

## 資料

## 心理学専攻者による操作的診断基準の評定者間信頼度

加藤 知佳子\* 渡辺 さちや\*\* 永田 俊明\*\*\* 北村 俊則\*\*\*\*

RELIABILITY OF PSYCHIATRIC DIAGNOSIS : APPLICATION  
OF OPERATIONALIZED DIAGNOSTIC CRITERIA

Chikako KATO, Sachiya WATANABE, Toshiaki NAGATA AND Toshinori KITAMURA

Agreement of diagnosis of psychiatric disorders by psychologists was examined in case vignette and inter-rater designs. In the case vignette design, the agreement of 3 graduate psychology students for the Research Diagnostic Criteria (RDC) diagnosis was high for the most of the RDC diagnostic categories, except for Schizotypal Features. In the inter-rater design, disagreement was only observed among 2 out of 11 psychiatric in-patients. This suggested that an operationalized criteria could enable psychologists to establish diagnosis of psychiatric disorders reliably and therefore communicate and collaborate with psychiatric researchers.

Key words : diagnostic criteria, psychiatric disorder, reliability, Research Diagnostic Criteria.

## 問 題

異常心理現象の研究を再現性あるものとするためには、その対象となる被験者を同定するのに使用する診断名が評定を行う者によって変動するような事態は避けなければならない。しかし最近の研究では、いわゆる臨床診断の一致度は低く (Kitamura, Shima, Sakio & Kato, 1989), さらに日本における診断と米国における診断の不一致についても指摘されている (Kitamura, Shima, Sakio & Kato, 1986)。そのため、操作的な診断概念の定義が必要になるわけだが、どのような形の検査で診断に必要な情報が収集されるにせよ、情報収集方法の標準化が同時になされるべきである。面接の形で

得られる情報をもとに診断することを前提にするならば、質問項目および質問方法の標準化が行われなければならない。

現在までに、いくつかの操作的診断基準とそれに対応した面接基準が作成されているが、中でも有名なものの1つとして Spitzer らによって考案された研究用診断基準 Research Diagnostic Criteria (Spitzer, Endicott & Robins, 1981) (RDC) と、感情病および精神分裂病用面接基準 Schedule for Affective Disorders and Schizophrenia (SADS) (Spitzer & Endicott, 1979) のセットがある。日本においても既に同席面接法 (北村・島, 1986), 症例要旨法 (Kitamura, Shima, Sakio & Kato, 1986) (須賀・森田・伊藤・北村, 1987), 再試験法 (北村・須賀・森田・伊藤, 1987) によって評定者間信頼度検定が行われ、ほとんどの診断概念および精神症状について高い一致率が得られている。

これまでの研究においては、当然のことながら、異常心理現象の評定にはもっぱら精神科医があたっていたわけだが、診断基準についての知識と理解が期待されるのは精神医学の近接領域に属する心理学専攻者に

- \* お茶の水女子大学 (Ochanomizu University)
- \*\* 日本女子大学カウンセリングセンター (Japan Women's University Counseling Center)
- \*\*\* 木嶋学園医療福祉専門学校 (Kijima Gakuen Medical & Welfare Institute)
- \*\*\*\* 国立精神・神経センター精神保健研究所 (National Institute of Mental Health)

とつても同様である。被験者あるいは調査対象を選定するにあたって、医学領域との共通言語を有することは、研究結果の交換、それぞれの分野での応用を行う上で必要不可欠である。北村(1988)によれば、臨床経験のない社会学者、人類学者なども各々の分野で異常心理現象の研究を行うにあたっては、一定の水準の診断を行うために操作的診断基準を用いるようになっており、ことに感情障害は一般人口中に多い精神障害であるため要請が高い。

教育心理学においても、例えば母親のうつ病が子どもに与える影響、*anorexia nervosa* やその他児童思春期に見られる精神障害、さらに精神障害における知覚・認知など様々な研究領域において、異常心理現象の知識と技術が必要となろう。ところが、日本においては操作的診断基準の存在についてあまり知られていない。RDCの心理学専攻者による信頼度検定はわずかに藤原ら(藤原・北村・生田・宗像・千葉・菅原・仲尾, 1988)の症例要旨法による報告があるのみであった。

藤原ら(1988)の研究では、臨床心理士として臨床経験1~11年の者を評定者としており、経験年数によって2群に分けて検討されているが(3年未満, 3年以上), すべての診断において著しい差異はなかった。そこで、我々は、心理学専攻者として大学院生を選び、RDCの評定者間信頼度を症例要旨法と同席面接法をもって検討したので、ここに報告する。

