

Propensity Score Matching 방법을
이용한 간경변증 위험 인자의 재평가

연세대학교대학원

의학전산통계학협동과정

의학통계학전공

노 성 유

Propensity Score Matching 방법을
이용한 간경변증 위험 인자의 재평가

지도 김도영 교수

이 논문을 석사 학위논문으로 제출함

2008년 6월 일

연세대학교 대학원

의학전산통계학협동과정

의학통계학전공

노성유

노성유의 석사 학위논문을 인준함

심사위원 _____ 인

심사위원 _____ 인

심사위원 _____ 인

연세대학교 대학원

2008년 6월 일

감사의 글

저에게 있어 대학원에서 보낸 2년이란 시간은 어디서도 얻지 못할 귀한 배움의 기회였고, 제 자신이 좀 더 깊이 있는 사람으로 거듭나기 위한 값진 훈련의 과정이었습니다. 쉽지만은 않았던 선택으로 시작하여 2년간의 결실을 맺기까지 그 안에는 말로는 다할 수 없을 만큼 감사한 분들이 많기에 이 글로 짧게나마 저의 마음을 전하고자 합니다.

먼저 바쁘신 와중에 인자한 웃음과 격려로 저의 논문심사를 해주신 김도영 선생님께 진심으로 감사드립니다. 항상 애정 어린 마음으로 2년 동안 통계학 수업을 지도해주신 임길섭 선생님, 조진남 선생님, 김동건 선생님, 그리고 한없이 부족한 저를 학문적으로 이끌어주시고 논문이 마무리되는 순간까지 세심하게 도움을 주신 송기준 선생님과 명성민 선생님께도 감사드립니다.

어려운 과정이 있을 때마다 화이팅 외쳐주신 무영오빠, 저의 쏟아지는 질문 공세에도 매번 하나하나 자세하게 설명해주신 원열오빠, 항상 변함없는 모습으로 우리의 본보기가 되어주신 은희언니께도 감사한 마음 전합니다.

지금은 졸업하고 각자의 맡은 바 자리에서 열심히 일하고 있는 선배들, 늘 친언니같이 챙겨주신 정운언니, 가장 가까운 곳에서 버팀목이 되어준 수희언니, 늘 밝은 모습으로 환하게 웃어준 영애, 내가 앞으로 나아가야 할 사회라는 출발선에서 첫 걸음을 떼는데 많은 도움이 되어준 진희, 조용하게 한결같이 마음써준 경화에게 고맙다고 말하고 싶습니다. 그리고 사랑스러운 후배들, 뭐든 들어주고 항상 든직하게 내편이 되어준 용진이, 성실하고 늘 먼저 배려해주는 마음이 참 예쁜 혜선이, 자진해서 어려운 일도 척척 도와주었던 낙훈오빠에게 너무 고맙고, 한 학기동안 바쁘다는 핑계로 많이 챙겨주지 못한 성훈씨, 성희씨, 수연이에게 미안한 마음 전합니다.

당신들과 함께해서 힘들어도 웃을 수 있었고, 당신들과 함께해서 나의 석사생활 2년이 환하게 빛날 수 있었습니다.

둘도 없는 평생지기 내 반쪽 헤미, 작은 투정까지도 다 받아주는 소중한 내 친구들 선화와 수현이, 내 인생의 감사한 선물들 경아, 유림이, 현정에게 늘 고마운 마음뿐입니다.

매일 아침 말씀 문자와 함께 때마다 잊지 않고 기도해 주었던, 내가 너무도 닮고 싶은 사람 세영언니, 언니의 따뜻한 기도 덕분에 무사히 잘 마칠 수 있었습니다.

마지막으로 처음부터 끝까지 매 순간마다 함께해주신 하나님께 감사드리며 나의 든든한 지원군인 부모님과 성은언니, 성우에게 이 작은 결실을 바칩니다. 가족들의 믿음과 격려는 제가 어떠한 상황에서도 흔들리지 않고 중심을 잡을 수 있도록 해 준, 세상에서 가장 크고 깊은 뿌리였습니다.

그런 가족들이 있어서 지금의 제가 있습니다. 많이, 아주 많이 사랑합니다.

2008년 8월

노성유 올림

차 례

제1장 서론	1
1.1 연구배경 및 목적	1
1.2 연구 내용 및 방법	2
1.3 논문의 구성	2
제 2장 Propensity score matching 방법의 고찰	3
2.1 관찰연구	3
2.2 처리할당에 대한 강한 무관성의 가정과 차원의 문제	3
2.3 Propensity score	5
2.4 Propensity score를 이용한 matching	7
2.4.1 마할라노비스 거리 matching	8
2.4.2 최근접 거리 matching	8
2.4.3 반경 내 최근접 마할라노비스 거리 matching	10
제 3장 간경변 발생 위험 인자의 재평가	13
3.1 자료에 대한 개요	13
3.2 분석에 쓰인 건강검진 항목	13
3.3 Propensity score matching 적용 전과 후의 일변량 분석	16
3.4 다중 로지스틱 회귀분석을 이용한 위험인자 선택	22
3.4.1 기존모형	23
3.4.2 보정모형	23
3.5 간경변증 발생 위험인자 재평가 결과	25
제 4장 결론 및 고찰	26
참고문헌	37
ABSTRACT	40

표 차 례

표 1. 건강검진 세부항목	14
표 2. 변수의 분리기준	16
표 3. 변수별 간경변 발생 분포	18
표 4. Propensity score matching 적용 전과 후의 변수별 일변량 분석	21
표 5. 기존모형과 보정모형의 로지스틱 회귀분석 결과	24
표 6. 기존모형과 보정모형에 대한 적합도 평가	25
표 7. 실제자료에서 위험 예측 확률 분포 비교	27
표 8. 기존모형과 보정모형을 이용한 위험 확률 추정치의 비교	30

그림 차 례

그림 1. 마할라노비스 거리 matching 알고리즘	12
그림 2. 기존모형의 ROC curve	26
그림 3. 보정모형의 ROC curve	26
그림 4. 실제자료의 위험 예측 확률 분포	28
그림 5. 기존모형과 보정모형의 위험인자 조합에 대한 확률 값의 분포	34

국문 요약

Propensity Score Matching 방법을 이용한 간경변증 위험 인자의 재평가

임상의학 분야에서 자주 수행되어지는 관찰연구(observational study)의 경우 '선택편의(selection bias)'라는 본질적 문제가 존재하게 된다. 이러한 문제들은 치료 집단과 비교집단의 공변량(covariate)들을 균형화 시켜주어 구조적인 집단 간 차이가 발생하지 않도록 하기 위해 matching 시켜주거나, 비슷한 개체끼리 층화시킨 후에 처리효과를 추정한다면 해결할 수 있다. 그러나 선택편의를 해결하기 위한 기존의 전통적인 방법들은 보정을 하기 위한 공변량의 수가 적어야 한다는 제한점이 있기 때문에 공변량 정보에 대한 스칼라(scalar)의 요약치를 제공하는 propensity score를 이용하면 왜곡된 치료효과를 효과적으로 평가할 수 있다.

본 논문에서는 propensity score matching을 이용하여 간경변증 발생의 위험인자들을 재평가하고자 1994년부터 2005년까지 건강검진센터에서 건강검진을 받은 검진자 중 병원에 내원하여 간경변증 발생 여부에 대한 진단을 받은 4,093명의 검진자료를 이용하였다. 기존 연구에서 밝혀진 간경변증에 대한 위험인자들의 재평가를 위해서 기존모형과 보정모형을 가지고 민감도, 특이도, 정확도, 위험확률 추정치 등의 측도를 통하여 비교하였다.

그 결과, 보정모형에서 선택된 6개의 인자, 즉, C형 간염항체, 알파-태아단백, B형간염 바이러스, 가족력, 음주력, 알칼리성 포스파타제 만으로도 기존모형에서 선택된 9개의 인자 못지않게 간경변증 발생 위험을 예측할 수 있었다. 이러한 결과를 통해 본 논문에서 제시하는 방법은 임상의학연구에 있어서 시간적, 경제적 비용의 측면을 고려하였을 때 유용하게 적용될 수 있으리라 판단된다.

핵심되는 말: propensity score, matching, 간경변증, 예측모형

제 1장 서론

1.1 연구 배경 및 목적

임상 의학 분야에서는 비교연구(comparative study)가 많이 수행되는데, 무작위 과정(randomization)을 거치지 않은 비교연구의 한 분야인 관찰연구(observational study)의 경우 '선택편의(selection bias)'라는 본질적 문제가 존재하게 된다. 이러한 선택편의는 치료집단과 비교대상이 되는 집단 간의 이질성으로 인해 발생하는 것으로서, 비교연구에서의 최종적인 관심인 “처리->반응”의 인과에 대한 추론을 그르치게 하거나 처리가 반응에 주는 효과를 과소 추정하거나 과대 추정하는 과오를 유발하게 된다. 따라서 공변량(covariate)들이 불균형(unbalanced) 자료일 경우 비교집단의 선정에 있어서 이를 통제해 줄 필요가 있다. 즉, 치료집단과 비교집단의 공변량들을 균형화 시켜주며 구조적인 집단 간 차이가 발생하지 않도록 하기 위해 matching 방법을 수행하거나 비슷한 개체끼리 층화시킨 후에 처리효과를 추정한다면 이러한 선택편의의 문제가 해결될 수 있다.

선택편의를 해결하기 위한 기존의 전통적인 보정방법들은 보정을 하기 위한 공변량의 수가 적어야 하기 때문에 매우 제한적이다. 그러나 공변량 정보에 대한 스칼라의(scalar) 요약치를 제공하는 propensity score는 이러한 제한점이 없다. 최근의 평가 방법론 분야에서는 왜곡된 치료효과를 효과적으로 보정하기 위해 propensity score를 이용하여 matching하는 방법, 층화(stratification)시키는 방법, 공분산(covariance)을 보정하는 방법 등의 개발에 집중되어지고 있다. 본 연구에서는 이러한 방법들 중에서 공변량들을 이용한 함수를 통해 얻어진 propensity score matching 방법을 기존의 임상적 위험요인 분석 자료에 적용해 질병발생의 위험요인으로 알려진 인자들을 재평가하고자 한다.

