

平成25年度
地方衛生研究所サーベイランス業務従事者研修
4月16日 14:30-15:30

実地疫学のための統計

疫学調査の基本ステップ

1. 集団発生の確認
2. “症例定義”の作成, 積極的な症例の探索
3. 現場および関連施設などの観察調査
4. 症例群の特徴を把握 : 時・場所・人
ラインリスティング→図式化
記述疫学
5. 感染源/感染経路やリスクファクターに関する仮説の設定
6. 仮説の検証
解析疫学
7. 感染拡大の防止策の実践、今後の予防策の提案
8. 報告書作成

(※必要な感染対策は適時に行なう)

内容

1. アウトブレイク調査データ解析のプロセス
2. 記述疫学で使う統計
3. 分析疫学で使う統計と検定
4. まとめ

アウトブレイク調査データ解析のプロセス

1. データクリーニング→データの質/精度を高める

- 異常値の取り扱い: 入力データの確認 (収集・コード化、入力値等)、欠損値などの取り扱い等

例) 1歳、95歳、その他 (10歳~60歳)

- 欠損値の確認: 欠損値の取り扱い方法を検討 (例: 職業で子供が欠損値、医療機関を受診していない者の検査データが欠損値→どう扱う?)

2. 測定する関連の計算、信頼区間

- アウトブレイク調査チーム: アウトブレイクの危険因子 (曝露と疾患の関連、発生の原因)・経路の測定

例) *E.coli*のアウトブレイクの原因を調べたところ、サラダまたはアルファアルファが下痢と強い関連があった→多くは相対危険度 (RR: Relative Risk) またはオッズ比 (OR: Odds Ratio)

3. 検定 (例: カイニ乗検定、分散分析等)

データクリーニング：数の分布

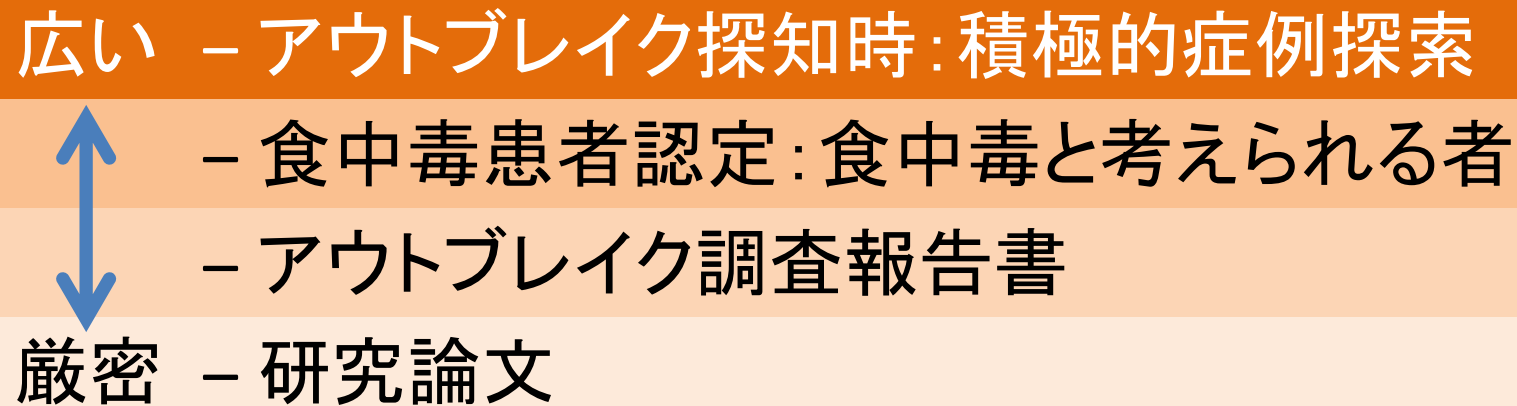
約40%は全体としての重要度は？

日時	人数	%	累積%	日時	人数	%	累積%
不明	29	38.7%	38.7%	4/19 12:00AM	2	2.7%	68.0%
4/18 3:00PM	1	1.3%	40.0%	4/19 12:30AM	7	9.3%	77.3%
4/18 9:00PM	1	1.3%	41.3%	4/19 1:00AM	10	13.3%	90.7%
4/18 9:15PM	1	1.3%	42.7%	4/19 2:00AM	3	4.0%	94.7%
4/18 9:30PM	2	2.7%	45.3%	4/19 2:15AM	1	1.3%	96.0%
4/18 9:45PM	2	2.7%	48.0%	4/19 2:30AM	2	2.8%	98.7%
4/18 10:00PM	1	1.3%	49.3%	4/19 10:30AM	1	1.3%	100.0%
4/18 10:15PM	1	1.3%	50.7%	合計	75	100.0%	100.0%
4/18 10:0PM	4	5.3%	56.0%				
4/18 11:00PM	5	6.7%	62.7%				
4/18 11:30PM	2	2.7%	65.3%				

症例定義の範囲

• 症例定義：目的により使い分ける

(CDC. Principles of Epidemiology in Public Health Practice, Third Edition)



感染症発生動向調査の定義に合致する症例

本事例で利用した症例定義

a. アウトブレイク探知時：積極的症例探索：326人

- 平成23年4月10日－29日に焼肉チェーン店舗を利用者かつ
- 消化器症状(下痢・血便・腹痛・嘔吐)・発熱のうち1つ以上を呈し、保健所へ相談あるいは有症苦情等の連絡をした者

b. 食中毒患者認定：181人

4月中に焼肉チェーン店を利用し喫食時間から10時間以上経過して発症し、次のいずれかに該当する者

- ① 血便を呈している者、
- ② 消化器症状(下痢、吐き気又は嘔吐、腹痛、渋り腹)が2つ以上あった者、
- ③ 消化器症状が1つとそれ以外の症状(発熱 37.5°C 以上、倦怠感、頭痛など)が1つ以上ある者、
- ④ 便から大腸菌O111あるいは腸管出血性大腸菌O111、O157を検出し、1つ以上の症状を呈する者

