

# 実務者のためのデータベース研究講座 コロナから学ぶ疫学研究実践

長崎大学大学院医歯薬学総合研究科

長崎大学医学部


臨床疫学分野

佐藤 泉美



March 7, 2023


# 疫学研究始めよう



僕 医療情報データベースを使って  
疫学研究やりたいんですよ



ほう  
どんな疫学研究したいのかね？



プロペンシテイスコア使いたいです  
なんかカッコいいじゃないですか



・・・ A子さん  
B男くんに疫学研究教えてあげて



喜んで!

# 疫学研究始めよう



コロナ感染症関連の研究がいいんじゃないかな？  
パンデミック後の数年で一連の疫学研究が  
実施されているし身近で頭に入りやすいだろう

パンデミックでは実践の疫学研究を  
学べた気がします  
じゃあざっと説明しますよ！

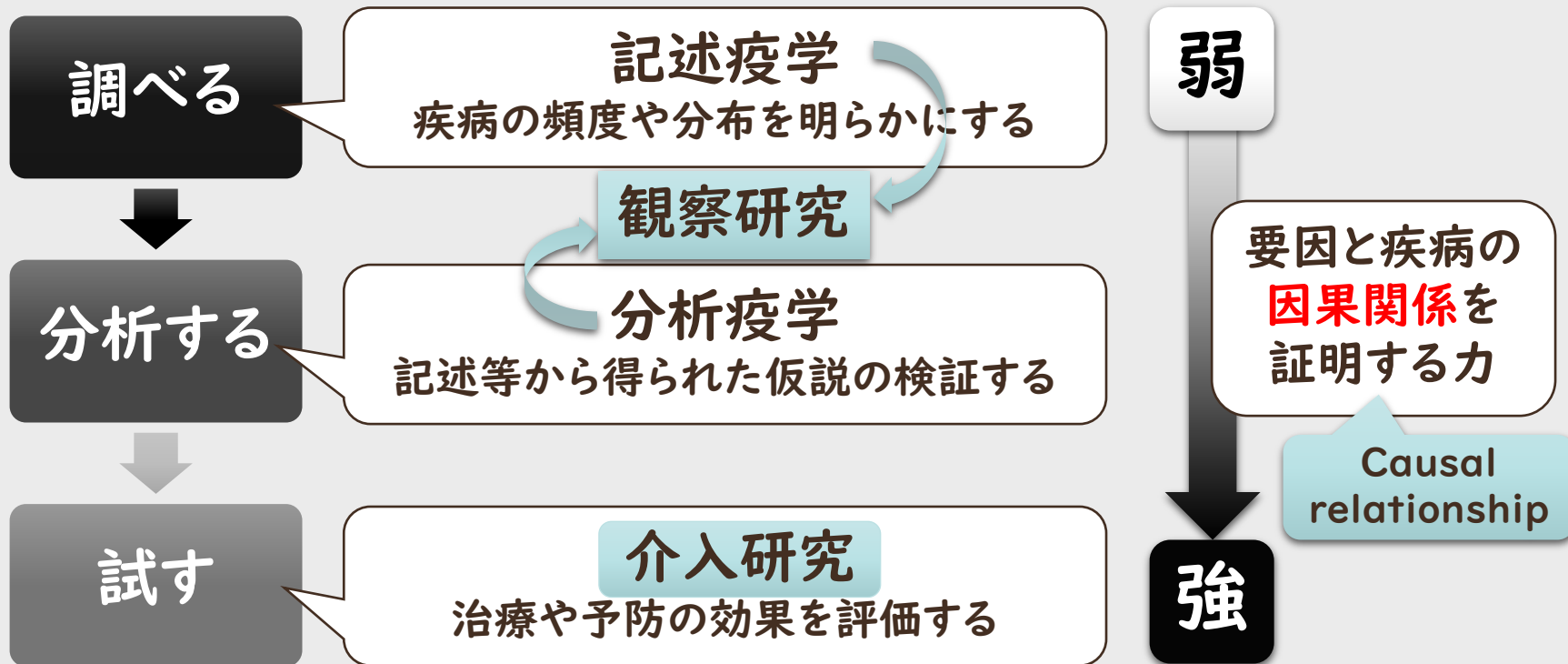


お願いします！  
押忍！

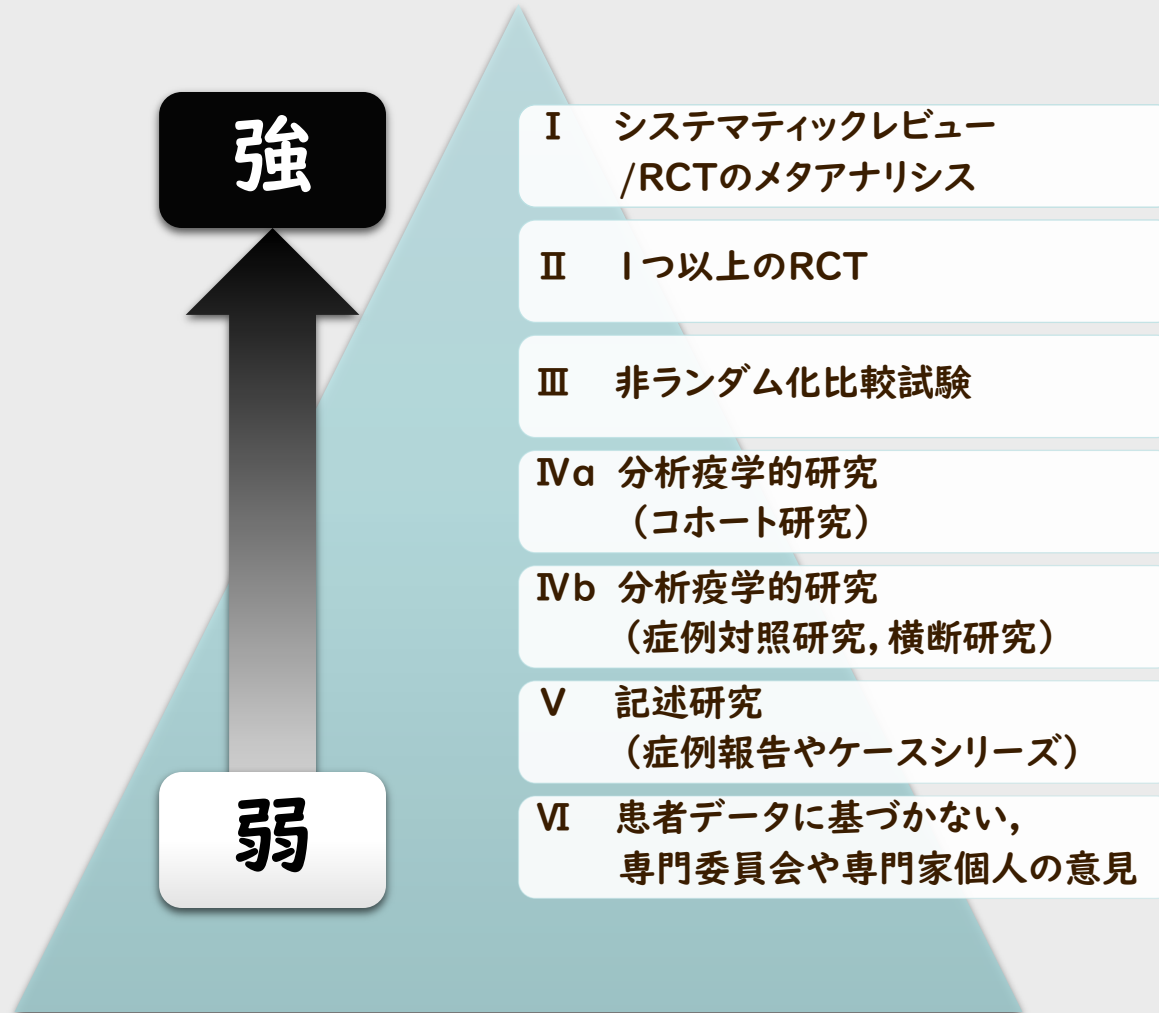
# 疫学研究とは？

・健康に関連する状態や事象の集団中の分布や決定要因を研究しかつその研究成果を健康問題の予防やコントロールのために適用する学問

(Last JM, A dictionary of epidemiology 2001)



# 研究デザインとエビデンスの強さ



# 觀察研究

## 記述疫学 (descriptive epidemiology)

- ・ 横断研究 (cross-sectional study)
- ・ 生態学的研究 (ecological study)

## 分析疫学 (analytical epidemiology)

- ・ コホート研究 (cohort study)
- ・ 症例対照研究 (case-control study)