1.2 연구 내용 및 방법

기존의 연구 결과 간경변증 발생 위험인자라고 알려진 24개의 변수들을 이용하여 각 인자에 대한 propensity score를 구한 후, propensity score matching 방법 중 가장 matching 능력이 탁월하다고 알려진 “반경 내 최근접 마할라노비스 거리 matching(Nearest available Mahalanobis metric matching in calipers)” 방법을 적용하여 변수 별로 원 자료를 matching 시킨다. matching된 자료로부터 구해진 변수 별 오즈비(odds ratio)와 보정되지 않은 기존의 인자 별 오즈비를 비교한다. 이를 토대로 각 방법 별로 간경변증 발생 위험 예측모형을 구축한 후, 이들 모형을 통하여 예측된 위험확률을 평가한다. 이를 통하여 재평가된 위험인자들을 통해 구축된 모형의 적절성을 확인한다.

1.3 논문의 구성

제 1장 서론에서는 연구의 배경 및 목적에 대해 소개하고 연구 내용 및 방법에 대해 제시한다. 2장에서는 propensity score의 이론적 개념과 이를 이용한 propensity score matching, 그리고 그에 따른 구체적인 방법과 알고리즘을 소개한다. 3장에서는 분석에 사용된 건강검진 센터에서 조사된 자료 및 2장에서 소개한 방법으로 비교평가 하였고, 4장에서는 결론 및 고찰에 대해서 논의하여, 앞으로 진행되어야 할 연구에 대한 제안 점을 서술한다.

제 2장 Propensity score matching 방법의 고찰

2.1 관찰연구

관찰연구란 두 개 이상의 집단을 통하여 어떤 문제의 원인이나 연관된 요소들을 밝힘으로써 문제를 설명하거나 이해하려는 시도이다. 일반적으로는 문제를 가지고 있는 집단과 가지고 있지 않은 집단을 비교함으로써 이루어지고 연구자가 어떤 중재행위를 하지 않고 관찰, 조사를 하는 연구가 되기 때문에 관찰연구라고 불린다. 많은 조사에서는 집단의 특성을 설명하는 것과 마찬가지로 비교하는 것에 중점을 두는 경우가 많다.

2.2 처리 할당에 대한 강한 무관성의 가정과 차원의 문제

Propensity score matching 방법은 Rosenbaum and Rubin(1983) 등에 의해 처음으로 소개되었다. Rosenbaum과 Rubin은 결측치가 없는 완벽한 자료에서, 개체 $i(i=1,\dots,N)$ 의 propensity score 는 관찰된 공변량들 x_i 의 벡터(vector)가 주어졌을 때 비교집단($Z_i = 0$)에 대한 치료집단($Z_i = 1$)의 조건부 확률로서 정의하였는데, 즉

$$e(x_i) = p(Z_i = 1 | X_i = x_i)$$

이다. 여기서, 주어진 X 와 Z_i 는 서로 독립이라고 가정한다.

$$pr(Z_i = z_i, \dots, Z = z_N | X_1 = x_1, \dots, X_N = x_N) = \prod_{i=1}^N e(x_i)^{z_i} \{1 - e(x_i)\}^{1-z_i}$$

만약 공변량이 존재할 때 처리할당에 대한 강한 무관성의 가정 성립할 경우, propensity score를 갖는 모든 개인들의 치료집단과 비교집단의 평균 차이는 propensity score의 평균 치료효과라 할 수 있다. 이 가정은 개인들에 대한 공변량들이 주어졌을 때, 치료집단에 대한 할당여부는 성과변수들과 독립적이고(가정 1), 치료집단과 비교집단 구성원들의 치료 할당 확률의 분포는 공통의 영역에 존재한다(가정 2)는 두 가지 가정을 만족시킬 때 충족된다. (Rosenbaum and Rubin, 1983; Dehijia and Wahba, 2002; 이석원, 2003)

가정 1: $(Y_0, Y_1) \perp Z | X$ [조건부 독립성의 가정(conditional independence assumption)]
 가정 2: $0 < Pr(Z = 1 | X) < 1$ [공통 영역의 가정(common support assumption)]

위의 가정들은 만약 이들이 충족된다면, 치료집단 할당에 대한 특성을 측정하는 변수들이 충분히 존재할 경우, 이들을 통제하는 것만으로 선택편의(selection bias)가 없는 효과 추정치를 산출해 낼 수 있음을 의미한다. 이론적으로 볼 때 연구자가 확보하고 있는 공변량(X)들이 가질 수 있는 값의 가능한 모든 조합별로 치료 집단과 비교집단의 구성원들을 짝지을 수 있을 것이다. 그러나 만약 공변량들의 수가 클 경우에는 현실적으로 이것이 불가능하다. 예를 들어 연구자가 10개의 공변량을 가지고 있고 이들이 모두 2개의 값을 가지는 가변수라고 할 때 가능한 조합은 2^{10} , 즉 1024가지나 되기 때문이다. 이것이 바로 차원의 문제인데 이 문제를 해결하기 위하여 propensity score를 matching의 기준으로 적용하는 방법이 propensity score matching 이다.

2.3 Propensity score

1절에서의 두 가정을 확장하여 Rosenbaum과 Rubin은 propensity score가 주어졌을 때 처리할당 여부는 공변량들과 무관하며(정리 1), 위의 가정 1, 2가 만족되는 경우 동일한 가정이 propensity score에 대해서도 성립함을 증명하였다.(정리 2)

정리 1. $X \perp Z \mid \mathcal{A}(X)$

정리 2. $(Y_0, Y_1) \perp Z \mid \mathcal{A}(X)$, $0 < \Pr(Z=1 \mid \mathcal{A}(X)) < 1$

위의 정리들은 강한 무관성의 가정이 성립할 경우 추정된 propensity score가 동일한 비교집단 구성원을 치료집단 구성원과 짝지어 비교하여, 양 집단 간의 관찰 가능한 특성들은 동일한 분포를 가지게 되고, 결과적으로 무작위 실험과 같이 선택편의가 없는 효과를 추정할 수 있음을 의미한다. 즉, 공변량 그 자체를 기준으로 matching할 때 발생하는 차원의 문제를 공변량들을 하나의 숫자로 요약한 함수 값인 propensity score에 의하여 matching 함으로써 해결할 수 있다는 의미이다. Rosenbaum과 Rubin(1983)은 propensity score가 가진 여러 특성들을 제시 하였는데, 그 내용은 다음과 같다.

1. Propensity score는 균형점수(balancing score)이며 공변량들의 ‘the coarsest function’이고, X 는 ‘the finest balancing score’이다.
2. Propensity score의 임의의 값에서 치료집단과 비교집단의 평균 차이는 평균처리효과(average treatment effect)의 불편추정량이다.
3. 통제변수들에 관하여 강한 무관성의 가정이 성립할 때, 이러한 통제변수들의 함수 값인 propensity score에 대해서도 역시 강한 무관성의 가정이 성립한다.

이러한 사실은 propensity score가 몇 가지 보정방법을 통해 관찰연구 자료에서 편의를 줄이는데 사용할 수 있음을 의미한다. propensity score의 어떤 특정한 값에서 공변량들이 주어졌을 때 처리할당의 개념을 무시할 수 있는 상황이라면, propensity score의 값들을 가진 모든 개체들의 치료집단과 비교집단 평균차이는 평균처리 효과의 불편추정치이다. 따라서 이러한 방법으로 matching하거나 층화, 혹은 공분산 보정 등을 통해 처리효과의 불편추정량을 구하는데 사용할 수 있다. (Rosenbaum and Rubin, 1983; Dehijia and Wahba 2002; 박근화 and 이관제, 2003)

Propensity score는 결측치가 없을 때 판별분석 또는 로지스틱 회귀분석을 사용하여 추정할 수 있다. 이 두 방법을 이용하면 관측된 공변량들이 주어진 조건 하에서 처리할당에 대한 확률의 추정치를 구할 수 있다. 이 논문에서는 로지스틱 회귀에 의한 방법을 이용하였는데 판별분석은 공변량들이 다변량 정규분포를 따른다는 가정을 하는 한편, 로지스틱 회귀분석은 그러한 가정에 비교적 자유롭고 판별분석에 의한 방법보다 더 편의를 줄이는 것으로 알려졌기 때문이다(Rubin, 1979). 물론 선형확률모형(linear probability model)이 가지고 있는 이분산성(heteroscedasticity)의 문제는 실제 분석에서는 큰 문제가 아니기 때문에 선형확률모형을 사용할 것을 제안하기도 하였다(Zhao, 2000). 그러나 이러한 propensity score를 logit으로 추정하든지 선형확률모형으로 추정하든지 비슷한 결과를 산출하는 것으로 알려져 있으므로 (Dehijia and Wahbar, 2002), 본 논문에서는 일반적으로 널리 쓰이는 다음과 같은 로지스틱 회귀모형을 이용하였는데, 즉

$$\log\left(\frac{P(X)}{1-P(X)}\right) = \alpha + \sum \beta_c X_c + \sum \gamma_d X_d^2 + \sum \theta_{ef} X_e X_f$$

이다. 강한 무관성의 가정에 근거하는 propensity score matching 방법은 효과를 성공적으로 추정하기 위해서 propensity score 추정식을 특정화하는 것이 중요한 문제로 대두된다. 일반적으로 propensity score 추정식에는 독립변수들과 함께 독립변수들의 이차항(X^2)과 교호작용항(interaction term)이 포함된다. 특히, 치료집

단과 비교집단 간의 균형(balancing)을 달성했는지 여부를 반복적으로 검증한 후 균형 여부에 따라 다시 수정되는데, 이차항과 교호작용항은 주로 균형화가 달성되지 않은 독립변수들에 한해서 식에 포함되는 것이 일반적이다.

2.4 Propensity score를 이용한 matching

치료집단과 비교집단의 비교에 있어서 교락요인(confounding factor)의 존재는 편의를 발생시킨다. 두 집단의 비교에서 교락요인의 효과를 통제하기 위하여 연구자가 취할 수 있는 조치에는 다음과 같은 접근 방식이 있다.

