θ 値の範囲で回答が存在することがうかがえる。EVCでも臨床患者群の見かけ上の高値が示されている。

項目10(図7)は、最も顕著なDIF項目である。臨床患者はすべて頭痛外来の患者であり、頭痛の自覚症状を有するために通院している。回答頻度を見ると、「まったくない」は164名中1名(0.6%)であり、「まれにある」回答でも7名(4.3%)しかない。最も多い回答は2番目に高い選択肢「頻繁に」への75名(45.7%)であり、次いで「いつも」の42名(25.6%)となっている。IRCCCの図で、臨床患者群はきわめて低い θ 値から「まれにある」回答が発現し、それに続く選択肢も全体として左側(低 θ 範囲)に位置しているのは、このような回答頻度が影響している。その結果として、非常に大きな得点の差異(EVC)がもたらされている。

項目11(図8)も項目10と類似したEVCが認められ、臨床患者が見かけ上高得点になっていることがうかがえるが、IRCCCは特徴的であった。各選択肢に対応したIRCCCが描かれるはずの左図だが、臨床患者のIRCCCは少ない。回答頻度を確認してみると、臨床患者の「いつも」回答はまったく存在せず、一般住民でも0.4%(76/21663)しかない。一つ軽い「頻繁に」でも一般住民では0.7%(154/21663)で、臨床患者では4.3%だが、総数164名なので実際の該当者は7名だけである。IRTに依拠したDIF検出は最も妥当な方法であるが、この項目のように稀有な回答選択肢が見られ

る場合には、推定が安定せず必ずしも適切ではないことを示しているものと考えられる。

項目 21 (図 9) は逆に一般住民が高得点となっていた。平均年齢を比較すると、一般住民の 63.5 ± 11.7 歳に比べて、臨床患者は 41.8 ± 13.0 歳と明らかに前者の方が高齢である ($t=21.4, p<0.001$)。加齢による頻尿傾向は広く認められるところであり、この項目 DIF 検出には回答者の年齢構成が影響している可能性がある。

男女間での DIF 検討では、項目 21 のみが DIF 傾向 (指標 $K > 0.40$) を示し、男性の方が見かけ上高得点傾向になる様子がうかがえた。加齢に伴う前立腺肥大の傾向が影響している可能性は否定できない。

D. 考察

令和 3 年度の研究として、中枢性感作症候群の国際的な標準測度となっている CSI-J に対する頭痛外来患者 168 名と一般地域住民 22,963 名の回答データを用いて、特異応答項目 DIF という測定バイアスの一種を探索的に検討した。解析の結果、6 項目 (#1 覚醒時残存疲労、#2 筋骨格痛み、#7 光過敏、#10 頭痛、#11 膀胱不快感・痛み、#21 頻尿) で DIF が検出された。

具体的に見ると、#1、#7 および #10 は、臨床患者の方が低い潜在特性での該当回答が認められ、高得点になりやすい項目であった。一方、#2 および #21 では逆に一般地域住民の方が低い潜在特性で該当回答が発現しており、地域住民の方が高くなる項目であった。また #11 は DIF を示すものの、高い選択肢の回答頻度が極端に少ないため、実質的な影響は乏しいと判断された。頭痛項目 (#10) を除くと、DIF は 2 項目ずつで相殺する関係になっているために、結果的に CSI-J 合計では見かけ上の得点差をもたらすものではないと考えられた。

CSI はこれまで欧州を中心にいくつもの言語版が開発されており、それに伴う心理測定法的特性 (信頼性・妥当性) の報告があるが、DIF の検討報告はほとんど見られない。本報告のような非常に大きな参照集団 (一般地域住民) との対比による DIF 検討は、これまでに例がなく、きわめて貴重な基礎データであると考えられる。昨年度報告した因子構造を参照し、今回の DIF 項目が特定の因子に関与しているのか否かを確認したが、顕著な DIF は認められなかった。

CSI の評価方法として、合計点に基づく段階分類があるが、その合計点を見かけ上高める、あるいは低めるような DIF 項目の影響は、少なくとも今回検討した頭痛外来患者との比較では、顕著ではないということが出来る。しかしながら、項目 11 の IRCCC (図 8) に見るように、実数が少ない

選択肢を含む場合の DIF 検討は、必ずしも適正とは限らないことに留意を要する。

今回の検討は、CSI のような患者の訴えに基づく健康関連指標が実証科学的ツールとなり得るのか否かを評価するために踏まえるべき基準、COSMIN チェックの一部に相当するものである。本研究班はさまざまな臨床像を呈する患者の CSI-J データを集めており、今後この臨床データの集約による DIF 検討を含む COSMIN チェックを行うことが望まれる。

COSMIN の評価領域 (図 1) に『反応性 Responsiveness』がある。時間経過とともに測定尺度のスコア変化を捉え評価する領域であるが、因子のスコア (評定値でも IRT の θ 値でも) の変動パターンから、このような CSS の病態に関する理解が得られる可能性もあると思われる。さらに、回答選択肢のスコアリングによっても尺度がもたらす情報量や鋭敏な θ のレベル等も異なることが知られている (Iwata et al., 2019)。すべての項目で 5 段階リカートスコアリングが最適な配点とは言えない可能性もある。今後さらに検討する必要がある。

