



AMENDED REAL WORD AND EPIDEMIOLOGY STUDY PROTOCOL

Association Between the Use of Hydrochlorothiazide and the Risk of Skin Cancers: A population-based retrospective cohort study in Taiwan Hydrochlorothiazide

Study type:	Retrospective non-international study based on the secondary use of pre-existing data
Study Number:	HCLTHC09553

Version Number	1.0
-----------------------	-----

Date	9 Nov 2021
-------------	------------

Total number of pages	42
------------------------------	----

Any and all information presented in this document shall be treated as confidential and shall remain the exclusive property of Sanofi (or any of its affiliated companies). The use of such confidential information must be restricted to the recipient for the agreed purpose and must not be disclosed, published or otherwise communicated to any unauthorized persons, for any reason, in any form whatsoever without the prior written consent of Sanofi (or the concerned affiliated MAH/MAH representative); 'affiliated MAH/MAH representative' means any corporation, partnership or other entity which at the date of communication or afterwards (i) controls directly or indirectly Sanofi, (ii) is directly or indirectly controlled by Sanofi, with 'control' meaning direct or indirect ownership of more than 50% of the capital stock or the voting rights in such corporation, partnership or other entity

Study / PASS information:

Title:	Association Between the Use of Thiazide Diuretics and the Risk of Skin Cancers: A Population-based Retrospective Cohort Study in Taiwan
Protocol version identifier	Amended protocol v1.0 version 1.0
Date of last version of protocol	9 Nov 2021
EU PAS register number (if applicable)	N/A
Active substance	Hydrochlorothiazide C03AA03, Trichlormethiazide C03AA06, Benzyhydrochlorothiazide C03AA, Indapamide C03BA11
Medicinal product	N/A
Product reference	N/A
Procedure number	N/A
Marketing authorization holder(s)	N/A
Joint PASS	No
Research questions and objective(s):	What is the association between hydrochlorothiazide and risk of skin cancers, including non-melanoma skin cancer (NMSC) (i.e., basal cell carcinoma and squamous cell carcinoma) and malignant melanoma (MM)?
Country(-ies) of study	Taiwan
Author	Edward Chia-Cheng Lai, Associate Professor, National Cheng Kung University Email: edward_lai@mail.ncku.edu.tw

MARKETING AUTHORIZATION HOLDER(S)

Marketing authorization holder(s) Sanofi-Aventis group
54 Rue La Boétie, 75008 Paris, France

MAH contact person Heike Schoepper, MD, PhD, MBA
Sanofi
1 Avenue Pierre Brossolette 91385 Chilly-Mazarin France
Tel: +49 69 305 15762
Email: Heike.Schoepper@sanofi.com

TABLE of CONTENT

1.	LIST OF ABBREVIATIONS	6
2.	RESPONSIBLE PARTIES	7
3.	ABSTRACT	8
4.	AMENDMENTS AND UPDATES	10
5.	MILESTONES	11
6.	RATIONALE AND BACKGROUND	12
7.	RESEARCH QUESTION AND OBJECTIVES	13
7.1	OBJECTIVES BY ENDPOINT	13
8.	RESEARCH METHODS	14
8.1	STUDY DESIGN	14
8.2	SETTING	14
8.3	VARIABLES.....	17
8.4	DATA SOURCES	19
8.5	STUDY SIZE	20
8.6	DATA MANAGEMENT.....	20
8.7	DATA ANALYSIS.....	21
8.8	QUALITY CONTROL.....	25
8.9	LIMITATIONS OF THE RESEACH METHODS	25
8.10	OTHER ASPECTS	26
9.	PROTECTION OF HUMAN SUBJECTS	26
9.1	INDEPENDENT ETHICS COMMITTEE/INSTITUTIONAL REVIEW BOARD	26
10.	MANAGEMENT OF REPORTING ADVERSE EVENTS/ADVERSE REACTIONS	27

11.	PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS	27
	FIGURE 1. STUDY COHORT SELECTION FLOWCHART.....	28
	FIGURE 2. STUDY DESIGN SCHEMA I (AIM 1 TO AIM 4)	29
	FIGURE 3. STUDY DESIGN SCHEMA II(AIM5).....	30
12.	APPENDICES.....	34
	APPENDIX A CODE FOR ANTIHYPERTENSIVE AGENTS EXPOSURE	34
	APPENDIX B CODE FOR EXCLUSION CRITERIA DEFINITION	35
	APPENDIX C COVARIATES OF BASELINE CHARACTERISTIC.....	36
	APPENDIX D ALGORITHM TO IDENTIFY PATIENT’S RESIDENCY CODE.	38
	APPENDIX E CODES FOR SKIN CANCER TOPOGRAPHY DEFINITION	39
	APPENDIX F CODES FOR SQUAMOUS CELL CARCINOMA MORPHOLOGY	40
	APPENDIX G CODES FOR BASAL CELL CARCINOMA MORPHOLOGY	41
	APPENDIX H CODES FOR MALIGNANT MELANOMA MORPHOLOGY	42

1. LIST OF ABBREVIATIONS

ICD-O-3	International Classification of Diseases for Oncology, 3 rd Edition
NHID	National Health Insurance Database
EPA	Environmental Protection Administration
UV	Ultraviolet
NMSC	Non-melanoma skin cancer
SCC	Squamous cell carcinoma
BCC	Basal cell carcinoma
MM	Malignant melanoma
HCTZ	Hydrochlorothiazide
TCRD	Taiwan Cancer Registry Database
IPTW	Inverse probability of treatment weighting
SMRW	Standardized mortality ratio weighting

2. RESPONSIBLE PARTIES

Role	Institution	Contact information
Sponsor	Sanofi R&D	Leila DJEBARRI PharmD, MPH PharmacoEpidemiologist SANOFI-AVENTIS R&D 1 avenue Pierre Brossolette, 91380 Chilly Mazarin Email: leila.djebarri@sanofi.com
Sponsor	Sanofi R&D	Denis GRANADOS MD, MPH PharmacoEpidemiology Head GenMed SANOFI-AVENTIS R&D 1 avenue Pierre Brossolette, 91380 Chilly Mazarin Email: Denis.Granados@sanofi.com
Principal investigator	National Cheng-Kung University	Edward Chai Cheng Lai Associate Professor Email: edward_lai@mail.ncku.edu.tw
Contracting vendor	IQVIA Solutions Taiwan, Ltd.	Kevin Lin Consultant Email: Kevin.Lin@iqvia.com

3. ABSTRACT

Title: Association Between the Use of Hydrochlorothiazide and the Risk of Skin Cancers: A population-based retrospective cohort study in Taiwan

Rationale and background: Studies predominantly based on Caucasian populations have reported that the use of hydrochlorothiazide (HCTZ) is associated with increased risk of non-melanoma skin cancer (NMSC), especially squamous cell carcinoma¹. Databases allowing for the assessment of the risk of NMSC with HCTZ in Asian population are scarce. The Taiwan National Health Insurance Database (NHID) has given the opportunity for such previous assessments. A case-control study from NHID has suggested no association between the use of HCTZ and skin cancers², although it lacked detailed analyses based on subtypes of skin cancers by histology.

Research question and objectives: To evaluate the association between HCTZ and risk of skin cancers by various histologic subtypes, including basal cell carcinoma (BCC), squamous cell carcinoma (SCC) and malignant melanoma (MM). Dose response analysis will also be conducted to evaluate the relationship.

Study design: Retrospective cohort study

Population: A cohort of patients aged 20+ years, and newly receiving antihypertensive agents will be identified from 2004 to 2015. Patients will be classified into: A. hydrochlorothiazide (HCTZ), B. thiazide diuretics other than HCTZ and C. all diuretics other than thiazide group, D. other antihypertensive agents other than diuretics, and E. a control cohort comprising one million randomly sampled patients which are not included in above cohort will be selected for comparison.