첫째, 처리배치를 확률화로 결정한다. 이렇게 하면 치료집단과 비교집단 간의 교락요인의 체계적인 수준 차가 존재하기 어렵게 된다. 그러나 이러한 확률화 실험은 위에서 논의한 바와 같이 여러 이유로 쉽게 진행되기 어려운 점이 있다.

둘째, 치료집단과 비교집단의 구성 단계에서 교락요인을 통제하는 방법으로 matching을 생각할 수 있다. matching이란 치료집단을 구성하는 모든 관측치들마다 가장 가까운 propensity score 값을 가지는 비교집단의 관측치를 짝지우는 작업으로써 사전에 집단 간 차이가 나지 않도록 통제하는 효과적인 방법이다.

연구자들은 제한된 치료집단의 환자 수와 치료집단에 비해 훨씬 많은 비교집단의 환자수로 이루어진 상황에 자주 직면하게 되는데, 이때 matching은 관찰된 공변량들을 고려하여 치료집단의 환자들로부터 짝을 이루는 비교집단의 환자들을 선택할 때 쓰이는 가장 보편적인 방법이다. 그러나 관심 있는 공변량의 수가 많을 때에는 중요한 모든 공변량들을 통제하여 matching 시키기는 현실적으로 힘들다는 어려움이 있다. Propensity score matching은 하나의 스칼라(scalar) 변수를 이용함으로써 동시에 많은 공변량들을 통제하여 치료효과의 불편추정량을 산출하기 위한 가중치를 제공하기 때문에 이러한 어려운 점을 해결할 수 있다.

본 논문에서는 propensity score matching을 이용한 여러 가지 방법 중 Rosenbaum과 Rubin이 제시한 최근접 거리 matching 방법, 마할라노비스 거리

matching 방법, 반경 내 최근접 마할라노비스 거리 matching 방법을 고려한다.

2.4.1 마할라노비스 거리 matching

이 방법은 propensity score matching을 사용하기 전에 가장 흔하게 쓰인 방법이다. 개체들을 무작위로 정렬시키고 첫 번째의 치료집단 환자와 모든 비교집단의 환자들 간의 거리를 계산하는데, 이때, 치료집단 개체 i 와 비교집단의 개체 j 간의 거리 $d(i, j)$ 는 다음과 같이 마할라노비스 거리로 계산된다. 즉,

$$d(i, j) = (u - v)^T C^{-1}(u - v)$$

이다. 여기에서,

u : 치료집단의 개체 i 에서 선택된 변수의 값

v : 비교집단의 개체 j 에서 선택된 변수의 값

C : 비교집단의 모든 개체들로부터 선택된 공분산 행렬(matrix)

$d(i, j)$ 가 가장 작은 비교집단의 개체 j 가 치료집단의 개체 i 와 matching 될 개체로 선택되고, 선택된 쌍은 pool에서 제외된다. 이 방법의 단점은 계산된 마할라노비스 거리의 차원의 수가 증가할수록 관측치들 간의 평균거리도 함께 증가하기 때문에 모형 안에 많은 공변량들이 존재하면 가까운 개체들끼리 matching하기가 힘들다는 것이다. 반면에 propensity score는 많은 공변량들을 이용하여 하나의 요약된 스칼라(scalar)를 이용하기 때문에 이러한 단점을 보완할 수 있다.

마할라노비스 거리 matching(Mahalanobis metric matching) 방법은 위에서 언급했던 마할라노비스 거리 계산법을 이용하여 다른 공변량들에 대한 propensity score의 logit ($\widehat{\alpha}(x)$) 값을 또 하나의 공변량으로 추가하여 이용한다. 또한 Rubin은 공변량들이 다변량 정규분포를 따르고 치료집단과 비교집단이 공통된 공분산을

가질 때, 이 방법의 편의(bias)는 equal per cent bias reducing(EPBR)이 된다는 것을 보여주었다. 즉, 모든 공변량에 대해서 줄어드는 percent bias가 같아지고 공변량들의 편의(bias)가 증가하지 않게 된다. 이 방법은 개체 별 표준화된 차이는 줄여주지만 propensity score에 따른 잠재적 차이를 남긴다는 단점이 있다.

2.4.2 최근접 거리 matching

최근접 거리 matching(Nearest available matching) 방법은 치료집단과 비교집단을 무작위로 정렬시킨 후, 첫 번째 치료집단에서의 개체와 가장 가까운 propensity score를 갖는 비교집단에서의 개체를 선택한다. 선택된 두 개체들은 pool의 범위에서 제외되고, 다음 치료집단에서의 개체가 선택된다. $\hat{q}(x)$ 는 정규분포에 가깝기 때문에 propensity score는 다음의 logit으로 추정된다. 즉,

$$\hat{q}(X) = \log[(1 - \hat{e}(X))/\hat{e}(X)]$$

이다. 이 방법은 계산법을 고려할 때 가장 쉽게 적용할 수 있는 방법이다.

2.4.3 반경 내 최근접 마할라노비스 거리 matching

이 방법은 propensity score를 이용한 최근접 거리 matching 방법과 마할라노비스 거리 matching 방법을 하나로 합쳐놓은 것이다. 치료집단의 개체들을 무작위로 정렬시키고 첫 번째로 하나의 치료집단 개체를 선택한다. 그 다음, 모든 비교집단 개체들 중 치료집단 개체의 사전에 정의된 propensity score (\hat{a}_i) 혹은 propensity score에 대한 logit (\hat{a}_i)의 반경(caliper) 안에 있는 개체를 선택하고, 두 집단 간의 마할라노비스 거리를 계산한다. 가장 가까운 치료집단과 비교집단의 쌍은 matching 되어 pool 안에서 제외되고, 이 과정을 치료집단의 개체가 0이 될 때까지 반복한다. 남겨진 모든 비교집단의 개체들은 다음 치료집단의 개체가 matching 될 때에 선택되어진다. 이때, 반경의 크기는 연구자에 의해 결정되는데 Cochran 과 Rubin은 반경의 크기를 얼마나 크게 할 것인지에 관해 치료집단과 비교집단의 공변량들에 대한 분산의 평균이 적당하다고 하였고, Rosenbaum과 Rubin은 사용된 propensity score 의 logit에 대한 표준편차의 4분의 1에 해당하는 값이 적당하다고 제안하였다.(Ralph B. D'agostino, 1998)

더불어 Rosenbaum과 Rubin은 이 방법이 치료집단과 대조집단의 공변량들 간 균형을 가장 잘 맞춰주기 때문에 세 가지 방법 중 가장 좋다고 제안하였다. 본 논문에서도 반경 내 최근접 마할라노비스 거리 matching(Nearest available Mahalanobis metric matching within caliper) 방법을 적용하였는데 이 방법의 알고리즘은 다음과 같다.

step 1. 모든 개체에 대한 propensity score는 선택된 처리에 영향을 미친다고 생각되는 가능한 모든 독립변수들을 포함하는 로지스틱 회귀분석(logistic regression)을 이용하여 구한다.

step 2. 치료집단의 개체를 무작위로 정렬시키고 치료집단으로부터 첫 번째 개체를 선택한다. 미리 지정한 반경 안에 propensity score가 존재하는 비교

집단에서의 개체를 matching 될 후보군으로 지정하게 되는데, 여기서 세 가지 경우의 상황이 존재한다.

(i) 반경 안에 존재하는 비교집단의 개체가 없는 경우, 처음으로 돌아가 다음 처리집단의 개체에 matching 할 비교집단의 개체를 결정한다.

(ii) 반경 안에 존재하는 비교집단의 개체가 하나인 경우, 그 개체가 최종적인 matching 될 개체로 확정된다.

(iii) 반경 안에 존재하는 비교집단의 개체가 두 개 이상일 경우, step 3의 절차를 따른다.

step 3. 중요한 변수들에 대하여 치료집단과 비교집단 간의 propensity score에 대한 마할라노비스 거리를 계산하고, 치료집단 개체와 가장 가까운 거리를 갖는 비교집단의 개체가 최종적으로 matching 될 개체로 선택된다. 선택된 쌍은 pool에서 제외되고 이 과정은 다음 치료집단의 개체에서도 반복된다. 이때, 비교집단에 남아있는 모든 개체들은 남아있는 matching 과정에 이용할 수 있다. 이상의 알고리즘을 도식화 하면 다음의 [그림 1]과 같다.

