

# PSMATCH Procedure × ニューラルネットワークの奇跡のマッチング ～Real World Dataの利用可能性を忖度する～

○馬場崇充、藤原正和、北西由武  
(塩野義製薬株式会社 解析センター)

## PSMATCH Procedure × Neural Network = Miracle Encounter ～Conjecturing the usability of Real World Data～

Takamichi Baba, Masakazu Fujiwara, Yoshitake Kitanishi  
Shionogi & Co., Ltd. Biostatistics Department

## 要旨:

観察データにおける比較による推測のために、PSMATCH Procedureによる傾向スコアマッチングが実装可能となった。傾向スコアをニューラルネットワークで推定することにより、より適切な観察データにおける比較が行えることを示す。

キーワード: 傾向スコアマッチング, ニューラルネットワーク, Real World Data

# Agenda

- リアルワールドデータをめぐる世界の動向
- 観察研究における比較・傾向スコアマッチングとは
- PSMATCH Procedureによる傾向スコアマッチング
- ロジスティック回帰モデルは傾向スコアの推定において最も適切か？
- IMSTAT Procedureにおけるneuralステートメントによるニューラルネットワークの実装
- シミュレーション
- 考察

## ICH Reflection on “GCP Renovation”: Modernization of ICH E8 and Subsequent Renovation of ICH E6

2. Proposed Annex 2: Non-Traditional Interventional Trials and/or data sources. Trial designs such as pragmatic clinical trials would be included in this annex, as would real world data sources to supplement or possibly replace new data collection within the trial itself. The study objectives could include evidence generation for regulatory review of approved products as well as for broader research questions, as appropriate. Principles for protocol compliance and trial monitoring would reflect the fact that approved, marketed products with better-documented and better-known safety profiles are being studied.

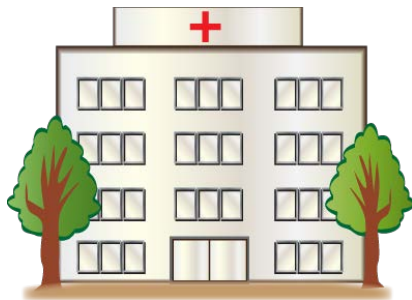
RWDを用いたPragmatic trialの検討  
→実臨床における薬の効果をみる

3. Proposed Annex 3: Non-Traditional Trial Designs. This annex would include designs other than RCTs and may include observational studies, patient registries, and other non-traditional trial designs that rely heavily on alternative data sources (e.g., EHRs, claims data, etc.). The studies may be designed to generate findings for important research objectives regarding health care practice and policy but could also be used to address regulatory questions (e.g., concerning product safety post-marketing). Principles for protocol compliance and trial monitoring laid out in this annex would be consistent with the data source and also, as in proposed annex 2, reflect the fact that marketed products with better-known safety profiles are being studied.

レジストリ, 電子医療情報を用いたNon-traditionalな臨床試験デザインの検討

ランダム化比較試験の一般化可能性の観点から, リアルワールドデータから得られるエビデンスを用いて, 一般化可能性を補強しようとする流れがある。

治験・臨床研究に最大限活用することを目的として構築される日本のレジストリネットワーク。  
※レジストリ: 患者が何の疾患で、どのような状態で存在しているかを集めた**観察データベース**



- (1)市場調査・治験の実施可能性調査
- (2)治験・製造販売後臨床試験の実施計画作成
- (3)治験・製造販売後臨床試験の候補患者のリクルート
- (4)例えば希少疾患における治験の対照群としての活用
- (5)製造販売後調査



第4回次世代医療機器開発推進協議会(H29.3.15)資料5より

観察データを利用していこうという流れはますます加速していくと考えられる。  
臨床試験から得られるデータとは異なる観察データをどう活用していけばいいのか？

活用を検討しているが、臨床試験のような比較ができない...どうすれば...