本事例で利用した症例定義

c. アウトブレイク調査報告書: 95人

- 2011年4月10日～29日、に焼肉チェーンで喫食者
かつ
- 消化器症状(下痢・、血便・、腹痛・または嘔吐)のうち、1つ以上を呈した者
かつ
- *E. coli* O157またはO111の分離・同定 [O111はVT産生の有無は問わず] またはO111抗LPS抗体が陽性者

d. 研究論文: 86人

- 2011年4月10日～29日、に焼肉チェーンで喫食者
かつ
- 消化器症状(下痢・、血便・、腹痛・または嘔吐)のうち、1つ以上を呈した者
かつ
- *E. coli* O111の分離・同定 [O111はVT産生の有無は問わず]) またはO111抗LPS抗体が陽性者

内容

1. アウトブレイク調査データ解析のプロセス
2. 記述疫学で使う統計
3. 分析疫学で使う統計と検定
4. まとめ

記述疫学で利用する統計

- 分布の特性を把握
 - 代表値の算出(対象者の属性等)
- 代表値
 - 人数・割合
 - 平均値: 分布の真ん中(重心)を表す
 - 標準偏差: 測定値のばらつき
 - 中央値・範囲: 全体の真ん中に来る測定値と最大値、最小値を記載(例: 年齢中央値64歳(範囲: 0-89歳))

対象者の属性

集団 No.	集団 の人数	データ 数	性別 (男)	年齢 中央値 (範囲) 歳	下痢	血便	STEC 感染 確定	HUS あり	症例定義 人(%)	症例定義 全て 満たす 人	罹患率 %
1	37	37	5	49 (45-57)	10	9	10	4	34 (92)	9	26
2	2	2	50	49 (46-52)	1	1	2	0	2 (100)	1	50
3	31	31	55	57 (45-68)	11	4	5	1	25 (81)	5	20
4	11	10	0	54 (53-55)	2	0	1	0	9 (90)	1	11
5	12	9	0	46 (43-48)	1	1	2	0	9 (100)	1	11
6	19	19	37	32 (15-49)	1	1	1	0	17 (89)	1	6
7	10	10	50	40 (40-42)	0	0	0	0	9 (90)	0	0
8	17	14	57	68 (65-73)	3	3	4	1	14 (100)	3	21
9	25	25	40	74 (69-75)	9	9	6	1	25 (100)	9	36
10	13	11	36	24 (21-48)	3	1	2	1	8 (73)	1	13
Total	177	168	32	53 (42-67)	41	29	33	8	152 (90)	31	20

記述疫学：曝露の状況を把握

Table 3. Relative Risk of Infection Associated with Sprouts and Other Raw Food Items in Univariable Analysis.

食 品	対象総数 人	対象集団 の曝露(%) 人(%)	曝露者中 の症例 (罹患率) 人(%)	対象者中 の非曝露 (率) 人(%)	対象中の 非曝露者 症例(罹患率) 人(%)	相対危険度 (95%信頼 区間)	P値
スプラウト	152	115 (76)	31 (27)	37 (24)	0	14.23 (2.55-∞)	0.001
トマト	152	50 (33)	14 (28)	102 (67)	17 (17)	1.68 (0.77-3.62)	0.18
きゅうり	152	50 (33)	14 (28)	102 (67)	17 (17)	1.68 (0.77-3.62)	0.18
白菜	152	45 (30)	13 (29)	107 (70)	18 (17)	1.72 (0.77-3.71)	0.17
赤チコリ	152	45 (30)	13 (29)	107 (70)	18 (17)	1.72 (0.77-3.71)	0.17
レタス	152	45 (30)	13 (29)	107 (70)	18 (17)	1.72 (0.77-3.71)	0.17

内容

1. アウトブレイク調査データ解析のプロセス
2. 記述疫学で使う統計
3. 分析疫学で使う統計と検定
4. まとめ

リスク比 (RR: Relative Risk)