Objectives: The main objective of this study is to evaluate the risk of non-melanoma skin cancer (NMSC) and malignant melanoma (MM) with the use of HCTZ versus control cohort.

Endpoints: The primary endpoints will be the non-melanoma skin cancer (NMSC) which includes basal cell carcinoma (BCC), squamous cell carcinoma (SCC). The secondary endpoints will further specify the subtypes of skin cancers which include: (1) SCC, BCC, and MM; (2) the topography of skin cancer (areas with different levels of sun exposure) and (3) morphology of melanoma (nodular melanoma, lentigo melanoma, superficial spreading melanoma, and acral lentiginous melanoma) will also be analyzed, finally, (4) a composite outcome including all kinds of skin cancers will be also assessed. We will perform intention-to-treat analysis and follow up patients from 2 years after

medication initiation until occurrence of study endpoints, death, or the end date of the database (Dec 31st, 2018). We will conduct an as-treated analysis analogously but will follow-up only for as long as patients have remained on the index of antihypertensive agents.

Variables: The variables include diagnosis code (International Classification of Diseases for Oncology, 3rd Edition, ICD-O3), ATC code and baseline covariate. The baseline covariates include patients demographics (age, sex, premium level of insurance, urbanization levels of residence, geographic area of residence), health resource utilization (number of emergency room visits, outpatient department visits, and hospitalizations), comorbidities, co-medication

Data sources: The study will extract the data from the largest real-world database in Taiwan NHID and will link to the Cancer Registry of Taiwan to verify the association between the use of thiazide diuretics and the various subtypes of skin cancers in the Asian population.

Study size: A pilot analysis using NHID shows that in the incidence rate of NMSC around 0.0023, we will need to study 246,818 patients to achieve the power of 80% and significant level 0.05.

Data Analysis: standardized mean difference (SMD) will be used to compare the baseline characteristics between groups. Hazard ratios and 95% confidence intervals derived from Cox-proportional hazard models will be used to compare the risks of study endpoints among the different groups. High-dimensional propensity score methods with matching, inverse probability of treatment weighting (IPTW), and standardized mortality ratio weighting (SMRW) will be performed to minimize measured and unmeasured confounding effects.

4. AMENDMENTS AND UPDATES

DOCUMENT HISTORY

Number	Date	Section of study protocol	Amendment or update	Reason
Amended protocol v1 version 1.0	09-Nov-2021	Section 8.2, Section 8.3, Section 8.4, Section 8.7 and Appendix C	Amendment	Change study time period
Original protocol v1	28-May-2021	Not applicable	Not applicable	Not applicable

The following sections have been updated:

- 8.2: SETTING
- 8.3: VARIABLES
- 8.4: DATA SOURCES
- 8.7: DATA ANALYSIS
- Appendix C: COVARIATES OF BASELINE CHARACTERISTIC

Protocol amendment summary of changes table

Section No. and Name	Description of Change	Brief Rationale
8.2: SETTING	Changed the time period (study entry date and follow up period) according to updated data period	TCRD is unavailable for 2018 due to the data administration policy, the data period for related datasets (NHID, National Death
8.3: VARIABLES	Changed the time period (study entry date and follow up period) according to updated data period	Registry and the EPA database) has been updated accordingly. Patient selection period of aim1-aim4 and aim5 has been changed
8.4: DATA SOURCES	Changed the data period for NHID, National Death Registry and the EPA	from 2004-2016 to 2004 -2015 and from 2004-2013 to 2004-2012 respectively, to ensure

Section No. and Name	Description of Change	Brief Rationale
8.7: DATA ANALYSIS	Changed the time period (study entry date and follow up period) according to updated data period	patients have at least 2-year follow-up for accurately classify the study cohort and ensure have enough time for outcome follow up.
Appendix C: COVARIATES OF BASELINE CHARACTERISTIC	List of comedication to be presented by the alphabetical order	

5. MILESTONES

Milestone	Planned date
Protocol approval	May 2021
Start date of the data extraction	Sep 2021
End date of the data extraction	Feb 2022
Final report of study results	May 2022

6. RATIONALE AND BACKGROUND

Skin cancer is a common cancer world-wide and among the Taiwanese population. According to data from the Taiwan Cancer Registry, the incidence of skin cancer has rapidly increased from 2 cases/100,000 patients in 1980 to 9.62 cases/100,000 patients in 2017 although it may be because of the improvement of registry.³ Among the different types of skin cancer, basal cell carcinoma is the most common histology constituting about 80-85% of all skin cancer. This is followed by squamous cell carcinoma which accounts for 15 to 20% of skin cancers⁴. The third most common type is melanoma, which accounts for only about 1% of skin cancers but is the leading cause of skin cancer death, due to its progressive and highly metastatic characteristics⁵.

The Fitzpatrick phototype system was initially developed for fair skinned individuals to gauge sun sensitivity based on self-reported erythema and ability to tan, following ultraviolet (UV) exposure⁶. It classifies skin into six different skin types based on skin color and reaction to sun exposure, ranging from very fair (skin type I) to very dark (skin type VI), whereby type I to III are at increased risk of sun damage, photo-aging and skin cancer, and type IV to VI are protected by pigmentation⁷. Although there have been many studies suggesting exposure to thiazide is associated with skin cancer risk, only a few have been based on an Asian population^{2,8}. A recent study by Pottegård et al.² applied case-control design and found no association between the use of thiazide and the risk of skin cancer in the Taiwanese population. The authors considered skin type to be one of the major factors accounting for this result, since the Taiwanese population has Fitzpatrick skin type III to IV. However, the study only analyzed skin cancer as a group without detailed analysis of the subtypes of skin cancer. A previous study had indicated that increased risk of skin cancer with the use of thiazide was more pronounced in SCC compared to BCC and MM⁹. Therefore, our study aims to extract data from a population-based database and link it with the Taiwan Cancer Registry to obtain information on the histology and morphology of the skin cancers, in order to address the issue of differential risk between subtypes of skin cancer.

7. RESEARCH QUESTION AND OBJECTIVES

The research question of this study is to evaluate the risk of skin cancer with the use of HCTZ versus control cohort and other antihypertensive agents.

Aim 1: To compare the risk of skin cancers, including BCC, SCC and MM among patients receiving HCTZ versus randomly sampled control cohort (**cohort A vs. cohort E**).

Aim 2: To compare the risk of skin cancers, including BCC, SCC and MM among patients receiving HCTZ versus other thiazide **diuretics (cohort A vs. cohort B)**.

Aim3: To compare the risk of skin cancers, including BCC, SCC and MM among patients receiving HCTZ versus non-thiazide -diuretics cohort (**cohort A vs. cohort C**).

Aim4: To compare the risk of skin cancers, including BCC, SCC and MM among patients receiving HCTZ versus hypertensive agents other than diuretics cohort (**cohort A vs. cohort D**).

Aim5: To evaluate the dose-response relationship between HCTZ use and the risk of skin cancers. (**Dose-response relationship**)

The detailed definition of cohort A, B, C, and D will be described in [8.2.1](#).

7.1 OBJECTIVES BY ENDPOINT

- A) Primary objective: The primary objective is the risk of non-melanoma skin cancer (NMSC) The primary objective will be compared between each specific cohort as our study aim definition.
- B) **Secondary objectives:** The secondary objectives will replicate the primary objectives by the following subgroups:
 - 1. Basal cell carcinoma (BCC)
 - i. Anatomy relationship analysis: BCC in sun exposed area/ sun protected area.
 - 2. Squamous cell carcinoma (SCC)
 - i. Anatomy relationship analysis: SCC in sun exposed area/ sun protected area.
 - ii. Lip SCC
 - 3. Malignant melanoma (MM)
 - i. Anatomy relationship analysis: MM in sun exposed area/ sun protected area.
 - ii. Acral lentiginous melanoma

- iii. Superficial spreading melanoma
 - iv. Nodular melanoma
 - v. Lentigo melanoma
4. Composite skin cancer (BCC, SCC, MM)

8. RESEARCH METHODS

8.1 STUDY DESIGN

This is a retrospective cohort study evaluating the risk of skin cancer among patient receive thiazide.

