Supplemental Material

Supplemental Note

Two types of personal identification numbers: ID1 and ID2

Because Japan has not deployed unique personal identification numbers in the healthcare system, unlike the U.S.[1] or Taiwan, [2] the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB) includes two types of personal identification number to link the insurance claims of individual patients: ID1 and ID2.[3] ID1 is a hash value generated from the insurer's ID, and the beneficiary's ID, date of birth, and sex; ID2 is a hash value generated from the beneficiary's name, date of birth, and sex.[3, 4] Because these IDs can change by some life events such as marriage, career change, retirement, etc., more robust personal identification number is preferable.

ID0

To improve personal identification, the improved personal identification number named ID0 has been proposed and is commonly used in NDB analysis. While the original method was described by Kubo et al.,[4] their method was limited in that it could link only claims of 2 years due to computer recourse.

In this study, we were able to link claims of over 8 years by modifying the original ID0 methods. Briefly, our modified method generates a number of chunks, and processes them one by one to effectively utilize the computer recourse. The details are described hereafter.

ID0 creation

ID1, ID2, the date of medical care, and medical care outcomes were extracted from the following 5 NDB categories: "medical inpatient claims," "medical outpatient claims," "medical inpatient claims subjected to bundled payment during diagnosis procedure combination (DPC) hospitalization," "DPC claims during DPC hospitalization," and "DPC claims subjected to bundled payment during DPC hospitalization".

We first searched for the same ID1 and determined that it corresponded to the same individual. The data linkage process ended when death was noted in the outcome section. When a series of claims tracked by ID1 was interrupted at a certain point, the corresponding ID2 at that point was used for further tracking. If more than one candidate ID2 was found, the data linking process was terminated to avoid linking data corresponding to different individuals.

Originally, data for all ages were extracted and linked simultaneously at this stage. However, the data volume was too large to use all the data of the 8 years from 2011 to 2018. Therefore, to deal with this problem, we extracted the data of the 8 years in small sections separated by age. To track individuals by ID1 or ID2, the same person's data must be extracted as a set of data; thus, we did so with great care.

Claims associated with an individual whose eligibility is under question are re-reviewed by the insurer through the Health Insurance Claims Review and Reimbursement Services. Such claims for re-review are not included in the NDB, and the old ID1 and a newly assigned ID1 can co-exist for approximately 3 months. To address this problem, the ID2 corresponding to the old ID1 was used to search data in the following month, the first preceding month, and the second preceding month to obtain an intermediate dataset table (medical and DPC).

The new ID1 was replaced by the old ID1 within a row so that a single ID1 was assigned (provisional ID0). When the replaced ID1 was the old ID1 of a different pair, it was replaced by the new ID1. This process was repeated until all second ID1s were replaced. The remaining ID1 was defined as a new variable ID0. In other words, a new data linkage variable ID0 was obtained by sequentially replacing an old ID1 with a new ID1. When there were no linkable claims because of only one medical visit (and thus no one-to-one ID1 pairing), the ID1 of the single claim served as ID0.

The ID0 generated in this way was used to follow up by connecting claims data across multiple months for the same patient. We identified the first and last month of claims data stored in the NDB for each identical ID0 of the target patient. Months for which no claims data occurred within that period were considered "periods for which insurance was not required."

1. Acquisti A, Gross R. Predicting Social Security numbers from public data. Proc Natl Acad Sci U S A 2009; 106: 10975-80.

2. Hsing AW, Ioannidis JP. Nationwide Population Science: Lessons From the Taiwan National Health Insurance Research Database. JAMA internal medicine 2015; 175: 1527-9.

3. Kato G. History of the secondary use of National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB). Trans Jap Soc Med Biol Eng 2017; 55: 143-50.

4. Kubo S, Noda T, Myojin T, et al. National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB): Outline and Patient-Matching Technique. bioRxiv 2018: 280008.

