

Validity and Reliability of Diagnosis from Hospital Records for Borderline Personality Disorder

Eita TONAI, Naoki UCHIDA, Kosuke YOSHIDA,
Nobuaki ETO, Ryoji NISHIMURA

Department of Psychiatry, Faculty of Medicine, Fukuoka University

Abstract: There are many controversies about long-term outcome of borderline personality disorder. We have to establish a retrospective diagnosis method with reliability in order to do a retrospective long-term follow-up study. Thus, the purpose of this study is that examination of validity and reliability of diagnosis from hospital records for borderline personality disorder reliability by retrospective DIB (Diagnostic Interview for Borderlines).

The subject is the patients who were admitted to Department of Psychiatry, Fukuoka university hospital more than eight weeks from April, 2000 to June, 2003 and their age was a 18 years old - 35 years old. We compared the results of DIB which were evaluated from hospitalization medical records of these patients and the results of DIB evaluated from real interview. As a result, DIB which was marked from hospitalization records showed enough concordance with a result of DIB by the interviews. It was suggested that validity of diagnosis from hospital records for Borderline personality disorder by DIB. In addition, inter-rater reliability of diagnosis from hospital records for Borderline personality disorder was suggested when two raters independently evaluated DIB from hospitalization records. Thus, it was suggested that an investigator could evaluate DIB from clinical record equally. It was shown that it was possible to use DIB from medical record of hospitalization as a diagnosis method of a retrospective study of borderline personality disorder.

Key words : Borderline personality disorder, retrospective diagnosis, medical record, inter-rater reliability, validity

境界性人格障害 (BPD) の入院診療録に基づく診断の 妥当性と信頼性に関する研究

藤内 栄太, 内田 直樹, 吉田 公輔,
衛藤 暢明, 西村 良二

福岡大学精神医学教室

要旨: 境界性人格障害の長期予後については議論のあるところが多い。長期の後ろ向き研究を行なうためには信頼性のある後ろ向き診断方法を確立する必要がある。そこで本研究では、入院診療録を境界パーソナリティ診断面接質問紙 (Diagnostic Interview for Borderlines: 以下 DIB) によって後ろ向きに境界性人格障害を診断することの外的妥当性と評価者間信頼性を検討することを目的とした。

平成 12 年 4 月から平成 15 年 6 月までに福岡大学病院精神神経科に 8 週間以上入院しており、年齢が 18 歳 ~ 35 歳の患者を対象とした。これらの患者の入院診療録から評価した DIB と実際の面接による DIB の結果を比較した。その結果、入院診療録から採点した DIB は面接による DIB の結果と十分な一致 ($\kappa = 0.71$)

を示した。このことより診療録を用いてDIBによって境界性人格障害を診断することの妥当性が示唆された。また、別々の評価者が独立に入院診療録からDIBを評価した場合にも十分な評価者間信頼性 ($\kappa = 0.72$) を示した。以上のことから複数の調査者が同様に診療録からDIBを評価できることが支持された。境界性人格障害の後ろ向き追跡研究の診断方法として診療録によるDIB診断を用いることの可能であることが示された。

キーワード：境界性人格障害，後ろ向き診断，診療録，信頼性，妥当性

はじめに

近年、わが国における境界性人格障害患者 (Borderline Personality Disorder: 以下、BPDと略) は増加していると思われるが¹⁾、現在有効な治療法が確立されていない上に、疾患の概念やその長期経過についても議論のあるところが多い²⁾。DSM-IV³⁾ や ICD-10⁴⁾ などの操作的診断基準に適応されている多くの精神科疾患はその原因が明らかになっておらず、臨床症状に基づいて症候群として規定されている。症候群としての疾患の概念を明らかにしてその疾患概念の妥当性を確立するためにはその長期予後を明らかにすることが必要である。BPDの長期予後についての研究は、西欧諸国では散見されるが^{5) 6) 7)}、守屋ら⁸⁾の研究を除くと本邦ではほとんど存在しないようである。

BPDの疾患概念はいくつかの変遷を経てきたために複雑で分かりにくいところがある¹⁰⁾。その上、BPDの疾患概念が精神科医に共有されることとなったのは、1980年にDSM-III⁹⁾が出版されてからであり、比較的最近のことである。川谷ら¹¹⁾が我々と同施設で行なった研究において臨床診断を再度診断したところ、今日、BPDと診断される患者のなかに過去において(特に1980年以前)他の診断名がつけられていたことが報告されている。また、診療録にある臨床診断だけに依拠してBPDの長期の後ろ向き研究を行なった場合にはその対象に偏りが生じる可能性が高く、より重症で典型的な群に偏ることが指摘されている¹²⁾。そこで追試可能なBPDの長期の後ろ向き研究を行なうためには、信頼性のある後ろ向きの診断方法が確立されなければならない。

本研究は、BPDの長期予後研究を可能とするための後ろ向き診断の方法を確立するための研究である。本研究では実際の面接で評価した境界パーソナリティ診断面接質問紙 (Diagnostic Interview for Borderlines: 以下、DIBと略)¹³⁾による診断と、入院診療録の情報のみから評価したDIBによる診断の相違を検討することによって後ろ向きの診断の妥当性を検討する。また、BPDの後ろ向き長期予後調査において膨大な量の診療録から多

くの患者を後ろ向きに診断するのは一人の調査者では困難であるため、複数の調査者が必要となる。その場合複数の調査者が同じ方法で同様に診断できるかどうかという疑問が生じる。その疑問に答えるためには、独立した評価者が別々に同じ入院診療録に記載されている精神病理所見をDIBで評価し、その相違を検討することによって評価者間の信頼性を検討することが必要となる。これまでのところ、欧米においてDIBを用いた後ろ向き診断に関する研究が行われてきた^{14) 15)}。そこでは十分に訓練を積んだ評価者がDIBを用いた際、診療録からBPDを診断することについての信頼性と妥当性は実際の面接と同様に高いことが示されている。