8.2 SETTING

8.2.1 Participants

A cohort of patients aged 20 years and older, newly receiving antihypertensive agents ([Appendix A](#)) from 2004-2015 will be identified and classified into four groups according to medication possession ratio (MPR) and each cohort is mutually exclusive: (A) receiving hydrochlorothiazide (HCTZ), (B) receiving thiazide diuretics except of HCTZ, (C) receiving other diuretic agents other than thiazide diuretics (loop diuretics, potassium sparing diuretics), (D) receiving other antihypertensive agents except of diuretic. Moreover, a random sample of 1 million as a control population (E) will be selected for comparison.

In the definition of medication possession ratio, we use fixed-medication possession ratio¹⁰ for the measurement and the formulation is following:

$$\frac{\text{Total prescription day of exposure drug within baseline period (e.g. HCTZ, thiazide, diuretics and other antihypertensive agent)}}{\text{Total duration of baseline period (730 days)}}$$

The classification flowchart of each cohort is defined as below:

Cohort (A): HCTZ cohort

Patients had 2 years exposure of antihypertensive agents with a HCTZ (monotherapy or combined) MPR above 80% will be selected. The study aim1-4 will include this cohort for analysis and the study design schema is presented in [Figure 1](#).



Cohort (B): Thiazide diuretics except of HCTZ

Patients who were exposed to thiazide, and thiazide-like diuretic for at least two years with a thiazide medication possession ratio (MPR) over 80% will be selected.



Cohort (C): Other diuretic agents other than thiazide diuretics (loop diuretics, potassium sparing diuretics)

Patients who were exposed to diuretics for at least two years with a medication possession ratio (MPR) over 80% will be selected.



Cohort (D): Other antihypertensive agents except of diuretic

Patients who were exposed to antihypertensive agents for at least two years with antihypertensive agents medication possession ratio (MPR) over 80% will be selected.



Cohort (E): Control cohort not treated with antihypertensives drugs

We will select our control cohort from all insured population from 2004-2015 of NHID. Patients aged above 20 years old and not included in cohort A-D will be randomly selected as control cohort. Each patient will be randomly assigned a study entry date, and then we will link the patient id with National Death Registry to exclude patient expired before study entry date to generate our one million cohort.

Exclusion criteria for all the cohorts:

Patients with some baseline comorbidity that may affect the result estimates described in previous studies¹¹ will be excluded. These conditions include any kind of cancer, human immunodeficiency virus

infection, organ transplantation, systemic lupus erythematosus, rheumatic arthritis or chronic kidney disease. We will check if patients have any record of these conditions within two years before the study entry date. In addition, patients receiving immuno-suppressant treatment for more than 6 months before the study entry date because many immunosuppressant agents have been associated with higher risk of skin cancer^{11,12} will be excluded. A list of conditions and immuno-suppressants for exclusion is presented in [Appendix B](#). Moreover, we will exclude patients who died within two years after drug initiation for Aims 1 and 2; we will exclude patients who died within five years after drug initiation for the fixed exposure group analysis of Aim 3. The study cohort selection flowchart is presented in [Figure 1](#).

8.2.2 Study time period

The study entry date for the cohorts (A, B, C) will be defined as two years after drug initiation, and for the control population (D) the study entry date will be two years after a randomly assigned date ([Figure 2](#)).

In order to clarify the association between dose and skin cancer as mentioned in our study **Aim 5**. We will do both fixed dose analysis and time varying analysis. In the fixed dose analysis, the study entry date will be defined as five years after drug initiation. Moreover, we will use time-varying approach to classify patients based on cumulative doses of HCTZ. The study entry date for the time-varying exposure analysis will be the first date of drug initiation. The study design schema for Aim 5 is presented in [Figure 3](#).

And the baseline period will be defined as 2 years after antihypertensive agents initiation in cohort A~C and 2 years after a randomly assigned date in cohort D. The variables in this baseline period will be collected as covariate and put into model for adjustment ([See section 9.3](#)).

Patient follow-up by intention-to-treat design will run from study entry date until the earliest date of occurrence of any kind of skin cancers, death or the end of the study period (31st Dec, 2017).

8.2.3 Overview of the linkage

We will extract entire datasets for the period from 2004 to 2017 from the National Health Insurance

Database (NHID), We will link the NHID to the Taiwan Cancer Registry and National Death Registry via the patient ID for ascertainment of the study endpoints. The data custodian will encrypted patient ID after the data is linked and created a unique anonymized ID for each patients. The study results will reflect real-world effects of drugs and their association with skin cancers. Through this high representativeness of the database and appropriate study design, the study can well be generalized to the entire population of Taiwan.

8.3 VARIABLES

8.3.1 Patients' characteristics

The baseline covariates will include patients' demographics (age, sex, level of insurance premium, urbanization levels of residence, geographic area of residence), health resource utilization (number of emergency room visits, outpatient department visits, and hospitalizations), comorbidities and co-medications. The details are listed in [Appendix C](#). Specifically, we will include an indicator, mean daily UV index, for the levels of sun exposure based on the area of residence of patients and drawn from the EPA environmental database. The geographic area of residence is available from the NHID eligibility data. Moreover, we will apply an algorithm developed by a previous study¹³ to verify patients' actual geographic area of residence, as shown in [Appendix D](#). We will also include medications that have been associated with photosensitivity in previous studies¹⁴⁻¹⁶. The list of photosensitivity medication is presented in [Appendix C](#). The variables will be considered in the regression models and propensity score methods for adjustments.

8.3.2 Exposure definition

Classification of patient's exposure status: The exposure of this study is the antihypertensive agents listed in Appendix A. Hypertensive patients will be classified into HCTZ group, thiazide group, and other antihypertensive agents group according their first 2 years MPR of exposure. Which means if patient's HCTZ MPR is above 80% then he/she will be included as HCTZ group, else if thiazide MPR is above 80% (also include HCTZ for MPR calculation) then he/she will be include as

thiazide groups but not in HCTZ groups, the other patients will include as other antihypertensive agents group if their antihypertensive agents MPR is above 80%.

Fixed exposure group analysis: For study Aim5, patients who received HCTZ from 2004 to 2012 will be classified into five groups as below, based on the ranges of cumulative doses of HCTZ within 5 years after drug initiation. This dose classification has been modified from a previous study [2,8](#) and adopted from prescription patterns of Taiwan, including (1) HCTZ <10,000 mg, (2) 10,000 mg ≤HCTZ <25,000 mg, (3) 25,000 mg ≤HCTZ <50,000 mg, (4) 50,000 mg ≤HCTZ <75,000 mg, and (5) HCTZ ≥75,000 mg. We will use the control cohort as the reference group for the dose response relationship analysis.

Also, we will include time varying approach for dose response analysis, the detail will be provided in [section9.7.2](#).

Discontinuation of the medication: in the as treated analysis, the grace period for medication discontinuation is 3 months.

8.3.3 Outcomes definition

Non-Melanoma skin cancer (NMSC) will be defined with ICD-O-3 codes from TCRD, which including basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and the codes for BCC and SCC is provided in [AppendixF&G](#). Regarding the malignant melanoma (MM), ICD-O-3 codes will also be used for defining MM and the detail code is provided in [Appendix H](#).