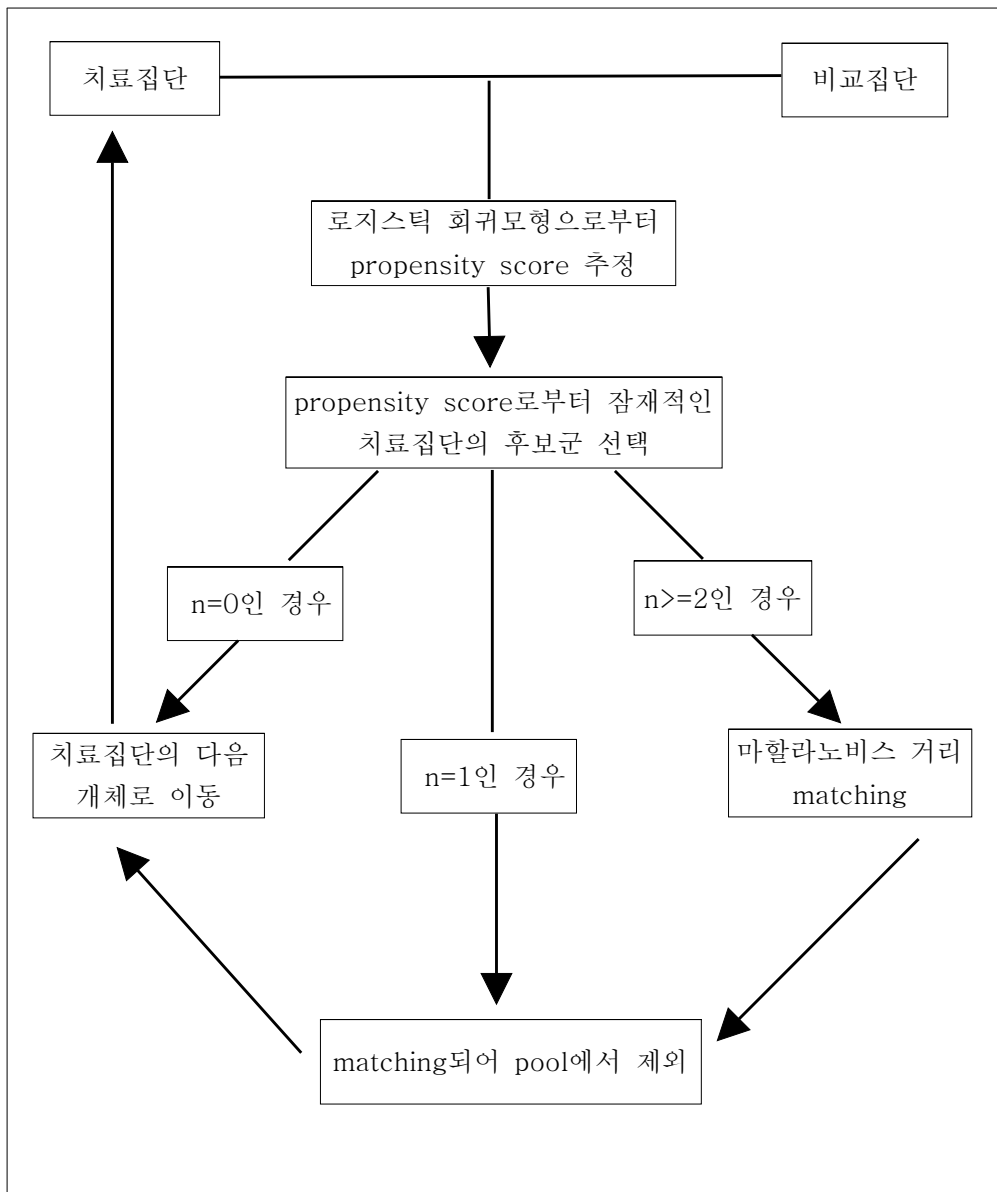


그림 1. 마할라노비스 거리 matching 알고리즘(Wuwei Wayne Freg, 2005)

제 3장 간경변 발생 위험 인자의 재평가

3.1 자료에 대한 개요

건강검진 센터에서 1994년 5월부터 2005년 9월 사이에 건강검진을 받은 총 124,121명의 건강검진 자료를 바탕으로 중복된 검진자들의 가장 최근의 정보를 담아 총 85,458명을 추출하였다. 이들 중 다시 병원에 내원하여 소화기내과 검진을 받은 8,031명 중에서 2000년부터 실시한 문진항목을 추가하고, 결측치가 존재하는 대상은 제외하여 최종적으로 4,093명을 분석 자료로 사용하였다. 4,093명 중 간경변 발생분포를 살펴보면 간경변 발생자는 501(87.76%)명이고, 비발생자는 3,592(12.24%)명이다.

3.2 분석에 쓰인 건강검진 항목

분석을 위해 기존에 알려진 간경변 발생 위험인자를 바탕으로 건강검진 항목에서 기초정보, 혈액검사, 간기능 검사, 혈청지질 검사, 종양혈청 검사, 대사 및 전해질 검사, 간염 검사, 뇨검사, 문진항목을 사용하였다. [표 1]에서 검진항목의 세부 내용을 정리하였다.

표 1. 건강검진 세부항목

검사항목	변수	설명
기초정보	sex	성별
	age	나이
혈액검사	RBC	red blood cell (적혈구)
	Hb	hemoglobin
	Hct	hematocrit
	MCV	mean corpuscular volume (평균적 혈구용적)
	MCH	mean corpuscular hemoglobin (평균 적혈구 색소량)
	MCHC	mean corpuscular hemoglobin concentration (평균 적혈구 혈색농도)
	WBC	백혈구
	LYM	lymphocyte (림프구)
	EOS	eosinocyte (호산구)
	BAS	basophil leukocyte (호염구)
	platelet	혈소판
대사 및 전해질	Na	sodium (나트륨)
	K	potassium (칼륨)
	Cl	chlorine (염소)
	CO ₂	carbon dioxide (이산화탄소)
	Ca	calcium (칼슘)
	P	phosphorus (인)
	glucose	blood glucose (혈당)
	BUN	blood urea nitrogen (혈중요소질소)
	creatinine	
	uric acid	요산

표 1. 건강검진 세부항목(계속)

검사항목	변수	설명
간기능검사	T.protein albumin	total protein (총단백)
	T.bilirubin	total bilirubin
	ALP	alkaline phosphatase
	AST	aspartate aminotransferase (아스파르테이트 아미노전이효소)
	ALT	alanine aminotransferase (알라닌아미노전이효소)
	r-GT	gamma-glutamyl transferase (감마-글루타밀전이효소)
	LDH	lactic dehydrogenase (젖산탈수소효소)
혈청지질	T.cholesterol	total cholesterol
	Tg	triglyceride (중성지방)
	HDL	high-density lipoprotein (고밀도콜레스테롤)
간염검사	HBsAg	hepatitis B virus (B형간염 바이러스)
	Anti-HBC	B형간염 C항체
	Anti-HCV	antihepatitis C virus (C형간염 항체)
종양혈청	α -FP	alpha-fetoprotein (알파-태아단백)
	CEA	carcinoembryonic (태아성암항원)
노검사	SG	specific gravity (비중)
	pH	hydrogen ion concentration (수소이온농도지수)
	protein	단백질
	urine glucose	요당
	ketone	ketone body (케톤체)
	blood	occult blood (잠혈)
	urobilinogen	
	bilirubin	
	nitrite	아질산염
	WBC	white blood cell (백혈구)
문진	family history	가족력
	drinking	음주력
	exercise	운동여부

3.3 Propensity score matching 적용 전과 후의 일변량 분석

4,093명의 전체 자료를 건강검진표의 기준을 기초로 하여 양성과 음성기준을 두어 51개의 변수를 두 개의 범주로 나누었다. 그 기준은 [표 2]에 정리하였다.

표 2. 변수의 분리 기준

변수	단위	양성기준	음성기준
sex		남	여
age	세	>=40	<40
RBC	$\times 10^6/\text{dL}$	남 <4.7 or >6.1 여 <4.2 or >5.4	남 4.7-6.1 여 4.2-5.4
Hb	g/dL	남 <14 여 <12	남 >=14 여 >=12
Hct	%	남 <42 or >52 여 <37 or >47	남 42-52 여 37-47
MCV	fL	<80	>=80
MCH	pg	<27	>=27
MCHC	g/dL	<33	>=33
WBC	$\times 10^6/\text{uL}$	>11	<=11
LYM	%	<19 or >48	19-48
EOS	%	>7	<=7
BAS	%	<1.5	<=1.5
platelet	$10^3/\text{uL}$	<130	>=130
Na	mM/L	<135 or >145	135-145
K	mM/L	<3.5 or >5.5	3.5-5.5
Cl	mM/L	<98 or >110	98-110
CO ₂	mM/L	>30	<=30
Ca	mg/dL	<8.8 or >11	8.8-11
P	mg/dL	<2.5 or >4.5	2.5-4.5
glucose	mg/dL	<70 or >110	79-110
BUN	mg/dL	<5 or >25	5-25

표 2. 변수의 분리 기준(계속)

변수	단위	양성기준	음성기준
creatinine	mg/dL	>1.4	<=1.4
uric acid	mg/dL	>6	<=6
T.protein	g/dL	<6	>=6
albumin	g/dL	<3.3 or >5.3	3.3-5.3
T.bilirubin	mg/dL	>=1.2	<1.2
ALP	IU/L	>115	<=115
AST/ALT	IU/L	>1	<=1
r-GT	IU/L	남 >55	여 <=55
		남 >35	여 <=35
LDH	mg/dL	>455	<=455
T.cholesterol	mg/dL	>=200	<200
triglyceride	mg/dL	<140	>=140
HDL	mg/dL	<30	>=30
HBsAg		양성	음성
Anti-HBC		양성	음성
Anti-HCV		양성	음성
α -FP	IU/L	>20	<=20
CEA	ng/mL	>=5	<5
SG		<1.003 or >1.030	1.003-1.030
pH		>8	<=8
protein		양성	음성
urine glucose		양성	음성
ketone		양성	음성
blood		양성	음성
urobilinogen	EU/dl	>=0.2	<0.2
bilirubin		양성	음성
nitrite		양성	음성
WBC		양성	음성
family history		유	무
drinking		하루평균 소주 1병 이상	그 외
exercise		운동한다	운동하지 않는다

이분형으로 나눈 51개의 변수 중 기존의 논문(하정윤, 2007; 유명애, 2007)에서 일변량 분석을 통해 sex, LYM, platelet, K, Ca, albumin, Tbilirubin, ALP, AST_ALT, rGT, LDH, Tcholesterol, triglyceride, HDL, HBsAg, Anti-HBc, Anti-HCV, aFP, CEA, ketone, urobilinogen, bilirubin, family history, drinking이 간경변 발생군과 비발생군 간에 유의한 차이가 있는 변수임을 밝혔다.

[표 3]은 51개의 변수에 대한 간경변 발생군과 비발생군의 발생 분포에 대한 결과이다. 결과를 보면 많은 변수들이 양성기준과 음성기준의 분포가 선택편의에 의해 음성기준으로 편중되어 있는 경향이 있음을 볼 수 있다.

