

## 〈原 著〉

佐藤 寛\*<sup>1)</sup>, 石川 信一\*, 下津 咲絵\*\*, 佐藤 容子\*

### 子どもの抑うつを測定する自己評価尺度の比較 —CDI, DSRS, CES-Dのカットオフ値に基づく判別精度—

児童青年精神医学とその近接領域 50(3); 307—317 (2009)

本研究の目的は、子どもの抑うつを測定する自己評価尺度の判別精度を受信者操作特性(ROC)分析と層別尤度比(SSLR)の観点から検討することであった。一般対象者の中学生286名に対し、抑うつの自己評価尺度であるCDI, DSRS, CES-Dの日本語版を実施した。加えて、DSM-IVに基づくうつ病(大うつ病, 気分変調症, 小うつ病)の半構造化面接を実施し、抑うつの自己評価尺度との比較を行った。本研究の対象者のうち、15名(5.2%)が面接時点で何らかのうつ病の診断に該当していた。ROC分析の結果、これらの自己評価尺度はいずれも中程度以上のうつ病の判別力を示しており、各尺度の最適なカットオフ値はそれぞれ、CDI 31点, DSRS 24点, CES-D 37点であることが明らかにされた。SSLRを算出したところ、CDIでは0~21点で0.51, 22~30点で1.57, 31~54点で108.40となった。DSRSでは0~15点で0.46, 16~23点で1.06, 24~36点で∞であった。CES-Dでは0~15点で0.40, 16~36点で1.00, 37~60点で54.20であった。各尺度の日本語版における従来のカットオフ値(CDI 22点, DSRS 16点, CES-D 16点)を満たしていた場合でも、得点が本研究のカットオフ値に満たない場合にはうつ病の検査後確率は検査前確率とほとんど変わらないことが示された。

**Key words** : adolescents, assessment, depression, receiver operating characteristics, stratum-specific likelihood ratios

#### I. 問題と目的

子どものうつ病<sup>1)</sup>の時点有病率は、児童期で0.1%~6.8%, 青年期で1.1%~8.0%であるとされている(佐藤ら, 2008)。子どものうつ病は深刻なQOLの低下を引き起こし(Bastiaansen et al., 2004), 治療されずに放置されると長期的に維持されやすく(Nolen-Hoeksema et al., 1992), 高い確率で再発する(Rao et al., 1995)。

子どものうつ病の診断評価を行う上で、客観的な評価尺度を用いたスクリーニングは有用である。Myersら(2002)の展望論文によれば、これまでに11以上の子どものうつ病の評価尺度

が標準化されている。その多くは自己評価式の尺度であり、うつ病の診断や治療効果の査定に広く用いられている。わが国においても、海外で開発された尺度の日本語版がいくつか作成されており、子どものうつ病の評価尺度として使用することができる。

Children's Depression Inventory (CDI: Kovacs, 1985)は、世界で最も頻繁に用いられている子ども用の抑うつ尺度である。もともとはBeck Depression Inventory (BDI: Beck et al., 1961)の子ども版として開発されたものであり、多くの研究において高い信頼性と妥当性が示されている(Kovacs, 2003)。また、日本語版についても十分な信頼性と妥当性が確かめら

\*宮崎大学教育文化学部

1) e-mail: hsato@cc.miyazaki-u.ac.jp

\*\*九州保健福祉大学社会福祉学部

<sup>1)</sup>本稿において「うつ病」という語は、大うつ病, 気分変調症, 小うつ病, 特定不能のうつ病の総称として用いている。

れている(真志田ら, 印刷中)。原版のマニュアルにおいてカットオフ値は20点とされているが(Kovacs, 2003), 日本語版ではそれよりもやや高い22点が採用されている(村田ら, 1988)。

Birlesonら(1987)によって作成された Depression Self-Rating Scale for Children (DSRS) は, わが国で近年最も頻繁に使用されている子どもの抑うつ尺度である。DSRS は, Zung (1965) の Self-rating Depression Scale (SDS) をもとに作成されており, CDI に比べると平易な表現が用いられていることが特徴的である。原版の DSRS にはある程度の信頼性と妥当性が確認されており(Charman, 1994; Fundudis et al., 1991), 日本語版についても信頼性と妥当性が示されていて(村田ら, 1996), 大規模な標準データが報告されている(傳田ら, 2004; 佐藤ら, 2006)。尺度のカットオフ値は原版では15点(Birleson et al., 1987), 日本語版では16点(村田ら, 1996)とされている。

国内で使用可能なもう1つの尺度として, Center for Epidemiologic Studies Depression Scale (CES-D: Radloff, 1977) が挙げられる。CES-D は, 本来は成人向けの尺度として開発されたものであるが, 海外においては青年期のうつ病を測定する尺度としても広く用いられている(Myers et al., 2002)。日本語版は島ら(1985)によって作成されており, 良好な信頼性と妥当性が報告されている。尺度のカットオフ値は, 原版・日本語版ともに16点とされている(Radloff, 1977; 島ら, 1985)。

以上のように, CDI, DSRS, CES-D の3つの自己評価尺度はいずれも高い信頼性と妥当性を有しており, 臨床レベルの抑うつ症状の有無を判断するカットオフ値が設定されていることから, 子どものうつ病のスクリーニングに有用であると考えられる。しかしながら, これらの尺度を子どものうつ病の診断に役立てるためには, いくつかの検討すべき点が残されている。