Subtypes of skin cancers by topography will be categorized into sun exposed position and sun protective position according to their location as below mention. The other topography code of skin cancer will be identified as no specific sun exposure status and not include in anatomy relationship analysis. (see [section:8.1](#))

- Sun exposed position: lip, eyelid, face, neck, upper limb, and shoulder.
 - Topography code: C440-C444, C446
- Sun protective position: lower limb, trunk, and hip.
 - Topography code: C445, C447

The four major morphology of melanoma (nodular melanoma, lentigo melanoma, superficial spreading melanoma, and acral lentiginous melanoma) will be selected for analyzed.

We will specifically evaluate lip SCC because thiazide has been shown to pose relatively high risk of lip SCC in previous studies. A composite endpoint including any kind of skin cancers will also be assessed.

8.4 DATA SOURCES

- (1) **National Health Insurance Database (NHID) 2003-2017** This study will utilize data from the National Health Insurance Database (NHID), which is provided by the Health and Welfare Data Science Center (HWDC) in Taiwan. The details of NHID have been described elsewhere¹⁷. Briefly, the NHID is derived from the National Health Insurance (NHI) program of Taiwan, containing records of approximately 23 million individuals (nearly 99.9% of the total population of Taiwan). The NHID includes records of diagnoses, medications and procedures from outpatient settings, inpatient settings, emergency rooms and contracted pharmacies. Many major disease diagnoses in the NHID have been validated by previous studies. We will link the NHID with the Taiwan Cancer Registry Database (TCRD) for cancer case ascertainment, as well as with the National Death Registry to verify patients' follow-up and mortality.
- (2) **Taiwan Cancer Registry Database (TCRD) 2006-2017:** the TCRD is also provided by the HWDC in Taiwan. The TCRD was established in 1979 to monitor cancer incidence at the national level¹⁸. Hospitals with more than 50 beds were requested to report newly diagnosed cancer patients. The TCRD includes information on patient diagnosis of a specific cancer, morphology, topography, and cancer staging. The detailed information of skin cancer topography and morphology is provided in [Appendix E-H](#). This database has a high completeness of around 95% to 98%, with morphological verification around 97%. The database is divided into a short-form database and a long-form database, whereby the long-form database includes 114 items consisting of cancer staging and detailed treatment and recurrence information, recorded since 2002. However, the long-form database only includes 16 specific cancers and skin cancer is not included. The short-form database covers all types of cancer (except in long-form database) and includes 42 items covering morphology, topography, and social behavior¹⁹.

- (3) **National Death Registry 2004-2017:** The National Death Registry is also maintained by the HWDC. The National Death Registry contains information of primary and contributing causes of death and date of death. We will link the NHID with the National Death Registry to verify if patients remained alive during our follow up period. The information is coded following ICD9 and the reliability is above 80%²⁰.
- (4) **Taiwan's Environmental Resource Database 2004-2017** The EPA is a publicly available database containing data from the 40 observation stations that have been maintained by the Environmental Protection Administration since 2004. The database provides information on 1,369 environmental and natural resource items (e.g., air quality, water pollution, greenhouse gas analysis) including the daily UV index from each observation station. Based on their geographic area of residence we will classify patients' UV index as being high or low UV exposure using a cut-off point of 6.

8.5 STUDY SIZE

We are already performing a pilot study consist of one million randomly sample population from NHID. The data indicate that the NMSC incidence among control cohort is 0.0023. If the true relative risk of skin cancer for thiazide recipient relative to controls is 1.5 and the standard deviation is 0.29, we will need to study 246,816 patients to be able to reject the null hypothesis that this relative risk equals 1 with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05.

8.6 DATA MANAGEMENT

To strengthen the protection of human health data, the Ministry of Health and Welfare (MOHW) of Taiwan has created large data repositories, the Health and Welfare Data Science Center (HWDC), and centralized health databases including the NHID, TCRD and National Death Registry. The HWDC manages these databases under tight supervision. Since November 2015, researchers have been required to visit an HWDC location in person to perform onsite analysis through remote connection to MOHW servers. No electronic devices may be brought into a center, and individual-level data are forbidden to be taken out. All analysis results to be brought out from a center must be reviewed by data custodians to prevent any disclosure of patient identity. Any number less than 3 is considered an identifiable number and thus cannot be brought out from an HWDC.

8.7 DATA ANALYSIS

8.7.1 Descriptive Analysis

Patient's baseline characteristics, covariates and outcomes will be described with number and percentage for categorical variables and mean with standard deviations, median and (interquartile range, IQR), for continuous variables. We will use the standardized mean difference to describe the differences in baseline covariates between cohorts. Standardized mean difference >10% represents a clinically significant difference.

8.7.2 Primary Analysis of the Endpoints

Study endpoints: The primary endpoint is the risk of NMSC (**Objective A**). The secondary endpoints (**Objective B**) are basal cell carcinoma (BCC), squamous cell carcinoma (SCC) and malignant melanoma (MM). Further, the subtypes of skin cancers by topography (See [8.3.3 Outcomes definition](#)) or morphology of melanoma (nodular melanoma, lentigo melanoma, superficial spreading melanoma, and acral lentiginous melanoma) will be analyzed as secondary endpoint. We will specifically evaluate lip SCC (Objective B.2.i) because thiazide has been shown to pose relatively high risk of lip SCC in previous studies. A composite endpoint including any kind of skin cancers (Objective B.4) will also be assessed. We will compare the incidence of study endpoints between patients receiving (1) HCTZ, (2) thiazide diuretics, and (3) other antihypertensive agents, and (4) a control cohort. Incidence rate will be calculated as the number of patients with an event of study endpoint divided by follow-up duration (described by per 100,000 person-years):

$$\frac{(Total\ number\ of\ patients\ with\ \geq\ 1\ event)}{(Sum\ of\ followed\ -\ up\ person\ -\ yrs\ to\ first\ event)} \times 100,000$$

The analysis of **Aim 5** will only focus on HCTZ and its dose-response relationship to each study endpoint. We will use both fixed exposure definitions and time-varying approach. In the fixed exposure group analysis, we will select patients who used HCTZ from 2004 to 2012 and classify them into five groups (<10,000 mg, 10,000-25,000 mg, 25,000-50,000 mg, 50,000 mg-75,000 mg and ≥75,000 mg) on the basis of the ranges of cumulative doses of HCTZ within 5 years after drug initiation

Time-varying exposure group analysis²¹: We will model the cumulative dose of HCTZ as a time-

varying variable, classify the patients into the five aforementioned categories by matching doses and cut-off points, and calculate the cumulative doses until the occurrence of the study endpoints. Since time varying is applied, the cumulative dose will be calculated since first prescription.

Survival Analysis: The Kaplan-Meier method will be used to plot unadjusted survival curves and relative risk will be estimated using the multivariable Cox proportional hazard model. The comparisons will be as follows: (1) thiazide type diuretics vs. control cohort and (2) thiazide type diuretics vs. other antihypertensive agents group. We will use the control cohort as the reference group for the risk assessment among different cumulative dose groups. Results are to be presented as adjusted hazard ratio (HR) with 95% confidence intervals (CIs) derived from the multivariable Cox models considering all specified covariates.

Propensity score analysis: To control confounding for measurable variables, propensity score (PS) matching will be used. The PS will be calculated for each patient in the study cohort using multivariable logistic regression analysis, conditional on all baseline covariates. The distribution of PS for both cohorts will first be described and the balancing of included covariates between the 2 cohorts before and after PS matching will be illustrated. Then, using the Greedy 8→1 digit algorithm, the propensity score will be used to match the persistent and non-persistent cohorts in a 1:1 ratio²². The survival probability in the PS matched cohort will be generated from the Cox model and presented as a survival curve.

8.7.3 Sensitivity Analysis

We will conduct several sensitivity analyses to examine the robustness of the study results, including:

- (1) **As-treated analysis:** We will conduct the analysis analogously, but follow-up will last only for as long as the patients remained on the index antihypertensive drugs. (Including a 90-day gap after the end of the medication supply to allow for late refills). That is, we will follow up patients from the date of drug initiation until occurrence of skin cancer, the end of the study period (2018/12), death, or discontinuation of medication, whichever came first.