	発症有り	発症なし	合計
曝露群	A	B	A+B
非曝露群	C	D	C+D

$$\text{リスク比(RR)} = \frac{[A/(A + B)]}{[C/(C + D)]}$$

- 曝露が発症と関連がない: RR=1
- 曝露が発症と正の関連がある: RR>1
- 曝露が発症と負の関連がある(予防因子含む): RR<1

リスク比を使う場合の調査デザイン

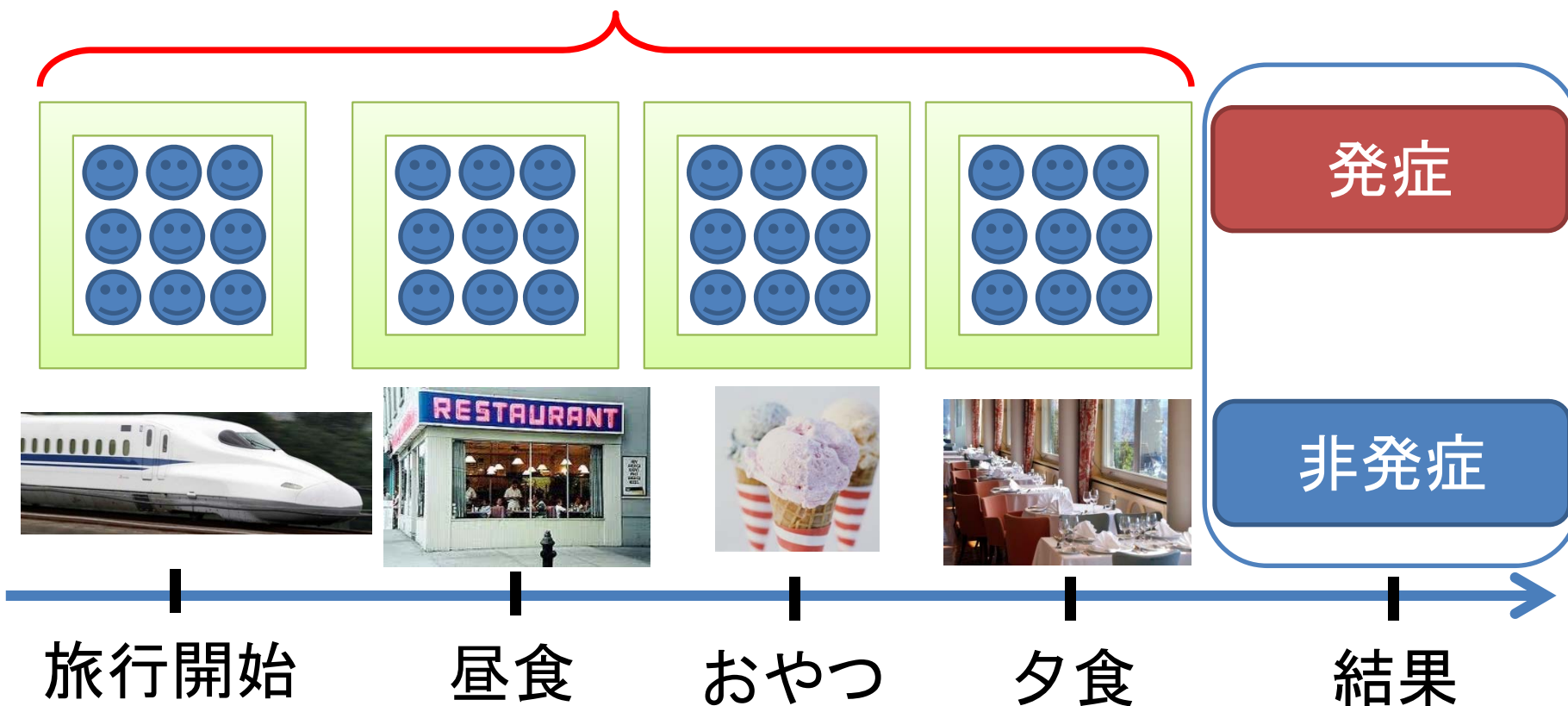
— 後ろ向きコホート研究 —

コホート研究の特徴

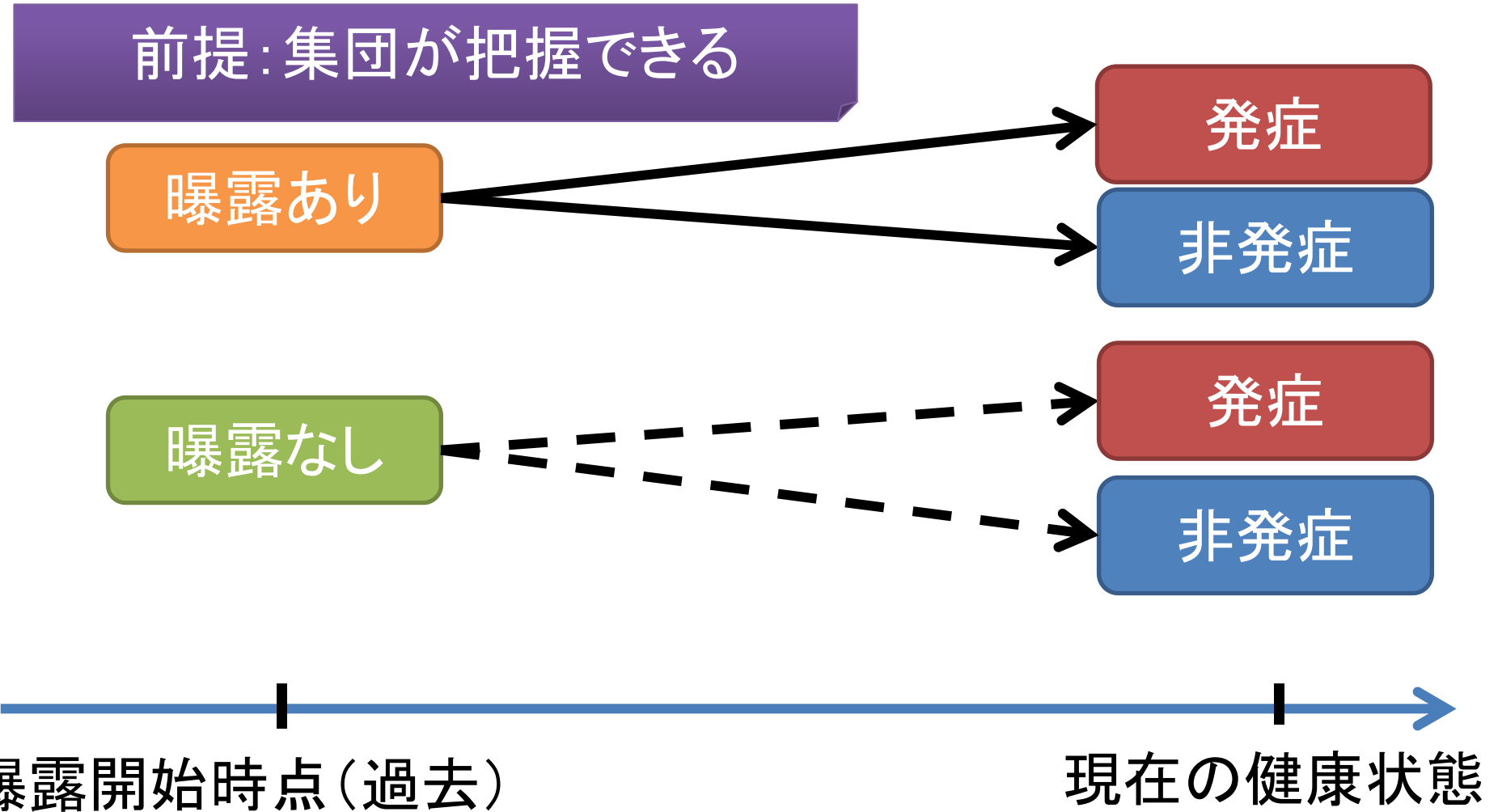
- 集団が定義できる場合に利用
- コホート研究の種類
 - 後ろ向きコホート研究：過去の曝露状況→現在の健康状態
 - 前向きコホート研究：現在の曝露状況→未来の健康状態
- 症例対照研究とコホート研究の比較
 - コホート研究：リスク比算出可能、原因と結果の検討可能
 - 症例対照研究：関連の検討
- 発症率の計算可能

コホート研究：集団の追跡可能

同一グループの曝露後の健康状況を追跡



後ろ向きコホート研究の時間的経過



リスク比 (RR: Relative Risk) の例

Table 3. Relative Risk of Infection Associated with Sprouts and Other Raw Food Items in Univariable Analysis.