## 方 法

心理学専攻の大学院生3名が評定者として参加した\*。

まず第1に、ニューヨーク州立精神医学研究所のEndicottらが作成した症例要旨集(全31症例)の日本語版を使用して、各評定者が独立にRDC診断を行った。

評定の一致度(信頼度)については、各評定者間の一致度と、各評定者とニューヨーク州立精神医学研究所の専門家による診断(以下正答とする)との一致度を、CohenのKappa( $\kappa$ )(Cohen, 1960)を使用して求めた。Kappa係数とは、偶然の一致率を除外できる方法として、近年診断の一致率の研究では広く使われている手法のひとつである(北村, 1988)。評定者が2名の場合の一致率に適しており、評価項目は2段階でも、それ以上でもよい。

測定された一致率を  $p_0$ 、偶然に起こりうる一致率を

\* 評定者3名の内訳は、大学院修士課程2年生、修士課程修了後研究生になった者、博士課程3年生であり、3名とも人格・臨床心理学を専門分野の1つとする者である。

$p_c$ とすると、kappa係数は次の式で求められる。

$$\kappa = \frac{p_0 - p_c}{1 - p_c}$$

もしも、 $p_0 < p_c$ ならば、偶然の一致よりも低い一致性を示し、 $\kappa < 0$ となる。 $p_0 > p_c$ ならば、偶然の一致よりも高い一致性を示し、 $\kappa > 0$ となる。 $p_0 = 1$ の場合は完全な一致を示し、 $\kappa = 1$ となる。

3名以上の評定者間の一致率を求める手法も存在するが(Fleiss, 1965)、本研究では、各評定者と正答間の一致率も必要なため、単純な3名以上の一致率計算法は使用せず、次のような方法を採用した。すなわち、評定者間の一致率は、すべての組み合わせで $\kappa$ を求め、それぞれ標準誤差を求め、統計的有意性を検討した。各評定者と正答間の一致率も同様に行った。なお、計算法の詳細については、Fleiss(佐久間訳, 1975)等を参照されたい。この際、Base rateが全症例の5%に満たない場合は $\kappa$ を求めなかった(Base rateとは、少なくとも1名の評定者が該当するとした症例の全症例に対する比率である)。

第2に、単科精神病院(丘の上病院)に入院後可及的早期の精神科患者11名を対象としSADSを使用した同席面接法でRDC診断を行った。3名の評定者が交代で2名1組になり、両評定者で病歴簿を通読した上で、2名のうち一方がSADSを用いて面接を行った。SADSは、RDC診断を行うのに必要な情報を得るために、質問内容、順序等を厳密に決めた“面接マニュアル”である。面接によって得られた細かい症状に関するチェックリストからは、RDC診断が自動的につけられるようになっているため、面接と診断を同一人物が行うことによるcarry-over effectsは、原則的に考慮する必要はない。もう一方の評定者は、面接の最後に必要があれば質問を加えた。面接終了後、各評定者は独立にRDC診断を行った。主治医による各患者の診断はいわゆる従来診断で行われており、前述の理由からこの従来診断とRDC診断を比較することは適当でないと考えた。

なお、3名の評定者は、今回の調査に先立ち、RDCの説明とビデオを使つてのRDC診断の練習、およびSADSを使用しての模擬面接を行った。

## 結 果

### (1)症例要旨法の場合

31症例中、欠損値があつた4症例を除いた27症例の現在挿話について、各診断概念ごとのBase rate、各評定者間の $\kappa$ 、各評定者と正答間の $\kappa$ を求め、それぞれの $\kappa$ について、標準誤差を求め統計的有意性を検討

した。TABLE 1 に示したのは、評定者間、評定者・正答間ともに、3通りの組み合わせによる  $\kappa$  の平均値であり、すべての組み合わせにおいて統計的に有意に高い一致度だったものを \* で示した。

TABLE 1 現在挿話の診断概念についての信頼度

診断概念	Base rate	(正答による比率)	評定者間の $\kappa$ (平均値)	評定者・正答間の $\kappa$ (平均値)
精神分裂病	0.22	(0.19)	0.67**	0.79**
分裂感情病 躁型	0.11	(0.11)	0.67**	0.85**
分裂感情病 抑うつ型	0.26	(0.15)	0.52 *	0.64**
躁病	0.07	(0.04)	0.70**	—
双極病 I	0.07	(0.07)	0.77**	0.88**
定型うつ病	0.30	(0.22)	0.79**	0.87**
準定型うつ病	0.07	(0.04)	0.77**	—
全般型不安症	0.07	(0.04)	0.77**	—
アルコール症	0.07	(0.07)	1.00**	1.00**
UFP+	0.15	(0.07)	0.52**	0.74**
OPD++	0.11	(0.15)	0.64**	0.63**
分裂型特徴	0.07	(0.11)	0.31	0.47**