- (2) **High-dimensional propensity score (hdPS):** We will use high-dimensional propensity score method with matching, inverse probability of treatment weighting (IPTW), and standardized mortality ratio weighting (SMRW). The hdPS will be used to address the potential impact of unmeasured confounding ([See section 9.10](#))

- (3) **Restriction analysis:** We will perform restriction analysis that will include patients with no record of non-skin cancer, human immunodeficiency virus infection, organ transplantation, systemic lupus erythematosus, rheumatic arthritis or chronic kidney disease during the follow up period. The analysis will extend the duration from the two-year baseline period to cover the entire study period, to capture the exclusion conditions and ensure these conditions do not affect the result estimates of the study.

- (4) **Confounding effect of smoking:** In order to evaluate the effect due to smoking, we will select a subgroup of male patients with COPD and repeat the analysis, with patients who are likely smokers.

- (5) **Alternate definition of chronic user:** To examine the assumption that MPR greater than 80% represents chronic users of antihypertensive agents and the effect thereof on the association assessment, we will use broader criteria and select a group of patients with MPR more than 60% for the sensitivity analysis.

8.7.4 Potential Impact of Residual Confounding and Bias

Three additional analyses will be conducted to address the potential impact of residual (or unmeasured) confounding. With regard to PS in the sensitivity analyses, we plan to calculate high-dimensional propensity score (hd-PS) in the sensitivity analysis. The hdPS algorithm is an automated technique that empirically identifies potential confounders or proxies for confounders in longitudinal data sets; the algorithm assesses thousands of diagnosis, procedure, and drug-dispensing codes recorded in administrative databases and then selects several hundred of those codes, transformed into binary covariates, which are called “empirical variables”²³. The hdPS algorithm consists of a 3 step, automated covariate identification, prioritization and causal treatment effect estimation using PS analysis. In claims data, common data dimensions include pharmacy claims, outpatient diagnoses, outpatient services, inpatient diagnoses, and inpatient services. From each dimension, the top n most prevalent codes are transformed into binary covariates and then individually considered for selection into a propensity score. With 5 dimensions and the default n = 200 and considering 3 levels of within-patient frequency of occurrence of each code (code occurred once, sporadically, or frequently), there are a possible 3,000 (200x3x5) indicator variables that could be added to a propensity score, a step called covariate identification. The hdPS algorithm then prioritizes each of these variables by its potential to bias the exposure-endpoint relation under study. Using the formula by Bross²⁴, the algorithm uses this approach to reduce a large number of candidate covariates by prioritization. By default, the algorithm will then include the top k = 500 of these covariates in a propensity score. Subsequently, hdPS will be used to examine the robustness of confounding controls by hdPS matching and weighting (including IPTW and SMRW).

We will calculate the E value using a 95% confidence interval for the adjusted hazard ratio. The calculation equation has been provided in a previous epidemiology study²⁵. The E-value is used to assess unmeasured confounders; to define the minimum degree of effect that an unmeasured confounder would need to explain the observed association between exposure and study endpoints. Instead of focusing on whether confounding of a specified strength would or would not suffice to overcome an effect estimate, the E-value focuses on the magnitude of the confounder associations that could produce confounding bias equal to the observed association^{25,26}.

Confounding by indication (i.e., pre-treatment variables that influence the treatment decision

and are also independent predictors of the outcome) is often the primary challenge for interpretation of a non-randomized study. The design of this study removes consideration for the initial treatment decision, which is one of the potential confounding variables. We will use hdPS and create more homogenous groups for comparison. Given the same hdPS, two groups may tend to have the same distribution of covariates and could be more exchangeable. Also, the active comparator study design will be applied to minimize the risk of immortal time bias. Immortal time refers to a period of follow-up during which, by design, death or the study outcome cannot occur²⁷. By using active comparator, we will reduce the immortal time bias between diagnosis and thiazide exposure²⁸. Moreover, the study entry date is the date 2 years (for Aim1 to Aim4) or 5 years after drug initiation (Aim5), which allows for a landmark time that can minimize the immortal time bias²⁹. Additionally, patients may undergo treatment modification in the early stage of therapy. By inclusion of the 2 or 5 year “lag time” we can confirm the patients’ exposure status to the drugs. We will select patients with at least 80% MPR of antihypertensive drugs for analysis to avoid possible misclassification bias.

8.8 QUALITY CONTROL

We will analyze the data on the basis of study protocol and statistical analytical plan (SAP), which are reviewed and refined by all investigators and clinician consultants. Two statisticians will perform and review the SAS programs to ensure that all analyses are correct.

8.9 LIMITATIONS OF THE RESEARCH METHODS

8.9.1.1 Limitation of database:

Some factors are not available from NHID, such as family history of skin cancer and life style (e.g. smoking or alcohol consumption), which have been reported to be related to skin cancer³⁰⁻³³.

- A systematic review study pooled data from a cohort study and 5 case-control studies and showed that smoking was associated with a 52% higher risk of SCC³². However, there was no association between smoking and the risk of BCC.
- In another large cohort study, smoking was associated with lower risk of melanoma in men and this association was both dose and duration dependent³⁴. The study also found smoking

only slightly increased the risk of BCC (OR: 1.02; 95% CI: 1.00–1.04) and SCC (OR: 1.08; 95% CI: 1.01–1.15).

- In one case-control study, smoking was associated with reduced risk of BCC and MM, but was not associated with increased risk of SCC³⁵.

To address these issues, we will select a subgroup of male patients with COPD and repeat the analysis (as a sensitivity analysis), because these patients are more likely to have been smokers. In Taiwan, the prevalence of smoking in females is very low (less than 5%).

8.9.1.2 Analytics approach limitations:

None

8.10 OTHER ASPECTS

8.10.1 Changes to the Protocol

None

8.10.2 Study Management

This study will be performed by IQVIA, with guidance, input, review and approval of Sanofi. The development of materials, data management, analysis and reporting will be insured by Pr. Lai Team.

9. PROTECTION OF HUMAN SUBJECTS

This study will not involve the collection, use or transmittal of individually identifiable patient / subject data. All the analyses will be conducted remotely in a branch of HWDC. To ensure the quality and integrity of the research, this study will be conducted under the guidelines for good pharmacovigilance practices (GVPs) and good pharmacoepidemiology practices (GPPs) issued by the International Society for Pharmacoepidemiology (ISPE), the Helsinki Declaration and its amendments, and any applicable national guidelines, laws and regulations.

9.1 INDEPENDENT ETHICS COMMITTEE/INSTITUTIONAL REVIEW BOARD

According to local regulations, the protocol will be submitted to the Institutional Review Board

(IRB) of NCKU/NCKU hospital for expedited review. Waiver of patient informed consent will be requested as all personal identifying information is encrypted within the NHID. Once IRB approval is obtained, the application for data (together with study protocol) will be submitted to the MoHW. Data access will be granted if the application is approved.

10. MANAGEMENT OF REPORTING ADVERSE EVENTS/ADVERSE REACTIONS

This study is a database analysis using anonymous individual information without any spontaneous reporting. Study outcomes will be reported in aggregate in the final study report, and no individual or expedited reporting is required, according to the EMA Guideline on good pharmacovigilance practices.

11. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

The study result and finding will be published in peer reviewed journal(s) and international conference(s) whenever possible. Authorship will follow guidelines established by Committee of Medical Journal Editors.

