

2011/02/09 CJC
Propensity score

担当:小尾 佳嗣

What is propensity score?

多くの因子のデータをもとに
多重回帰解析を用いて算出した、
ある介入が行われる確率。

Why propensity score ?

- 最も確実に介入効果を検証するには、RCTに代表される実験的な研究が適当である。
 - ただし、RCTが現実的ではない状況がある。
 - 倫理的問題
 - 費用的問題
 - ICUの重症患者における右心カテーテルモニタリングの効果
Conors A, et al. JAMA. 1996;276:889-897
 - CADの既往/疑い症例に対するAspirinの生命予後に対する効果
Gum P, et al. JAMA. 2001;286:1187-1194
 - RCTを実行する前の予備的な仮説の検証
 - RCTでも、補正が必要となる状況もある。
 - プロトコル逸脱などによるランダム化の失敗
 - Cluster randomized control trial
- 観察研究から、可能な限りBiasを除去する手法が必要。

Why not multivariate analysis using OLS regression?

- モデル化したOutcomeと変数の関係が本当に正しいのか？
 - 複雑な関数や交互作用が存在すると、結果は不正確になる。
 - 変数が多くなれば多くなるほど、この確率は高くなる。
- イベント発生数：変数の比が、10:1より小さいときは、多変量解析によるモデルは飽和してしまう。(バイアスが大きく不正確)
 - Logistic regressionでは、7:1以下ならpropensity scoreを用いる方が正確という報告も。(Cepeda AS, et al. Am J Epidemiol. 2003;158:280-287.)



多くの因子を一つにまとめたPS という変数を使った
マッチングや層別化、weighting により、RCTに擬似的な解析
(quasi-randomized experiment)を行うことができる。

Protocol for PS analysis

I. PS analysisに適したモデルの作成

- ① PSの仮定に基づいた変数の選択とdataの収集
- ② PS調整前における背景dataの比較 [E(+)群 vs E(-)群]
- ③ PS estimation
- ④ PSによるマッチング or 層化後における背景dataの比較

II. PSを用いた回帰分析

Assumptions of PS

A) Strongly ignorable treatment assignment

割付けは観測した変数に依存し、アウトカムには依存しない。

- 変数の測定が介入の割り当て以前に行われること
 - ✓ 介入後に生じた変数の変動は、アウトカムに影響するため。
- 観測した変数によって、割付けを説明できること
 - ✓ C-indexでCheck
- 観測されていない共変数・交絡要因が割付けに影響を与えないこと
 - ✓ 理想的には、割付けや予後に影響を与える因子を全て事前に列挙する。
 - ✓ Retrospective に行う際や追加研究では、実際に観測した全変数である場合が多い。

B) Overlap assumption

比較しうる症例の重なりが存在する。

- ガイドラインに沿って割り当てが決定されていれば、同じような条件の症例がほとんどいないために、両者を比較することができない。

PS estimation

- アウトカムや割り当てに関連する可能性のある因子は全て含める。
- 集めた全ての変数を用いてExposureをアウトカムとしたLogistic regressionを行うことが一般的
 - 連続変数を正規分布に近づけるよう適切に変換、もしくは区分変数化する。
 - 例外的に、Over fittingは許容される。(10:1を超えて良い)
- モデルの作成方法は、色々あります。(後述)

それでは、過去の実際の論文を見ていきましょう！

Aspirin Use and All-Cause Mortality Among Patients Being Evaluated for Known or Suspected Coronary Artery Disease

Gum P, et al. JAMA. 2001;286:1187-1194

...

[Design and Setting]

Prospective, nonrandomized, ***observational cohort study*** conducted between 1990 and 1998 at an academic medical institution, with a median follow-up of 3.1 years.

[Patients]

Of 6174 consecutive adults undergoing stress echocardiography for evaluation of known or suspected coronary disease, ***2310 (37%) were taking aspirin***. Patients with significant valvular disease or documented contraindication to aspirin use, including peptic ulcer disease, renal insufficiency, and use of nonsteroidal antiinflammatory drugs, were excluded.

[Main Outcome Measure]

All-cause mortality according to aspirin use.

[Results]

During 3.1 years of follow-up, 276 patients (4.5%) died. ...

