

【研究論文】

# 知能テストを用いた障害診断の精緻化

——ベイズの定理による誤診率の低減——

緒 方 康 介\*

キーワード：知能テスト ベイズ診断 知的障害 シミュレーション

要約：知的障害の診断における知能テストの利活用は極めて重要である。特に、測定誤差をどのように考慮するのかが誤診率との関係からも整理の必要な課題である。確率論的な診断を下す上では、結合尤度比により事後確率を更新できるベイズの定理が有用であり、複数の知能テスト（WISC-IV、WISC-III、K-ABC）を用いた際に誤診率が低減するのかをシミュレーションにより分析した。複数の知能テストが同値となる仮定を置くと、単独実施の場合に比較して、複数実施は第二種の誤りに対する改善を示した。標準化調査に基づく相関をシミュレーションに組み込むと、相関の低い K-ABC を含めた際には、むしろ複数実施により誤診率が増加した。相関係数の違いによる誤診率への影響を確認したところ、2つの知能テストの相関係数が0.9以上の場合には改善が認められ、0.7以下の場合には改善はなく、0.8程度であれば初回の知能テスト結果に基づいて判断するのが有用との知見が得られた。

## 1. 緒言

### 知的障害の疫学

文部科学省（2018）によると、特別支援教育を必要とする児童生徒数は近年増加している（Figure 1）。特別支援学校の在学者に限定すると、障害種別は1974年（28%）に肢体不自由を上回って以降、知的障害が常に最大比率である（2017年：91%）。厚生労働省（2019）が実施する「生活のしづらさなどに関する調査（旧知的障害児（者）基礎調査）」によると、全国の知的障害者数は、385,100（1990年）、413,000（1995年）、459,100（2000年）、547,000（2005年）、622,000（2011年）、962,000（2016年）と推計されている。総務省統計局（2017）の「国勢調査」では、2016年10月1日時点での人口は126,933,000人と算出されており、知的障害者数は概ね0.76%程度となる。ただしこの統計は療育手帳所持者の数であり、すべての知的障害者が療育手帳を取得するわけではないため、厳密な「知的障害者数」とは異なる。

---

\*大阪大谷大学人間社会学部

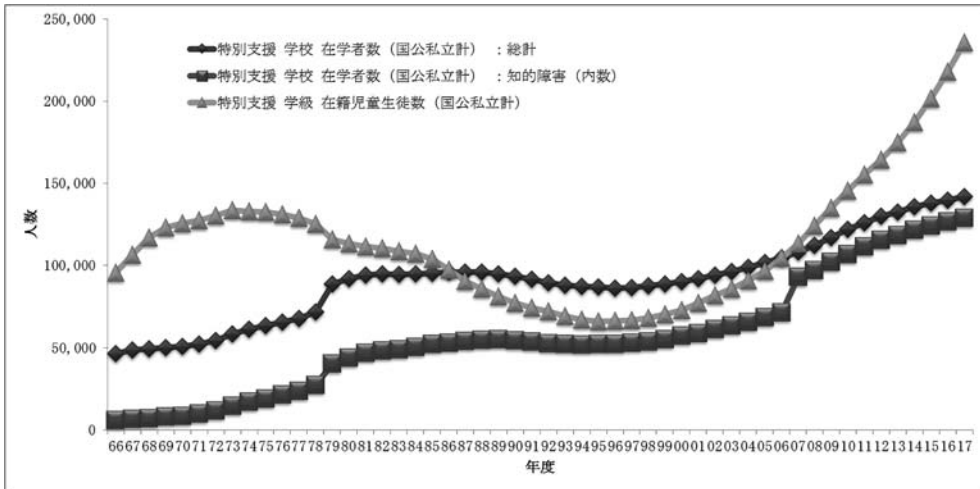


Figure 1

過去半世紀における特別支援学校（学級）在学（在籍）者数の推移：特別支援学校は幼稚部、小学部、中学部、高等部の合計、特別支援学級は小学校、中学校の合計（文部科学省初等中等教育局特別支援教育課，2018）

総人口に占める知的障害者数の疫学調査は、アイスランド（0.23%）とデンマーク（0.09%）で実施された19世紀の報告にまで遡る。ただしこれらは20世紀初頭に開発された知能テスト以前の調査であり、ときの権力者でもあった聖職者による主観的な印象報告が判断基準とされていた（Maulik & Harbour, 2010）。『国際疾病分類（International Classification of Diseases: ICD）』や『精神障害の診断と統計マニュアル（Diagnostic and Statistical Manual of Mental Disorders: DSM）』といった診断システムの確立およびBinet, A. とSimon, T. による知能テスト開発以降、客観的な知能測定が発達に伴って疫学研究も進展してきた。

これまで世界各地で行われた疫学調査は膨大な数に上るものの、1980～2009年までの知見をメタ分析したMaulik, Mascarenhas, Mathers, Dua, and Saxena (2011) による報告は頑健なエビデンスである。(1) 過去に重複した報告がない、(2) 数値での記載がある、(3) ダウン症や自閉症などの疾患に伴う知的障害ではない、(4) 特別支援学校や精神科クリニックなどの偏ったサンプリングではない、(5) 障害に係る重症度の記載がある、(6) 判定基準はICDやDSMなどの診断システムか知能テストなどの臨床アセスメントに基づいている、以上6点を除外/内包基準として5,072の文献から最終的には52の報告が選び出された。分析の結果、総人口に占める知的障害者の比率は1.037%（95%信頼区間：0.955～1.188%）と推定された。入手可能な研究知見のなかでは最大規模のメタ分析であり、本研究では1.037%を総人口に占める知的障害者比率として採用した。

## 知的障害の定義

知的障害の定義は診断システムごとに少しずつ異なっているが、本邦で入手可能な最新版の DSM-5 では以下のように記述されている。“知的能力障害（知的発達症）は、発達期に発症し、概念的、社会的、および実用的な領域における知的機能と適応機能両面の欠陥を含む障害である。以下の3つの基準を満たさなければならない。