SAS STAT 14.2より新たにPSMATCH Procedureが実装

観察データにおける  
比較が可能に！



利用しているとある疑問が...傾向スコア推定にロジスティック回帰モデルを用いることは常に適当か？



リアルワールドデータの活用のために傾向スコアマッチングを学び、適用し、浮かんできた問題を機械学習を用いて克服していく！

## 設定

- 処置群 (treatment group) ... 製薬会社の臨床試験における実薬群
- 対照群 (control group) ... たとえば患者レジストリデータ

## Notation

$Y(0)$ : 対照群に属していれば得られる応答

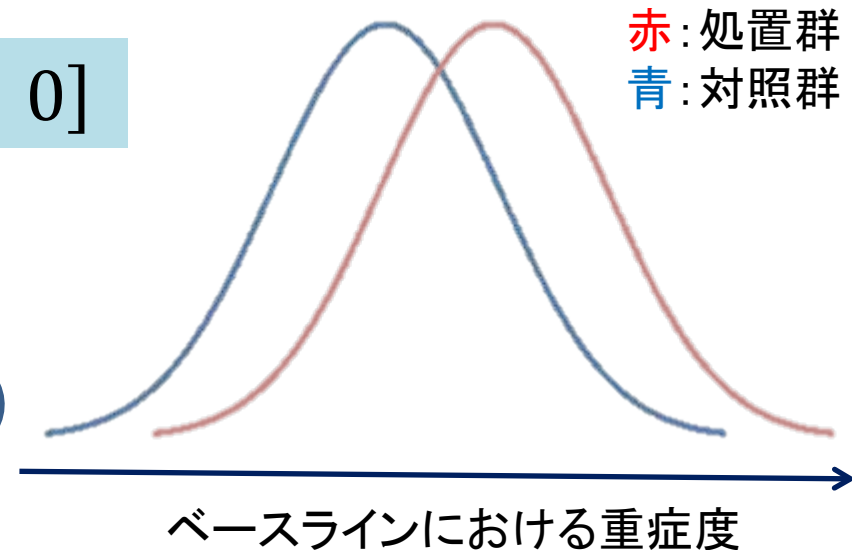
$Y(1)$ : 処置群に属していれば得られる応答

$T$ : 被験者が対照群に属せば0, 処置群に属せば1となる割り当て変数

潜在変数

$$E[Y(1)|T = 1] - E[Y(0)|T = 0]$$

処置の効果を表すのか？



重症度が同じ被験者同士で比較を行う

$$E[Y(1)|重症度, T = 1] - E[Y(0)|重症度, T = 0]$$



比較する上で考慮したい因子をZ(共変量)とおく

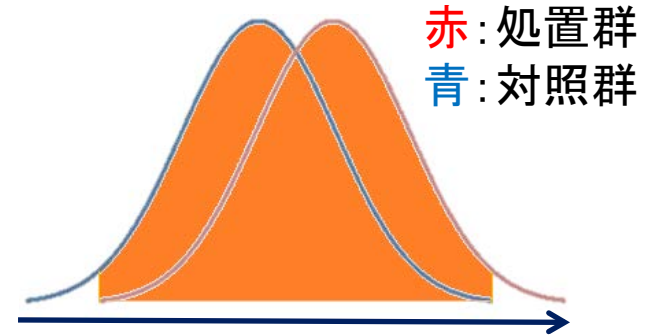
$$E[Y(1)|Z, T = 1] - E[Y(0)|Z, T = 0]$$



考えるべき共変量Zが高次元である場合は？



次元の呪いにより、共変量の調整が困難になる



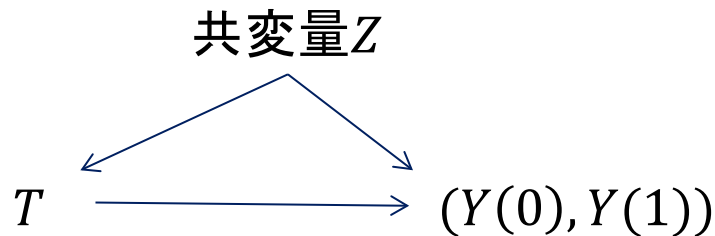
ベースラインにおける重症度

Ex. 年齢, BMI, バイタルサイン, 臨床検査値...etc.・・・が同じ

処置群		対照群
○		×
○	?	×
○		×
○		×
○		×



Rosembaum & Rubin (1983)



強く無視できる割り当て条件

$$(Y(0), Y(1)) \perp\!\!\!\perp T \mid Z$$

傾向スコア

$$e(Z) = P(T = 1 \mid Z)$$

傾向スコアの値が似た被験者同士を抽出すると、共変量の分布がある程度重なった被験者集団ができる

$$E[Y(1) \mid Z, T = 1] - E[Y(0) \mid Z, T = 0] = E[Y(1) \mid e(Z), T = 1] - E[Y(0) \mid e(Z), T = 0]$$