Supplemental Table

Supplementary Table. Correspondence Table of Codes used in the current study

A. NDB diagnostic codes

NDB diagnostic code	Disease Name	ICD-10 code
3624004	Central Serous Choriopathy	H35.7
3630002	Central Serous Retinochoriopathy	H35.7
8837586	Central Chorioretinopathy	H30.0
8837587	Central Retinopathy *	H30.0
3624006	Central Chorioretinitis *	H30.0
3630004	Central Retinitis *	H30.0
8837585	Central Choroiditis *	H30.0

*These disease names are historical synonyms for Central Serous Chorioretinopathy.

The ICD-10 code H35.7 and H30.0 correspond to six and eleven NDB diagnostic codes, respectively. We employed only two of six and five of eleven NDB diagnostic codes which correspond to CSC.

B. NDB drug codes

NDB drug code	Drug Name	ATC code
620001909	Verteporfin	S01LA01
620008448	Pegaptanib	S01LA03
620009103	Ranibizumab	S01LA04
621894901	Ranibizumab	S01LA04

622352001	Ranibizumab	S01LA04
622199401	Aflibercept	S01LA05

C. NDB procedure codes

NDB procedure code	Procedure Name	ICD-9-CM code
150244110	Laser photocoagulation	14.55

Supplemental Figure

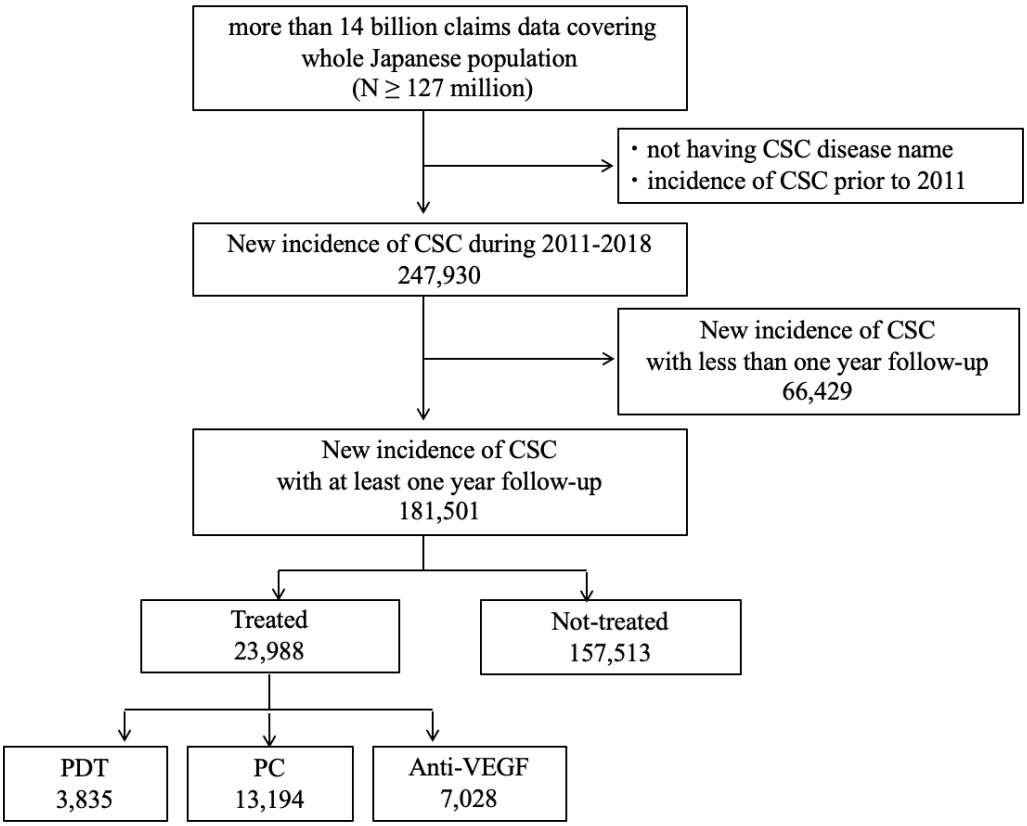
Supplementary Figure 1. Flow-diagram.

