

都道府県別コロナ関連データセット

2022年5月9日

仲田泰祐・岡本亘（東京大学）

重要ポイント

- 日本における新型コロナウイルス感染症に関する定量分析や定量的なディスカッションを行う際に、データの収集・加工は大きなハードルとなっている。
 - 理由①：フォーマット（PDF、EXCEL等のファイル形式やデータ構造）が統一されていない。
 - 理由②：同じ名前のデータでも定義が異なる場合がある。
例：重症患者数（国基準）、重症患者数（旧都基準・新都基準）、重症患者数（大阪府基準）
 - 理由③：データを公表する機関・タイミングがバラバラである。
例：ワクチンデータを収集する場合
 - ワクチン（一般接種）→デジタル庁
 - ワクチン（医療従事者、2021/04/09以前のデータ）→厚生労働省
 - ワクチン（医療従事者、2021/04/12以降のデータ）→首相官邸

分析

- そこで、新型コロナウイルス感染症に関連するデータを自動で収集し一元化・都道府県（+全国）パネルデータとして無償で提供。
 - 日次データと、日次データを加工した週次データを同時に提供。
 - 収集データは人口, 感染状況, 医療提供体制, 行動制限有無, ワクチン, 人流, 気象, etc. など多岐にわたる。
 - 都道府県別にこれだけの情報量を盛り込んだデータセットはこれまでに発表されていない。
- 本データセットは因果推論等のデータ分析やモデル分析に自由に活用可能。
 - データはサーバー上で毎日（※一部データは毎週）自動的に更新。
 - 本データセットを利用することにより、分析のために必要なデータ収集・加工等の煩わしい作業を、大幅に短縮できる。

利用方法・利用規定

利用方法 ①

- https://docs.google.com/spreadsheets/d/1OOwRFo5sh_kaDQF79BdpAHhI_WXXcXpV5tj4NXYQBHk/edit?usp=sharing にブラウザでアクセスします。
- 下図のように【ファイル】>【ダウンロード】>【Microsoft Excel】と順にクリックすることで、エクセルデータとしてダウンロードできます。

The screenshot shows a Google Sheet interface for a 'COVID-19 Standard Dataset'. The 'File' menu is open, and the 'Download' option is selected, which has opened a sub-menu where 'Microsoft Excel (.xlsx)' is highlighted. The spreadsheet data is visible in the background, with columns for dates, cumulative counts (cumN), tested cases, ICU cases, and deaths (dD).

	E	F	G	H	I	J	K	L	M	N
	N	cumN	tested	I	ICU	D	dD	R	V_o1	
2020-01-26	0	4	0	0	0	0	0	0	0	0
2020-01-27	0	4	1	0	0	0	0	0	0	0
2020-01-28	3	7	1	0	0	0	0	0	0	0
2020-01-29	1	8	0	0	0	0	0	0	0	0
2020-01-30	4	12	0	0	0	0	0	0	0	0
2020-01-31	12	12	1	0	0	0	0	0	0	0
2020-02-01	12	12	5	0	0	0	0	0	0	0
2020-02-02	12	12	0	0	0	0	0	0	0	0
2020-02-03	12	12	0	0	0	0	0	0	0	0
2020-02-04	13	13	0	0	0	0	0	0	0	0
2020-02-05	15	15	0	0	0	0	0	0	0	0
2020-02-06	15	15	1	0	0	0	0	0	0	0
2020-02-07	15	15	0	0	0	0	0	0	0	0
2020-02-08	15	15	0	0	0	0	0	0	0	0
2020-02-09	15	15	0	0	0	0	0	0	0	0
2020-02-10	0	15	5	0	0	0	0	0	0	0
2020-02-11	2	17	1	0	0	0	0	0	0	0
2020-02-12	0	17	0	0	0	0	0	0	0	0
2020-02-13	4	21	0	0	0	0	0	0	0	0
2020-02-14	7	28	3	0	0	0	0	0	1	0
2020-02-15	12	40	15	0	0	0	0	0	1	0
2020-02-16	6	46	77	0	0	0	0	0	1	0
2020-02-17	6	52	13	0	0	0	0	0	1	0
2020-02-18	8	60	32	0	0	0	0	0	1	0
2020-02-19	9	69	94	0	0	0	0	0	1	0
2020-02-20	8	77	90	0	0	0	0	0	1	0
2020-02-21	13	90	49	0	0	0	0	0	1	0
2020-02-22	24	114	50	0	0	0	0	0	1	0

利用方法 ②

- 本データセットを csv としてエクスポートするための直リンクを指定できます。
- 日次データ：
https://docs.google.com/spreadsheets/d/1OOwRFo5sh_kaDQF79BdpAHhI_WXXcXpV5tj4NXYQBHk/export?format=csv&gid=1016364865
- 週次データ：
https://docs.google.com/spreadsheets/d/1OOwRFo5sh_kaDQF79BdpAHhI_WXXcXpV5tj4NXYQBHk/export?format=csv&gid=1350668593