표 3. 간경변증 발생여부에 관한 위험요인들의 기술통계량

변수	양성기준	비간경변(N=3,592) 빈도(%)	간경변(N=501) 빈도 (%)
sex	남	1,919 (53.42)	329 (65.67)
age	>=40	2,908 (80.79)	404 (80.64)
RBC	남: <4.7 or >6.1 여: <4.2 or >5.4	1,822 (50.72)	259 (51.70)
Hb	남: <14 여: <12	571 (15.90)	67 (13.37)
Hct	남: <42 or >52 여: <37 or >47	1,116 (31.07)	141 (28.14)
MCV	<80	63 (1.75)	6 (1.20)
MCH	<27	78 (2.17)	6 (1.20)
MCHC	<33	402 (11.19)	50 (9.98)
WBC	>11	65 (1.81)	8 (1.60)
LYM	<19 or >48	379 (10.55)	86 (17.17)
EOS	>7	42 (1.17)	5 (1.00)
BAS	<1.5	8 (0.22)	1 (0.20)
platelet	<130	52 (1.45)	85 (16.97)
Na	<135 or >145	14 (3.17)	13 (2.59)
K	<3.5 or >5.5	28 (0.78)	9 (1.80)
Cl	<98 or >110	67 (1.87)	13 (2.59)
CO ₂	>30	120 (3.34)	18 (3.59)
Ca ²	<8.8 or >11	77 (2.14)	24 (4.79)
P	<2.5 or >4.5	132 (3.67)	25 (4.99)
glucose	<70 or >110	454 (12.64)	64 (12.77)

표 3. 변수 별 간경변 발생 분포(계속)

변수	양성기준	비간경변(N=3,592) 빈도(%)	간경변(N=501) 빈도 (%)
BUN	<5 or >25	29 (0.81)	7 (1.40)
creatinine	>1.4	125 (3.48)	10 (2.00)
uric acid	>6	831 (23.13)	114 (22.75)
T.protein	<6	13 (0.36)	3 (0.60)
albumin	<3.3 or >5.3	19 (0.53)	9 (1.80)
T.bilirubin	>=1.2	513 (14.28)	13 (20.56)
ALP	>115	187 (5.21)	74 (14.77)
AST/ALT	>1	1,858 (51.73)	228 (45.51)
r-GT	남자: >55 여자: >35	714 (19.88)	184 (36.73)
LDH	>455	195 (5.43)	56 (11.18)
T.cholesterol	>=200	1,444 (40.20)	148 (29.54)
triglyceride	<140	2179 (60.66)	366 (73.05)
HDL	<30	48 (1.34)	17 (3.39)
HBsAg	양성	151 (4.20)	245 (48.90)
Anti-HBC	양성	2,117 (58.94)	403 (80.44)
Anti-HCV	양성	27 (0.75)	65 (12.97)
a-FP	>20	1 (0.03)	33 (6.59)
CEA	>=5	232 (6.46)	56 (11.18)
SG	<1.003 or >1.030	4 (0.11)	0 (0.00)
pH	>8	36 (1.00)	8 (1.60)
protein	양성	381 (10.61)	66 (13.17)
urine glucose	양성	101 (2.81)	15 (2.99)
ketone	양성	168 (4.68)	40 (7.98)
blood	양성	1,154 (32.13)	143 (28.54)
urobilinogen	>=0.2	117 (3.26)	47 (9.38)
bilirubin	양성	112 (3.12)	48 (9.58)
nitrite	양성	25 (0.70)	4 (0.80)
UWBC	양성	435 (12.11)	67 (13.37)
family history	유	198 (5.51)	243 (48.50)
drinking	하루평균 소주 1병 이상	548 (15.26)	164 (32.73)
exercise	운동한다	1,968 (54.79)	296 (59.08)

본 논문에서는 위의 24개의 변수를 포함한 자료에 propensity score matching 방법을 적용하여 적용 전에 자료에서 유의한 차이가 있다고 보였던 변수들을 적용 후 분포를 보정시킨 후에도 실제로 유의한 차이가 있는지 비교해 보았다. 그 결과 24개의 변수들 중에서 ALP, HBsAg, Anti-HCV, aFP, family history, drinking 이렇게 6개 변수만이 간경변 발생에 영향을 미치는 것으로 나타났다. [표 4]는 propensity score matching 적용 전과 후의 변수 별 일변량 분석 결과이다.

표 4. Propensity score matching 적용 전과 후의 변수별 일변량 분석

변수	propensity score matching 적용 전		propensity score matching 적용 후	
	오즈비 (95% 신뢰구간)	유의 확률	오즈비 (95% 신뢰구간)	유의 확률
sex	1.66 (1.37-2.02)	0.0001	1.13 (0.87-1.48)	0.3423
LYM	1.75 (1.36-2.26)	0.0001	1.08 (0.76-1.52)	0.6596
platelet	13.90 (9.70-19.93)	0.0001	1.37 (0.81-2.31)	0.2331
K	2.32 (1.09-4.96)	0.0243	2.65 (0.73-9.54)	0.1267
Ca	2.29 (1.43-3.66)	0.0003	0.57 (0.29-1.11)	0.0975
albumin	8.79 (3.77-20.45)	0.0001	1.25 (0.33-4.63)	0.7385
Tbilirubin	1.55 (1.22-1.96)	0.0002	1.05 (0.75-1.47)	0.7334
ALP	3.15 (2.36-4.20)	0.0001	1.68 (1.07-2.64)	0.0232
AST_ALT	0.77 (0.64-0.94)	0.0091	1.08 (0.86-1.36)	0.4802
rGT	2.33 (1.91-2.85)	0.0001	1.13 (0.85-1.49)	0.3893
LDH	2.19 (1.60-2.99)	0.0001	0.77 (0.50-1.18)	0.2349
Tcholesterol	0.62 (0.50-0.76)	0.0001	0.88 (0.69-1.11)	0.2986
triglyceride	0.52 (1.48-2.40)	0.0001	1.31 (0.95-1.80)	0.0913
HDL	2.59 (1.47-4.54)	0.0006	1.66 (0.68-4.06)	0.2621
HBsAg	0.00 (17.16-27.70)	0.0001	8.23 (5.64-12.00)	0.0001
Anti-HBC	2.86 (2.27-3.60)	0.0001	0.95 (0.91-1.59)	0.7658
Anti-HCV	19.68 (12.43-31.17)	0.0001	8.71 (4.42-17.17)	0.0001
aFP	253.21 (34.5-1855.65)	0.0001	11.88 (1.41-100.00)	0.0062
CEA	1.82 (1.33-2.48)	0.0001	1.04 (0.68-1.59)	0.8926
ketone	1.76 (1.23-2.53)	0.0016	0.84 (0.51-1.40)	0.5214
urobilinogen	3.07 (2.16-4.37)	0.0001	0.86 (0.51-1.46)	0.5909
bilirubin	3.29 (2.315-4.68)	0.0001	1.27 (0.69-2.33)	0.4401
family history	16.14 (12.87-20.24)	0.0001	3.70 (2.74-5.01)	0.0001
drinking	2.70 (2.19-3.32)	0.0001	1.64 (1.23-2.17)	0.0005

3.4 다중 로지스틱 회귀분석을 이용한 위험인자 선택

Propensity score matching을 적용하지 않았을 때 선택된 24개의 변수와 propensity score matching을 적용했을 때 선택된 6개의 변수를 각각 독립변수로 넣은 뒤 로지스틱 회귀분석을 하였다. 회귀분석을 통하여 얻어진 두 모형에 서의 최종적인 간경변증 위험인자를 찾고, 이들 모형에 대한 정확도를 파악하였다.

24개의 변수 중 9개의 위험인자가 선택된 모형(이하 기존모형)과 6개의 변수 중 6개의 위험인자가 모두 선택된 모형(이하 보정모형)을 평가하기 위해서 민감도, 특이도, 위양성율, 위음성율, 정확도를 사용하였다. 민감도는 실제 간경변 발생자를 예측을 통해서도 간경변 발생자로 분류하는 비율을 나타내는 것이고, 특이도는 실제 간경변 비발생자를 예측을 통해서 간경변 비 발생자로 분류하는 비율을 나타낸다. 위양성율은 실제 간경변 비발생자를 간경변 발생자로 잘못 분류한 비율이고, 위음성율은 실제 간경변 발생자를 간경변 비 발생자로 잘못 분류한 비율이다. 또한 모형의 적합도를 평가하기 위해서 수정된 결정계수와 ROC 곡선 면적, 그리고 Hosmer-Lemeshow 적합도 검정을 수행하였다.

3.4.1 기존모형

기존모형에서는 Anti-HCV, aFP, HBsAg, family history, drinking, platelet, ALP, rGT, triglyceride가 간경변 발생의 위험인자로 선택되었는데, 이를 이용하여 구한 간경변 발생 위험확률 예측식은 다음과 같았다. 즉,

$$P = \frac{e(x)}{1 + e(x)}$$

이다. 예측식을 기술하면,

$x = -4.2715 + (3.7149 * \text{Anti-HCV}) + (2.9965 * \text{aFP}) + (3.0329 * \text{HBsAg}) + (2.3688 * \text{family history}) + (1.1442 * \text{drinking}) + (0.8669 * \text{platelet}) + (0.8255 * \text{ALP}) + (0.4140 * \text{rGT}) + (0.5973 * \text{triglyceride})$ 를 나타낸다.

3.4.2 보정모형

보정모형에서는 Anti-HCV, aFP, HBsAg, family history, drinking, ALP 모두가 간경변 발생의 위험인자로 선택되었다. 이는 기존모형의 위험인자에서 platelet, rGT, triglyceride가 제외된 결과와 같다. 이를 이용한 간경변 발생 위험확률 예측식은 다음과 같았다. 즉,

$$P = \frac{e(x)}{1 + e(x)}$$

이다. 예측식을 기술하면,

$x = -3.7402 + (3.8234 * \text{Anti-HCV}) + (3.3602 * \text{aFP}) + (3.1471 * \text{HBsAg}) + (2.4106 * \text{family history}) + (0.8255 * \text{ALP}) + (0.4140 * \text{rGT}) + (0.5973 * \text{triglyceride})$

history)+(1.1835*drinking)+(1.0409*ALP)를 말한다.

[표 5]는 기존모형과 보정모형에 대한 오즈비, 회귀계수, 유의확률 결과이다.

