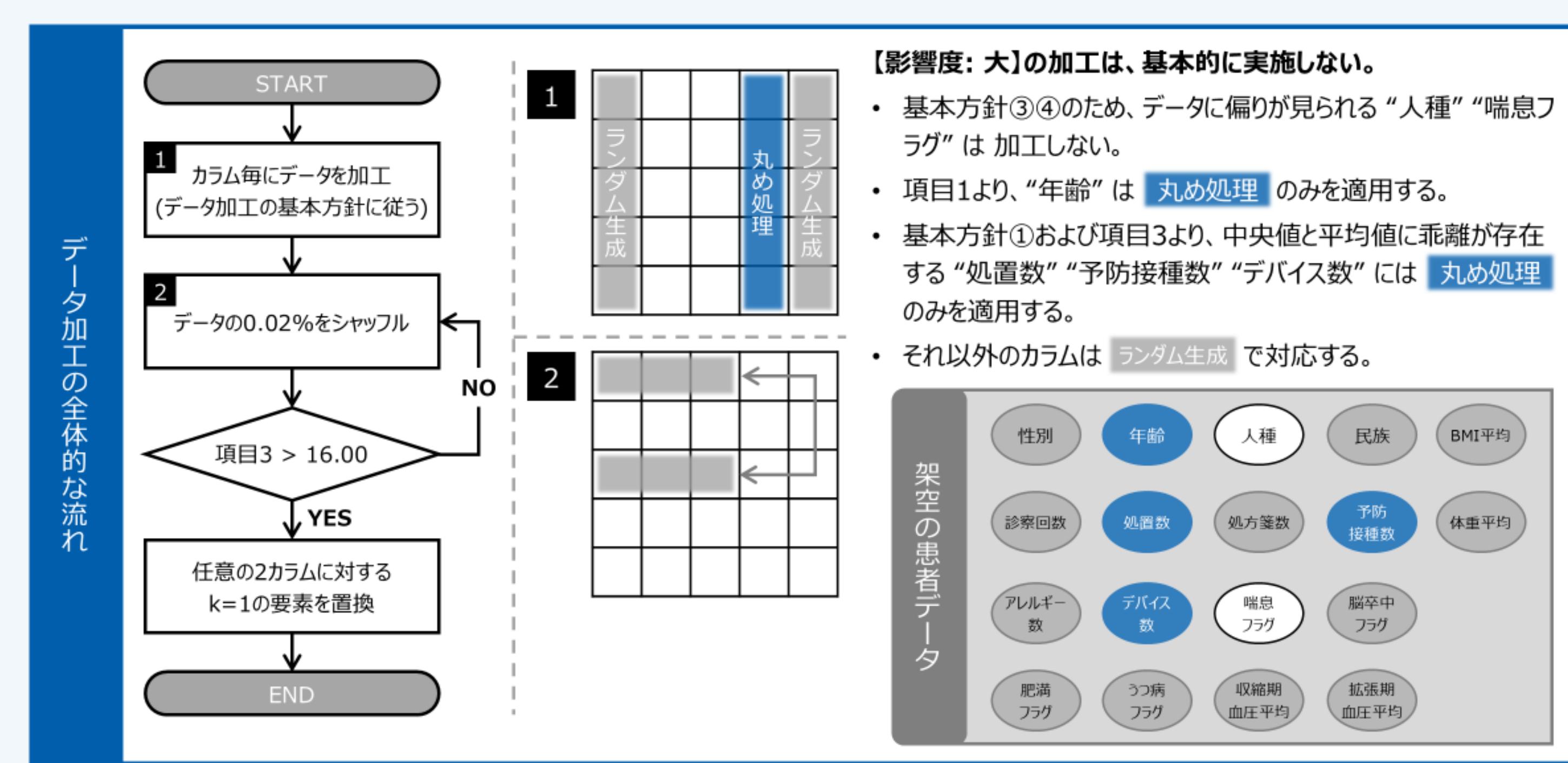
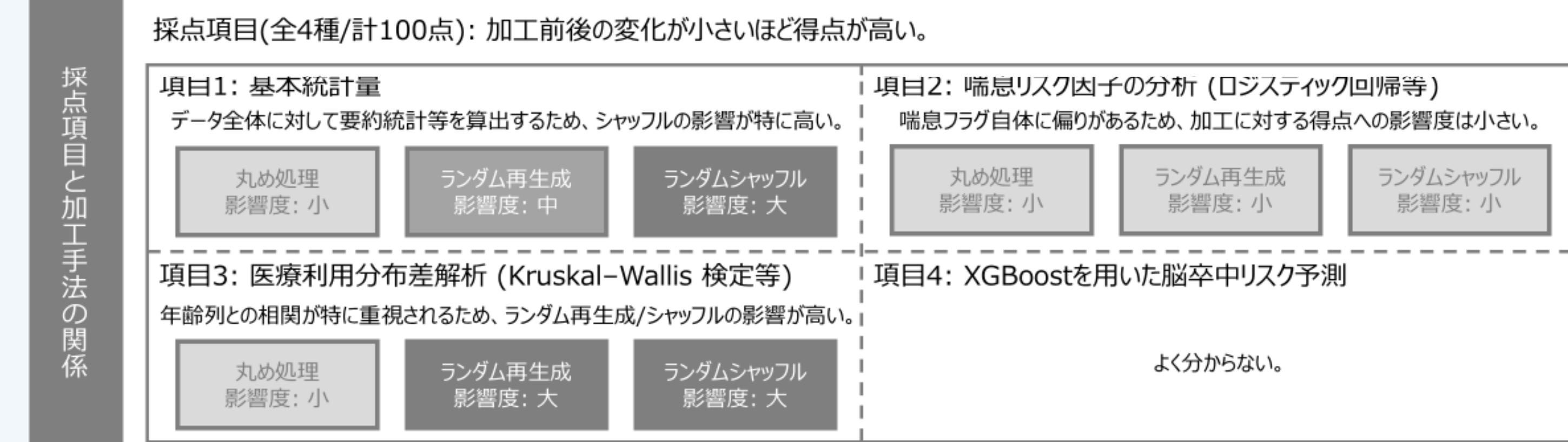