そこで、今回、日本版DIBでも原版のDIBと同じく診療録に対して用いてもその診断の妥当性と評価者間信頼性が十分に示されるかどうかを検討する。このように後ろ向きの診断の信頼性と妥当性が確立されることによって、長期後ろ向き研究での結果を解釈する際に曖昧さを減少させ、正確さの上限を提示することができると思われる。

対象と方法

〈対象〉

平成12年4月から平成15年6月までに福岡大学病院精神神経科に8週間以上入院しており、年齢が18歳～35歳の患者を対象とした。入院台帳よりこれらの患者を選択し、器質性精神病、明らかな精神病、双極性障害の症状を持つ患者は除外した。この結果47名が選出された。

入院診療録によるDIB評価の妥当性の検討ではこれらの条件に合致する患者のうち退院後2年を経過していない者を対象とした。これらの患者に電話または、現在の当科通院中または入院中の患者には主治医から連絡して研究参加の意思について確認した。入院中の患者は退院時に面接を行なった。連絡がついた患者は33名でそのうちの21名が研究に参加する意志を表明した。

入院診療録によるDIBの評価者間信頼性の検討では47名の患者のうち妥当性の検討に導入されなかった26名を対象とした。

〈方法〉

1) 境界パーソナリティ障害質問紙 (Diagnostic Interview for Borderlines : DIB)¹³⁾ について

DIB⁵⁾ は Gunderson らが 1981 年に発表した境界性人格障害診断のための半構造化面接方法である。半構造化という意味は、もともとの決められた質問に対する答えや判断を明確にするために、質問表にはないが、しかしそれと関連して必要と思われる質問を追加してもよく、場合によってはそれまでの病歴記録も参考にしてよいとされていることである。

面接による評価は社会適応性、衝動-行為のパターン、感情、精神病、対人関係の5つのセクションにわたる。採点方法について表1に概略を示したが、DIBには166の評価項目があり、112項目は患者に質問してその答えを面接者が評価する。残りの54項目は評価者の観察による評価項目である。それらの評価から5つのセクションにわたる合計29個のステートメントとよばれる項目を面接者が採点する。各ステートメントスコアは、3段階(0, 1, 2)で点数化される(一部、[-2, -1, 0]で採点するステートメントもある)。ステートメントスコアをセクションごとに合計してそのセクションの合計セクションスコアを算出する(-4~12)。その得点は各セクションで定められた方法によって3段階(0, 1, 2)のスケールドスコアに変換される。それを5つのセクションで合計し、0~10点のトータルスコアを算出する。

Gunderson ら¹³⁾ はトータルスコア7点以上を境界性人格障害あり、6点以下はその他の臨床症候群とするのが妥当であるとした。原版のDIBはその後、信頼性及び妥当性の検討も重ねられてきた^{16) 17) 18)}。米国においては、改訂が重ねられDIB-Revised¹⁹⁾となっているが、今回はDIB第2版を日本語版に翻訳したものを使用した。DIB第2版日本語版はすでに三宅らによって十分検

討され、高い信頼性と妥当性が示されている²⁰⁾。日本語版でも感度、特異度検討したところ、7点以上を境界性人格障害と考えることが妥当であると示された。

DIBを用いるためには十分な訓練を受け評価者間信頼性を高める必要があり、著者と共同研究者(吉田)は三宅らの共同研究者から直接訓練を受け高い評価者間信頼性を達成することができた。三宅らは直接DIBの開発者であるGundersonと評価者間信頼性を合わせている。

元来DIBは面接の中で用いられることを目的として作成されたために、後ろ向きに用いるためにはいくつかの変更が必要であった。最も大きな変更点は、DIBの多くの項目は最近2年または3ヶ月にあったことについて質問しているため、退院時に評価したと仮定してそこからさかのぼって2年間または3ヶ月間の状態を診療録から採点するとしたことである。また、今回の対象とならなかった境界性人格障害患者の診療録を5冊、著者と共同研究者によって評価し、不一致の点については検討を重ねた。その結果、DIBの面接用のマニュアル²¹⁾をもとにして診療録から採点する場合の細かいマニュアルを作成した。さらにこのマニュアルは研究開始後も改訂を重ねた。

2) 入院診療録によるDIB評価の妥当性の検討方法

研究に参加の意志を表明した患者に著者が研究の内容について説明を行ない、書面にて同意を得た。研究に同意したすべての対象患者に対して著者が平成15年6月から同年9月の間に直接DIBの面接を行なった。すでに退院していた患者で当科通院中の患者は当科外来にて面接を行ない、他の病院に入院中の患者はその病院の面接室を利用してDIB面接を行なった。他院に通院中の患者は当科へ来院してもらい面接した。実際の面接には約60分から90分の時間を要した。次に対象となった患者について全く知らず、面接によるDIBの結果につい

表1 DIBの採点方法

セクション	ステートメント(数)	セクションスコア	スケールドスコア
社会適応	S 1 - S 4 (4)	0 ~ 8	0, 1, 2
衝動行為	S 5 - S 9 (5)	0 ~ 10	0, 1, 2
感情	S10 - S14 (5)	-2 ~ 8	0, 1, 2
精神病	S15 - S22 (8)	-4 ~ 12	0, 1, 2
対人関係	S23 - S29 (7)	-2 ~ 12	0, 1, 2

でも関知しない別の評価者（吉田）が、DIBによって直接面接を受けた患者について入院診療録からのみDIBを採点した。実際の面接によるDIB得点と入院診療録から採点したDIB得点を比較検討した。

