

# 精神疾患診断の問題点と操作診断の必要性

北村 俊則

国立精神・神経センター

精神保健研究所

社会精神保健部

精神科診断学, 11(2); 191-218, 2000.

## I. はじめに

精神疾患の操作的診断基準 **operational diagnostic criteria** が開発されてから 20 年の時間が経過し、日本においても広く知られるようになってきた。しかし、研究目的以外に、操作的診断基準を臨床現場に導入することには懐疑的・否定的意見も少なくない。こうした意見にはいくつかの根拠があるだろうが、その多くのものは診断という行為そのものを拒否するのではなく、「従来診断 **conventional diagnosis**」や彼らの言うところの「臨床診断 **clinical diagnosis**」を重視し、操作的診断を硬直的であり臨床の用に立たないと批判するものと考えられる。1997 年 5 月 30 日に東京国際フォーラムで行われた第 93 回日本精神神経学会総会でのディベート「精神医学の対立点〈操作診断の功罪〉」にて発言する機会を得たが、このときの議論をもとに、本稿では「従来診断」や「臨床診断」の位置付けを再考し、操作診断の必要性について検討する。

## II. 従来診断の問題点

### 1. 精神疾患の診断一致度

まず始めに検討しなければいけないことは、精神科における病名診断が、医師間でどれほど一致するかということである。Kraus (1996) は、「直感が、経験のほぼ土台である診察者の感受性や理論的な態度にある程度依存しているにもかかわらず、明らかに精神科医の間には、分裂病性の幻覚や被影響体験の特質、うつ病性の気分変調の特性について、大きな一致がある」と主張しているが、特別な判断の一致率の数値を示しているわけではない。古典的精神病理学の流れをくみ、操作診断に反対するこれらの論者に共通することは、診断の一致率のデータに触れることなく、強い確信を持って、(少なくとも経験ある)精神科医の臨床判断は驚くほど一致すると述べていることである。

操作的診断基準が導入される前は、精神科の病名診断の統一性が存在していなかった。過去には、英国と米国の精神疾患の臨床統計値が異なることから両国の共同研究が実施され、個別の精神疾患の範囲が異なることが原因であると指摘された (Kendell, 1971)。世界保健機構が主催して行われた国際精神分裂病パイロット研究 **International Pilot Study of Schizophrenia (IPSS)** では、ロンドンセンターの診断に比べモスクワセンターとワシントンセンターのうつ病診断の範囲は狭く、その分、精神分裂病の範囲が広いことが示された (Leff, 1977)。また、Saugstad ら (1983) は 1970 年に ICD-8 が導入される前後の欧米各国の精神科入院統計を検討し、各国の疾患比率の差が変化しておらず、米国では精神分裂病が多く、英国では躁うつ病が多数を占めていることを報告している。こうした諸研究は、世界の各国の精神科診断がそれぞれ独自の方法を用いており、使用している病名は同一でも、かならずしも統一された病態を示すものではなかったことを示唆している。

ニューヨーク州立精神医学研究所 New York State Psychiatric Institute で作成された研究用診断基準 Research Diagnostic Criteria (RDC) (Spitzer ら, 1978a) による症例要旨集を、その正答を見ないで 5 名の日本の精神科医が診断し、その最終合意診断を RDC 診断と比較したところ、米国に比べ、日本の精神分裂病は RDC の分裂感情病や一部の定型うつ病まで含むかなり広い病態を含むことが認められた (Kitamura ら, 1986) (表 1)。

#### 表 1 を挿入

同じ国の中でも精神科医による診断は一致しないことが認められている。20 名の日本人精神科医に 29 例の症例要旨 case vignette を提示し、従来診断による診断を求めた Kitamura ら (1989b) は、診断の一致率 (intraclass correlation coefficient (ICC) で求めた) が全般的に低いことを報告している (表 2)。ICC は 0.8 以上であればほぼ完璧といえ、0.6 ~ 0.8 なら十分であり、0.4~0.6 では中程度であり、0.4 を下れば問題がなしとは言えなくなる (Landis ら, 1977)。ICC が 0 というのは、診断する者同士の判断が同じであっても単なる偶然の一致であることを意味する。この調査で最も高い一致率を得られた病名は精神分裂病の 0.75 と強迫神経症の 0.73 であり、他の病名はすべて 0.70 を下回っていた。ことに境界例の 0.12、躁病の 0.22 は実務上まったく使用に耐えないことを示していた。また、非疾患 (正常) の判断も十分に一致しているのではなく、ICC がわずかに 0.52 であった。このことは、どこまでを精神疾患としどこからは正常とするかについても、日本の精神科医の判断に差異があることを意味している。つまり、ある個人は医師 A の診察では (なんらかの) 精神疾患ありとされ、医師 B の診察では正常 (精神疾患は存在しない) とされる可能性が低くないのである。

#### 表 2 を挿入

Kitamura ら (1989b) の研究に参加した精神科医は某大学医学部精神神経科の同窓生より無作為に選ばれているから、卒前卒後の教育バイアスも考慮しなければならない。しかし、この参加者の中には、日本の精神医学において指導的立場に立っている者も少なからず、含まれており、必ずしもこのことが日本全体の状況から大きく偏寄しているとは考えられない。

ICD-10 の草案の日本における実地試行において約 600 名の精神科医が参加した多施設共同研究 (Kitamura ら, 1994) において、ICD-10 診断に加えいわゆる従来診断を記載するよう求めた研究でも、同席面接法で非定型精神病、抑うつ神経症、器質性精神疾患で診断一致率 ( $\kappa$  で計算) が 0.7 を下回り、再試験法では、これに加えてうつ病、躁病、不安神経症、心気症の  $\kappa$  が 0.7 に達しなかった (表 3)。上記の Kitamura ら (1989b) は症例要旨を精神科医に読ませたものであり、直接の面接を行ったものではない。しかし Kitamura ら

(1994) は、各施設における外来および入院の新患を順に直接面接したものである。病名によって診断一致率に変化があるが、いずれも従来診断では十分高い診断一致率が得られないことは指摘できる。

### 表 3 を挿入

この Kitamura ら (1989b) の研究は、1980 年代の前半に、Kitamura ら (1994) の研究は 1980 年代の後半に実施されており、DSM-III が発表された 1980 年からあまり時間も経過していない時点でのものである。参加した精神科医も DSM-III の詳細については熟知していたとは思えない。同様の調査を現時点で行えば、恐らく異なった所見が得られるであろう。しかし、操作診断を拒否する従来診断派については上記の所見はなお現在も該当する可能性が高いと思われる。

次に、なぜ臨床精神科医の従来診断は一致しないのかについて考察しよう。これには少なくとも 2 つの理由がある。

第 1 に、使用する病名とその数が異なる。例えば非定型精神病という病名は、いかなる症例にもこれを用いないとする精神科医が少なからず存在する (Kitamura ら, 1989a)。仮に、従来の考え方という内因性精神病という概念を認め、この分類を考えてみよう。ある精神科医は、内因性精神病を精神分裂病と躁うつ病の 2 群に分け、またある精神科医はこれに非定型精神病を加え 3 群に分けたとする。こうした状況で両者の診断名は一致しなくても驚くことではない。上記 Kitamura ら (1989b) の調査では、成人に見られる機能的な精神疾患（器質性疾患と物質性疾患を除く）の病名で ICD-9 に掲載されている病名すべてと、病院・診断所のカルテ調査から抽出された病名を一覧表にし、これを 20 人の精神科医に提示し、「日常的に使用する」「日常的に使用はしないがそうした症例に遭遇すれば使用する」「これまでに使用したこともないし、これからも使用しない」のいずれかの選択肢で、各病名の使用状況を確認した。その結果、その病名は「決して使用しない」とする精神科医の率は、分裂病潜伏型 65%、分裂病残遺型 65%、急性精神分裂病 60%、分裂感情病 55%、妄想性人格障害 50%、情動性人格障害 50%、反応性錯乱 45%、分裂病性反応 45%、分裂病型人格障害 45%、反社会性人格障害 45%などが高く、ほぼ半数の精神科医が「使用しない」と回答している。ICD-9 がその時点の公的分類基準であるにもかかわらず、多くの精神科医はこれに従っていないのである。

すべての精神科医が使用する病名をまとめた Kitamura ら (1989a) は、日本の精神科医が共通して用いている機能的な精神疾患 functional psychiatric disorders の従来診断名は図 1 のようになると報告している。つまり、日本の精神科医が従来診断に共通して用いている病名は 20 に満たず、この表に掲載されていない病名を使用した場合には、診断が一致しなくなる可能性が発生するのである。

## 図 1 を挿入

先に述べた「診断は一致するもの」という確信が多くの日本の精神科医に支持されていたひとつの理由は、こうした基礎分類法においてコンセンサスが得られていたからで、このコンセンサスがあれば診断の評定者間信頼度は高いと誤解されたのであろう。

第 2 に、同一病名を使用してもそれが含む臨床像の範囲に個人差がある。前述の Kitamura ら (1989a) の調査では、20 名の精神科医に様々な種類の症例要旨を提示し、これに従来診断を付けさせた。この症例はニューヨーク州立精神医学研究所で RDC に従った診断を下した症例である。この 29 症例に対して、日本の精神科医が精神分裂病の診断を下した症例の範囲は個人差が著しかった。ある精神科医は RDC で慢性経過を取る精神分裂病に限定して従来診断の精神分裂病の診断を下していたが、一方、別の精神科医は、RDC の慢性精神分裂病、急性精神分裂病、分裂感情病、躁病の症例に従来診断の精神分裂病の診断を下していた。精神分裂病の範囲を広く取る精神科医は、その分、他の疾患の範囲が狭くなる。また、精神分裂病の範囲を狭く取る精神科医は、他の疾患の範囲が広がるか、別の病名を用いて境界の領域の病態を記述しようとするのであろう。この所見は当然、参加精神科医の各種属性の影響を受けるものであろうから、これで全般化した議論をするには慎重でなければならない。しかし先に述べたように、この調査に参加した医師達は一教室の同窓生から無作為に抽出しており、少なくともその大学の卒業生の特徴を示していると思われる。

いずれにせよ、「診断は一致するもの」という日本の精神科領域での誤解が生まれたもうひとつの理由は、このように典型例では診断が一致し、辺縁例では個人差が出ることから、基本的には診断がかわらないと思ってしまったのであろう。

1980 年代の Kitamura ら (1989a; 1989b) の調査をまとめると、(1) 日本における精神科医の従来診断の一致率は低い (2) 日本の精神科医の用いている診断名は共通したものではなく、個人差がある (3) 同じ病名を用いても、それが包含する臨床病態の範囲は個人によって大きく異なるといえる。精神科医の持つ「常識」が「多くの場合、ほぼ正確な判断を私たちに提供してくれる」との主張の根拠は失われているのである。

ではどうすれば良いのか。精神科医の間の診断一致率を向上させるには (1) 共通の診断分類を採用し、(2) それぞれの診断名の包含する臨床病態の範囲に明文の規定を与えれば良いのである。これが操作診断の出発点である。

## 2. 「勘」と「経験」と「権威」

一部の精神科医は、診断が一致する必要はないかのような発言をしている。「先輩の精神科医の診断を盗め」と新人を指導している卒後研修指導者も少なからずいるであろう。この場合、先輩医師の診断が「正しい診断」であり、これになるべく近付くことが新人には

期待されている。先輩医師と新人医師の診断に差異があっても、それは問題ではなく、新人が克服すべきハードルなのである。従って、同一の患者の病態について複数の医師の診断に差異があっても当たり前であり、改善を要求されることではないと考えるのである。

精神科に限らず、日本の医療において診断は、「勘」と「経験」と「権威」によって行われていた。「見立て」の良い医師は「勘」を働かせ、多少とも直観的な判断で診断を行っていた。臨床判断に困った場合は、長年積み重ねてきた様々な症例の「経験」から判断を行ってきた。こうした「勘」と「経験」は客観的に比較対比することは難しいので、おのずとそれになんらかの「権威」を与えることで医療と医学の領域の「政治的」安定を計ってきたのであろう。

しかし、こうした土壌で実証主義的臨床態度(古川ら, 1992)は育たない。医師に只々すがり、頼ることで医療を受けてきた時代の患者には「勘」と「経験」と「権威」による診断で事足りたのであろう。だが、現在の患者が医師に求めているものは、あてずっぽう的な「勘」でも、単に長年その領域で仕事をしているというだけの「経験」でも、職名やタイトルで表されるだけの「権威」でもない。患者自身の健康回復に役に立つ情報を有している医師である。日本においても、一般国民が医療における患者の自己決定権を尊重し、情報の開示を希望することは、これまで考えられてきた以上に強いものになってきている(Hayashi ら, 2000; Hasui ら, 2000)。そもそも、「勘」はその評価者個人の中でも時によって変動するであろう。「経験」と言っても自分が診察した患者をすべて記憶している医師は少ない。記憶に残っている症例のみから判断することは危険であろう。また、一人の医師が経験した症例は数も少なく、施設の特性等からバイアスもかかっている。