食 品	対象総数 人	対象集団 の曝露(%) 人(%)	曝露者中 の症例 (罹患率) 人(%)	対象者中 の非曝露 (率) 人(%)	対象中の 非曝露者 症例(罹患率) 人(%)	相対危険度 (95%信頼 区間)	P値
スプラウト	152	115 (76)	31 (27)	37 (24)	0	14.23 (2.55-∞)	0.001
トマト	152	50 (33)	14 (28)	102 (67)	17 (17)	1.68 (0.77-3.62)	0.18
きゅうり	152	50 (33)	14 (28)	102 (67)	17 (17)	1.68 (0.77-3.62)	0.18
白菜	152	45 (30)	13 (29)	107 (70)	18 (17)	1.72 (0.77-3.71)	0.17
赤チコリ	152	45 (30)	13 (29)	107 (70)	18 (17)	1.72 (0.77-3.71)	0.17
レタス	152	45 (30)	13 (29)	107 (70)	18 (17)	1.72 (0.77-3.71)	0.17

オッズ比 (OR: Odds Ratio)

	発症有り	発症なし	合計
曝露あり	A	B	A+B
曝露なし	C	D	C+D

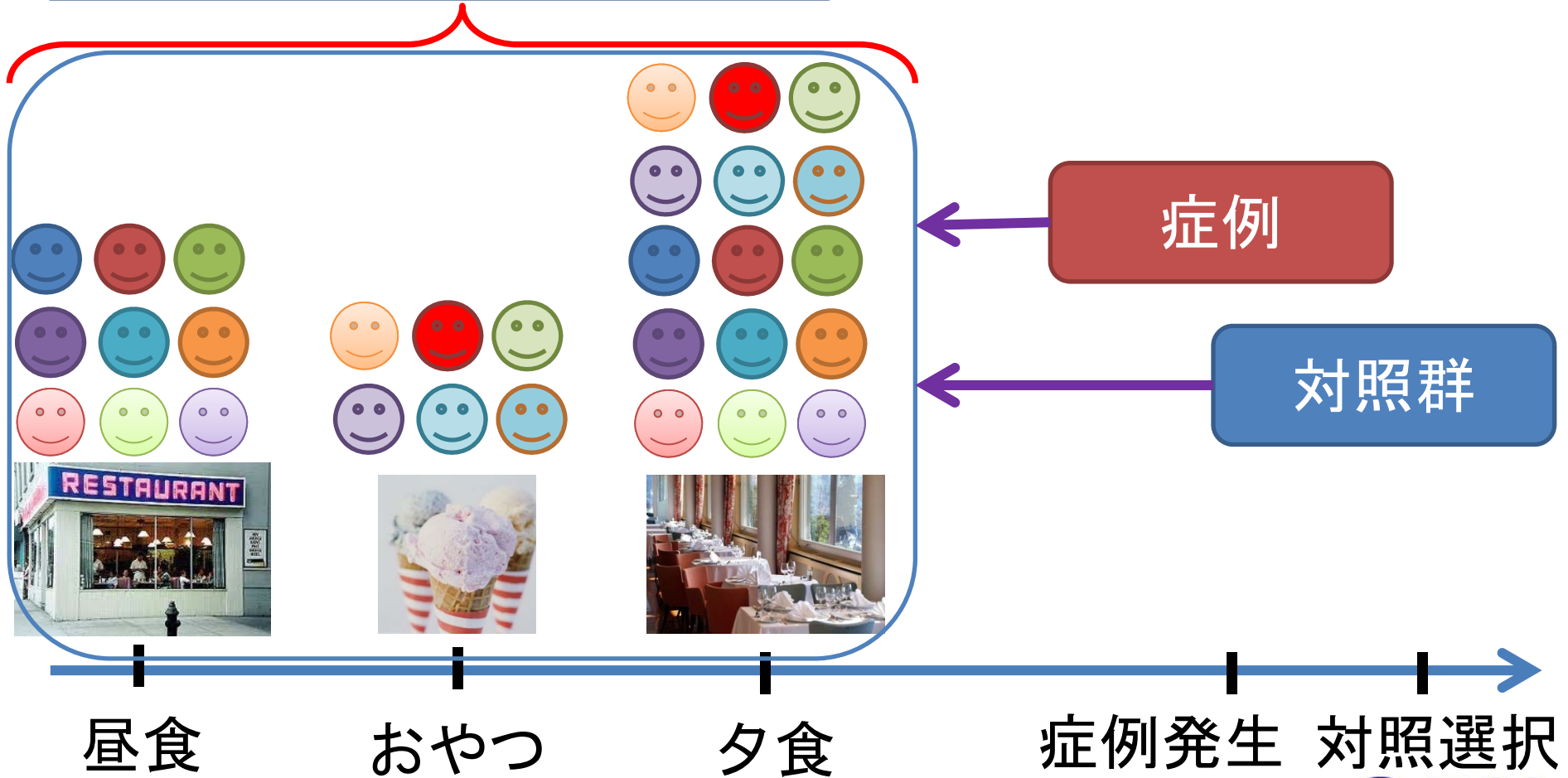
$$\text{オッズ比(OR)} = \frac{(A/C)}{(B/D)} = \frac{(A \times D)}{(B \times C)}$$

