第 2 期 医薬安全性研究会
第17回定例会 2015 年 11 月 14 日

アソシエーション分析の
医薬品副作用報告データベースへの適用

藤原 正和　北西 由武　都地 昭夫　渡辺 秀章

塩野義製薬株式会社 解析センター

要約:アソシエーション分析は，項目間の関連性を評価するために，主にマーケット分野を中心として古くから用いられており，適用を工夫することで，様々な場面で有用であると考えられる．SAS® (SAS Institute Inc., Cary, NC, USA) では，IMSTAT (In-Memory Statistics for Hadoop) という新機能が追加され，その中の ARM (Association Rule Mining) ステートメントを用いることで，アソシエーション分析が実行可能となった．本報告では，まずアソシエーション分析の理論的背景の説明を行った．次に医薬品副作用報告データベースを用いた副作用間の関連性評価の事例として，抗うつ薬の副作用を対象として副作用間の関連を評価した．副作用の関連ルールに対して得られた Support の値は低かったが，Confidence は 0.8 を超えており，Lift も 1 を超えていたことから，挙がってきている関連ルールの関連性は強いと考えられた．さらに抗うつ薬と他の薬剤との併用の有無による副作用の関連の違いを検討し，併用の有無により副作用の関連ルールが異なる結果を得た．さらにアソシエーション分析の考えを基にした，レコメンデーションシステムを適用し，患者の背景情報から今後発現すると考えられる副作用が予測できることを示した．R を用いても同様の解析を行うことができることを紹介した．アソシエーション分析を用いることで，大規模な副作用情報の中から，関連ルールを抽出することが可能となり，アソシエーション分析の JADER への適用は有益であると考えられた．ただし，データベースの性質を把握しておく必要があり，解釈には注意が必要である．

キーワード: アソシエーション分析，医薬品副作用データベース，JADER，SAS，R

目 次

1. ビッグデータと製薬企業 ---------- 4
2. アソシエーション分析---------- 9
3. JADER に対する適用結果----------20

第 2 期 医薬安全性研究会
第17回定例会 2015 年 11 月 14 日

アソシエーション分析の
医薬品副作用報告データベースへの適用

第2期 医薬安全性研究会　総会及び第17回定例会

2015年11月14日

塩野義製薬株式会社

藤原 正和，北西 由武，都地 昭夫，渡辺 秀章

要旨：

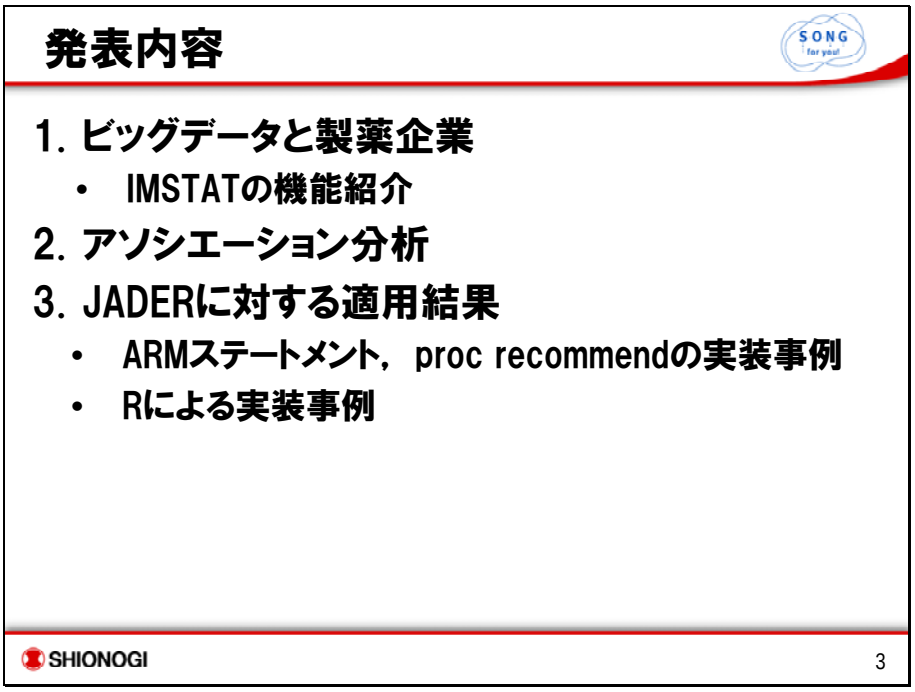
アソシエーション分析は、項目間の関連性を評価するために、主にマーケット分野を中心として古くから用いられており、適用を工夫することで、様々な場面で有用であると考えられる。

SAS**でも、**IMSTAT**という新機能が追加され、その中の**ARM(Association Rule Mining)**ステートメントを用いることで、アソシエーション分析が実行可能となった。**
本講演ではアソシエーション分析の理論的背景の説明に加えて、製薬業界における適用例として、医薬品副作用報告データベース(JADER)**を用いて副作用間の関連性評価を行った事例紹介，及びアソシエーション分析の考えを基にした、レコメンデーションシステムの紹介を行う。**

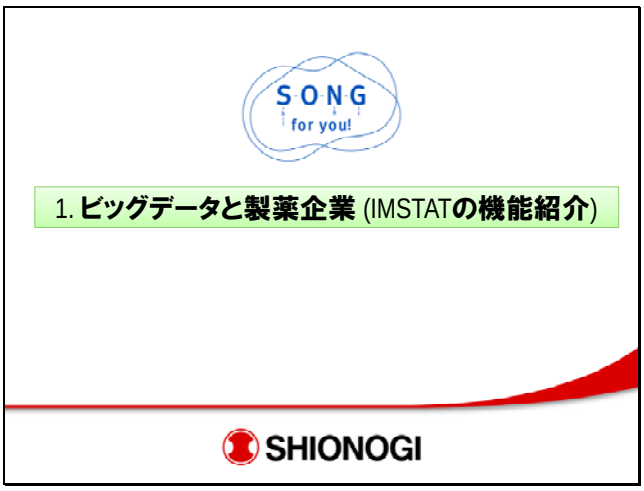
加えて、SAS **だけでなく、**R**を用いたアプローチについても取り上げる。**