A 臨床的評価および個別化、標準化された知能検査によって確かめられる、論理的思考、問題解決、計画、抽象的思考、判断、学校での学習、および経験からの学習など、知的機能の欠陥。B 個人の自立や社会的責任において発達のおよび社会文化的な水準を満たすことができなくなるといふ適応機能の欠陥。継続的な支援がなければ、適応上の欠陥は、家庭、学校、職場、および地域社会といった多岐にわたる環境において、コミュニケーション、社会参加、および自立した生活といった複数の日常生活活動における機能を限定する。C 知的および適応の欠陥は、発達期に間に発症する。”（American Psychiatric Association, 2013 高橋・大野監訳 2014, p.33）。ただし、DSM-5 では臨床評価に重点が置かれたため、操作的定義の色彩が薄くなっており、A、B、Cの基準が不明瞭である。旧版の DSM-IV-TR でも B 基準は操作的に定義されていなかったものの、C 基準の「発達期」は「18歳以前」、A 基準の「知的機能の欠陥」も「明らかに平均以下の知的機能」と明確であり、「明らかに平均以下」はさらに詳しく「平均より約2標準偏差低い」と操作的に定義されていた（American Psychiatric Association, 2000 高橋・大野・染矢訳 2002）。知的機能に関して、DSM-IV-TR と ICD-10 はどちらも標準化された個別式の知能テストによって Intelligence Quotient (IQ) を測定すべきとしており、知的障害の診断における IQ の重要性は言を俟たない。しかしながら、「精神遅滞」（かつての診断システムにおける表記）の定義における IQ の用いられ方、とりわけ重症度基準への適用には Table 1 に示したように両診断システム間で若干の差異がある。ここで DSM-IV-TR に IQ の幅があるのは測定誤差を考慮しているためであり、診断基準としては大雑把に「およそ5点」の測定誤差が考慮されている（American Psychiatric Association, 2000 高橋・大野・染矢訳 2002）。

Table 1 診断システムにおける「精神遅滞」の重症度基準

	DSM-IV-TR <sup>1)</sup>	ICD-10 <sup>2)</sup>
軽度	50～55 からおよそ 70	50 から 69
中度	35～40 から 50～55	35 から 49
重度	20～25 から 35～40	20 から 34
最重度	20～25 以下	20 未満

1) American Psychiatric Association (2000, 高橋・大野・染矢訳 2002)

2) World Health Organization (1992, 融・中根・小見山・岡崎・大久保監訳 2005)

## 知能テストの測定誤差

測定誤差とは、次式で計算される観測された個人の得点における誤差の大きさの推定値であり、信頼性が高ければ高いほど小さくなる（Wechsler, 2003 日本版 WISC-IV 刊行委員会 2010）。

$$SEM = SD \sqrt{1 - r_{xx}} \quad (1)$$

ただし、Standard Error of Measurement (SEM) は測定誤差、Standard Deviation (SD) は尺度の標準偏差、 $r_{xx}$  は尺度の信頼性係数である。

測定誤差が小さいことは尺度得点に基づく判断の安定性を保証するため、知的障害を診断する知能テストにおいても極めて重要な計量心理学的特性の1つである。たとえば、本邦で使用頻度の高い Wechsler Intelligence Scale for Children (WISC) の最新版 WISC-IV における IQ の測定誤差は 3.38 (Wechsler, 2003 日本版 WISC-IV 刊行委員会 2010)、旧版の WISC-III では 3.44 (Wechsler, 1991 日本版 WISC-III 刊行委員会 1998) である。Kaufman Assessment Battery for Children (K-ABC) の最新版 KABC-II における IQ 相当値である CHC 総合尺度 (認知総合尺度) の測定誤差は 2.58 (3.35) (Kaufman & Kaufman, 2004 日本版 KABC-II 制作委員会 2013)、旧版の K-ABC (認知処理過程) では 4.20 (Kaufman & Kaufman, 1983 松原・藤田・前川・石隈訳 1993) である。このように本邦で標準化された知能テストにも、測定された IQ には 3~4 点程度の誤差があり、知的障害の診断における知的機能の欠陥 (A 基準) を評価する際には十分留意しなければならない。しかしながら、知的障害の診断において測定誤差が適切に考慮されることは稀である。DSM-IV-TR での「およそ 5 点」といった評価は、障害の有無を判定する境界値 70 に対して 75 までを診断可能としているが、どのように測定誤差を適用するのかについての記載はない (American Psychiatric Association, 2000 高橋・大野・染矢訳 2002)。

## 誤診率と測定誤差

本来、臨床の障害診断は医師による判定であり、「有」か「無」の 2 択となる。ただし診断システムの基準に照らせば、後に医師の判定が誤診であると判明する可能性もゼロではない。すなわち、臨床医によって「知的障害」と判定されても、厳密に診断基準を適用すれば「知的障害」が否定されることもある。通常、誤診率は「有」と判定された患者に障害が「無」だったケース数とその逆のケース数を全体のケース数で除算して算出する個人間の統計量である。しかしながら、臨床場面では個人間の誤診率ではなく、当該個人に対する誤診率 (あるいは正

診率)に関心がある。誤診率を正しく把握しておくことは、臨床の評価だけでなく、裁判所の証拠採用に際しても極めて重要であり、科学鑑定に対する判断基準(ドーバート基準)には「どの程度の誤りが生じるかが明らかにされていること」が含まれている(勝又, 2008)。