3) 入院診療録によるDIBの評価者間信頼性の検討方法

著者と比較的経験の少ない精神科医（衛藤）が別々に独立して診療録からDIBを採点した。別の評価者（衛藤）は著者と同程度のDIB面接での評価者間信頼性を得るために著者から訓練を受けた。別の評価者（衛藤）は境界性人格障害の概念、DIBの評価方法についての詳細なレクチャーを受けた後、著者が行なったDIBの面接を録画したビデオテープ7本を評価して著者との差を検討した。さらに実際に別の評価者（衛藤）が2人の患者の面接を行ないその過程と評価を著者がスーパーバイズした。これらのことによってマニュアルで推奨される評価者間の一致²¹⁾を得ることができた。つまり、面接におけるDIBの評価者間信頼性が十分に得られたという条件のもと、2人の評価者が独立して診療録からDIBを評価した。

〈統計〉

評価者間のDIBのステートメントスコアにおける一致とBPD診断の一致といったカテゴリカル(categorical)データについては κ 統計量を用いて算出した。DIBのセクションスコアとトータルセクションスコアといった次元性(dimensional)データの一致については級内相関係数(interclass correlation coefficient: 以下ICCと略す)を用いて算出した。 κ 統計量または

ICCの値を評価するとき、Fleiss²²⁾が提示したガイドラインを適応し0.75を超えたものを「優れている」、0.40未満を「改善が必要」とした。なお、統計パッケージはSPSS Ver. 11.0を使用した。

結 果

1) 入院診療録によるDIB評価の妥当性

面接によるDIBと診療録によるDIBの一致を検討することで診療録DIBの外的妥当性を検討した。この研究の対象となった患者の特徴を表2に示した。対象は男性2名、女性19名の合計21名であった。入院時平均年齢は、 26.6 ± 4.7 歳、調査対象となった入院の平均入院期間は 5.9 ± 6.5 ヶ月であった。対象の第一臨床診断はBPD6名、うつ病6名、神経性大食症2名、神経性無食欲症2名、適応障害2名、解離性障害2名、強迫性障害1名であった。面接によるDIB評価を退院時に行なうことができた患者と退院後数ヶ月経って面接によるDIB評価をできたものが存在する。すべての患者は少なくとも退院後2年以内にDIBの面接を受けたが、退院時からDIB評価までの平均期間は 6.7 ± 9.1 ヶ月であった。退院からDIBの評価期間と、評価者間(藤内と吉田)DIB評点の差に相関がないか検討したところ有意な相関は認められなかった。これらの患者において著者が面接で評価したDIBと独立した別の評価者(吉田)が診療録からDIBを採点してその結果の一致について κ 統計量を算出した。その結果を表3に示した。DIBのトータルスコアで7点以上となったものをBPDと診断するが、診断の一致は0.71と十分なものであった。各セク

表2 対象患者の特徴

	診療録 DIB の 妥当性研究	診療録 DIB の 評価者間信頼性研究	平均
対象人数	21	26	
入院時年齢	26.6 ± 4.7	25.2 ± 4.8	25.5 ± 4.6
性別	男性 (%)	3 (14.3%)	6 (23.1%)
	女性 (%)	18 (85.7%)	20 (76.9%)
発症年齢	20.9 ± 4.4	20.4 ± 4.7	20.5 ± 4.5
先行する平均治療施設数	1.4 ± 1.4	1.6 ± 1.2	1.5 ± 1.2
治療歴 (月)	25.1 ± 25.8	18.2 ± 21.7	21.4 ± 23.7
初診年齢	22.5 ± 5.1	22.8 ± 5.3	22.7 ± 5.1
対象となった入院期間の平均 (月)	5.9 ± 6.5	3.5 ± 1.5	4.6 ± 4.6
平均入院回数	2.5 ± 2.5	2.0 ± 1.5	2.2 ± 2.0
平均合計入院期間 (月)	8.9 ± 7.9	6.0 ± 5.3	7.3 ± 6.7
退院から DIB 面接までの期間 (月)	6.7 ± 9.1		

平均値±標準偏差

表3 診療録 DIB の妥当性と診療録 DIB での評価者間信頼性

DIB ステートメント	診療録 DIB 妥当性	診療録 DIB 評価者間信頼性
I. 社会適応	.64	.52
S1. 職場または学校での不安定	.77	.66
S2. 特別な達成	— ^b	.34
S3. 社交性	.38	.47
S4. 概観の適切さ	.49	.51
II. 衝動行為のパターン	.78	.65
S5. 自傷行為	.42	.80
S6. 自殺企図による他者操作	.46	.79
S7. 薬物乱用	.60	.50
S8. 性的逸脱	.44	.46
S9. 他の衝動行為	.49	.75
III. 感情	.50	.64
S10. 抑うつ	.26	.30
S11. 怒り	.48	.41
S12. 過剰な欲求	.33	.53
S13. 慢性のディスフォリア・アンヘドニア・空虚感	.50	.41
S14. 情緒の平板または高揚	.64	.71
IV. 精神病	.49	.49
S15. 非現実感	.36	.43
S16. 離人症	.53	.47
S17. 短期の抑うつ精神病	.33	.49
S18. 短期の妄想体験・妄想様体験	.24	.39
S19. 薬物による精神病体験	— ^a	— ^a
S20. 幻覚または虚無・誇大・奇怪な妄想	.33	.90
S21. そう状態または広範・持続性の妄想・幻覚	.52	1.00
S22. 治療中の一過性精神病様体験・入院中の行動上の退行	.48	.50
V. 対人関係	.69	.60
S23. 孤独回避	.50	.43
S24. 孤独	— ^a	.55
S25. 世話をする／されるの葛藤	.40	.43
S26. 不安定な二者関係	.46	.32
S27. 親密かつ不安定な関係	.51	.42
S28. 依存・マゾヒズム	.47	.25
S29. スタッフを巻き込む／著しい逆転移誘発	.56	.45
DIB BPD 診断 (≥7)	.71	.72

