

判別関数による糖尿病治療群の判別

○高木 広文, 豊川 裕之, 東京大学医学部疫学教室

目的: 糖尿病は治療上特にインスリンを必須とするか否かによって, インスリン依存性糖尿病と非依存性糖尿病とに大別することが行なわれている¹⁾。本研究は糖尿病患者の初診時における諸検査項目などから, 患者のそれ以後の治療法の判別を行い, 病型による各項目の差異を明らかにすることを目的とした。

方法: 分析資料には厚生省特別研究「糖尿病治療に用いる薬剤と心血管系障害の推移に関する研究」のデータを用いた²⁾。この研究には, 朝日生命成人病研究所, 東京都済生会中央病院, 国立病院医療センター, 虎ノ門病院, 東京女子医大, および東京大学第三内科の6施設が参加し, 5年以上の治療経過の明らかな患者1,358例(男762例, 女596例)についての資料が収集された。糖尿病の病型分類は, 5年間のうち80%以上の期間にわたってインスリンが治療に用いられた場合にインスリン治療群(以下I群と称す)とし, その他の場合を非インスリン治療群(以下N群と称す)とした。この病型を外的基準とし, 表1, 2, 3に示した変数および性を説明変数として判別関数を求めた。変数選択の基準にはウィルクスの Λ 統計量³⁾の減少が最大になる変数を採用した。また選択された変数がどのような意味をもつのかを検討するために, 主成分分析を行った。カテゴリ変数の各カテゴリには表4のように数値を与えた。なお以後の分析では欠損値のあるケースは分析から除外したが, 線型判別関数の算出には, 欠損値に各群の平均値を代入して用いた。実際の計算には, 東京大学大型計算機センターのSPSS⁴⁾を用いた。

結果:

1. 各項目の2群間の比較: 表1に糖尿病の発症年齢, 罹病期間, 発症前の最大肥満度およびその時の年齢, 初診時における主要検査の2群における平均値とt検定の結果を示した。大部分の項目において有意差($p < 0.05$ とすると)が認められたが, 血清トリグリセリド, 血清クレアチニンでは有意差は認められなかった。またこれら2項目は有効な標本数が極めて少なく, 以後の分析から除外した。表2に初診時における眼底, 心電図および尿たん白所見を示した。眼底および尿たん白は, I群においてその異常者の割合が高く, 統計学的にも有意であったが, 心電図所見での異常はI群とN群の間に差はなかった。表3に家族歴, 既往歴および喫煙歴を示した。有意差の認められたものは, 高血圧の既往歴および喫煙歴のみであり, いずれもN群で高率であった。これらの項目を用いてI群かN群かの判別を行ったが, 空腹時血糖値と発症年齢は重要な役割りをもつと考えられるので⁵⁾, これら2項目に欠損のあるケースは分析から除外した。この結果, I群257例(男133例, 女124例)N群1,091例(男622例, 女469例)となったが, このうちN群から261例(男147例, 女114例)を無作為抽出し判別関数算出に用い, 残りは判別関数のチェックに用いた。

2. 線型判別関数: 表5に選択された変数とその順位, 全例での外的基準(I群に0, N群に1を与えた)との単相関係数, 重相関係数, マハラノビスの汎距離および各段階での正診率を示した。表6には判別得点算出のための判別係数(素データに対する)と標準化判別係数を示し, 図1に2群の判別得点の累積分布を示した。自由度再調整済みマハラノビスの汎

表1 各項目の2群の平均値・標準偏差およびt検定

項 目	群	N	Mean	S.D.	t*	p
発 症 年 齢 (歳)	I	257	41.6	13.6	8.82	0.000
	N	1099	49.7	11.5		
罹 病 期 間 (年)	I	257	5.3	6.6	5.28	0.000
	N	1100	3.0	4.9		
最 大 肥 満 度 (%)	I	238	22.5	20.2	3.44	0.001
	N	986	27.6	20.6		
最大体重時年齢 (歳)	I	200	33.0	12.0	6.65	0.000
	N	796	39.3	12.1		
空腹時血糖値 (mg/dl)	I	257	202.1	75.1	11.03	0.000
	N	1093	147.7	51.1		
血清コレステロール (mg/dl)	I	194	232.5	59.4	2.26	0.024
	N	928	222.1	51.9		
血清トリグリセリド (mg/dl)	I	54	110.7	69.5	0.18	0.861
	N	290	112.6	90.7		
血清尿酸値 (mg/dl)	I	115	17.6	5.1	2.73	0.007
	N	621	16.0	5.7		
血清クレアチニン (mg/dl)	I	24	1.6	3.1	0.91	0.371
	N	76	1.0	0.3		
初診時肥満度 (%)	I	243	2.9	19.0	6.37	0.000
	N	1015	11.4	17.1		
収 縮 期 血 圧 (mmHg)	I	175	128.2	26.6	4.27	0.000
	N	934	137.4	23.5		
拡張期血圧 (mmHg)	I	176	77.0	13.3	4.36	0.000
	N	935	81.7	13.0		