・ 特定不能の機能性精神病    ・ その他の精神医学的障害

\* すべての  $\kappa$  で  $p < .05$     \*\* すべての  $\kappa$  で  $p < .01$

(注) RDC 診断は重複診断を許しているため、正答による比率の和は 1 を越える。

評定者間の信頼度は大部分の診断概念において高い値が得られた。やや低かったのは分裂感情病抑うつ型であり、分裂型特徴はかなり低かった。

評定者・正答間の信頼度はすべての診断概念において高い値が得られた。

次に RDC 診断に必要な精神症状を、躁症状群 8 項目、定型抑うつ症状群 12 項目、精神分裂病様症状群 3 項目に整理し、診断概念の場合と同様、それぞれの信頼度を求め、TABLE 2 に示した。

精神症状評価の評定者間信頼度は、おおむね高い値が得られたが、いくつかの精神症状についてかなり低い値が目立った。

躁症状群については注意散漫、判断力の貧困の信頼度が非常に低かった。定型抑うつ症状群の内因性定型うつ病に特徴的な症状の信頼度は全体的に低く、特に反応性の欠如、気分の質的变化が低かった。精神分裂病様症状群については思考形式の障害が低かった。

評定者・正答間の信頼度は、注意散漫、反応性の欠如を除いて、高い値が得られた。

(2) 同席面接法の場合

TABLE 2 精神症状についての信頼度

診断概念	Base rate	(正答による比率)	評定者間の $\kappa$ (平均値)	評定者・正答間の $\kappa$ (平均値)
<sup>1)</sup> 躁症状群				
基準 I 高揚気分	0.19	(0.15)	0.70 **	0.80 **
基準 II 活動的	0.15	(0.15)	0.53 *	0.70 **
多弁	0.19	(0.15)	0.67 **	0.85 **
思考促進	0.15	(0.07)	0.60 *	0.69 **
誇大性	0.19	(0.15)	0.67 **	0.78 **
睡眠量減少	0.11	(0.07)	0.85 **	0.85 **
注意散漫	0.11	(0.07)	0.30	0.46
判断力の貧困	0.11	(0.04)	0.26	—
<sup>2)</sup> 定型抑うつ症状群				
基準 I 抑うつ気分	0.50	(0.52)	0.82 **	0.80 **
基準 II 食欲不振	0.37	(0.37)	0.83 **	0.89 **
不眠	0.50	(0.50)	0.85 **	0.90 **
精力低下	0.30	(0.26)	0.72 **	0.67 **
興味の喪失	0.30	(0.33)	0.87 **	0.85 **
自責感	0.26	(0.26)	0.56 *	0.63 **
集中力低下	0.41	(0.41)	0.95 **	0.90 **
自殺傾向	0.37	(0.41)	0.77 **	0.79 **
激越・制止	0.37	(0.37)	0.70 **	0.80 **
内因性定型うつ病に特徴的な症状				
広範囲の興味の喪失	0.22	(0.26)	0.43 *	0.56 **
反応性の欠如	0.15	(0.37)	0.23	0.11
気分の質的变化	0.15	(0.00)	0.12	—
<sup>3)</sup> 精神分裂病様症状群				
妄想	0.41	(0.37)	0.65 **	0.75 **
幻覚	0.37	(0.33)	0.69 **	0.76 **
思考形式の障害	0.30	(0.11)	0.38	0.55 **

\* すべての  $\kappa$  で  $p < .05$     \*\* すべての  $\kappa$  で  $p < .01$

現在挿話についての両評定者の RDC 診断は 11 名の被験者のうち 9 名 (82%) について一致していた。それぞれの被験者の性別、年齢と各評定者による診断は TABLE 3 の通りである。

診断が一致しなかった 2 症例について、その原因を簡単にまとめると TABLE 4 のようになる。

患者 6 は、軽い不快気分と定型抑うつ症状群に特徴的な症状を呈していたが、臨床症状の前景となっている不快気分の解釈が TABLE 4 に示したようにくい違ったため、診断も一致しなかった。