値はすべて κ 係数を示している。

^a : スコアがすべて一定のために統計量が算出されなかった。

^b : 片方の評価者のスコアが一定のために統計量が算出されなかった。

ションのスケールドスコアの一致について見ると、IV 精神病セクションのスケールドスコアで 0.49 が最低で、最も高い一致を示したのはII 衝動行為のパターンで 0.78 であった。衝動行為のパターンでの評価が診療録による DIB と面接による DIB の間で最も一致しやすかったが、その理由として、衝動行為は診療録に記載する医療スタッフにとっても患者にとっても最もインパクトがある事実であり、このセクションでは評価者は患者が報告したことに対して複雑な解釈を行なう必要がないためと思われる。

ステートメントごとに見ていくと、いくつか一致が低い (κ 係数が 0.40 未満) ために改善が必要と思われたステートメントがあった。それらは S3 社交性、S10 抑うつ、S12 過剰な欲求、S15 非現実感、S17 短期の抑うつ精神病、S18 短期の妄想体験・妄想様体験、S20 幻覚または虚無・誇大・奇怪な妄想であった。これらの項目は 1 つを除いてすべてIII 感情セクションまたはIV 精神病セクションに属するステートメントであった。3 つの項目では κ 係数を算出することができなかった。S 2 特別な達成では片方の評価者がすべて 0 と評価しておりこのため κ 統計量が算出できなかった。S19 薬物による精神病体験、S25 孤独ではどちらの評価者もすべての対象に 0 をスコアしていたために κ 統計量が算出できなかった。

表 4 に診療録 DIB の妥当性を検討する際にそれぞれの次元 (dimensional) なセクションスコアの一致について検討した結果を ICC (interclass correlation coefficient: 級内相関係数) の値で示した。DIB 合計セクションスコアと各セクションスコアの ICC はすべて 0.75 を超えており、優れた一致を示したといえる。

今回の面接による DIB 評価でもともと境界性人格障害 (BPD) と診断されていた 6 名に加えて 3 名が BPD と診断された。それら 3 名の第一臨床診断は 2 名がうつ病、1 名が適応障害であった。また、診療録による DIB 診断でも、もともと診断されていた 6 名に加えて 3 名が

BPD と診断された。その中の 1 名は面接による DIB でも BPD と診断された適応障害の患者であったが、残りの 2 名は神経性大食症の患者であった。この 2 名は面接による DIB では BPD と診断されなかった。

2) 診療録による DIB 評価の評価者間信頼性の検討

診療録による DIB 評価の評価者間信頼性研究の対象となった患者の特徴を表 2 に示した。対象は、男性 6 名、女性 20 名の合計 26 名であった。入院時平均年齢は 25.2 ± 4.8 歳、調査対象となった入院の平均入院期間は 3.5 ± 1.5 ヶ月であった。第一臨床診断は、BPD 4 名、分裂病型人格障害 2 名、分裂病質人格障害 2 名、依存型人格障害 1 名、回避性人格障害 1 名、特定不能の人格障害 1 名、神経性大食症 6 名、神経性無食欲症 1 名、うつ病 3 名、PTSD (外傷後ストレス症候群) 1 名、全般性不安障害 1 名、急性ストレス障害 1 名、性同一性障害 1 名、パニック障害 1 名であった。これらの患者を著者と別の評価者 (衛藤) が評価したときの一致を表 3 に示す。DIB のトータルスコアで 7 点以上となったものを境界性人格障害と診断するが、診断の一致は κ 統計量で 0.72 と十分なものがあった。セクションごとのスケールドスコアの一致をみると最も低値で IV 精神病の 0.49 であった。スケールドスコアが最も高かったのは衝動性のパターンで 0.65 であった。

ステートメントごとに見ていくと、改善が必要な κ 係数を示したのは、S2 特別な達成、S10 抑うつ、S18 短期の妄想体験・妄想様体験、S26 不安定な二者関係、S28 依存・マゾヒズムの 5 つのステートメントであった。その他の S19 薬物による精神病体験を除く 23 のステートメントでは評価者間のスコアに十分な一致を認めた。S19 薬物による精神病体験は、違法な薬物で精神病体験を引き起こしたことがある患者が全くおらず、どちらの評価者もすべて 0 とスコアしたために κ 係数を算出することができなかった。

次元的 (Dimensional) なスコアで示されるセクシ

表 4 診療録 DIB の妥当性と診療録 DIB 評価者間信頼性における次元信賴性

DIB デイメンショナルスコア	診療録 DIB 妥当性	診療録 DIB 評価者間信頼性
社会適応セクションスコア	.82	.90
衝動行為のパターンセクションスコア	.91	.93
感情セクションスコア	.78	.82
精神病セクションスコア	.77	.78
対人関係セクションスコア	.80	.87
DIB 合計セクションスコア	.90	.96

数値はすべて ICC の値である。

ンスコアの一致について検討した結果を ICC の値で表 4 に示した。DIB 合計セクションスコアと各セクションスコアの ICC はすべて 0.75 を超えており、優れた評価者間の一致を示した。

今回の研究で著者の診療録 DIB の評価では、もともと BPD と診断されていた患者 4 名に加えて新たに 5 名が著者によって BPD と診断された。これらの患者の第一臨床診断は、神経性大食症 2 名、分裂病質人格障害 1 名、依存性人格障害 1 名、性同一性障害 1 名であった。別の評価者(衛藤)では新たに 4 名を BPD と診断したが、診療録に記載された第一臨床診断で BPD となっていたにもかかわらず、別の評価者(衛藤)の DIB で BPD とならなかったものが 1 名いた。診療録による DIB で別の評価者(衛藤)が新たに BPD と診断した患者は、著者が新たに BPD と診断した患者とすべて重なっていたが、別の評価者(衛藤)が診断せず著者が BPD と診断した患者が 1 名いた。その第一臨床診断は神経性大食症であった。

