

操作的診断基準と疾病分類の 功罪

過去・現在・未来

北村俊則*

抄録

精神科医の診断の不一致が問題となり、操作的診断基準が作成された。しかし、疾患単位の妥当性の議論が積み残し課題となった。Taxometrics による研究ではほとんどの診断名は範疇的性質ではなく、次元的なものであることが確認された。現在の診断学は因子分析を行うことで症状構造を確認し、クラスター分析、潜在クラス分析などで質的に異なる群を同定することが主流になっている。さらに、精神疾患の病名が直接に偏見の形成に寄与していること、さらに多数が「健康的」で、少数が「病的」という考え方の矛盾を考えると、精神科診断の現代的課題は未解決の部分が多い。

KEY WORDS

操作的診断基準 (operational diagnostic criteria), 疾病分類 (nosology), 精神医学の歴史 (history of psychiatry)

歴史的展望

かつて精神疾患の診断は精神科医療や精神医学の外部の人間には、そして内部の人間にも分かりにくいものであった。筆者が大学の精神神経科医局に入局したときも、精神症状の記載方法(症状学)はかなり詳細に指導を受けたが、疾患の診断の方法については指導されなかった記憶がある。その時代は、そもそも各疾患のイメージについての記載はあったが、診断手続き(アルゴリズム)は存在していなかった。そのため、精神科医が下す診断は一致しないことがむしろ普通であった。同じ診断名(たとえば当時の「精神分裂病」)でも、臨床医によってイメージする範囲が大きく異なっていることが報告された¹⁾。そして、当然に評定者間一致率は低いものであった。当時、我々が行った調査でも、臨床精神科医のさまざまな病態の診断一致率²⁾を intraclass correlation (ICC) で求めると、すべて当時の診断名で、精神分裂病 0.75, 感情障害 0.51, 非定型精神病 0.26, 妄想状態 0.45, 神経症 0.64, 人格障害 0.54, 境界例 0.12, 正常 0.52 であり、うつ病と抑うつ神経症を合わせた場合、やっと 0.73 にまで上昇する³⁾。ICC は通常、研

* 北村メンタルヘルス研究所(〒151-0063 東京都渋谷区富ヶ谷 2-26-3A)

究では0.8以上、臨床では0.7以上が望ましいと考えられる。こう考えると、当時の日本国内の精神科医による診断は著しく一致率の低いものであり、これは海外でも同様であった。

これは外部からの批判的になった。精神科医療において強制入院(本人の希望に反した入院)が多く、多くの国で日常的に行われていたこともこれに拍車をかけた⁴⁾。上記の研究で「正常」、つまり精神疾患でないものの診断の一致率がわずか0.52であったことは大きな問題であった。この汚名を晴らし、名誉を挽回するには、精神科医の診断がともかく一致することが必要であった。つまり、評定者間一致率を担保することが必要であった。この努力の産物が操作的診断基準であった。最も大掛かりな取り組みがDSM-IIIであった。当時行われた多くの研究が、信頼度(つまり診断一致率)に集中していたのもこうした背景があったからである。一方、診断概念が妥当かどうかの議論は進んでいなかった。

DSM-IIIの「下敷き」になった操作的診断基準はファイナー基準である⁵⁾。この時点で、操作的診断基準は今後の改善を前提とした一時的取り決めであるとの認識が研究者の間では認められていた。Feighnerら⁵⁾は論文の初めに次のように述べている。

This communication is meant to provide common ground for different research groups so that diagnostic definitions can be emended constructively as further studies are completed. (建設的に病名診断の定義を作成することで研究がさらに先に進むよう、さまざまな研究グループに共通する基盤を提供することが本研究の意図である)

さらに以下のように付け加えている。

These criteria are not intended as final for any illness. (これらの診断基準は、いかなる疾患についても最終的なものとして意図されていない)

しかし、いったんDSM-IIIが「錦の御旗」になってしまうと、診断基準の作成は保守的になり、変更を拒むようになった。精神科医たちは、DSM-IIIが規定しているのが疾患であり、そこにある疾患を我々が規定するという哲学を忘れてしまった。特に、診断基準が設定している各疾患の構成概念の検討、すなわち疾患概念の妥当性の探査に消極的になったのである。これが操作的診断基準の(1970年代の診断学の研究者たちが思いつかなかった)副作用であった。

操作的基準の妥当性：次元性

ところで、精神疾患は本当に存在するのであろうか。身体疾患であれ精神疾患であれ、他から質的に区別できる存在が確認されるときに、疾患概念が確認されたと考えられる⁶⁾。精神医学では、ある変数の分布に希少点⁷⁾が存在すれば、疾患の存在を想定してもよいと考える。ある臨床指標に基づいてケースの分布図を見て、一峰性の図であれば正規分布と考えることが可能だが、**図1**のように二峰性の分布であれば、2つの質的に異なる母集団から集まったサンプルであると推定できる。この時に、疾患概念の存在が推定され得るのである。つまり、評価している変数(症状などの臨床指標)の正規性が破綻していることが疾患の存在を示唆している。ところで、ほとんどの疾