Strongly ignorable treatment assignment

- Data on baseline demographics, medical history, cardiovascular risk factors, and medication use (including regular use of aspirin) were **collected prospectively at the time of testing**. **All data were entered online prior to the start of the stress test** by either a physician or a trained exercise physiologist, with those entering data into the database **blinded to the hypothesis of this study as well as to the results of the subsequent stress test and stress echocardiogram**.
 - ここで選択された変数は、同様に死亡をアウトカムとして運動負荷エコー時の心拍変時能との関係を見た先行研究で測定されたもの。
 - ✓ アスピリンの使用に関わる変数として選択されたものではない。
 - 全て観察開始時点で、各変数を測定している。

PS estimation

- The propensity for aspirin use was determined *without regard to outcome, using multivariable logistic regression analysis. A full non-parsimonious model* was developed that included *34 covariates*, some of which are listed in TABLE 1. This model yielded a *c statistic of 0.83*, indicating a strong ability to differentiate between aspirin users and nonusers. *A propensity score for aspirin use was then calculated from the logistic equation for each patient.* This score ranged from 0.03 to 0.98 and, in effect, represented the probability that a patient would be using aspirin.

- 観察期間中の死亡が276例なので、34もの変数を生存解析に直接投入することは困難なため、PSを用いている。
- フルモデルでPSを算出し、C-indexも提示。
- Resultで、Interactionもモデルに組み込んでいることが記載されている。

Comparison baseline characteristics between 2 groups

Table 1. Baseline and Exercise Characteristics According to Aspirin Use*

Variable	Aspirin (n = 2310)	No Aspirin (n = 3864)	P Value
Demographics			
Age, mean (SD), y	62 (11)	56 (12)	<.001
Men, No. (%)	1779 (77)	2167 (56)	<.001
Clinical history			
Diabetes, No. (%)	388 (17)	432 (11)	<.001
Hypertension, No. (%)	1224 (53)	1569 (41)	<.001
Tobacco use, No. (%)	234 (10)	500 (13)	.001
Prior coronary artery disease, No. (%)	1609 (70)	778 (20)	<.001
Prior coronary artery bypass graft, No. (%)	689 (30)	240 (6)	<.001
Prior percutaneous coronary intervention, No. (%)	667 (29)	148 (4)	<.001
Prior Q-wave MI, No. (%)	369 (16)	285 (7)	<.001
Atrial fibrillation, No. (%)	27 (1)	55 (1)	.04
Congestive heart failure, No. (%)	127 (6)	178 (5)	.12
Medication use			
Digoxin use, No. (%)	171 (7)	216 (6)	.004
β-Blocker use, No. (%)	811 (35)	550 (14)	<.001
Diltiazem/verapamil use, No. (%)	452 (20)	405 (10)	<.001
Nifedipine use, No. (%)	261 (11)	283 (7)	<.001
Lipid-lowering therapy, No. (%)	775 (34)	380 (10)	<.001
ACE inhibitor use, No. (%)	349 (15)	441 (11)	<.001
Cardiovascular assessment and exercise capacity			
Body mass index, mean (SD), kg/m ²	29 (5)	30 (7)	<.001
Ejection fraction, mean (SD), %	50 (9)	53 (7)	<.001
Resting heart rate, mean (SD), beats/min	74 (13)	79 (14)	<.001
Resting blood pressure, mean (SD), mm Hg			
Systolic	141 (21)	138 (20)	<.001
Diastolic	85 (11)	86 (11)	.04
Purpose of test to evaluate chest pain, No. (%)	300 (13)	468 (12)	.31
Mayo Risk Index ≥1, No. (%)†	2021 (87)	2517 (65)	<.001
Peak exercise capacity, mean (SD), METs			
Men	8.6 (2.4)	9.1 (2.6)	<.001
Women	6.6 (2.0)	7.3 (2.1)	<.001
Heart rate recovery, mean (SD), beats/min	28 (11)	30 (12)	<.001
Ischemic ECG changes with stress, No. (%)	430 (24)	457 (14)	<.001
Echocardiographic left ventricular ejection fraction ≤40%, No. (%)	321 (14)	226 (6)	<.001
Stress-induced ischemia on echocardiography, No. (%)	495 (21)	436 (11)	<.001
Fair or poor physical fitness for age and sex, ¹³ No. (%)	714 (31)	1248 (38)	.26

*MI indicates myocardial infarction; ACE, angiotensin-converting enzyme; MET, metabolic equivalent task; and ECG, electrocardiogram.

†The Mayo Risk Index is described in the "Methods" section.