先輩医師の診断が正しいとすれば、それはその先輩が「勘」と「経験」と「権威」をもっているからではなく、単純に正しいから正しいのである。新人医師の診断も正しければ正しいのであり、「先輩医師の診断を盗め」といった家元制度のような発想は撤廃すべきであろう。もし、今後も「勘」と「経験」と「権威」による診断が日本の主流を占めるのであれば、真に患者が必要とする医学・医療情報を蓄積できる精神医学・精神科医療は死に絶える。

では、こうした従来診断を用いるということとはどのような問題を有しているのであろうか。次のいくつかの章では、このことが抱えている諸問題を考察する。

### Ⅲ. 診断の持つ社会的責任

#### 1. インフォームド・コンセントと病名告知

精神疾患の病名診断は操作化すべきか否かのこれまでの議論でひとつ欠けていた視点は、病名は患者に知らせるものであるという側面である。昔の医療体制においては、病名は患者に知らせるものではなかった。患者の知る権利を守るインフォームド・コンセント(告

知同意)の理念が導入されるに従って、病名告知が行われることが多くなってきた。インフォームド・コンセントは、患者が自らの価値基準に従って医療の内容とその開始・停止について自己決定権を行使するものであり (Weisbard, 1986)、患者が判断を行うためには十分な関連医療・医学情報が提供されなければならない (Appelbaum ら, 1987; Grisso, 1986)。この医療・医学情報の中には病名が含まれる。医療関係者の中には、病名までは告知する必要はないとする意見もある。しかし、提案された治療法から期待できる効果を新たに理解するには、その治療法の適用が開示されねばならず、患者の状態がその適用に合致しているか否かについての判断には病名が開示されなければならない。例えばガンの病名を告知されていない患者が、どうして抗癌剤投与の利益を理解できるのでしょうか。このことは精神疾患についても同様である (Miller, 1994)。

情報を受ける患者が考えることは、(1) 自分の病名は何か (2) その病気の予後はいかがか (3) どのような治療法が存在するのか (4) 各治療法の長短は何か (5) 各治療法の治療成績はいかほどか等である (Hasui ら, 2000)。こうした情報の基礎には病名の開示とその内容の説明が不可欠である。病名告知を受ける患者に「この病名は他の医師は使わない」「精神疾患の病名分類は医師によって異なる」「この病名は医師であれば皆使うが、あなたの病名は違う医師が診れば違う病名をつける」といった説明をして、一体どのような理解を患者はすればよいのか。使用する診断名が違うのに予後の比較をどうやって行うのか。「自分が診た患者では…」と説明しても、その医師が一体何名の症例を何年間追跡調査したのであるか。医師が「自分の治療経験」から最も良いと考えた治療方針を提示しても、その治療経験の評価自体にバイアスがかかっている。

医師によって診断結果は異なると聞かされた患者は、自分の目の前にいる医者は藪医者ではないかと思うであろう。そもそも使用する診断名も医師によって異なると聞かされた患者は、精神医学そのものが時代遅れの学問だと思ってしまうであろう。患者の予後判定や治療法選択基準の重要な部分は病名診断であることは一般国民の広く認知しているところであるから、この病名診断が医師によって異なり、使用する病名分類も医師によって違うと言われれば、患者の立場では精神科医と精神医学に懐疑的にならざるを得ない。患者の立場からすれば、「正しい」診断を受けたいというのは当然の希望である。2名の医師が異なる診断を下してもいずれも正しいといわれては、患者は当惑するばかりであろう。身体疾患で置き換えてみれば、医師 A は胃癌の診断を下し、医師 B は胃潰瘍だという診断をして、なお A も B も正しいという状況は、常識的には理解の限度を超えている。2名の医師の診断が異なれば少なくとも 1名の診断は誤診である (両名の診断とも誤診の可能性ももちろんある)。病名告知を含めたインフォームド・コンセントの実行が医師に課せられた責任であるとすれば、診断分類の不統一と診断の不一致は社会に対する責任を放棄していることになる。こうやって自らと自らが所属する業界の社会的認知を抹殺する精神科医はまさに時代遅れである。

## 2. 強制入院と病名診断

精神科医療では医療保護入院と措置入院という 2 種類の強制入院制度が日常的に用いられている。強制入院制度は、その目的がもっぱら当該患者の医療と保護にあるとは言うものの、個人の自由の一時的剥奪であることから民事収容 *civil commitment* (*Harvard Law Review*, 1974) とも称されている。日本においても精神保健福祉法で強制入院させられる患者の人権を守る多くの法的措置が取られているのも、自由の剥奪という事柄の重大性が認識されているからと考えられる。他の身体疾患であれば（一部感染症などの例外的疾患を除けば）、本人の同意なくして強制的に入院させることはあり得ない。たとえ放置すれば重大な障害を残したり、場合によっては死に至る可能性が高くとも、（意識消失の状態や救急医療の対象を除けば）本人が治療や入院を拒否すれば、医療者が入院を強行することはないし、またできない。ところが、精神保健福祉法では「精神障害」でありかつ自傷他害のおそれがあれば措置入院の対象となり、あるいは医療と保護の必要があれば医療保護入院の対象となり、患者本人が同意せずとも（一定の手続きを踏めば）入院させることができる。したがって、いったんなんらかの精神疾患の診断が付けば、強制入院の対象となり得るのである。一方、精神疾患でなければ自傷他害の疑いがあっても強制入院はさせられない。

強制入院は、たとえ本人の健康のためとはいえ自由の剥奪を個人に迫るのであるから、その手続きと内容は明瞭かつ公平なものでなければならない。「明瞭である」とは、強制入院が実施される理由がだれの目から見ても分かりやすいということであり、それが一部の専門家にのみ理解できるものではないことである。「公平である」とは、事前の条件が揃っているのであれば強制入院が実施されるか否かが状況の特徴や担当者の恣意性に左右されないことである。こうした強制医療の手続きの公平性は法的に要請されるものである。しかし、最近では入院させられる患者の側が強制を感じるかいなかに注目した研究が行われている。精神科患者が自分の入院に強制が働いたと感じるかは、それが法的に自由入院だったか強制入院だったかと一致するものではない。文献の概観 (Monahan ら, 1995) と少数例の予備研究 (Hoge ら, 1993) をした後で、Lidz ら (1995) は 157 名の患者が入院の直後 (48 時間以内) にどれほど強制されたと感じたかを *MacArthur Perceived Coercion* 尺度 (Gardner ら, 1993) で評価した。この尺度は、入院についての他からの影響、自己のコントロールの可能性、自己の選択があったか、自己の行為に自由があるかについてその感じ方を問うものである。Lidz ら (1995) の研究は、この *MacArthur Perceived Coercion* 尺度の得点を規定する要素を検討するものであった。説明変数は、施設、診断、入院形態などのほかに、手続き的正義尺度を組み込んだ。この尺度は、患者から見て、相手の動機が患者のことを配慮したものか、相手は患者を尊重したのか、相手は患者の発言を真剣に考慮したのか、相手は患者を公平に扱ったかの 4 点を採点するものである。*MacArthur Perceived Coercion* 尺度の得点には、他の指標を統制した後もなお手続き的正義尺度が有意の関与を示していた。

つまり、入院形態が強制入院であっても、医療スタッフの患者に対する態度が公平で患者の発言に耳を貸すものであれば、患者はさほど強制を感じないということである。一方、入院形態が自由入院であっても、医療スタッフが患者の言うことに耳も貸さず質問にも答えないのであれば、患者は自分の入院に強制感を持つのである。

強制入院の手続き的公平性でまず要請される点は、なぜその患者に強制医療が実施されるかの説明であり、この説明には病名の説明が含まれる。ことに精神保健福祉法でいう精神障害に該当するかについて、医師は明瞭な説明を与えなければならない。ところが、すでに見たように日本の精神科医に従来診断により症例要旨の診断を付けさせた際の「非疾患（正常）」の評定者間一致率は 0.52 である。このことは、ある患者が精神保健指定医の診察によって強制入院に服される時に、「精神疾患か正常か」の診断は医師によってずいぶん異なることを意味する。入院を拒否する患者にとって、自分が強制的に病院に入れられたのは、当たった医者が悪かったということになる。これでは、明瞭で公平であるとはいいがたい。

また、自傷他害のおそれがあっても精神疾患（精神保健福祉法でいう精神障害）でなければ医師が患者を強制的に病棟に収容することはできない。この場合、患者が入院を拒否すれば、患者の自由に任せるしかない。その結果、予測したとおりの自傷行為や他害行為があったとしよう。そして、患者の家族や他害行為の被害者が医師の注意義務違反を理由に賠償を訴えたならば、医師はどのように対応するであろうか。担当医は、患者が精神疾患に罹患していなかったことを理由に抗弁するであろう。しかし、「精神疾患か正常か」の信頼度が ICC にして 0.6 を下回るという事実は、他の医師であればなんらかの精神疾患の診断をつけていた可能性が低くないことを強く示唆している。信頼性の低い従来診断を用いた精神科医は、その低信頼度ゆえに訴訟を失うことにもなるのであろう。

ところで、なぜ精神疾患を有する者のみが強制入院に服さしめられるのかについては法的議論があるところである（例えば北村, 1995; Kitamura ら, 1998）。一般的な考え方は自傷のおそれがある場合は国親機能 *parens patriae* で、他害のおそれがある場合は警察権能 *police power* で、国家が患者の収容を行うというものであろう（Brakel, 1985）。また医療と保護の必要がある場合の医療保護入院については、患者本人が治療の必要性を合理的に判断できる能力 *competency* がないことがその根拠となる。欧米においては判断能力のある場合は精神科患者についても治療拒否権を認める動きがあり、判断能力の臨床評価は重要な課題になっている。

精神科における治療同意判断能力の評価方法についてはいくつかの手法が海外で開発されている。例えば、Appelbaum ら (1981) の Competency Questionnaire、Roth ら (1982) の Two-part Consent Form、Weithorn ら (1982) の Measure of Competency to Render Informed Treatment Decision、Belter ら (1984) の Recognition of Rights Violating Counselling、Grisso ら (1992) の Manual of Understanding Treatment Disclosure、Janofsky ら (1991) の Hopkins Competency Assessment Test、Bean ら (1994) の Competency Interview Schedule などがそれ

である。日本においても Kitamura ら (1993) が臨床場面で容易に使用可能な Structured Interview for Competency Incompetency Assessment Testing and Ranking Inventory (SICIATRI) を作成し、信頼性と妥当性を Tomoda ら (1997) が報告している。従来こうした判断は臨床医の裁量に任されてきた。しかし臨床判断の一致率は低い (Kitamura ら, 2000)。前述のような患者の権利が強く認められるにつれ、単なる印象判断で強制入院が決められない状況になりつつある。患者自身が希望しない治療が強制されるか拒否できるかは、まさに臨床医が患者の判断能力をどのように診断するかに掛かっており、治療同意判断能力について、臨床医の信頼性の高い評価法が求められているのである (北村ら, 1994)。

インフォームド・コンセントにおける診断も強制入院手続きにおける診断も、いずれも精神科医が社会に対して有している専門家としての責任を考えれば、決して評価の一致率が低くて済む問題ではないことは明らかである。妥当性を損なうことなく信頼度を向上させる努力が、業界としての精神科医療に求められているのである。

### 3. 精神保健専門家相互の情報伝達

個々の精神科医により使用する診断名 (病名) が異なることは、専門家相互の情報交換に多大な支障を生ずる (北村, 1990)。前述のようにある病名を半数の精神科医は使用し、他の半数の精神科医は用いない場合を想定してみよう。ある症例の紹介状を書く際に、まず自分はいかなる疾患分類体系を用いているかを明示しなければならなくなる。「私は非定型精神病という病名を使用するが、それでもこの患者は躁病であると診断した」とか「この患者はうつ病だと診断したが、私は非定型精神病の病名を用いないので、あなたのように非定型精神病という病名を用いればこの患者はそれに該当するだろう」といった紹介状は余りに繁雑になる。個別症例ではなく、さまざまな治療方法の成績を報告する論文において診断名の統一がなければ、論文の内容を読者が自分の臨床で応用することも困難であろう。

かつての精神科医療は 1 名の医師がほとんどすべての医療作業を行っていたが、近年のように医療チームによる医療が行われるようになると、看護婦、臨床心理士、作業療法士、ケースワーカーなど多くの異なる職種の専門家との情報交換をしなければならない。この際、「診断は医師の専権行為であり、その内容は他の職種の人々は知らなくても良い」といった態度は通用しない。また「医師の診断は医師ごとで異なる」といった説明も納得を得られない。

心理学を専攻したものについて、基礎的教育を行えば、診断基準に従った診断はかなり高い正確さをもって実施できることは報告されている (藤原ら, 1988; Hasui ら, 1999; 加藤ら, 1990; Sugiura ら, 1998)。藤原ら (1998)、加藤ら (1990)は、まず短い教育を行った上で、症例要旨法により心理学を専攻した者の RDC 診断の評定者間信頼度を求め、同じ方法による精神科医のそれ (須賀ら, 1987) に劣らない成績であることを報告している (表 4)。ま