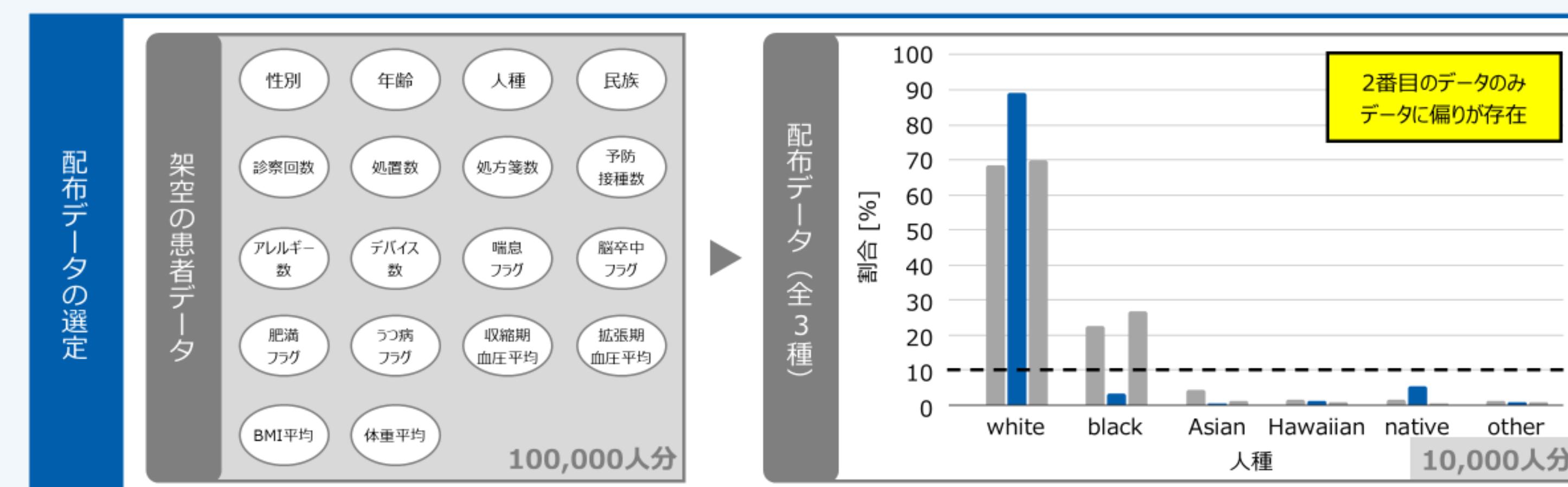