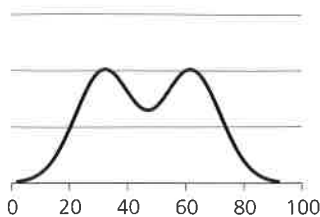


図1 二峰性の分布
(文献⁹⁾より改変)

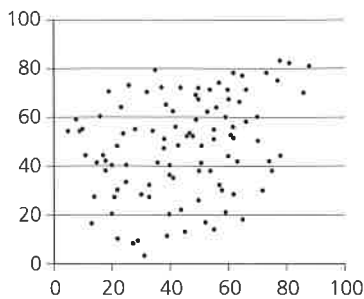


図2 多変量正規性がある
(文献⁹⁾より改変)

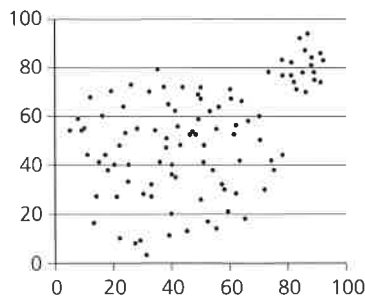


図3 多変量正規性がない
(文献⁹⁾より改変)

患の診断基準は複数の症状あるいは臨床指標から作られている。仮に2つの指標があったとしよう。いずれの変数も、それだけ見れば正規分布を示している。そこで、平面上にケースをプロットする。そして図2のような分布であれば、多変量正規性があると考えられる。しかし、図3に見るように、右上の部分の少数のケースが他の多数のケースから隔絶しているように見える。このことを多変量正規性が欠けているという。

身体的病理所見で疾患の存在を同定できないのであるから、精神疾患の存在はそれを特徴付ける臨床指標(たとえば症状、持続期間、検査所見)の多変数正規性が崩れた際に推定できる。もし、多変数正規性が存在するのであれば、臨床指標は次元のなものであり、低い極から高い極に連続的に配置されている。そして、(科学的意味での)区分点は存在しない。この場合、臨床実務から区分点を設定しても、それはあくまで便宜的なものである。このような希少点を見つける方法には taxometrics のさまざまな手法^{8,9)}、潜在クラス分析¹⁰⁾、クラスター分析のさまざまな手法¹¹⁾がある。Taxometrics の種々の手法による研究は、当該疾患概念が次元のなものであるか範疇的のものを峻別できる。全体の傾向をまとめる指標が comparison curve fit index (CCFI) であり、これが0.5を下回ると次元的现象であると考えられる。Ruscio ら¹²⁾は CCFI が0.5を下回る診断は物質障害だけであり、その他の疾患はすべて次元のなものであると述べている。

診断概念の妥当性を検討する作業は統計学的には構成概念妥当性 (construct validity) の検討に等しい。これまでその外的指標として家族負因、内分泌変化、脳波の変化、脳の画像所見、脳血流量、心理検査、各種の社会的要因など、おびただしい数の要因が研究されてきた。ところがいずれも決定的な所見を得ていない。ほとんどの診断基準は「あり・なし」の名義尺度である。たしかに、臨床場面では「薬物を投与するかしないか」「入院治療か外来治療か」「障害者年金を与えるか与えないか」など、「はい・いいえ」の判断が求められる状況があり、何らかの区分点を設定せざるを得ない。しかし、これらはすべて実務上の便宜を考えた処置である。こうした便宜的処置による二分類が定着すると、臨床かあたたかもそこに質的区分点があるかのような錯覚に陥るのである。この錯覚を権威が補強する。妥当性を求める概念が本当は比尺度であるにもかかわらず人為的(恣意的)区分点をつけてあたかも名義尺度であるかのように扱えば、出るべき所見も有意のものとして出なくなるのは当然である。精神科診断学の

大きな間違いの1つがここにある。

従来、精神病理学は症状学(symptomatology)と疾病分類学(nosology)の両輪で行われていた。これは精神科に限られたものではなく、医学全般に行われてきた古典的思考である。これを近代の診断学に翻訳すれば、症状学が因子分析であり、疾病分類学がクラスター分析(および類似の手法)である。現行の精神科診断学の報告と精神科医療の著しい乖離が、双方の発展を阻害しているのである。