第一に, これらの尺度の日本語版(および原版)では, うつ病患者群と一般対照群との比較によってカットオフ値が定められている。この

ように典型的な患者の群と健常対象者の群とを比較する方法は, 対象者に偏りが生じるために診断精度が過大評価されてしまうという範囲バイアス(spectrum bias: Ransohoff et al., 1978)が生じる。適切なカットオフ値を設定するためには, 実際に検査の対象となりうるすべての範囲の対象者を含めた分析を実施することが求められる。

第二に, スクリーニングテストの判別精度を評価するための方法としては, 受信者操作特性(Receiver Operating Characteristics: ROC)分析が適用されることが多い。ROC分析では, ROC曲線下領域(Area Under the ROC Curve: AUC)<sup>2</sup>, 感度, 特異度, 陽性的中率, 陰性的中率, 誤判別率などといった多様な指標を算出し, 検査の判別精度について多角的な評価を行う。中でも, 陽性的中率と陰性的中率は, 検査で陽性・陰性となった結果が実際の疾患の有無とのくらいの割合で一致するかを表す指標であり, 日常的な臨床場面においても有用性が高い。しかしながら, CDI, DSRS, CES-Dといった尺度の日本語版では, 判別精度の指標として感度と特異度が報告されているのみである。したがって, これらの尺度においてもROC分析に基づいた, より厳密な判別精度の検証を行う必要があるといえる。

そして第三に, 近年ではより妥当性の高い判別精度の評価法として, 層別尤度比(Stratum-Specific Likelihood Ratio: SSLR)<sup>3</sup>が適用され始めている。検査結果が陽性・陰性の2値データしか示さない場合には, 従来の感度や特異度によるアプローチもある程度の妥当性がある。しかしながら, 抑うつ尺度のように連続値をと

<sup>2</sup> AUC は, 1-特異度を横軸, 感度を縦軸にとったグラフ(ROC曲線)の下部領域の面積である。AUC が大きいほど, その検査の診断精度が高いことを示す。

<sup>3</sup> SSLR は, 疾患を有する者のうちその検査値をとる者の割合を, 疾患を有しない者のうちその検査値をとる者の割合で割ったオッズである。すなわち, SSLR=1の場合, その検査の判別精度は偶然の確率に等しいことを表し, SSLRが1よりも小さくなればなるほど疾患の可能性が低くなることを, 1よりも大きくなればなるほど疾患の可能性が高くなることを示す。詳しくは古川(2000)を参照のこと。

る検査の場合、検査結果をいくつかの段階に層別化し、段階ごとの評価を行うことが有用である(古川, 2000)。従来のアプローチでは、カットオフ値を上回った対象者はすべて等質の集団とみなされていたが、SSLRを導入することによってカットオフ値を段階的に設定・評価することが可能になり、それぞれの段階ごとに疾患の確率を算出できるようになる。わが国でSSLRを適用した先行研究としては、精神科初診患者にCES-Dを実施した例(Furukawa et al., 1997)や、一般人口における精神疾患の簡便なスクリーニングテストを検討した例(古川ら, 2003)がある。

以上の観点に基づき、本研究では一般サンプルにおけるCDI, DSRs, CES-D日本語版の判別精度をROC分析とSSLRの観点から評価し、比較・検討することを目的とする。

## II. 方法

### 1. 対象者

宮崎県の公立中学校に在籍する一般中学生347名(男子189名, 女子158名, 年齢範囲12~14歳, 平均年齢 $13.3 \pm 0.6$ 歳)を対象に、質問紙と面接の実施を依頼した。いずれかの回答に不備があった対象者、もしくはいずれかを実施した日に学校を欠席していた対象者を除く286名(男子155名, 女子131名, 年齢範囲12~14歳, 平均年齢 $13.3 \pm 0.6$ 歳)を分析の対象とした(回答率82.4%)。

### 2. 質問紙尺度

抑うつを測定する自己評価尺度として用いたのは、以下の3つの尺度の日本語版である。まず、CDI(Kovacs, 1985)は27項目3件法で構成されており、過去2週間の抑うつ症状について質問する尺度である。得点範囲は0~54点である。次に、DSRS(Birleson et al., 1987)は18項目3件法の尺度であり、過去1週間の抑うつ症状について尋ねる。得点範囲は0~36点となる。CES-D(Radloff, 1977)は20項目4件法からなっており、過去1週間の抑うつ症状を質

問している。得点範囲は0~60点である。質問紙は学級内において配布・実施を行った。回答した質問紙は個別に配布された封筒に入れた状態で回収した。

### 3. 診断面接

DSM-IVに基づく半構造化面接法であるAnxiety Disorders Interview Schedule for Children (ADIS-C: Silverman et al., 1996)日本語版の中から、大うつ病, 気分変調症, 小うつ病<sup>4</sup>の診断面接を実施した。ADIS-Cは、本来は子どもの不安障害を診断するために開発された構造化面接であるが、うつ病を含む関連した精神疾患の診断も可能であり、良好な再検査信頼性と評定者間信頼性が確認されている(Lynehan et al., 2007; Silverman et al., 2001)。日本語版は佐藤ら(2008)によって評定者間信頼性と併存的妥当性が確認されている。ADIS-Cでは、DSM-IVにおいて定義されている大うつ病の9つの症状、および気分変調症の7つの症状を順次質問し、回答を評価することによってうつ病の診断を行う。面接は質問紙回収後1~2週間以内に学校内の共同利用教室を用いて実施した。対象者は一人ずつ面接室に入り、一人あたりおよそ10分前後の面接を受けた。