ところが個人に下された障害診断の誤診率を求める際、先行知見である個人間の誤診率をそのまま適用することは不適切である。たとえば過去に100人が調査されており、そのうち10人に誤診が判明したとして、誤診率10%を、当該個人の受けた診断が誤っている確率と捉えようと、すべての個人に対して同じ誤診率を適用することになる。知能テストによるIQ測定を活用して知的障害を診断する場合、IQが75の個人と50の個人では誤診率が異なり、境界値付近の75よりも50の方が知的障害である確率(正診率)は高いと推定できる。測定誤差を考慮してもIQが50であれば、±5の範囲でいずれにせよ知的障害と判定できる。一方、IQが75の場合は測定誤差を考慮すると、知的障害の可能性が否定される場合もある。したがって、個人を対象とした診断における正誤の確率を計算するためには、個人間の誤診率をそのまま適用するのではなく、測定誤差を利用するなどの工夫が必要となる。

### ベイズ診断

頻度主義に立つ客観確率と異なり、1度限りの現象に対して主観確率を計算する方法にベイズ統計がある。臨床医学の診断では、従来からベイズの定理を利用して定量的に診断確率を算出してきた(佐々木, 2013)。ベイズ診断では、障害の事前確率と尤度から事後確率を求める。たとえば、知的障害のベイズ診断では、目の前にいる児童が知的障害である事前確率 $P(D)$ 、知的障害がある場合に知能テストで知的障害と判定される確率 $P(T|D)$ 、そして知能テストが知的障害と判定するすべての確率 $P(T)$ 、以上を基にして、知能テストで知的障害と判定された場合にその児童が本当に知的障害である事後確率 $P(D|T)$ が計算できる。

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} \quad (2)$$

しかしながらベイズ診断に際しては、上式の確率型よりも下式のオッズ型を利用の方が実用的である(佐々木, 2013)。

$$\frac{P(D|T)}{P(\bar{D}|T)} = \frac{P(T|D)}{P(T|\bar{D})} \times \frac{P(D)}{P(\bar{D})} \quad (3)$$

オッズ型で表現されたベイズの定理では、右辺第1因子は尤度比となる。尤度比は、知能テスト結果において、「障害がある確率」を「障害がない確率」で除算した値である。右辺第2因子は事前オッズであり、知的障害の疫学的な有病率から計算する。最終的に算出された左辺

の事後オッズを「1+オッズ」で除算して確率に再変換することにより、知能テストを実施した後、当該児童が知的障害である確率を計算する。以上のようにベイズ診断を利用すれば個人内の誤診率（正診率）を算出可能となる。

ここで、測定誤差を考慮して尤度比を計算することを考える。知能テストを実施して得られるIQ値には測定誤差があるため、IQの測定値は「真値+誤差」となる。したがって当該のIQ値が得られた場合の「知的障害である確率」および「知的障害でない確率」は分布から計算できる。たとえばWISC-IVを実施してIQが65と算出された場合、65を中心とした真値の分布を考える。正規分布が仮定できるため、WISC-IVの測定誤差を標準偏差にした $M=65$ 、 $SD=3.38$ の分布を想定する（Figure 2）。DSM-IV-TRでは、測定誤差を考慮しなければ $IQ \leq 70$ が知的障害の基準である（American Psychiatric Association, 2000 高橋・大野・染矢訳2002）。それゆえ「知的障害である確率」は、真値の分布における $IQ \leq 70$ である確率を計算し、「知的障害でない確率」はその補集合となる。Figure 2の例では「知的障害である確率」93.047%、「知的障害でない確率」6.953%となり、尤度比は13.382（ $=0.93047/0.06953$ ）と計算できる。事前オッズは、Maulik et al.（2011）のメタ分析を用いて1.037%の事前確率を変換して0.010とする。これをオッズ型ベイズの定理で積算し、事後オッズ0.140から逆変換により12.298%と事後確率を算出する。

このようにベイズ診断を用いれば個人内における「知的障害である確率」を計算できるた

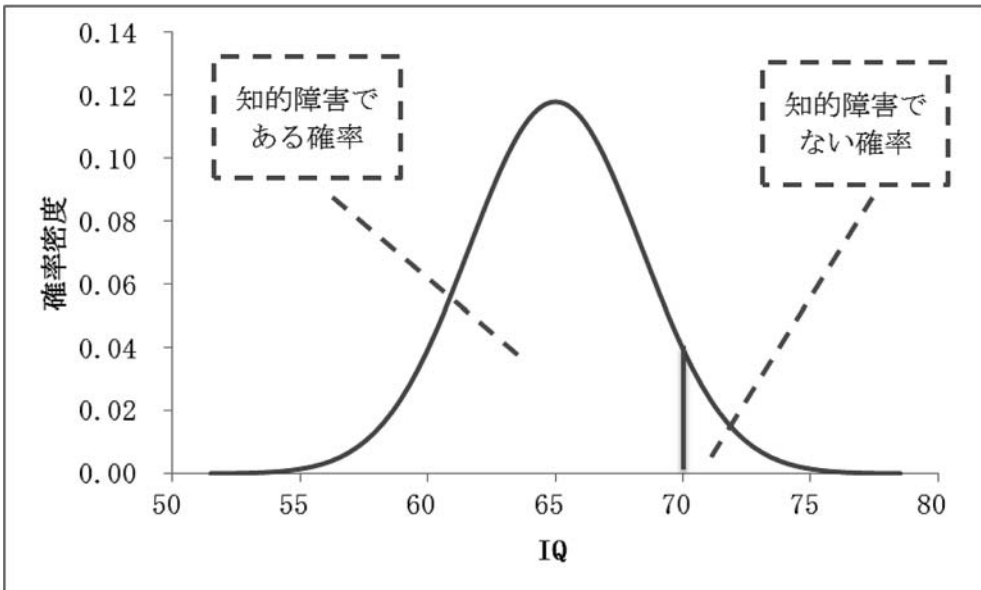


Figure 2

WISC-IVによってIQ=65と測定された場合の真値の分布と知的障害診断の尤度

め、同時に誤診率も推定可能である。先の例に沿うならば、WISC-IV による知能テストで IQ が 65 と算出されれば、DSM-IV-TR の診断基準 A について、知的機能の欠陥を認定できる。ところが B と C 基準が満たされている条件下で、この知能テストの結果だけを判断材料とすると誤診率が 87.702% もあることになる。ベイズの定理では事前確率が考慮されるため、そもそも生起頻度の稀な事象では事後確率が低く抑えられがちとなる。ただし事前確率には、専門家の判断を加味することもできるため、事前確率を最適化することでこの問題を解決することも可能である。たとえば「知的障害」の認定を受けるために知能テストが実施される場合、病院や児童相談所に訪れた子どもが「知的障害」を有している確率は一般人口中の疫学的な有病率よりも明らかに高い。そこで病院や児童相談所では、当該機関における経験的な統計値を事前確率に適用することで正診率を向上させることができる。