例1. Python の pandas ライブラリでデータを読み込むことができます。

```
df = pd.read_csv("https://docs.google.com/spreadsheets/d/1OOwRFo5sh_kaDQF79BdpAHhI_WXXcXpV5tj4NXYQBHk/export?format=csv&gid=1350668593")
```

例2. 別の Google スプレッドシート で以下の関数を用いることで、データを呼び出すことができます。

```
= IMPORTDATA("https://docs.google.com/spreadsheets/d/1OOwRFo5sh_kaDQF79BdpAHhI_WXXcXpV5tj4NXYQBHk/export?format=csv&gid=1350668593")
```

例3. curl コマンドや wget コマンドによりデータを落とすこともできます。

```
curl -L "https://docs.google.com/spreadsheets/d/1OOwRFo5sh_kaDQF79BdpAHhI_WXXcXpV5tj4NXYQBHk/export?format=csv&gid=1016364865" > data.csv
```

利用規定

- 本データセットの使用ライセンスは MIT License に準拠します。
 - 本データセットは無償で利用・加工・再配布が可能です。
 - コンテンツを利用する際は必ず出典を明記してください。記載例を以下に示します。
出典：仲田泰祐・岡本亘（2022）「COVID-19 Standard Dataset」
https://docs.google.com/spreadsheets/d/1OOwRFo5sh_kaDQF79BdpAHhl_WXXcXpV5tj4NXYQBHk/edit?usp=sharing
- 次頁に掲載する免責事項を事前に読み、理解した上でのご利用をお願いいたします。

免責事項

- 本データセットに掲載されている情報の正確性については万全を期していますが、我々は利用者が本データセットの情報をを用いて行う一切の行為について責任を負うものではありません。
- また、利用者が本データセットを利用したことにより発生した利用者の損害及び利用者が第三者に与えた損害に対して、責任を負うものではありません。
- 本データセットに掲載されている情報は、予告なしに変更又は削除することがあります。
- 本データセットは、東京大学の仲田泰祐・岡本亘が研究者個人として開発しているものです。東京大学は本データセットに関する一切の責任を負いません。

データの一覧・出典

変数名	説明	単位	週次データへの換算方法	出典
prefecture	都道府県名（ローマ字）	-	-	-
都道府県	都道府県名（漢字）	-	-	-
population	全人口	人	-	総務省「令和2年国勢調査」
population_e	高齢者（65歳以上）人口	人	-	
date	日付	-	-	-
N	新規陽性者数	人	該当日から6日前までの平均値	厚生省「各都道府県の検査陽性者の状況（空港・海港検疫、チャーター便案件を除く国内事例）」
cumN	累積陽性者数		該当日の値	
tested	検査実施人数		該当日から6日前までの平均値	
I	入院治療等を要		該当日の値	
ICU	重症患者数（自		該当日の値	
D	死亡者数		該当日の値	
dD	新規死亡者数		該当日から6日前までの平均値	
R	退院者数	該当日の値		
V_o1	高齢者以外（65歳未満）の1回目ワクチン累積接種回数	回	該当日の値	デジタル庁「都道府県別接種回数詳細」 首相官邸「これまでのワクチン総接種回数（医療従事者等、日別）」 首相官邸「これまでのワクチン総接種回数（医療従事者等、都道府県別）」
V_o2	高齢者以外（65歳未満）の2回目ワクチン累積接種回数			
V_o3	高齢者以外（65歳未満）の3回目ワクチン累積接種回数			
V_e1	高齢者（65歳以上）の1回目ワクチン累積接種回数			
V_e2	高齢者（65歳以上）の2回目ワクチン累積接種回数			
V_e3	高齢者（65歳以上）の3回目ワクチン累積接種回数			
V_u1	年齢区分不詳者の1回目ワクチン累積接種回数			
V_u2	年齢区分不詳者の2回目ワクチン累積接種回数			
V_u3	年齢区分不詳者の3回目ワクチン累積接種回数			
V_m1	医療従事者の1回目ワクチン累積接種回数			
V_m2	医療従事者の2回目ワクチン累積接種回数			
V_l1	全体の1回目ワクチン累積接種回数			
V_l2	全体の2回目ワクチン累積接種回数			
V_l3	全体の3回目ワクチン累積接種回数			
V_l1_rate	全体の1回目ワクチン接種率	%	該当日の値	
V_l2_rate	全体の2回目ワクチン接種率			
V_l3_rate	全体の3回目ワクチン接種率			
V_e1_rate	高齢者（65歳以上）の1回目ワクチン接種率			
V_e2_rate	高齢者（65歳以上）の2回目ワクチン接種率			
V_e3_rate	高齢者（65歳以上）の3回目ワクチン接種率			