# 記述疫学

## 疾病の発生状況の分布の記述 頻度やパターン

人

性別, 年齢, 人種,  
BMI, 生活習慣等

時間

時間的推移, 季節  
変動, 日内変動等

場所

国, 州, 県, 市等

原因仮説の設定も探索的に行われる  
因果関係の証明は困難

# 記述疫学

## 横断研究 (cross-sectional study)

Prevalence

- ・ 一時点の曝露と有病状態（有病率）を同時に観察
- ・ 因果関連の特定は困難；時間的關係が不明確
- ・ 時間とコストや労力の削減；比較的实施が容易

## 生態学的研究 (ecological study)

- ・ 集団単位で曝露状況や疾病頻度を観察・比較
  - ・ 2つの要因の相関関係を見る
- ・ 主に仮説検証よりも仮説設定のため

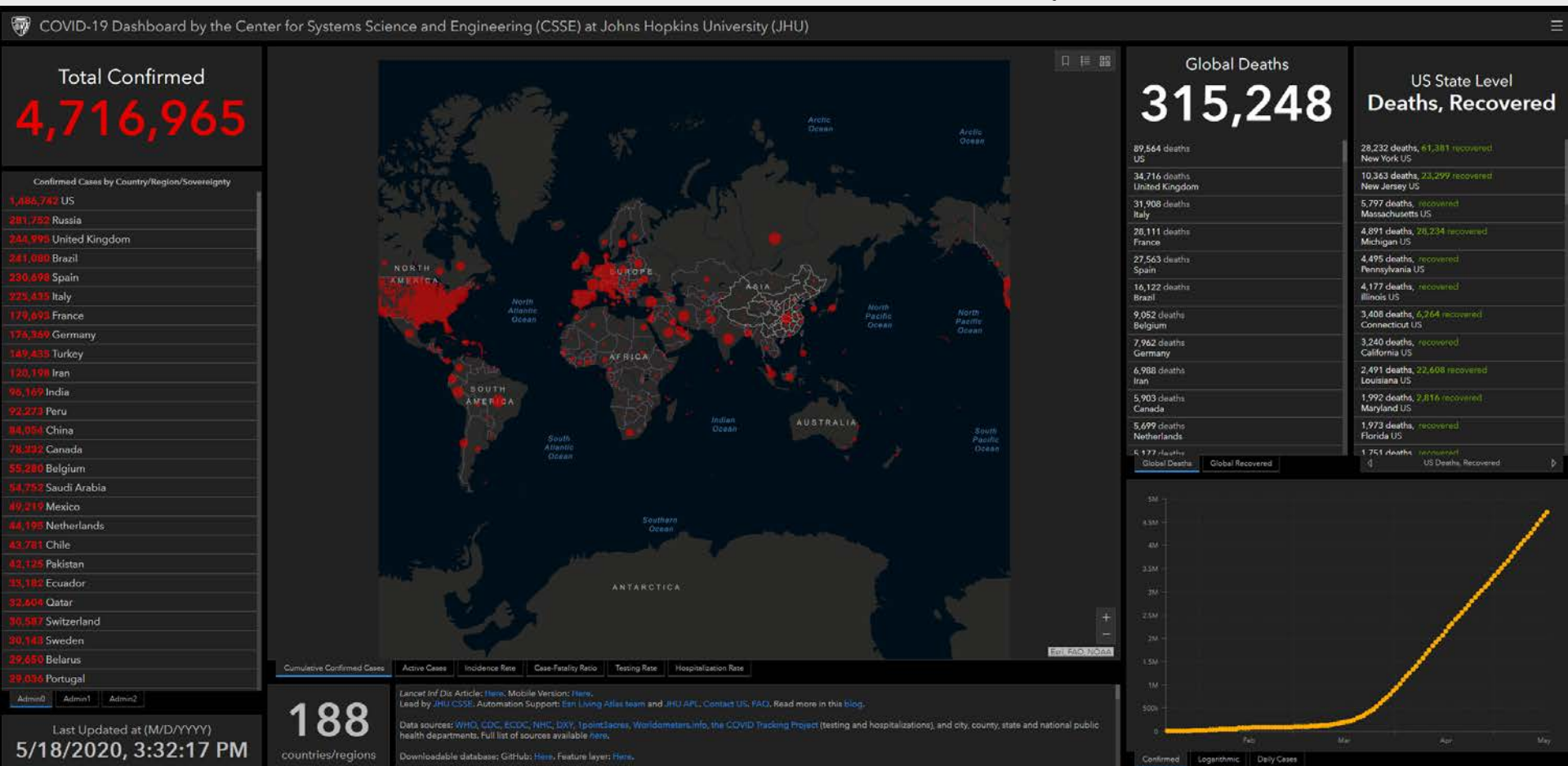


# 色々な横断研究

- ・ 症例研究(case series)
  - ・ 患者集団を対象にし疾患の特徴を記述
    - ・ 強いリスク要因を検出できる可能性
    - ・ 例) サル痘患者
- ・ 連続横断研究(serial prevalence study)
  - ・ ある期間ごとに定期的に実施される横断研究
  - ・ 経時変化の推論
    - ・ 同じ集団ではないのでコホート研究ではない
    - ・ 実施される毎に集団の特性が異なる

# 例) 身近な記述疫学

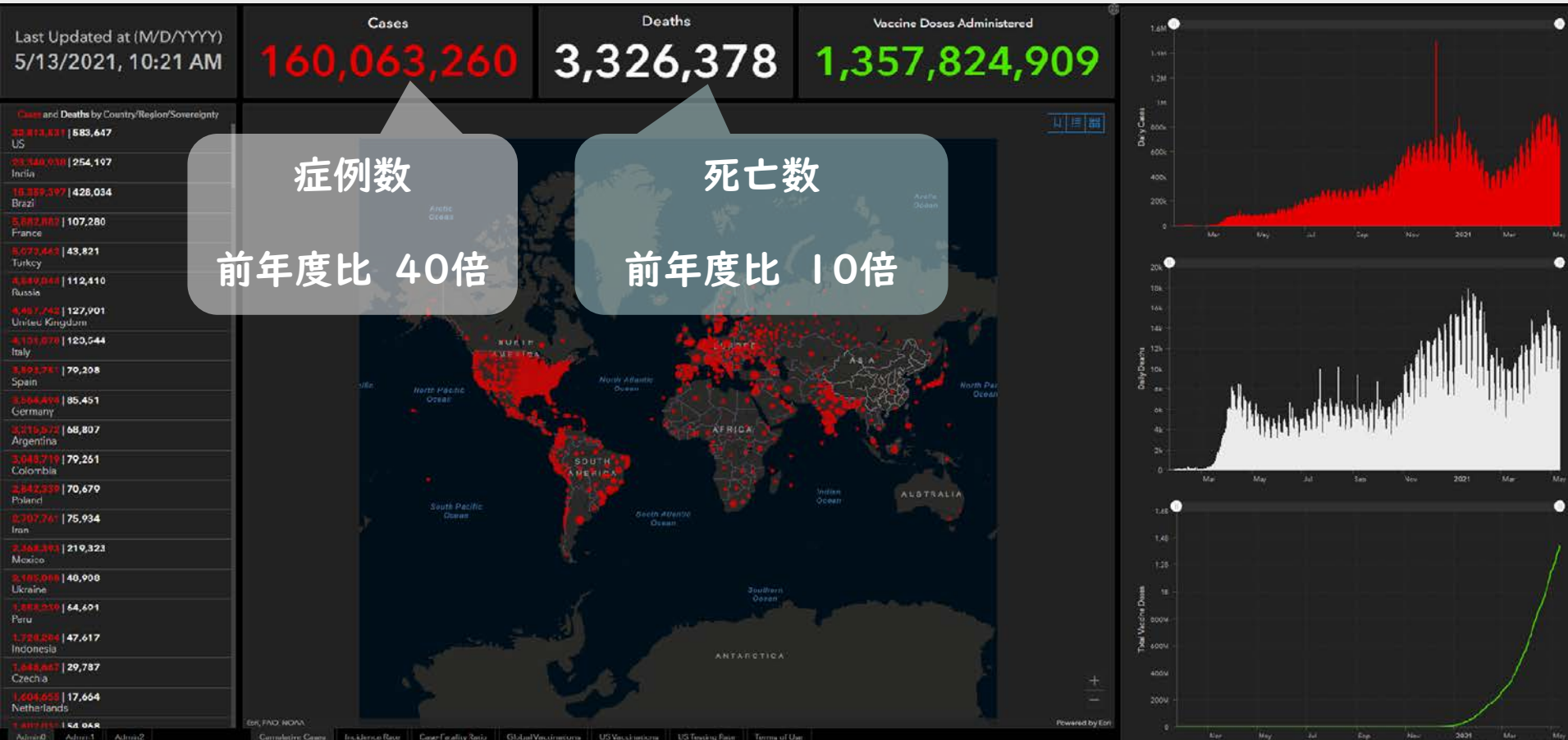
## · Covid-19 cases & deaths on May 18, 2020