### 4. 面接実施者および倫理的配慮

面接は臨床心理士3名、および臨床心理学を専攻する大学院生5名によって実施された。3名の臨床心理士は精神科医療と教育心理臨床に3年以上従事しており、児童青年期の精神保健を専門とした臨床経験を有していた。5名の大学院生は、精神科の医療機関において2年間の実務を伴う実習の経験があった。面接実施者全員に対し、ADIS-Cの実施方法や、自殺の危険があるなどの緊急性の高い子どもへの対応法について、ロールプレイを伴う6時間のトレーニングを行った。緊急性の高い子どもについては、

<sup>4</sup>小うつ病の診断基準はADIS-Cには含まれていないが、大うつ病の診断基準をもとに症状の数が2つ以上5つ未満認められる場合に小うつ病の基準を満たすものとして判断した。

学校長を含む管理職、養護教諭、および担当のスクールカウンセラーにただちに連絡した。さらに、担当のスクールカウンセラーによる緊急面接を実施した上で、近隣の医療機関への紹介を行った。

面接は地域の中学校を対象とした「心の健康診断」として行われた。この事業は県教育委員会の主催するスクールカウンセラー緊急配置事業に伴って実施されたものであり、対象校における通常の教育活動の一環として計画された。本研究の目的と方法および個人情報の取り扱いについて学校長に説明を行い、実施の同意を得た。

対象者には、①面接は普段の気分や体の調子などについて尋ねるものであること、②質問に正しい答えやまちがった答えはないこと、③面接で答えた内容は、家族や友人、教師などに漏れることはなく個人のプライバシーは守られること、④質問への回答は強制ではないこと、などといった点について書面と口頭で説明を行い、参加への同意を得た。

なお、本研究実施当時、著者の所属する宮崎大学教育文化学部には倫理委員会が未開設であったため、教育臨床心理専修講座内に設けられた研究会議において、著者・共著者を除く複数の専門家に本研究の倫理的側面に関する精査・検討を依頼し承認を得た。

### Ⅲ. 結果

#### 1. うつ病の半構造化面接の結果

ADIS-Cによる面接の結果、対象者286名のうち15名(5.2%)が面接時点においていずれかのうつ病の診断基準に該当した<sup>5</sup>。診断の内訳は、大うつ病1名(0.3%)、気分変調症13名(4.5%)、小うつ病4名(1.4%)であった。大うつ病の診断に該当した1名と、小うつ病の診断がついた4名のうち1名については、気分変調症の基準も同時に満たしていた。本研究では、面接時点

<sup>5</sup> 本研究のデータのうち、うつ病の有病率について疫学的観点から検討を行ったデータは佐藤ら(2008)において詳しく述べられている。

においていずれかのうつ病の診断を満たしていた15名をうつ病群とし、うつ病の診断に当てはまらなかった271名を非うつ病群として分析を行った。

#### 2. ROC 曲線による尺度の比較

CDI, DSRS, CES-DのROC曲線を図1に示した。AUCと95%信頼区間をそれぞれ算出したところ、CDIではAUC=0.85(95%CI:0.76-0.93)、DSRSではAUC=0.84(95%CI:0.74-0.95)、CES-DではAUC=0.82(95%CI:0.71-0.93)であった。Swets(1988)によれば、AUCは0.5~0.7で低い判別力、0.7~0.9で中程度の判別力、0.9以上で高い判別力を表すとされている。このことから、3つの尺度はいずれも中程度以上の判別力があり、判別力の高い方からCDI, DSRS, CES-Dの順となることが示唆された。

#### 3. 最適なカットオフ値の検討

尺度のカットオフ値を設定する際に、従来は感度と特異度のバランスによってカットオフ値を決定することが多かった。しかしながら、適切なカットオフ値を定めるためには、感度と特異度だけでなく、偽陽性率と偽陰性率のバランスや、有病率も考慮に入れた指標を用いることがより妥当である。このような観点から、Kraemer(1992)は検査と診断の一致度を感度と特異度のバランスによって重みづけした評価