Table 3. Selected Baseline and Exercise Characteristics According to Aspirin Use in Propensity-Matched Patients*

Variable	Aspirin (n = 1351)	No Aspirin (n = 1351)	P Value
Demographics			
Age, mean (SD), y	60 (11)	61 (11)	.16
Men, No. (%)	951 (70)	974 (72)	.33
Clinical history			
Diabetes, No. (%)	203 (15)	207 (15)	.83
Hypertension, No. (%)	679 (50)	698 (52)	.46
Tobacco use, No. (%)	161 (12)	162 (12)	.95
Cardiac variables			
Prior coronary artery disease, No. (%)	652 (48)	659 (49)	.79
Prior coronary artery bypass graft, No. (%)	251 (19)	235 (17)	.42
Prior percutaneous coronary intervention, No. (%)	166 (12)	147 (11)	.25
Prior Q-wave MI, No. (%)	194 (14)	206 (15)	.52
Atrial fibrillation, No. (%)	21 (2)	24 (2)	.65
Congestive heart failure, No. (%)	79 (6)	89 (7)	.43
Medication use			
Digoxin use, No. (%)	115 (9)	114 (9)	.94
β-Blocker use, No. (%)	352 (26)	358 (26)	.79
Diltiazem/verapamil use, No. (%)	223 (17)	223 (17)	>.99
Nifedipine use, No. (%)	127 (9)	144 (11)	.28
Lipid-lowering therapy, No. (%)	281 (21)	271 (20)	.63
ACE inhibitor use, No. (%)	209 (15)	214 (16)	.79
Cardiovascular assessment and exercise capacity			
Body mass index, mean (SD), kg/m ²	29 (6)	29 (6)	.83
Ejection fraction, mean (SD), %	51 (8)	51 (9)	.65
Resting heart rate, mean (SD), beats/min	77 (13)	76 (14)	.13
Resting blood pressure, mean (SD), mm Hg			
Systolic	141 (21)	141 (21)	.68
Diastolic	85 (11)	86 (11)	.57
Purpose of test to evaluate chest pain, No. (%)	153 (11)	159 (12)	.72
Mayo Risk Index ≥1, No. (%)†	1108 (82)	1110 (82)	.92
Peak exercise capacity, mean (SD), METs			
Men	8.7 (2.5)	8.3 (2.5)	.01
Women	6.5 (2.0)	6.7 (2.0)	.13
Heart rate recovery, mean (SD), beats/min	28 (12)	28 (11)	.82
Ischemic ECG changes with stress, No. (%)	231 (22)	223 (21)	.64
Echocardiographic left ventricular ejection fraction ≤40%, No. (%)	147 (11)	156 (12)	.50
Stress-induced ischemia on echocardiography, No. (%)	239 (18)	259 (19)	.32
Fair or poor physical fitness for age and sex, ¹³ No. (%)	445 (33)	459 (34)	.57

*MI indicates myocardial infarction; ACE, angiotensin-converting enzyme; MET, metabolic equivalent task; and ECG, electrocardiogram.

†The Mayo Risk Index is described in the "Methods" section.

マッチング後、男性のpeak exercise capacity以外は全て有意差なし！

Result

単変量解析では有意差は認められなかったが、多変量解析では有意になっている。

PSによるmatchingを行った後に様々なモデルで検討しても、その結果はほぼ変わらず認められることから、アスピリンの有効性がより確かなものとなった。

Figure 1. Kaplan-Meier Curve Relating Aspirin Use to Time to Death Among Propensity-Matched Patients

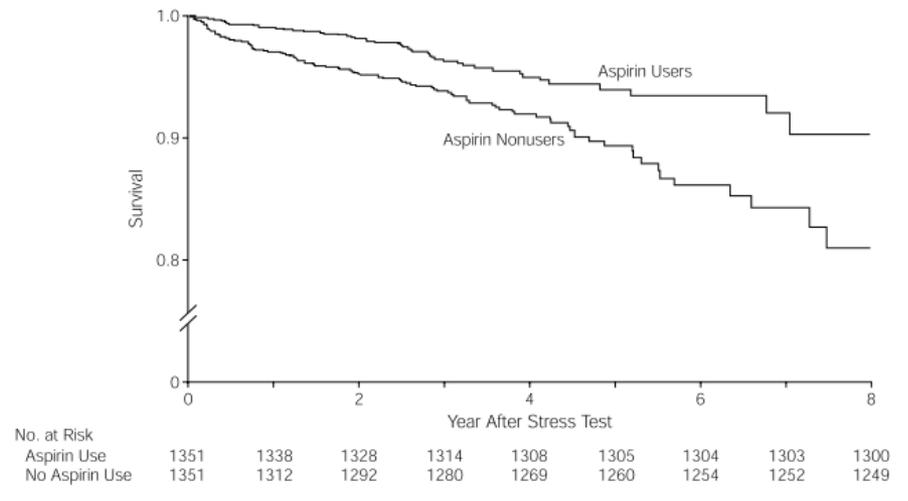


Table 2. Cox Proportional Hazards Analyses of Time to Death Among Patients Using Aspirin (N = 6174)*