た RDC の正答との一致率も心理学専攻者の成績は精神科医のそれにほぼ肩を並べるものであった（表 5）。精神科看護婦、作業療法士、ケースワーカーなどの精神科診断能力に関する実証的研究は余りみかけない。しかし、どのような職種の者でも、一定の教育と訓練を積めば、一定水準の診断能力は得られると考えるほうが常識的であろう。医学部出身者の場合は、身体疾患に関する教育を多く受けているから、精神疾患のなかでも器質性疾患の実際の診断能力は優れているが、これはそうしたことに係る教育量に依存したものと考えられる。

#### 表 4 表 5 を挿入

以前の精神科医療はある医師がある患者を「抱え込む」スタイルのものであった。こうした医療サービス供給方法であれば、患者の診断の信頼性や妥当性が問われる余地はなかった。患者が医療機関を自由に変更したり、一つの医療機関での治療に多くの専門家が関わるようになったため、従来隠されていた診断の信頼性や妥当性の問題が明確化されてきたのであろう。

仮に、臨床心理士の行った心理テストの結果を評定者間一致度が ICC で 0.2 ほどであったら、臨床精神科医はどのような批判をするであろうか。信頼性のないテスト結果は臨床のようには立たないと判断するであろう。そうであれば、臨床精神科医の診断の信頼度は低くて良いのであろうか。「臨床心理士の評価は技術であり、臨床精神科医の診断は芸術であるから、その両者を同一の平面で語るのは誤っている」というような発言をする医師はもはやいないであろうが、そうした医師がいれば余りに不遜な態度であろう。

### IV. 信頼性と妥当性

#### 1. 精神科診断の「芸術」と「科学」

そもそも精神医学で扱う諸概念は操作化に馴染まないとの意見もあろう。生気抑うつ、肛門期への固着、人格水準の低下、プレコックス感など、たしかに操作化の対象になりにくい概念は精神医学に少なくない。こうした概念を扱う場合、評価の再現性はほとんど問題にされない。多くの論者は、「信頼性の確保のために妥当性を犠牲にすることは誤りである」と主張している。しかし、信頼性の低い概念の妥当性を強く主張することも賢明とは言えない。再現性の低い概念を無批判に使用し、他の研究者や臨床家はその概念を理解できないことを問題としない態度は誤りであろう（藤縄ら, 1991）。

筆者は 1976 年から 4 年間英国の単科精神病院に病棟医として勤務する機会があったが、その際に入院患者の表情から観察できる感情鈍麻 **blunted affect** は日本でも英国でも同じだという体験を得た。これをきっかけに、感情鈍麻の評価が言語的情報にどれほど左右さ

れるかについて研究を行った (Kitamura ら, 1984)。この研究の対象は英国の精神病院に入院中の 20 名 (男性 16 名、女性 4 名) の慢性精神分裂病患者で、まず 2 名の精神科医が同席面接で Brief Psychiatric Rating Scale (BPRS) (Overall ら, 1962) と Symptom Rating Scale (SRS) (Wing, 1961) の評価を行い、この面接の様子をビデオで録画した。2 月後にこのビデオテープの画面を暗くして音声情報のみで BPRS と SRS の項目 (のうち音声だけでも評価できる項目: 音声項目) の評価を行った (audio review)。さらに 2 月後に今度は音声を消して視覚情報だけで同じビデオテープの症例の BPRS と SRS の項目 (のうち視覚だけでも評価できる項目) の評価を行った (visual review)。最後にさらに 2 月後に、音声と視覚情報をもとに与えた上で、同じビデオテープの症例の BPRS と SRS の全項目の評価を行った (audio-visual review)。その結果、(a) 感情鈍麻得点は面接時点と audio-visual review で visual review より高い、(b) 感情鈍麻得点と BPRS あるいは SRS の音声項目得点の相関は、面接時点と audio-visual review で高く、audio review 時の音声項目と visual review 時の感情鈍麻得点の相関はずっと低い、(c) 症状評価の医師間の一致率は面接時点と audio-visual review で高く、audio review では低かった。このことは、言語による情報 (例えば幻覚や妄想の訴え) が情動の評価にバイアスとなって関与することを示唆している。つまり、臨床判断をする際に、患者が幻覚や妄想などの症状を言語的に訴えると、我々はそれに引きずられて感情鈍麻も重くつけてしまうのである。恐らく、言語的訴えから患者の診断を付けてしまい、「精神分裂病なのだから感情も鈍麻しているに違いない」との無意識の判断が働くのかもしれない。

この研究は、精神分裂病の症状として重要であるといわれてきた情動鈍麻という概念の評価が不安定で、その評価には評価者が有している患者情報が強く影響することを示している。その後、行った陰性症状評価尺度 Scale for the Assessment of Negative Symptoms (SANS) (Andreasen, 1984) の研究 (北村ら, 1984, 1987) では、我々は極力陽性症状についての情報は聴取せずに行い、さらにアンカーポイントの整備を行うことでその信頼度を得ることに努力したのである。

すでに述べたように、臨床精神科医の診断上の判断は一致するものという誤った認識が、国内海外を問わず認められる。上記の研究から推測すれば、患者の非言語的特徴の判断は、例えば幻覚や妄想のような言語的情報が与えられていればそれに引きずられてしまうことから、2 名の精神科医の判断が通常の臨床場面では結構一致しているかのような感覚 (錯覚) を与えてしまっているのであろう。しかしこれは例えば明瞭な 1 級症状がある例で、その 1 級症状の判断が一致すると、それ以降、「1 級症状がある」→「精神分裂病に違いない」(診断の一致) →「そうであれば情動も鈍麻しているはずだ」(症状評価の一致) →「そうであれば病前の適応も悪かったはずだ」(行動評価の一致) という経路を経て、「常識と経験のある精神科医の判断は概ね一致する」という見掛上正しそうな結論に至るのであろう。

操作的評価方法を準備せずとも高い信頼度が (少なくとも訓練を受けた評定者の間では)

得られるのであれば、その概念はそのまま使用してもよい。しかし、それでは十分な信頼度が得られないのであれば (a) その概念に対して新しく操作的な定義と評価方法を作成するか (b) その概念を一切破棄するかの二者択一が迫られよう。繰り返して述べれば、概念がより主観的であれば信頼性も低いとはいえないし、構造化面接がすべて具体的な項目をその基準として要求しているわけではない。具体的な内容であっても信頼度の低い症状項目も存在する。例えば北村ら(1986)は同席面接法による2名の医師の **Schedule for Affective Disorders and Schizophrenia (SADS)** (Spitzer ら, 1978b) を用いた面接をうつ病関連の症状のうち、抑うつ感情、厭世観、食欲不振、体重減少、自殺傾向、不眠は 0.8 以上の  $\kappa$  係数を示したが、自責感の  $\kappa$  係数は低いことを報告している。米国での **SADS** 評価の信頼度検定ではほとんどの症状の  $\kappa$  が 0.8 を越えたもののやはりいくつかの症状は低く出ており、例えば行動増加の  $\kappa$  は同席面接法で 0.51 であった (McDonald-Scott ら, 1986b)。

操作的診断基準を作成してもなお信頼度の低い病名の例として家族歴研究診断基準 **Family History-Research Diagnostic Criteria (FH-RDC)** (Andreasen ら, 1977; Endicott ら, 1978) を挙げてみよう。親族のなかでの精神疾患発症の診断をつける際は、直接その親族に面接して診断する方法 (**family study method**) と、他のものからの情報をもとに推測する方法 (**family history method**) がある。時間的あるいは人的効率からすれば後者の方が優れているが、その診断には **false negative** が多いことが指摘された。そこで、病名診断の「閾値」を低くすることでこの問題を解決しようとして開発されたのが **FH-RDC** である。日本における **FH-RDC** の信頼度は症例要旨法で行われているが、うつ病、躁病、反社会性人格は高い信頼度を示したが、分裂感情病の信頼度は低かった (北村ら, 1984b)。同様の成績は米国でも報告されており (Andreasen ら, 1977)、家族歴という断片的で限られた範囲の情報の中で評価をするのは、分裂感情病の基準は困難すぎるということが伺えたのである。

重要なのは、いかに事柄の本体を表す概念であってもその評価のばらつきが大きければ以降の臨床的検討の用には立たないという点である。もし評価のばらつきが小さければ抽象的・主観的概念であっても躊躇なく診断基準にとりこんでよい。ある疾患について (比較的抽象的な) 定義を行ったなら、まず信頼性係数を求める研究を行い、その上でそれが臨床の場で多くの評価者が共通して用いることができる (比較的具体的な) 基準を準備し、その基準の信頼度を高めるような道具を作成することが、精神科臨床研究に求められているのである。

繰り返しになるが、病名診断にとって妥当性と信頼性はいずれも重要である。妥当性のない診断名は、いかに信頼性が高くとも何の価値もない。一方、信頼性のない診断名は、その妥当性を検証しようにもできない (診断名の誤差範囲が著しく広いため、妥当性指標に有意の所見を得るのが不可能に近くなる)。従って、当初は「勘」と「経験」と「権威」で作成したとしても、そうした診断基準の信頼性の向上に努め、その上で診断の妥当性検討に入ることになる。症状や診断の評価の信頼性を確認せずに、妥当性の議論ばかり行うことはなんら生産性のない作業である。症状や概念は抽象的であるほど信頼性に欠ける傾

向にあることは事実であるが、既に述べたように、具体的な症状であっても信頼性が低いものもあるし、抽象的概念であっても高い信頼性を得られるものもある。高い信頼性が得られるのであれば、精神科医療と臨床精神医学にとってはいずれも良いのである。例えば北村ら (1984a) は操作的診断基準の初期の試みである Feighner 基準の症例要旨法による信頼性を検討し、原発性感情病は高い信頼度 (ICC が 0.88) を示すのに対し、精神分裂病のそれは非常に低い (ICC が 0.24) ことを見出した。その原因は残遺期と寛解期の区別において意見が一致しないことにあった。操作診断でも信頼度が低くなることはあり得る。しかし、このように信頼度検定をすることは、評価の不一致の存在箇所を同定し、その原因を追求することができ、診断教育の役にも立つのである。使用している診断手法が評定者によって一致するのかわそれとも異なるのかわを知ることは大変重要である。信頼性検討そのものを否定する臨床家や研究者には、こうした視点が欠如している。

## 2. 構造化面接から診断基準へ

操作的診断基準に対する懐疑的・否定的意見のなかには、操作的診断基準ではなく精神科の構造化面接 **structured interview** に対する批判であることがある。本来、操作的診断基準と構造化面接は別個の事柄である。しかし、構造化面接は診断の操作化とともに発達してきたという歴史的経緯 (北村ら, 1984; 北村, 1988a; McDonald-Scott ら, 1986a) があり、不可分の関係にあるので、ここで検討する価値があろう。

2名の評定者 **rater** が下した診断の不一致は、評定者が同席して面接にあたった場合も、別の時点で面接にあたった場合も発生する。前者の一致率を評定者間信頼度 **inter-rater reliability**、後者の一致率を再試験信頼度 **test-retest reliability** という。診断の不一致の発生源として Spitzer ら (1975) は、被検者分散 **subject variance**、状況分散 **situation variance**、情報分散 **information variance**、基準分散 **criterion variance**、観察分散 **observation variance** の5つがあるとした。患者の状態が時間の経過とともに変化するため再試験信頼度が低下するのが被検者分散である。患者の症状そのものは同じでも、面接を実施する状況が異なるために患者の供述が変化するため再試験信頼度が低下するのが状況分散である。面接者の質問が異なるために患者の供述が異なり、再試験信頼度が低下するのが情報分散である。患者の供述やその他の情報は全く同じであるが、そうした情報を統合する方法が面接者によって異なるために評定者間信頼度が低下するのが基準分散である。また同じように、患者の供述やその他の情報は全く同一であるが、重症度や閾値の設定が面接者によって異なるために評定者間信頼度が低下するのが観察分散である。

こうした診断の不一致の要因をどのように回避するかにはいくつかの手法がある。まず、被検者分散は真の変化であるから、回避のしようがないが、信頼度検定にあたっては評価と次の評価の時間間隔をあまり開けない工夫が必要である。状況分散を回避するには同席面接を行えばよい。通常、同一の患者群に対する面接をした場合、再試験信頼度より評定

者間信頼度がやや高く現れるのはこれに起因しているのであろう。基準分散を回避する手法が操作的診断基準である。同じ情報を用いてもどの部分をどの様に用いて診断に至るかに個人差があれば、結果も一致しないことは十分あり得る。得られた臨床情報から診断に至るアルゴリズムを統一するのが操作的診断の主要目的である。一方、情報分散を回避する手法が構造化面接である（図 2）。必要最小限度の項目について質問することを義務付けるのが、構造化面接の主要目的である。言葉を代えれば、重要項目の聞き漏らしを避けることで情報化分散を低下させるのである。最後の観察分散を回避する手法が構造化面接の各項目に明文の定義を付けることとアンカーポイントの設定である。普段より何分早く目が覚めれば早朝覚醒とするのかといった点の定義付けがここでの主な作業になる。