This is a flow-diagram showing the process of selecting central serous chorioretinopathy patients in the current study.

PDT: photodynamic therapy

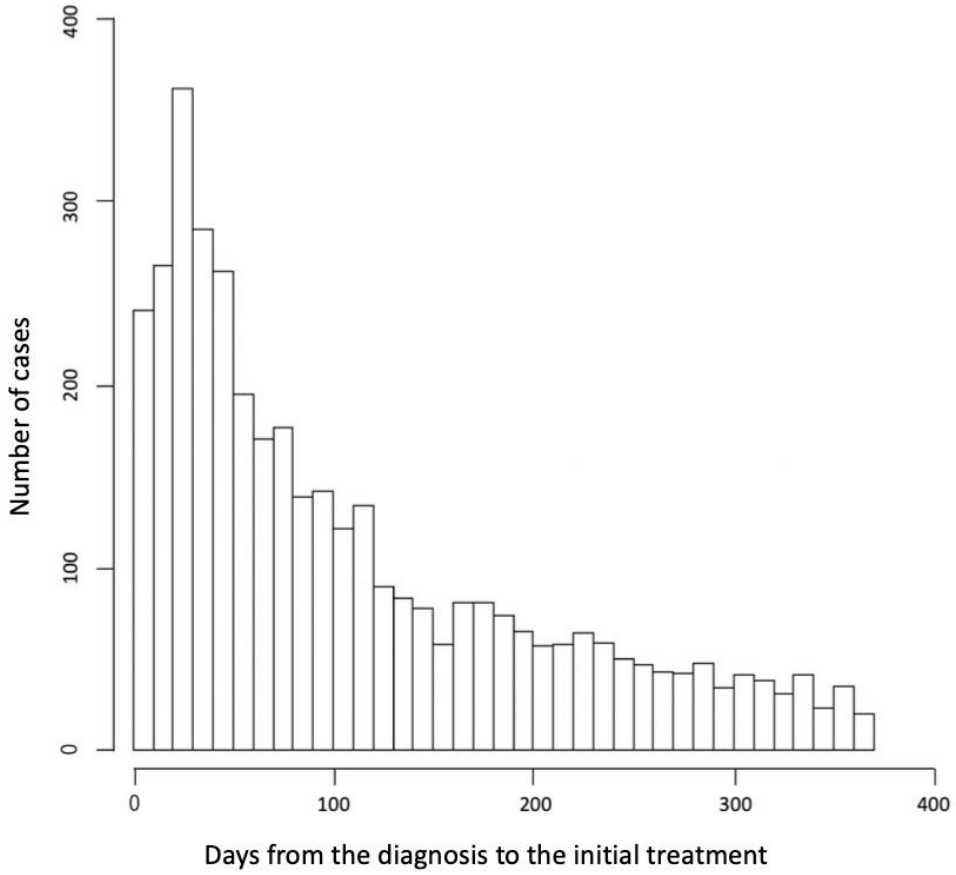
PC: laser photocoagulation

Anti-VEGF: anti-vascular endothelial growth factor



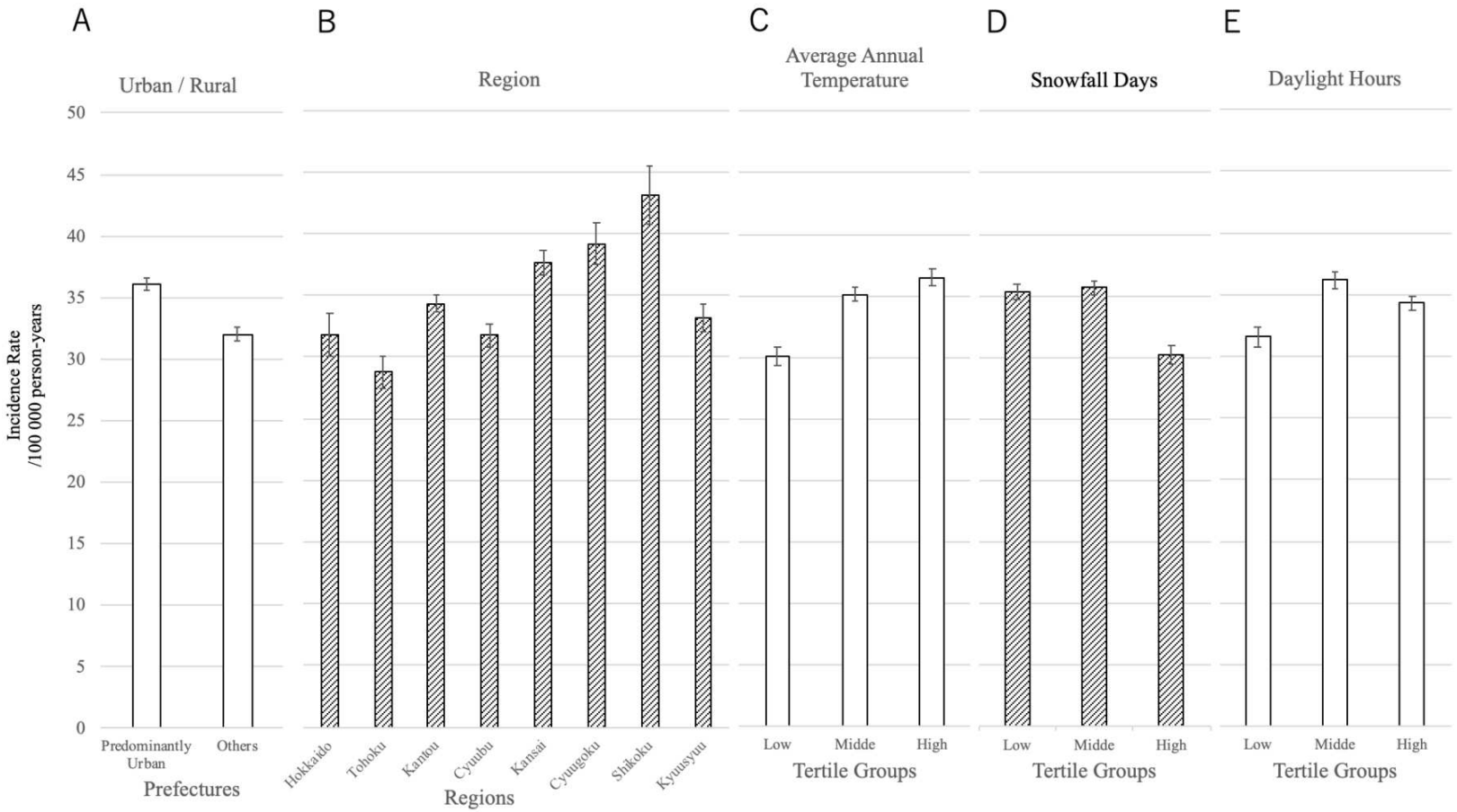
Supplementary Figure 2. Days from the diagnosis of central serous chorioretinopathy to the initial treatment with photodynamic therapy.

The histogram shows the number of cases per 10-days group from the diagnosis of central serous chorioretinopathy to the initial treatment with photodynamic therapy with verteporfin. The peak period of initial treatment 30-40 days from the diagnosis. The overall distribution was skewed to the right.