* 等分散性が棄却された場合にはWelchの方法によって計算した。

表2 2群における眼底、心電図および尿たん白所見

項 目	所 見	I	N	Chi-sq.	p
眼 底	No	85(57.8)	573(75.0)	18.504	0.000
	(1)*	43(29.3)	139(18.2)		
	(2)**	19(12.9)	52(6.8)		
心 電 図	Normal [#]	158(91.3)	727(91.6)	0.003	0.959
	Abnormal	15(8.7)	67(8.4)		
尿たん白	No	139(62.6)	751(77.1)	20.365	0.000
	(+)	69(31.1)	179(18.4)		
	(++) 以上	14(6.3)	44(4.5)		

* Scott I, II or Wagener I, II

** Scott III, IV, V or Wagener III, IV, V

Q, S T, T変化に異常のないものを正常, 他のもを異常とした。

表3 2群の家族歴・既往歴および喫煙歴

家族歴・既往歴・喫煙歴	I	N	Chi-Sq.	p
糖尿病：家族歴	No	161(66.0)	0.046	0.831
	Yes	83(34.0)		
虚血性心疾患：家族歴	No	206(94.5)	2.761	0.097
	Yes	12(5.5)		
脳卒中：家族歴	No	156(66.4)	1.815	0.070
	Yes	79(33.6)		
高血圧：家族歴	No	155(70.5)	0.901	0.368
	Yes	65(29.5)		
虚血性心疾患：既往歴	No	243(99.2)	2.696	0.101
	Yes	2(0.8)		
脳卒中：既往歴	No	246(98.4)	0.024	0.876
	Yes	4(1.6)		
高血圧：既往歴	No	214(86.3)	7.147	0.008
	Yes	34(13.7)		
胃炎・ネフローゼ：既往歴	No	234(92.9)	0.016	0.900
	Yes	18(7.1)		
肝炎・肝硬変：既往歴	No	234(95.9)	0.678	0.410
	Yes	10(4.1)		
胃切：既往歴	No	243(99.2)	0.013	0.911
	Yes	2(0.8)		
甲状腺疾患：既往歴	No	236(99.2)	0.001	0.976
	Yes	2(0.8)		
喫 煙 歴	No	134(73.6)	4.867	0.027
	Yes	48(26.4)		

() : %

距離は全段階で増加しており、選択された15変数は何らかの意味で判別に寄与することがわかる。変数選択の順位が必ずしも判別への寄与の順位とは一致しないので、標準化判別係数の大きさを寄与をみるものとした。最も重要な変数は空腹時血糖値であり、これのみで67.6%の正診率が得られ(表5)、測定値が高い場合にI群となる。ついで血清尿酸値、発症年齢、初診時肥満度、眼底所見、罹病期間の順であった。最大体重時年齢は外の基準との単相関は大きい、表7に示したように発症年齢との相関係数が、0.581あり、選択順位も寄与順位もやや下がった。収縮期血圧は選択順位に比べ、寄与順位は上に

なっていた。選択順位が10番目以降は全てカテゴリ変数であり、寄与も大体この順であるが、尿たん白、喫煙歴で比較的寄与が高く、血清コレステロールと尿たん白はほぼ同等の寄与であった。実際の正診率の推移をみると(表5, 図2), I群, N群ともに変数の増加とともに、正診率が単調に増加するわけではなく、I群では9変数のときに最も高く、N群では11変数のときに最も高かった。両群あわせた場合には、15変数の場合に77.8%と正診率は最も高いが、これとて9変数の場合の77.6%と0.2%の差しかなく、人数にして僅か1例の差であった。この15変数による判別得点を用いて、N群のうち判別関数の算出に用いなかった830例について判別を行ったところ(分割点は0), 81.0%の正診率が得られた。

3. 主成分分析：判別分析に用いた説明変数と治療法および眼底悪化に関する変数を1つ加え、表8に示したように28変数による主成分分析を行い、判別関数に選択された変数のもつ