SHIONOGI 2

本講演の要旨である．アソシエーション分析は，項目間の関連性を評価するために，主にマーケット分野を中心として古くから用いられており，適用を工夫することで，様々な場面で有用であると考えられる．SAS では IMSTAT という新機能が追加され，その中の ARM (Association Rule Mining) ステートメントを用いることで，アソシエーション分析が実行可能となった．本講演ではアソシエーション分析の理論的背景の説明に加えて，製薬業界における適用例として，医薬品副作用報告データベース (Japanese Adverse Drug Event Report Database: JADER) を用いて副作用間の関連性評価を行った事例紹介，及びアソシエーション分析の考えを基にした，レコメンデーションシステムの紹介を行う．加えて，SAS と R を用いたアプローチを紹介する．

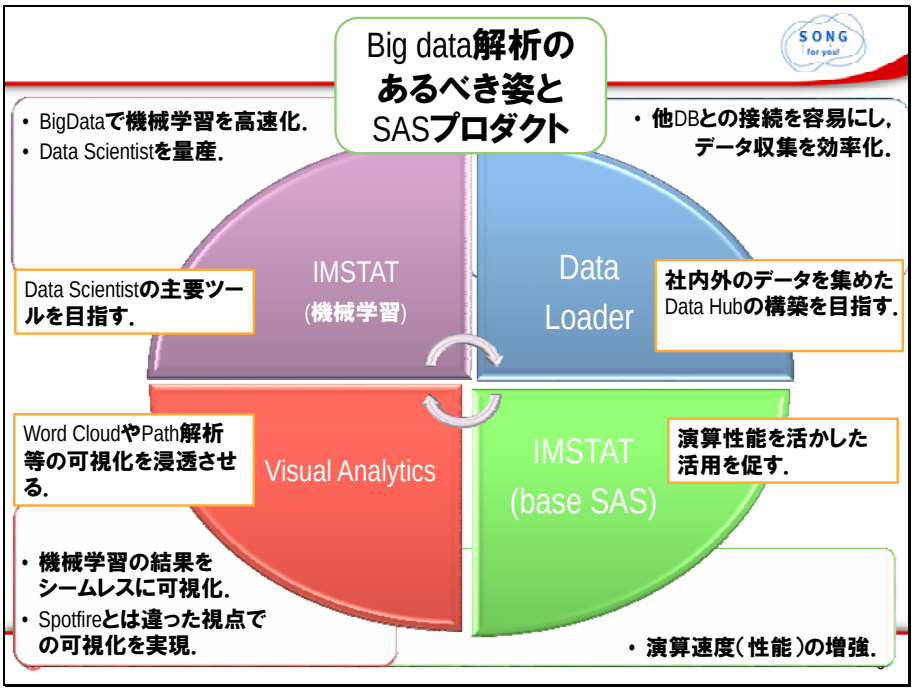


発表の構成としては，まず製薬業界とビッグデータについて取り上げる．次にアソシエーション分析の理論について説明して，最後に医薬品医療機器総合機構 (PMDA) が公開している JADER などの副作用自発報告データベースに適用した結果を紹介する．

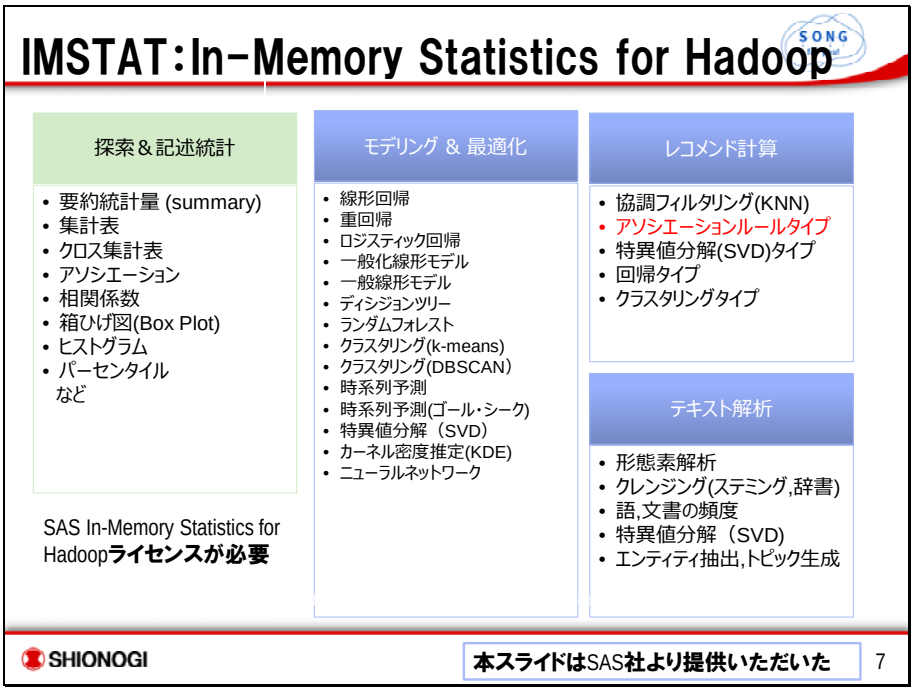


まず, ビッグデータと製薬企業 (IMSTAT の機能紹介) についてである.