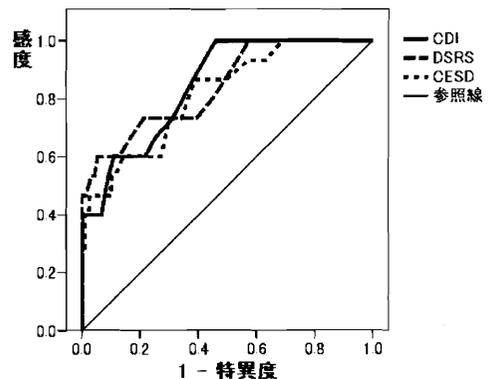


図1 CDI, DSRS, CES-DのROC曲線

表1 従来のカットオフ値と本研究におけるカットオフ値の判別精度の比較

	$x(w, 0)^*$	陽性的中率	陰性的中率	感度	特異度	誤判別率
CDI						
≥22 (従来のカットオフ値)	0.26	25.0%	97.2%	53.3%	91.1%	10.8%
≥31 (本研究のカットオフ値)	0.57	85.7%	96.8%	40.0%	100%	3.5%
DSRS						
≥16 (従来のカットオフ値)	0.20	20.9%	97.5%	60.0%	87.5%	14.0%
≥24 (本研究のカットオフ値)	0.69	100%	97.1%	46.7%	100%	2.8%
CES-D						
≥16 (従来のカットオフ値)	0.07	10.7%	97.8%	73.3%	66.1%	33.6%
≥37 (本研究のカットオフ値)	0.53	75.0%	96.8%	40.0%	99.3%	3.8%

\*ただし,  $w=0.5$ とする。

指標である  $x(w, 0)$  を推奨している。この指標は、以下の数式で表わされる。

$$x(w, 0) = \frac{w[\text{Pr}(D+)\text{Pr}(T-)x(1, 0)] + (1-w)[\text{Pr}(D-)\text{Pr}(T+)x(0, 0)]}{w[\text{Pr}(D+)\text{Pr}(T-)] + (1-w)[\text{Pr}(D-)\text{Pr}(T+)]}$$

ここで,  $\text{Pr}(T+)$ は検査が陽性となる確率,  $\text{Pr}(T-)$ は検査が陰性となる確率,  $\text{Pr}(D+)$ は疾患を示す確率,  $\text{Pr}(D-)$ は疾患を示さない確率である。また,  $x(1, 0) = (\text{感度} - \text{Pr}(T+)) / \text{Pr}(T-)$  であり,  $x(0, 0) = (\text{特異度} - \text{Pr}(T-)) / \text{Pr}(T+)$  とする。 $w$  は感度に対する重みづけの大きさを表しており,  $w=1$  とすると感度のみを重視した指標,  $w=0$  とすると特異度のみを重視した指標,  $w=0.5$  とおくと感度と特異度を同程度重視した指標となる。本研究では, 感度と特異度を同程度に重視した指標を用いることとし,  $w=0.5$  においてカットオフ値を検討した。

各尺度について, 最小得点から最大得点までカットオフ値を1点ずつ移動させ, それぞれの得点をカットオフ値とした場合の  $x(w, 0)$  を算出した。その結果,  $x(w, 0)$  が最も高い値を示したカットオフ値は, CDIでは31点, DSRSでは24点, CES-Dでは37点となった。いずれの尺度においても, 最適なカットオフ値は先行研究において推奨されている得点 (CDI=22点, DSRS=16点, CES-D=16点) よりも高いことが示唆されていた。

従来のカットオフ値と本研究におけるカットオフ値にそれぞれ基づいた場合の  $x(w, 0)$ , 陽性的中率, 陰性的中率, 感度, 特異度, 誤判別率を表1に示す。これを見ると, いずれの尺度においても従来のカットオフ値は陽性的中率が低く, 誤判別率が高い。一方, 本研究において算出されたカットオフ値を適用すると,  $x(w, 0)$ , 陽性的中率, 誤判別率は大きく改善され, 特異度もやや向上することがわかる。また, 陰性的中率は本研究のカットオフ値を用いた方がわずかに低くなるが, 確率としてはほとんど変わりがなかったことが示された。なお, 感度については従来のカットオフ値を用いた方が比較的高い値を示していたが, いずれのカットオフ値を用いた場合においても感度の数値はあまり良好なものではなかった。うつ病の診断が陽性・陰性であった対象者ごとに, CDI, DSRS, CES-Dの尺度得点のヒストグラムを図2~4に示す。

以上の指標を3つの尺度間で比較すると, 本研究のカットオフ値を用いた場合にはDSRSが最もうつ病の判別能力が高く, 次いでCDI, CES-Dの順であることが示された。

#### 4. SSLR による判別精度の段階的評価

本研究において算出されたカットオフ値は, いずれの尺度においても従来のカットオフ値よりも高い得点であった。そこで, それぞれの尺度の得点範囲を, 従来のカットオフ値未満, 従来のカットオフ値以上かつ本研究のカットオフ

