

丸山ワクチン 15 万症例のデータマイニング

1202720 成蹊大学 *新村秀一 SHINMURA Shuichi

日本医科大学付属病院ワクチン療法施設 飯田和美, 平井敏之, 永積惇

1 はじめに

日本医科大学付属病院ワクチン療法研究施設では、1964年からSSM(人型結核菌体抽出物質、略称丸山ワクチン)による癌治療をはじめた。そして、1979年からSSMの治験患者をコンピュータに登録している。新村・飯田・丸山(1987)では、1985年7月までに登録された118,301症例を用いた帰無仮説モデルによる成果を発表している。

今回の研究では、2000年9月までに登録された152,989症例のデータをデータマイニングの代表的な手法の決定木で分析した結果を報告する。

2 分析対象と研究方法

2・1 分析対象

表1は、区間幅5年で、治験患者の重要なイベント年を単純集計したものである。癌治療は、診断が確定されることから始まり、その病状に合わせて手術、放射線療法、化学療法などが

表1 年代別治験患者数

	診断確定年	手術年	投与開始 ¹	最終外来 ²
-1980	771	16849	16356	7714
1981-1985	6113	20307	48002	29342
1986-1990	21536	19561	38086	28091
1991-1995	18866	15152	29970	19533
1996-2000	10719	7474	20548	15756
小計	58005	79343	152962	100436
欠損値	94984	73646	27	52553
総計	152989	152989	152989	152989

行われる。癌患者や家族の中には、この他にSSMの投与を希望するものもいる。表中の「投与開始」は、SSMの投与開始年であり、「最終外来年」は家族が最後にSSMをとりに来た年である。表に見る通り、残念ながら欠損値が多い。

2・2 分析データと目的変数の設定

種々の分析を試行した後データを、手術日から外来最終日までのある47,439症例をA群とし、この症例を省いた投与日から外来最終日までのある52,275症例をB群とし、それ以外をC群とした。A群では手術日から外来最終日までの生存日数を、B群では投与日から外来最終日までの生存日数を目的変数とし、30項目以上の癌患者の各種属性(性、年齢、診断結果、治療法など)を説明変数として用いた。

3 分析方法と分析結果

分析方法としては、一般的な解析手法による分析の後、決定木分析を行った。分析ソフトとして、StatisticaとSPSSのAnswerTreeを用いた。

Exhaustive CHAIDを用いたA群の分析で、表2のように4レベルで18個のノードに

分割された。各ノードにおける生存日数の平均値の大小順に並べ替えてある。一番成績の良かったノード15は、第1層と2層でリンパ節転移がなく、第3層で女性、第4層で手術方法の種類で特定された4,964例である。平均生存日数が1,782日で、95%点が5,035日である。

表2 18ノードの生存日数

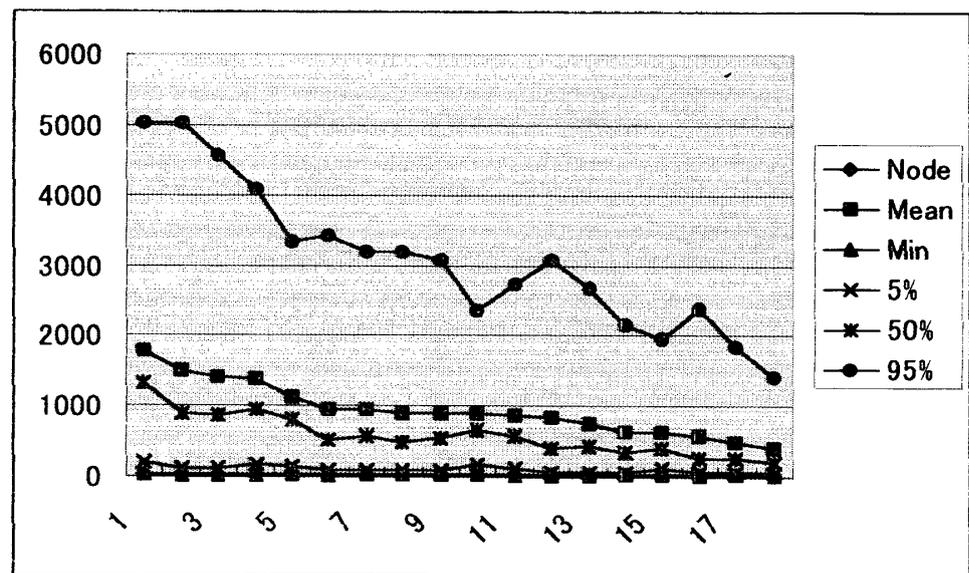
Rank	Node	Case	Mean	Min	5%	50%	95%
1	15	4964	1782	15	208	1316	5035
2	24	1595	1481	4	115	879	5021
3	20	5469	1404	1	105	870	4579
4	18	2534	1381	12	167	952	4074
5	23	6097	1129	21	157	791	3336
6	21	3137	958	6	75	520	3423
7	16	1049	939	28	96	565	3180
8	17	1084	904	34	88	485	3192
9	31	1120	889	32	98	538	3071
10	27	516	887	31	162	673	2350
11	28	2963	862	8	110	563	2715
12	25	6804	822	2	66	413	3065
13	19	1967	754	7	68	425	2659
14	22	659	632	29	69	351	2164
15	29	1562	624	21	106	410	1949
16	30	1476	561	3	50	251	2393
17	26	1759	486	15	64	261	1849
18	32	2684	389	9	51	213	1398

図1は、これを折れ線で表した。横軸は18個のノードをランク順に並べてあり、縦軸は生存日数、6個の系列は値の小さなものから最小値、5%点、中央値、平均値、95%点に対応している。最小値と5%点の生存日数はほとんど変わらない。*で表された中央値は、平均値に平行している。一方、95%点は大きな差異が認められた。

4 まとめ

95%点にみるように、決定木での分類で、長期生存例を良く分類できることが分かった。例えば、ノード15で生存日数は5,035日(13.8年)、ノード18で1398日(3.8年)であった。しかし、5%点は全てのノードが208日以下であり、癌という

図1 18ノードの各生存日数



業病の難しさが改めて明らかにされた。

文献) 新村秀一, 飯田和美, 丸山千里(1987). SSMの癌治療における帰無仮説モデルによる評価, 医療情報学 7(3), 263-275.