製薬企業は様々な観点でビッグデータを利用している. Signal Detection, Text Mining などがその 1 例である.

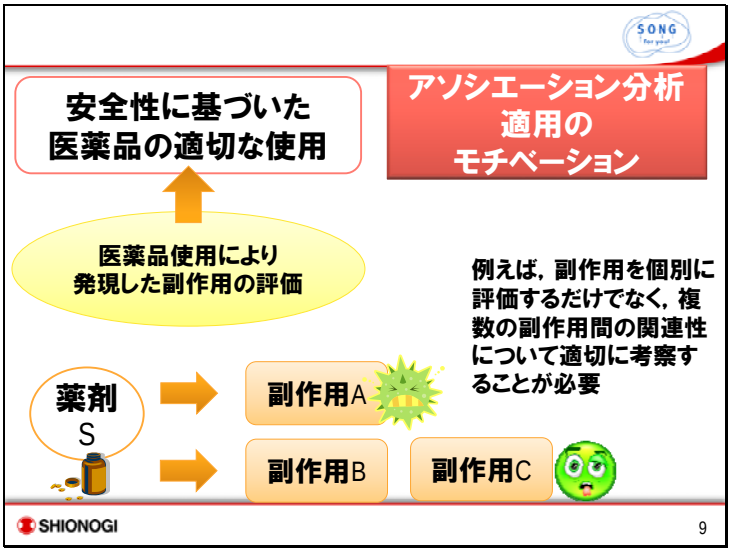


SAS を用いた Big data 解析のあるべき姿としては, Data Loader® (SAS Institute Inc., Cary, NC, USA)でデータの蓄積や結合, そして Base SAS® (SAS Institute Inc., Cary, NC, USA)や IMSTAT を用いた演算, ならびに機械学習を行い, 新しい知見を導くことが大事である. そして, 視覚的に結果を眺めることも重要であるため, その観点で Visual Analytics® (SAS Institute Inc., Cary, NC, USA)の利用も有益であると考えられる.

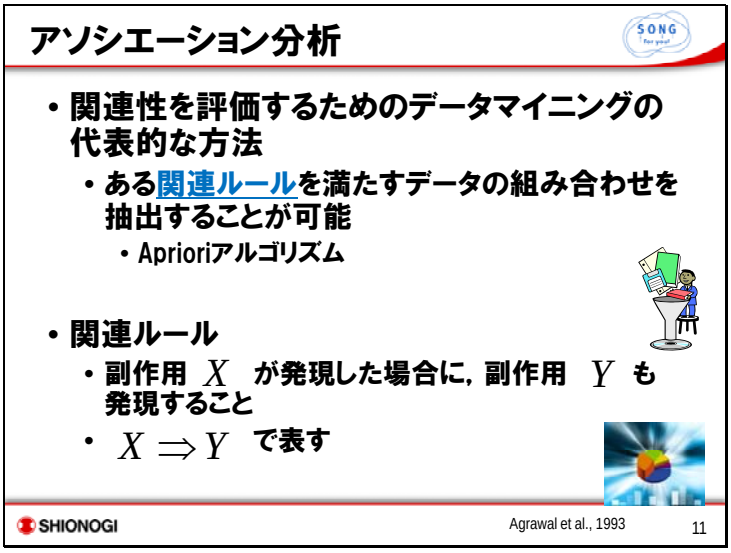


本スライドは SAS 社から提供いただいた．IMSTAT ではこれらの機能が備わっている．本講演では赤字で示しているアソシエーションルールタイプの関数を用いた．

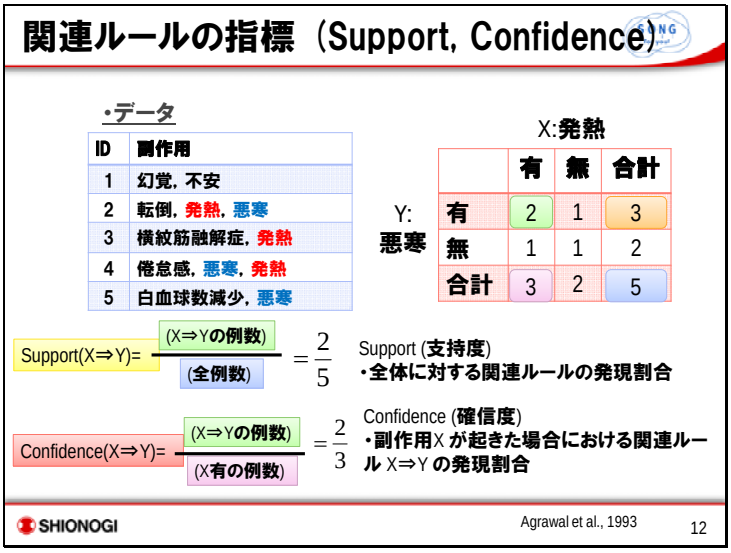
アソシエーション分析の理論について述べる.