- 曝露が発症と関連がない: OR=1
- 曝露が発症と正の関連がある: OR>1
- 曝露が発症と負の関連がある(予防因子含む): OR<1

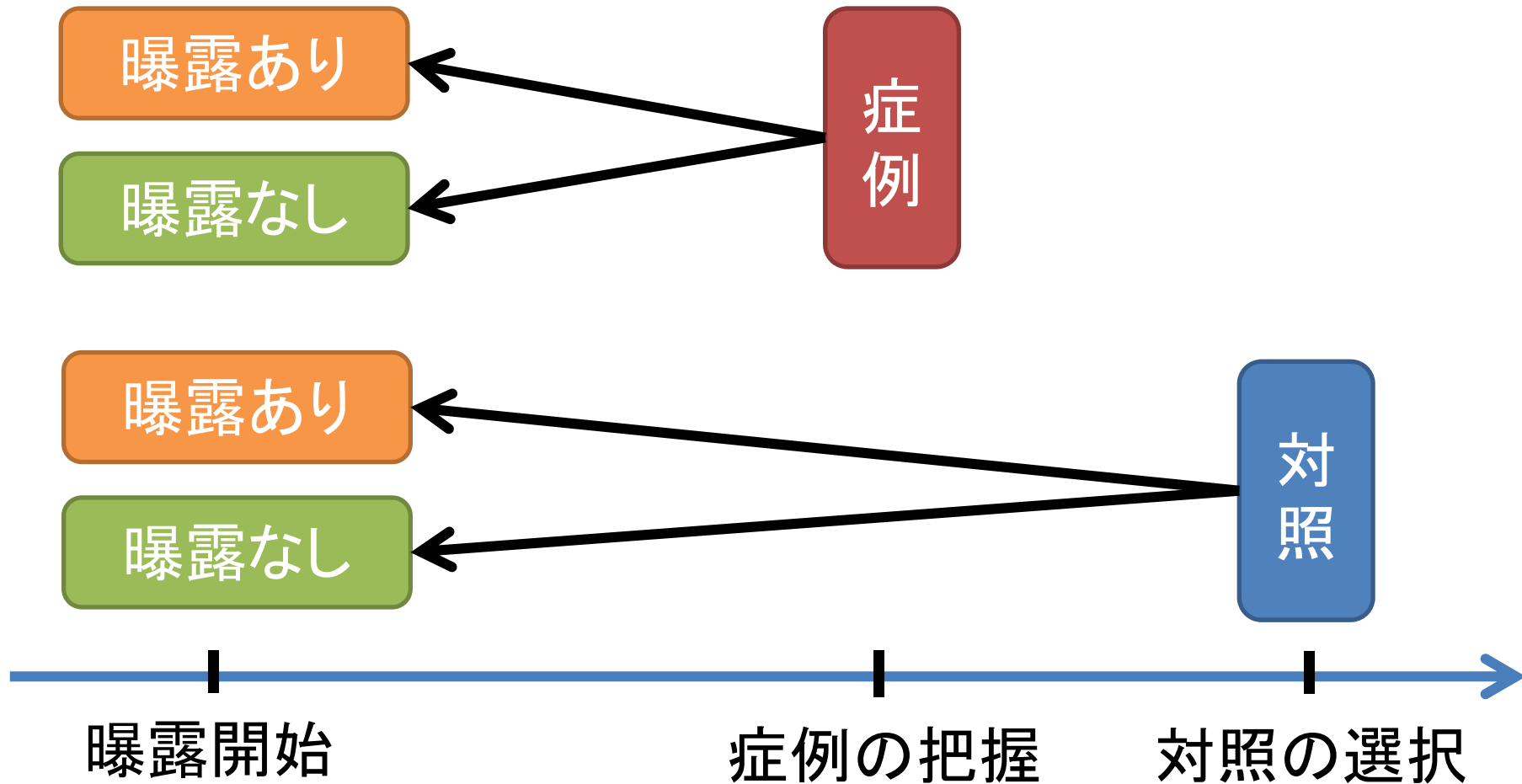
オッズ比を使う場合の調査デザイン — 症例対照研究 —

症例対照研究：集団の追跡不可能

施設により利用者異なる



症例対照研究の時間的経過



オッズ比

オッズ比 = 7.84
(95%信頼区間: 1.52-54.2)

TABLE 4

Comparison between the secondary attack rate in siblings and parents, pandemic H1N1 influenza outbreak, Japan, May-June 2009 (n=143)

	症例	対照	合計	OR(95%CI)	P-value
兄弟	10	51	61	7.84	<0.01*
両親	2	80	82	(1.52-54.2)	
	12	131	143		

CI: confidence interval; OR: odds ratio.
*Chi-square test

TABLE 5

Comparison between household contacts receiving antiviral prophylaxis and those not, pandemic H1N1 influenza outbreak, Japan, May-June 2009 (n=293)

	症例	対照	合計	OR(95%CI)	P-value
予防投薬あり	1	121	122	0.10	<0.05*
予防投薬なし	13	158	171	(0-0.75)	
合計	14	279***	293		
予防投薬あり(20歳未満)	1	23	24	0.15	0.09**
予防投薬なし(20歳未満)	11	39	50	(0.01-1.30)	
合計	12	62	74		
予防投薬あり(20歳以上)	0	90	90	0	0.50**
予防投薬なし(20歳以上)	2	105	107	(0-4.86)	
合計	2	195	197		

CI: confidence interval; OR: odds ratio.

* Chi-square test

** Fisher's exact test

*** Including 14 without prophylaxis and eight with prophylaxis for whom the age was not known.

マッチングした場合のオッズ比

		対照		
		曝露有り	曝露なし	合計
症例	曝露あり	e	f	$e+f$
	曝露なし	g	h	$g+h$
合計		$e+g$	$f+h$	

$$\text{オッズ比(OR)} = \frac{\text{症例の曝露ありで対照で曝露なし}}{\text{症例の曝露なしで対照の曝露あり}} = \frac{f}{g}$$

- 曝露が発症と関連がない: $OR=1$
- 曝露が発症と正の関連がある: $OR>1$
- 曝露が発症と負の関連がある(予防因子含む): $OR<1$

マッチングしたオッズ比 (症例対照研究)

N Engl J Med. 2011 Nov 10;365(19):1763-70.