考 察

1) 境界性人格障害の診断 (BPD) について

境界性人格障害 (BPD) の概念は Gunderson²³⁾ によると大きく分けて 3 つの流れが統合されたものであるため、その疾患概念を理解することは難しい。1 つは、Hoch ら²⁴⁾ による神経症と統合失調症、あるいは躁うつ病との境界上のスペクトラムにある症例という意味での“borderline”の使われ方である。この意味で“borderline”が用いられた場合は、DSM-III でいう BPD だけではなく分裂病型人格障害も含めるために BPD の疾患概念とは大きな隔りがある。第 2 の流れは、1983 年に Stern が²⁵⁾ 報告してその後 Kernberg²⁶⁾ や Masterson²⁷⁾ に受け継がれた流れである。これは精神療法を行なっている人々の中に比較的広く普及していた精神力動および発達の理解とその治療論に基づく“borderline”の概念である。第 3 の流れは Grinker²⁸⁾、Gunderson ら²⁹⁾ のように症候群として borderline を定義するような流れで DIB によって把握される内容はこれらのものに最も近い。この第 3 の流れを中心にして、ここにあげた概念を統合するように DSM-III で BPD の診断基準が作成されていった。このようにいくつかの流れを統合してできた疾患概念のために 1980 年代でもその理解は困難であったと思われる。

DSM は操作的診断基準であり、診断基準にある臨床症状の有無によって診断がつけられる。その最も大きな意義は診断的判定の信頼性の向上であり、その分類の妥当性ではない。ここでの妥当性とは、同じ診断を下された場合、病因、治療反応性、臨床症状、長期経過などが

同様であることを指す。多くの精神科疾患はその病因については明らかになっていないため、疾患概念の妥当性を確立するためには前述した要因(長期予後など)について実証的な調査が必要となる。つまり、BPD が DSM-III に記載された時点で BPD の疾患概念の妥当性は低く、妥当性の確立は将来の課題であった。

DSM では 5 つの軸による多軸分類を採用しているが、その目的は臨床情報を組織化し伝達すること、臨床的状況の複雑さを捉えること、同じ診断を示す各個人の非均質性を記述することなどである。I 軸は人格障害と精神遅滞以外のすべての障害や疾患を記録するためのものである。操作的診断基準であるために患者が複数の I 軸疾患を持つこともたびたびありうる。次に II 軸には人格障害と精神遅滞を記録するが、これらは持続的であるためエピソード的な I 軸疾患とはその性質が異なる。しかし、同様に II 軸疾患でも重複診断は起こることは多い。DSM 診断ではその作成上、重複診断が起こりやすいのである。

次にどの診断を主診断とするかということについて問題となるが、これについて明確な基準はない。DSM-IV では主診断として標準的には受診理由または、主な治療の対象となった疾患を記録することを勧めているが、それは I 軸でも II 軸でもよいとされている。

今回の我々の調査では臨床診断と DIB 診断の不一致がいくつかの患者で認められた。これは前述した重複診断と主診断の問題に起因すると考えられる。精神科疾患では合併症 (comorbidity) が存在することが多いが、合併症を持つ患者の臨床診断は、治療の対象となる疾患名、または優勢な病状を臨床診断とすることが多いため、臨床診断と DIB による診断の不一致が生じたと考えられる。DIB は BPD であるかどうかを診断するものであり、DIB によって新たに BPD と診断されたからといって臨床診断を否定するものではない。例えば、神経性大食症については多くのところで境界性人格障害との合併が言われている^{30) 31) 32)}。また、人格障害を 2 つ以上合併することも多いと言う報告³³⁾もある。本研究でもその影響として第一臨床診断としては依存性人格障害と分裂病質人格障害の患者が DIB で改めて BPD と診断されたと考えられる。

2) 診療録による DIB 評価の妥当性

後ろ向きの診断方法の妥当性について研究した報告は著者の知見が及ぶ限りでは 2 つしか存在しない^{14) 15)}。McGlashan ら¹⁵⁾ は BPD を DSM-III⁹⁾、DIB でそれぞれ診断した後に、BPD と診断された群とその他の診断群で統計学的な特徴を比較して診断の妥当性を示した。さらに McGlashan らは彼らの施設においての“Borderline”という診療録での診断と DIB、DSM-III との一致を κ

係数で示した。DIBでは0.58、DSM-Ⅲで0.40であった。しかし、彼らの入院診療録が記載された当時はBorderlineの疾患概念自体が曖昧で、DSM-Ⅲが出版される以前であり、DIBやDSM-Ⅲにおける境界性人格障害の概念と彼らのBorderlineの概念が大きく異なっている可能性が多くあるため高い κ 係数を求めるのは困難であった。我々の研究では診療録による診断でも外的妥当性の基準とする診断としても共通にDIBにおける境界性人格障害の疾患概念を用いているために今回我々が検討した結果では0.71とMcGlashanの報告よりも高くなっていると思われる。Ameriusらの研究¹⁴⁾は、面接によるDIBと診療録によるDIBのトータルスコアの相関が高いことで妥当性を示したが、今日ではこの方法によって一致を示すことは統計学的に意味をなさないとされている³⁴⁾。しかし、彼らは臨床診断との一致を症例ごとに示していたのでそのデータから我々が計算すると、その κ 係数は我々の結果とほぼ同レベルで0.74と十分なものであった。

現在DIBは一般的に境界性人格障害診断においてはゴールドスタンダードと考えられ、多くの境界性人格障害の研究で使用されている^{35) 36) 37)}。一方、Spitzerら³⁸⁾はLEAD (Longitudinal, expert, all data diagnosis: 経験ある臨床家による縦断的で多角的な情報源からの診断)を標準的な診断として考えることを推奨しているが、そもそも人格障害の診断は、臨床家個人が標準化されていない方法で行なうとその信頼性が非常に低くなるため標準化された診断面接が作成されるようになった⁸⁾。その経緯を考えるとSpitzerの提案には疑問を持たざるをえない。また、Cornellら¹⁷⁾によると境界性人格障害の概念についてそれぞれの施設で異なることも多いとされているため、我々の研究においてはLEADを妥当性の検討における外的妥当性の基準診断とするよりは、面接によるDIB診断の結果を基準診断とした方が正確であると考えた。