\*The center for systems Science and Engineering at Johns Hopkins University

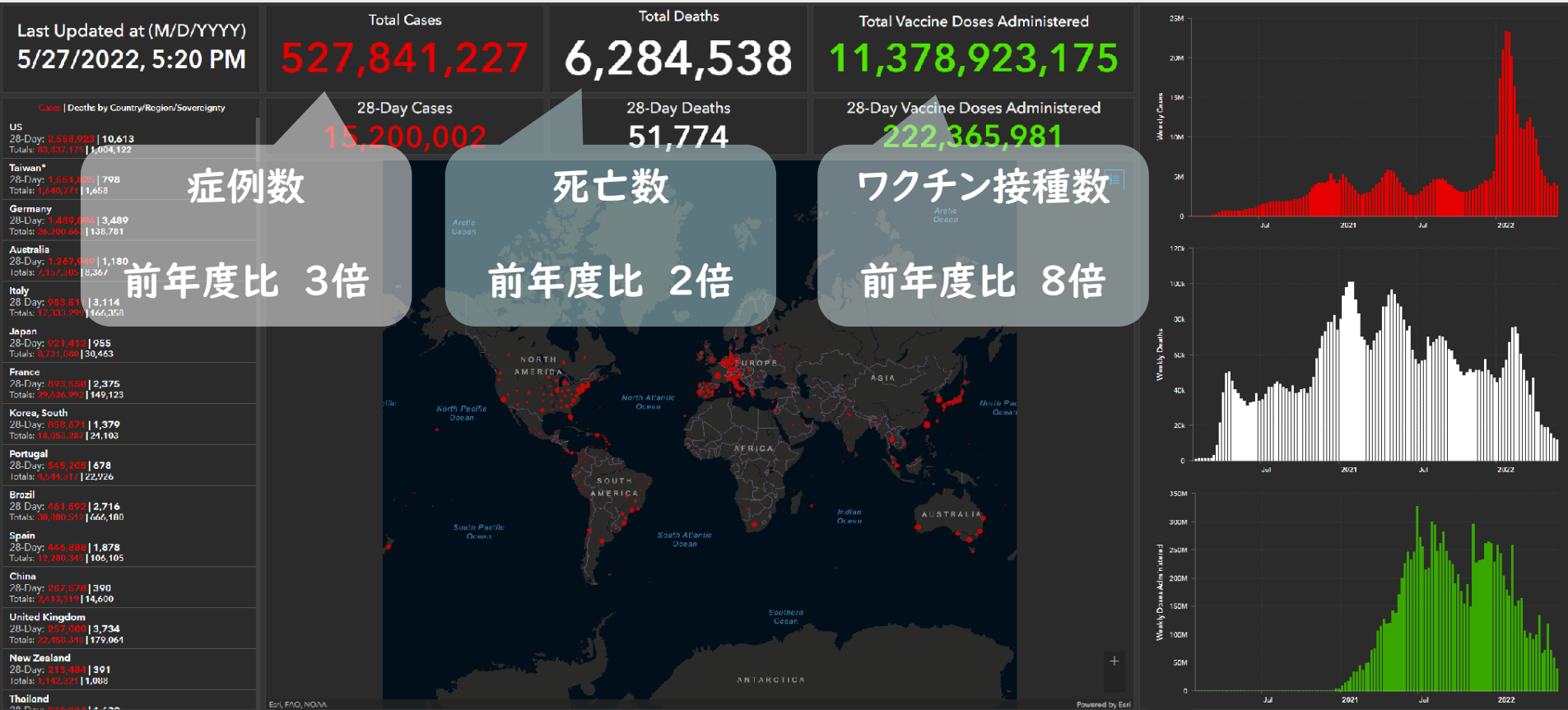
# 例) 身近な記述疫学

- Covid-19 cases & deaths on May 18, 2021



# 例) 身近な記述疫学

## · Covid-19 cases & deaths on May 27, 2022



# 例) 記述疫学研究

## 症例研究

2020年1月1~20日に武漢の1病院で  
COVID-19と診断された患者99人の特徴

患者背景因子  
臨床症状  
検査値

Patients (n=99)	
Age, years	
Mean (SD)	55.5 (13.1)
Range	21-82
≤39	10 (10%)
40-49	22 (22%)
50-59	30 (30%)
60-69	22 (22%)
≥70	15 (15%)
Sex	
Female	32 (32%)
Male	67 (68%)
Occupation	
Agricultural worker	2 (2%)
Self-employed	63 (64%)
Employee	15 (15%)
Retired	19 (19%)
Exposure to Huanan seafood market*	49 (49%)
Long-term exposure history	47 (47%)
Short-term exposure history	2 (2%)
Chronic medical illness	50 (51%)
Cardiovascular and cerebrovascular diseases	40 (40%)
Digestive system disease†	11 (11%)
Endocrine system disease†	13 (13%)
Malignant tumour	1 (1%)
Nervous system disease	1 (1%)
Respiratory system disease	1 (1%)
Admission to intensive care unit	23 (23%)
Clinical outcome	
Remained in hospital	57 (58%)
Discharged	31 (31%)
Died	11 (11%)

Data are n (%) unless specified otherwise. 2019-nCoV=2019 novel coronavirus.  
\*Long-term exposure is having worked at or lived in or around Huanan seafood market, whereas short-term exposure is having been to Huanan seafood market occasionally. †12 were diabetic.

**Table 1: Demographics, baseline characteristics, and clinical outcomes of 99 patients admitted to Wuhan Jinyintan Hospital (Jan 1-20, 2020) with 2019-nCoV pneumonia**

Patients (n=99)	
<b>Signs and symptoms at admission</b>	
Fever	82 (83%)
Cough	81 (82%)
Shortness of breath	31 (31%)
Muscle ache	11 (11%)
Confusion	9 (9%)
Headache	8 (8%)
Sore throat	5 (5%)
Rhinorrhoea	4 (4%)
Chest pain	0 (0%)
Dyspnoea	0 (0%)
Nausea	0 (0%)
Mucous sputum	0 (0%)
Conjunctivitis	0 (0%)
Urticaria	0 (0%)
Bleeding	0 (0%)
Multifocal consolidation	0 (0%)
<b>Treatment</b>	
Oxygen therapy	75 (76%)
Mechanical ventilation	
Non-invasive (ie, face mask)	13 (13%)
Invasive	4 (4%)
CRRT	9 (9%)
ECMO	3 (3%)
Antibiotic treatment	70 (71%)
Antifungal treatment	15 (15%)
Antiviral treatment	75 (76%)
Glucocorticoids	19 (19%)
Intravenous immunoglobulin therapy	27 (27%)

2019-nCoV=2019 novel coronavirus. ARDS=acute respiratory distress syndrome. ECMO=extracorporeal membrane oxygenation. CRRT=continuous renal replacement therapy.

**Table 2: Clinical characteristics and treatment of patients with 2019-nCoV pneumonia**

Patients (n=99)	
<b>Blood routine</b>	
Leucocytes (×10 <sup>9</sup> per L; normal range 3.5-9.5)	7.5 (3.6)
Increased	24 (24%)
Decreased	9 (9%)
Neutrophils (×10 <sup>9</sup> per L; normal range 1.8-6.3)	5.0 (3.3-8.1)
Increased	38 (38%)
Lymphocytes (×10 <sup>9</sup> per L; normal range 1.1-3.2)	0.9 (0.5)
Decreased	75 (76%)

Patients (n=99)	
(Continued from previous column)	
<b>Infection-related biomarkers</b>	
Procalcitonin (ng/mL; normal range 0.0-5.0)	0.5 (1.1)
Increased	6 (6%)
Interleukin-6 (pg/mL; normal range 0.0-7.0)	7.9 (6.1-10.6)
Increased	51 (52%)
Erythrocyte sedimentation rate (mm/h; normal range 0-0-15-0)	49.9 (23.4)
Increased	84 (85%)
Serum ferritin (ng/mL; normal range 21.0-274.7)	808.7 (490.7)
Increased	62 (63%)
C-reactive protein (mg/L; normal range 0.0-5.0)*	51.4 (41.8)
Increased	63/73 (86%)
<b>Co-infection</b>	
Other viruses	0
Bacteria	1 (1%)
Fungus	4 (4%)

Data are n (%), n/N (%), mean (SD), and median (IQR). Increased means over the upper limit of the normal range and decreased means below the lower limit of the normal range. 2019-nCoV=2019 novel coronavirus. \*Data available for 73 patients.