Table 1. Vegetables or Fruits Evaluated in a Case–Control Study in the German Outbreak.*

食 品	症例の曝露 人数/合計人数(%)	対照の曝露 人数/合計人数(%)	マッチングした オッズ比 (95%信頼区間)	P値
スプラウト	6/24 (25)	7/80 (9)	4.35 (1.05–18.0)	0.04
きゅうり	22/25 (88)	52/79 (66)	3.53 (0.96–12.9)	0.06
りんご	22/24 (92)	57/81 (70)	3.91 (0.86–17.7)	0.08
コショウ	16/24 (67)	35/80 (44)	2.66 (0.90–7.9)	0.08
いちご	19/26 (73)	43/81 (53)	2.33 (0.90–6.0)	0.08

条件付きロジスティック回帰分析で算出可能

区間推定・検定

- 95%信頼区間(区間推定)
 - 95%の範囲内に真の値が存在
 - 確からしさの指標: 95%信頼区間の上限と下限の間に1を含まない
 - 範囲が広いとデータのばらつきが大きい
- 検定(P値)
 - $P < 0.05$: 帰無仮説を棄却する

マッチングしたオッズ比 (症例対照研究)

N Engl J Med. 2011 Nov 10;365(19):1763-70.

Table 1. Vegetables or Fruits Evaluated in a Case–Control Study in the German Outbreak.*

食 品	症例の曝露 人数/合計人数(%)	対照の曝露 人数/合計人数(%)	マッチングした オッズ比 (95%信頼区間)	P値
スプラウト	6/24 (25)	7/80 (9)	4.35 (1.05–18.0)	0.04
きゅうり	22/25 (88)	52/79 (66)	3.53 (0.96–12.9)	0.06
りんご	22/24 (92)	57/81 (70)	3.91 (0.86–17.7)	0.08
コショウ	16/24 (67)	35/80 (44)	2.66 (0.90–7.9)	0.08
いちご	19/26 (73)	43/81 (53)	2.33 (0.90–6.0)	0.08

点推定

区間推定

データの種類

- 質的データ(カテゴリカル・データ、質的変数)
 - 名義データ: 数値が区別の意味しか持たない(例: 疾病の有無、性別、学歴、居住地域、移動手段[徒歩、車、自転車、バイク])
- 量的データ(量的変数)
 - 連続データ: 実測値、数値で表される(例: 身長、体重など)
 - 離散データ: 整数で表されたり、区間で区切られているもの(電車の運賃、他)

統計方法の選択

- 分布を仮定する(パラメトリック)方法
 - 母集団の正規性や等分散性が仮定される
 - 平均値の差の検定(t 検定)や分散分析(F 検定)等
 - ※標本サイズが小さい場合、統計量の分布型が不正確なことが多いためパラメトリックな手法の適用は不適切になる事が多い → 分布を仮定しない方法
- 分布を仮定しない(ノンパラメトリック)方法
 - 母集団の分布型(母数)について一切の仮定を設けない(=分布を仮定しない)

尺度・変数・データの種類

- 名義尺度:単に区別するために用いられている尺度
例)性別を1. 男、2. 女としたり、血液型を1. A型、2. B型、3. O型、4. AB型としたもの。変数の値に意味はない(平均値などの意味ない)。
- 順序尺度:大小関係にのみ意味がある尺度
例)患者の転帰を1. 退院、2. 入院中、3. 死亡を数値に対応させたもの。平均値は定義できないが中央値は定義できる。
- 間隔尺度(距離尺度):数値の差のみ意味あり(順序尺度の性質もあり)
例)温度が 10°C から 15°C になった時、50%の温度上昇と表現しない。ともに 5°C の温度上昇と表現する。 5°C の数値に意味がある。
- 比例尺度(比尺度):数値の差とともに数値の比にも意味がある尺度(順序尺度・間隔尺度の性質もあり)
例)体重が50kgから60kgに増加と100kgから110kgに増加は同じ10kgの増加でも、前者は20%増、後者は10%増と表現する。比が定義できる。絶対零点を持つことと同じことを表す。

分布の仮定での分類

	分布仮定	分布仮定しない
対象とする 統計量	平均値, 分散 積率相関係数 (Pearson)	中央値, 範囲, 度数 順位相関係数 (Spearman)
尺度水準	間隔尺度, 比例尺度	名義尺度, 順序尺度 間隔尺度, 比例尺度
母集団の 分布型	正規分布を仮定 等分散性を仮定	不問
標本サイズ	小さすぎてはいけない	不問

検定手法の選択方法

検定目的	パラメトリック	ノンパラメトリック	
		名義尺度	順序尺度以上 ^{a)}
独立性	<ul style="list-style-type: none"> 相関係数の検定 	<ul style="list-style-type: none"> χ^2検定 フィッシャーの正確確率検定 	<ul style="list-style-type: none"> χ^2検定 フィッシャーの正確確率検定
比率の差		<ul style="list-style-type: none"> χ^2検定 フィッシャーの正確確率検定 	<ul style="list-style-type: none"> χ^2検定 フィッシャーの正確確率検定
対応のない2標本の差	<ul style="list-style-type: none"> 平均値の差のt検定 		<ul style="list-style-type: none"> マン・ホイットニーのU検定 ウィルコクソン符号付順位和検定
対応のないK標本の差	<ul style="list-style-type: none"> 一元配置分散分析 		<ul style="list-style-type: none"> クラスカル・ウォリス検定

a) 順序尺度・間隔尺度・比例尺度のいずれかである場合

検定の選択方法

- χ^2 検定VSフィッシャーの正確検定
 - χ^2 検定: セルの期待値が ≥ 5
 - フィッシャーの正確検定: セルの期待値 < 5 が1つでもある場合
- 差の検定
 - 量的データ分布仮定できる
 - ✓ t検定(2群)、一元配置分散分析(3群以上)
 - 量的データ分布仮定できない
 - ✓ 対応のない2群の差の検定: マンホイットニーのU検定
 - ✓ 対応のある2群の差の検定: ウィルコクソン順位和検定
 - ✓ 対応のない3群以上の差の検定: クラスカル・ウォリス検定
 - ✓ 対応のある3群以上の差の検定: フリードマン検定