### 図 2 を挿入

ところで、アンカーポイントの設定はこれまで十分な議論がなされていないように思える。DSM-III において痴呆の診断基準が導入されたことを機に北村ら (1985) は 3 名の精神科医による同席面接法で痴呆診断の信頼度検定を行った。3 名の医師 (A、B、C) 相互の診断一致率 ( $\kappa$ ) は AB が 0.68、AC が 0.69、BC が 0.48 であり、医師 B と医師 C の意見が強く食い違っていることが認められた。3 名の医師の臨床経験を調べたところ、医師 A は、ごく短期間老人病棟を受けもったことがあるが、主として成人の機能的な精神疾患を扱うことが多い者であった。医師 B は、定期的に老人クラブにおける診察を続けていた。そして医師 C は、外来の診察に加えて、精神病院における老人患者を扱うことが多かった。つまり、医師 B は日常、軽症の痴呆例に接することが多く、医師 C は逆に重症の痴呆患者に接することが多く、医師 A がその中間に位置している。したがって臨床経験の内容からは A と B、A と C は近いが、B と C は遠いのである。おそらく B は軽い状態も「症状あり」と取り、C はよほど重くないと「症状あり」とは取らないのであろう。これが 2 名の精神科医の診断一致率に反映していると推論できた。従って、操作的診断基準を設定しても、各症状の重症度のどこに「あり」と「なし」の区分点を決めるかで個人的差異が発生すれば、最終診断は合わなくなるのである。操作的診断基準の各項目に該当する症状の評価について、さらに具体的な評価方法が設定されなければならない。

以上の説明から分かるように、個別患者の診断の確度は、操作的診断基準を用いればすべて担保できるというものではない。入力された情報に誤りがあれば、出力された診断も正確とはいえない。一方、入力される情報が豊富でも、診断にいたるアルゴリズムが異なれば最終結果も異なり得る。こうした観点から、操作的診断基準と構造化面接は相補的な関係にある。実際の臨床場面では、構造化面接を用いることで高い操作診断の信頼度を得ることができる (北村ら, 1986)。

ところで、構造化面接に対する批判の多くは、それが硬直的であり、患者の心理内界を理解するものではないというものである。質問の順番が決められているため、面接の流れ

が人工的になる。構造化面接では決められたことしか質問せず、患者の体験に共感していない。共感がなければ患者の精神内界の真の理解はできない。構造化面接に記載のない症状や体験については興味を示さない。場合によっては構造化面接に挙げられていない事項について患者が話し始めると面接者がさえぎることがある。

これらは構造化面接の初心者にはしばしば見られる問題であり、批判自体は正当である。しかし、スピード違反の自動車事故を理由に車の存在が罪悪だと主張することと同じで、下手な構造化面接をする面接者の存在を理由に構造化面接そのものを否定することは当を得ていない。本来、優秀な臨床医の診断面接はかなり構造化されていた。診断のために聴取の必要な症状はおおよそ決まっていたし、「訴えのない症状は存在しない症状ではない」と言うことは、医学全般の診断技法で教えられてきた。患者に確認しても否定されてはじめてその症状は存在しないといえる。共感を示す態度でなければ患者は自己の体験について語りたがらなくなる。構造化面接は非共感的な態度を維持せよとはどこにも書いていない。項目の質問順序は初心者では決められた順を守るよう要請されるが、習熟した面接者では、構造化面接のなかで（必要に応じた）自由な移動が許されているし、そのほうが望ましい。例えば Present State Examination (PSE) (Wing ら, 1974) や SADS では項目の移動の許可が明示されている。

### 3. 定性的精神病理学と定量的精神病理学

従来診断を擁護し操作診断を否定する論調の一部には、症状の段階評価（とおそらくはそれをもとにした統計的解析方法）に対する批判があるようである。木村 (1994) は次のように述べている。

「精神医学にも「計量精神医学」というのがある。患者の気分の変動とか不安や妄想の程度とか、行動の不自然さまで、「五段階評価」のようなもので計測して、その評点の上下でもって病気の悪化や改善を測定しようとするのだが、そんな方法で「こころ」の症状が測定できるなどということを「常識」のある精神科医ならけっして考えないだろう。」

そして、精神症状の測定不可能性について木村 (1994) はさらに解説を加えている。

「血液検査にしてもその他の検査にしても、すべて正常値というものが決まっていて、どの検査でどの値がどれくらい正常値から外れているかによって、大体は病気の診断がつく。治療をして病気が治ってくると、検査の成績もすべて正常値の範囲内に戻ってくる。…精神医学ではそうはいかない。精神活動を数字に置き換えるなどということができないからである。躁病の気分変動だった

ら、まだしもその程度を数字なり曲線の上下なりで表すこともできるだろう。彼は今日はちょっとハイだとか、彼女は今のところ落ち込んでいるとかいう言いかたは、そのひとなりの平均的な気分の水準を一応の「正常値」と見なした上で、そのレベルからの上がり下がりを行い表している。しかしこのような気分の逸脱は身体医学の異常値のように客観的な判断基準にはなりにくい。「正常値」がその人その人で違って、普段から物静かで口数の少ない人も陽気で活動的な人もあるからだ。」

身体現象は測定可能だが精神現象はそれが困難であるとの主張は根深いものがある。しかし、測定という点に関して身体現象と精神現象はそれほど異なるのであろうか。血液検査には正常値が決まっているというのは国民全般の平均値とその分散が分かっているという意味であり、個体の正常値を示しているのではない。もともと体温の低いものが感染症に罹患し、体温が「正常値」の範囲内であっても「発熱」といえる状態はしばしば遭遇する。身体現象も心理現象も集団の平均値が個体の正常値とは必ずしもいえないのは共通している。「正常値」を越える血圧でも、マラソン直後に測定したものであればそれは病的とはいわない。ふだん静かな人が環境に応じて陽気になったとしてもそれだけで異常とは考えない。この環境を考慮に入れるのである。測定という観点で見れば、身体現象も心理現象もいずれも同様の性質を持った測定対象であるといえる。

第 2 に、段階評価が心理現象の評価に馴染まないと主張する研究者も、特定の個人が特定疾患に罹患している・いないの判断や、行動が自然だ・不自然だの判断や、自殺の危険がある・ないの判断などは行っている。段階評価はある事柄の程度を複数の段階（アンカーポイント）に分けて評価する方法である。開発したものの名をとってリカート尺度ともいわれる (Likert, 1932)。例えば BPRS 各症状は (1) なし (2) ごく軽度 (3) 軽度 (4) 中等度 (5) やや高度 (6) 高度 (7) 非常に高度の 7 つの程度が示されている。血圧であれば水銀柱の高さ (ミリメートル) で表示できるが、精神症状はこうした指標がないので、アンカーポイントの定義に従って評価者の主観で決めるのである。しかし、例えば作為体験が「あるかないか」の判断をするには 7 段階のアンカーポイントはいらない。「なし」と「あり」に対応する「1」と「2」、あるいは「0」と「1」で十分である。であるから、こうしたカテゴリカルな現象の評価はリカート尺度の特殊型である。上記の行動が「自然だ」「不自然だ」の判断なども 2 段階のリカート尺度である。旧来の精神病理学は、精神現象をいわば定性的に研究し、臨床の現場では 2 件法による段階評価を行ってきたのであり、一方、構造化面接で多く使用される症状評価は多段階による評価を行っているのであり、評価の本質は同一である。

症状の段階評価に対する従来診断家の批判は、評価尺度にも向けられている。患者という人間の全体像をいくつかの項目に切り分け、各項目ごとの評価を行い、その後それを足し算しても、その結果は患者の全体像ではないと言うのが、こうした批判の骨子である。

では従来診断を行っているものは全体像をどの様に把握しているのでしょうか。病歴簿の記載を見れば、例えば「ゆううつそうな顔貌」「希死念慮」「作為体験」など要素心理学 *faculty psychology* にもとづいた記載が並んでいるのではないかと考えれば、古典的精神病理学こそ、人間の心理現象を多くの要素（例えば、思考、感情、知覚、知能など）に分類した上で、各要素の評価を行っていたのである。先に述べた通り、旧来の精神病理学は定性的評価であり、近年の評価尺度は定量的評価が多いという違いしかない。加えて、細部に分けずに全体を評価する場合、どうしても重要な細部の評価に目が届かなくなるという欠点がある。おそらく全体的評価には、評価する側に一定の理論的枠組みがあり、それに従って判断を下しているのかもしれない。そうであれば、臨床上は誤謬に気付きにくい危険な方法であり、研究上は新しい発見の少ない閉塞した方法である。

## V. 診断と治療

### 1. 病名診断と診断的フォーミュレーション

操作的診断基準の必要性について検討する時に、診断 *diagnosis* という用語の意味を考えなければならない。医学・医療において診断は (1) 病名診断 *nosological categorisation* と (2) 診断的フォーミュレーション *diagnostic formulation* という 2 つの意味を持っている (表 6) (北村, 1991)。

#### 表 6 を挿入

なんらかの病的状態になった個人の症状や状態はひとつとして同一のものはない。例えば糖尿病という病名診断がついた人々をみても個人によって症状や経過はすべて異なる。しかし、患者の数だけ病名があるとは普通考えない。異なる病状を呈していても、その中で臨床上役に立つ共通項を見出し、それをを用いてひとつの病名をつけることが病名診断である。共通項の妥当性を担保する様々な臨床的指標のことを外的基準 *external criteria* と呼ぶ。ある病名診断をひとつの疾病単位として認めることの妥当性の高低は、何を外的基準に持ってくるかで大きく左右されよう。また、病名診断では患者それぞれの個人的特徴は切り捨てられる。つまり、病名診断は多くの人口のデータを集計してはじめて行える作業である。その成果をもとに、個々の患者にあてはめを行うことが、病歴簿の「診断名」の欄へ病名を記載することに繋がる。従って、病名診断では診察時点までの臨床経過を後方視的 *retrospective*、事後的に観察し、事前に決めてある病名一覧表のなかから当該患者の有する病態がいずれに当てはまるかを事実認定的に決定する。だからこそ操作化 *operationalization* が可能になる。

一方、診断的フォーミュレーションは、個別患者の個別的特徴のうち、予後 prognosis 判定と治療方針の決定という臨床目的にとって重要な情報をまとめあげる作業であり、疾病分類による病名診断は重要ではあるが診断的フォーミュレーションの一部分でしかない。身体各科においても、入院させるか外来で診るか、薬物を開始するか、開始するならどの薬剤を選択するか、選択した薬剤の容量をいかほどにするか等の決定を、病名診断だけで行う者はいない。つまり、病名診断は予後判定と治療方針の決定に大きく関与するが、そのみが排他的に関与するものではない。診断的フォーミュレーションの対象はあくまで個人であり、その個人が今後どの様になるかを考え、医療従事者が何をすべきかを、前方視的 prospective に検討する作業である。当然、これは仮説設定であるといえる。その診断的フォーミュレーションに沿って治療を開始しても、予想と大きく反する進展を見た場合、当初の診断的フォーミュレーションを見直すことになる。「見立て違い」があったら途中で潔く変更することは医療の伝統であった。診断的フォーミュレーションが変更される場合は、その構成要素である病名診断も当然変更の対象となりうる。個別症例の他とは異なった特徴を記載するのであるから、診断的フォーミュレーションはひとつの名詞で表現できるものではなく、数行から十数行の文章になる。こうしたことから、診断的フォーミュレーションの操作化は容易ではないし、特別に操作化する必要もない。

なお、操作診断は一時点での診断を行う方法で、操作診断でいったんついた病名は以降一切変更されないと誤解される向きもあるが、そうではない。新しい情報、新しい状況で、操作診断による診断は変更されるものである。なぜなら入力されるデータに変更があれば、出力結果も当然に変化するからである。

以前に中安 (1991) と北村 (1991) の間で操作診断に関する議論があり、その中で「診断」の定義について意見の不一致があった。中安(1991)は、病名診断は疾病分類学 nosology であり、彼のいう「診断」ではないと主張している。しかし、ほとんど大部分の臨床医は病歴簿の「診断名」の欄に××病という病名を記載し、決して診断的フォーミュレーションを長々と記入してはいない。通常の病歴簿の診断名欄は診断的フォーミュレーションを記入するだけの空白は用意されていない。「診断」という用語を診断的フォーミュレーションに限定するのであれば、中安(1991)は日本国内では例外的存在と思われる。一方、私が知る限り、日本の医学教育で診断的フォーミュレーションの理念は十分教育されているとは思えない。従来の医学教育のなかは「診断」というと、胃癌、肝硬変、糖尿病などの病名の決定と同義語として使用されているように思える。

「診断は治療の侍女」という表現がある。治療に役だってはじめて診断は価値を持つという意味であろう。ここでいう「診断」は病名診断であろうか、診断的フォーミュレーションであろうか。おそらく後者であろう。だから日本にも診断的フォーミュレーションは存在したのである。ただそれが「診断」としての位置を与えられず、むしろ「みたて」といった漠然とした概念で考えられてきた。

ところでこの「みたて」を行うのはどのような情報が与えられた上での話しであろうか。

何等情報がなく、単に患者の顔を5秒見ただけでは「みたて」は行えない。もちろん「みたて」は患者によって変わるもので、変わるからこそ「主観的な」「総合的な」判断が必要なのである。ではあるが、なお基本的・共通的な情報は存在する。医学の伝統には病歴の記載方法があり、(a) 主訴 (b) 現病歴 (c) 既往歴 (d) 家族歴 (e) 生活史 (f) 現在症などの順に記載するのである。この病歴のエッセンスを抽出したものが診断的フォーミュレーションである。だから、診断的フォーミュレーションの構成要素は（純粹に個別的な事項もあるが）病歴記載の項目であるといえる。