**Table 3: Laboratory results of patients with 2019-nCoV pneumonia**

# 例) 記述疫学研究

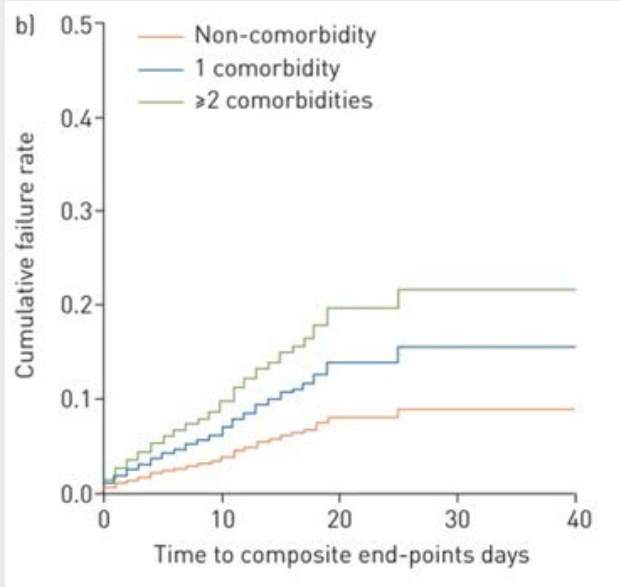
- 2020年1月29日までに中国30都市552病院から集められたCOVID-19患者1,099人の特徴



Characteristic	All Patients (N = 1099)	Disease Severity	
		Nonsevere (N = 926)	Severe (N = 173)
<b>Age</b>			
Median (IQR) — yr	47.0 (35.0–58.0)	45.0 (34.0–57.0)	52.0 (40.0–65.0)
Distribution — no./total no. (%)			
0–14 yr	9/1011 (0.9)	8/848 (0.9)	1/163 (0.6)
15–49 yr	557/1011 (55.1)	490/848 (57.8)	67/163 (41.1)
50–64 yr	292/1011 (28.9)	241/848 (28.4)	51/163 (31.3)
≥65 yr	153/1011 (15.1)	109/848 (12.9)	44/163 (27.0)
Female sex — no./total no. (%)	459/1096 (41.9)	386/923 (41.8)	73/173 (42.2)
<b>Smoking history — no./total no. (%)</b>			
Never smoked	927/1085 (85.4)	793/913 (86.9)	134/172 (77.9)
Former smoker	21/1085 (1.9)	12/913 (1.3)	9/172 (5.2)
Current smoker	137/1085 (12.6)	108/913 (11.8)	29/172 (16.9)
<b>Coexisting disorder — no. (%)</b>			
Any	261 (23.7)	194 (21.0)	67 (38.7)
Chronic obstructive pulmonary disease	12 (1.1)	6 (0.6)	6 (3.5)
Diabetes	81 (7.4)	53 (5.7)	28 (16.2)
Hypertension	165 (15.0)	124 (13.4)	41 (23.7)
Coronary heart disease	27 (2.5)	17 (1.8)	10 (5.8)
Cerebrovascular disease	15 (1.4)	11 (1.2)	4 (2.3)
Hepatitis B infection¶	23 (2.1)	22 (2.4)	1 (0.6)
Cancer	10 (0.9)	7 (0.8)	3 (1.7)
Chronic renal disease	8 (0.7)	5 (0.5)	3 (1.7)
Immunodeficiency	2 (0.2)	2 (0.2)	0

# 例) 記述疫学研究

- 2019年12月11日~2020年1月31日に中国31都市575病院のいずれかに入院した1,590人のCovid19感染患者: 併存疾患の影響



Features	HR (95% CI)	p-value
Type of comorbidities		
COPD	2.681 (1.424–5.058)	0.002
Diabetes	1.586 (1.028–2.449)	0.037
Hypertension	1.575 (1.069–2.322)	0.022
Malignant tumour	3.501 (1.604–7.643)	0.002
Comorbidities n		
1	1.789 (1.155–2.772)	0.009
≥2	2.592 (1.611–4.171)	<0.001

# 例) 関連研究

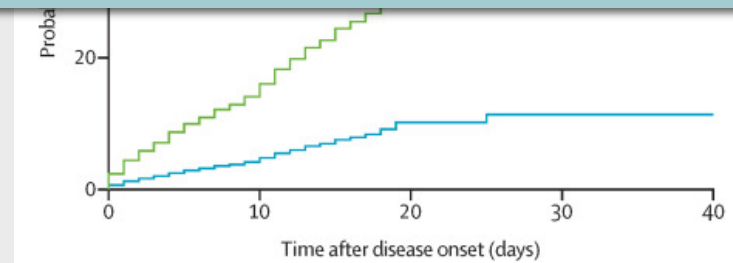
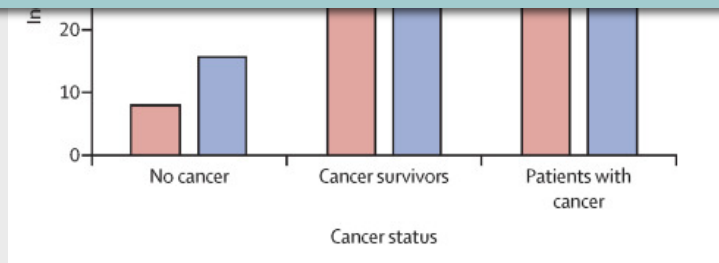
Covid 19感染者1,590人のうち18人ががん患者

- 一般集団よりがん発生率が高い
- 重症イベント(死亡やICU入室)のリスクが高い

## この結論を否定

12名はがん治療からとっくに回復している  
年齢層もnon-cancer患者より高い

むしろがん患者が医療サービスが受けられない問題に注目





# 例) 記述疫学研究

- 2020年3月1日~4月4日にCOVID19感染でNYCの12病院のいずれかに入院した5,700人の特徴

Table 1. Baseline Characteristics of Patients Hospitalized With COVID-19

	No. (%)
Demographic information	
Total No.	5644
Age, median (IQR) [range], y	47 (37-57) [18-90]
Sex	
Female	2700 (47.8)
Male	2944 (52.2)
Race <sup>a</sup>	
No.	% Median (IQR)
African American	1162 (20.6) 47 (37-57)
Asian	95 (1.7) 47 (37-57)
White	3500 (61.7) 47 (37-57)
Other/multiracial	187 (3.3) 47 (37-57)

Table 2. Presentation Vitals and Laboratory Results of Patients Hospitalized With COVID-19

	No. (%)	No.	Reference ranges
Triage vitals <sup>a</sup>			
Temperature >38 °C	1734 (30.7)	5644	
Temperature, median (IQR), °C	37.5 (36.9-38.3)		
Oxygen saturation			
<90%	1162 (20.6)		
% Median (IQR)	95 (91-99)		
Received supplemental oxygen at triage			
Respiratory rate >24 breaths/min	1584 (28.1)		
Heart rate			
≥100 beats/min	2457 (43.5)		
Median (IQR)	97 (85-110)		
Initial laboratory measures, median (IQR) <sup>b</sup>			
White blood cell count, ×10 <sup>9</sup> /L	7.0 (5.2-10.0)		

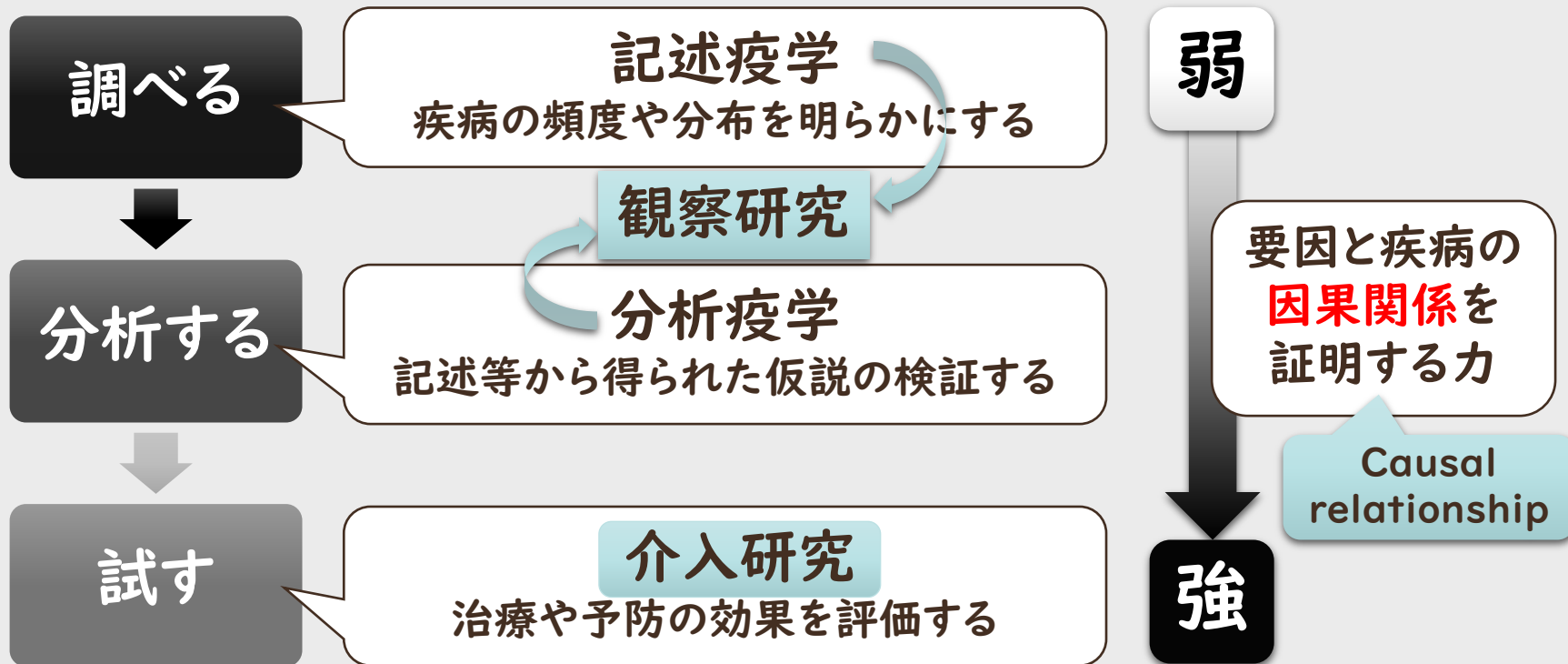
Table 4. Discharge Disposition by 10-Year Age Intervals of Patients Hospitalized With COVID-19