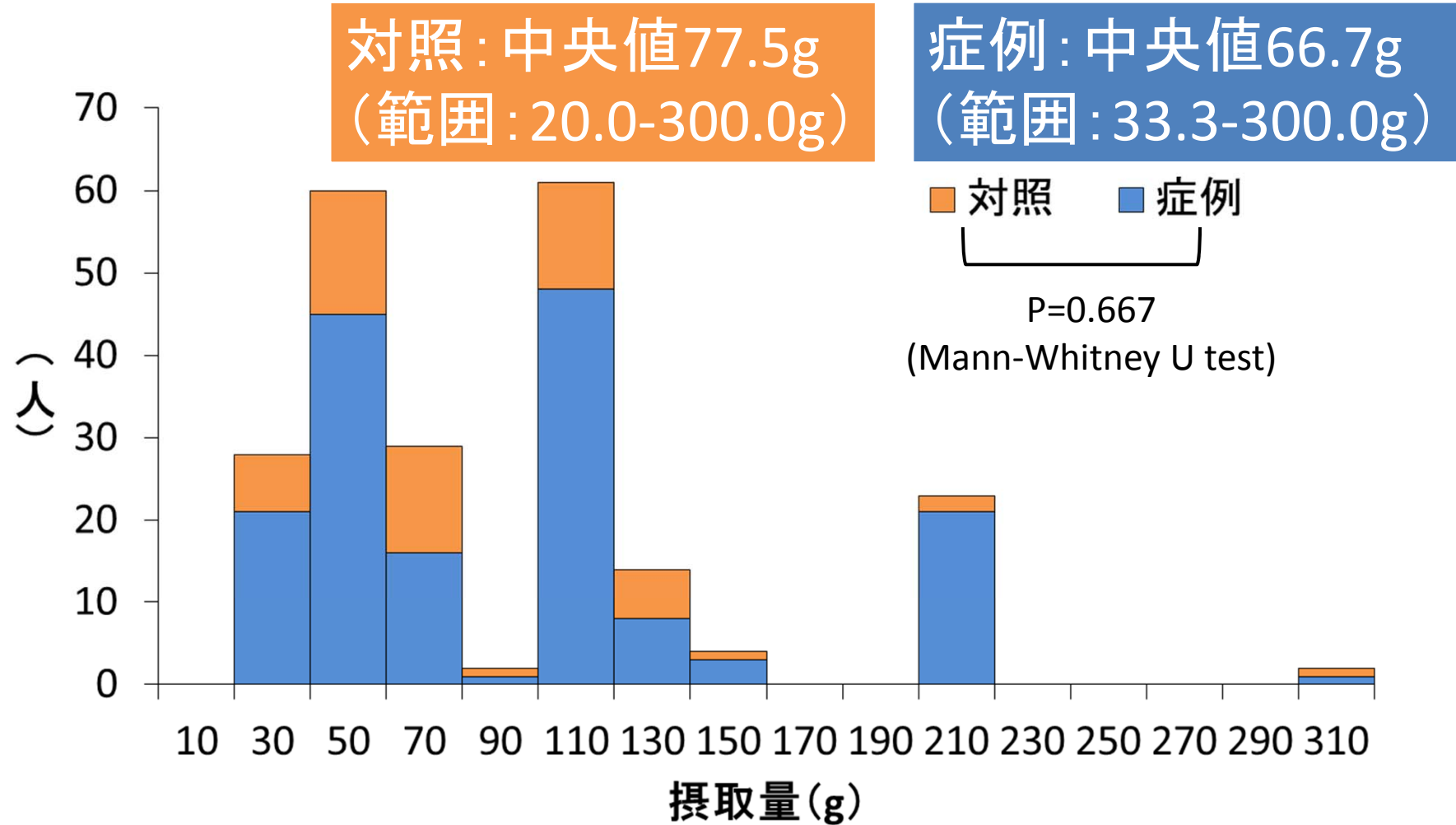
2 × 2表で利用される統計検定の 使い分け方法

検定の種類	大まかな基準
Pearsonの χ^2 検定	サンプルサイズ>100 セルの期待値>10
Yatesの χ^2 検定	サンプルサイズ>30 セルの期待値 \geq 5
Mantel-Haenszel χ^2 検定	サンプルサイズ>30 順序変数
Fisher's正確テスト	サンプルサイズ<30 セルの期待値<5

検定の選択方法

- χ^2 検定VSフィッシャーの正確検定
 - χ^2 検定: セルの期待値が ≥ 5
 - フィッシャーの正確検定: セルの期待値 < 5 が1つでもある場合
- 差の検定
 - 量的データ分布仮定できる
 - ✓ t検定(2群)、一元配置分散分析(3群以上)
 - 量的データ分布仮定できない
 - ✓ 対応のない2群の差の検定: マンホイットニーのU検定
 - ✓ 対応のある2群の差の検定: ウィルコクソン順位和検定
 - ✓ 対応のない3群以上の差の検定: クラスカル・ウォリス検定
 - ✓ 対応のある3群以上の差の検定: フリードマン検定

ヒラメの喫食量 (2010年事例)



中央値の比較

マンホイットニーのU検定 (Mann-Whitney's U test)

Table 4. Hospital Course in the 2008 *Escherichia coli* O111 Outbreak in Oklahoma^a