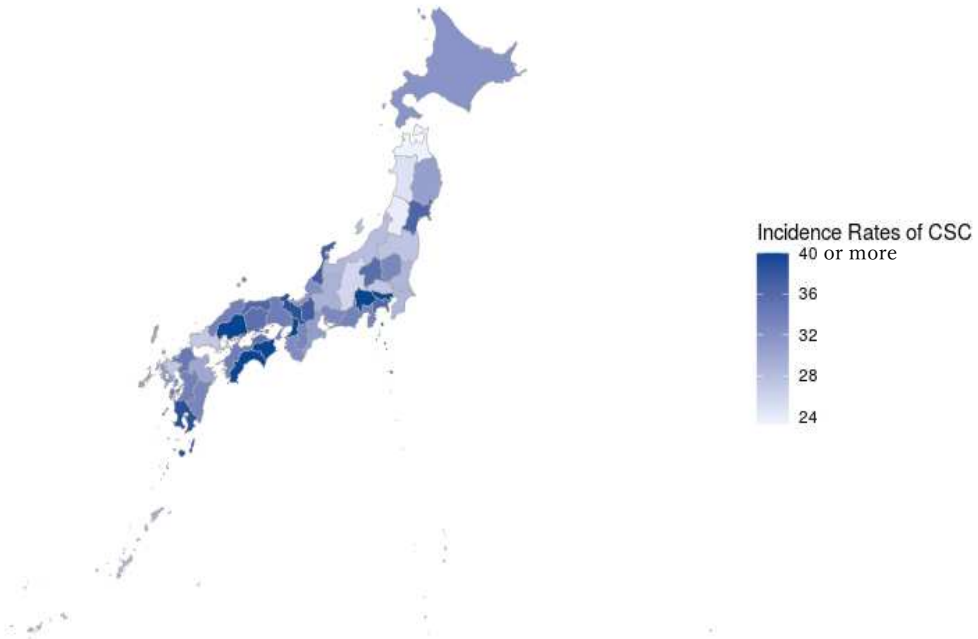
Supplementary Figure 3. Geographical and climatic variation of the incidence rate of central serous chorioretinopathy (CSC) per 100,000 person-years.

A: Comparison between predominantly urban or rural prefectures. The incidence rate of CSC was significantly higher in predominantly urban prefectures than in rural prefectures. **B:** Comparison among the eight main Japanese regions. Regional variation of incidence rate of CSC was apparent. The lowest and highest incidence rates were observed in Tohoku and in Shikoku, respectively. **C:** Comparison among tertile groups of the average annual temperature. The incidence rate of CSC increased in line with the rise in average temperatures. **D:** Comparison among tertile groups of snowfall days in a year. The incidence rate of CSC was lowest in the group with the highest snowfall days per year. **E:** Comparison among tertile groups of daylights hours in a year. The incidence rate of CSC was lowest in group with the lowest total daylight hours per year. Incidence rate of CSC is calculated as incidence / total population at risk in each group. Error bars represent 95% confidence intervals.


A
Urban / Rural
B
Region
C
Average Annual Temperature
D
Snowfall Days
E
Daylight Hours
Incidence Rate /100 000 person-years
50
45
40
35
30
25
20
15
10
5
0
Predominantly Urban
Others
Prefectures
Hokkaido
Tohoku
Kantou
Cyuubu
Kansai
Cyuugoku
Shikoku
Kyuusyuu
Regions
Low
Midde
High
Tertile Groups
Low
Midde
High
Tertile Groups
Low
Midde
High
Tertile Groups

Supplementary Figure 4. Geographical distribution of incidence rates of central serous chorioretinopathy by the Japanese 47 prefectures.

The incidence rates of central serous chorioretinopathy across the 47 Japanese prefectures are presented as a heat map. The incidence rates ranged from 23.9 to 69.7 per 100,000 person-years in Aomori and Kochi, respectively. High incidence rates were observed in urban or western prefectures.



Supplementary Figure 5. Seasonal variation in incidence of central serous chorioretinopathy (CSC). Dashed lines represent average annual monthly incidence of CSC for the entire year. The distribution of CSC in a year was significantly different ($P < 0.001$, analysis of variance). Incidence from March to June and from October to November were higher than the annual average monthly incidence.