我々は研究期間を通して直接面接した著者と診療録から診断した別の評価者間で1人の患者を評価するごとに評価者間で生じた差について採点を正確にするようディスカッションを重ねた。診療録による情報の解釈と評価については十分に話し合う必要があった。

我々の努力にもかかわらず、いくつかのステートメント (S3 社交性, S10 抑うつ, S12 過剰な欲求, S15 非現実感, S17 短期の抑うつ精神病, S18 短期の妄想体験・妄想様体験, S20 幻覚または虚無・誇大・奇怪な妄想・幻覚) では十分な κ 係数は認められなかった。今回我々が用いた妥当性の研究デザインにおいて入院診療録からのDIB評価と面接によるDIB評価に差が生じる原因として3つ挙げられる。

1つは異なった時期を対象として評価を行なっている

と言う事である。入院診療録から評価したDIBは退院時を基点としてそこからさかのぼって患者を評価したが、面接によるDIBは退院時に評価したものもあるが、11例は退院後数ヶ月経過してDIBによる評価を受けた。これは評価対象期間を2年間としている項目であれば大きな問題は生じないが、3ヶ月として区切った項目ではそれぞれの評価で結果が大きく食い違う原因となったかもしれない。2つの評価方法の十分な一致が認められなかったステートメントはいずれも「この3ヶ月」を調査対象としている項目が多いⅢ感情セクションとⅣ精神病セクションに属していた。

差が生じる2つ目の原因は、異なる情報源から情報を得たことがあげられる。診療録によるDIBでは主治医または看護スタッフを含むその他の医療スタッフによる記載をもとに評価するが、面接によるDIBでは、一部決められた項目においては診療録の情報も用いられるが、多くは患者本人の報告から評価する。このため評価の対象となる情報源が異なり評価に差が生じることが考えられる。低い一致を示したステートメントはすべてⅢ感情セクションまたはⅣ精神病セクションに属していたが、特にⅢ感情セクションでは面接中に実際に観察される患者の感情も評価の対象となるため診療録からの評価は困難であったと思われる。直接の面接によるDIB評価は評価者が経験と訓練によって得られた技術を駆使して診断するため、直接面接しなければ得られない感情や細かな表情などの情報を多く含んでいる。面接で得られた情報でもそれが診療録になければ2つの評価方法には差が生じることになる。診断にとって重要な情報が入院診療録に記載されていない可能性は否定できない。さらに入院診療録は主に主治医が記載する部分が多く、主治医の評価が偏っていた場合には、我々は正確な評価が困難である。しかし、そのようなことが疑われる診療録では他の看護スタッフとの評価の一致を考えて評価するようにした。

評価方法間で差が生じる3つ目の原因としては、別々の評価者が得られた事実をどのように解釈するのかと言う点でも差が生じることがあげられる。この原因には面接者のパーソナリティや経験の差が含まれている。この問題の技術的な部分については、事前に避けることができるようディスカッションを重ね最大限努力した。しかし、臨床的な事実の判断はケースバイケースとなり評価者間での判断のズレは避けがたいものであった。三宅ら²⁰⁾はDIB日本語版の妥当性と評価者間信頼性を検討しているが、我々の結果と同様にS12 過剰な欲求 ($\kappa = 0.33$)、S18 短期の妄想体験・妄想様体験 ($\kappa = 0.315$) では低い一致を認めている。また、Gundersonらの研究¹³⁾でもS18 短期の妄想体験・妄想様体験では他のステートメントに比較して低い一致を認めている。

さて、DIBの合計セクションスコアと各セクションスコアといった次元的 (dimensional) なスコアについてICCを算出した。すべてのセクションと合計セクションスコアで優れたレベルのICCの値を示した。この値は各セクションのスケールドスコアや各ステートメントのスコアなどのカテゴリカル (categorical) な一致の値を上回っていた。これはGundersonら¹³⁾やZanariniら¹⁹⁾がDIBの信頼性について検討した結果と同様であった。

3) 診療録によるDIB評価の評価者間信頼性について

臨床経験があり十分な訓練を受けた評価者 (藤内) と、臨床経験が不十分であるがDIBの訓練を十分に積んだ評価者 (衛藤) が別々に診療録からDIBを採点した。皆川ら²¹⁾は、DIB面接を録画した少なくとも5本のテープで評価者間の一致について検討し、0~10点のトータルスコアの80%以上が1点差以内の誤差におさまり、BPD診断も80%以上一致するところまで続けることを勧めている。著者とは別の評価者 (衛藤) は7本のDIBを録画したビデオテープと2つのDIB面接を行なうことで皆川らが勧める一致を超えることができた。

本研究でのBPD診断の評価者間の一致は0.72と十分なものであった。McGlashan¹⁵⁾の研究では評価者間の信頼性は、診療録によるDIB診断では診療録によるDSM-III診断の0.73より高く0.75であった。本研究で得られた信頼性の値はMcGlashanらのものと同等であると考えられる。今回の結果で十分な評価者間の一致が得られたことから精神科医として経験が不十分、または境界性人格障害の治療に多く携わっていないものでもDIBを採点するための訓練を十分積めばDIBを後ろ向きに用いて十分に診断できることが示唆された。

各セクションの一致は、診療録によるDIBの妥当性の研究と同様にII衝動行為のパターンで高く、IV精神病で低かった。ステートメントごとにみえていくと改善が必要 ($\kappa < 0.40$) と思われる κ 係数を示したのは、S2特別な達成、S10抑うつ、S18短期の妄想体験・妄想様体験、S26不安定な二者関係、S28依存・マゾヒズムであった。

S26不安定な二者関係とS28依存・マゾヒズムはV対人関係のセクションのステートメントである。V対人関係のセクションは三宅ら²⁰⁾の研究でも他のセクションに比較して低い κ 係数を認めたように、このセクションは最も臨床経験が必要とされる評価の難しい項目である。評価者はいくつかの事実から患者の防衛スタイルや隠された動機について推測し評価しなければならない。このため十分にDIBの訓練を積んだとしても臨床経験が不十分な場合、V対人関係セクションのステートメントのいくつかでは十分な一致が得られないことがあると思われる。