共変量の分布がそろうところ  
での観測値の期待値の差

傾向スコアの分布がそろうところ  
での観測値の期待値の差

PSMATCH ProcedureがSTAT14.2から実装可能に！！

共変量

共変量

観察データ

PSMATCH  
Procedure

擬似介入データ

コード

```
proc psmatch data=マッチングしたい観察データ;
```

```
class 群を表す変数;
```

```
① psmodel 群を表す変数(treated=処置群名)= 共変量;
```

```
② match method=greedy(order=ascending) stat=lps caliper=0.2;
```

```
③ assess lps var=(共変量) / plots=(boxplot barchart);
```

```
output out(obs=match)=マッチングデータ;
```

```
run;
```

①傾向スコアをロジスティック回帰モデルによって推定

②マッチングアルゴリズムを規定

③マッチング後の共変量のバランスを確認

```
match method=greedy(order=ascending) stat=lps caliper=0.2;
```

i

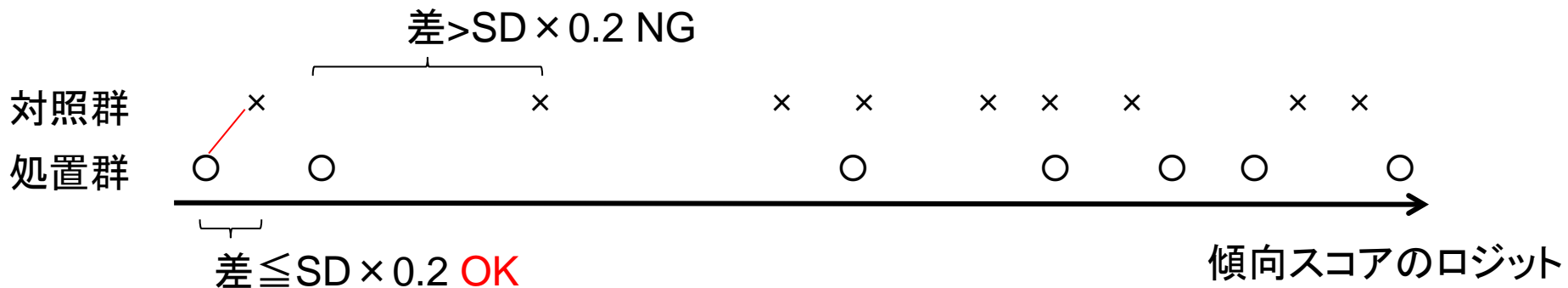
ii

iii

- i マッチングアルゴリズムを規定. 上記はascending greedy nearest neighbor matching
- ii マッチング因子を傾向スコアのロジットとする. 状況によってはこれは正規分布に従うことが期待される.
- iii マッチングする際の傾向スコアのロジットの差の絶対値の許容幅を規定. 許容幅の中で差の絶対値が小さいものをマッチングする. 許容幅は傾向スコアのロジットのSD×0.2が推奨される(Austin 2011).

例: 1対1のascending greedy nearest neighbor matching

処置群において、傾向スコアのロジットが小さい順からマッチングの相手を探していく。



### ③ マッチング後の共変量のバランスを確認

共変量のバランスを確認するために2つのアウトプットが出る

①ある1つの共変量 $z$ の群間における距離は**Standard Variable Difference**が用いられる(Austin 2007)