今後の方向性

操作的診断基準の功罪を考え、今後の方向性を考えるのであれば、まずは原点に戻り、人間の心理活動(従来の症状)の実証的評価の再確認をしなければならない¹³⁾。評価の対象はさまざまな症状である。しかし、同じ概念の症状であっても、その表現型は個人で異なる。そして、その評価方法は面接者の評価であったり、患者(対象者)の自己申告であったりする。評価法によって捉え方も変わるものである。そこで症状構造の検討は multitrait-multimethod の観点から行うことが肝要である¹⁴⁾。多数項目で評価する概念は探索的因子分析で求め、そのモデルの交差妥当性を確認的因子分析で明らかにする。しかし確認的因子分析における適合度に不具合が生じることもある。そのときは複数項目から合成変数を作成し、それを因子分析の解析の対象とする。これをパーセル化といい、その手法は重要な役割を持つことになる¹⁵⁾。いくつかの項目(症状)が単一次元の性質を持っていることは、explained common variance(ECV)の手法で確認することも有用であろう¹⁶⁾。当然、そうした概念測定法の配置不変性・測定不変性・構造不変性の確認も必須である¹⁷⁾。特に、個人の属する属性が異なる場合には配置不変性・測定不変性・構造不変性が認められなければ、そうした症状測定の妥当性が疑われる。その上で、多くの観察項目(症状)の因子構造の決定を行う。この際、従来行われてきた単純な一次因子構造ではなく階層的因子構造(二次因子構造や階層的因子分析)は考慮にいれなければならない¹⁸⁾。

たとえば産後うつ病の症状は anhedonia, anxiety, dysphoria という3因子から構成されており、さらに全体の重症度を表現する全般的因子(general factor : g factor)が存在することが明らかになってきた¹⁹⁾。ここまでは症状の分類を行うものである。

症状の因子構造が確認できた後に、クラスター分析などの手法によって対象者の分類を行う。これが疾患単位の設定である。多くの精神疾患は次元的なものであると述べた。しかし、近年の産後ボンディング障害のクラスター分析で明確な(病的)ボンディング障害のクラスターが同定されている²⁰⁾。また我々は子どもの気質をEASIで測定し、まずその因子構造が理論通りの4因子になることと、いくつかの項目は削除すべきことを見出し²¹⁾、次にこの因子構造から導き出される下位尺度得点を用いて、two-step cluster 分析によって児の気質の特徴から対象児が4グループに分類できることを報告した²²⁾。クラスター分析はクラスター数の設定に恣意的な部分があることが批判されていた。two-step cluster 分析はこれを排除するため一定のアルゴリズムでクラスター数を決定できる方法である。こうした精緻な研究は多く行われなければならない。

頑強性の確認できた因子構造とクラスター分析による対象者の分類が行われたなら、そこで外的基準との関連を見た構成概念妥当性の検討を行う。こうした作業が新しい(そして自然科学的)診断基準作成の動きに繋がりを、こうした方向性の研究が真に臨床に有用な情報を提供するのである。かつての、勘と経験と権威による診断学が、実証的データに基づきそれに移行できるのである。

自然科学から社会科学へ

一部の精神現象に診断を与えることは、本来、支援方策(入院、加療、退院など)の決定のため行われていた。しかし、必然的といってよいほど、精神疾患の名称は偏見を惹起してきた。同一の統合失調症事例を提示し、それに対する偏見的態度を調査した研究がある²³⁾。ここでは一般人の対象者を無作為に2群に分け、一方の群には事例の経過のみを提示し、もう一方の群には同一の経過の最後に「そしてAさんは精神科を受診したところ、『統合失調症』と診断されました」という文章を加えた。偏見的態度は有意に後者の群で高かった。つまり、疾患に対する偏見が診断名そのものからきていたのである。診断名を作成すること自体が偏見を生んでいることは、精神科医療にとって見過ごせない事実である。診断が与えられることは、その当事者が治療の対象であることを意味し、治療が必要なのはそれが不健康であるという含意がある。

仮にある心理状態(のグループ)を用いて複数のクラスターが確認できた際に、少数のグループを病的と判断する根拠は脆弱である²⁾。これはう歯(虫歯)を考えれば明らかであろう(虫歯のない人のほうが病気?)。また、機能障害の存在が精神疾患の基準に組み込まれているが、機能障害は生活する社会(対人・文化)環境の中で規定されるものである。多数例が作っている環境に、少数例の人々が「生きづらさ」を感じた、つまり機能障害があるといっても、tautology であるという批判にさらされるだけである。同性愛や性同一性障害をみれば明白である。この点がかつての反精神医学の主張であった²⁴⁾。加えて、多数例の人々は少数例の人々の「医療と保護」を目的に、医療を強制する。強制的根拠が精神疾患の存在であれば、ふたたび tautology であるという批判にさらされる。つまり、精神疾患の存在という仮説の妥当性が臨床家に与えられた大きな課題になっている。