1. 加工フェーズ

【基本方針】データの有用性よりも匿名性

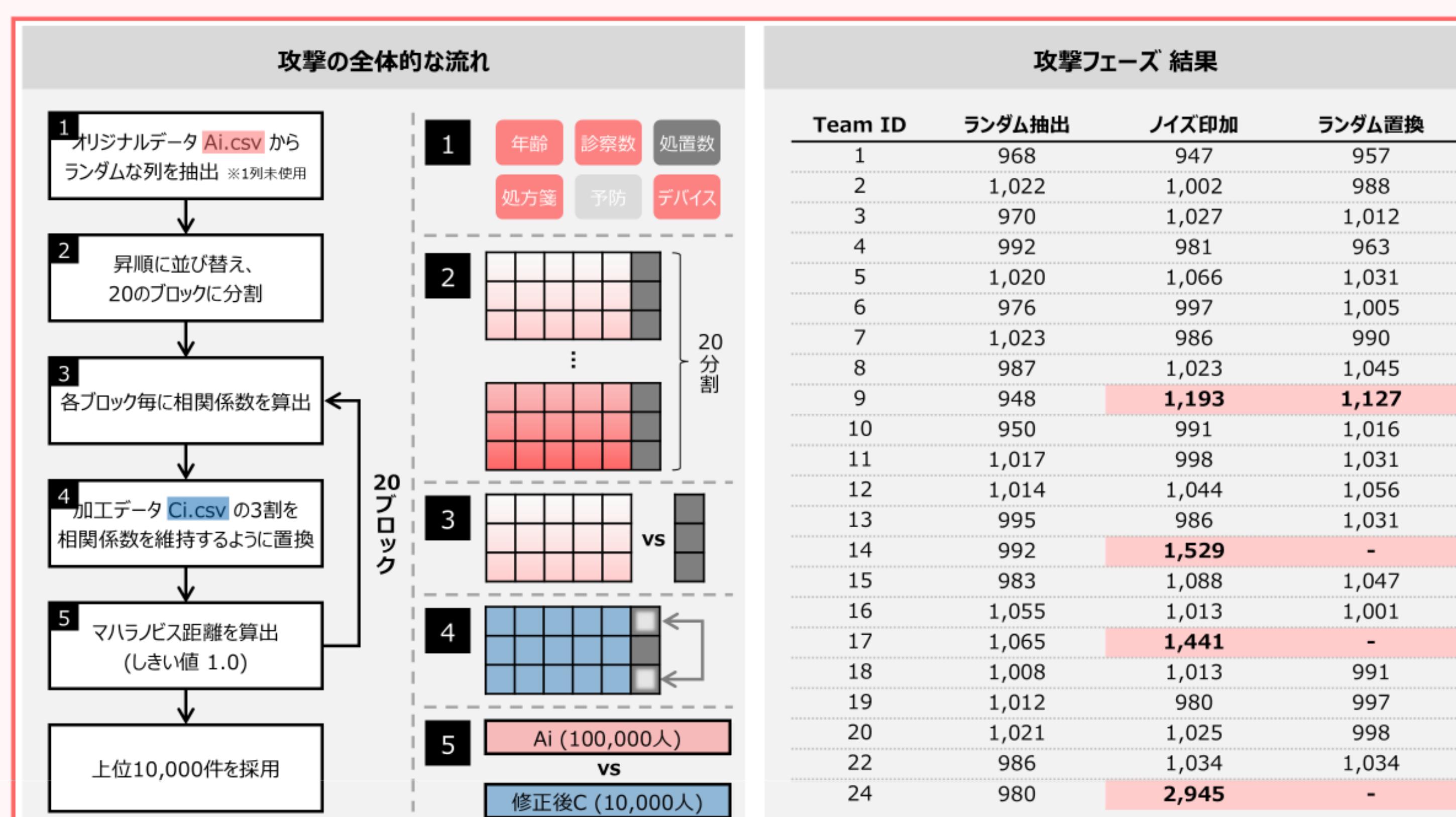
- ① 数値データの下位/上位5%は、匿名性が無いデータとみなし、置換対象とする。
- ② 最終的にデータはランダムシャッフルで調整する。
- ③ そのため、データの頻度に偏りが見られれば **匿名性が担保** されているとして加工しない。(配布データの選定方針)
- ④ 今回は、このしきい値を 90%以上 とし、カラム毎に確認する。
- ⑤ しきい値を下回るカラム (90%未満) に関しては、ランダム再生成 or 丸め処理($k \geq 2$)-匿名化 で対応する。