患者 8 は躁的な時期に普段より過活動の傾向があり、請けおすぎた仕事をかたづけするため早朝から出勤するということがあった。これを過活動のため仕事上の支障をきたしたと解釈したのが面接者で、いつもより働かねばならなかったにせよ支障をきたしたわけではないとしたのが観察者である。ただ、この解釈の違いは診断上は躁病の重篤度の違いに納まった (双極病 II の方が軽躁)。

TABLE 3 各患者の面接者、観察者による診断

患者番号	性別	年齢	面接者の診断	観察者の診断
1	♀	55	精神分裂病	精神分裂病
2	♂	45	アルコール症 薬物常用症	アルコール症 薬物常用症
3	♂	39	定型うつ病	定型うつ病
4	♂	25	定型うつ病	定型うつ病
5	♀	31	分裂感情病躁型	分裂感情病躁型
6	♂	30	*全般的な不安型	定型うつ病
7	♂	39	定型うつ病	定型うつ病
8	♂	30	*双極病 I 強迫症	双極病 II 強迫症
9	♂	30	定型うつ病 OPD(非定型人格障害)	定型うつ病 OPD(非定型人格障害)
10	♀	24	分裂感情病抑うつ型 OPD(大食症) 分裂型特徴(疑診)	分裂感情病抑うつ型 OPD(大食症) 分裂型特徴(疑診)
11	♀	27	定型うつ病	定型うつ病

\* 診断が一致しなかった症例

TABLE 4 診断不一致の原因となった臨床症状と各評定者の解釈

患者番号	臨床症状	面接者の解釈	観察者の解釈
6	軽い不快気分	不安気分	抑うつ気分
8	躁エピソードにおける過活動の結果	機能障害あり	機能障害なし

よって、大きく診断が分かれたのは患者6のみであると言えよう。

## 考 察

操作的診断基準の信頼度を検討する場合、基準そのものが内包している問題点以外に、評定者側の要因がかなり影響を与える。それは主に、評定者がどれくらい診断概念とそれを支える精神症状とを理解しているか、ということにかかっている。

今回調査に参加した心理学専攻者は、精神医学に関する知識が全くなかったわけではないが、診断概念や診断に至るまでの手順に関しては今回使用したRDCによって学んだ。従って、RDC以外の基準で診断を下すことがないため、操作的診断基準の使い手としては、むしろ理想的だったはずである。

実際、症例要旨法で行った信頼度検定の結果でも、精神科医が行ったもの(須賀他, 1987)に劣らない一致率が得られた。藤原ら(1988)による先行研究もほぼ同様

の所見を示している。同席面接法による診断では、不一致例は1名のみと言ってよい。

しかし、一部の診断概念および精神症状には、理解が難しいもの、評価が難しいものがあった。例えば、症例要旨法の結果において、分裂感情病抑うつ型、分裂型特徴の一致率が高くなかったのは、精神病像の理解が十分でなかったことが関連している。具体例を挙げると、3名の評定者のうち少なくとも1名が分裂感情病抑うつ型と診断した症例は7例あったが、そのうちの3例は正答によると定型うつ病であった。その3例中2例は、思考形式の障害をまちがって「あり」としたものである。もう1例は、正答では分裂型特徴を重複診断しているもので、はっきり幻覚とは言い難い症状を基準に合った幻覚としてしまったものである。

RDCにおいて、思考形式の障害は、どのように障害されている場合「あり」とするかが厳密に決められている。また、妄想、幻覚も、特に感情障害との関係において細かい弁別が要求されている。そのため実際の評定にあたってはことばの上での定義を覚えるだけでは難しく、そのような定義、弁別が要求されている背景や、典型例とそうでない例との違いについて、具体的な学習が必要のように思われた。

ちなみに、分裂感情病抑うつ型の信頼度は精神科医においてもあまり高くない(須賀他, 1987)が、一致しなかった内訳が心理学専攻者の場合とは異なる。精神科医の場合、正答によると分裂感情病抑うつ型の症例に対し精神分裂病と診断する傾向が見られた。これは分裂病様症状の存在を確認した時点で診断を下してしまう日本の従来診断の方法に慣れてしまっていることによると思われるが、RDCによってのみ診断手順を学習した今回の評定者には、このような傾向は見出せなかった。

躁症状群に属する活動的、注意散漫、判断力の貧困などの症状は、それぞれの典型例が重症度に関する評定の基準にされるという一面があり、今回の評定者は精神科医に比べて具体例に関する知識が少ないことが、低い一致率の一因になっていると考えられる。しかし、これらの症状は、基準IIの下位項目であることから、躁病関係の診断名の一致率に直接影響を与えてはいない。