Model	Hazard Ratio (95% CI)	P Value
Unadjusted	1.08 (0.85-1.39)	.50
Adjusted for age and sex	0.75 (0.58-0.96)	.02
Adjusted for age, sex, and history of CAD	0.57 (0.44-0.74)	<.001
Multivariable adjusted†	0.67 (0.51-0.87)	.002
Adjusted for age and sex among prespecified strata		
Normal LV function	0.75 (0.56-1.01)	.06
Abnormal LV function	0.54 (0.34-0.84)	.006
No history of prior CABG surgery	0.74 (0.54-1.08)	.06
History of prior CABG surgery	0.56 (0.35-0.89)	.01

*CI indicates confidence interval; CAD, coronary artery disease; LV, left ventricular; and CABG, coronary artery bypass graft.

†Adjusted for age, sex, body mass index, resting heart rate, resting systolic blood pressure, use of antihypertensive medications, digoxin, β -blockers, lipid-lowering therapy, nitrates, angiotensin-converting enzyme inhibitors, dihydropyridine and nondihydropyridine calcium channel blockers, congestive heart failure, smoking, atrial fibrillation, left and right bundle-branch block, pathologic Q waves, prior CABG surgery, prior percutaneous coronary intervention, chronic lung disease, peripheral vascular disease, exercise capacity, chronotropic response, heart rate recovery, left ventricular ejection fraction, echocardiographic evidence of myocardial ischemia, and failure of the left ventricle to decrease in size with exercise.

Table 4. Cox Proportional Hazards Analyses of Aspirin Use and Mortality Among Propensity-Matched Patients (n = 2702)*

Model	Hazard Ratio (95% CI)	P Value
Unadjusted	0.53 (0.38-0.74)	.002
Adjusted for propensity	0.53 (0.38-0.74)	<.001
Adjusted for propensity and selected variables†	0.59 (0.42-0.83)	.002
Adjusted for propensity and all covariates‡	0.56 (0.40-0.78)	<.001

*CI indicates confidence interval.

†Selected variables included prior coronary artery disease, prior coronary artery bypass grafting, prior percutaneous intervention, and ejection fraction $\leq 40\%$.

‡For a list of covariates, see Table 2 footnote (†).

Pitfall

Comparison of the baseline characteristics

A) マッチング後における解析

➤ 独立したサンプルではないことを考慮する必要がある！

- ✓ Paired t-test (Not t-test)
- ✓ Willcoxon matched pairs signed-rank test (Not Willcoxon rank sum test)
- ✓ McNemar test (Not Chi squared test)
- ✓ Absolute standardized difference in covariate means
 - Sample sizeも考慮に入れた比較方法

B) マッチング以外の手法を用いるときの背景比較

- 一般的には5群のQuintileに分けて、各層で比較
- ✓ Nが少ないときは3～4層になることも
 - ✓ 通常の比較方法でOK

Conclusions

- Benefits:
 - ある介入効果に注目した観察研究において、quasi-randomized experimentを行うことが出来る。
 - 多くの変数での調整が必要であるにも関わらず、これに対してアウトカム達成数が比較的少ない際に有用である。
 - マッチングを用いたデザインの研究では、変数をひとつに統一することが可能であるため、非常に便利。
- Limitation:
 - 観察できなかつた因子に関しては調整不可能。
 - 均等な分布を得るためには、結局大きなサンプルが必要になることも。
 - マッチングや層化に恣意性が残る。

各種PS model作成法

Author	Procedure
Rosenbaum & Rubin (1984)	<ol style="list-style-type: none">① 変数選択法 [$t > 1.5$ (df 20で$P = 0.15$)] でPS model を作成② E(+)群 vs E(-)群 で有意な変数があれば、2乗項や2変数間の交互作用項を含め、再度PSを作成する。
Eichler & Lechner (2002)	最初から2乗項と全ての1次交互作用項を含める。
Dehejima & Wahba (1999)	<ol style="list-style-type: none">① 全ての1次項で作成したPSで層別化し、各層で群間比較を行う。② 群間に有意差があれば、より高次の項や交互作用項を含める。
Hirano & Imbens (2001)	<ol style="list-style-type: none">① 有意差をいくつか設定 (e.g. $t=0, 1, 2, 4, 8, 16, \infty$)② 各変数の1次項で、介入に対するLogistic regressionを行う③ 同様に 各変数の1次項で、Outcomeの適切なregressionを行う④ 一方でも有意な変数を用いて、設定した各t値におけるPSを作成⑤ それぞれのPSで解析を行う。(Sensitivity analysisのようなもの)