$$d = \frac{\bar{z}(1) - \bar{z}(0)}{\sqrt{(V(0) + V(1))/2}}$$

$\bar{z}(0)$  : 対照群における共変量 $z$ の標本平均

$\bar{z}(1)$  : 処置群における共変量 $z$ の標本平均

$V(0)$  : 対照群における共変量 $z$ の標本分散

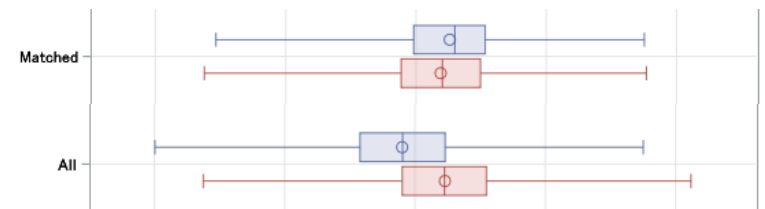
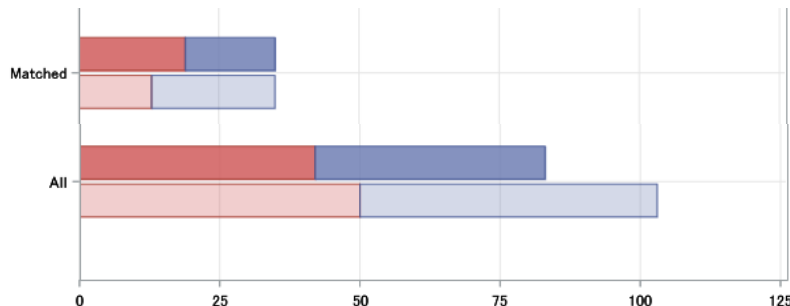
$V(1)$  : 処置群における共変量 $z$ の標本分散

PSMATCH Procedureではマッチング前, 後のStandard Variable Differenceが出力される

②グラフィックスによる共変量のバランスの確認

```
assess lps var=(共変量) / plots=(boxplot barchart);
```

赤 : 処置群      共変量が連続値であればboxplot  
青 : 対照群      離散値であればbar chartを出力



psmodel 群を表す変数(treated=処置群名)= 共変量;

共変量を指定するだけでロジスティック回帰モデルによる傾向スコアの推定を行う  
⇒非常に簡単に傾向スコアの推定が可能

しかし一方で...容易さの反面

傾向スコアのモデルをきちんと考える必要があるのではないか？

Lee et al.(2010)

- 共変量が高次元であり、次元の呪いを避けるために傾向スコアを用いている。ただし、傾向スコアの推定は高次元でうまくいくのか？
- 共変量はお互いに相関をもち、割り当て変数に対して交互作用を持つことはないのか？
- 共変量の割り当て変数に対するモデルは線形か？

傾向スコアをロジスティック回帰モデルで推定することは適切ではない場合があるかもしれない

Westreich et al.(2010)

傾向スコアの推定において、ロジスティック回帰モデルよりも機械学習を用いた方が正確な群間差の推測ができることが示唆されている。

Lee et al.(2010)

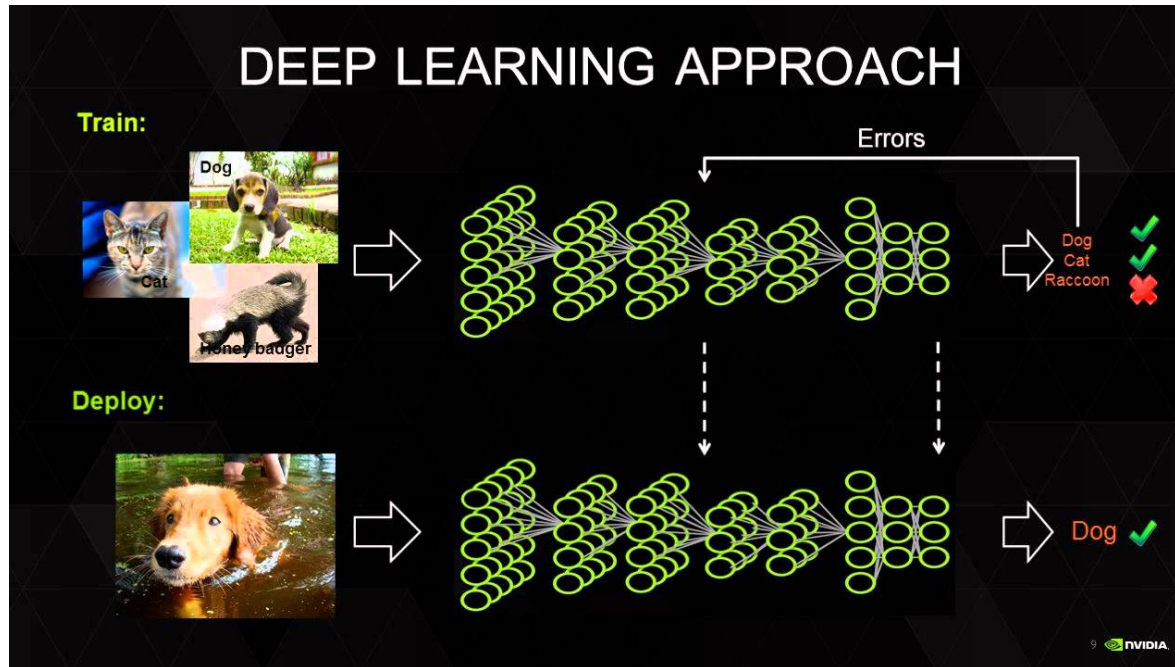
シミュレーションによって、傾向スコアの重み付けによる群間差の推定 (inverse-probability-weighting法) において、傾向スコアが共変量の**交互作用**や、**非線形性**をもつモデルであるときに、ロジスティック回帰モデルで傾向スコアを推定すると、群間差の推定精度が悪いことが示された。