### 結合尤度比

事前確率自体を最適化する方法以外にも、ベイズ診断を利用して誤診率を低減させる工夫は可能である。結合尤度比を利用する方法である。結合尤度比とは、複数のテスト結果が独立に測定されている条件下で、複数のテスト結果の尤度比を積算して事後オッズを更新していく方法である（佐々木，2013）。たとえば「知的障害である確率」を求める場合、1 回目に WISC-IV を実施し、2 回目に KABC-II を実施することで、2 つの知能テストの結果を総合して結合尤度比を計算し、最終的な事後確率へと導くのである。再び先の例を用いるならば、1 回目の WISC-IV で IQ が 65 と算出され、2 回目に KABC-II を実施したところ CHC 総合尺度も 65 であったとする。この場合、WISC-IV の結果が出た時点での「知的障害である確率」は 12.298% であるが、KABC-II の結果が得られることによって 83.842% にまで上昇する。事前確率が稀な事象においても、複数のテストが同じ方向の結果を示すことで事後確率は更新されて、誤診率も改善されるのである。結合尤度比を用いる方法は、複数のテストを実施しなければならず非効率的であるが、事前確率を主観的に設定する恣意性を回避できる意味では有用である。

### 研究の目的

知能テストによる知的障害の判定における誤診率を低減する上で、複数の知能テストを用いることの効果を検証することが本研究の目的である。結合尤度比による事後確率の更新で、どの程度の誤診率低減が可能かを定量的に把握する。特に、(1) いくつの知能テストを用いるのが効率的か、(2) どの程度の相関のある知能テストを用いるのが効果的かについて、シミュレーションデータを分析するなかで明らかにする。

## 2. 方法

### シミュレーション

本研究では障害診断の精緻化を目指して複数の知能テストによる事後確率の推移を比較分析する。シミュレーションには Microsoft 社の Excel 2010 を使用して、NUMERICAL TECHNOLOGIES 社の Excel アドイン型のフリーソフト NtRand 3.3 を実行した。NtRand 3.3 は、Mersenne Twister 法による疑似乱数を生成する。汎用ソフトの Excel 上で実行可能であることに加えて、複数の知能テスト間の相関をモデル化した多変量正規分布からの乱数生成が可能であるため、本研究では NtRand 3.3 を使用することにした。

NtRand 3.3 で多変量正規分布からの乱数を生成するためには、各変数の平均および変数間の共分散を指定する必要がある。ほとんどの場合、現代の知能テストは、平均 100、標準偏差 15 となるように作成されているため、NtRand 3.3 への指定入力に際しても、平均は 100、標準偏差は 15 で固定した。相関係数の算出公式では、1つ目と2つ目の変数の標準偏差を掛け合わせた値で共分散を除算して求める。そのため、分析の目的に応じて異なるものの、指定入力が必要な共分散は、既知の相関係数に  $15 \times 15$  を乗じて算出した。

シミュレーションでは、100,000 ケースの乱数を生成し、指定入力した平均値と相関係数(共分散)が再現されているのかを確認してから分析を進めた。知能テストにおける IQ は、多くの場合、整数で 40~160 の範囲となるが、乱数生成では小数点以下の値が生じる上に、100,000 ケースものデータを生成すると、40 未満や 161 以上の数値も出現する。そのため、小数点第一位で四捨五入し、整数化した後、40 未満の場合は 40、161 以上の値に対しては 160 に固定して調整した。

乱数は測定値のシミュレーションであるため、各乱数を Table 2 の尤度比に換算し、事前確率 1.037% のオッズを掛け合わせるにより事後確率のオッズを算出後、分析ごとに 100,000 ケース個々の事後確率に再換算して本研究の分析指標とした。

## 3. 結果

### 尤度比の算出

まず3つの知能テストの測定誤差から、(1) WISC-IV、(2) WISC-III、(3) K-ABC の尤度比を計算した (Table 2)。IQ 値 70 を基準として知的障害の有無を判別し、小数点以下は第三位までを表示した。IQ 値が概ね 85 以上となれば、いずれの知能テストでも尤度比は 0.000 未満となった。Excel の計算能力に限界があり、WISC-IV では 43、WISC-III では 42 以下の場