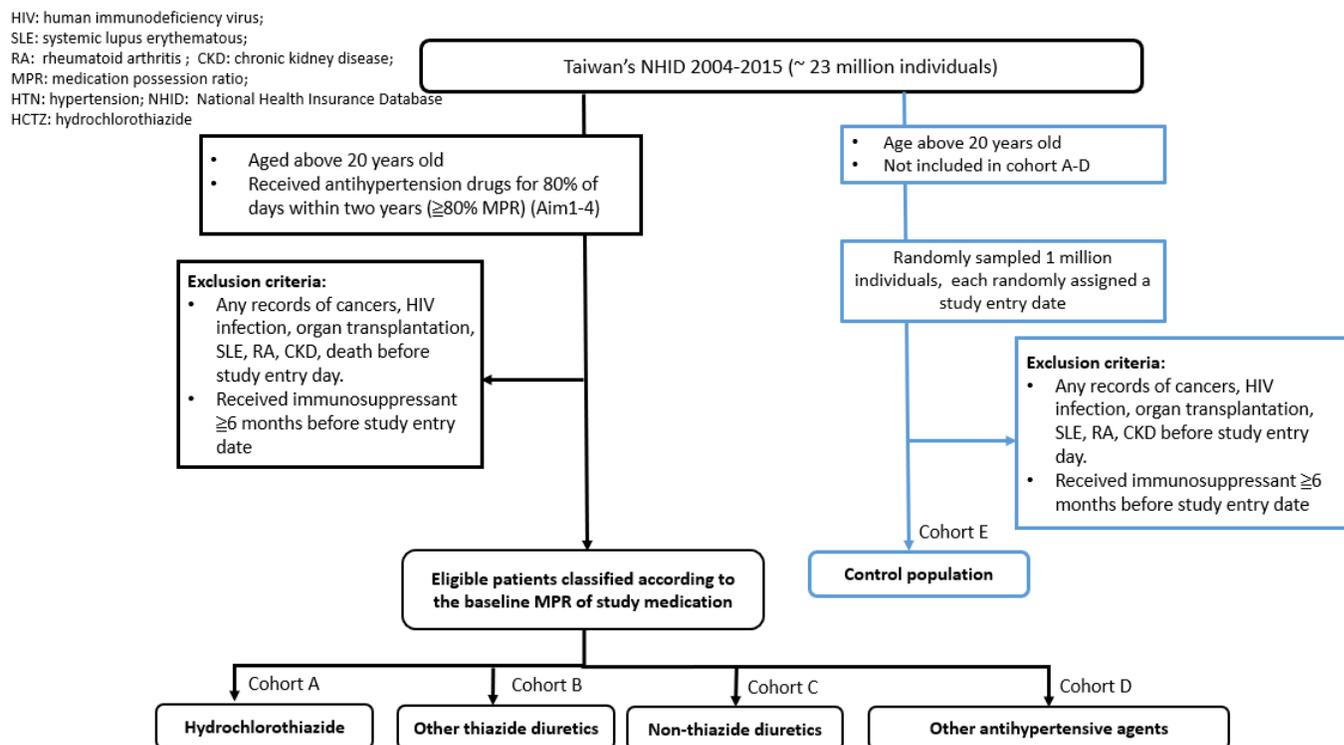


Figure 1. Study Cohort Selection Flowchart

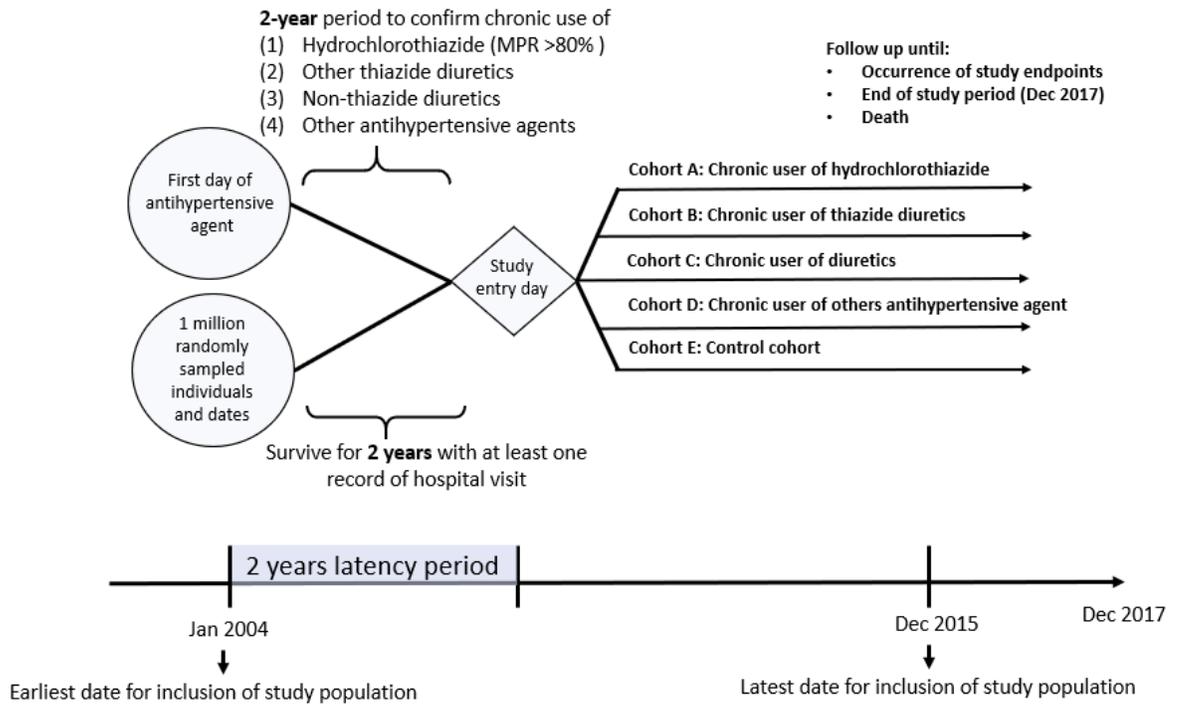


Figure 2. Study Design Schema I (Aim 1 to Aim 4)

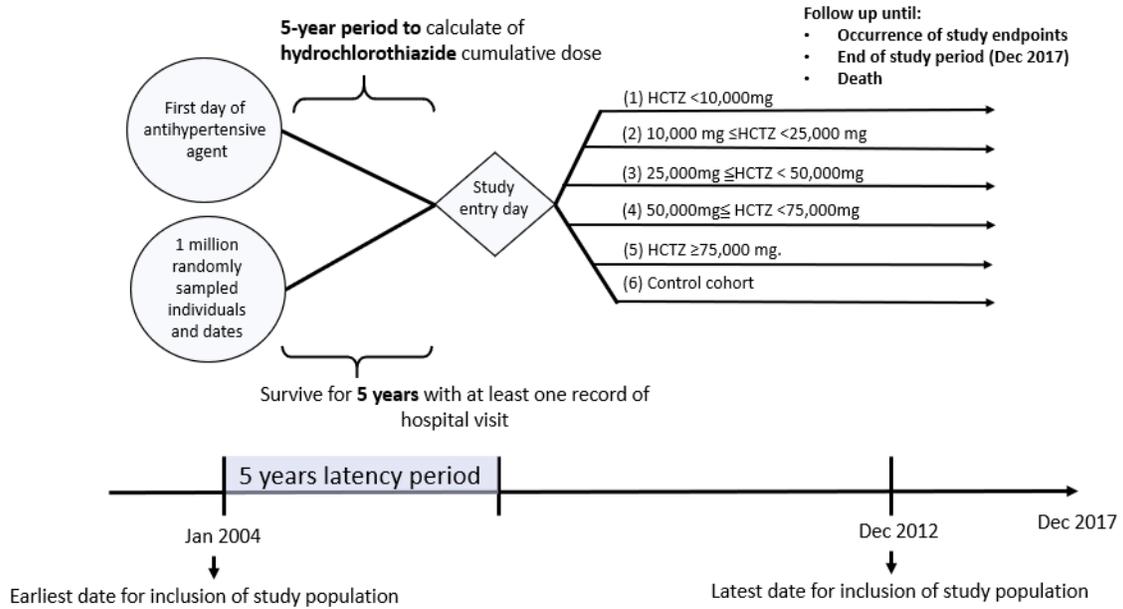


Figure 3. Study Design Schema II(Aim5)