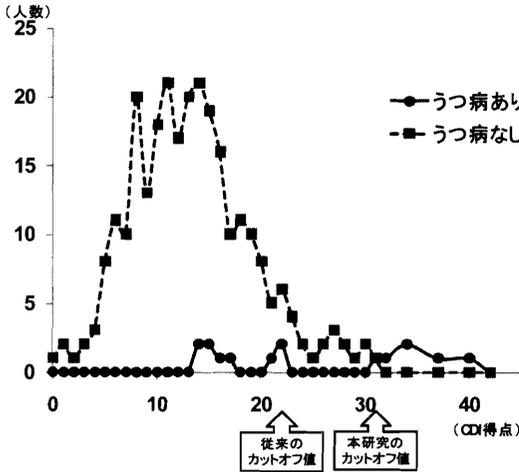


図2 CDI得点のヒストグラムとカットオフ値

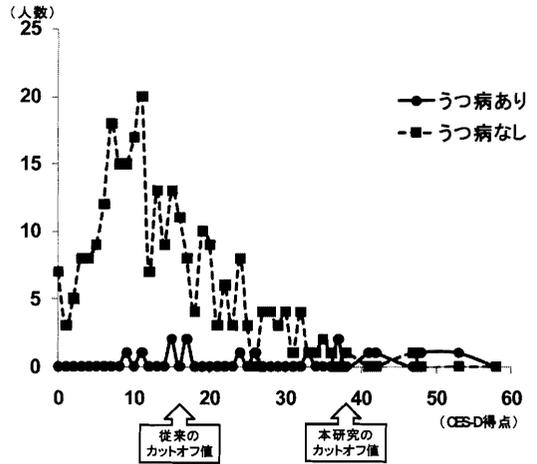


図4 CES-D得点のヒストグラムとカットオフ値

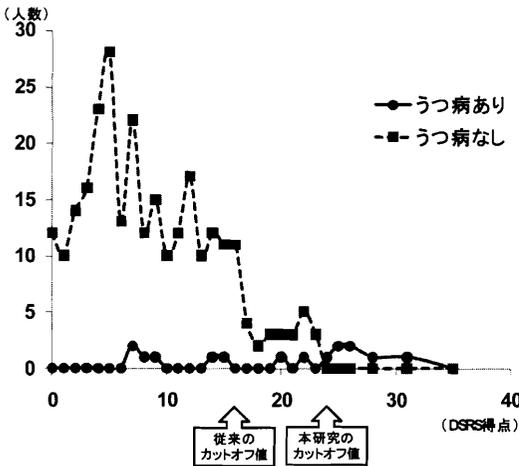


図3 DSRS得点のヒストグラムとカットオフ値

0.5~2の SSLR はあまり意味を持たないとされている。

古川 (2000) の基準に従うと, CDI では本研究におけるカットオフ値 (31点) を上回った場合には SSLR=108.40 とかなり大きな値を示していた。しかしながら, 従来のカットオフ値を上回っていても本研究のカットオフ値より低い得点であった場合 (22~30点) には, SSLR=1.57 とあまり大きな値は認められなかった。一方, 従来のカットオフ値よりも低い得点を示していた場合には, SSLR=0.51 であり, 疾患の確率はやや低まる可能性が示された。同様の傾向は DSRS と CES-D についても認められた。いずれの尺度においても, 本研究のカットオフ値を超える得点を示した場合には高い SSLR の値を示しており (DSRS : ∞, CES-D : 54.20), うつ病が強く疑われることが示唆された。しかし, 従来のカットオフ値を上回っていても本研究のカットオフ値を下回る得点であれば SSLR の値はほぼ 1 に近く (DSRS : 1.06, CES-D : 1.00), 診断を左右するほどの値ではなかった。また, 従来のカットオフ値よりも得点が低ければ SSLR の値は低まり (DSRS : 0.46, CES-D : 0.40), うつ病の可能性は低くなることが示された。なお, 3つの尺度の間で SSLR を比較すると, 本研究のカットオフ値を用いた場合に

値未満, 本研究のカットオフ値以上の3つの段階に層別化し, これらの段階について SSLR を算出した。各尺度について算出された SSLR と, その95%信頼区間を表2に示す。

SSLR は, 1 よりも大きければ大きいほど, その検査結果が出た場合の疾患の可能性が高くなり, 1 よりも小さければ小さいほど, その検査結果が出た場合の疾患の可能性は低くなることを意味する指標である。また古川 (2000) によれば, 10以上もしくは0.1以下の SSLR はしばしば確定診断に至らせ, 5以上もしくは0.2以下であれば疾患の確率を大きく変化させるが,

表2 SSLRによるCDI, DSRS, CES-Dの判別精度の段階的評価

	SSLR	SSLRの 95%信頼区間	一般対象者における 検査後確率 <sup>a</sup>	精神科患者における 検査後確率 <sup>b</sup>
CDIの得点範囲				
0~21	0.51	0.31-0.86	2.7%	16.0%
22~30	1.57	0.47-5.21	7.9%	36.9%
31~54	108.40	19.80-593.39	85.6%	97.6%
DSRSの得点範囲				
0~15	0.46	0.25-0.82	2.4%	14.5%
16~23	1.06	0.33-3.45	5.5%	28.3%
24~36	∞		100%	100%
CES-Dの得点範囲				
0~15	0.40	0.18-0.88	2.2%	13.0%
16~36	1.00	0.50-2.01	5.2%	27.2%
37~60	54.20	13.83-212.42	74.8%	95.3%

<sup>a</sup>一般対象者のうつ病の有病率(検査前確率)を5.2%とした場合(本研究のデータより)

<sup>b</sup>精神科患者のうつ病の有病率(検査前確率)を27.1%とした場合(傳田ら, 2001のデータより)

はDSRSが最も値が高く, 次にCDI, そしてCES-Dという順であった。

ところで, SSLRと疾患の検査前確率(対象母集団における有病率)が明らかな場合, Faganのノモグラム(Fagan, 1975)を用いることで疾患の検査後確率(検査結果を受けて推定される疾患の確率)を推定することが可能になる。臨床現場において役に立つ検査とは, 疾患の検査後確率を検査前確率よりも大きく高めることのできる検査であると言い換えることができる。そこで, 本研究において得られた一般対象者におけるうつ病の有病率(5.2%), および傳田ら(2001)によって報告されている児童青年期の精神科受診患者におけるうつ病の有病率<sup>6</sup>(27.1%)をもとに, CDI, DSRS, CES-Dを用いた場合の検査後確率の推定を行った(表2)。