Table 2 知的障害基準に係る 3つの知能テストの尤度比

IQ	WISC-IV (SEM = 3.38)			WISC-III (SEM = 3.44)			K-ABC (SEM = 4.20)		
	70 以下 (%)	71 以上 (%)	尤度比	70 以下 (%)	71 以上 (%)	尤度比	70 以下 (%)	71 以上 (%)	尤度比
40	100.000	0.000	<b>138572296226783</b>	100.000	0.000	<b>474063118670577</b>	100.000	0.000	2187806474311
...	...	...	...	...	...	...	...	...	...
45	100.000	0.000	14297141674191	100.000	0.000	5458908639235.960	100.000	0.000	756803581.999
...	...	...	...	...	...	...	...	...	...
50	100.000	0.000	610596598.725	100.000	0.000	327791613.257	100.000	0.000	1042892.213
...	...	...	...	...	...	...	...	...	...
55	100.000	0.000	220133	99.999	0.001	154098.502	99.982	0.018	5632.178
...	...	...	...	...	...	...	...	...	...
60	99.845	0.155	646.123	99.818	0.182	547.035	99.137	0.863	114.822
...	...	...	...	...	...	...	...	...	...
65	93.047	6.953	13.382	92.696	7.304	12.690	88.307	11.693	7.552
...	...	...	...	...	...	...	...	...	...
70	50.000	50.000	1.000	50.000	50.000	1.000	50.000	50.000	1.000
...	...	...	...	...	...	...	...	...	...
75	6.953	93.047	0.075	7.304	92.696	0.079	11.693	88.307	0.132
...	...	...	...	...	...	...	...	...	...
80	0.155	99.845	0.002	0.182	99.818	0.002	0.863	99.137	0.009
...	...	...	...	...	...	...	...	...	...
85	0.000	100.000	0.000	0.001	99.999	0.000	0.018	99.982	0.000
...	...	...	...	...	...	...	...	...	...
90	0.000	100.000	0.000	0.000	100.000	0.000	0.000	100.000	0.000
...	...	...	...	...	...	...	...	...	...
95	0.000	100.000	0.000	0.000	100.000	0.000	0.000	100.000	0.000
...	...	...	...	...	...	...	...	...	...
100	0.000	100.000	0.000	0.000	100.000	0.000	0.000	100.000	0.000
...	...	...	...	...	...	...	...	...	...

※…太字は Excel の計算能力に限界があったため、最近似値で代替。小数点第 3 位までの表記において 100 以上はすべて同じ結果

合、尤度比が算出されなかったため、いずれも計算可能な最大値で代替した。

### 複数テストの理論的効用

複数の知能テストを実施した際、いずれの測定結果においても IQ が同値であったと仮定した場合の事後確率を Table 2 で計算した尤度比を単純結合させて算出した (Figure 3)。WISC-IV を基軸にして、(1) WISC-IV 単独、(2) WISC-IV と WISC-III、(3) WISC-IV と K-ABC、(4) WISC-IV と WISC-III と K-ABC を実施した場合の事後確率推移を比較した。

事前確率 1.037% を基準にした事後確率の立ち上がり方は (1)~(4) で異なっていた。WISC-IV 単独では IQ が 60 を下回らないと 90% を超えないが、2つの知能テストを用いると 63、3つの知能テストでは 65 辺りの IQ 値で事後確率が 90% を上回っており、急峻な立ち上がりが確認された。

誤診率は 2つの誤った診断確率の和として定義できる。1つは、知的障害が無いのに「有」

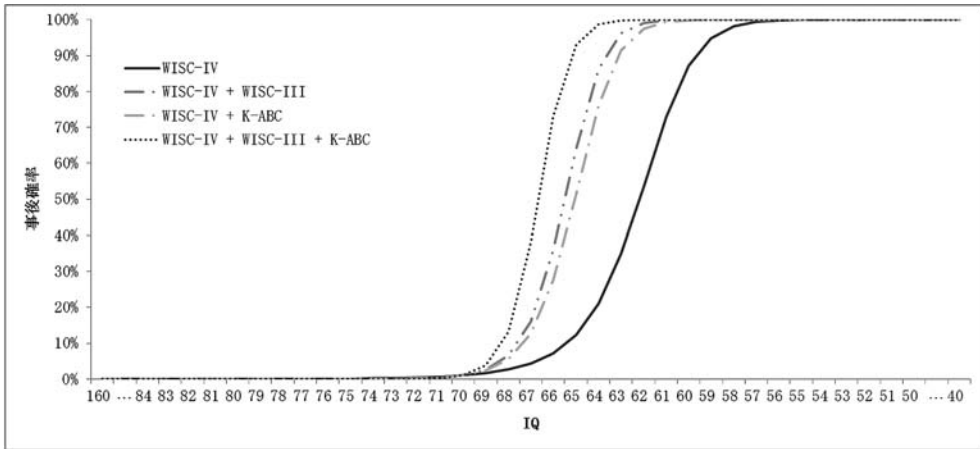


Figure 3

複数テストの結合尤度比を用いた事後確率の比較：IQ が3つの知能テストで同値であると仮定している。IQ が160～84 までと 50～40 までにグラフ上の変化はないため省略している。

Table 3 理論的シミュレーションにおける事後確率の平均値比較と誤診率の低減効果

	IQ 帯域		改善比		
	160～71	70～40	第一種の誤り	第二種の誤り	誤診率の低減度
WISC-IV	0.018%	73.906%	1.000	1.000	1.000
WISC-IV + WISC-III	0.007%	84.115%	1.000	1.138	1.069
WISC-IV + K-ABC	0.008%	82.746%	1.000	1.120	1.060
WISC-IV + WISC-III + K-ABC	0.004%	87.766%	1.000	1.188	1.094

※改善比は WISC-IV 単独実施を基準とした場合の比率であり、第一種の誤りについては、IQ 帯域 160～71 における事後確率の平均値の補集合を算出した上で計算している。誤診率の低減度は第一種と第二種の誤りにおける改善比を平均したものである。

と判断する場合であり（第一種の誤り）、もう1つは、知的障害が有るのに「無」と間違える場合である（第二種の誤り）。知能テストの単独実施と複数実施における有用性を比較検討する本研究の目的に即せば、IQ 値が70以下の帯域で事後確率が高まり、71以上の帯域で低くなれば誤診率が低減したと操作的に定義できる。この観点から、IQ が40～70と71～160の帯域における事後確率の平均値ならびに単独実施に比較した場合の改善比を Table 3 に示した。第一種の誤りに対する改善比は小さかったものの、第二種の誤りに対しては知能テストを複数実施することで1.1倍以上の低減効果があった。

### 複数テストの経験的効用

Figure 3 と Table 3 から明らかなように、同時に実施する知能テストの数が増えれば、それだけ事後確率が急上昇し、改善比は大きく、誤診率を低減できるように思われる。ただし現