扱う問題はズバリ

**共変量の高次元, 交互作用, 非線形性**



## 今流行のAIで用いられているディープラーニング



<https://www.youtube.com/watch?v=4A14mOsT6vQ>

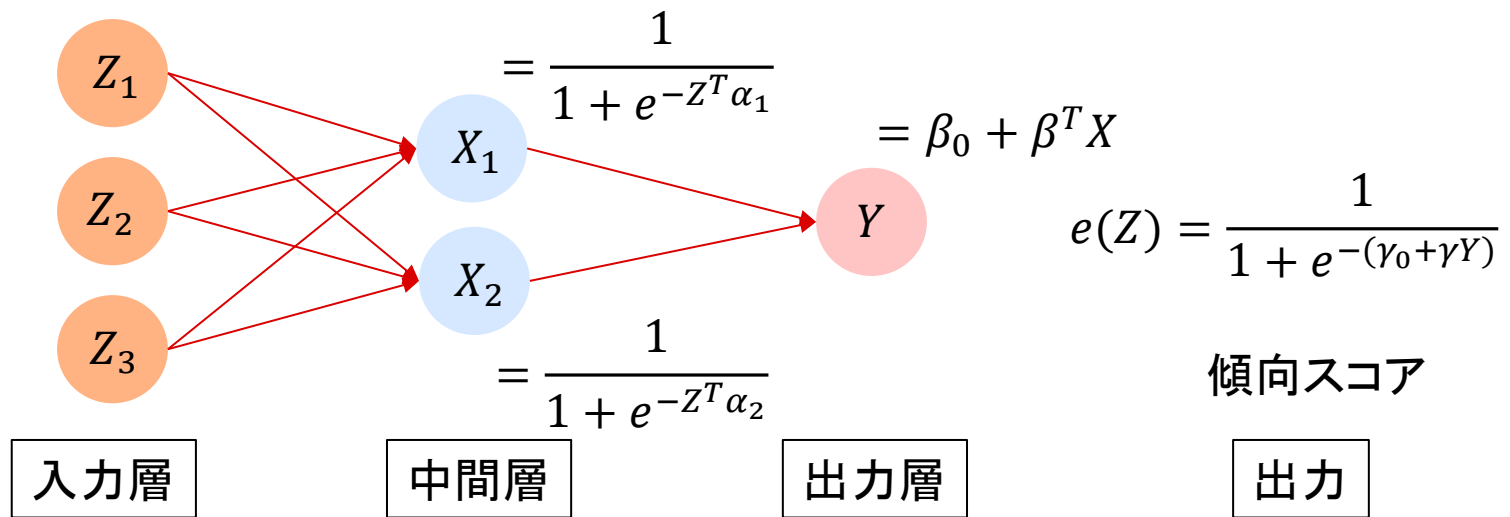
大規模で、複雑なデータに対して、柔軟なモデルを構築できる  
(高次元) (相互作用) (非線形性)

ディープラーニングの基となっているのは **ニューラルネットワーク**



例: 共変量3, 中間層1でノードの数2, 応答2値

共変量  $Z = (1, Z_1, Z_2, Z_3)^T$  中間層  $X = (1, X_1, X_2)^T$  活性化関数はロジスティック



イメージ

中間層, 出力層の各因子は, 共変量の線形和からできている  
 ⇒ 共変量間の相関, 交互作用を考慮  
 中間層は共変量の非線形な項になっており, 出力はさらにその線形和で構成される  
 ⇒ 非線形性を考慮



```
proc imstat data=データ;  
  neural 群を表す変数/ seed=初期値シード  
  input=(共変量)  
  nominal=(群を表す変数)  
  activation=(ノードの関数)  
  hidden=(中間層の因子数)  
run;
```