2. 攻撃フェーズ

【基本方針】全チーム自分と似たような加工をしていると仮定

- ① 自チームの加工方針を鑑み、新たに生成した「身長平均」を加えた 7カラムで攻撃を行う。
- ② 差分プライバシノイズを除去する手段として、ランダム置換を採用した。
- ③ 数値データを主として扱う攻撃のため、マハラノビス距離を主軸にメンバーシップ推定を実施する。



3. まとめ

【加工フェーズ】

- 有用性を犠牲^{※1}にしても関わらず、匿名性を担保する^{※2}ことが出来なかった。
- ※1: 本戦(有用性)の得点は66点であり、約34の減点。50点が元のデータとほぼ相関が無い状態。(所感)
- ※2: 匿名性を担保 = 本戦(匿名性)のスコアが約90点。完全ランダムでデータを推定した際の結果と同値。
- 単一カラムに対する丸め処理(k -匿名化)では、匿名性が担保されず、有用性だけが犠牲になった。
- 複数カラムに対する丸め処理(k -匿名化)が個票データの匿名性に繋がる。
- 我々のチームへのデータ暴露に特に貢献した **Team06** の攻撃手法を学び、改善に繋げる。

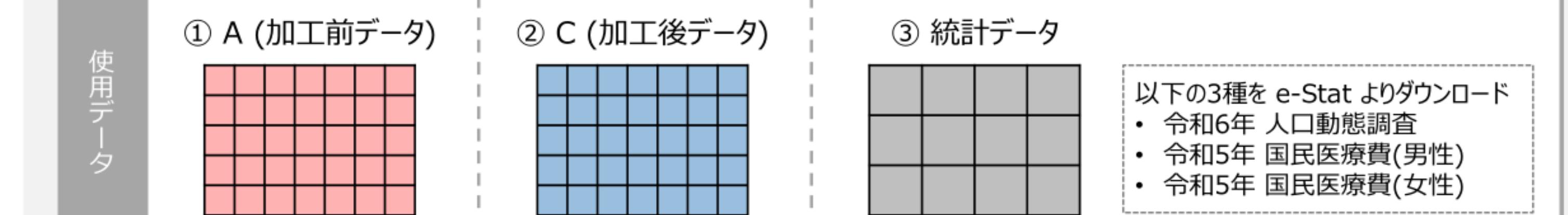
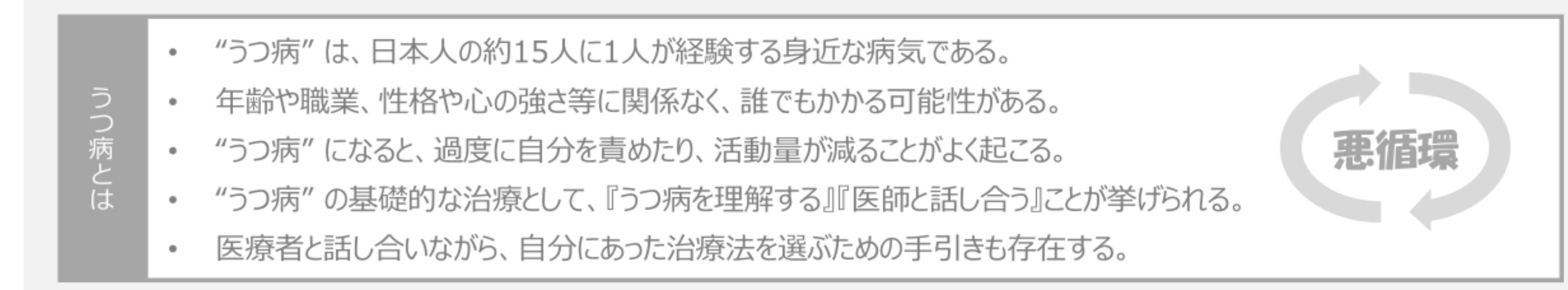
【攻撃フェーズ】

- ながら攻撃フェーズの着想は良かったと自負している。
 - 「平均身長」カラムの作成
 - 並び替えとブロック化による相関行列とデータ置換の併用
- なお、感覚から脱することが出来ず、理論化まで至れなかった点は反省するべき。
- 我々のチームへのデータ暴露に特に貢献した **Team06** の攻撃手法を学び、改善に繋げる。

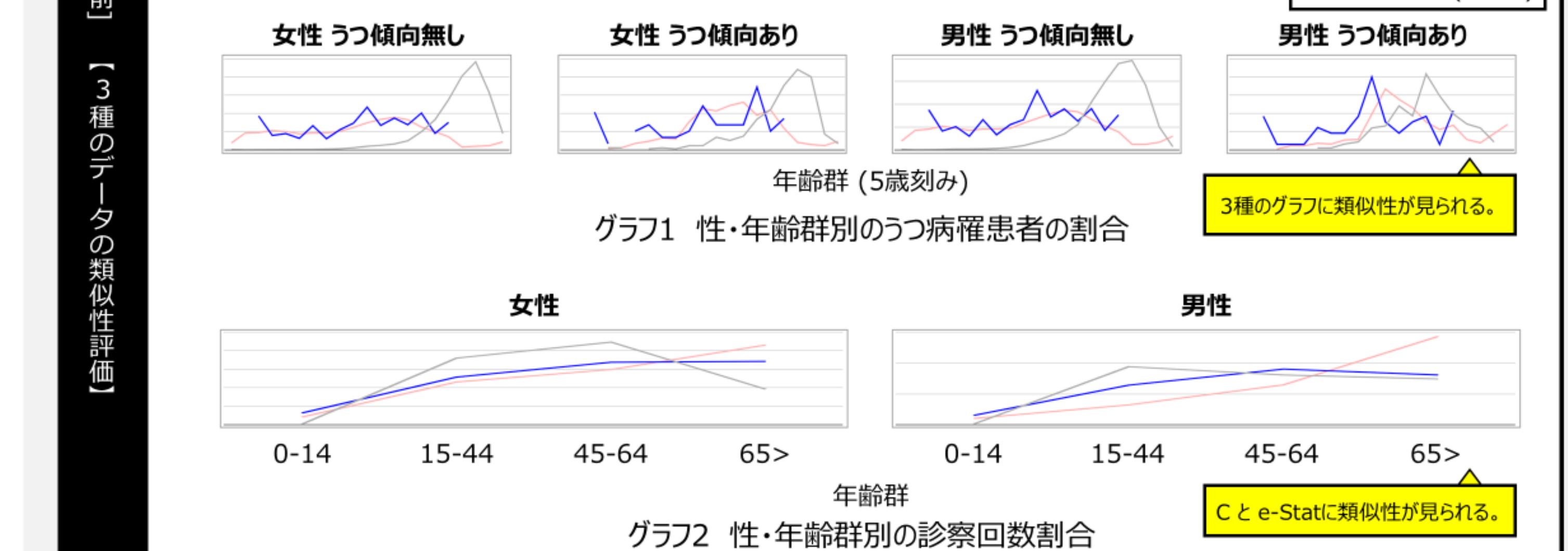
順位	本戦 (有用性)	本戦 (匿名性)	本戦 (攻撃)
19	66.36	82.29	10248