Reference list

1. Pedersen SA, Gaist D, Schmidt SAJ, Holmich LR, Friis S, Pottegard A. Hydrochlorothiazide use and risk of nonmelanoma skin cancer: A nationwide case-control study from Denmark. *J Am Acad Dermatol*. 2018;78(4):673-681 e679.
2. Pottegard A, Pedersen SA, Schmidt SAJ, et al. Use of hydrochlorothiazide and risk of skin cancer: a nationwide Taiwanese case-control study. *Br J Cancer*. 2019;121(11):973-978.
3. Guy GP, Jr., Machlin SR, Ekwueme DU, Yabroff KR. Prevalence and costs of skin cancer treatment in the U.S., 2002-2006 and 2007-2011. *Am J Prev Med*. 2015;48(2):183-187.
4. Orthaber K, Pristovnik M, Skok K, Perić B, Maver U. Skin Cancer and Its Treatment: Novel Treatment Approaches with Emphasis on Nanotechnology. *Journal of Nanomaterials*. 2017;2017:1-20.
5. Domingues B, Lopes JM, Soares P, Populo H. Melanoma treatment in review. *Immunotargets Ther*. 2018;7:35-49.
6. Hogue L, Harvey VM. Basal Cell Carcinoma, Squamous Cell Carcinoma, and Cutaneous Melanoma in Skin of Color Patients. *Dermatol Clin*. 2019;37(4):519-526.
7. de Vries E, Trakatelli M, Kalabalikis D, et al. Known and potential new risk factors for skin cancer in European populations: a multicentre case-control study. *Br J Dermatol*. 2012;167 Suppl 2:1-13.
8. Park E, Lee Y, Jue MS. Hydrochlorothiazide use and the risk of skin cancer in patients with hypertensive disorder: a nationwide retrospective cohort study from Korea. *Korean J Intern Med*. 2020;35(4):917-928.
9. Jensen AO, Thomsen HF, Engebjerg MC, Olesen AB, Sorensen HT, Karagas MR. Use of photosensitising diuretics and risk of skin cancer: a population-based case-control study. *Br J Cancer*. 2008;99(9):1522-1528.
10. Sperber CM, Samarasinghe SR, Lomax GP. An upper and lower bound of the Medication Possession Ratio. *Patient Prefer Adherence*. 2017;11:1469-1478.
11. Green AC, Olsen CM. Cutaneous squamous cell carcinoma: an epidemiological review. *Br J Dermatol*. 2017;177(2):373-381.
12. Collins L, Quinn A, Stasko T. Skin Cancer and Immunosuppression. *Dermatol Clin*. 2019;37(1):83-94.
13. Lin MH, Yang AC, Wen T-H. Using regional differences and demographic characteristics to evaluate the principles of estimation of the residence of the population in National Health Insurance Research Databases (NHIRD). *Taiwan Journal of Public Health*. 2011;30:347-360.
14. Monteiro AF, Rato M, Martins C. Drug-induced photosensitivity: Photoallergic and phototoxic reactions. *Clin Dermatol*. 2016;34(5):571-581.

15. Ibbotson S. Drug and chemical induced photosensitivity from a clinical perspective. *Photochem Photobiol Sci.* 2018;17(12):1885-1903.
16. Blakely KM, Drucker AM, Rosen CF. Drug-Induced Photosensitivity-An Update: Culprit Drugs, Prevention and Management. *Drug Saf.* 2019;42(7):827-847.
17. Hsieh CY, Su CC, Shao SC, et al. Taiwan's National Health Insurance Research Database: past and future. *Clin Epidemiol.* 2019;11:349-358.
18. Chiang CJ, You SL, Chen CJ, Yang YW, Lo WC, Lai MS. Quality assessment and improvement of nationwide cancer registration system in Taiwan: a review. *Jpn J Clin Oncol.* 2015;45(3):291-296.
19. Chiang CJ, Wang YW, Lee WC. Taiwan's Nationwide Cancer Registry System of 40 years: Past, present, and future. *J Formos Med Assoc.* 2019;118(5):856-858.
20. Lu T-H, Lee M-C, Chou M-C. Accuracy of cause-of-death coding in Taiwan: types of miscoding and effects on mortality statistics. *International Journal of Epidemiology.* 2000;29(2):336-343.
21. Zhang Z, Reinikainen J, Adeleke KA, Pieterse ME, Groothuis-Oudshoorn CGM. Time-varying covariates and coefficients in Cox regression models. *Ann Transl Med.* 2018;6(7):121.
22. Parsons L. Reducing Bias in a Propensity Score Matched-Pair Sample Using Greedy Matching Techniques. 2001.
23. Schneeweiss S. Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. *Clin Epidemiol.* 2018;10:771-788.
24. Bross ID. Spurious effects from an extraneous variable. *J Chronic Dis.* 1966;19(6):637-647.
25. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med.* 2017;167(4):268-274.
26. Haneuse S, VanderWeele TJ, Arterburn D. Using the E-Value to Assess the Potential Effect of Unmeasured Confounding in Observational Studies. *JAMA.* 2019;321(6):602-603.
27. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol.* 2008;167(4):492-499.
28. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol.* 2015;11(7):437-441.
29. Mi X, Hammill BG, Curtis LH, Greiner MA, Setoguchi S. Impact of immortal person-time and time scale in comparative effectiveness research for medical devices: a case for implantable cardioverter-defibrillators. *J Clin Epidemiol.* 2013;66(8 Suppl):S138-144.
30. Berlin NL, Cartmel B, Leffell DJ, Bale AE, Mayne ST, Ferrucci LM. Family history of skin cancer is associated with early-onset basal cell carcinoma independent of MC1R genotype. *Cancer Epidemiol.* 2015;39(6):1078-1083.

31. Kallini JR, Hamed N, Khachemoune A. Squamous cell carcinoma of the skin: epidemiology, classification, management, and novel trends. *Int J Dermatol.* 2015;54(2):130-140.
32. Leonardi-Bee J, Ellison T, Bath-Hextall F. Lifestyle factors of smoking, BMI and alcohol on the risk of Non-Melanoma Skin Cancer in adults: a systematic review. *JBI Libr Syst Rev.* 2012;10(6):352-398.
33. Leonardi-Bee J, Ellison T, Bath-Hextall F. Smoking and the risk of nonmelanoma skin cancer: systematic review and meta-analysis. *Arch Dermatol.* 2012;148(8):939-946.
34. Song F, Qureshi AA, Gao X, Li T, Han J. Smoking and risk of skin cancer: a prospective analysis and a meta-analysis. *Int J Epidemiol.* 2012;41(6):1694-1705.
35. Morales DR, Pacurariu A, Slattery J, Kurz X. Association between hydrochlorothiazide exposure and different incident skin, lip and oral cavity cancers: A series of population-based nested case-control studies. *Br J Clin Pharmacol.* 2020;86(7):1336-1345.