製薬メーカーとしては，安全性に基づいた医薬品の適切な使用を促していく必要がある．安全性に基づいた医薬品の適切な使用のための 1 つの手段として，医薬品使用により発現した副作用の評価が挙げられる．副作用の評価として，1 つの副作用に注目して評価する方法もあるが，例えば，副作用を個別に評価するだけでなく，複数の副作用間の関連性について適切に考察することも有益であると考えられる．



複数の副作用間の関連性について適切に考察するために，アソシエーション分析を用いる．アソシエーション分析とは，関連性を評価するためのデータマイニングの代表的な方法である．Apriori アルゴリズム (Agrawal et al., 1993) を用いることで，ある関連ルールを満たすデータの組み合わせを抽出することが可能である．本講演では，副作用 X が発現した場合に，副作用 Y も発現することを X⇒Y で記載する．



関連ルールを評価するために，指標が必要である．本講演では 4 つの指標を取り上げる．まず Support についてである．Support は支持度とも呼ばれ，全体に対する関連ルールの発現割合を意味する．本スライドの例では，2/5 となる．また，Confidence (確信度) という指標は副作用 X が起きた場合における関連ルール X⇒Y の発現割合を意味する．この例では 2/3 となる．

Support (X⇒Y), Confidence (X⇒Y) の考察

- 値が高いほど関連性が強いと判断される
- 関連性の評価にはConfidenceだけでは不十分
  - X有が1例で，それがX有，Y有であったときにConfidenceは1になってしまうため，Supportと共に考えていく必要がある
- ただし，Support，Confidenceの値が高い場合でも，意味のないルールが抽出されてしまうことがある

例

X:発熱

Y:悪寒

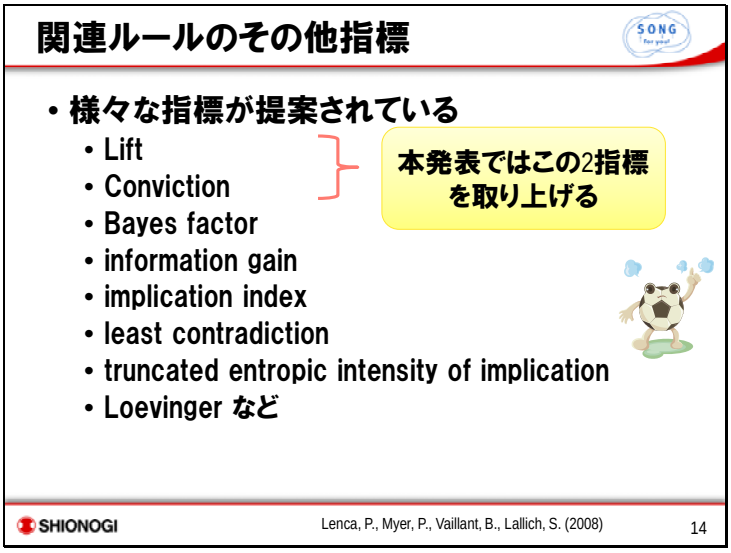
	有	無	合計
有	a	c	Y
無	b	d	N-Y
合計	X	N-X	N

Y/N の値が高い場合に，a/X，a/Nの値も大きくなることがある

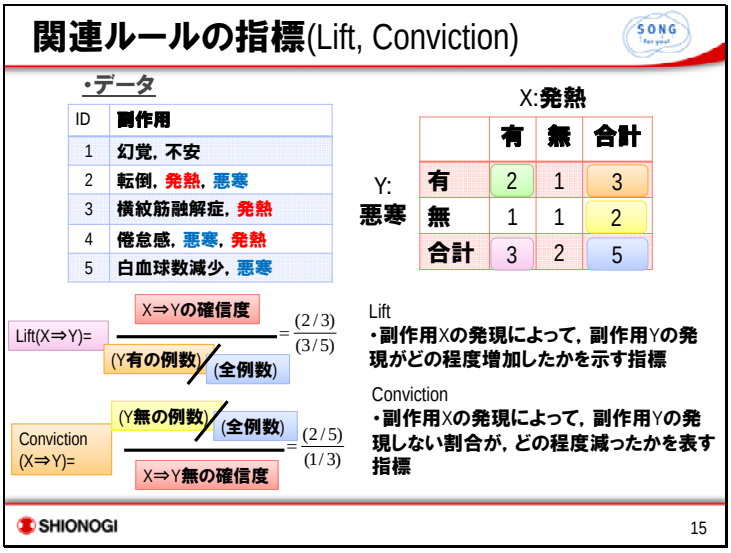
つまり，発熱に関係なく，悪寒を発現した例数が多い場合にも，Support(X⇒Y)，Confidence(X⇒Y)は高い値を取ってしまう

SHIONOGI 13

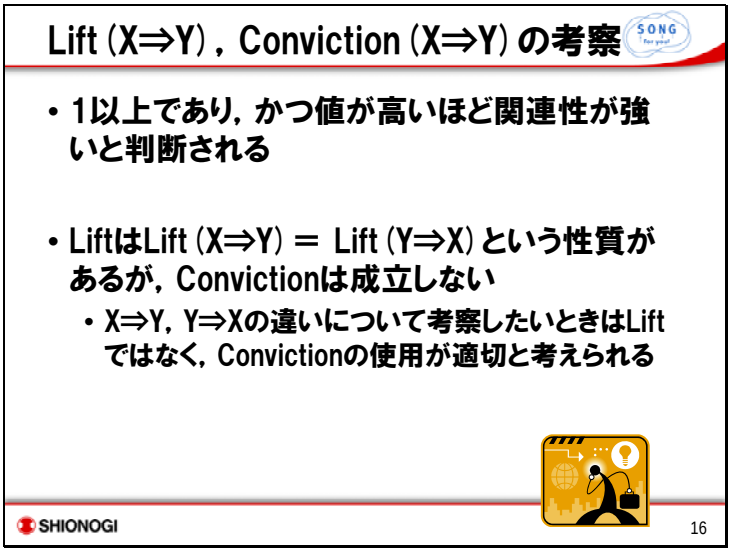
Support(X⇒Y)，Confidence(X⇒Y)の考察である．これらの指標は，値が高いほど関連性が強いと判断されるが，関連性の評価には Confidence だけでは不十分である．X 有が 1 例で，それが X 有，Y 有であったときに Confidence は 1 になってしまうため，Support と共に考えていく必要がある．ただし，Support，Confidence の値が共に高い場合でも，意味のないルールが抽出されてしまうことがあるため，注意が必要である．それは Y/N の値が高い場合に，a/X，a/N の値も大きくなることがあるためである．