4. 加工後データを使用したデータ分析

- 概要
- 今回のデータ項目を皮切りに、改めて「うつ病」に対する調査を実施
 - 調査から「うつ病」を正しく理解できていなかつことが発覚 << 自己流の療法で改善する病と誤認
 - 今回の加工後データを用いることで、「うつ傾向」の早期発見と対策が出来ないか検討
 - データ分析より、**高BMI**かつ**低血圧(120/80以下)**で「うつ傾向」が高まることが発覚
 - 適正BMIや正常血圧に近づけるために、「岡山瀬戸内ばなな」や「MOMOTARO JEANS」が有効であると推察
 - 現在、上記検討の正確性を評価中...



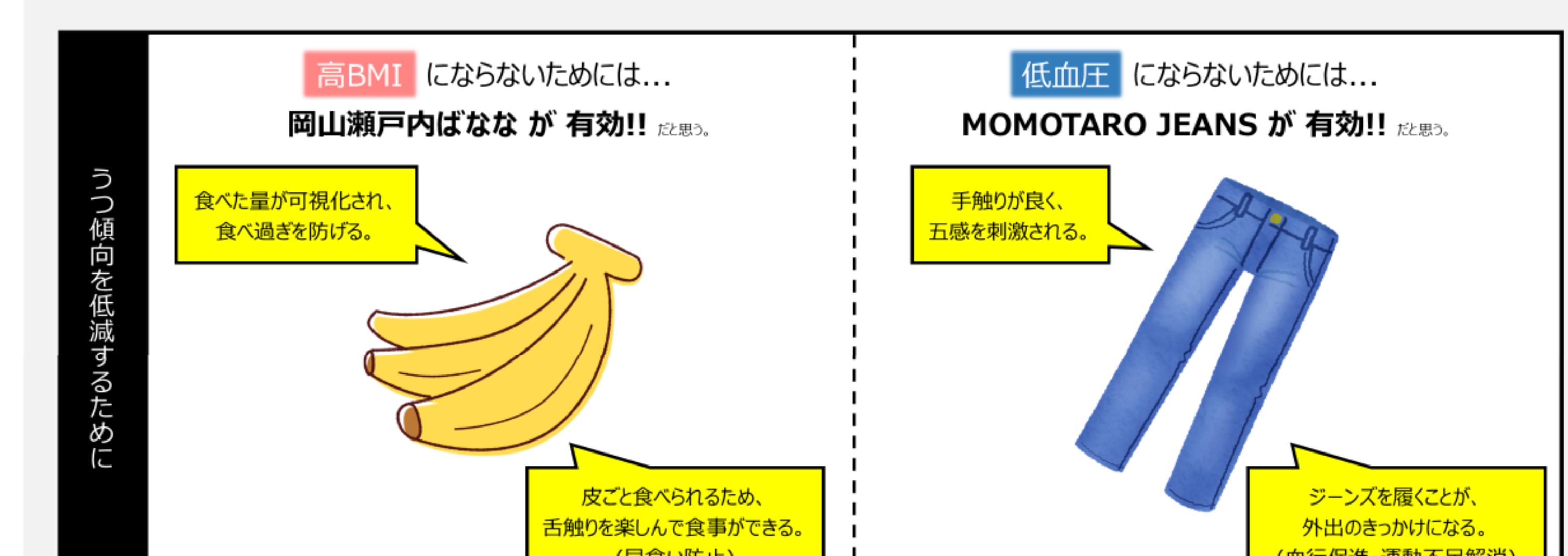
- データ分析「前」
- 個票データの大部分は日本国外のデータであるため、第一に **日本国内と国外の「うつ傾向」の違いを評価** する。
 - グラフ1は、「性・年齢群別のうつ病罹患者の割合」である。
 - グラフ2は、「性・年齢群別の診察回数割合」である。
 - グラフ1より、男性のうつ傾向の類似性が、グラフ2より、男性の診察回数の類似性が確認できる。
従って、**日本国内と国外で男性のうつ傾向に違いは無い** と言える。
 - 以後、加工後データCと日本の個票データを同一のデータと仮定し、データ分析を実施する。