표 5. 기존모형과 보정모형의 로지스틱 회귀분석 결과

변수	기존모형			보정모형		
	회귀계수 (표준오차)	유의확률	오즈비 (95% 신뢰구간)	회귀계수	유의확률	오즈비 (95% 신뢰구간)
Anti-HCV	3.71 (0.27)	0.0001	41.05 (23.96-70.32)	3.82 (0.27)	0.0001	45.76 (26.85-77.99)
aFP	2.99 (1.36)	0.0273	20.01 (1.39-286.50)	3.36 (1.25)	0.0076	28.79 (2.43-340.021)
HBsAg	3.03 (0.16)	0.0001	20.75 (15.36-28.05)	3.14 (0.15)	0.0001	23.26 (17.31-31.26)
family history	2.36 (0.15)	0.0001	10.68 (7.94-14.36)	2.41 (0.14)	0.0001	11.14 (8.32-14.90)
drinking	1.14 (0.16)	0.0001	3.14 (2.31-4.25)	1.18 (0.14)	0.0001	3.26 (2.45-4.35)
ALP	0.82 (0.25)	0.0005	2.28 (1.43-3.63)	1.04 (0.22)	0.0001	2.83 (1.83-4.37)
platelet	0.86 (0.30)	0.0036	2.38 (1.32-4.26)			
triglyceride	0.59 (0.17)	0.0003	1.81 (1.31-2.51)			
rGT	0.41 (0.16)	0.0106	1.51 (1.10-2.07)			

3.5 간경변증 발생 위험인자 재평가 결과

[표 6]은 기존모형과 보정모형을 이용하여 간경변 발생에 대한 민감도, 특이도, 위양성율, 위음성율, 정확도, ROC 곡선 면적, Hosmer-Lemeshow 적합도 검정, 수정된 결정계수를 구하여 정리한 것이다. 그 결과, 민감도와 위양성율에서는 보정모형이 약간 더 높았고 나머지 적합도 부분에서도 보정모형이 약간 낮긴 했지만 전체적으로 기존모형과 큰 차이가 없는 것을 확인할 수 있었다.

표6. 기존모형과 보정모형에 대한 적합도 평가

	기존모형(%)	보정모형(%)
Sensitivity	79.6	80.0
Specificity	90.2	89.5
False negative rate	20.4	20.2
False positive rate	9.8	10.5
Total accuracy	88.9	88.3
Area under the ROC curve(AUC)	0.897	0.891
Hosmer-Lemeshow Goodness of fit	0.6425	0.5273
Adjusted R-square	0.5323	0.5230

[그림 2]와 [그림 3]에서 보여주는 바와 같이 기존모형과 보정모형의 ROC 곡선은 거의 같은 곡선의 형태를 그리고 있음을 알 수 있다.

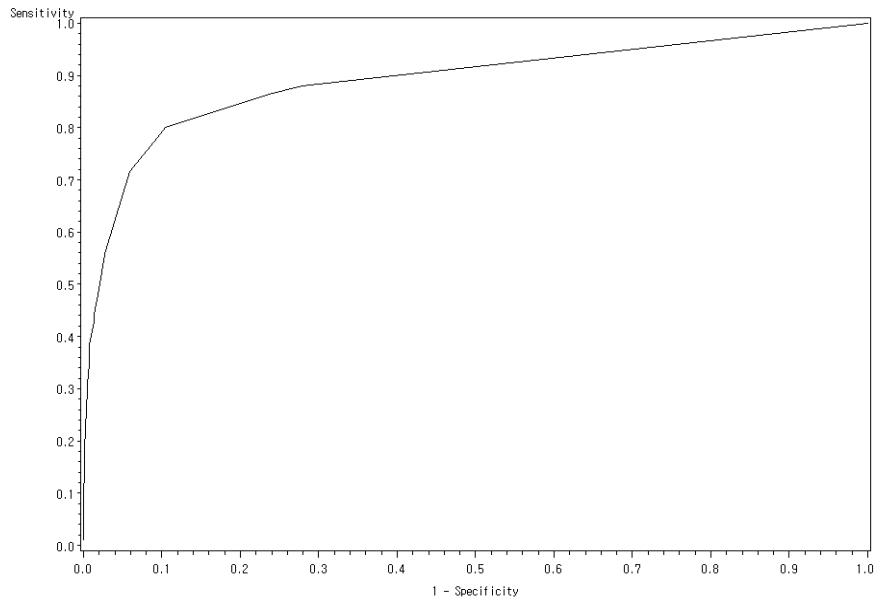


그림 2. 기존모형의 ROC곡선

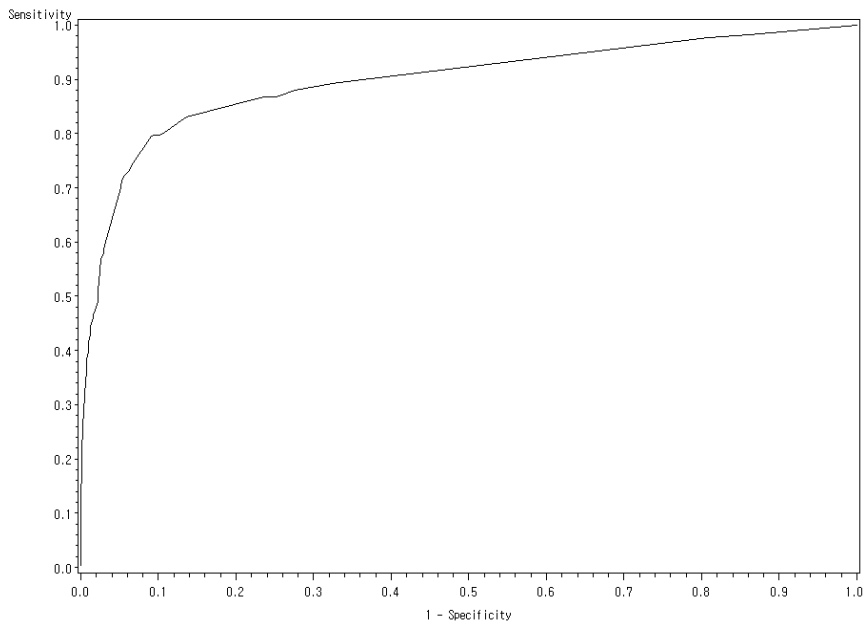


그림 3. 보정보형의 ROC곡선

[표 7]은 4,093명의 실제 자료에 대하여 두 모형을 이용했을 때 예측된 위험 확률의 분포를 비교한 결과이다. 결과를 살펴보면 확률분포를 10%씩 10개의 구간으로 나뉘었을 때 기존모형과 보정모형의 확률분포가 두 모형 간에 큰 차이가 없음을 확인할 수 있었다. 두 모형의 위험확률에 대한 빈도차가 10-19%인 구간에서 2.5%로 가장 컸고, 이를 제외한 나머지 위험 확률에서는 빈도차가 모두 1% 미만인 것을 알 수 있었다.

[그림 4]는 [표 7]의 결과를 히스토그램으로 나타낸 것이다. 여기서도 마찬가지로 기존모형과 보정모형을 실제자료에 적용했을 때, 두 모형의 위험확률에 따른 빈도분포가 전체적으로 비슷하게 나타났다.

표 7. 실제자료에서 위험 예측 확률 분포 비교

위험확률(%)	기존모형	보정모형
<10	3,234(79.01)	3,315(80.99)
10-19	133(3.25)	30(0.73)
20-29	174(4.25)	179(4.37)
30-39	157(3.84)	182(4.45)
40-49	71(1.73)	68(1.66)
50-59	57(1.39)	47(1.15)
60-69	31(0.76)	50(1.22)
70-79	38(0.93)	27(0.66)
80-89	80(1.95)	97(2.37)
>90	118(2.88)	98(2.39)
계	4,093(100)	4,093(100)