これを見ると, 一般対象者に対してこれらの尺度を実施した場合, 本研究において得られたカットオフ値を超える得点を示していれば, かなり高い確率でうつ病の診断に当てはまることがわかる(CDI: 85.6%, DSRS: 100%, CES-D: 74.8%)。一方で, 従来のカットオフ値を超えていたとしても, 本研究のカットオフ値未満

の得点であれば, うつ病の診断に該当する確率はさほど高くはない(CDI: 7.9%, DSRS: 5.5%, CES-D: 5.2%)。

同様に, 傳田ら(2001)のデータをもとに精神科患者にこれらの尺度を適用した場合を仮定すると, 本研究のカットオフ値を上回っていればうつ病の診断基準を満たす可能性は極めて高い(CDI: 97.6%, DSRS: 100%, CES-D: 95.3%)。しかしながら, 従来のカットオフ値と本研究のカットオフ値の間の得点を示す場合には, 検査後確率は検査前確率とほぼ同程度で, あまり大きな違いはなかった(CDI: 36.9%, DSRS: 28.3%, CES-D: 27.2%)。

#### IV. 考察

本研究の結果から, ①CDI, DSRS, CES-Dの日本語版はいずれも子どものうつ病に対して中程度以上の判別力を持つ, ②しかしながら, これらの尺度の最適なカットオフ値は先行研究において推奨されていたものよりも高い得点であり, CDIでは31点, DSRSでは24点, CES-Dでは37点が最も適切なカットオフ値と考えられる, ③一般対象者と精神科患者のいずれにおいても, 本研究のカットオフ値を上回っていた場合にはかなり高い確率でうつ病の診断が疑われ

<sup>6</sup>本研究の分析に傳田ら(2001)の有病率データを使用することについては著者より承諾を得ている。

る、④従来のカットオフ値を超える得点を示している、本研究のカットオフ値より低い得点であればうつ病の検査後確率は検査前確率とあまり変わらない、といった点が示された。

本研究において、CDI, DSRS, CES-Dといった自己評価尺度が子どものうつ病のスクリーニングテストとして有用であることが明らかにされた。自己評価尺度は、他者評価による測定尺度に比べて簡便に実施することが可能であることから、スクリーニングの初期の段階で用いるのに適している。本研究において示されたカットオフ値に基づく判断の陽性的中率は高く、スクリーニングにいずれの尺度を用いた場合でも、これらのカットオフ値以上の得点を示した対象者はうつ病の診断を考慮する必要があると考えられる。

CDI, DSRS, CES-Dの判別精度を本研究の結果に基づいて比較すると、ほぼすべての指標においてDSRSの判別精度が最も高いという結果が得られている。このことから、一般対象者における子どものうつ病のスクリーニングテストとしてはDSRSが最も推奨され、次いでCDI, CES-Dの順であるということが出来る。ただし、これらの尺度の判別精度を示す指標はいずれも高い水準に達しており、実際にはどの尺度も十分な判別精度を有していると考えられる。

一方で、これらの尺度の得点が本研究のカットオフ値に満たなかったからといって、うつ病の診断が除外されることを意味するわけではないことには留意すべきである。これらの尺度を一般対象者(有病率を5.2%と仮定する)に適用すると、検査結果が陰性となった対象者の3%前後が偽陰性であると推定される。また、うつ病の有病率が一般対象者よりも高い精神科患者(有病率を27.1%と仮定する)に適用した場合を想定すると、16~18%程度の偽陰性が生じることが推定される。このような偽陰性が生じてしまう理由は、尺度の感度の低さに見て取ることができる。本研究のカットオフ値に基づく、CDIの感度は40.0%、DSRSでは46.7%、CES-

Dでは40.0%といずれも低い感度を示している。すなわち、実際にうつ病の診断基準を満たす子どもたちの半数以上が、自己評価式の尺度には低い得点しか示さないという実態を反映したものであると考えられる。

感度の低さは、これらの尺度をスクリーニングテストとして用いる際に大きな問題となる。良質なスクリーニングテストは感度と陽性的中率がいずれも高いことが要求される。しかしながら、CDI, DSRS, CES-Dといった自己評価尺度は、この2点を同時に満たすことが困難である。それぞれの尺度の感度を70%以上に維持しようとする、本研究のデータに基づくとそのカットオフ値はCDI 16点、DSRS 14点、CES-D 17点となる(いずれも感度73.3%)。この場合、各尺度の陽性的中率はCDIで11.6%、DSRSで16.2%、CES-Dで12.0%となり、どの尺度もかなり多くの偽陽性を生じてしまい、判別精度は大幅に低下することとなる。この点については、海外の先行研究における報告とも一致している。これらの尺度の原版においても、カットオフ値に基づく感度と陽性的中率は同時に十分な値を示さないことが報告されている(CDI:感度14.2%、陽性的中率13.8%、DSRS:感度66.7%、陽性的中率15.0%、CES-D:感度85.2%、陽性的中率10.3%、Birlleson et al., 1987; Matthey et al., 2002; Roberts et al., 1991)。したがって、感度と陽性的中率が両立しないことは、わが国の文化や本研究のサンプルに特有の問題というよりも、これらの尺度が抱える本質的な問題である可能性が高い。