Support, Confidence だけでは関連性を評価する上で不十分であるため，関連ルールの指標として他にも数多く提案されている．本講演では Lift, Conviction (Lenca, P., et al., 2008) を紹介する．



Lift は副作用 X の発現によって，副作用 Y の発現がどの程度増加したかを示す指標であり，この例では(2/3)/(3/5)=10/9 である．Conviction は副作用 X の発現によって，副作用 Y の発現しない割合が，どの程度減ったかを表す指標であり，この例では(2/5)/(1/3)=6/5 である．



Lift(X⇒Y), Conviction(X⇒Y) の考察である．これらは値が 1 以上であり，かつ値が高いほど関連性が強いと判断される．これらの指標の使い分けであるが，Lift は Lift(X⇒Y)＝Lift(Y⇒X)という性質があるが，Conviction は成立しない．X⇒Y，Y⇒X の違いについて考察したいときは Lift ではなく，Conviction の使用が適切と考えられる．

複数の副作用間の関連ルールについて

・データ

ID	副作用
1	幻覚，不安
2	転倒，発熱，悪寒
3	横紋筋融解症，発熱
4	倦怠感，悪寒，発熱
5	白血球数減少，悪寒，転倒

X:発熱，悪寒

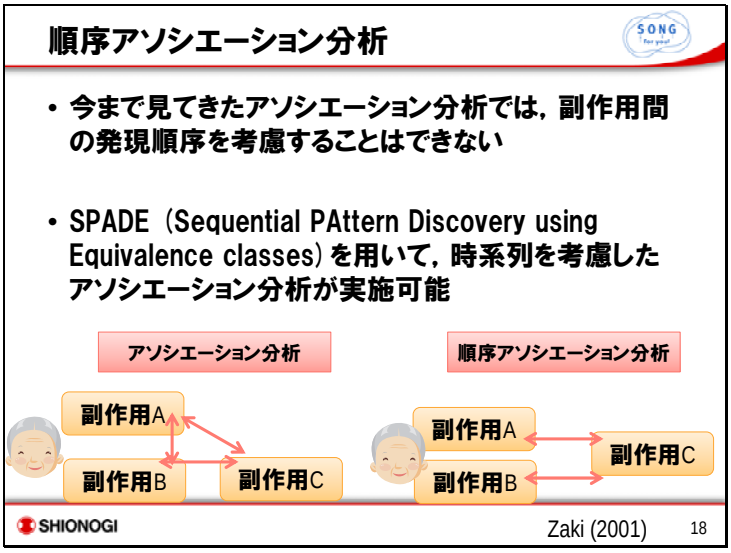
Y: 転倒	有	無	合計
有	1	1	2
無	1	2	3
合計	2	3	5

・Support(X⇒Y), Confidence(X⇒Y), Lift(X⇒Y), Conviction(X⇒Y)
は同様に計算可能

・X，Yを上記のように定義することで，複数の事象間の関連ルールについても考察が可能となる

SHIONOGI 17

複数の副作用間の関連ルールについてである．X:発熱，悪寒，Y:転倒とすれば，これまで見てきたように，Support(X⇒Y)，Confidence(X⇒Y)，Lift(X⇒Y)，Conviction(X⇒Y) は同様に計算可能である．つまり，X，Y を上記のように定義することで，複数の事象間の関連ルールについても評価が可能となる．



今まで見てきたアソシエーション分析では，副作用間の発現順序を考慮することはできない．そこで，SPADE (Sequential PAttern Discovery using Equivalence classes)を用いて，時系列を考慮した順序アソシエーション分析が実施可能である (Zaki, 2001).

順序アソシエーション分析の指標

・データ

ID	副作用 (報告時点1)	副作用 (報告時点2)
1	幻覚	不安
2	転倒，発熱	悪寒
3	横紋筋融解症，発熱	----
4	倦怠感，発熱	悪寒
5	白血球数減少	悪寒，転倒

Y: 発熱後に悪寒

X:発熱

	有	無	合計
有	2	0	2
無	1	2	3
合計	3	2	5

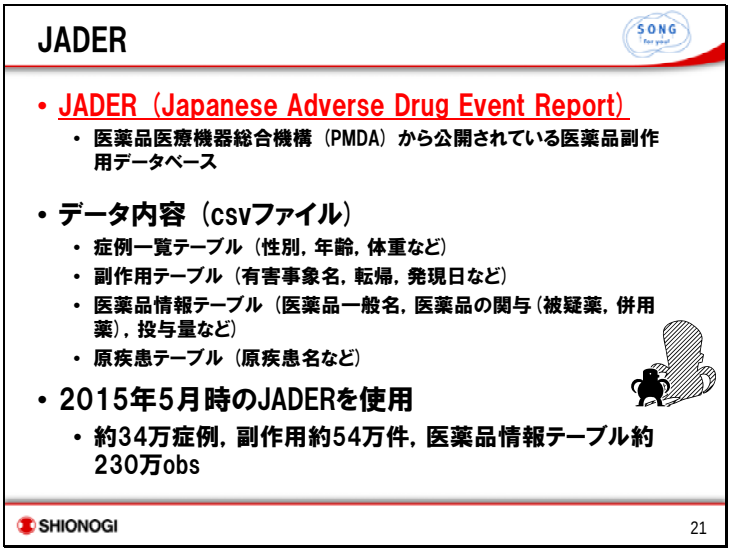
・Support(X⇒Y), Confidence(X⇒Y), Lift(X⇒Y), Conviction(X⇒Y) **は計算可能**