CSS は難治性であり、一般地域住民の中にも、その症状に苦しむ人は少なからず存在すると思われる。精度の高い測定尺度・評価法を構築し、地域における症状保有率や直接・間接の関連因子の解明につなげていくとともに、臨床患者のデータとの対比から各 CSS 病態の疾病概念の明確化や鑑別診断の一助につなげていきたい。

E. 結論

中枢性感作症候群の国際的な標準測度となっている CSI-J に対する頭痛外来患者 168 名と一般地域住民 22,963 名の回答データを用いて、特異応答項目 DIF という測定バイアスの一種を探索的に検討した。解析の結果、6 項目 (#1 覚醒時残存疲労、#2 筋骨格痛み、#7 光過敏、#10 頭痛、#11 膀胱不快感・痛み、#21 頻尿) で DIF が検出された。臨床患者の方が見かけ上高得点となりやすい項目は 3 項目、逆に地域住民の方が高得点となりやすい項目は 2 項目で頭痛項目 (#10) を除くと、DIF は相殺する関係になっているために、結果的に CSI-J 合計では見かけ上の得点差をもたらすものではないと考えられた。

F. 健康危険情報

該当せず。

G. 研究発表

1. 論文発表
なし。

2. 学会発表

岩田 昇, 菊地賢一, 小橋 元. 中枢神経感作症調

査票 CSI-J の COSMIN チェック(1). 第 85 回日本心理学会, 2021 年 9 月, 東京(Web 開催).

H. 知的財産権の出願・登録状況(予定を含む。)該当せず。

I. 参考文献

- Iwata N, Kumagai R, Saeki I. Do mothers and fathers assess their children's behavioral problems in the same way as do their children?: an IRT investigation on the Strengths and Difficulties Questionnaire. *Japanese Psychological Research* 2020; 62: 87-100. (doi: 10.1111/JPR.12268)
- Iwata N, Tsutsumi A, Wakita T, Kumagai R, Noguchi H, Watanabe N. The effect of alternative scoring procedures on the measurement properties of a self-administered depression scale: an IRT investigation on the CES-D scale. *European Journal of Psychological Assessment* 2019; 35: 55-62. (doi: 10.1027/1015-5759/a000371)
- 熊谷龍一. 統合的 DIF 検出方法の提案 –“EasyDIF”の開発–. *心理学研究* 2012; 83: 35-43. (doi: 10.4992/jjpsy.83.35)
- Mayer TG, Neblett R, Cohen H, Howard KJ, Choi YH, Williams MJ, Perez Y, Gatchel RJ. The development and psychometric validation of the Central Sensitization Inventory (CSI). *Pain Practice* 2012; 12: 276-285. (doi:10.1111/j.1533-2500.2011.00493.x).
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research* 2010; 19: 539-549. (doi: 10.1007/s11136-010-9606-8)
- Neblett R. The Central Sensitization Inventory: a user's manual. *Journal of Applied Biobehavioral Research* 2018; 23: e12123. (doi: 10.1111/jabr.12123)
- Nishigami T, Tanaka K, Mibu A, Manfuku M, Yono S, Tanabe A. Development and psychometric properties of short form of central sensitization inventory in participants with musculoskeletal pain: a cross-sectional study. *PLoS ONE* 2018; 13: e0200152. (doi: 10.1371/journal.pone.0200152)
- Clauser BE, Mazor KM. Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice* 1998; 17: 31-44. (doi: 10.1111/j.1745-3992.1998.tb00619.x)
- Holland PW, Thayer DT (1988). Differential item performance and the Mantel-Haenszel procedure. In: Wainer H, Baun HI. (Eds.), *Test Validity*. Erlbaum, Hills-dale, NJ, 1988: pp. 129-145.
- Iwata N (2002). Are the Japanese more depressed than the U.S. individuals? – an IRT-DIF study –. In: (eds.) H. Yanai, A. Okada, K. Shigemasu, Y. Kano, J.J. Meulman, “New Developments on Psychometrics” pp. 231-238, Springer-Verlag: Tokyo.
- Iwata N, Buka S (2002). Race/ethnicity and depressive symptoms: a cross-cultural/ethnic comparison among university students in East Asia, North and South America. *Social Science & Medicine* 55: 2243-2252 (doi: 10.1016/S0277-9536(02)00003-5).
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph* 1969; Supplement 17: 1-100. (doi: 0.1007/BF03372160)