表4 カテゴリ変数に与えた数値

項目	数値	0	1	2
性	Male	Female	—	—
家族歴・既往歴	No	Yes	—	—
喫煙歴	No	Yes	—	—
尿たん白	No	(+)	(++)	—
心電図	Normal	Abnormal	—	—
眼底	Normal (1)	(2)	—	—

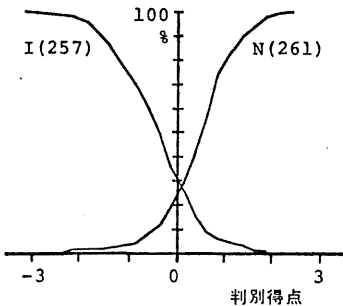


図1 2群の判別得点の累積分布

意味を検討した。5年間の治療法は、第1, 第2主成分の負荷量が大きく、第1主成分は収縮期血圧, 拡張期血圧, 発症年齢, 最大体重時年齢などの主に「年齢」に関係した変数で負荷量が大きかった。第1主成分の負荷量の大きなものからは、収縮期血圧, 初診時肥満度, 発症年齢, 最大体重時年齢の4変数が選択され、これ

らの測定値が小さい場合にI群となる。第2主成分の負荷量の大きなものは、眼底所見, 罹病期間, 尿たん白, 空腹時血糖値であり、これらは糖尿病性細小血管障害⁶⁾に関する変数であり、測定値が大きい(重度である)場合にI群となる。第3主成分は性に関する変数で、喫煙歴が選択され、第4主成分は家族歴のうち高血圧, 脳卒中に関するもので、発症年齢も負荷量が大きく、高血圧の家族歴が選択された。第5主成分は循環器系に関するもので、心電図所見, 脳卒中の既往歴が選択された。これら5主成分のうち特に高い負荷量がないもので選択された変数は、血清コレステロール, 糖尿病の家族歴, 血清尿素窒素であった。

考察: 初診時の検査, 問診などで得られたデータを用いて、インスリン治療群と非インスリン治療群の判別分析を行い、さらに主成分分析により説明変数の意味を検討した。本研究で用いた説明変数にはカテゴリ変数が含まれており、これに対して線型判別関数を用いるのは問題がないことはない。ダミー変数を用いる方法があるが⁷⁾、これは二分法的属性に対しては0, 1の数値を与えることに対応するので、本研究で用いた変数で問題となるのは尿たん白, 眼底所見の2変数のみである。実際、これらの2変数についてもダミー変数の扱いをして、分析したところ、注目すべき改善とはならず、表4に示したような扱いのままにした。これとは逆に林の数量化2類を用いる方法もあるが、これは全変数がカテゴリ変数の場合にはダミー変数を用いた判別分析に一致するが⁷⁾、本研究のように連続量を含む場合、区分の仕方に問題が

表5 治療群の判別に用いた変数 (前進選択法による)

変数	r*	R ₁ **	D _m ² †	D _m /2	正診率(%)				D _m ² ‡	D _m ² §
					I群	N群	両群	両群		
1. 空腹時血糖値	-0.354	0.394	0.731	0.427	67.6	67.7	67.6	66.5	0.722	0.713
2. 発症年齢	0.255	0.484	1.222	0.553	66.9	74.7	70.9	71.0	1.201	1.181
3. 血清尿素窒素	-0.100	0.545	1.684	0.649	68.1	77.4	72.9	74.2	1.651	1.619
4. 眼底所見	-0.138	0.580	2.018	0.710	70.4	75.1	72.8	76.1	1.972	1.926
5. 初診時肥満度	0.188	0.601	2.254	0.751	70.4	76.2	73.4	77.4	2.193	2.134
6. 罹病期間	-0.170	0.625	2.559	0.800	72.8	77.0	74.9	78.8	2.483	2.409
7. 最大体重時年齢	0.206	0.639	2.757	0.830	72.4	76.6	74.5	79.7	2.666	2.577
8. コレステロール	-0.073	0.651	2.931	0.856	72.0	78.2	75.1	80.4	2.824	2.721
9. 収縮期血圧	0.139	0.664	3.140	0.886	75.1	80.1	77.6	81.2	3.065	2.897
10. 尿たん白	-0.114	0.675	3.338	0.914	73.5	80.8	77.2	82.0	3.197	3.061
11. 喫煙歴	0.075	0.684	3.513	0.937	73.2	81.2	77.2	82.6	3.353	3.200
12. 高血圧・家族歴	0.031	0.691	3.641	0.954	73.5	80.8	77.2	83.0	3.464	3.296
13. 心電図所見	-0.003	0.695	3.725	0.965	73.9	80.5	77.2	83.3	3.531	3.348
14. 糖尿病・家族歴	0.008	0.699	3.801	0.975	74.3	80.5	77.4	83.5	3.590	3.390
15. 脳卒中・既往歴	0.004	0.701	3.856	0.982	74.7	80.8	77.8	83.7	3.628	3.414