このような問題が起きる理由としては、尺度の項目内容による影響が考えられる。たとえばCDIの項目を見ると、否定的なボディイメージや友人の欠如といった、うつ病の診断には直接関係のない項目が多く認められる。DSRSとCES-Dにおいても、腹痛や家庭内不和、対人関係上の問題といったうつ病の基準には含まれない項目が一定の割合で存在する。これらの項目が多数含まれることで、うつ病ではない子どもたちでもある程度得点が高くなり、うつ病の子

どもたちの得点が相対的に低められてしまっている可能性がある。

以上の点を考え合わせると, CDI, DSRS, CES-D の適正な使用方法について示唆を得ることができる。これらの自己評価尺度は, 基本的には本研究で算出されたカットオフ値に基づいて用いることが望ましい。ただし, カットオフ値に基づく検査結果は, 「陽性であればうつ病が強く疑われるが, 陰性であってもうつ病の可能性を除外できない」ものとして解釈されるべきである。これらの尺度は, いずれも一定の割合のうつ病の子どもを見逃す危険性をはらんだツールであることを理解して用いる必要がある。

また, 一般対象者におけるスクリーニングのような高い感度が求められる場面では, 従来のカットオフ値を用いることについても考慮する必要があるかもしれない。ただし, 感度を高めようとすれば陽性的中率は著しく低下するため, この方法はかなり多くの偽陽性が生じることが許容される場面に限定されるべきである。もしくは, 自己評価尺度に加えて, 親, 教師, 治療者といった他者評価による情報を複合的に用いることで, スクリーニングの感度を向上させることができる可能性がある。これらの方法についても, 実際のデータを踏まえて評価する意義があるといえる。

最後に, 本研究の限界と今後の課題について述べる。第一に, 本研究は単一の集団に対して尺度の検討がなされており, 同様の他の集団を対象とした交差妥当化の手続きはとられていない。本研究の結果のエビデンスとしての強固性を補強するためにも, 交差妥当化による再検証が求められる。第二に, 本研究では一般対象者のデータのみを用いている。傳田ら (2001) のデータを参照した検査後確率の推定はなされているが, 実際に精神科患者に対してこれらの尺度を適用した場合の判別精度についても確認されるべきである。中でも特に, 不安障害, 統合失調症, 行為障害といった他の精神疾患との弁別的妥当性について検討することは重要であると考えられる。第三に, 本研究では12~14歳の

年齢層を対象としている。他の年齢層に対する結果の一般化可能性を探る上でも, より幅広い年齢層を対象としたデータを収集し, 尺度の判別精度を検討することが必須であるといえる。

## 文 献

- Bastiaansen, D., Koot, H. M., Ferdinand, R. F. et al. (2004): Quality of life in children with psychiatric disorders: Self-, parent, and clinician report. *Journal of the American Academy of Child and Adolescent Psychiatry*, **43**, 221-230.
- Beck A. T., Ward, C, Mendelson, M. et al. (1961): An inventory for measuring depression. *Archives of General Psychiatry*, **4**, 53-63.
- Birleson, P., Hudson, I., Buchanan, D. G. et al. (1987): Clinical evaluation of a self-rating scale for depressive disorder in childhood (depression self-rating scale). *Journal of Child Psychology and Psychiatry*, **28**, 43-60.
- Charman, T. (1994): The stability of depressed mood in young adolescents: A school-based survey. *Journal of Affective Disorders*, **30**, 109-116.
- 傳田健三, 佐々木幸哉, 朝倉聡他 (2001): 児童・青年期の気分障害に関する臨床的研究. 児童青年精神医学とその近接領域, **42**, 277-302.
- 傳田健三, 賀古勇輝, 佐々木幸哉他 (2004): 小・中学生の抑うつ状態に関する調査: Birleson 自己記入式抑うつ評価尺度 (DSRS-C) を用いて. 児童青年精神医学とその近接領域, **45**, 424-436.
- Fagan, T. J. (1975): Nomogram for Bayes's theorem. *New England Journal of Medicine*, **293**, 257.
- Fundudis, T., Berney, T. P., Kolvin, I. et al. (1991): Reliability and validity of two self-rating scales in the assessment of childhood depression. *British Journal of Psychiatry, Supplement*, **11**, 36-40.
- 古川壽亮 (2000): エビデンス精神医療: EBP の基礎から臨床まで. 東京, 医学書院.
- 古川壽亮, 大野裕, 宇田英典他 (2003): 一般人口中の精神疾患の簡便なスクリーニングに関する研究 平成14年度厚生労働科学研究費補助金 (厚生労働科学特別研究事業) 心の健康問題と対策基盤の実態に関する研究 研究協力報告書.