実の測定場面では、必ずしも複数の知能テスト結果が同値とはならず、一定の範囲内で変動することが予測される。この状況をシミュレーションによって確認するため、3つの知能テスト間の相関係数を考慮して乱数データを100,000ケース生成した。相関係数は標準化調査の結果に基づいて指定した。5～16歳の子ども91名に対して実施されたWISC-IVの標準化調査では、WISC-IVとWISC-IIIのIQにおける相関係数は0.86であり、同じ調査で、5～12歳の子ども83名に対して実施されたWISC-IVのIQとK-ABCの認知処理過程尺度との相関係数は0.74であった（Wechsler, 2003 日本版WISC-IV刊行委員会 2010）。WISC-IIIの標準化調査によると、5～14歳の子ども28名に対して実施されたWISC-IIIのIQとK-ABCの認知処理過程尺度との相関係数は0.66であった（Wechsler, 1991 日本版WISC-III刊行委員会 1998）。

生成された乱数データから、WISC-IVを基軸にして、(1) WISC-IV単独、(2) WISC-IVとWISC-III、(3) WISC-IVとK-ABC、(4) WISC-IVとWISC-IIIとK-ABCを実施した場合の平均事後確率をFigure 4に示した。複数の知能テスト結果が全く同値になるという非現実的な強い仮定を置いた理論的シミュレーション（Figure 3）との違いは主に2つであった。1つ目は、IQ値がある帯域より低くなると、WISC-IVを単独で用いる場合よりも複数の知能テストを実施する方が、むしろ事後確率を低下させる第二種の誤りが生じてくることである。2つ目は、3つより2つの知能テストを実施した場合の方が、どのIQ帯域においても平均事後確率が高かったことである。これらの現象は複数の知能テストが完全相関していないために生じている。

Table 4に経験的シミュレーションにおける事後確率の平均値ならびに改善比を示した。わ

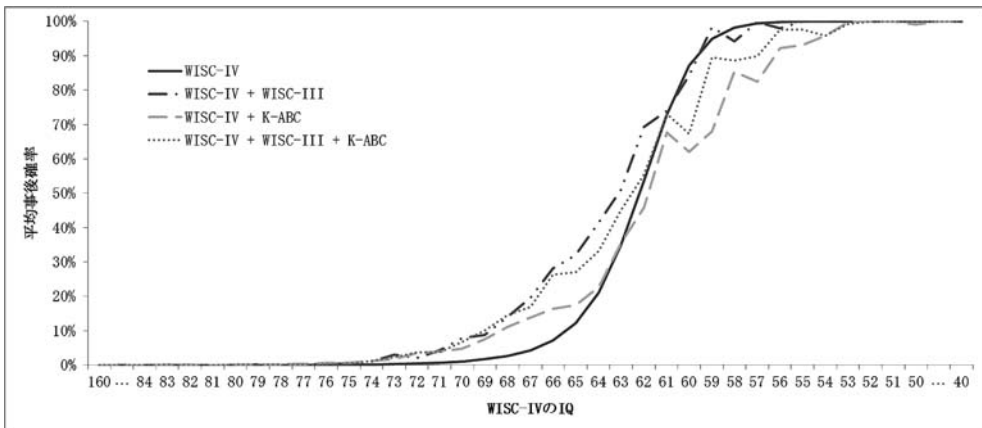


Figure 4

シミュレーションによる複数テストの結合尤度比を用いた平均事後確率の比較：IQ間の相関は標準化調査に基づいて推定している。IQが160～84までと50～40までにグラフ上の変化はないため省略している。

**Table 4** 経験的シミュレーションにおける事後確率の平均値比較と誤診率の低減効果

	IQ 帯域		改善比		
	160～71	70～40	第一種の誤り	第二種の誤り	誤診率の低減度
WISC-IV	0.018%	73.906%	1.000	1.000	1.000
WISC-IV + WISC-III	0.130%	78.106%	0.999	1.057	1.028
WISC-IV + K-ABC	0.139%	71.637%	0.999	0.969	0.984
WISC-IV + WISC-III + K-ABC	0.136%	75.277%	0.999	1.019	1.009

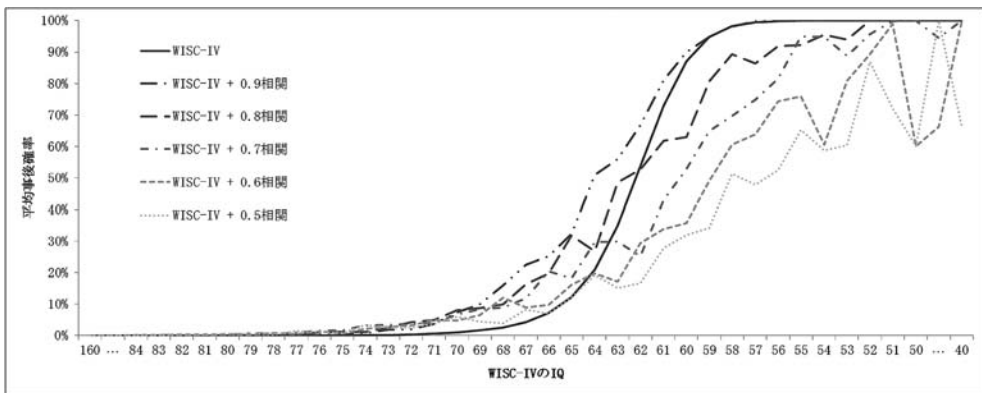
※改善比は WISC-IV 単独実施を基準とした場合の比率であり、第一種の誤りについては、IQ 帯域 160～71 における事後確率の平均値の補集合を算出した上で計算している。誤診率の低減度は第一種と第二種の誤りにおける改善比を平均したものである。