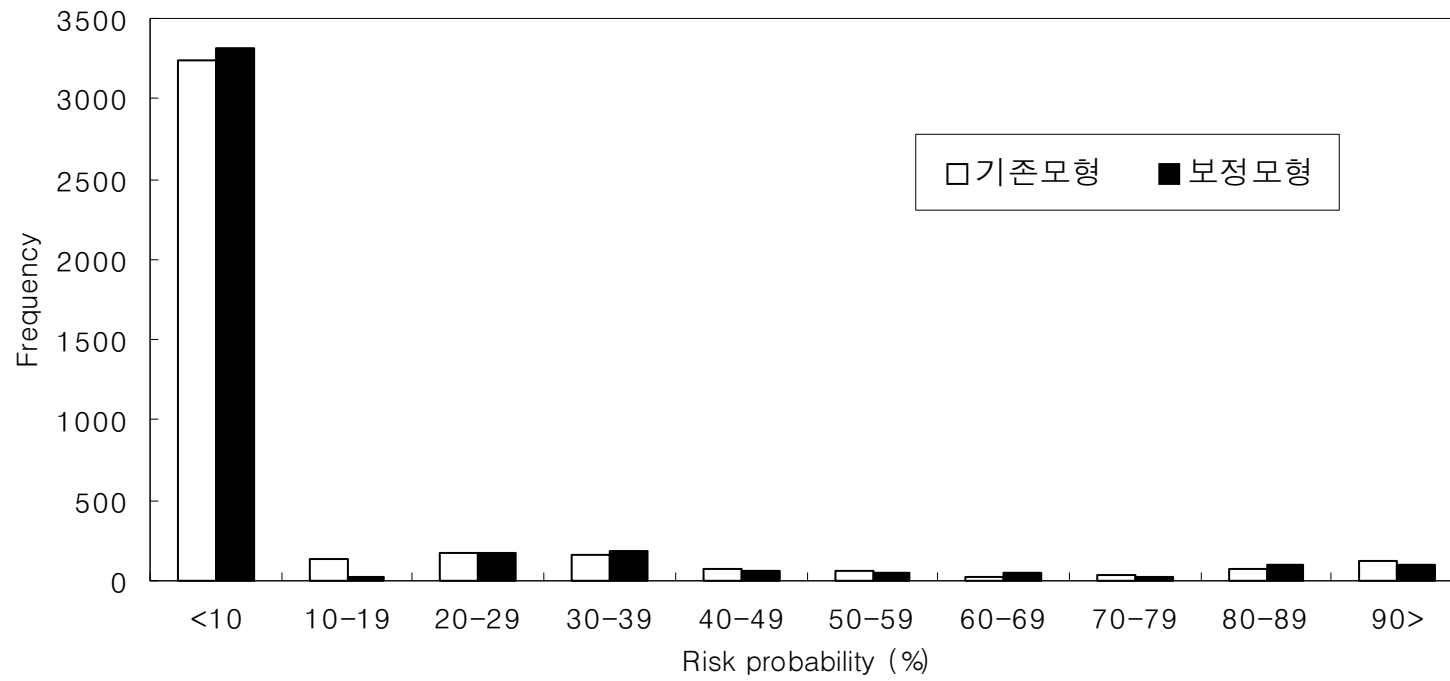


그림 4. 실제자료의 위험 예측 확률 분포

기존모형을 기준으로 보정모형의 적절성을 평가하기 위해 모든 변수들이 이분형으로 되어있는 변수 별 모든 조합을 구하고 그 조합에 대한 확률 값을 구하여 비교하였다. 모형 별 조합은 9개의 인자를 고려한 기존모형에서는 $2^9=512$ 개의 조합으로, 6개의 인자를 고려한 보정모형에서는 $2^6=64$ 개의 조합으로 구성된다. 결과적으로 기존모형이 보정모형과 중복된 6개의 위험인자 이외에 3개의 위험인자가 추가된 모형을 갖기 때문에, 기존모형에서는 보정모형과 같은 64개 조합 외에 그 64개의 조합 별로 $2^3=8$ 개의 조합이 추가된다.

본 논문에서는 64개의 조합에 대하여 기존모형에서 8개 확률 값의 중간 값과 보정모형에서 1개의 확률 값을 비교하였다.

[표 8]에서는 기존모형과 보정모형의 모든 조합 별 위험확률 추정치를 비교해 본 결과, 두 모형 간의 위험확률 추정치 값은 큰 차이가 없었고, 이는 두 모형에 대한 일표본 t-검정의 유의확률이 모두 0.05보다 큰 것을 통해 확인할 수 있었다. [그림 5]는 [표 8]을 바탕으로 64개의 조합 별 예측 확률의 분포를 나타낸 그래프이다. 그래프를 보면 두 모형의 조합 별 확률 값의 분포가 같은 양상을 보이고, 확률 값이 높은 조합일수록 보정모형이 기존모형과 거의 일치하는 반면 확률 값이 낮은 조합일 경우 보정모형이 기존모형과 미약한 차이가 존재하는 것을 볼 수 있다.

표 8. 기존모형과 보정모형을 이용한 위험확률 추정치의 비교

Number of combination	Anti-HCV	aFP	HBsAg	Family history	drinking	ALP	기존모형	보정모형	유의확률
							중앙값 (최소값, 최대값)		
1	1	1	1	1	1	1	99.998 (99.995, 99.999)	99.999	0.4376
2	1	1	1	1	1	0	99.995 (99.987, 99.998)	99.996	0.5248
3	1	1	1	1	0	1	99.993 (99.983, 99.997)	99.996	0.5249
4	1	1	1	1	0	0	99.985 (99.961, 99.994)	99.988	0.5249
5	1	1	1	0	1	1	99.977 (99.941, 99.991)	99.985	0.5249
6	1	0	1	1	1	1	99.957 (99.890, 99.983)	99.962	0.525
7	1	1	0	1	1	1	99.955 (99.886, 99.983)	99.969	0.5251
8	1	1	1	0	1	0	99.948 (99.866, 99.980)	99.958	0.5251
9	1	1	1	0	0	1	99.928 (99.816, 99.972)	99.952	0.5252
10	0	1	1	1	1	1	99.912 (99.775, 99.966)	99.939	0.5253
11	1	0	1	1	1	0	99.902 (99.750, 99.962)	99.891	0.5256
12	1	1	0	1	1	0	99.898 (99.741, 99.960)	99.912	0.5255
13	1	0	1	1	0	1	99.865 (99.657, 99.947)	99.875	0.5256
14	1	1	0	1	0	1	99.860 (99.644, 99.945)	99.899	0.5258
15	1	1	1	0	0	0	99.836 (99.582, 99.936)	99.863	0.526
16	0	1	1	1	1	0	99.799 (99.489, 99.922)	99.828	0.5263
17	0	1	1	1	0	1	99.724 (99.298, 99.892)	99.801	0.5268
18	1	0	1	1	0	0	99.693 (99.219, 99.880)	99.646	0.5271

표 8. 기존모형과 보정모형을 이용한 위험확률 추정치의 비교(계속)

Number of combination	Anti-HCV	aFP	HBsAg	Family history	drinking	ALP	기존모형	보정모형	유의확률
							중앙값 (최소값, 최대값)		
19	1	1	0	1	0	0	99.681 (99.191, 99.875)	99.714	0.5271
20	1	0	1	0	1	1	99.542 (98.841, 99.821)	99.574	0.5282
21	1	1	0	0	1	1	99.526 (98.799, 99.814)	99.656	0.5283
22	0	1	1	1	0	1	99.371 (98.412, 99.754)	99.439	0.5294
23	1	0	0	1	1	1	99.115 (97.773, 99.653)	99.115	0.5313
24	0	1	1	0	1	1	99.066 (97.651, 99.634)	99.325	0.5317
25	1	0	1	0	1	0	98.962 (97.393, 99.592)	98.804	0.8532
26	1	1	0	0	1	0	98.924 (97.299, 99.577)	99.031	0.5327
27	1	0	1	0	0	1	98.577 (96.449, 99.440)	99.623	0.5353
28	1	1	0	0	0	1	98.525 (96.322, 99.420)	98.884	0.5357
29	0	0	1	1	1	1	98.265 (95.689, 99.316)	98.274	0.5376
30	0	1	0	1	1	1	98.202 (95.536, 99.291)	98.601	0.5381
31	1	0	0	1	1	0	98.002 (95.057, 99.211)	97.533	0.5396
32	0	1	1	0	1	0	97.893 (94.795, 99.168)	98.112	0.5404
33	1	0	0	1	0	1	97.273 (93.325, 98.918)	97.166	0.5451
34	0	1	1	0	0	1	97.125 (92.979, 98.859)	97.829	0.5462
35	1	0	1	0	0	0	96.811 (92.246, 98.731)	96.196	0.5486
36	1	1	0	0	0	0	98.696 (91.982, 98.685)	96.903	0.0272

표 8. 기존모형과 보정모형을 이용한 위험확률 추정치의 비교(계속)

Number of combination	Anti-HCV	aFP	HBsAg	Family history	drinking	ALP	기존모형	보정모형	유의확률
							중앙값 (최소값, 최대값)		
37	0	0	1	1	1	0	96.125 (90.673, 98.452)	95.262	0.5539
38	0	1	0	1	1	0	95.987 (90.361, 98.396)	96.136	0.5549
39	0	0	1	1	0	1	94.747 (87.607, 97.883)	94.575	0.5646
40	0	1	0	1	0	1	94.563 (87.206, 97.807)	95.570	0.5661
41	1	0	0	1	0	0	93.986 (85.963, 97.565)	92.371	0.5706
42	0	1	1	0	0	0	93.671 (85.295, 97.432)	94.087	0.5732
43	1	0	0	0	1	1	91.294 (80.425, 96.413)	90.950	0.5925
44	0	0	1	1	0	0	88.768 (75.588, 95.295)	86.027	0.6136
45	0	1	0	1	0	0	88.400 (74.910, 95.129)	88.397	0.6168
46	0	0	1	0	1	1	84.137 (67.505, 93.146)	83.635	0.6542
47	0	1	0	0	1	1	83.645 (66.701, 92.910)	86.347	0.6568
48	1	0	0	0	1	0	82.128 (64.281, 92.171)	78.018	0.6724
49	1	0	0	0	0	1	76.970 (56.682, 89.540)	75.475	0.7208
50	0	0	0	1	1	1	73.202 (61.798, 87.492)	70.988	0.7575
51	0	0	1	0	1	0	69.924 (47.642, 85.617)	64.346	0.7903
52	0	1	0	0	1	0	69.153 (46.735, 85.162)	69.072	0.7981
53	0	0	1	0	0	1	62.839 (39.817, 81.231)	61.012	0.8633
54	0	1	0	0	0	1	61.986 (38.948, 80.670)	65.946	0.8722

표 8. 기존모형과 보정모형을 이용한 위험확률 추정치의 비교(계속)

Number of combination	Anti-HCV	aFP	HBsAg	Family history	drinking	ALP	기존모형		보정모형	유의확률
							중앙값	(최소값, 최대값)		
55	1	0	0	0	0	0	59.436	(36.433, 78.945)	52.079	0.2667
56	0	0	0	1	1	0	54.497	(31.897, 75.393)	46.354	0.9518
57	0	0	0	1	0	1	46.557	(25.404, 69.019)	42.832	0.963
58	0	0	1	0	0	0	42.577	(22.468, 65.466)	35.592	0.9205
59	0	1	0	0	0	0	41.691	(21.840, 64.639)	40.613	0.9111
60	0	0	0	1	0	0	27.639	(12.980, 49.388)	20.923	0.7659
61	0	0	0	0	1	1	20.405	(9.097, 39.566)	18.008	0.6959
62	0	0	0	0	1	0	10.100	(4.200, 22.286)	7.198	0.6041
63	0	0	0	0	0	1	7.553	(3.089, 17.253)	6.301	0.583
64	0	0	0	0	0	0	3.456	(1.377, 8.369)	2.320	0.5506