S10抑うつとS18短期の妄想様体験のステートメント

の一致は、今回の妥当性に関する検討の結果と共通して低い。S18はGundersonらの研究¹³⁾でも他のステートメントに比較して低い一致を示していることから、精神病理学的に判断・評価の困難な項目であると思われる。一方、S10抑うつの一一致は他の研究では低く報告されていないために診療録で評価するとき特有の問題と考えられる。この問題について診療録によるBPD診断の妥当性を検討した先行研究^{14) 15)}のなかで各ステートメントの妥当性についてふれられていないので参考となる報告はない。しかし、その原因として前述したが、診療録では細かな表情や感情についての情報が得られにくいことがあげられる。加えてS10抑うつを評価するときその有無ではなく連続して抑うつが認められるかが問題となる。診療録で抑うつ感情の有無は容易に判断できるが、その持続性と重症度に関して判断する記載についてはDIB評価に十分なほどはなかったように思われる。

S2特別な達成は、当てはまる患者が少ないことが影響していると考えられる。一人の評価者 (藤内) は3人に1をスコアして別の評価者 (衛藤) は2人に1をスコアしたが、1人の患者についてしか一致していなかった。

DIBの合計セクションスコアと各セクションスコアといった次元的 (dimensional) なスコアの評価者間の一致についてICCを用いて算出した。信頼性の研究でも妥当性の研究と同様にすべてのセクションと合計セクションスコアで優れたレベルのICCの値を示した。この値は各セクションのスケールドスコアや各ステートメントのスコアなどのカテゴリカル (categorical) なデータの一致の値を上回っていた。

最後に本研究の結果を一般化するためにはいくつかの問題点があげられる。当科では、入院時に人格障害が疑われた場合、主治医による定期的な面接が行なわれ、その多くは週2回30分以上である。このため入院診療録のなかに医師による詳細な情報が含まれていると考えられる。また、入院診療録の冒頭には詳細な病歴ファイルが添え付けられており本研究で入院診療録からDIBを採点する際、このファイルから多くの情報を得ることができた。当科における入院治療、または入院診療録の特殊性のために十分な妥当性、及び評価者間信頼性が得られた可能性は否定できない。さらに著者を含めて今回の研究に参加した評価者は、DIBの評価について十分な訓練を受け、評価に不一致が生じた場合にはディスカッションを重ねるよう努めた。以上のことより、十分な記載がなされた診療録を使用した我々の今回の結果を一般化することには限界があると思われる。また、対象患者が少ないために κ 統計量が計算できなかったり、低くでてしまった可能性がある。本研究にさらに厳密さを求めるならば、診療録によるDIBの妥当性を検討する際、

すべての患者を退院時に面接することができればより正確な結果が得られたかも知れない。

ま と め

入院患者を対象に診療録から評価した DIB と実際の面接による DIB の結果を比較した。その結果、入院診療録から採点した DIB は面接による DIB の結果と十分な一致 ($\kappa = .71$) を示した。このことより診療録を用いて DIB によって境界性人格障害 (BPD) を診断することの妥当性が示唆された。また、別々の評価者が独立に入院診療録から DIB を評価した場合にも十分な評価者間信頼性 ($\kappa = .72$) を示した。以上のことから複数の調査者が同様に診療録から DIB を評価できることが示唆された。境界性人格障害 (BPD) の後ろ向き追跡研究の診断方法として診療録による DIB 診断を用いることが可能であることが示された。

謝 辞

本研究を行なうにあたり、DIB の評価者間信頼性を合わせるために境界性人格障害および DIB について丁寧なご指導を頂いた皆川邦直教授 (法政大学現代福祉学部)、統計に関してご助言頂いた吉永一彦先生 (福岡大学医学部社会医学総合研究室)、福岡大学精神医学教室の皆様は心より感謝する。

参考文献

- 1) 西園昌久：人格障害の入院治療。わが国における人格障害の治療に関する実態調査と入院治療：人格障害の治療体系構築のために。精神分析研究。41：404-419, 1997.
- 2) Sanislow, C.A. and T.H. McGlashan：Treatment outcome of personality disorders. Can J Psychiatry, 43: 237-250, 1998.
- 3) The American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders., IV. Washington, D.C. (高橋三郎訳：DSM- IV 精神障害の診断と統計マニュアル。医学書院：東京), 1994.
- 4) World Health Organization：The ICD-10 Classification of Mental and Behavioral Disorders. (ICD-10 精神および行動の障害。医学書院：東京), 1992.
- 5) McGlashan, T.H.：The Chestnut Lodge follow-up study. III. Long-term outcome of borderline personalities. Arch Gen Psychiatry, 43：20-30, 1986.
- 6) Paris, J., Brown, R., Nowlis, D.：Long-term follow-up of borderline patients in a general hospital. Compr Psychiatry, 28：530-535, 1987.
- 7) Stone, M.H., Stone, D.K., Hurt, S.W.：Natural history of borderline patients treated by intensive hospitalization. Psychiatr Clin North Am, 10: 185-206.
- 8) 守屋直樹：青年期境界例の長期経過。思春期青年期精神医学, 7：155-16, 1997.
- 9) The American Psychiatric Association. Diagnostic and