ずかではあるものの、第一種の誤りに対する誤診率は増加傾向にあったが、第二種の誤りに対しては K-ABC だけの追加実施を除いて改善があった。

#### 複数テスト間相関の影響

Figure 4 ならびに Table 4 の分析結果を受けて、複数の知能テストを実施して障害診断の事後確率を計算する上で、テスト間の相関係数が異なると、どの程度の影響が現れるのかをシミュレーションにより検証した。まず 1 つ目の知能テストとして WISC-IV を実施した後、相関の程度が異なる別の知能テストを実施したと仮定した場合に、平均事後確率がどのように推移するのかを調べた。シミュレーションでは、相関係数を 0.1 ずつ 0.5 から 0.9 まで変化させて、再度 100,000 ケースを乱数生成して平均事後確率の変遷を確認した (Figure 5)。

Table 5 に事後確率の平均値ならびに改善比を示した。第一種の誤りに対してはわずかでは



**Figure 5**

相関係数の異なる仮想テストを用いたシミュレーションにおける平均事後確率の比較：IQ が 160～84 までにグラフ上の変化はないため省略している。

**Table 5** 相関係数の異なるシミュレーションにおける事後確率の平均値比較と誤診率の低減効果

	IQ 帯域		改善比		
	160～71	70～40	第一種の誤り	第二種の誤り	誤診率の低減度
WISC-IV	0.018%	73.906%	1.000	1.000	1.000
WISC-IV + 0.9 相関	0.107%	79.055%	0.999	1.070	1.034
WISC-IV + 0.8 相関	0.176%	73.469%	0.998	0.994	0.996
WISC-IV + 0.7 相関	0.229%	68.238%	0.998	0.923	0.961
WISC-IV + 0.6 相関	0.206%	60.439%	0.998	0.818	0.908
WISC-IV + 0.5 相関	0.214%	53.111%	0.998	0.719	0.858

※改善比は WISC-IV 単独実施を基準とした場合の比率であり、第一種の誤りについては、IQ 帯域 160～71 における事後確率の平均値の補集合を算出した上で計算している。誤診率の低減度は第一種と第二種の誤りにおける改善比を平均したものである。

あるものの、相関係数の低下に伴い第二種の誤りが生じる確率は増していた。ただし相関係数が 0.9 の知能テストがある場合、総合的に誤診率は改善されていた。

#### 4. 考察

##### 複数テストによる効用

複数の知能テストが同値になるという強い仮定を置いた理論的シミュレーションの結果に鑑みれば (Figure 3 と Table 3)、知的障害の診断において 2 つ以上の知能テストを実施することに不利益はなく、誤診率低減の観点からも完全に推奨できる。特に感度が約 10% 高まっており、知的障害のある者を知的障害でないと誤診する確率を低減できていた。他方、知的障害でない者を知的障害でないと正しく診断する特異度に対しては、わずかな影響しかなく、複数の知能テストを実施しても 1% 程度しか改善しなかった。

しかしながら、現実の臨床場面は複数の知能テストの結果が常に完全に同値となるわけではない。そのため標準化調査で報告された相関係数に基づく経験的シミュレーションによって、複数の知能テストによる平均事後確率の推移を検証した。Figure 4 と Table 4 に鑑みると、3 つよりも 2 つの知能テスト、特に WISC-IV に WISC-III を加えた場合に、最も誤診率を低く抑えられていた。これは K-ABC をテストバッテリーに含むかどうかの影響と考えられる。標準化調査における WISC-IV と WISC-III の相関係数は 0.86 であったが、WISC-IV と K-ABC では 0.74、WISC-III と K-ABC では 0.66 と低かった。つまり、相関係数の低い知能テストが含まれることによって測定値の変動が大きくなり、誤診率の改善に対して寄与が小さくなったものと考えられる。逆に言えば、Table 4 に示されたように、相関係数の高い知能テストを追加実施するならば (WISC-IV と WISC-III)、相関係数の低い知能テスト (K-ABC) を含めた 3

つのテストバッテリーを使用する場合よりも誤診率を低減させられる可能性がある。したがって、複数の知能テストを用いて障害診断を行う場合、いくつの知能テストが最適かという問いはほとんど無意味であり、知能テスト間の相関の強さこそが検討すべき重要な指標といえる。

そこで Figure 5 ではシミュレーションの利点を活かして、現実には存在しない知能テストを想定し、相関係数の変化に伴う平均事後確率の推移を分析した。その結果、相関係数が 0.9 程度の知能テストがある場合、WISC-IV の単独実施に比較して、特異度にごくわずかな低下を含みつつ、感度が 5% 以上高まることから、総じて誤診率の低減効果が認められた。しかし 0.8 程度の相関では、WISC-IV の単独実施とほとんど変わらない誤診率にとどまり、2 つ目の知能テストを実施する分だけ費用対効果は下がっていた。相関係数が 0.7 以下では、この傾向が顕著となり、とりわけ感度の低下が著しく誤診率が増加してしまう危険性が懸念され、知能テストの複数実施を推奨できない結果となった。

こうした結果となったのは、現実の知能テストが完全相関していないためである。しかしながら Figure 5 と Table 5 の経験的シミュレーションは臨床的に重要な示唆を与えている。相関が 0.7 以下の場合に誤診率の低減効果は全くなかったが、たとえば相関が 0.8 の場合は次のような臨床応用が可能である。完全相関ではないため、低い IQ 帯域 (40~60) で、知能テストの単独実施に比較して複数実施による感度の高まりはなかった。つまり、初回の知能テストで IQ が 60 を下回るような低い値が出た場合、診断精度を高める (誤診率を下げる) という目的に限れば、2 つ目の知能テストを実施する意味はほとんどないといえる。しかしながら、とりわけ確定診断が難しい 60~70 の IQ 帯域では、単独実施 (平均事後確率 21.171%) に対して複数実施 (平均事後確率 28.478%) の感度が高く、0.8 程度の相関しかない知能テストであったとしても、2 つ目を実施する意味は認められる。