感染状況

ワクチン

H	入院患者数	人	該当日の値	厚労省「療養状況等及び入院患者受入病床数等に関する調査」
H_bed	確保入院病床数	床		
ICU_national	重症患者数（国基準）	人		
ICU_national_bed	確保重症病床数（国基準）	床		
H_occupancy_rate	入院病床使用率	%		
ICU_national_occupancy_rate	重症病床使用率	%		
retail_and_recreation_percent_change_from_baseline	Google Mobility（小売、娯楽）	%	該当日から6日前までの 平均値	Google「COVID-19: コミュニティ モビリティ レポート」
grocery_and_pharmacy_percent_change_from_baseline	Google Mobility（食料品店、薬局）			
parks_percent_change_from_baseline	Google Mobility（公園）			
transit_stations_percent_change_from_baseline	Google Mobility（乗換駅）			
workplaces_percent_change_from_baseline	Google Mobility（職場）			
residential_percent_change_from_baseline	Google Mobility（住宅）			
rainfall	降水量	mm	該当日から6日前までの 平均値	気象庁「過去の気象データ検索」【地域気象観測システム（アメダス）】
temperature	平均気温	°C		
humidity	相対湿度	%		
soe	緊急事態宣言ダミー	-	該当日の値	各自治体の報道発表資料等を参照し、手入力により作成しました。
manbo	まん延防止等重点措置ダミー	-		

医療提供体制

人流

気象

行動制限

使用例

データ可視化：COVID-19 Standard Dashboard

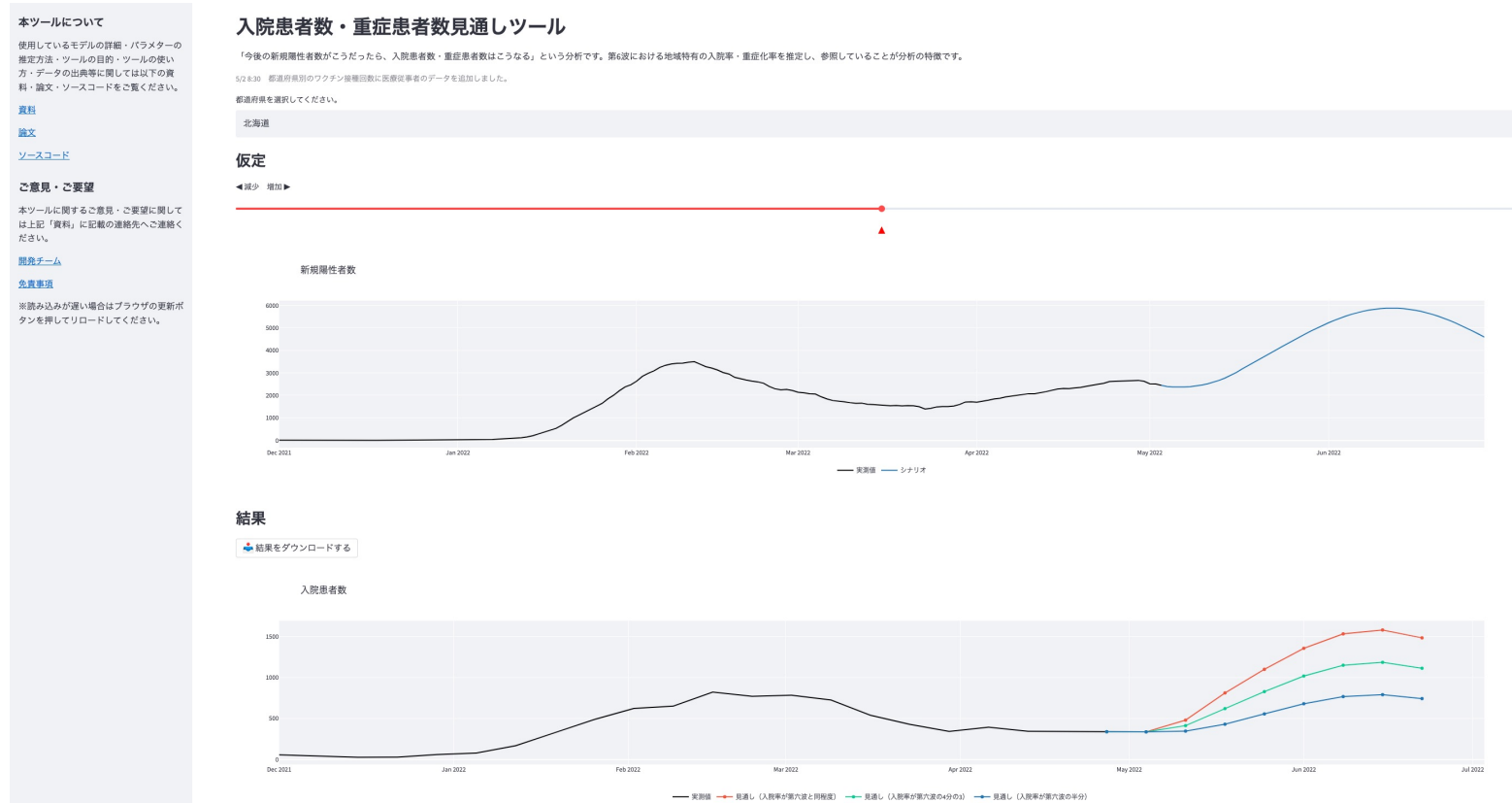
- 本データセットを可視化するダッシュボードを構築。
- 今後、ダッシュボード上で人流と経済活動に関するレポートを提供する予定。