病歴記載の項目が診断的フォーミュレーションを基礎付けるのであれば、病歴記載のおおのこの項目を操作化することも可能である。DSM-III以降用いられている多軸診断 *multiaxial diagnosis* の基本理念はここにある。DSMで採用されている5軸は、臨床症候群、人格障害（発達障害）、身体疾患、心理社会的ストレス、機能状態の5つである。DSMにおける「診断」を病名診断に限定して考えると、必要な軸は最初の2つの軸だけである。日本における研究報告でDSMを用いる場合、第1軸と第2軸のみの報告が多いのもこうした理解（誤解）があるからであろう。しかし、DSMにおける診断を診断的フォーミュレーションであると考え、その構成要件は病歴記載の項目であると理解すれば、多軸診断は日本の臨床にも馴染まないものではない。主訴と現病歴は第1軸（あるいは第2軸）であり、既往症のなかには第3軸が含まれ、生活史では第4軸と第5軸が主要な情報を与える。これまでの5軸以外にも他の軸を追加すべきという主張もあるのは、現在の5軸では診断的フォーミュレーションに必要な情報のすべてを網羅できないからであろう（北村, 1993）。計量的評価を否定する者も、初診患者の診察では伝統的病歴聴取を行っているはずであり、この点でDSM風の多軸診断を行っているのである。

## 2. 病名診断と治療的介入

上記の議論から明らかなように、本論の検討課題である操作診断は、多くの場合、病名診断を指す。私の上述の議論はこうした病名診断において操作化がなぜ必要かに焦点を当てたのである。操作的診断基準が診断的フォーミュレーションの役に立たないとの理由からその弊害を主張する論者に対しては、その範囲の限りにおいて、私は賛成する。操作的診断基準による病名診断のみで診断的フォーミュレーションは行えないのは当然である。操作的診断基準の使用のみで臨床診断が完了したとする臨床医がいればその者の考えは誤っている。操作的診断基準による病名診断はあくまで診断的フォーミュレーションの一構成要素である。操作的診断基準を用いた病名診断のみで個別患者に最も適した治療方針を立てることはできない。もしそのようなことを行う臨床家がいれば余りに短絡的である。しかしこれらの事柄は、操作診断の欠陥であったり、罪状ではない。これらは、操作診断による病名診断の本来の目的を理解せず、操作診断を誤用している臨床医の欠陥であり、責任であり、罪状である。法定規制速度を大幅に越えたスピードで自動車事故を起こした

からといって、「だから車には欠陥がある、社会に車があることは罪悪だ」と主張することは当を得ない。欠陥があるのは無謀運転する人間にあるのであって、車にはない。操作診断も車もあくまで道具であり、それを正しく使用するのも、誤用・乱用・悪用するのすべて使用する人間の問題である。操作診断の誤用・乱用・悪用の擁護の必要はない。

中安 (1991) は操作的診断基準が治療を放棄したものだとして批判したことがある。彼は特に診断基準の持続期間について、もし早い段階に治療を開始すれば短期で快復してしまい、診断に十分な症状の持続が得られず、(操作的診断基準を用いる限り) 治療が診断を不可能にしてしまうと主張している。しかし、操作的診断基準の側からこの議論を見ると、中安 (1991) は診断もつかない内に断定的な治療を始めているといえる。

中安 (1991) は精神分裂病を例にこの議論をしているが、ここではうつ病を例に考えてみよう。良く知られている通り DSM-IV の大うつ病 *major depression* の診断基準で要求されている持続期間は 2 週間である。症状項目を満たしていてもその状態が 2 週間を越えて始めて大うつ病の診断が許される。2 週間を越えない場合は、「その他」の診断となる。なぜ、2 週間かについては実証的データは乏しいが、ファイナー基準 (Feighner ら, 1973) 以来の伝統である。ところで、ある女子大生が抑うつ感情といくつかの大うつ病関連症状を訴えて学生相談室に来所したとしよう。抑うつ感情は、問題ないと思っていた学科の前期の試験が驚いたことに不合格だった日の翌日に発生した。たまたま彼女の担当医が同じ大学で行っていた授業を取っていたこともあり、この女子大生は学生相談室にやってきた。症状はこの 4 日間続いていた。女子大生は「こんなことは実家の親にも言えないし、自分の成績を知っている高校時代の友人にはとても恥ずかしくて話せない、授業に出ても講義を集中して聴けない」と訴えた。この女子大生には、まだ 4 日しか持続していない挿話であるから DSM-IV の大うつ病の診断は下せない。

中安 (1991) の論法を用いれば、5 日目の「うつ病」患者に治療的介入 (精神療法・薬物療法・電気けいれん療法などいくつかの治療法の選択肢があるが、その適否はここでは問わない。仮に三環系抗うつ薬を投与したとしよう) を行って「治癒」に至ったなら、治療をしたがために本来の「うつ病」の診断はつけられずに終わったのであり、もし DSM-IV の診断を確定したければあと少なくとも 10 日は全くの無治療で様子を観察し、上記の症状が 14 日間持続することを見届けなければならない。つまり、DSM-IV は 10 日間にわたって治療を放棄したのである。初期診断を否定し、硬直した持続期間基準を設定した DSM-IV 診断の臨床上的欠点がここにある。

中安 (1991) の論法では、4 日しか持続していない抑うつ挿話であっても、「うつ病」(ここでは DSM-IV の大うつ病) の診断が適当で、治療的介入は妥当であることになる。しかし、本当にそう言えるのであろうか。そういうだけの合理的根拠を我々は提示できるのであろうか。中安風に考えれば、そもそも DSM-IV の大うつ病基準の持続期間条項は不要であるばかりか有害でさえあるのではないか。実にこれに答えるだけの合理的根拠はほとんどないのである。これに答えるにはいくつかの疫学調査結果が必要である。第 1 に、持続

は 4 日間であっても症状基準をすべて満たす抑うつ状態は無治療の場合はほとんど 14 日以上持続を示すという疫学所見があれば、上記の批判は合理的といえる。地域住民中で見られるうつ病の 9 割は医療機関を受診しない (藤原, 1995) ことから、こうした研究は実行可能と考えられる。第 2 に、仮に持続期間が 14 日未満の抑うつ状態であっても、症状項目が DSM-IV の大うつ病のそれに合致しているなら、いくつかの外的基準について、14 日以上持続の大うつ病と同じであるという所見があれば、これも上記の批判は合理的といえる。この外的基準については詳しく後述するが、ここでは抑うつ状態の再発率を取り上げて見よう。もし、14 日未満の持続の挿話と 14 日以上持続する挿話の 2 年以内の再発率がほぼ同様で、いずれも抑うつ状態のない人口 (対照群 control) のそれよりも目だって高いことを示すことができれば、(かなり強く) うつ病診断で持続期間基準の日数を低めるべきであると主張できよう。当然に、では 1 日の持続でもうつ病とよいかとの疑問が発生する。従って、上記の研究課題は「2 年以内の再発率を区別する持続期間の区分点は何日であるか」になる。もし、未治療でも持続期間 14 日未満の抑うつ状態の多くが 14 日以内に回復するという成績が得られたとすると、上記の女子大生に三環系抗うつ薬を投与したことには何の価値もなかったことになる。「失敗したといっても前期の試験だし、後期にしくじらなければいいのだ」といった認知的対処行動を取り、抑うつ状態から回復することも十分想定できる。無駄な治療で医療費を請求したのであれば独善であり、社会的に糾弾されても致し方あるまい。こうした基礎的研究は大変少ないのが現状である。従来診断を主張し、初期診断の重要性を主張する精神科医は、まずもってこうした基礎研究を推進すべきであろう。それがなければ臨床家の単なる自己満足に終わってしまう。

十分な情報がないまま治療を開始しなければならない場合や、緊急時については、診断基準に則った病名診断はできない。この場合は、鑑別診断を行い、臨床上最も適切と思われる介入を行う。これも十分な診断的フォーミュレーションの作業である。操作診断による病名診断は診断的フォーミュレーションを禁止するものではなく、個別症例における臨床上の要請は病名診断を越えるものである。ただし、不十分な情報のみで鑑別診断を行うのであるから、最終診断とは異なることも少なからずあり得る。この場合の法的責任は医療者にある。かりに訴訟になった場合、操作診断を行うのに十分な情報が得られない状況だったということは医療者に過失がないことを示唆するが、一方、十分な情報が得られており操作診断でも診断可能な疾患を見過ごした場合は医療者に過失が問われ得る。つまり、操作診断を臨床の場に導入することは、医療者と患者の双方に法的保護を与える基礎となり、医療者の独善的診療行為に一定の歯止めを与えることになる。

#### IV. 病名の妥当性

##### 1. 症候群と病名診断

操作診断には状態像 *Zustandbild* 診断が欠けているとの批判がある (中安, 1991)。中安 (1991) は、状態像を「(いくつかの) 症状のたんなる複合とは異なるもの」で、例えば幻覚妄想状態なら「幻覚と妄想の存在は当然」として、「そのほかの精神内容のすべてを、また不安げな、また時に猜疑的、時に敵意を抱いた目つきや表情、硬く閉ざされたような姿勢のあり様等にはじまる表出のすべてを一塊のものとして示すパターンである」と規定している。そして状態像診断はパターン認知であり、状態像のエッセンスを取り出したものが症候群 *syndrome* であることも述べている。中安 (1991) は疾患の診断の前提が状態像の把握であり、*DSM-III-R* にはこうした状態像診断がないと批判しているのである。たしかに、*DSM-III-R* に状態像診断がないのは事実である。しかし、状態像 (症候群) の評価は中安 (1991) がいうほどには簡単なものではない。

そもそも、従来診断に十分なパターン認識に基づいた状態像診断が存在していたであろうか。一体、従来診断における症候群は何種類あるものであろうか。この点を明示した論文は意外と少ないように思う。かつて標準的な教科書といわれた西丸(1957)の教科書では、状態像として、神経衰弱状態、減動増動状態、幻覚妄想状態、錯乱状態、記憶減退状態、欠陥状態の6つが挙げられている。しかし、各状態像がどのようなパターン認識から形成されたのかについての記載はない。おそらくは、従来からの伝統(「権威」と自らの創案(「経験」)を折衷(「勘」)した上で作成されたのであろう。例えば、幻覚妄想状態は「幻覚と妄想が主となる状態」とされているが、これは幻覚と妄想の両者が存在しなければ幻覚妄想状態といわないのであろうか。例えば、幻聴のみを症状とし妄想を欠く症例や幻覚を伴わない追跡妄想の症例は西丸 (1957) の分類ではどの状態像とされるのであろうか。恐らく幻覚妄想状態とされるのであろう。そうであれば「幻覚または妄想が主となる状態」といった説明にすべきである。その場合、幻聴のみを症状とし妄想を欠く症例と幻覚を伴わない追跡妄想の症例は同一のグループに分類する合理的根拠はどこにあるのであろうか。また、ひとつの状態像 (症候群) に属するとされた症状は同時的に存在する頻度が高いのであろうか。例えば、欠陥状態に挙げられている「感情が鈍い」「興味や関心が失われている」「意欲が減じている」という症状は、共に現れやすいのであろうか。そのうちある症状Xがあれば同時に他の症状Yも現れる可能性が、症状Xが見られない場合に症状Yが現れる可能性より高いという所見があるのだろうか。

第2に、各状態像 (症候群) は相互に排他的 *mutually exclusive* なカテゴリーであるのか、1人の患者で重複が許されるのかについての厳密な規定がない。重複診断 *multiple diagnosis* が許されるのであれば、ある患者の状態像 (症候群) は「幻覚妄想状態+減動増動状態+欠陥状態」ということもしばしば起こるはずである。だとすると「…を主とする状態」という説明は不要で、単に「…が見られる状態」でよくなる。だが、多数の重複診断が状態像 (症候群) に与えられるとなると、中安 (1991) の言うような、疾患の診断の前提としての状態像を把握する価値はなくなる。一方、西丸 (1957) の6つの状態像 (症候群) が相互排他的な関係にあり、1時点の1名の患者は1つの状態像 (症候群) 診断しか

与えられないとすると、例えば幻覚妄想があり同時に興奮・思考奔逸がある患者は、幻覚妄想状態とするのか減動増動状態とするのか。おそらく、臨床像の前景に立っている症状で判断するのであろう。では、どうやって前景に立つか立たないかを評価するのであろうか。

第 3 に、ひとつの状態像（症候群）のなかにある特質の両極に位置する症状を含めている場合がある。減動増動状態がそうで、「口数が多いー少ない」「表情が動かないー活発」「動作が少ないー多い」と言うように、通常は（気分障害の混合状態 *mixed state* などの例外を除き）同時に存在しえない症状が含まれている。これらの対極的症状が 1 つの状態像（症候群）を構成するというのであれば、1 症例の中でその経過中に一方の症状があれば他の時点で対極の症状が出現する（あるいは出現する頻度が高い）という事実が示されなければならない。