- Furukawa, T., Hirai, T., Kitamura, T. et al. (1997): Application of the Center for Epidemiologic Studies Depression Scale among first-visit psychiatric patients: A new approach to improve its performance. *Journal of Affective Disorders*, **46**, 1-13.
- Kovacs, M. (1985): The Children's Depression Inventory (CDI). *Psychopharmacological Bulletin*, **21**, 995-998.
- Kovacs, M. (2003): *Children's Depression Inventory: Technical manual update*. New York, Multi-Health Systems.
- Kraemer, H. C. (1992): *Evaluating medical tests: Objective and quantitative guidelines*. California, Sage Publications.
- Lyneham, H. J., Abbott, M. J. & Rapee, R. M. (2007): Interrater reliability of the Anxiety Disorders Interview Schedule for DSM-IV: Child and parent version. *Journal of the American Academy of Child and Adolescent Psychiatry*, **46**, 731-736.
- 真志田直希, 尾形明子, 大園秀一他 (印刷中): CDI (Children's Depression Inventory) 日本語版作成の試み. 行動療法研究.
- Matthey, S. & Petrovski, P. (2002): The Child's Depression Inventory: Error in cutoff scores for screening purposes. *Psychological Assessment*, **14**, 146-149.
- 村田豊久, 小林隆児 (1988): 児童・思春期の抑うつ状態に関する臨床的研究 厚生省: 児童・思春期精神障害の成因に関する研究 昭和62年度研究報告書, pp. 69-81.
- 村田豊久, 清水亜紀, 森陽二郎他 (1996): 学校における子どものうつ病: Birleson の小児期うつ病スケールからの検討. 最新精神医学, **1**, 131-138.
- Myers, K. & Winters, N. C. (2002): Ten-year review of rating scales. II: Scales for internalizing disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, **41**, 634-659.
- Nolen-Hoeksema, S., Girgus, J. S. & Seligman, M. E. P. (1992): Predictors and consequences of childhood depressive symptoms: A 5-year longitudinal study. *Journal of Abnormal Psychology*, **101**, 405-422.
- Radloff, L. S. (1977): The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, **1**, 385-401.
- Ransohoff, D. F. & Feinstein, A. R. (1978): Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *New England Journal of Medicine*, **299**, 926-929.
- Rao, U., Ryan, N. D., Birmaher, B. et al. (1995): Unipolar depression in adolescents: Clinical outcome in adulthood. *Journal of the American Academy of Child and Adolescent Psychiatry*, **34**, 566-578.
- Roberts, R. E., Lewinsohn, P. M. & Seeley, J. R. (1991): Screening for adolescent depression: A comparison of depression scales. *Journal of the American Academy of Child and Adolescent Psychiatry*, **30**, 58-66.
- 佐藤寛, 永作稔, 上村佳代他 (2006): 一般児童における抑うつ症状の実態調査. 児童青年精神医学とその近接領域, **47**, 57-68.
- 佐藤寛, 下津咲絵, 石川信一 (2008): 一般中学生におけるうつ病の有病率: 半構造化面接法を用いた実態調査. 精神医学, **50**, 439-448.
- 島悟, 鹿野達男, 北村俊則他 (1985): 新しい抑うつ性自己評価尺度について. 精神医学, **27**, 717-723.
- Silverman, W. K. & Albano, A. M. (1996): *Anxiety Disorders Interview Schedule for DSM-IV-Child version*. Texas, Psychological Corporation.
- Silverman, W. K., Saavedra, L. M. & Pina, A. A. (2001): Test-retest reliability of anxiety symptoms and diagnoses with anxiety disorders interview schedule for DSM-IV: Child and parent versions. *Journal of the American Academy of Child and Adolescent Psychiatry*, **40**, 937-944.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285-1293.
- Zung, W. W. K. (1965): A self-rating depression scale. *Archives of General Psychiatry*, **12**, 63-70.

## CDI, DSRS, AND CES-D AS MEASURES OF DEPRESSION AMONG ADOLESCENTS IN JAPAN: ROC ANALYSES AND STRATUM-SPECIFIC LIKELIHOOD RATIO ANALYSIS

Hiroshi SATO, Shin-ichi ISHIKAWA

*Faculty of Education and Culture, University of Miyazaki*

Sakie SHIMOTSU

*Department of Social Welfare, Kyushu University of Health and Welfare*

Yoko SATO

*Faculty of Education and Culture, University of Miyazaki*

The three depression self rating scales, the Children's Depression Inventory (CDI), the Depression Self-Rating Scale for Children (DSRS), and the Center for Epidemiologic Studies Depression Scale (CES-D), are used to screen for depression in Japanese adolescents. The present study incorporating these three scales and a semi-structured interview for determining DSM-IV depressive disorders aimed to test the ability of the scales to identify depressive disorders in a community sample of junior high school students in Japan. The receiver operating characteristics (ROC) analyses and stratum-specific likelihood ratios (SSLRs) were applied to the data sets of 286 community adolescents aged 12 to 14 years old. The ROC analyses revealed moderate convergent validity of

these scales in detecting depressive disorders. The optimal cut-off points suggested by the ROC analyses were 31 for CDI, 24 for DSRS and 37 for CES-D, which were all higher than traditional cut-off points. Results of the SSLRs further demonstrated that these three scales were useful in screening for depressive disorders in Japanese community adolescents, applying the optimal cut-off points as noted.

Author's address:

H. Sato

Faculty of Education and Culture,  
University of Miyazaki

1-1 Gakuen Kibanadai Nishi,  
Miyazaki-Shi 889-2192, JAPAN