<https://datastudio.google.com/s/iFQSL0z0In4>

モデル分析：入院患者数・重症患者数見通しツール

- 本データセットを利用して、入院患者数・重症患者数見通しツール・レポートを提供。
- 「47都道府県における病床見通し：レポートとツールの解説」（仲田泰祐・岡本亘）
- https://covid19outputjapan.github.io/JP/files/NakataOkamoto_Briefing_20220413.pdf



<https://covid19-icu-tool.herokuapp.com/>

データ分析：まん延防止等重点措置の効果に関する定量分析の例

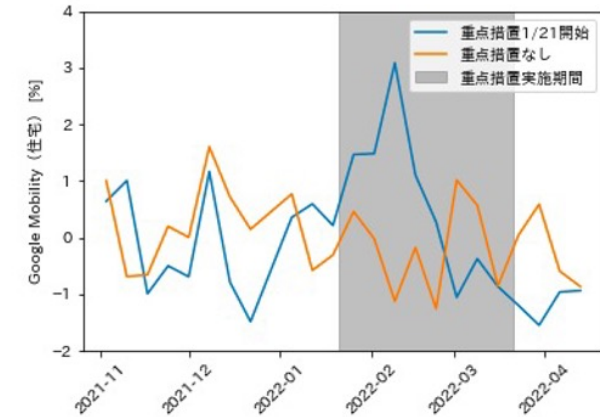
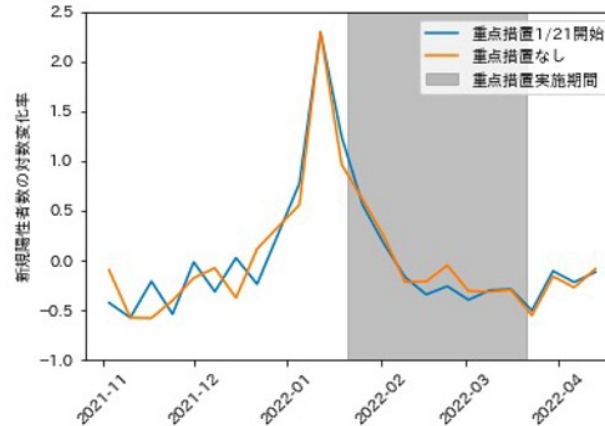
Maeda, Nakata & Okamoto (2022)

■ 手法

Kitamura (2021) の手法を簡略化し、目的変数を人口比新規陽性者数から新規陽性者数の対数変化率(週次)に変更、説明変数として都道府県固定効果、重点措置ダミー、重症病床使用率等を追加したラグ付き固定効果モデルを推定。第六波において東京都と同じく1/21に重点措置を開始した9都道府県と実施しなかった11都道府県を比較することで、DID的に重点措置の効果を検証。重症病床使用率は医療逼迫度合いを表し、これが人々の行動変容効果を捉えると想定して追加。

■ データ：

重症病床数、重症患者数、
新規陽性者数、Google Mobility



■ 重要ポイント

- ① まん延防止等重点措置が新規陽性者数の変化率に与える効果は統計的に有意な形では確認できなかった。
- ② ①の結果は、Kitamura (2021) の結果と整合的である。
- ③ 重症病床使用率が1%増加すると、変化率は1%低下する。
- ④ 一方で、まん延防止等重点措置開始後、在宅率が大きく上昇していることが確認された。(ただし、これは因果関係を立証するものではない。)

■ その他の情報

- 分析結果をまとめたスライドは本人たちに要連絡

- 毎週火曜日分析を更新

<https://Covid19OutputJapan.github.io/JP/>

- 質問・分析のリクエスト等
 - dfujii@e.u-tokyo.ac.jp
 - taisuke.nakata@e.u-tokyo.ac.jp