

予後因子偏りの調整：Covariate Adjustment

randomized clinical trial を行ったにもかかわらず、患者の予後因子に偏りよう生じることがあります。例えば治療群で重症例が少なく、無治療群で重症例が多く、結果は治療群の方が若干成績が良いという状態です。たしかに randomization を行ったのですが、数が十分でなかったせいか偏ってしまいました。我々はこの治療が有効であったと結論してしまってもよいのでしょうか？

Adjusted Comparison: Poststratification

1974年 Armitage と Gehan によって行なわれた破傷風抗毒素血清の臨床試験についてみてみましょう。

	患者数	生存数	生存率
全員	79	30	
治療群	41	21	51.2 ± 7.8
無治療群	38	9	23.7 ± 6.9
差			27.5 ± 10.4

79人の患者さんを対象とし49人、およそ6割の患者さんが亡くなっており、破傷風は極めて致死率の高い病気であることがわかります。抗毒素血清を投与された49人中21人が助かり(51%)、投与を受けなかった38人中9人(24%)しか助かっていません。単純に $p = x/n$, $SE(p) = \sqrt{pq/n}$, としますと point estimate と standard error を算出することができます(上表)例; $[0.512 \times (1 - 0.512) / 41] = 0.078$ 。そして治療効果は、estimate = $51.2 - 23.7 = 27.5\%$, standard error = $\sqrt{(7.8^2 + 6.9^2)} = 10.4\%$ とすることができます。

$$t = (d - 0) / se = (0.275 - 0) / 0.104 = 2.6 > 1.96$$

よって抗毒素は患者さんの生存に関係すると結論できます。

しかし実際には治療群には発症後進行の早い重症例が多いかもしれません。逆に無治療群に重症例が多ければ抗毒素の効き目がなくても治療軍の生存率が高くなってしまいかもしれません。それでは治療群と無治療群で重症度に偏りがなかったか検討してみましょう。

	患者数	重症患者数	重症患者の割合
治療群	41	28	68.3 ± 7.3
無治療群	38	30	78.9 ± 6.6
差			-10.7 ± 9.8

どうやら無治療群に重症患者さんが多そうです。これを重症患者さんの偏り(下表)に応じてウェイトをかけて adjustment してみましょう。

	重症	軽症	Adjusted difference
治療群	10/28 = 35.7 ± 9%	11/13 = 84.6 ± 10%	
無治療群	4/30 = 13.3 ± 6%	5/8 = 62.5 ± 17%	
差	22.4 ± 11.0%	22.1 ± 19.8%	
ウェイト	58/79 = 0.734	21/79 = 0.266	
ウェイト平均	0.734 x 22.4	+ 0.266 x 22.1	= 22.3 ± 9.6%

元々の差は 27.5%の改善でしたから重症患者さんの偏りに基き調整したところ 22.3%

にまで減少しました。しかしそれでも 1.96 より大きく治療効果はあります。臨床研究の表 1 は 2 群間の臨床的特徴を示してあることが多く (potentially influential baseline variables (covariates)、偏りがあった場合上のように adjustment を行なうことが大切です。しかし予後を左右する因子は数多くあり、ある因子はよく知られていても他の因子は知られてさえいないかもしれません。また予後因子に数も無数に近くあるかもしれません。どれを予後因子として選択し adjustment するかは単に研究者の裁量にかかっています。基本的には研究をデザインした時点で covariates を設定しデータを収集するべきですが、解析してからそのような covariates が浮かび上がることがあるのが現実です。特に randomization を行なったとしても対象数が少ない場合には確率的に偏りを生じることもあり、これは confounder などの bias を生じる原因となります。そのような場合には重要な予後因子の偏りを調整しておく必要があります。

次に University Group Diabetes Program (UGDP)の調査(1970)によるデータを参考に adjustment について考えてみましょう。

	placebo	Tolbutamide	Insulin (standard)	Insulin (variable)
対象数	205	204	210	204
心血管疾患				
心筋梗塞	0	10	3	2
突然死	4	4	4	5
他の心疾患	1	5	1	2
心以外の血管疾患	5	7	5	3
心血管疾患合計	10	26	13	12
非心血管疾患				
癌	7	2	4	2
上記以外の原因	3	2	2	3
原因不明	1	0	1	1
全ての死亡	21	30	20	18
心血管疾患 (%)	4.9	12.7	6.2	5.9
全て (%)	10.2	14.7	9.5	8.8

上の表でインスリン以外の治療、すなわちプラシーボとトブタミンに焦点をあてて考えてみます。破傷風の例と異なり、糖尿病の場合、死亡までの時間は非常に長くなります。上のデータは 5 年間かけて参加者を募り、平均 8 年半かけて集めたデータです。しかも各個人の経過観察期間は異なっています。しかしプラシーボとトブタミンに関してはランダムに振り分けているため基本的にはバイアスを生じません。心血管疾患による死亡に着目するとプラシーボよりトブタミンの方がだいぶ多くなっています。本当にトブタミンが悪いのでしょうか、それとも covariates に偏りがあったのでしょうか？実際のデータをみてみましょう。