* 全ケース(1368例)における単相関係数。

** I群257例, N群261例における重相関係数。

† マハラノビスの汎距離。

‡ 自由度調整済みマハラノビスの汎距離。

§ 自由度調整済みマハラノビスの汎距離。

表6 前進選択法による15変数の判別係数

変数	判別係数	標準化判別係数
1. 空腹時血糖値	-0.00598	-0.422
2. 発症年齢	0.01751	0.230
3. 血清尿素窒素	-0.07026	-0.248
4. 眼底所見	-0.41835	-0.217
5. 初診時肥満度	0.01204	0.225
6. 罹病期間	-0.03846	-0.217
7. 最大体重時年齢	0.01652	0.176
8. 血清コレステロール	-0.00299	-0.150
9. 収縮期血圧	0.00953	0.206
10. 尿たん白	-0.26986	-0.151
11. 喫煙歴	0.36697	0.143
12. 高血圧・家族歴	0.24617	0.106
13. 心電図所見	-0.31011	-0.074
14. 糖尿病・家族歴	0.18969	0.087
15. 脳卒中・既往歴	-0.59434	-0.073
定数項	0.38792	

表7 治療群の判別に用いた15変数の相関係数

1. 空腹時血糖値	1.000													
2. 発症年齢	-0.139	1.000												
3. 血清尿素窒素	0.049	0.038	1.000											
4. 眼底所見	-0.116	-0.033	0.096	1.000										
5. 初診時肥満度	-0.129	0.157	-0.077	0.003	1.000									
6. 喫煙歴	0.063	-0.242	0.037	0.250	0.002	1.000								
7. 最大体重時年齢	-0.127	0.581	-0.017	-0.070	0.244	-0.132	1.000							
8. コレステロール	0.115	0.125	0.050	0.000	0.140	0.015	0.056	1.000						
9. 収縮期血圧	-0.111	0.247	-0.026	0.170	0.300	0.179	0.175	0.117	1.000					
10. 尿たん白	0.164	-0.002	0.037	0.267	0.029	0.099	-0.020	0.119	0.182	1.000				
11. 喫煙歴	-0.083	0.030	0.078	-0.054	-0.105	0.024	-0.026	-0.119	-0.048	-0.031	1.000			
12. 高血圧・家族歴	-0.034	-0.108	-0.033	-0.053	0.037	0.001	-0.042	-0.002	0.027	-0.033	0.006	1.000		
13. 心電図所見	0.024	0.139	-0.002	-0.035	0.117	0.070	0.070	0.138	0.125	0.072	-0.055	0.020	1.000	
14. 糖尿病・家族歴	0.007	-0.184	0.013	-0.006	-0.018	0.031	-0.110	-0.037	-0.059	-0.002	-0.034	0.053	-0.010	1.000
15. 脳卒中・既往歴	-0.060	0.025	-0.009	0.075	0.045	0.010	-0.031	0.009	0.087	0.033	0.090	-0.004	0.170	0.002
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.

表8 28変数による主成分分析 (第5主成分まで)