Age intervals, y	Patients discharged alive or dead at study end point		Length of stay among those who died, median (IQR), d <sup>a</sup>	Discharged, No./No. (%)
	Died, No./No. (%)			
	Male	Female		
0-9	0/13	0/13	NA	13/13 (100)
10-19	0/1	0/7	NA	1/1 (100)
20-29	3/42 (7.1)	1/55 (1.8)	4.0 (0.8-7.4)	39/42 (92.9)
30-39	6/120 (5.0)	7/81 (8.6)	7.8 (7.4-2.6)	117/112 (104.5)

# 疫学研究とは？

・健康に関連する状態や事象の集団中の分布や決定要因を研究しかつその研究成果を健康問題の予防やコントロールのために適用する学問

(Last JM, A dictionary of epidemiology 2001)

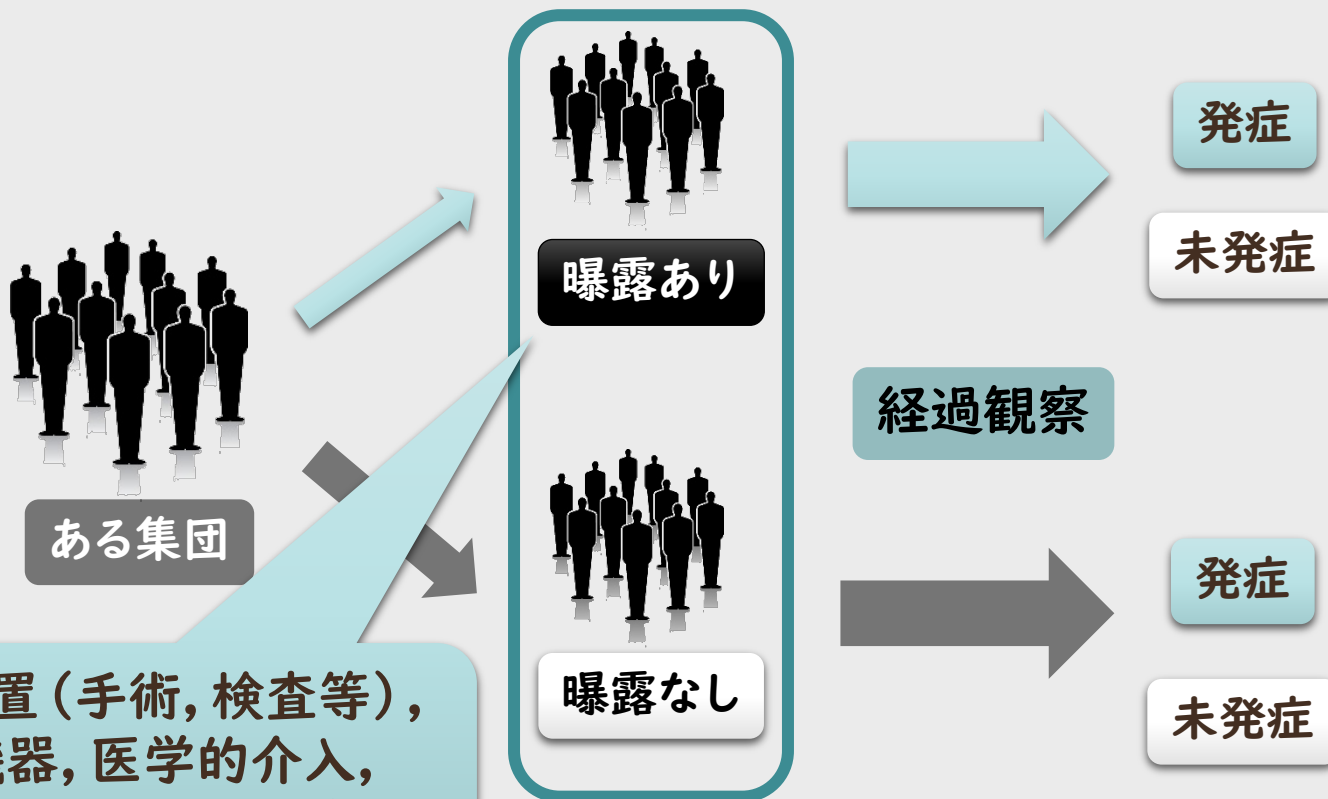


# 分析疫学

- ・ 要因と疾病の関連を検討
  - ・ 時間の順序を考慮
    - ・ 要因の曝露時点は疾患発症よりも前
  - ・ 仮説の検証
    - ・ 記述疫学研究などで検討した仮説
  - ・ 因果関係の強い結論付けは難しい
    - ・ バイアスの制御が難しい
    - ・ 記述疫学よりは検討できる
- ・ 研究の種類は主に2つ
  - ・ コホート研究
  - ・ 症例対照研究

# コホート研究

- ・人々を要因の有無で分け疾患の発症を一定期間観察する
- ・曝露とイベント（疾患発生）の関連を検討



薬剤, 処置 (手術, 検査等),  
医療機器, 医学的介入,  
これらの組合せなど

# バイアスとの闘い

## 偶然誤差 (random error)

- ・ 全くの偶然で生じる
- ・ サンプルサイズを大きくすると小さくなる

## 系統誤差 (systematic error / bias)

- ・ 疾患リスクに対する曝露の効果について誤った推測の原因となる
- ・ 特定の要因が影響してある方向に偏って生じる
  - ・ 選択バイアス
  - ・ 情報バイアス
  - ・ 交絡

# 妥当性

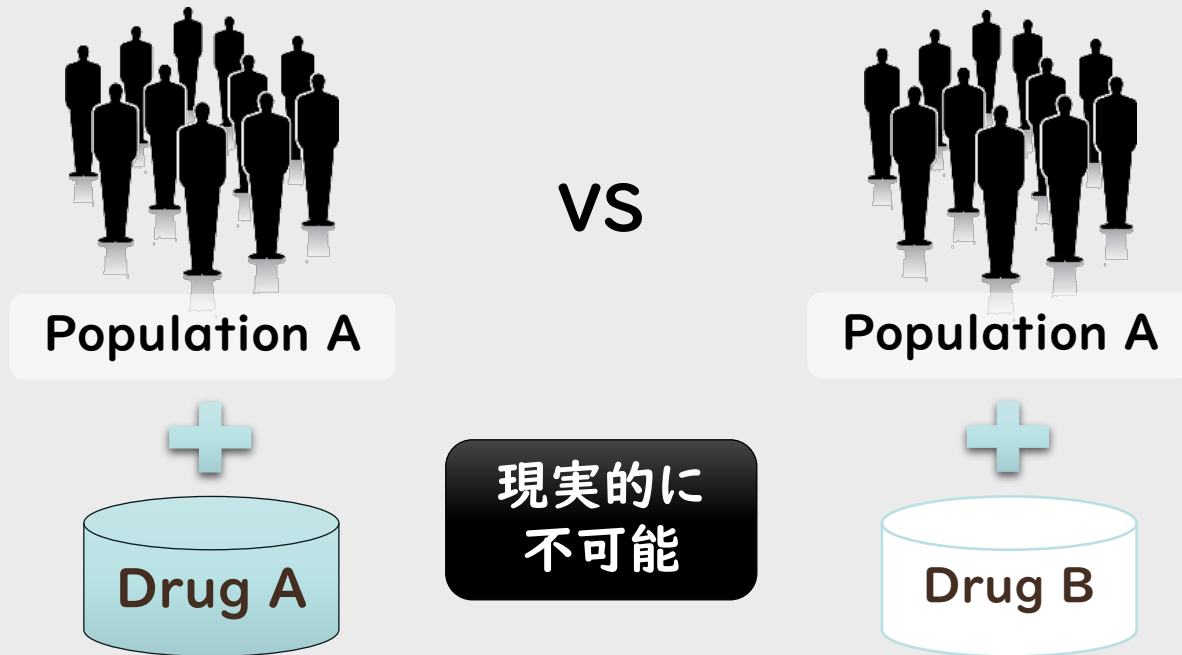
- ・ 内部妥当性 (internal validity)
  - ・ 研究対象集団による検討で導かれた結論は研究条件や結果も併せて合理的か？
- ・ 外部妥当性 (External validity)
  - ・ 研究対象集団から得られた結果を研究対象集団以上の集団に一般化できるか？
  - ・ 一般化可能性 (Generalizability)