図1 COSMIN check process

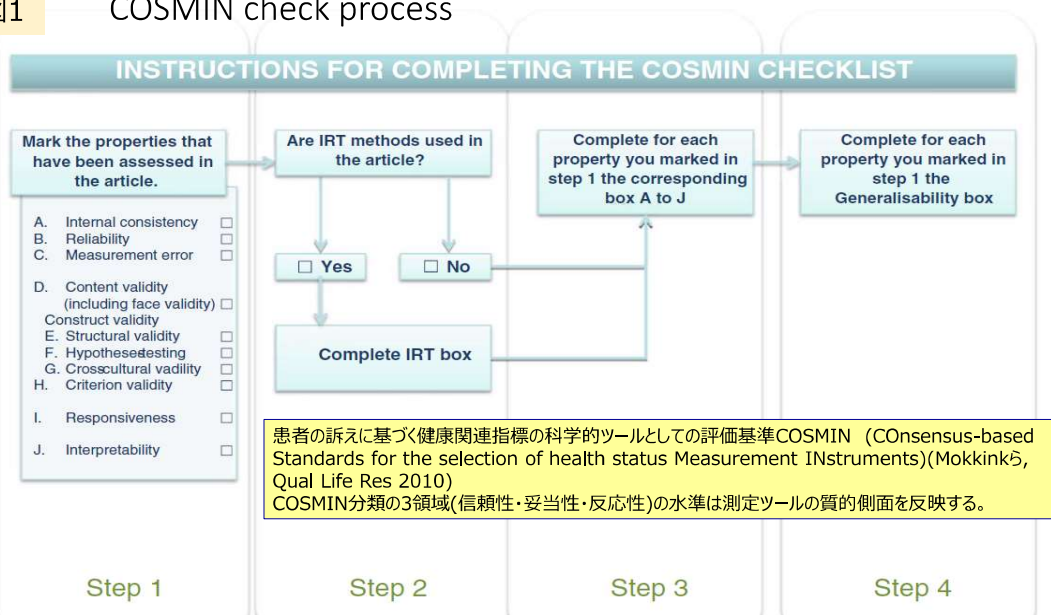


図2

DIF(特異応答機能)の例
Positive Affect (逆転項目)への日米回答の大きな相違

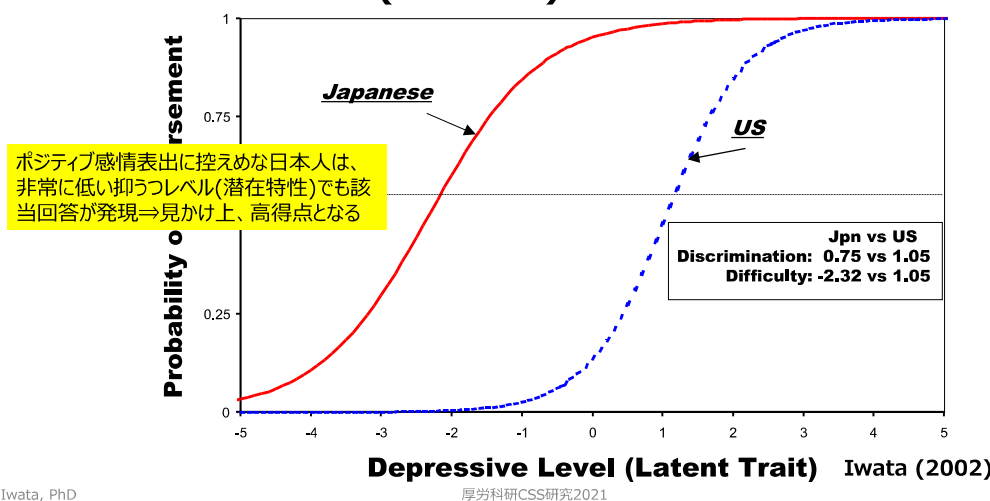


表1

CSI-JのDIFチェック (#1-13) Index K

項目	臨床 vs 一般	男性 vs 女性
1 目が覚めた時に、疲れが残るすっきりしない	.986	.221
2 筋肉に硬さや痛みを感じる	.813	.184
3 不安発作がある	.134	.046
4 歯を食いしばったり、歯ぎしりをしたりする	.179	.063
5 下痢や便秘の問題を抱えている	.119	.011
6 普段の生活での動作を行う上で、助けが必要である	.110	.014
7 明るい光に過敏である	.679	.124
8 身体を動かすと、すぐに疲れる	.176	.030
9 全身のあらゆるところに痛みを感じる	.281	.020
10 頭痛がある	1.785	.279
11 膀胱の不快感と排尿時にじわんとする痛みのいずれか/両方がある	1.763	.137
12 よく眠れない	.277	.067
13 集中することが難しい	.145	.122

表2

CSI-JのDIFチェック (#14-25) Index K

項目	臨床 vs 一般	男性 vs 女性
14 乾燥肌やかゆみ、発疹など、皮膚の問題がある	.332	.110
15 ストレスで身体の症状が悪化する	.262	.148
16 悲しかったり、ゆううつな気分になる	.145	.105
17 元気が出ない	.102	.033
18 首と肩の筋肉が張った感じがする	.141	.213
19 あごに痛みがある	.032	.022
20 香水などのある特定の匂いでめまいや吐き気がする	.220	.168
21 おしっこが近い(頻尿に排尿したくなる)	.917	.464
22 夜寝る時、あしに不快感や落ち着かない感じがある	.256	.046
23 物事を思い出すことが難しい	.372	.198
24 子どもの頃に心の大きな傷(トラウマ)を経験した	.095	.035
25 骨盤周辺に痛みがある	.252	.036