変数	I	II	III	IV	V	2重和
△収縮期血圧	0.683	0.285	0.292	0.056	-0.283	0.716
△初診時肥満度	0.682	0.096	-0.048	0.020	0.000	0.536
△発症年齢	0.620	0.242	-0.231	-0.063	0.079	0.507
△収縮期血圧	0.585	0.175	0.278	0.127	-0.392	0.620
△発症年齢	0.544	-0.308	0.035	-0.464	0.150	0.630
△最大体重時年齢	0.539	-0.289	-0.060	-0.345	0.102	0.507
△高血圧・既往歴	0.472	0.020	0.178	0.295	0.032	0.343
△糖尿病・家族歴・5年	-0.092	0.652	0.165	-0.128	-0.073	0.483
△初診時眼底所見	0.039	0.604	0.161	-0.079	-0.060	0.402
△喫煙歴	-0.031	0.514	0.287	0.177	0.071	0.384
△尿たん白	0.119	0.496	0.053	-0.088	0.116	0.284
△空腹時血糖値	-0.250	0.478	-0.221	-0.094	0.163	0.375
△5年間の治療法	0.369	-0.461	0.113	-0.024	-0.164	0.389
△性 別	0.171	0.165	-0.722	0.180	0.070	0.615
△喫 煙 歴	-0.122	-0.203	0.668	-0.077	0.087	0.516
△高血圧・家族歴	0.092	-0.087	-0.033	0.686	0.030	0.489
△脳卒中・家族歴	0.128	-0.149	0.008	0.591	0.057	0.391
△虚性心疾患・既往	0.133	-0.080	0.140	0.066	0.562	0.364
△心電図所見	0.291	0.083	0.005	-0.040	0.552	0.398
△脳卒中・既往歴	0.126	0.084	0.312	0.116	0.443	0.330
△血清コレステロール	0.230	0.204	0.210	0.115	0.139	0.171
△糖尿病・家族歴	-0.132	0.042	-0.085	0.295	0.036	0.115
△血清尿素窒素	-0.108	0.182	0.170	-0.232	0.098	0.137
△虚性心疾患・家族	0.116	-0.048	0.100	0.220	-0.036	0.075
△肝臓・既往歴	-0.020	-0.018	-0.085	-0.048	0.236	0.067
△肝臓・既往歴	-0.029	-0.065	0.149	0.100	0.202	0.078
△胃 切 除	-0.051	-0.101	0.191	0.030	0.198	0.089
△腎臓・ネフローゼ	0.095	0.103	-0.008	0.101	0.121	0.045
総 変 異	2.984	2.324	1.755	1.620	1.370	10.05
解 釈 率	10.7%	8.3%	6.3%	5.8%	4.9%	35.9%

△ 判別分析において変数選択されたもの。
■ 判別分析における外的基準。

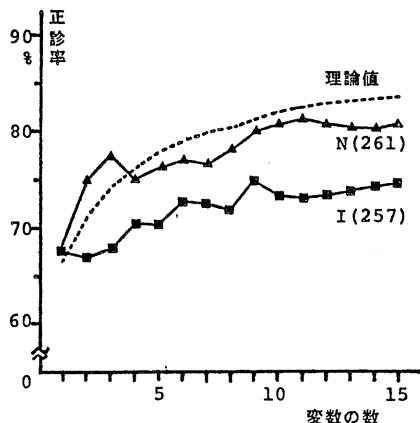


図2 変数追加による正診率の推移

あり、各種の方法で区分して実際に一通り計算したが、判別関数以上の効率をあげられなかった。多変量データには欠損値が含まれることが、実際の場面では多い。今回は各群の平均値を欠損部に代入して用いたが、重回帰式を用いて欠損値を予測するなどが考えられ、この点についてさらに検討する必要がある。

結論：糖尿病患者の初診時のデータを用いて、患者の治療にインスリンが必須か否かの判別分析を行い、また判別関数に選択された変数の意味を検討するため主成分分析を行った。選択された変数は、収縮期血圧・初診時肥満度・発症年齢・最大体重時年齢（第1主成分）、初診時眼底所見・罹病期間・尿たん白・空腹時血糖値（第2主成分）、喫煙歴（第3主成分）、高血圧の家族歴（第4主成分）、心電図所見・脳卒中の既往歴（第5主成分）、血清コレステロール、糖尿病の家族歴および血清尿素窒素の15変数であった。判別の正診率は77.8%であり、判別関数の算出に用いなかったデータに応用したところ81.0%であった。これらの結果に主に方法論的な考察を加え、また問題点を指摘した。

終りに、貴重な資料を提供していただいた東京大学医学部第三内科の小坂樹徳教授をはじめ「糖尿病治療に用いる薬剤と心血管系障害の推移に関する研究」班員および関係各位に感謝します。

参考文献：1) 小坂樹徳, 現代医療, 9 (8-9), 1015-1025, 1977 2) 小坂樹徳 他: 厚生省特別研究「糖尿病治療に用いる薬剤と心血管系障害の推移に関する研究」, 総括研究報告書, 1977 3) C.R.Rao: Linear Statistical Inference and its Applications, 2nd edit., 奥野忠一 他訳, 統計的推測とその応用, 東京図書 (東京), 1977 4) 三宅一郎 他, SPSS統計パッケージ, 11 解析編, 東洋経済 (東京), 1977 5) 王子亘由 他, 糖尿病, 16, 430, 1973 6) 小坂樹徳, 内科, 37, 364, 1976 7) 安田三郎, 社会統計学, 丸善 (東京), 1969