・X，Y**を上記のように定義することで，発現順序を考慮した関連ルールについても考察が可能となる**

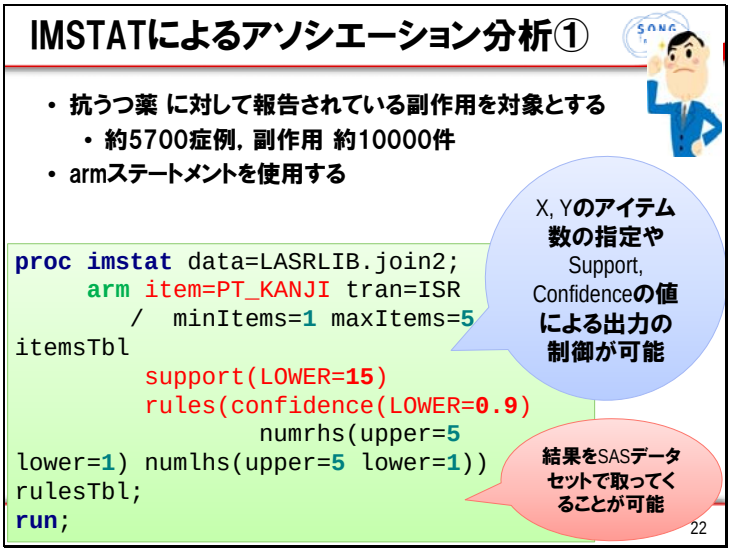
SHIONOGI 19

順序アソシエーション分析の指標については，これまで見てきたものと同様の考えで，Support(X⇒Y), Confidence(X⇒Y), Lift(X⇒Y), Conviction(X⇒Y) の計算が可能である.

JADER への適用結果を述べる.



JADER は医薬品医療機器総合機構 (PMDA) から公開されている医薬品副作用データベースである. 中は csv 形式であり, 症例一覧テーブル (性別, 年齢, 体重など), 副作用テーブル (有害事象名, 転帰, 発現日など), 医薬品情報テーブル (医薬品一般名, 医薬品の関与(被疑薬, 併用薬), 投与量など), 原疾患テーブル (原疾患名など) の計 4 つが利用可能である. 本講演では, 2015 年 5 月更新時の JADER を使用した.



IMSTAT によるアソシエーション分析のプログラム例である．抗うつ薬に対して報告されている副作用を対象として解析を行った．約 5,700 症例に対して，副作用は約 10,000 件報告されていた．IMSTAT によるアソシエーション分析では，X, Y のアイテム数の指定や Support, Confidence の値による出力の制御が可能であり，結果を SAS データセットで取ってくることが可能であり，その後のデータハンドリングも可能である．

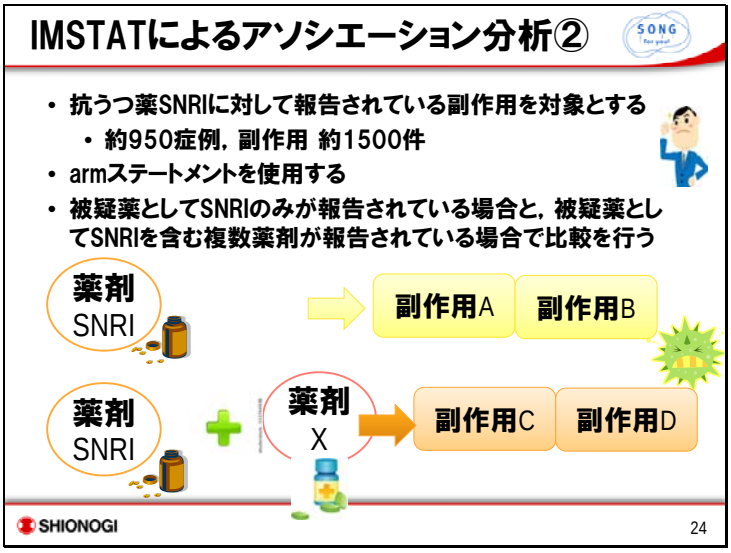
JADERに対する適用結果
（Support降順の上位10位の中で主要な関連ルール）

関連ルール	Support	Confidence	Lift
{AST**増加**} => {ALT**増加**}	0.0054	0.94	151.0
{**血中乳酸脱水素酵素増加**} => {ALT**増加**}	0.0016	0.82	131.5
{**頭痛**, **浮動性めまい**} => {**悪心**}	0.0014	0.89	46.3
{**易刺激性**,**不眠症**} => {**不安**}	0.0014	0.80	57.1
{**妄想症**} => {**幻覚**}	0.0012	0.87	59.6

- Support**の値は低かった**
- **ただし，**Confidence**は**0.8**を超えており，**Lift**も**1**を超えていたことから，挙がってきている関連ルールの関連性は強いと考えられた**

SHIONOGI 23

解析結果である．関連ルールに対して得られた Support の値は低かった．ただし，Confidence は 0.8 を超えており，Lift も 1 を超えていたことから，挙がってきている関連ルールの関連性は強いと考えられた．



次に，抗うつ薬 (Serotonin norepinephrine reuptake inhibitor: SNRI) に絞り，被疑薬として SNRI のみが報告されている場合と被疑薬で SNRI を含む複数薬剤が報告されている場合でそれぞれ副作用の組み合わせを評価した.