強制医療の根拠は精神疾患の存在を根拠にすることはできない。精神疾患といわれるものが名義尺度でなく比尺度であることから、いずれに区分点(「これ以上の重症度であれば強制入院をさせる」)を設定しても恣意的である。強制医療の根拠はむしろ、治療同意判断能力(民事上の意思能力)で決めるべきである。これは疾患の性質や重症度と強い関連を持つものではない。治療同意判断能力の評価も精神科診断の大きな課題として残っている。

文献

- 1) Kitamura T, Shima S, Sakio E, et al : Psychiatric diagnosis in Japan. I. a study on diagnostic labels used by practitioners. *Psychopathol* 22 : 239-249, 1989

- 2) 北村俊則：精神症状測定の実論と実際—評価尺度、質問票、面接基準の方法論的考察 第2版。海鳴社，1995
- 3) Kitamura T, Shima S, Sakio E, et al : Psychiatric diagnosis in Japan. II. reliability of conventional diagnosis and discrepancies with RDC diagnosis. *Psychopathol* 22 : 250-259, 1989
- 4) 北村総子，北村俊則：精神科医療における患者の自己決定権と治療同意判断能力。学芸社，2000
- 5) Feighner JP, Robins E, Guze SB, et al : Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry* 26 : 57-63, 1973
- 6) 北村俊則：精神科診断学概論—病理所見のない疾患の概念を求めて。北村メンタルヘルス研究所，2013
- 7) Kendell RE, Brockington IF : The identification of disease entities and the relationship between schizophrenic and affective psychoses. *Br J Psychiatry* 137 : 324-331, 1980
- 8) Meehl PE : Bootstraps taxometrics : solving the classification problem in psychopathology. *Am Psychologist* 50 : 266-274, 1995
- 9) 北村俊則：精神に疾患は存在するか。星和書店，2017
- 10) Lubke GH, Miller PJ : Does nature have joints worth carving? A discussion of taxometrics, model-based clustering and latent variable mixture modelling. *Psychol Med* 45 : 705-715, 2015
- 11) Borgen FH, Barnett DC : Applying cluster analysis in counselling psychology research. *J Couns Psychol* 34 : 456-468, 1987
- 12) Ruscio J, Ruscio AM, Meron M : Applying the bootstrap to taxometric analysis : generating empirical sampling distributions to help interpret results. *Multivariate Behav Res* 42 : 349-386, 2007
- 13) 北村俊則：精神・心理症状学ハンドブック 第3版。日本評論社，2013
- 14) Campbell DT, Fiske DW : Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 56 : 81-105, 1959
- 15) Little TD, Cunningham WA, Shahar G : To parcel or not parcel : exploring the question, weighting the merits. *Structural Equation Modeling* 9 : 151-173, 2002
- 16) Rodriguez A, Reise SP, Haviland MG : Evaluating bifactor models : calculating and interpreting statistical indices. *Psychol Methods* 21 : 137-150, 2016
- 17) Vandenberg RJ, Lance CE : A review and synthesis of the measurement invariance literature : suggestions, practices, and recommendations for organizational research. *Organizational Res Meth* 3 : 4-70, 2000
- 18) Brunner M, Nagy G, Wilhelm O : A tutorial on hierarchically structured constructs. *J Pers* 80 : 796-846, 2011
- 19) Hada A, Kubota C, Imura M, et al : The Edinburgh Post-natal Depression Scale : model comparison of factor structure and its psychosocial correlates among mothers at one month after childbirth in Japan. *Open Fam Stud J* 11 : 1-17, 2019
- 20) Matsunaga A, Takauma F, Tada K, et al : Discrete category of mother-to-infant bonding disorder and its identification by the Mother-to-Infant Bonding Scale : a study in Japanese mothers of a 1-month-old. *Early Hum Dev* 111 : 1-5, 2017
- 21) Ohashi Y, Kitamura T : The EASI : factor structure and measurement and structural invariance between the parent's gender, the child's age, and two measurement time points. *Psychol* 10 : 2177-2189, 2019
- 22) Ohashi Y, Kitamura T : Typology of temperament of Japanese children aged 3 and 4. *Psychol* 11 : 955-965, 2020
- 23) Matsunaga A, Kitamura T : The effects of symptoms, diagnostic labels, and education in psychiatry on the stigmatization towards schizophrenia : a questionnaire survey among a lay population in Japan. *Mental Illness* 8 : 6344, 2016
- 24) Szasz T : *The Myth of Mental Illness*. Harper, New York, 1974

Title

Pros and Cons of Operational Diagnostic Criteria and Classification : Past, Present, and Future

Author

Toshinori Kitamura : Kitamura Institute of Mental Health, Tokyo, Japan