まずは内部妥当性が大事

# 適切な比較対照群の設定

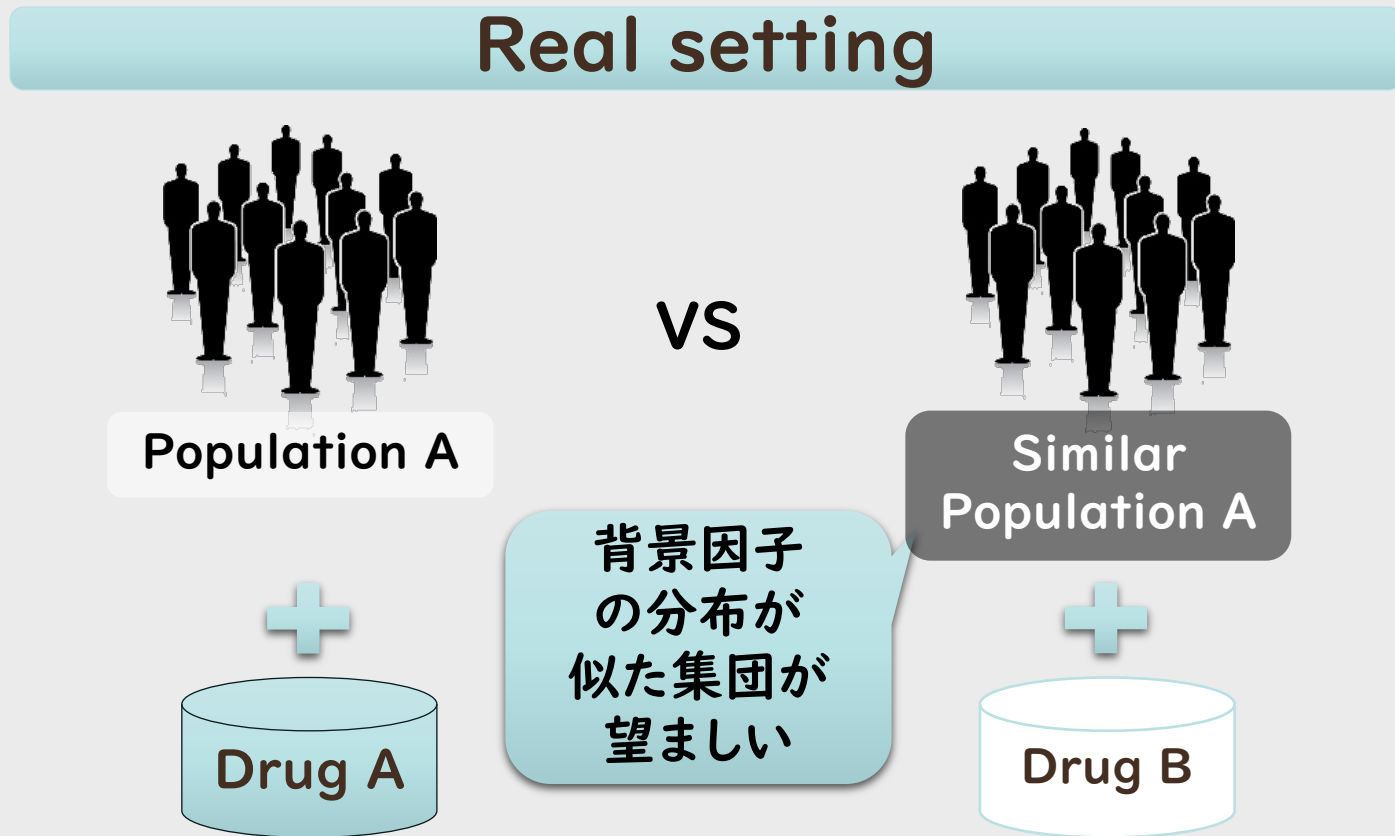
- ・ A薬とB薬の治療効果の比較

Ideal setting



# 適切な比較対照群の設定

- ・ A薬とB薬の治療効果の比較





# 対照集団の設定

- ・曝露あり集団と同じような集団を探す
  - ・目的のアウトカムに対する交絡を最小にするような集団
    - ・背景因子：年齢, 性別
    - ・同様の疾患に罹患
    - ・同様の重症度
    - ・同じ薬剤クラスの薬剤を使用
    - ・同じ治療目的の薬剤を使用
    - ・同じような有害事象を起こす薬剤を使用
    - ・同じ治療目的の介入（手術, 検査等）を受ける

無治療の患者や  
服用薬なしの患者は  
いい対照群にならないことが多い

# 例) コホート研究

- ヒドロキシクロロキンの使用と呼吸不全の関連を検討@米
- 北部マンハッタンにある急性期病院のclinical data warehouse
- 2020年3月7日～4月8日にSARS-Cov2陽性で入院した1,376人が対象
- ベースライン時から挿管または死亡までの時間を当該薬使用の有無で評価
- 主解析, 感度解析共にヒドロクロロキンと死亡・挿管に有意な関連無し

Analysis	Intubation or Death
No. of events/no. of patients at risk (%)	
Hydroxychloroquine	262/811 (32.3)
No hydroxychloroquine	84/565 (14.9)
Crude analysis — hazard ratio (95% CI)	2.37 (1.84–3.02)
Multivariable analysis — hazard ratio (95% CI)*	1.00 (0.76–1.32)
Propensity-score analyses — hazard ratio (95% CI)	
With inverse probability weighting†	1.04 (0.82–1.32)
With matching‡	0.98 (0.73–1.31)
Adjusted for propensity score§	0.97 (0.74–1.28)

# 関連研究

- ヒドロキシクロロキンの非盲検ランダム化比較試験@英国
- COVID19入院した患者をヒドロキシクロロキン1,561人, 通常治療3,155人に割付け, 28日死亡率を比較
- 両群に差は見られなかった

**Table 2.** Primary and Secondary Outcomes.

Outcome	Hydroxychloroquine (N = 1561)	Usual Care (N = 3155)	Rate or Risk Ratio (95% CI)
	<i>no./total no. (%)</i>		
Primary outcome: 28-day mortality	421/1561 (27.0)	790/3155 (25.0)	1.09 (0.97–1.23)*
Secondary outcomes			
Discharge from hospital in ≤28 days	931/1561 (59.6)	1983/3155 (62.9)	0.90 (0.83–0.98)*
Invasive mechanical ventilation or death†	399/1300 (30.7)	705/2623 (26.9)	1.14 (1.03–1.27)‡
Invasive mechanical ventilation	128/1300 (9.8)	225/2623 (8.6)	1.15 (0.93–1.41)
Death	311/1300 (23.9)	574/2623 (21.9)	1.09 (0.97–1.23)

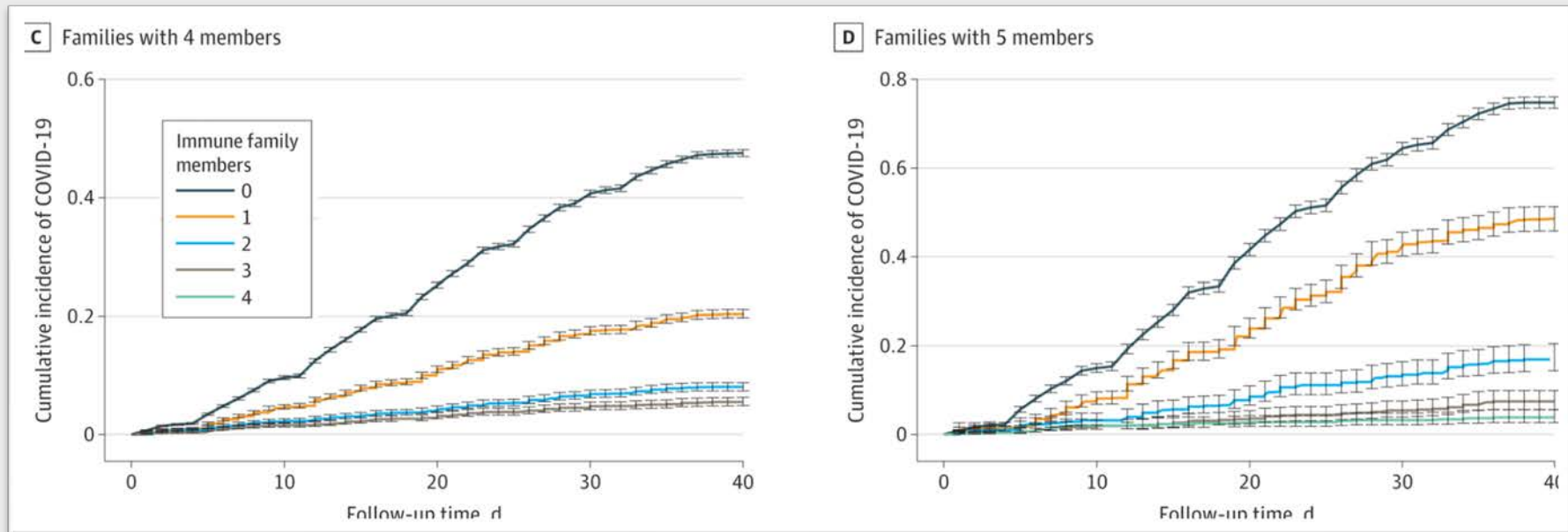
\* The between-group difference was calculated as a rate ratio.

† Patients who were receiving invasive mechanical ventilation at randomization were excluded from this analysis.

‡ The between-group difference was calculated as a risk ratio.