従来の伝統的精神病理学を主張する研究者はこうした疑問に答える研究結果を示してはいない。

ある特定の症状 X（例えば不安感）が他の特定の症状 Y（例えば強迫念慮）と同時に出現する可能性が高いかどうかを見るには、患者集団（例えば 100 名の外来新患）のなかで X と Y の有無（と出来ればその重症度）を測定し、両者の相関係数を求めればよい。X と Y の相関係数が高ければそれだけ両症状が 1 つの状態像（症候群）（西丸がいうところの神経衰弱状態）を形成する可能性が高いと推論できる。多くの症状の関係を同時に検討するにはそれぞれの 2 症状間の相関係数を求めても全体像は把握できない。この場合、因子分析 *factor analysis* が有効である。因子分析からはいくつかの因子 *factor* が抽出され、ある因子に高い因子負荷量 *factor loading* を示す複数の項目（症状）がひとつのまとまりを有すると推定できる。ところで、中安 (1991) が言うように、(a) 何らかの幻覚 (b) 何らかの妄想 (c) 不安げな、猜疑的、敵意を抱いた目つきや表情 (d) 硬く閉ざされたような姿勢の出現は、すべてひとつの因子に高い因子負荷量をしめるという所見が報告されているのであろうか。旧来の精神病理学者は、それは体験的に正しいと主張するであろう。しかし、たとえば不安げな、猜疑的、敵意を抱いた目つきや表情は、不安神経症や躁病の易怒気分などでも起こるのであろうし、硬く閉ざされたような姿勢はうつ病の精神運動制止でも見られるであろう。こうした実証的所見がないまま個人の臨床体験のみから 1 つの状態像（症候群）を規定するのは慎重を欠いているとの非難を免れないであろう。

精神症状を因子分析などの多変量解析の手法で群分けする（症候群を規定する）研究は 1970 年代以降多く行われてきた。この進展はひとつには大型計算機の出現と汎用化に負うところが大きい。実際に 1970 年代以前の因子分析は手計算（あるいは手回し計算機）を行い、ひとつの結果を出すのに 1 日をかけていたものが、最近の大型計算機を用いれば、（入力が完了さえすれば）計算自体は 1 秒もかからずに終了する。精神科の症候群に関するこれまでの因子分析研究をすべてここで概観することはできないので、操作診断に関連する部分を検討する材料になるもののみを紹介する。

Kitamura ら (1995) は、584 名の精神病院入院患者について、入院時の症状評価の結果を因子分析し、5 つの因子を得た。各因子に高い因子付加量を有する項目から、各因子は (1) 躁症状 (2) 抑うつ症状 (3) 陰性症状と思考形式の障害 (4) 陽性症状 (5) 緊張病症状の 5 つであると解釈できた。罪業妄想は抑うつ症状に属し、誇大妄想は躁症状に属していた。

この研究から、何らかの幻覚と何らかの妄想がひとつのグループ（症候群）を作るのではないということが分かる。罪業妄想はうつ病に関連する症状と共に出現する可能性のほうが高い。誇大妄想は躁病に関連する症状と共に出現する可能性のほうが高い。従って、幻覚妄想状態という状態像（症候群）を規定するにあたっては、「何らかの幻覚および何らかの妄想」ではなく「…を除く幻覚および妄想」という規定のほうが厳密であろう。従来の精神病理学者はおそらくこうしたことは常識ある精神科医であればだれでも知っていることだと反論するであろう。この反論は正しくない。もし、そうであれば最初から「…を除く」という但し書きを挿入しておくべきであろう。また、躁病にみられる誇大妄想は臨床像の前景に立つものではないとの反論もあろう。はじめに述べたように、「前景に立つ」ということの明分の定義はなかった。「前景に立つ」という不明瞭な記載ではなく、「…を除く幻覚および妄想」であればだれの目にも明瞭であろう。

精神分裂病に見られる症状の分類について、Crow (1980, 1985) の陽性症状 positive symptoms・陰性症状 negative symptoms 2 分論は大きな影響を臨床家と研究者に与えた。精神分裂病に関する因子分析も多く発表されてきた。このなかで、北村ら (1989) は BPRS と SANS を用いて慢性精神分裂病患者の症状の因子構造を解析し、その症状がおよそ陽性症状と陰性症状に分けられることを確認した。しかし、Crow が主張するように 2 分論ではなく、3 つの独立した群があるとの意見が少なからず出されるようになってきた (Liddle, 1987; Liddle ら, 1990; Peralta ら, 1992; Thompson ら, 1993)。こうした因子分析では、陽性症状から思考形式の障害を中心とした解体症状が独立してきている。こうした所見は、旧来の方法による状態像診断では分からなかった点であろう。

神経症の領域では図 1 に示した古典的亜型分類は近年の操作診断のなかではさらに細かく分類されてきている。こうした流れに対抗して、そもそも神経症はひとつのまとまった病態であるという generalist の主張も存在する (Looney ら, 1979; Andrews ら, 1990; Larkin ら, 1992; Sims, 1986)。診療所初診患者の症状評価を行った Kitamura ら (1997) は、その因子分析の結果から、神経症症状は大きく 3 つのカテゴリーに分類できることを示した。西丸のいう神経衰弱状態が現在の神経症の様々な状態像に該当するのであろうが、それがいくつかの症状群に分けるべきかについて因子分析はこのようにひとつの方向性を与えることができる。少なくとも、旧来の「勘」と「経験」の症状学だけでは進展は期待できない。

因子分析は、対象のおき方、扱う項目（症状）の設定によって結果がかなり変化するという特徴がある。因子がいくつあるのかについて異論が多く出されることはまれではない。この場合、従来の探索的因子分析 exploratory factor analysis でなく近年の確認的因子分析

confirmatory factor analysis の手法を用いれば、いくつの因子を設定することが最も妥当であるかを定めることができる。Lenzenweger ら (1996) の研究は精神分裂病の症状を 4 つに群分けすることの妥当性を確定的因子分析で示した画期的な報告である。Lenzenweger ら (1996) の因子は (1) 妄想・幻覚からなる reality distortion (2) 思考形式の障害と奇異な行動からなる disorganization (3) 陰性症状 (4) 病前社会不適応の 4 つである。

操作診断には状態像診断が欠けているとの批判 (中安, 1991) はその通りであるが、上記の記述から分かるように、従来の状態像 (症候群) 類型には実証的証拠が少なく、最近の因子分析などを用いた研究では従来の類型とは異なる症状の組み合わせが提案されてきている。従来診断を主張する医師は、こうした症状の群分けは従来直感的に行ってきたと主張するであろう。しかし、常識と直感に頼った症状の群分けには限度があり、誤謬が発生しやすい。もちろん、因子構造の解釈をする上でこれまでの精神病理学の知識は不可欠である。しかしだからといって精神病理学者の直感だけでは病因や治療法に結びつく知識の蓄積にはいたらない。状態像 (症候群) を診断に組み込むことは従来の精神病理学者が思っているほどには容易な作業ではないのである。

## 2. 診断基準の妥当性

これまで病名診断の議論では病名自体が臨床上妥当なものであるかには触れなかったもので、ここで、従来診断と操作診断の対比を行うに必要な範囲で、ごく簡単に検討しておきたい (Robins ら, 1989)。

ある基準を用いて複数の患者をひとつの識別名 (病名、診断名) で「くくる」作業が妥当であることを説明する方法には医学の伝統がある。これを近代的な術語で書き直したのが Robins ら (1970) に代表される St. Louis のワシントン大学の研究グループであった (彼らはセントルイス学派とも呼ばれる)。

その第 1 の方法は症状のまとまりであり、症候群についてはすでに述べた。特定の複数の症状をもってひとつの識別名 (病名、診断名) としたならば、これらの症状が同じ個体 (患者) にほぼ同時に出現する頻度は偶然の可能性を大きく上回っているはずである。

第 2 の方法は、同一識別名 (病名、診断名) で陽性とされる者が血縁に、これも偶然の可能性を大きく上回って多く出現するという家族集積性である。この現象は遺伝性のことも環境性のこともあるが、いずれの場合も家族集積性のある「くくり方」は病名の独立性を強く示唆すると考えられてきた。

第 3 の方法は、経過の同一性である。横断面での症状から同一識別名 (病名、診断名) で陽性とされる者のその後の変化の様子がほぼ同じであれば、これもその「くくり方」が病名の独立性を強く示唆すると考えられてきた。例えば、反社会性人格障害のおおくの成人は児童期に反社会的行動を示すが、反社会的行動を示す多くの児童は成人になって反社会性人格障害の診断がつくわけではないこと (Robins, 1978) から、児童期の行為障害

conduct disorder は成人の反社会性人格障害とは別のカテゴリーであると考えられる。精神分裂病が当初早発痴呆 dementia praecox と呼ばれたのはこの経過に重点を置いたためであるという記載は多くの教科書の示すところである。

第 4 に、特定の治療への反応性が同一であれば、その「くくり方」は妥当であると考えられる。

第 5 に、もし手に入れられるのであれば何らかの検査所見が同一であれば、これも識別名（病名、診断名）の妥当性を強く示唆する指標である。

多くの場合、疾患の診断方法は臨床医が共通のイメージを持つことから始まり、やがてそれに概念規定を与えようとする努力が続き、その上で明文の規定を決める作業が行われる。いずれの定義であっても、その定義による疾患概念が臨床単位として妥当であることを主張するには、上記の 5 つの基準のうちいくつかを示すことが要請される。特定の疾患の定義はこれまでにいくつも提案されてきている。例えばうつ病については古典的な内因性・反応性の 2 分類が実証的所見とは合致しないことが指摘され (Klerman ら, 1979; 千葉ら, 1987)、多数の亜型分類方法が提案されている (北村, 1988b)。そのうちいずれが良いかを検討する手段として、古典的な妥当性指標は現在でもその価値を減じてはいない。

前記の記載から分かるように、現在我々の目の前にある診断基準で完璧なものはない。そもそも医学のいかなる領域においても改善の余地のないくらい診断基準が備わっている疾患は例外的であろう。内科疾患の診断基準が国際的に変更になっても喫驚する者は少ない。診断基準の変更は学問の進歩を反映したものであり、むしろ歓迎すべき事柄である。旧来の診断基準の欠点を修正することで、診断基準はさらに臨床に有用なものとなって行く。例えば DSM-IV の特定の疾患の基準に不都合があるからといって、操作的診断基準の有効性を否定する傾向は、1 回の事故で自動車の存在を悪と決め付けることに似ていよう。

DSM-III-R では、診断基準を構成する症状のなかで当該症例で認められる症状数をもって大うつ病挿話の重症度を定める方針を取っている。つまり診断を下すに必要な症状の数に余分があることはほとんどなければ軽症、診断を下すに必要な症状の数より数個の余分があれば重症とされている。簡便さが魅力ではあるが、その根拠はなんであろうか。実証的所見は余りなかった。そこで Kitamura ら (1993) は 75 名のうつ病入院患者について RDC の診断基準症状数、Hamilton うつ病評価尺度の総合点による重症度、Global Assessment Scale (GAS) で測定した機能障害の程度の相関を見た。すると診断基準の症状数は Hamilton うつ病評価尺度 Hamilton Rating Scale for Depression (HRSD) 得点と過去 1 週間の GAS 得点とは 0.4 ほどの相関を示したが、本来強い相関を示すべき挿話中の最悪の週の GAS 得点とは 0.2 ほどの相関しか示さなかった。この所見は、RDC の診断基準症状数、Hamilton うつ病評価尺度得点、GAS 機能障害得点の 3 つは別個の現象を評価しているものではなく、診断基準症状数で重症度を規定することには慎重さが必要なことを示唆している (坂本, 1995)。

診断基準の改定は、当然であるがすでに述べた診断基準の妥当性をさらに強くすること

が条件である。こうした作業を行うに際しても、多くの専門家が情報と意見を誤解なく交換できるシステムが必要であり、そのひとつが操作的診断基準である。

## VII. まとめにかえて

### 操作診断の未来

本論ではいわゆる従来診断の問題に触れ、精神科診断においてなぜ操作診断が必要であるかについて述べてきた。最後に、操作診断の今後のあるべき姿についてしばらく、私見を述べたい。

これまでの歴史的流れから明らかのように、操作診断の基本的構成要件は症状とその持続期間であった。そして極力原因については（器質性疾患は例外として）言及しないよう心掛けてきていた。精神疾患の原因論については、生物学的、心理学的、社会学的に多くの説が対立し、結論に至っていなかった状況から考えれば、ともかくも操作診断を定着させる目的から考えればこの方策は正しかったといえる（Bayer ら, 1985）。しかし、これまで機能性と呼ばれていた精神疾患のかなりなものについてその生物学的、心理学的、社会学的原因（発症要因）が実証的に示されるようになってきた。こうした所見は操作診断の構成項目に注意深く組み込まれるようになるだろう。