- Statistical Manual of Mental Disorders III. The American Psychiatric Association. Washington, D.C., 1980
- 10) 小此木啓吾：境界例の概念とその変遷。精神医学., 34: 293-300, 1992.
 - 11) 川谷大治, 牛島定信, 鈴木智美, 田上伸興, 逸見嘉之介, 古賀康彦, 藤川英昭, 兵藤和郎, 門田一法, 福井敏, 他：福岡大学病院における境界例診断の変遷と治療について。精神経誌., 92: 830-837, 1990.
 - 12) Zimmerman, M.：Diagnosing Personality Disorders: A review of issues and Research methods. Arch Gen Psychiatry., 51：225-245, 1994.
 - 13) Gunderson, J.G., Kolb, J.E., Austin, V.：The diagnostic interview for borderline patients. Am J. Psychiatry., 138：896-903, 1981.
 - 14) Amerius, B.A., Kullgren, G., Renberg, E.：Borderline diagnosis from hospital records: Reliability and validity of Gunderson's diagnostic interview for borderlines (DIB). J Nerv Ment Dis., 173：32-34, 1985.
 - 15) McGlashan, T.H.：The borderline syndrome. I. Testing three diagnostic systems. Arch Gen Psychiatry, 40：1311-1318, 1983.
 - 16) Kroll, J., Pyle, R., Zander, J., Martin, K., Lari, S., Sines, L.：Borderline personality disorder: Interrater reliability of the diagnostic interview for Borderlines. Schizophr Bull., 7：269-272, 1981.
 - 17) Cornell, D.G., Silk, K.R., Lundolph, P.S. Lohr, N.E.：Test-retest reliability of the diagnostic interview for borderlines. Arch Gen Psychiatry., 40：1307, 1983.
 - 18) Frances, A., Clarkin, J.F., Gilmore, M., Hurt, S.W., Brown, R.：Reliability of criteria for borderline personality disorder: A comparison of DSM- III and the Diagnostic Interview for borderline patients. Am J Psychiatry., 141: 1080-1084, 1984.
 - 19) Zanarini, M.C., Gunderson, J.C., Frankenburg, F.R., Chaunecy, D.L. The revised diagnostic interview for borderlines: Discriminating BPD from other axis II disorders. J Personal Disord., 3: 10-18, 1989.
 - 20) 三宅由子, 皆川邦直, 守屋直樹, 生田憲正, 宮本政於, 小此木啓吾：DIB (Diagnostic interview for borderlines) による境界例診断の試み。精神科治療学., 2：401-409, 1987.
 - 21) 皆川邦直, 三宅由子, 守屋直樹, 生田憲正, 西園マールハ文：DIB ガイドライン (マニュアルと解説)。皆川邦直, 三宅由子 (編)：境界例, pp. 46-69, 医学書院 (東京), 1993.
 - 22) Fleiss, J.L.：Statistical methods for rates and proportions (2nd ed.). New York Wiley, 218, 1981.
 - 23) Gunderson, J.G.：The borderline diagnosis, in Borderline personality disorder: clinical guide, J.G. Gunderson, Ed. American psychiatric publishing, inc: Washington, DC. p. 1-14, 2001.
 - 24) Hoch, P.H., Cattel, J.P., Strahl, M.O., Pennes, H.H.：The course and outcome of pseudoneurotic schizophrenia. Am J Psychiatry., 119: 106-115, 1962.
 - 25) Stern, A.：Psychoanalytic therapy in the borderline group of neurosis. Psychoanal Q, 17: 527-528, 1938.
 - 26) Kernberg, O.：Borderline personality organization. J Am Psychoanal Assoc, 15: 641-685, 1967.
 - 27) Masterson, J.F.：The Narcissistic and borderline disorders: An integrated developmental approach. Burner/ Mazel, inc., 1981.
 - 28) Grinker, R.R., Werble, B., Drrye, R.C.：Borderline

- syndrome. New York, Basic books, 1968.
- 29) Gunderson, J.G. and J. Kolb : Discriminating features of borderline patients. *Am J Psychiatry*, 135: 792-796, 1978.
 - 30) Grilo, C.M., Sanislow, C.A., Skodol, A.E., Gunderson, J.G., Stout, R.L., Shea, M.T., Zanarini, M.C., Bender, D.S., Morey, L.C., Dyck, I.R., McGlashan, T.H. : Do eating disorders co-occur with personality disorders?; Comparison groups matter. *Int J Eat Disord*. 33: 155-164, 2003.
 - 31) Murakami, K., Tachi, T., Washizuka, T., Ikuta, N., Miyake, Y. : A comparison of purging and non-purging eating disorder patients in comorbid personality disorders and psychopathology. *Tokai J Exp Clin Med*. 27: 9-19, 2002.
 - 32) Matsunaga H, Kiriike N, Nagata T, Yamagami S. : Personality disorders in patients with eating disorders in Japan. *Int J Eat Disord*. 23: 399-408, 1998.
 - 33) Grilo CM, Sanislow CA, McGlashan TH. : Co-occurrence of DSM-IV personality disorders with borderline personality disorder. *J Nerv Ment Dis*. 190: 552-554, 2002.
 - 34) 古川 壽亮 : エビデンス精神医療 (EBP の基礎から臨床まで). 医学書院 (東京), 2000.
 - 35) Zanarini, M.C., Frankenburg, F.R., Hennen, J., Silk, K.R. : The longitudinal course of borderline psychopathology: 6-year prospective follow-up of the phenomenology of borderline personality disorder. *Am J Psychiatry*, 160 : 274-283, 2003.
 - 36) Bateman, A. and Fonagy, P. : Effectiveness of partial hospitalization in the treatment of borderline personality disorder: a randomized controlled trial. *Am J Psychiatry*., 156: 1563-1569, 1999.
 - 37) Sanislow, C.A., Grilo, C.M., Morey, L.C., Bender, D.S., Skodol, A.E., Gunderson, J.G., Shea, M.T., Stout, R.L., Zanarini, M.C., McGlashan, T.H. : Confirmatory factor analysis of DSM-4 criteria for borderline personality disorder: Findings from the Collaborative longitudinal personality disorders study. *Am J Psychiatry*., 159: 284-290, 2002.
 - 38) Spitzer R.L. : Psychiatric diagnosis: are clinicians still necessary?. *Compr Psychiatry*. 24: 399-411, 1983.
(平成 24. 1. 11 受付, 平成 24. 3. 8 受理)