JADERに対する適用結果（抜粋）

SNRI単剤

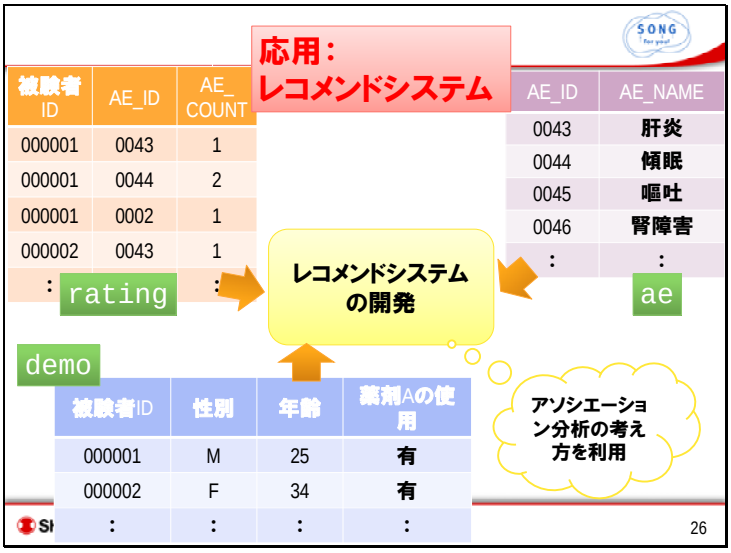
関連ルール	Support	Confidence	Lift
{不安} => {易刺激性}	0.0072	1	139.67
{AST増加} => {ALT増加}	0.0048	1	139.67
{易刺激性} => {不安}	0.0072	1	139.67

併用薬による影響の考察

SNRI+併用被疑薬

関連ルール	Support	Confidence	Lift
{肺水腫} => {心肺停止}	0.0058	1	172.0
{乳腺炎} => {高プロラクチン血症}	0.0039	1	172.0
{歩行不能} => {ジスキネジア}	0.0039	1	129.0
{不安,不眠症} => {易刺激性}	0.0039	1	129.0
{自殺行為} => {自殺念慮}	0.0039	1	86.0

結果を示す. Support の値が低いものの, これらの副作用の組み合わせの違いは併用薬による影響が含まれていると考えられる.



アソシエーションの考え方を応用して，レコメンドシステムの構築を検討した．これを用いることで，ある副作用が観測された被験者が次に発現すると考えられる副作用を予測することができる．

応用：レコメンドシステム (proc recommend)

```
proc recommend  recom = LASRLIB.movielens;
   add LASRLIB.movielens/
   item = PT_CODE user = isr rating = count;
 addtable LASRLIB.rating  /type = rating
       vars=(isr PT_CODE PT_KANJI  count);
 addtable LASRLIB.ae       /type = item;
 addtable LASRLIB.demo   /type = user;
run;
 method knn / label = "knn" k = 20 positive
 similarity = pc seed = 1234;
run;
 predict / method = knn  label="knn" Num = 5
             users = ("000001"~"0XXXX");
run;
```

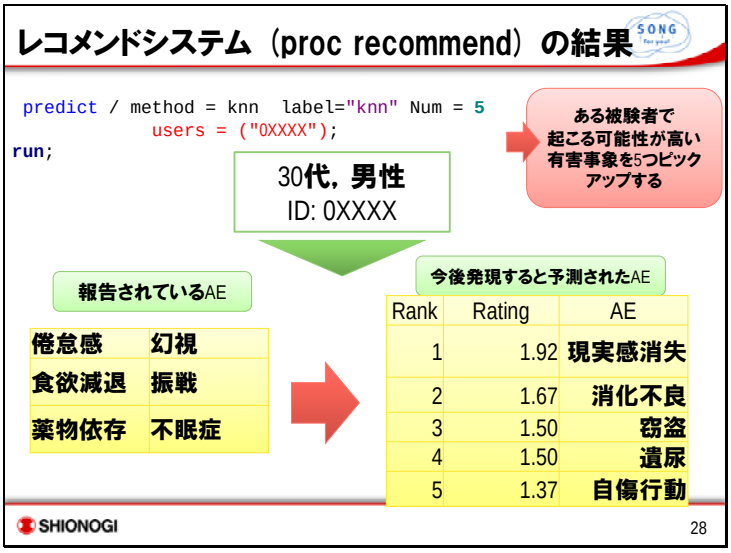
3つのデータセットを指定する

類似している被験者を探すアルゴリズム指定

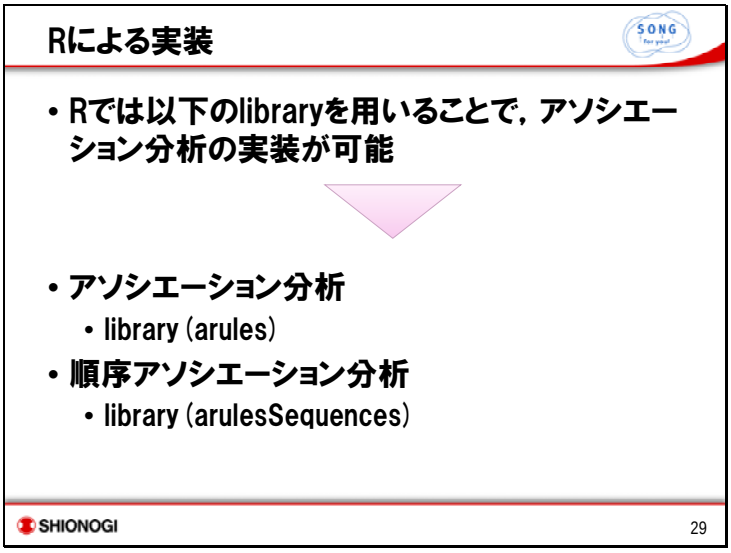
ある被験者集団で起こる可能性が高い有害事象を5つピックアップする
(高齢者，女性など)