 傾向スコアの推定値

psdataステートメントで傾向スコアの推定値を利用

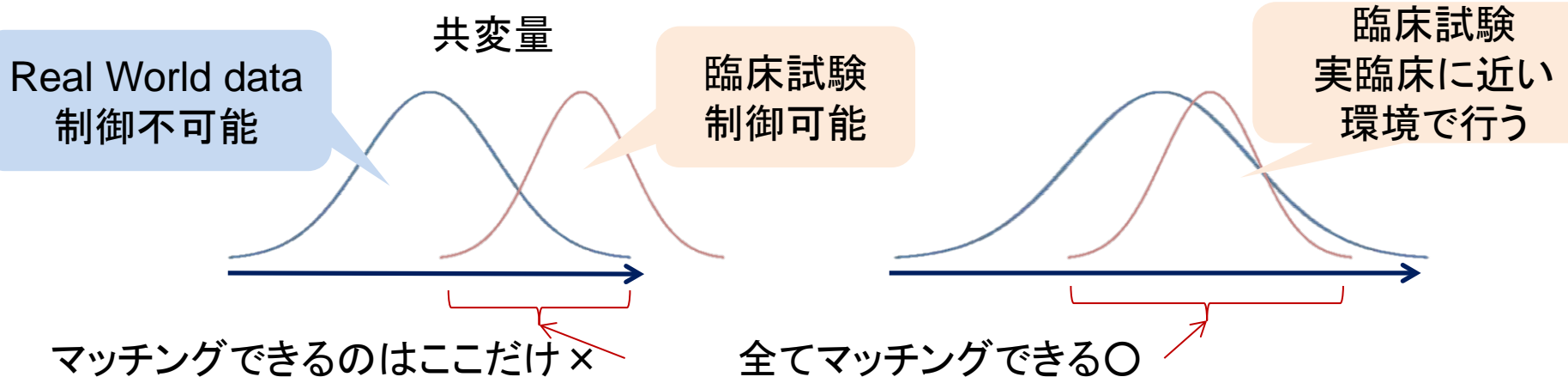
```
proc psmatch data=マッチングしたい観察データ(傾向スコアの推定値有);  
  class 群を表す変数;  
  psdata treatvar=群を表す変数(treated=処置群名) PS=推定した傾向スコアの変数;  
  match method=greedy(order=ascending) stat=lps caliper=0.2;  
  assess lps var=(共変量) / plots=(boxplot barchart);  
  output out(obs=match)=マッチングデータ;  
run;
```

PSMATCH Procedure × ニューラルネットワークにより, 適切な傾向スコアマッチングを行う

数値実験の結果は当日公開する

## Real World Dataの利用可能性を忖度する

- 1対多のマッチングの多をどれだけとる必要があるかは注意する必要がある
- 倫理的な観点から処置群の被験者は全てマッチングできることが好ましい



- 臨床試験もReal Worldに近い環境にして、**Effectiveness**を見ることに利用できるのでは？
- 有効性のみではなく、有害事象の発生割合の比較といった**安全性の解析**にも利用可能.
- 同一疾患、異なる医薬品の臨床試験をマッチングで比較することで精度よく効果の差を見積もることも可能.

- ICH Reflection on “GCP Renovation”: Modernization of ICH E8 and Subsequent Renovation of ICH E6 (2017)
- 第4回次世代医療機器開発推進協議会(H29.3.15)資料5  
(<http://www.kantei.go.jp/jp/singi/kenkouiryuu/kaihatsu/dai4/siryuu5.pdf>)
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Austin, P. C. (2011). Optimal caliper widths for propensity - score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, 10(2), 150-161.
- Lee, B. K., Lessler, J., & Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in medicine*, 29(3), 337-346.
- Westreich, D., Lessler, J., & Funk, M. J. (2010). Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of clinical epidemiology*, 63(8), 826-833.
- Austin, P. C. (2007). Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. *The Journal of Thoracic and Cardiovascular Surgery*, 134(5), 1128-1135.

傾向スコアマッチングをReal World Data活用のための1つの武器として適用を考えていきましょう



ご清聴ありがとうございました

End of Slide