本来症状のみによる診断は臨床単位として独立したものとは言いにくい。同一の原因を有していても異なる臨床像を示すこともあるし、同一の臨床像を示していても異なる原因を有していることも可能性として否定できないからである。同一の臨床像だからといって同一の病態ではない可能性について 1 例を挙げよう。精神分裂病の経過中に抑うつ症状が出現すること（Barnes ら, 1989）も、うつ病の患者に陰性症状が見られること（Kulhara ら, 1978）も、ともにしばしば報告されることから、両症状群の異同が問題になる。Kitamura ら（1991）は、RDC の定型うつ病（MDD）、躁病、分裂感情病うつ型（SADD）、分裂感情病躁型、精神分裂病（SCH）のいずれかの診断基準を満たす 193 名の入院患者について HRSD と SANS の評価を行い、両尺度得点の相関を各診断群ごとに求めた。すると両尺度間の相関は定型うつ病群では有意に高かったが、精神分裂病群ではほとんど相関していなかった（表 7）。患者全体で行った HRSD と SANS の全項目の因子分析では抑うつ症状と陰性症状は別の因子構造を示した。ここから推論できることは、抑うつ症状をと陰性症状は精神分裂病においては質的に異なる症状である一方、うつ病においては両症状は同じ事柄の別の表現であると考えられる。症状レベルでの評価方法の確度を高めても、そうした方法で評価した現象が均質の物であるという保証はない。複数の症状がひとつの因子構造を有することは、それが特定の病態生理を反映している可能性を示唆するものではあるが、それを保証するものでは決してない。今後の研究は、むしろ全体の枠組みのなかで個別症状の研究を見なおす必要があるであろう。どのような診断枠組みを準備すべきかについては慎重な

配慮が必要であろう。

表 7 を挿入

すでに述べたように、ある診断基準でくくられた集団についていくつかの外的妥当性指標を用いて他の集団と区別されるものであるという所見を提示することで、従来の操作診断の妥当性の検定は行われていた。しかし、過去 20 年近くのこうした診断研究は一定の成果を上げたものの、臨床的知識を大きく増加させるポテンシャルは失いかけている。今後は、ある外的指標の区分けを明示できるような臨床症状の組み合わせを模索することで新しい診断基準を作成することを真剣に考えなければならない。どの疾患領域についてどの外的指標を持ってくるかの良否が研究成果に大きく影響しよう。これまでの診断研究者と臨床家の「経験」と「勘」が問われるのである。

文 献

- Andreasen, N. C. (1984). *Scale for the Assessment of Negative Symptoms*. Iowa: Department of Psychiatry, Iowa University.
- Andreasen, N. C., Endicott, J., Spitzer, R. L. and Winokur, G. (1977). The family history method using diagnostic criteria. *Archives of General Psychiatry*, 34, 1229-1235.
- Andrews, G., Stewart, G., Morris-Yates, A., Holt, P. and Henderson, S. (1990). Evidence for a general neurotic syndrome. *British Journal of Psychiatry*, 157, 6-12.
- Appelbaum, P. S., Lidz, C. W. and Meisel, A. (1987). *Informed Consent: Legal Theory and Clinical Practice*. Oxford University Press: New York.
- Appelbaum, P. S., Mirkin, S. A. and Bateman, A. L. (1981). Empirical assessment of competency to consent to psychiatric hospitalization. *American Journal of Psychiatry*, 138, 1170-1176.
- Barnes, T. R. E., Curson, D. A., Liddle, P. F. and Patel, M. (1989). The nature and prevalence of depression in chronic schizophrenic in-patients. *British Journal of Psychiatry*, 154, 486-491.
- Bayer, R. and Spitzer, R. L. (1985). Neurosis, psychodynamics, and DSM-III: history of the controversy. *Archives of General Psychiatry*, 42, 187-196.
- Bean, G., Nishisato, S., Rector, N. A. and Glancy, G. (1994). The psychometric properties of the competency interview schedule. *Canadian Journal of Psychiatry*, 39, 368-376.
- Belter, R. W. and Grisso, T. (1984). Children's recognition of rights violation in counseling. *Professional Psychology*, 15, 899-910.
- Brakel, S. (1985). Involuntary institutionalization. In (eds.) S. J. Brakel, J. Parry and B. A. Weiner. *The Mentally Disabled and the Law*, Chicago: American Bar Foundation.
- Crow, T. J. (1980). Molecular pathology of schizophrenia: more than one dimension of pathology? *British Medical Journal*, 280, 66-68.
- Crow, T. J. (1985). The two-syndrome concept: origins and current status. *Schizophrenia Bulletin*, 11, 471-486.
- 千葉浩彦, 佐藤達哉, 北村俊則 (1987). RDC 定型うつ病の症状論的研究. *精神科治療学*, 2, 77-86.
- Endicott, J., Andreasen, N. C. and Spitzer, R. L. (1978). 加藤元一郎, 北村俊則 (訳) (1985). 家庭歴研究診断基準 (FH-RDC). *社会精神医学*, 8, 55-64.
- Feighner, J. P., Robins, E., Guze, S. B., Woodruff, J. P., Winokur, G. and Munoz, R. (1973). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, 26, 57-63.
- 藤縄昭, 北村俊則 (1991). 精神医学における操作的疾病分類学: 症状学の効用と限界. *精神科治療学*, 6, 549-554.
- 藤原茂樹 (1995). 一般人口におけるうつ病の頻度および発生要因に関する疫学的研究. *慶應医学*, 72, 511-528.

- 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 精神科診断の一致努度の比較. *臨床精神医学*, 17, 1081-1087
- 古川壽亮, 北村俊則 (1992). 精神医学における実証主義と操作化と計量的手法. *臨床精神病理*, 13, 89-100.
- Gardner, W., Hoge, S. K., Bennett, N., Roth, L. H., Lidz, C. W., Monahan, J. and Mulvey, E. (1993). Two scales for measuring patients' perceptions for coercion during mental hospital admission. *Behavioral Sciences and the Law*, 11, 307-321.
- Grisso, T. (1986). *Evaluation Competencies: Forensic Assessments and Instruments*. Plenum Press: New York.
- Grisso, T. and Appelbaum, P. S. (1992). *Manual of Understanding Treatment Disclosure*. Psychiatric Department, University of Massachusetts: Worcester.
- Harvard Law Review (1974). Developments in the law: civil commitment of the mentally ill. *Harvard Law Review*, 87, 1190-1406.
- Hasui, C., Hayashi, M., Tomoda, A., Kohro, M., Tanaka, K., Ageo, F. and Kitamura, T. (2000). Patients' desire to participate in decision-making in psychiatry: a questionnaire survey in Japan. *Psychological Reports*, 86; 389-399.
- Hasui, C., Sugiura, T., Tanaka, E., Sakamoto, S., Sugawara, M., Kitamura, T. and Aoki, Y. (1999). Reliability of childhood mental disorder diagnoses by Japanese psychologists. *Psychiatry and Clinical Neurosciences*, 53, 57-61.
- Hayashi, M., Hasui, C., Kitamura, F., Murakami, M., Takeuchi, M., Katoh, H. and Kitamura, T. (2000). Respecting autonomy in difficult medical settings: a questionnaire study in Japan. *Ethics and Behavior*, 10, 51-63.
- Hoge, S. K., Lidz, C., Mulvey, E., Roth, L., Bennett, N., Siminoff, L., Arnold, R. and Monahan, J. (1993). Patient, family, and staff perceptions of coercion in mental hospital admission: an exploratory study. *Behavioral Sciences and the Law*, 11, 281-293.
- Janofsky, J. S., McCarthy, R. J. and Folstein, M. F. (1992). The Hopkins Competency Assessment Test: a brief method for evaluating patients' capacity to give informed consent. *Hospital and Community Psychiatry*, 43, 132-136.
- 加藤知佳子, 渡辺さちや, 永田俊明, 北村俊則 (1990). 心理学専攻者による操作的診断基準の評定者間信頼度. *教育心理学研究*, 38, 413-417.
- Kendell, R. E. (1971). Psychiatric diagnosis in Britain and the United States. *British Journal of Hospital Medicine*, 6, 147-155.
- 木村敏 (1994). *病理を考える*. 東京, 岩波出版社.
- 北村總子 (1995). 疾患の医療における自己決定権とその制約原理. *法学政治学論究*, 24, 37-64.

- Kitamura, F., Tomoda, A., Tsukada, K., Tanaka, M., Kawakami, I., Mishima, S. and Kitamura, T. (1988). Method for assessment of competency to consent in the mentally ill: rationale, development, and comparison with the medically ill. *International Journal of Law and Psychiatry*, 21, 223-244.
- 北村俊則 (1988a). 症状測定 of 理論と実際. 東京: 海鳴社.
- 北村俊則 (1988b). 感情障害の診断基準 現代精神医学大系年刊版 88 B, 3-72. 東京: 中山書店.
- 北村俊則 (1990). 構造化面接の効果と限界. *精神科診断学*, 1, 473-478.
- 北村俊則 (1991). 神医学における操作的診断基準と DSM-III-R: 従来診断の実証的視点の欠落に触れて. *精神科治療学*, 6, 21-531.
- 北村俊則 (1993). DSM-IV の多軸診断について. *精神科診断学*, 38-488.
- Kitamura, T. (1994). The ICD-10 chapter on mental disorders versus conventional diagnosis in Japan. In J. E. Mezzich, Y. Honda. and M. C. Kastrup (eds.) *Psychiatric Diagnosis: A World Perspective*. New York: Springer.
- Kitamura, T., Kahn, A. and Kumar, R. (1984). Reliability of clinical assessment of blunted affect. *Acta Psychiatrica Scandinavica*, 69, 242-249.
- Kitamura, T. and Kitamura, F. (1993). *Structured Interview for Competency Incompetency Assessment Testing and Ranking Inventory (SICIATRI)*, Ichikawa: Department of Sociocultural Environmental Research, National Institute of Mental Health, NCMNP.
- 北村俊則, 北村總子 (1994). 精神科医療における治療同意の判断能力評価手法について. *精神科診断学*, 5, 233-242.
- Kitamura, T. and Kitamura, F. (2000). Reliability of clinical judgement of patients' competency to give informed consent: a case vignette study. *Psychiatry and Clinical Neurosciences*, 54, 245-247.
- 北村俊則, 丸山普, 大塚俊男, 下仲順子, 中里克治 (1985). DSM-III 痴呆診断および柄澤式ぼけ評価尺度の評定者間信頼度. *老年精神医学*, 12, 774-777.
- Kitamura, T., Nakagawa, Y. and Machizawa, S. (1993). Grading depression severity by symptom scores: is it a valid method for subclassifying depressive disorders? *Comprehensive Psychiatry*, 34, 280-283.
- Kitamura, T., Nakamura, M., Miura, I. and Fujinawa, A. (1997) Symptoms of neuroses: profile patterns and factor structure of clinic attenders with non-psychiatric disorders. *Psychopathology*, 30, 191-199.
- Kitamura, T., Okazaki, Y., Fujinawa, A., Yoshino, M. and Kasahara, Y. (1995). Symptoms of psychoses: a factor-analytic study. *British Journal of Psychiatry*, 166, 236-240.
- 北村俊則, 島悟 (1984). 慢性精神分裂病における陰性症状評価尺度の評価者間信頼度. *慶應医学*, 61, 277-283.

- 北村俊則, 島悟 (1986). 感情病および精神分裂病様面接基準 (SADS) 研究用診断基準 (RDC) の評価者間信頼度. *精神医学*, 28, 41-45.
- 北村俊則, 島悟, 加藤元一郎, 岩下覚, 神庭重信, 白土俊幸, 藤原茂樹, 市川洋子, 加藤雅高, 神庭靖子, 飯野利仁, 生田憲正, 宮岡等, 竹井茂樹, 樋山光教, 越川裕樹, 柘野雅之, 千葉忠吉 (1987). 慢性精神分裂病に対する評価尺度の信頼度. *精神医学*, 29, 933-940.
- 北村俊則, 島悟, 加藤元一郎, 岩下覚, 神庭重信, 白土俊幸, 藤原茂樹, 市川洋子, 加藤雅高, 神庭靖子, 飯野利仁, 生田憲正, 宮岡等, 竹井茂樹, 樋山光教, 越川裕樹, 柘野雅之, 千葉忠吉 (1989). 慢性精神分裂病の陽性症状と陰性症状. *精神医学*, 31, 131-136.
- 北村俊則, 島悟, 崎尾英子, 加藤元一郎 (1984a). 症状要旨を用いたファイナー診断基準の信頼度検定. *精神医学*, 26, 1203-1207.
- 北村俊則, 島悟, 崎尾英子, 加藤元一郎 (1984b). 症状要旨法による家族歴研究診断基準 (FH-RDC) の信頼度検定. *社会精神医学*, 7, 308-312.
- Kitamura, T., Shima, S., Sakio, E. and 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. and Kato, M. (1989a). Psychiatric diagnosis in Japan. I. a study on diagnostic labels used by practitioners. *Psychopathology*, 22, 239-249.
- Kitamura, T., Shima, S., Sakio, E. and Kato, M. (1989b). Psychiatric diagnosis in Japan. II. reliability of conventional diagnosis and discrepancies with Research Diagnostic Criteria diagnosis. *Psychopathology*, 22, 250-259.
- 北村俊則, 島悟, 崎尾英子, 高橋龍太郎, 加藤元一郎 (1984). 症状チェックリストとしての「感情および精神分裂病用面接基準」(SADS) と「現在症診察表」(PSE) の比較. *臨床精神医学*, 13, 293-299.
- Kitamura, T., and Suga, R. (1991). Depressive and negative symptoms in major psychiatric disorders. *Comprehensive Psychiatry*, 32, 88-94.
- Klerman, G. L., Endicott, J., Spitzer, R. and Hirschfeld, R. M. A. (1979). Neurotic depressions: a systematic analysis of multiple criteria and meanings. *American Journal of Psychiatry*, 136, 57-61.
- Kraus, A. (1996). 角田京子, 津田均 (訳) 精神科診断の疾病分類における直感の重要性. *精神医学*, 38, 501-510.
- Kulhara, P., and Chadda, R. (1987). A study of negative symptoms in schizophrenia and depression. *Comprehensive Psychiatry*, 28, 229-235.
- Landis, J. R. and Koch, G. G. (1997). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Larkin, B. A., Copeland, J. R. M., Dewey, M. W., Davidson, I. A., Sanders, P. A., Sharma, V. K., McWilliam, C. and Sullivan, C. (1992). The natural history of neurotic disorder in an elderly