もう 1 つ、臨床的な応用に際して提言できることがある。0.9 という高い相関係数は、2 つの異なる知能テスト間ではなかなか得られにくい。しかしながら、知能テストの IQ 値に関して、再検査信頼性を測定すると、ほとんどの場合 0.9 以上の相関は確認される。つまり、同じ知能テストを 2 度実施することで、対象者の知的障害をより精確に診断することが理論上は可能となるのである。臨床的には、練習効果による結果の歪みが懸念されるため、間隔を空けず、同じ対象者に同じ知能テストを実施することは稀である。ただし、0.9 程度の相関がある知能テストの反復には、感度の高まりがあったため、60~70 という確定診断の難しい帯域に初回の IQ 値が落ちた場合、同じ知能テストを再度実施することで事後確率を精査するという方法論は誤診率の低減に寄与するものと考えられる。

以上の結果を総合すると、(1) 相関係数が 0.9 程度の知能テストがあるならば複数実施 (あるいは再実施) が推奨される、(2) 相関係数が 0.7 以下の知能テストしかないならば複数実施は推奨されない、(3) 相関係数が 0.8 程度の知能テストがあるならば、1 つ目の知能テスト結

果に基づいて、2つ目の実施そのものを選択・判断することが臨床的に有効と結論できる。

## 研究の限界と今後の課題

本研究はシミュレーションに根差しており、計量心理学の観点から、理論と実践の結合点として有意義な知見が得られたものと考えられる。しかしながら、方法論的にいくつかの限界もある。シミュレーションが依拠した複数の知能テスト間における相関係数は、標準化調査に基づいているものの、テスト開発時に妥当性の併存的証拠として分析される複数テストの実施において、標本サイズは通常極めて小さい。本研究で引用した WISC-IV、WISC-III、K-ABC においても、28～91名のデータに基づいて算出されている。そもそもの相関係数が不安定であれば、シミュレーションの信頼性が危うくなる。続いて、本研究知見における誤診率の低減度はあくまでも相対的な改善比に留まっている点に限界がある。知能テストの単独実施に比較した場合の複数実施による相対的効果を検証しているのであり、たとえば、精神科医による確定診断のあるケースを分析対象にして、知能テストだけによる絶対的な誤診率を計算したわけではない。とりわけ、2つの知能テストで矛盾する結果が得られた際、事後確率による診断が精神科医による確定診断と合致することを確認できれば、臨床的には極めて有用である。そのため、既に診断を受けているケースに対して、ベイズの定理を応用した障害診断の有効性を確認していく研究が今後の課題となる。

## 文献

- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders (4th ed., Text)*. Washington, DC: Author. (高橋三郎・大野裕・染矢俊幸 (訳) (2002). *DSM-IV-TR 精神疾患の診断・統計マニュアル* 医学書院)
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. Washington, DC: Author. (高橋三郎・大野裕 (監訳) (2014). *DSM-5 精神疾患の診断・統計マニュアル* 日本精神神経学会, 医学書院)
- 勝又義直 (2008). 裁判所における科学鑑定の評価について 日本法科学技術学会誌, 13, 1-6.
- Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children (K-ABC)*. Circle Pines: American Guidance Service. (松原達哉・藤田和弘・前川久男・石隈利紀 (訳) (1993). *K-ABC 心理・教育アセスメントバッテリー 解釈マニュアル* 丸善メイツ)
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children Second Edition*. U.S.A.: NCS Pearson, Inc. (日本版 KABC-II 制作委員会 (訳) (2013). *日本版 KABC-II マニュアル* 丸善出版)
- 厚生労働省 (2016). 「平成 28 年生活のしづらさなどに関する調査 (全国在宅障害児・者等実態調査)」 厚生労働省社会・援護局障害保健福祉部企画課  
〈[https://www.mhlw.go.jp/toukei/list/seikatsu\\_chousa\\_h28.html](https://www.mhlw.go.jp/toukei/list/seikatsu_chousa_h28.html)〉 (2019 年 9 月 20 日)
- Maulik, P. K., Harbour, C. K. (2010). Epidemiology of Intellectual Disability. In: J. H. Stone, M. Blouin, (eds). *International Encyclopedia of Rehabilitation*. Available online :

- <http://cirrie.buffalo.edu/encyclopedia/en/article/144/>
- Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: A meta-analysis of population-based studies. *Research in Developmental Disabilities*, 32, 419-436. doi: 10.1016/j.ridd.2010.12.018
- 文部科学省 (2018). 特別支援教育資料 (平成 29 年度) 文部科学省初等中等教育局特別支援教育課  
〈[http://www.mext.go.jp/a\\_menu/shotou/tokubetu/material/1406456.htm](http://www.mext.go.jp/a_menu/shotou/tokubetu/material/1406456.htm)〉 (2019 年 9 月 20 日)
- 佐々木春喜 (2013). 診断推論と確率－ベッドサイドでのベイズの定理 日本プライマリ・ケア連合学会誌, 36, 191-197.
- 総務省統計局 (2017). 人口推計 (平成 28 年 10 月 1 日現在) 総務省統計局  
〈<http://www.stat.go.jp/data/jinsui/2016np/index.html>〉 (2019 年 9 月 20 日)
- Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children-Third Edition*. U.S.A.: The Psychological Corporation. (日本版 WISC-III 刊行委員会 (訳) (1998). 日本版 WISC-III 知能検査法 日本文化科学社)
- Wechsler, D. (2003). *Technical and Interpretive Manual for the Wechsler Intelligence Scale for Children-Fourth Edition*. U.S.A.: NCS Pearson. (日本版 WISC-IV 刊行委員会 (訳) (2010). 日本版 WISC-IV 知能検査 理論・解釈マニュアル 日本文化科学社)
- World Health Organization. (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organization. (融道男・中根允文・小見山実・岡崎祐士・大久保善朗 (監訳) (2005). ICD-10 精神障害および行動の障害 臨床記述と診断ガイドライン 医学書院)