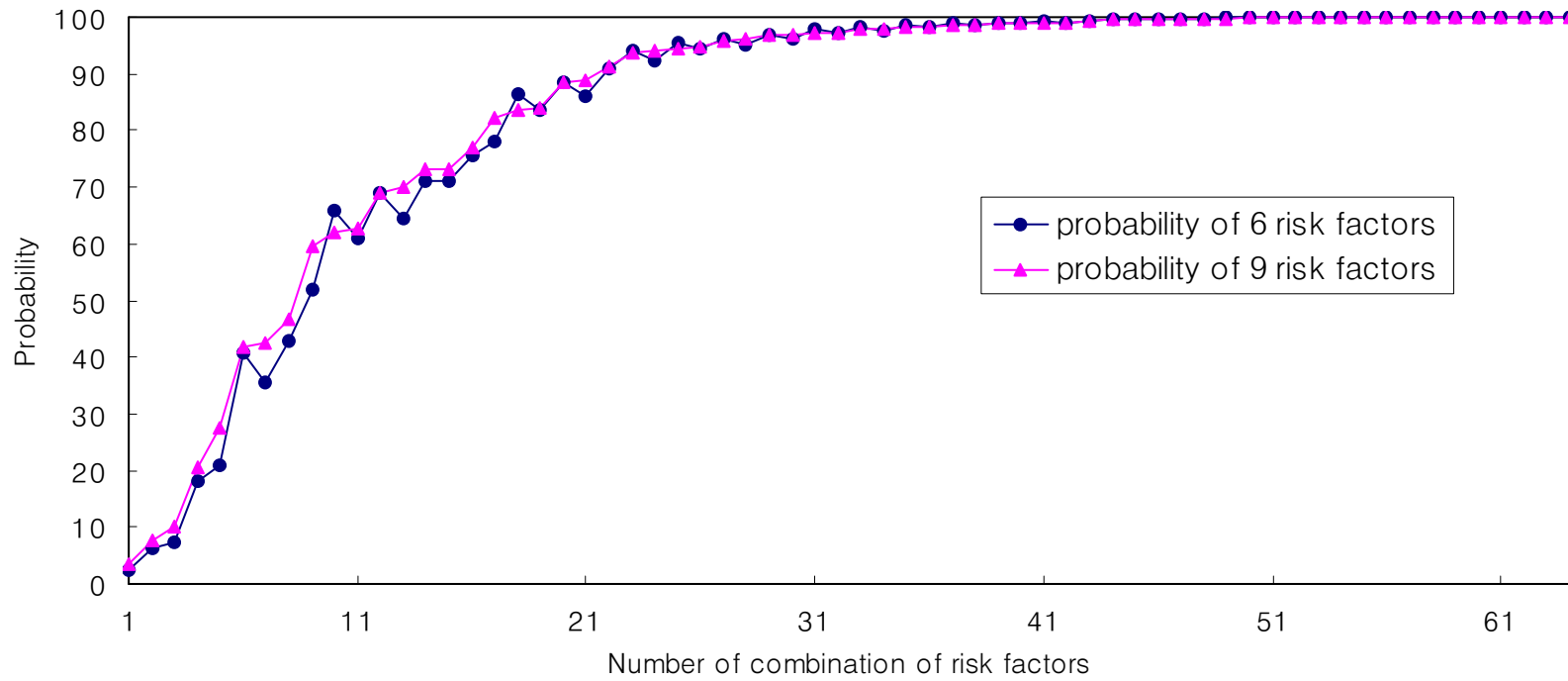


그림 5. 기존모형과 보정모형의 위험인자 조합에 대한 확률값의 분포

제 4장 결론 및 고찰

본 연구에서는 1994년부터 2005년 사이에 건강검진을 받은 자 중에서 병원에 내원하여 간경변 발생 여부에 대한 진단을 받은 4,093명을 대상으로, 음성기준으로 편중되어있는 분포를 보정시키기 위해 propensity score matching 방법을 적용하여 기존의 연구에서 밝혀졌던 간경변증을 발생시키는 위험 인자들을 재평가해보았다.

Propensity score matching 적용 전의 기존모형에서 선택된 9개의 위험인자와 적용 후의 보정모형에서 선택된 6개의 위험인자를 비교, 평가해본 결과 오히려 민감도와 위양성율에서는 보정모형이 약간 더 높았고 나머지 적합도 부분에서도 보정모형이 약간 낮긴 했지만 전체적으로 기존모형과 큰 차이가 없는 것으로 보아 보정모형에서 선택된 6개의 위험인자 만으로도 기존모형에서의 9개 위험인자 못지않게 간경변증 발생 위험을 예측할 수 있다는 결론을 내릴 수 있었다. 실제 4,093명의 자료에 대하여 두 모형을 이용했을 때 예측된 위험 확률의 분포를 비교한 결과에서도 기존모형과 보정모형의 확률분포가 큰 차이가 없게 나타났다. 마지막으로 기존모형을 기준으로 보정모형의 적절성을 평가하기 위해 모든 변수들이 이분형으로 되어있는 변수 별 조합을 구하여 그 조합에 대한 확률 값을 비교해 본 결과에서도 역시 두 모형의 64개 조합 별 확률 값의 분포가 같은 양상을 보였고 확률 값이 낮은 조합의 경우에는 보정모형이 기존모형과 미약한 차이가 존재하기는 했지만 확률 값이 높은 조합일수록 보정모형이 기존모형과 거의 일치하는 결과가 나타났다. 그러므로 간경변증 임상연구 수행 시 제한적일 수밖에 없는 시간적, 비용적인 측면 등을 고려하였을 때, 굳이 기존모형에서 선택된 9개의 위험인자를 측정하지 않고 보정모형에서 선택된 6개의 인자 즉, Anti-HCV, aFP, HBsAg, family history, drinking, ALP를 측정하는 것만으로도 충분히 간경변증을 평가할 수 있다는 결론을 얻을 수 있었다.

본 연구는 그 동안의 연구되었던 propensity score matching 방법에 대한 이론적인 정리와 함께 실제 임상자료를 이용하여 분석에 적용해 보았다. 이러한 임상

학자료, 혹은 다른 분야의 자료들에 대하여 propensity score matching 방법을 이용한다면 비실험 연구에 있어서 더욱 객관적인 비교연구가 가능하게 될 것이다.

그런데 propensity score를 이용하여 matching 할 때 공변량으로 고려해야할 변수들이 많아지게 되면, 실제 matching 된 자료의 수가 극단적으로 작아질 수도 있기 때문에 matching을 하기 전의 전체 자료수가 다소 커야 된다는 문제가 발생할 수 있다. 향후에는 이러한 제한점을 해결하기 위해 propensity score를 구할 때 적절한 공변량의 수를 결정하는 방법, 혹은 공변량이 고정되었을 때 적절한 표본의 크기를 선택할 수 있는 방법에 대한 연구가 필요할 것이라 생각된다.

참고문헌

이관제, 국세정. propensity score를 이용한 비관측 편의에 관한 연구. 한국통계학회 춘계학술대회, 1980

이현지. 성향점수(Propensity score)를 이용한 우울증 주부의 혈당 비교연구

김성은. LASSO를 이용한 간경변 발생 예측 모형 연구. 2006

하정윤. MDR을 이용한 간경변증 발생 고위험군 분류. 2007

유영애. 간경변증 발생 위험군 분류를 위한 SPAN의 유용성 평가. 2007

Rubin D. B. The use of matching and regression adjustment to remove bias in observational studies. *Biometrics*, 1973;29:185-203.

Rubin D. B. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 1979;74:318-328.

Rubin D. B. Bias reduction using Mahalanobis-metric matching. *Biometrika*, 1980;36:293-298.

Rosenbaum P. R., Rubin D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983;70:41-55.

Rosenbaum P. R., Rubin D. B. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 1984;79:516-524.

Ralph B., D'Agostino Jr. propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, 1998;17:2265-2281.

Dehijia R. H., Wahba S. propensity score matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*, 2002;84:151-161.

Rosenbaum P. R. *Observational studies*. Springer, New York, 2002

Sherry Weitzen, Lapane K. L., Toledano A. Y., Hume A. L. et al. Principles for modeling propensity score s in medical research: A systematic literature review. *Pharmaco Epidemiology and Drug Safety*, 2004;13:841-853.

Austin P. C., Mamdani M. M., Stukel T. A., Anderson G. M. et al. The use of the propensity score for estimating treatment effects: Administrative versus clinical data. *Statistics in Medicine*, 2005;24:1563-1578.

Austin P. C., Mamdani M. M. A Comparison of propensity score Methods: A case-study estimating the effectiveness of post-AMI statin use. *Statistics in Medicine*, 2006;25:2084-2106.

Cepeda M. S., Boston Ray, Farrar J. T., Storm B. L. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *American Journal of Epidemiology*, 2003;158:280-287.

Foster E. M. propensity score matching (An illustrative analysis of dose response). *Medical Care*, 2003;41:1183-1192.

Zanutto E. L. A comparison of propensity score and linear regression analysis of complex survey data. *Journal of Data Science*, 2006;4:67-91.

ABSTRACT

Reassessment of Risk Factors for the Development of Liver Cirrhosis Based on Propensity Score Matching Methods

Noh, Sung Yoo

Dept. of Biostatistics and Computing

The Graduate School

Yonsei University

In a case of observational study to be frequently conducted in a clinical trial field, an essential problem of 'selection bias' exists. These problems can be solved by executing a matching method preventing a difference between the structural groups to balance the treatment group and the control group's covariates, or if the treatment effect is estimated after stratifying similar individuals. However, because the typical methods for solving the selection bias control only the small number of covariates for correction, they have some critical limitations. In this situation it a propensity score which provides a summary value of scalar on the covariates information is used, the distorted treatment effect can be effectively examined.

In this thesis, we reassessed the risk factors for the development of liver cirrhosis using the propensity score matching method. We used the data of 4,093 subjects who received the screening test in first and visited medical

center for check up liver cirrhosis from 1994 to 2005.

In order to reassess the risk factors for liver cirrhosis which were reported in the previous studies, the predictabilities of the existing model and adjusted model were compared through a sensitivity, specificity, accuracy, and the risk probability estimate and others.

As the result of comparison, it could be concluded that only the 6 risk factors, that is, Anti-HCV, aFP, HBsAg, a family history, drinking, and ALP could predict the risk of liver cirrhosis successfully. Also, it seems to be more efficient that these approaches are applied in the clinical researches which are needed to take time and cost relatively expensive.

Key words: propensity score, matching, liver cirrhosis, prediction model