なお、分裂型特徴や内因性定型うつ病に特徴的な症状については、精神科医が行った研究においても一致率が低く(須賀他, 1987)、理解が難しいものとされている。今回の調査でもこれらの一致率は低く、特に後者においてそれがより顕著に現れた。

異常心理学や臨床心理学の領域における研究・臨床を行うに際して、精神障害の診断技術の確立は、心理学専攻者が近接領域である精神医学の研究者との「共通言語」を用いるためには不可欠であり、今後ともその重要性は増加するであろう。本研究はRDCのような操作的診断基準がこれを可能にすることを示唆している。

## 結 論

心理学専攻者が評定者になり、症例要旨法および同席面接法によってRDC診断の信頼度検定を行った。

その結果、精神症状の具体例に関する知識不足がハンディにはなったものの、おおむね高い一致率が得られた。精神科医に比べて一致率の低かった分裂病様症状やいくつかの躁症状については、具体例に基づいて弁別の基準を学習することによって、より一致率が向上すると思われる。

従って、学習に際して多少の工夫が必要ではあるが、心理学専攻者によっても操作的診断基準は使用可能であると言える。精神医学領域とコミュニカティブな研究を提供するために、心理学においても、幅広く使用されることが期待される。

## 引用文献

- Cohen, J. 1960 A coefficient of agreement of nominal scales. *Educational Psychological Measurement*, **20**, 37—46.
- Fleiss, J.L. 1965 Estimating the accuracy of dichotomous judgments. *Psychometrika*, **30**, 469—479.
- フライス, J.L. 佐久間昭(訳) 1975 計数データの統計学 初版 東京大学出版会  
(Fleiss, J.L. 1973 *Statistical methods for rates and proportions*. New York: John Wiley & Sons.)
- 藤原茂樹・北村俊則・生田憲正・宗像恒次・千葉浩彦・菅原健介・仲尾唯治 1988 臨床心理士と精神科医による症例要旨法を用いたRDC精神科診断の一致度の比較 *臨床精神医学*, **17**, 1081—1087.
- 北村俊則 1988 感情障害の診断基準 現代精神医学大系 年間版 '88—B, 53—72. 中山書店
- 北村俊則・島悟 1986 感情病および精神分裂病用面接基準(SADS)と研究用診断基準(RDC)の評

定者間信頼度 *精神医学*, **28**, 41—45.

Kitamura, T., Shima, S., Sakio, E. & Kato, M. 1986 Application of Research Diagnostic Criteria and International Classification of Diseases to case vignettes. *Journal of Clinical Psychiatry*, **47**, 78—80.

Kitamura, T., Shima, S., Sakio, E. & Kato, M. 1989 Psychiatric diagnosis in Japan. II. Reliability of conventional diagnosis and discrepancies with Research Diagnostic Criteria diagnosis. *Psychopathology*, **22**, 250—259.

北村俊則・須賀良一・森田昌宏・伊藤順一郎 1987 操作的診断基準の信頼性とその問題点——III. 再試験法による研究用診断基準(RDC), ハミルトンうつ病評価尺度, 陰性症状評価尺度の信頼度検定 *精神医学*, **29**, 579—585.

Spitzer, R.L. & Endicott, J. 1979 *Schedule for Affective Disorders and Schizophrenia (SADS)*. 3rd ed. New York: Biometrics Research Department, New York State Psychiatric Institute.

(保崎秀夫(監訳) 北村俊則・加藤元一郎・崎尾英子・島悟・高橋龍太郎(訳) 1983 感情病および精神分裂病用面接基準 星和書店)

Spitzer, R.L., Endicott, J. & Robins, E. 1981 *Research Diagnostic Criteria (RDC) for a selected group of functional disorders*. 3rd ed. New York: Biometrics Research Department, New York State Psychiatric Institute.

(本多裕・岡崎祐士(監訳) 安西信雄・平松謙一・亀山知道・池淵恵美・増井寛治・宮内勝(訳) 1982 精神医学研究用診断マニュアル 国際医書出版)

須賀良一・森田昌宏・伊藤順一郎・北村俊則 1987 操作的診断基準の信頼性とその問題点——I. 症例要旨法による研究用診断基準(RDC)の評定者間信頼度検定 *精神医学*, **29**, 373—378.

## 付 記

本研究を行うにあたって多大なご援助をいただきました丘の上病院院長延島信也先生ならびに同病院の医局・看護科・ケースワーク室の皆様へ深謝いたします。  
(1989年10月2日受稿)