SHIONOGI 27

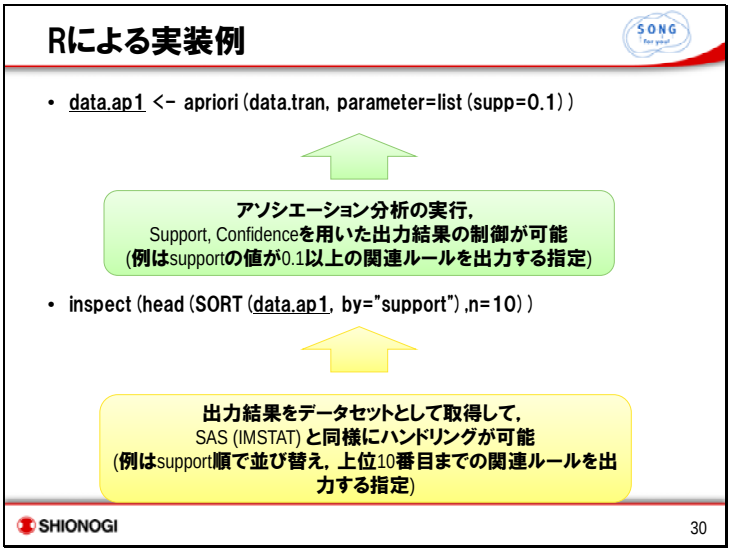
IMSTAT の recommend プロシジャを用いることでレコメンドシステムの実装が可能である．類似している被験者を探し，その被験者情報を用いて，該当する被験者に対して次に起こる可能性が高い有害事象をピックアップすることが可能である．



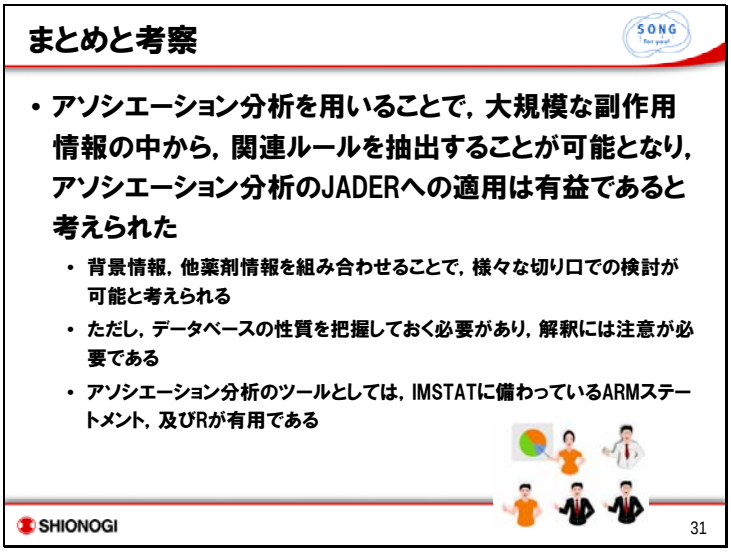
結果を示す. ここでは現実感消失等の副作用が挙がっている.



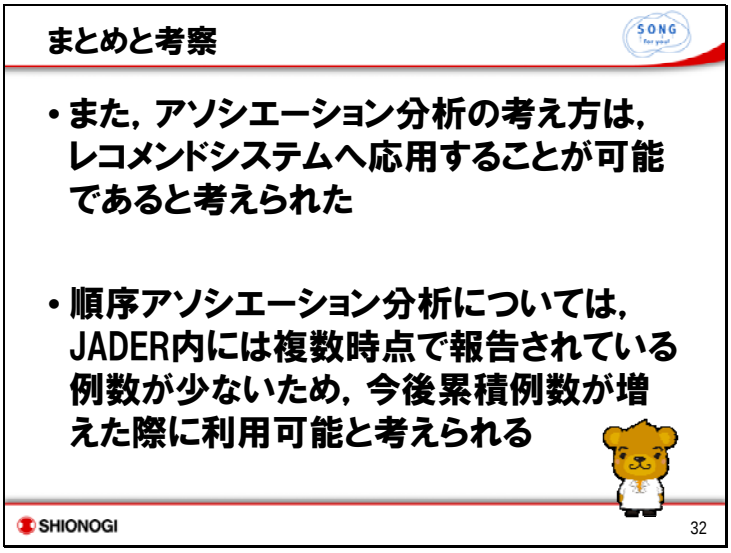
R (The R Foundation for Statistical Computing, Vienna, Austria) による実装についても紹介する．R ではアソシエーション分析は library (arules) を，順序アソシエーション分析は library (arulesSequences) を用いることで実装可能である．



プログラム例を示す．アソシエーション分析の実行の場合，Support，Confidence を用いた出力結果の制御が可能である．また，出力結果をデータセットとして取得して，ハンドリングすることも可能である．



まとめと考察である．アソシエーション分析を用いることで，大規模な副作用情報の中から，関連ルールを抽出することが可能となり，アソシエーション分析の JADER への適用は有益であると考えられた．ただし，データベースの性質を把握しておく必要があり，解釈には注意が必要である． アソシエーション分析のツールとしては，IMSTAT に備わっている ARM ステートメント，及び R が有用である．



また，アソシエーション分析の考え方は，レコメンドシステムへ応用することが可能であると考えられた．複数回報告されている症例が少ないため，順序アソシエーション分析に関してはもう少し集積例数が増えた際に有益な結果が得られると考えられる．

参考文献

- Agrawal, R., Imielinski, T., Swami, A., 1993. Mining association rules between sets of items in large databases. In: Buneman, P., Jajodia, S. (Eds.), Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. Washington, D.C., pp. 207-216.
- Philippe Lenca, Patrick Meyer, Beno î t Vaillant, St é phane Lallich. On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. European Journal of Operational Research, 184(2):610-626, 2008.

SHIONOGI 33

追加:

M. J. Zaki (2001). SPADE: An efficient algorithm for mining sequences. In Proc. of Machine Learning, 42, 31-60.

SONG for you!

End of Slides

SHIONOGI