- データ分析「後」
- 下記表より、血圧やBMIの値が適正值から外れるほど、うつの傾向の高まりが確認できる。
 - 低血圧(120/80以下)**は睡眠不足やストレスなどによる自律神経の乱れ、**高BMI**は心理的ストレスによる影響だと推察する。

表 カラム別のうつ傾向割合

データ分析「後」	拡張時血圧別の うつ傾向 割合		収縮時血圧別の うつ傾向 割合		平均BMI別の うつ傾向 割合	
	うつ傾向 なし	うつ傾向 あり	うつ傾向 なし	うつ傾向 あり	うつ傾向 なし	うつ傾向 あり
「うつ傾向分析」	< 80	53.49%	57.72%	< 120	53.82%	50.41%
	80 - 85	19.23%	12.20%	120 - 125	12.96%	11.38%
	85 - 90	13.89%	19.51%	125 - 130	10.92%	8.13%
	90 - 95	7.30%	4.88%	130 - 135	8.87%	13.01%
	95 - 100	3.60%	2.44%	135 - 140	5.88%	8.94%
	100 - 105	1.70%	2.44%	140 - 145	3.64%	1.63%
	105 - 110	0.60%	0.81%	145 - 150	2.08%	4.07%
	> 110	0.19%	0.81%	150 - 155	0.97%	1.63%
	総計	100.00%	100.00%	155 - 160	0.58%	0.81%
				160 - 165	0.19%	
総計	100.00%	100.00%	165 - 170	0.08%		



- 今後の展望
- 今回、加工後データを引数としたデータ分析であるため、仮説まみれの結論を導いた。
 - 仮説1つ1つの何が正しく、何が誤っているのかを、自身の身体に対して検証していく。
 - 特に、岡山瀬戸内ばなな や MOMOTARO JEANS は、せっかくの機会なので購入しようと思う。

- 参考文献
- e-Stat: 人口動態調査 / 人口動態統計 確定期 死亡
https://www.e-stat.go.jp/stat-search/files?page=1&query=%E3%81%86%E3%81%A4&layout=dataset&stat_infid=000040316542&metadata=1&data=1
 - e-Stat: 国民医療費 / 令和5年度国民医療費 統計表
https://www.e-stat.go.jp/stat-search/files?page=1&query=%E3%81%86%E3%81%A4&layout=dataset&stat_infid=000040358481&metadata=1&data=1
 - 厚生労働省: うつ病の認知療法・認知行動療法 (患者さんのための資料)
<https://www.mhlw.go.jp/bunya/shougaiken/kokoro/dl/04.pdf>
 - 日本うつ病学会
<https://www.secretariat.ne.jp/jsmid/ippan/shiryo.html>
 - MOMOTARO JEANS
<https://momojeans.com/>
 - 瀬戸内ばなな
<https://setoubanana.com/>