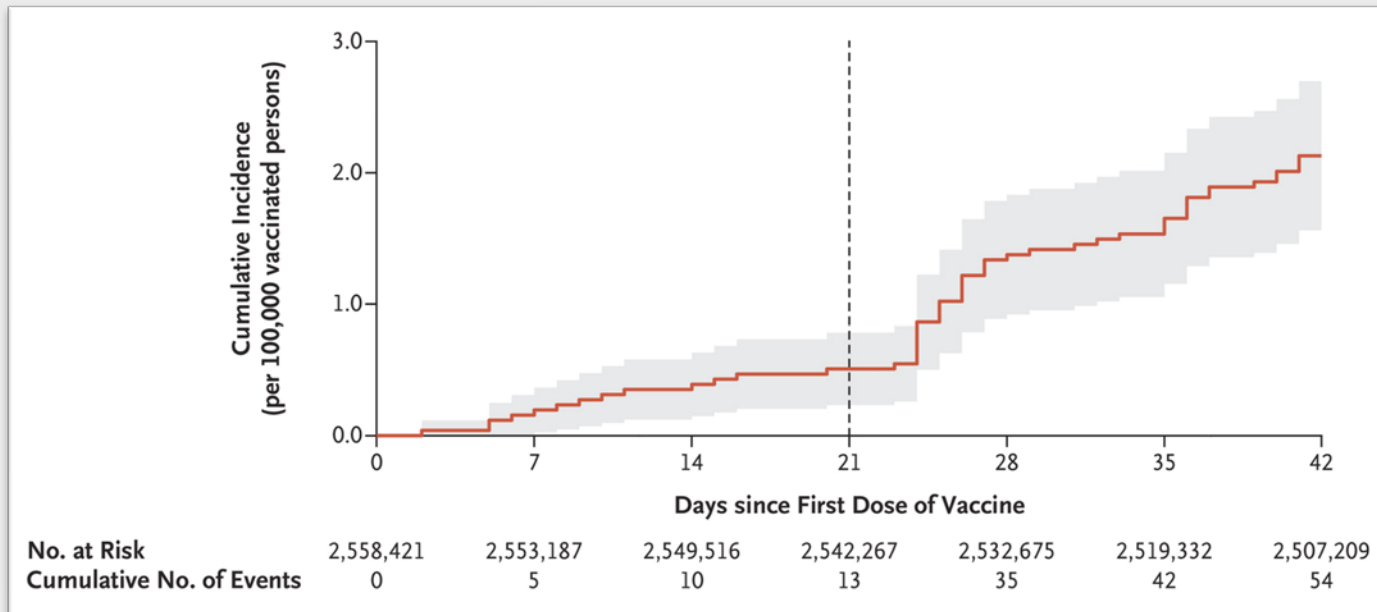
# 例) コホート研究

- 家族内でのCOVID-19免疫と非免疫家族の感染リスクとの関連@スウェーデン
- スウェーデンの全国登録から得られたデータ
- 2021年5月26日までにCOVID-19感染かワクチン接種で免疫を獲得したすべての人とその家族1,789,728人(約81万家族)
- 免疫獲得家族の数が増加するに伴い非免疫家族の感染リスクは減少



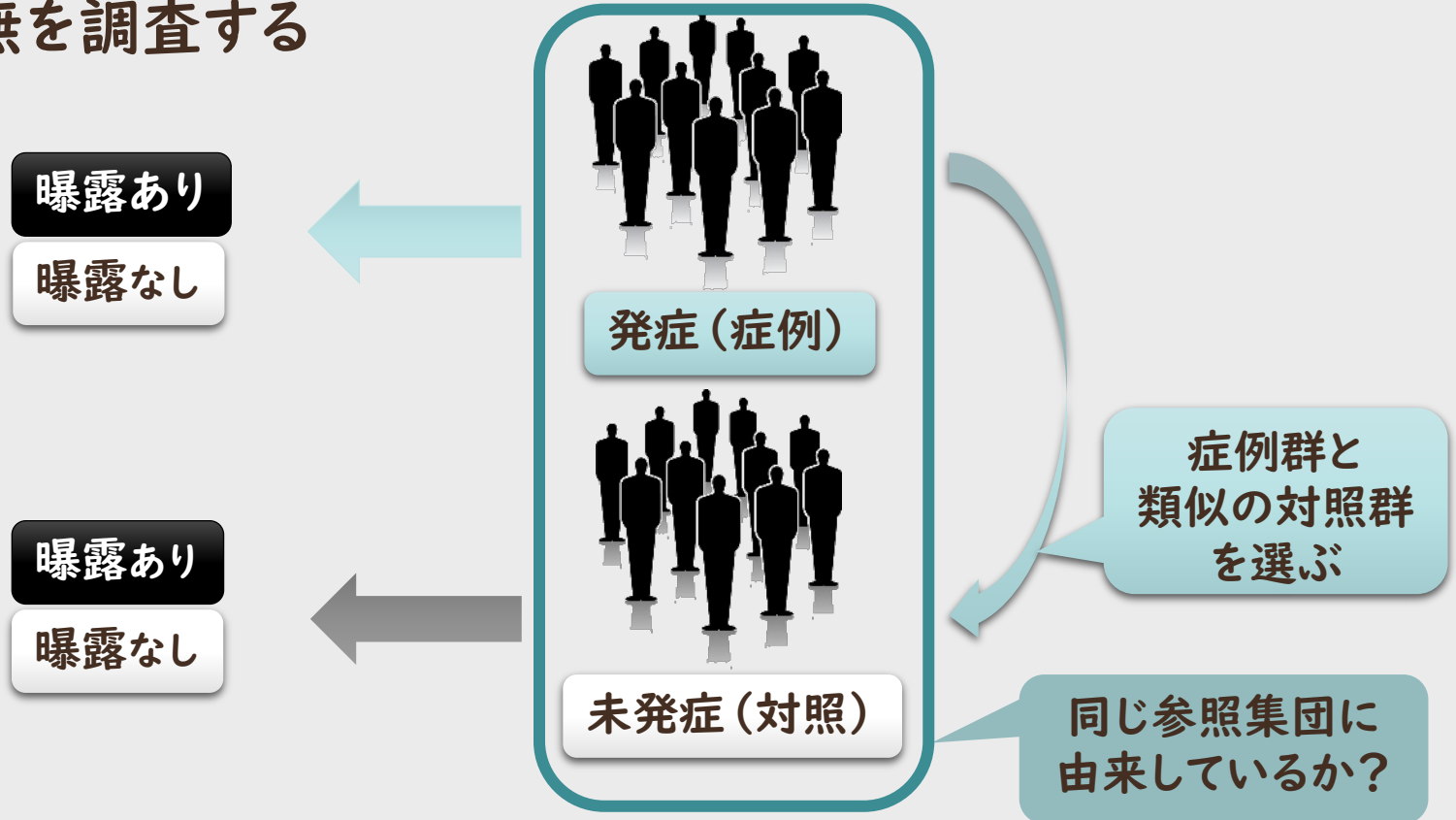
# 例) コホート研究

- mRNAワクチンの初回接種から42日後までの心筋炎の発症率の調査@イスラエル
- 国内最大の医療機関であるClalit Health Services社のデータベース
- mRNAワクチンを接種した250万人以上の16歳以上の保険加入者が対象
- 心筋炎の推定発生率は2.13例(10万人当たり)
- 16~29歳の男性が最も高い10.69例



# 症例対照研究

- ある疾患を発症した人と発症していない人の曝露の有無を調査する



# コントロール選択の原則

- ・ケースと同じ疾患を発症するリスクがあり  
かつ その時点ではその疾患を発症していない  
かつ 発症していないこと以外はケースと同じ人々
  - ・同じ疾患を発症したらケースに含まれる人々
  - ・ケースが選択された同じ参照集団から選ばれた人々

研究者が最も頭を悩ませる  
サンプリングバイアスが発生するかも  
結果に重大な影響を及ぼす

ケースやコントロールが何らかの理由で  
ケースやコントロールにならない場合  
例) 死亡や回復, 経済格差による受診行動の違い

# コントロールの選び方

## 院内コントロール

- ・ 例) 長大病院の患者からケースとコントロールを選ぶ
- ・ ある疾患で治療中の患者は研究対象の疾患を発症したら長大病院を受診するだろう

## ポピュレーションベースのサンプリング

- ・ 例) 地域行政や医療保険組織による疾患登録システムからケースとコントロールを選ぶ
- ・ ケースは地域住民の対象疾患患者の母集団を代表すると考えられるのでコントロールはその地域住民から選択すればよい

## 複数のコントロール群を設定

- ・ 例) 同じ救命救急科の治療を受けた患者, 同じ病院の入院患者, 同じ所属組織, 地域の電話帳から選択

## マッチングで選択



# コントロールの選び方

## ・マッチング

- 症例に対して交絡になりうる因子（性別や年齢等）が一致する対照を選ぶ

症例群 (ケース群)



対照群 (コントロール群)