	重症 (リスク因子2つ以上)	軽症 (リスク因子2つ未満)	Adjusted difference
治療群	18/114 = 15.7 ± 3%	5/75 = 6.7 ± 3%	23/189 = 12.2 ± 2%
無治療群	8/97 = 8.2 ± 3%	1/88 = 1.1 ± 1%	9/185 = 4.9 ± 2%
差	7.6 ± 4.4%	5.6 ± 3.1%	7.3 ± 2.2%
ウエイト	211/374 = 0.546	163/374 = 0.436	
ウエイト平均	0.564 x 7.6	+ 0.436 x 5.6	= 6.7 ± 2.8%

指摘の通りトプタミン治療群に重症例が偏っていました。そこで先と同様に adjustment すると、2者間の差は7.3%から6.7%まで低下しています。元々ランダムに振り分けているため偏りはさほど極端にはならないのが普通です。このような現象は多くの臨床試験で認めることができ、大抵 adjustment を行なっています。しかし結果が逆転してしまうようなことは対象数がある程度以上多い限りほとんど無いはずです。

1980年に行なわれた PARIS study について検討してみましょう。少なくとも1回以上心筋梗塞を経験した患者さんを対象に、アスピリン+ペルサンチン投与群(P/A)、アスピリンのみ投与群(ASA)、プラシーボ群(PLB)に振り分けました。それぞれの人数は810、810、406人でした。ここではP/AとPLBの死亡率の差に焦点をあてて検討していきます。

	患者数	死亡数	死亡率
ペルサンチン・アスピリン	810	87	10.7 ± 1.1
プラシーボ	406	52	12.8 ± 1.7
差			-2.1 ± 2.0

若干治療群の方で死亡率が低くなっていますが、有意な差ではありません。治療群とプラシーボ群で予後因子に偏りがいかどうか検討したところ、特殊心電図の異常が治療群に多い傾向にありました。

	患者数	心電図異常	率
ペルサンチン・アスピリン	810	63	7.8 ± 0.9
プラシーボ	406	20	4.9 ± 1.1
差			2.9 ± 1.4

確かに僅かながら特殊心電図異常が治療群に多い傾向にあります。前と同じように adjustment を行なってみましょう。

	心電図異常あり	心電図異常なし	Adjusted difference
治療群	10/63 = 15.9 ± 5%	77/747 = 10.3 ± 1%	
プラシーボ群	4/20 = 20.0 ± 9%	48/386 = 12.4 ± 2%	
差	-4.1 ± 10.1%	-2.1 ± 2.0%	
ウエイト	83/1216 = 0.068	1133/1216 = 0.932	
ウエイト平均	0.068 x -4.1	+ 0.932 x -2.1	= -2.3 ± 2.0%

2.1%から2.3%まで増えました。しかし結論をひっくり返す程大きな改善ではありません。

さていくつかの例で偏りの調整を練習してきました。偏りを調整(adjustment)して、よ

り信頼度の高いデータに仕上げるのは非常に大切なことですが、この計算を行なってもそんなに大きな影響はありませんでした。しかしながら、いくら randomization を行なっても covariate に偏りを生じることがありえるので、そのような場合には adjustment を行なうべきです。

サブグループ解析

臨床試験を行ない予想した結果が単純に得られればよいのですが、必ずしもそうとは限りません。例えば先の UGDP の例で、心血管疾患による死亡率を示した下の表をみてください。

	トブタミン	プラシーボ	χ^2 test
53 歳以下の男性	5/26 = 19.2%	1/28 = 3.6%	1.9
54 歳以上の男性	6/37 = 16.2%	6/36 = 16.7%	0
53 歳以下の女性	1/71 = 1.4%	1/85 = 1.2%	0
54 歳以上の女性	14/70 = 20.0%	1/58 = 3.4%	6.7 (p < 0.01)

基本的にトブタミンとプラシーボ群の間に有意差はありませんが、54 歳以上の女性に限って見た場合、トブタミン使用は有意に心血管疾患による死亡率を増加させています。また若い男性においてもその傾向は認められますが、人数が少なく統計学的有意差は得られていません。このようにサブグループ解析は非常に重要なのですが、しばしば人数が少なくなり過ぎて統計学的に解析しにくくなります。しかし、サブグループ解析の結果をもって次の研究計画を組む事も可能です。例えば先に挙げた PARIS study の結果より最近心筋梗塞を起した患者さんに対するアスピリンの効果が大きかったため、PARIS II においては最近心筋梗塞を発症した患者さんに焦点をあてて研究しています。

複雑解析

破傷風の解析は比較的単純なものでした。比較的短期間で結果がでたからです。一方UGDPの例では経過観察期間なども患者さんによってまちまちでした。そこで生存曲線を用いて比較したらどうなるのでしょうか？読む人にとっては理解しやすいと思います。Cox model (Cox 1972) も広く用いられている解析法です。Proportional Cox model では、2つの曲線の形が似ており、さらにその対数がお互い比で表せることを前提としています。例えば外科治療を行なうと最初死亡が多く、一定期間を乗り越えてしまえば晩期死亡は少ないのに対して、内科治療では常に一定の死亡がみられるとすれば、両曲線は交叉するかもしれません。このような場合両者を Proportional Cox model で比較することは不適切であるということです。Cox model を用いた解析は multiple regression のそれと似た効用がありますが、Cox model では censor を含んだデータでも解析できる点が大きな強みです。