- urban population: findings from the Liverpool longitudinal study of continuing health in the community. *British Journal of Psychiatry*, 160, 681-686.
- Leff, J. (1977). International variation in the diagnosis of psychiatric illness. *British Journal of Psychiatry*, 131, 329-338.
- Lenzenweger, M. F. and Dworkin, R. H. (1996). The dimensions of schizophrenia phenomenology: not one or two, at least three, perhaps four. *British Journal of Psychiatry*, 168, 432-440.
- Liddle, P. F. (1987). The symptoms of chronic schizophrenia: a re-examination of the positive-negative dichotomy. *British Journal of Psychiatry*, 151, 145-151.
- Liddle, P. F. and Barnes, T. R. D. (1990). Syndromes of chronic schizophrenia. *British Journal of Psychiatry*, 157, 558-561.
- Lidz, C. W., Hoge, S. K., Gardner, W., Bennett, N., Monahan, J., Mulvey, E. P. and Roth, L. H., (1995). Perceived coercion in mental health hospital admission; pressure and process. *Archives of General Psychiatry*, 52, 1034-1039.
- Likert, R. A. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1-55.
- Looney, J. G. and Gunderson, E. K. E. (1979). A longitudinal study of neurosis in young men. *Hospital & Community Psychiatry*, 30, 470-474.
- McDonald-Scott, 北村俊則, 島悟 (1986a). 精神医学研究における SADS の役割: I. 臨床評価の不一致の原因と SADS の成立. *精神医学*, 28, 743-749.
- McDonald-Scott, 北村俊則, 島悟 (1986b). 精神医学研究における SADS の役割: II. SADS の信頼性と他の面接基準との比較. *精神医学*, 28, 889-895.
- Miller, R. D. (1994). The US Supreme Court looks at voluntariness and consent. *International Journal of Law and Psychiatry*, 17, 239-252.
- Monahan, J., Hoge, S. K., Lidz, C., Roth, L. H., Bennett, N., Gardner, W., and Mulvey, E., (1995). Coercion and commitment: understanding involuntary mental health hospital admission. *International Journal of Law and Psychiatry*, 18, 249-263.
- 中安信夫 (1991). DSM-III-R に見る臨床診断的視点の欠落: 精神医学における臨床診断のあり方に触れて. *精神科治療学*, 6, 511-520.
- 西丸四方 (1957). *精神医学入門*. 東京: 南山堂.
- Overall, J. E. and Gorham, D. R. (1962). The brief psychiatric rating scale. *Psychological Report*, 10, 799-892.
- Peralta, V., De Leon, J., and Cuesta, M. J. (1992). Are there more than two syndromes in schizophrenia? a critique of the positive-negative dichotomy. *British Journal of Psychiatry*, 161, 335-343.
- Robins, E. and Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *American Journal of Psychiatry*, 126, 983-987.

- Robins, L., (1978). Sturdy childhood predictors of adult antisocial behavior: replications from longitudinal studies. *Psychological Medicine*, 8, 611-622.
- Robins, L.N. and Barrett, J. E. (eds.) (1989). *The Validity of Psychiatric Diagnosis*. New York: Raven Press.
- Roth, L., Lidz, C. W., Meisel, A., Soloff, P. H., Kaufmann, K., Spiker, D. G. and Foster, F. G. (1982). Competency to decide about treatment or research: an overview of some empirical data. *International Journal of Law and Psychiatry*, 5, 29-50.
- 坂本真士 (1995). 憂うつと不安における正常と異常. *精神科診断学*, 6,131-142.
- Saugstad, L. F. and Odegard. O. (1983). Persistent discrepancy in international diagnostic practice since 1970. *Acta Psychiatrica Scandinavica*, 68, 501-510.
- Sims, A., (1986). Perspectives in the study of neurosis in contemporary psychiatric practice. *Psychiatric Development*, 4, 273-287.
- Spitzer, R. L. Endicott, J. and Robins, E. (1975). Clinical criteria for psychiatric diagnosis and DSM-III. *American Journal of Psychiatry*, 132, 1187-1192.
- Spitzer, R. L., Endicott, J. and Robins, E. (1978a). *Research Diagnostic Criteria (RDC) for a selected group of functional disorders*. 3rd ed. New York: Biometric Research, New York Psychiatric Institute.
- Spitzer, R. L., Endicott, J. and Robins, E. (1978b). *Schedule for Affective Disorders and Schizophrenia (SADS)*, (3rd ed.), New York: Biometric Research, New York Psychiatric Institute.
- 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.
- Sugiura, T., Hasui, C., Aoki, Y., Sugawara, M., Tanaka, E., Sakamoto, S. and Kitamura, T. (1998). Japanese psychology students as psychiatric diagnosticians: Application of criteria of mood and anxiety disorders to written case vignettes using the RDC and DSM-IV. *Psychological Reports*, 82, 771-781.
- Thompson, P. A. and Meltzer, H.Y. (1993). Positive, negative, and disorganization factors from the Schedule for Affective Disorders and Schizophrenia and the Present State Examination: a three-factor solution. *British Journal of Psychiatry*, 163, 344-351.
- Tomoda, A., Yasumiya, R., Sumiyama, T., Tsukada, K. , Hayakawa, T., Matsubara, K., Kitamura, F. and Kitamura, T. (1997). Validity and reliability of Structured Interview for Competency Incompetency Assessment Testing and Ranking Inventory (SICIATRI). *Journal of Clinical Psychology*, 53, 443-450.
- Weisbard, A. J. (1986). Informed consent: the law's uneasy compromise with ethical theory. *Nebraska Law Review*, 65, 749-767.

- Weithorn, L. A. and Campbell, S. B. (1982). The competency of children and adolescents to make informed treatment decision. *Child Development*, 53, 1589-1598.
- Wing, J. K. (1961). A simple and reliable subclassification of chronic schizophrenia. *Journal of Mental Sciences*, 107, 862-875.
- Wing, J. K., Cooper, J. E. and Sartorius, N. (1974). *Measurement and Classification of Psychiatric Symptoms: An Introduction to the PSE and CATEGO programme*. Cambridge: University Press.

表1. RDC診断と日本の従来診断

従来診断	RDC診断				
	精神分裂病	分裂感情病	定型うつ病	準定型うつ病	その他
精神分裂病	5	5	2	0	0
躁うつ病	0	0	7	0	0
妄想状態	1	0	0	0	0
神経症	0	0	0	0	0
人格障害	0	0	0	1	1
アルコール症	0	0	1	0	0
合計	6	5	10	1	1

Kitamura ら(1986)より改変して引用

表 2. 症例要旨法による日本人精神科医の診断一致率

病名	診断一致率 ( I C C )
精神分裂病	0.75
非定型精神病	0.26
うつ病+抑うつ神経症	0.73
躁病	0.22
妄想状態	0.45
その他の精神病	0.06
神経症 (非抑うつ性神経症)	0.66
人格障害	0.54
境界例	0.12
その他の疾患	0.19
非疾患 (正常)	0.52

Kitamura ら(1989b) より改変して引用

表 3. 直接面接法による日本人精神科医の診断一致率

病名	診断一致率 (κ)	
	同席面接	再試験面接
精神分裂病	0.90	0.82
非定型精神病	0.67	0.56
うつ病	0.75	0.50
躁病	0.80	0.26
不安神経症	0.84	0.23
抑うつ神経症	0.55	0.52
心気症	0.72	0.66
心因反応	0.81	0.70
器質性疾患	0.69	0.82

Kitamura ら(1994)より改変して引用

表 4. 症例要旨法による精神疾患の評定者間信頼度

RDC もしくは DSM-IV による診断名	精神科医	心理学専攻 者	心理学専攻 者	心理学専攻 者
	(n = 12)	(n = 7)	(n = 3)	(n = 11)
	須賀ら, 1987	藤原ら, 1988	加藤ら, 1990	Sugiura ら, 1998
精神分裂病	0.82	0.50	0.67	—
分裂感情病・躁型	0.74	0.64	0.67	—
分裂感情病・抑うつ型	0.72	0.49	0.52	—
躁病	0.91	0.92	0.70	0.73
定型うつ病	0.82	0.75	0.79	0.72
準定型うつ病	0.56	0.43	0.77	—
アルコール症	0.93	1.00	1.00	—
パニック障害	—	—	—	0.92
強迫性障害	—	—	—	0.66
恐怖性障害	—	—	—	0.73
全般性不安障害	—	—	—	0.51

表5. 症例要旨法による精神疾患の正答との一致率

RDC もしくは DSM-IV による診断名	精神科医	心理学専攻 者	心理学専攻 者	心理学専攻 者
	(n = 12)	(n = 7)	(n = 3)	(n = 11)
	須賀ら, 1987	藤原ら, 1988	加藤ら, 1990	Hasui ら, 1999
精神分裂病	0.80	0.60	0.79	—
分裂感情病・躁型	0.82	0.72	0.85	—
分裂感情病・抑うつ型	0.51	0.38	0.64	—
躁病	0.95	0.95	—	—
定型うつ病	0.89	0.78	0.87	0.91
準定型うつ病	0.72	0.48	—	—
アルコール症	0.96	1.00	1.00	—
パニック障害	—	—	—	—
強迫性障害	—	—	—	—
恐怖性障害	—	—	—	—
全般性不安障害	—	—	—	0.64

表6. 病名診断と診断的フォーミュレーション

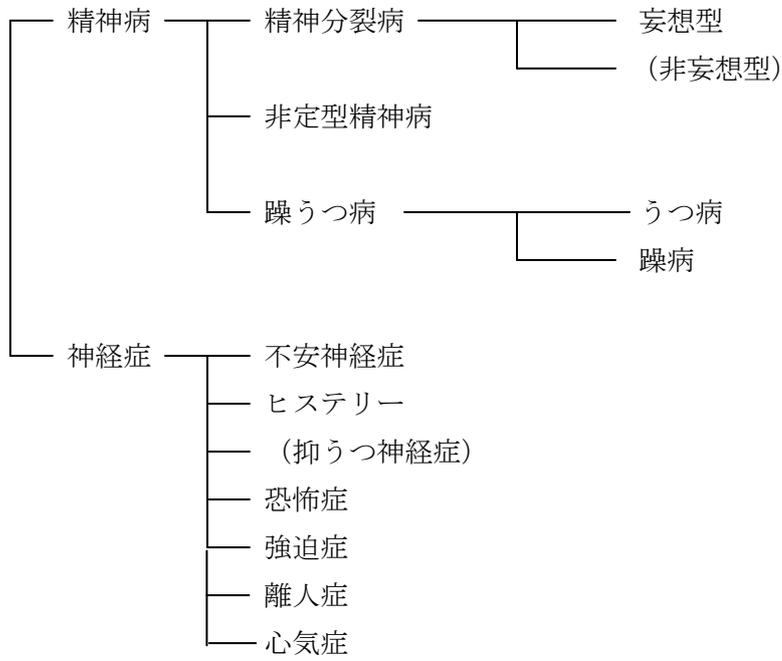
	病名診断	診断的フォーミュレーション
対象	疾患を有する者の集団	個別の症例
方向性	後方視的	前方視的
観点	事後的	事前的
作業	事実認定	仮説認定
目的	疾患単位の同定	予後判定・治療方針決定
操作化	可能	困難

表 7. 重症精神疾患における抑うつ症状と陰性症状

HRSDとの相関	M D D (n=75)	S A D D (n=24)	S C H (n=57)
<b>SANS</b>			
総合得点	0.51***	0.64**	0.15
情動の平板化・情動鈍麻	0.52***	0.67***	0.20
思考の貧困	0.47***	0.66***	0.10
意欲・発動性欠如	0.50***	0.50**	0.29*
快感消失・非社交性	0.55***	0.67***	0.29*
注意の障害	0.42***	0.54**	0.12

Kitamura (1991) より引用 \* P<.05; \*\* P<.01; \*\*\* P<.001

図1. 日本の精神科医が共通して使用している従来診断名



Kitamura ら(1989a) より改変して引用

図2. 構造化面接と診断基準