マッチング因子で交絡することはない  
選択バイアスが生じる可能性がある

曝露と相関がある因子でマッチング  
対照群と曝露を独立に選択する原則を無視

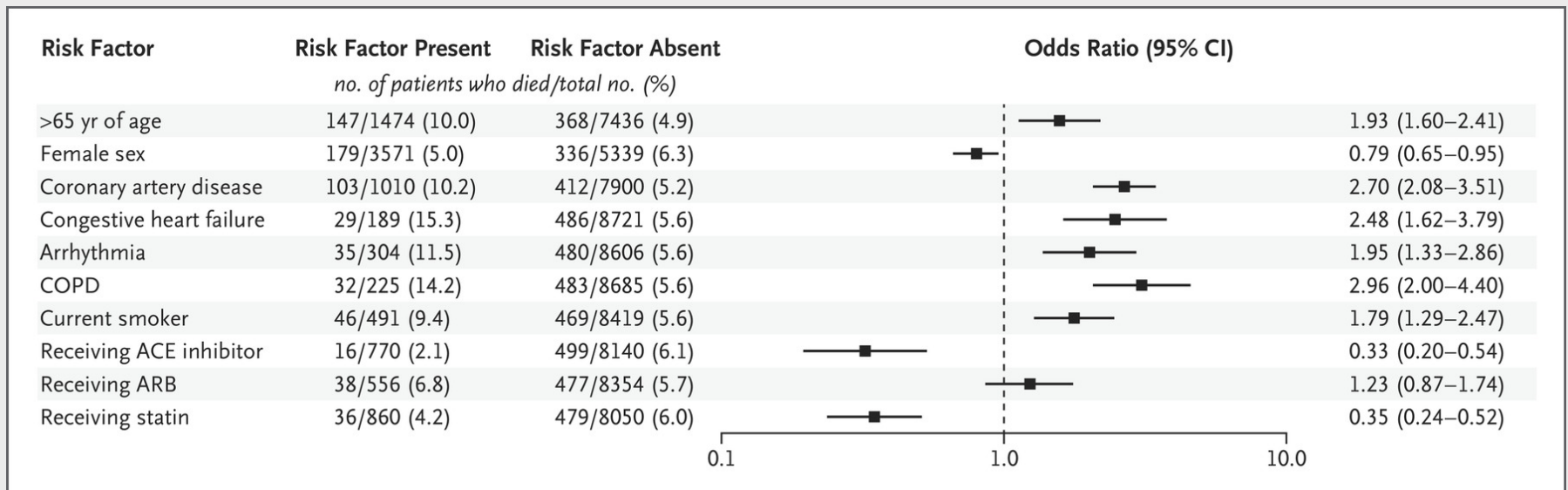
# 例) 症例対照研究

- ARB/ACEとSARS-CoV2感染の関連を調べる@イタリア
- イタリアロンバルディア州の国民健康保険のデータとCovid19感染情報
- 2020年2月21日から3月11日にCovid19感染記録のあった国民健康保険加入の40歳以上の患者6,272人とそのコントロール30,759人が対象
- ケース:SARS-CoV2感染者
- コントロール:性別, Index dateの年齢, 居住エリアでマッチした最大5名
- 曝露:Index date前の5年間(2019年中処方, 3回連続処方, 最終四半期1回)

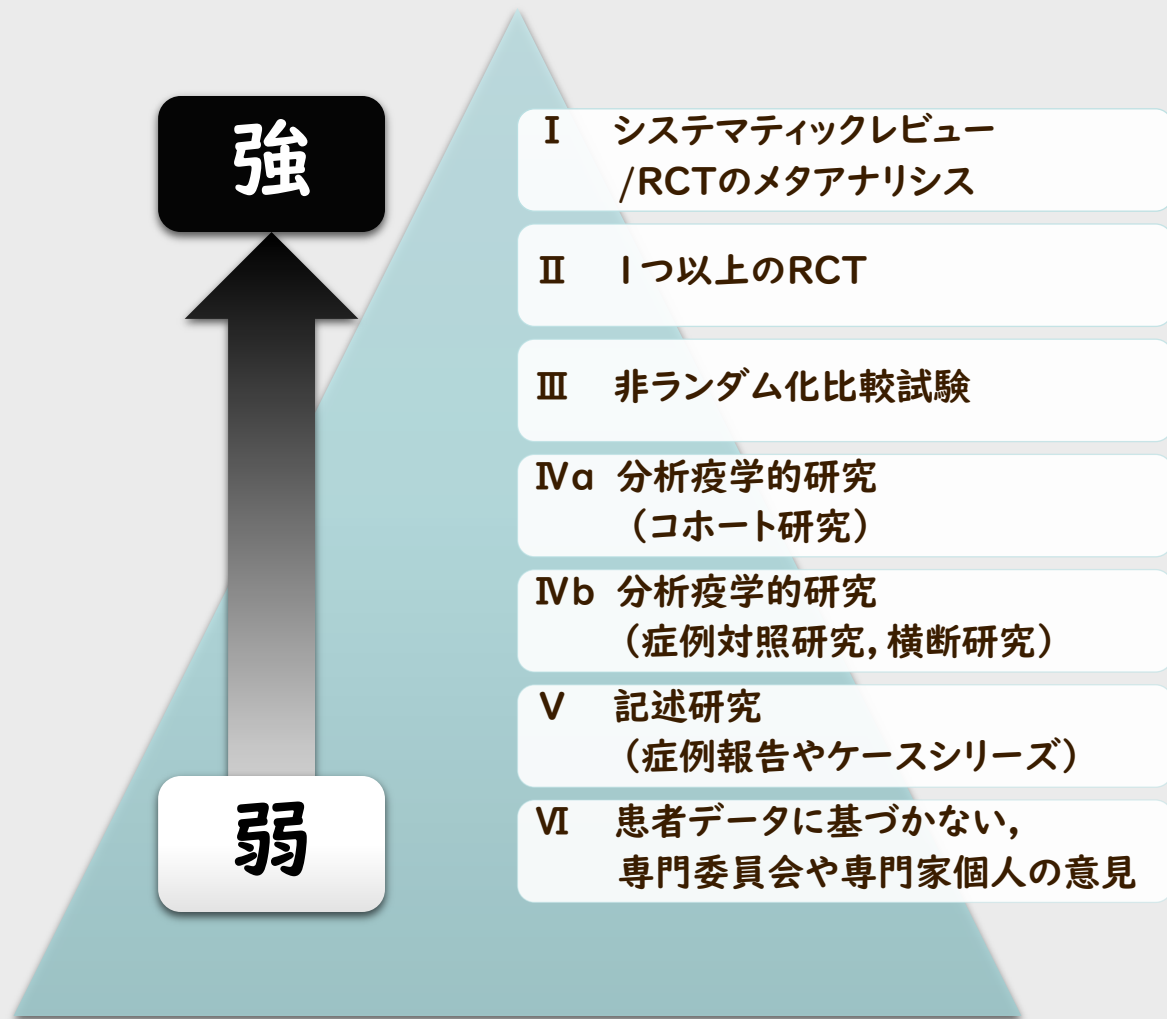
Variable	Odds Ratio for Covid-19 (95% CI) <sup>†</sup>	
	Unadjusted	Adjusted
<b>Drugs<sup>‡</sup></b>		
Antihypertensive drugs overall	1.53 (1.43–1.63)	
ACE inhibitors	1.16 (1.08–1.24)	0.96 (0.87–1.07)
ARBs	1.20 (1.12–1.29)	0.95 (0.86–1.05)
Calcium-channel blockers	1.28 (1.18–1.38)	1.03 (0.95–1.12)
Beta-blockers	1.42 (1.33–1.51)	0.99 (0.91–1.08)
Diuretics as a whole	1.69 (1.57–1.83)	
Thiazide or thiazide-like diuretics	1.09 (1.01–1.17)	1.03 (0.86–1.23)
Loop diuretics	2.01 (1.83–2.20)	1.46 (1.23–1.73)
Mineralocorticoid-receptor antagonists	1.59 (1.37–1.85)	0.90 (0.75–1.07)

# 例) ARB関連研究

- Covid19入院患者の心血管疾患および薬剤治療と院内死亡の関連を評価@アジア, 北米, 欧州11カ国169病院
- Surgical Outcomes Collaborative registryのデータ
- 2019年12月20日から2020年3月15日にCovid19感染で入院した記録のあった8,910人が対象
- 入院中死亡者と生存退院者で既往や使用薬剤を評価
- ACE 阻害薬オッズ比 0.33



# 研究デザインとエビデンスの強さ



# メタアナリシス（メタ分析）

- ・ システマティックレビューの過程で実施
- ・ 複数の研究結果（オッズ比等の疫学的指標）を定量的に統合
  - 個々ではサンプルサイズが小さい研究でも複数の研究結果を統合することで精度が上がる

Effects of remdesivir in patients hospitalised with COVID-19:

**A systematic review and individual patient data meta-analysis of randomised controlled trials**

Global Diabetes Prevalence in COVID-19 Patients and Contribution to COVID-19- Related Severity and Mortality:

**A Systematic Review and Meta-analysis**

The Impact of the COVID-19 Pandemic and Associated Control Measures on the Mental Health of the General Population :

**A Systematic Review and Dose-Response Meta-analysis**