変数	全入院者		HUS発症例		HUS非発症例		P値 ^b
	人	中央値(範囲)	人	中央値(範囲)	人	中央値(範囲)	
入院日数	72	5.0 (1-55)	26	17.0 (3-55)	46	3.5 (1-36)	<.001
ICU日数	72	0 (0-41)	26	7.5 (0-41)	46	0 (0-15)	<.001
透析期間	18	6.0 (1-31)	17	6.0 (1-31)	1	12.0 (12)	.40
透析日数	18	9.5 (1-27)	17	9.0 (1-25)	1	27.0 (27)	.14
最大白血球数, / μ L	72	16 400 (5200-51 600)	26	24 700 (10 100-51 600)	46	13 800 (5200-26 300)	<.001
最小ヘモグロビン, g/dL	72	11.2 (5.1-15.4)	26	7.4 (5.1-12.2)	46	11.8 (7.7-15.4)	<.001
最大血小板数, $\times 10^3 / \mu$ L	72	180.0 (11-444)	26	31.5 (11-139)	46	216.0 (61-444)	<.001
最大血清クレアチニン値, mg/dL	72	1.1 (0.2-8.3)	26	3.7 (1.0-8.3)	46	0.8 (0.2-7.1)	<.001

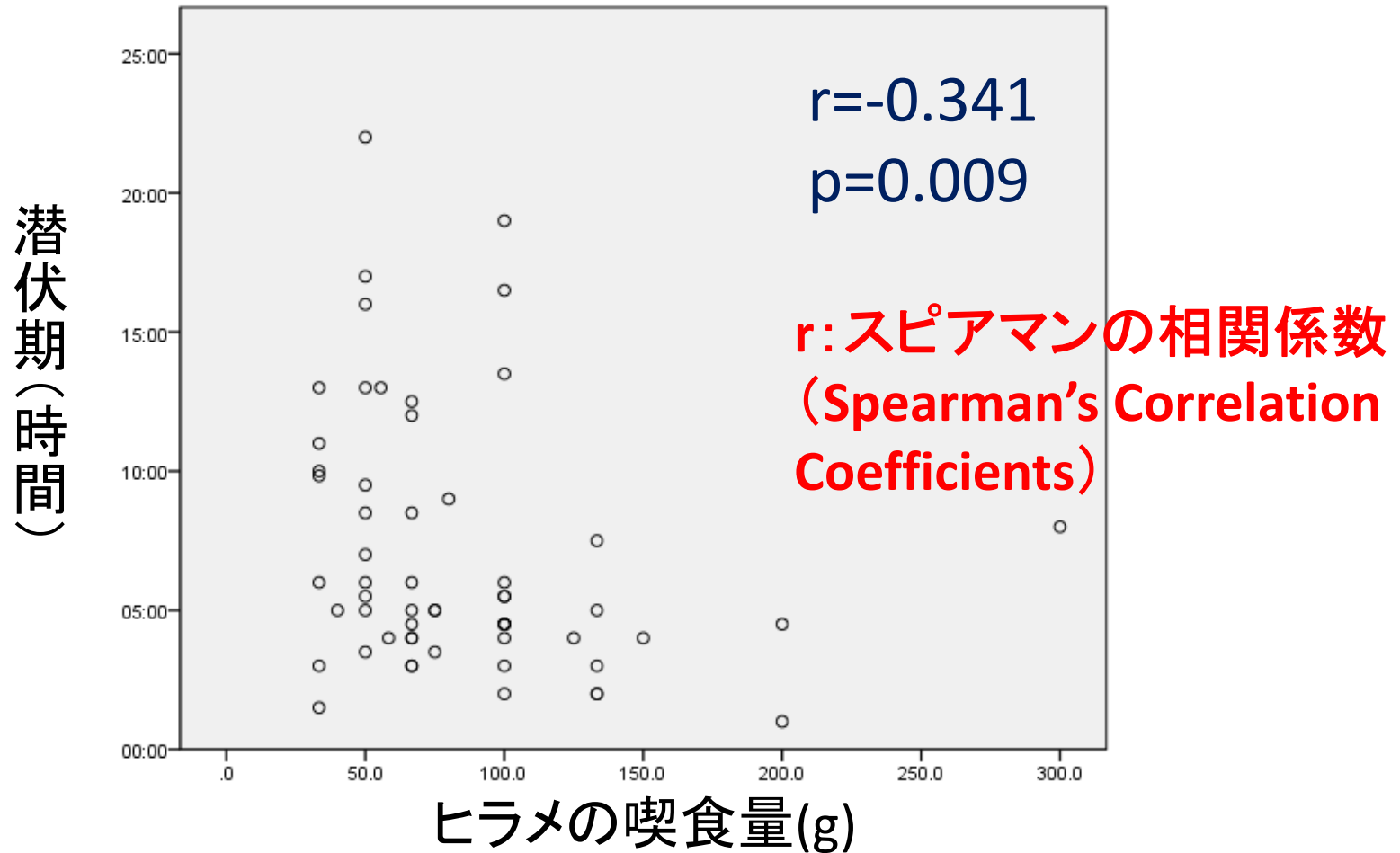
Abbreviation: HUS, hemolytic uremic syndrome.

SI conversion factors: To convert creatinine level to micromoles per liter, multiply by 88.4; hemoglobin level to grams per liter, multiply by 10.0; platelet count to $\times 10^9/L$, multiply by 1.0; white blood cell count to $\times 10^9/L$, multiply by 0.001.

^a $P < .05$ for boldfaced rows.

^b Calculated using the 2-sided Wilcoxon rank sum test.

散布図・相関係数 (喫食と潜伏期の関連)



内容

1. アウトブレイク調査データ解析のプロセス
2. 記述疫学で使う統計
3. 分析疫学で使う統計と検定
4. まとめ

まとめ

- データ: クリーニング、精度の検討
- 記述疫学:
 - 対象の属性や曝露に関する代表値を算出
 - 仮説の検討
- 分析疫学:
 - 解析方法: データの形式、分布等で使い分ける
 - 点推定、区間推定の解釈
 - 検定方法の選択