12. APPENDICES

Appendix A Code for antihypertensive agents exposure

Class	Drug	ATC code
Cohort A: Hydrochlorothiazide	Hydrochlorothiazide	C03AA03
	HCTZ + angiotensin converting enzyme inhibitors	C09BA02
	HCTZ + angiotensin receptor blockers	C09DA01, C09DA03, C09DA04, C09DA06, C09DA07, C09DA08
	HCTZ + calcium channel blockers and angiotensin receptor blockers	C09DX01, C09DX03
	HCTZ + potassium sparing diuretics	C03EA01
	HCTZ + renin inhibitor and calcium channel blockers	C09XA52
	HCTZ + vasodilator	C02LA01, C02LA51
Cohort B: Thiazide diuretics other than HCTZ	Benzylhydrochlorothiazide	C03AA
	Chlorthalidone + beta blockers	C07CB03, C07BB03
	Chlorthalidone + angiotensin receptor blockers	C09DA09
	Indapamide	C03BA11
	Indapamide + angiotensin converting enzyme inhibitor	C09BA04
	Metolazone	C03BA08
	Trichlormethiazide	C03AA06
Cohort C: Diuretics other than thiazide	Loop diuretics	C03CA
	Potassium sparing diuretics	C03D
Cohort D: Other antihypertensive agents other than diuretics	Alfa-1 blockers	C02CA, G04CA03
	Angiotensin converting enzyme inhibitors	C09A
	Angiotensin converting enzyme inhibitors + calcium channel blockers	C09BB
	Angiotensin receptor blockers	C09C
	Angiotensin receptor blockers + calcium channel blockers	C09DB
	Beta blockers	C07A
	Calcium channel blockers	C08C (Exclude C08CA06, Nimodipine), C10BX03
	Centrally acting drugs	C02A
	Direct vasodilators	C02D
	Renin inhibitor	C09XA02

Appendix B Code for exclusion criteria definition

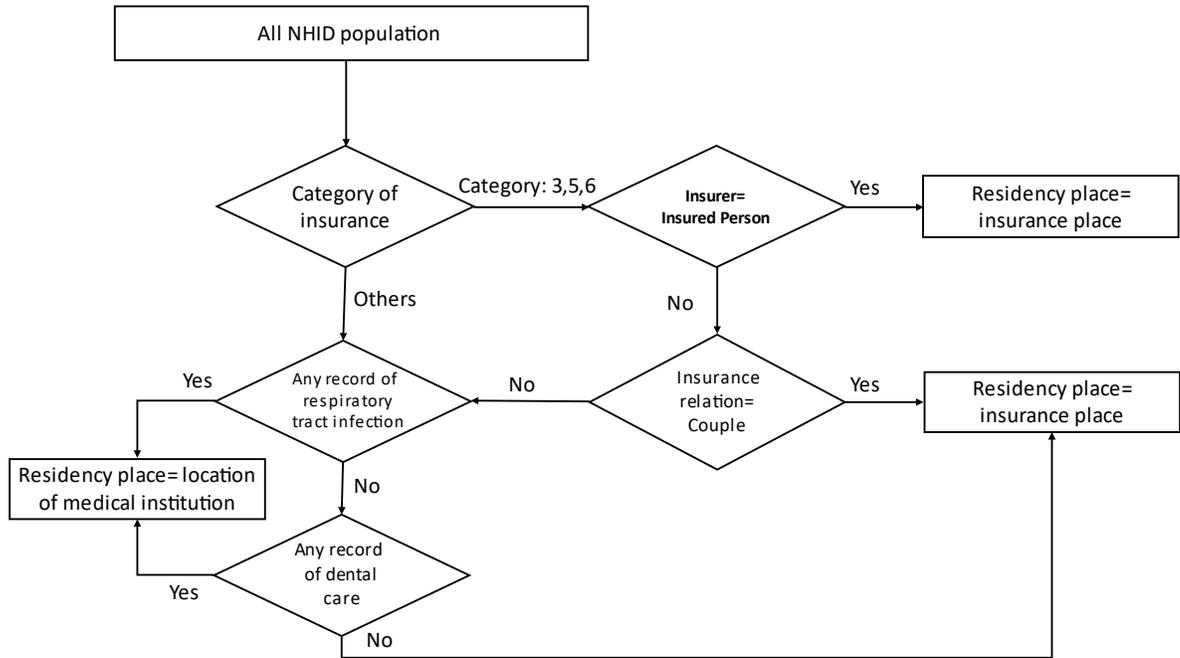
Disease	ICD-9	ICD-10
Any kind of cancer	140-239	C00-C97, D04
HIV infection	042	B20
Organ transplantation	V42	Z94
Systemic lupus erythematosus	710	M32
Rheumatoid arthritis	714	M05, M06
Chronic kidney disease	585	N18
Immunosuppressant	ATC code	
Azathioprine	L04AX01	
Corticosteroid	H02AA02, H02AB01, H02AB02, H02AB04, H02AB06, H02AB10, H02BX	
Cyclophosphamide	L01AA01	
Cyclosporine	L04AD01	
Everolimus	L04AA18, L01XE10	
Mycophenolate	L04AA06	
Methotrexate	L01BA01	
Sirolimus	L04AA10	
Tacrolimus	L04AD02	

APPENDIX C Covariates of baseline characteristic

Variable	Code types	Code	
Age	continuous		
Sex	categorical		
Income level	categorical		
Urbanization levels	categorical		
Health resource utilization			
ER visits	continuous		
Outpatient visits	continuous		
Hospitalizations	continuous		
Comorbid conditions		ICD9	ICD10
Asthma	Dichotomous	493	J45
Bipolar disease	Dichotomous	296	F0631, F31
Chronic obstructive pulmonary disease	Dichotomous	491, 492,496	J43, J44
Congestive heart failure	Dichotomous	428	I50
Coronary artery disease	Dichotomous	410, 411, 412, 413, 414	I20, I21, I22, I23, I24, I25
Cataract	Dichotomous	366	H25, H26, H27, H28
Depression	Dichotomous	2962, 2963, 3004, 30112, 3090, 3091, 311	F32, F33, F341, F432
Dementia	Dichotomous	290, 2940, 2941, 3310, 3311, 3312, 33182, 33183, 33189	F00, F01, F02, F03, G30, G310, G311, G3183, G3184, G319
Diabetes mellitus	Dichotomous	250	E10, E11, E13, E14
Glaucoma	Dichotomous	365	H40
Hemorrhagic stroke	Dichotomous	430, 431, 432, 436	I60, I61, I62, I64, I690, I691, I692, I694
Hypertension	Dichotomous	401, 402, 403, 404, 405	I10, I11, I12, I13, I15
Ischemic stroke	Dichotomous	433, 434, 436, 4371	I63, I64, I693, I694
Osteoporosis	Dichotomous	7330	M80, M81, M82
Peptic ulcer	Dichotomous	530, 531, 532, 533, 534	K221, K25, K26, K27, K28
Parkinsonism	Dichotomous	332	G20, G21
Pterygium	Dichotomous	3724	H110
Schizophrenia	Dichotomous	295,30122	F20, F21, F25
Geographic area of residence	Categorical	North, Central, South and East, by zip codes	
Concomitant antihypertensive agents	Dichotomous	See appendix F	
Medication history		ATC code	

Aspirin	Dichotomous	B01AC06, N02BA01
Amiodarone	Dichotomous	C01BD01
Chlorpromazine	Dichotomous	N05AA, N05AA01
Dronedarone	Dichotomous	C01BD07
Fenofibrate	Dichotomous	C10AB05, C10BA03
Flupentixol	Dichotomous	N05AF01, N06CA02
Griseofulvin	Dichotomous	D01BA01
Hydroxychloroquine	Dichotomous	P01BA02
Isoniazid	Dichotomous	J04AC01, J04AM02, J04AM05, J04AM06, J04AM07
Macrolide	Dichotomous	J01FA01, J01FA09, J01FA10
Nonsteroidal anti-inflammatory drug	Dichotomous	M01A
Quinolones	Dichotomous	J01MA01, J01MA02, J01MA06, J01MA12, J01MA14, J01MA15
Quinidine	Dichotomous	C01BA01
Retinoid	Dichotomous	D10AD01, D10BA01, L01XX14
Statin	Dichotomous	C10AA (The 1~4th level of ATC code), C10BA01, C10BA02, C10BA03, C10BA05, C10BX03
Sulfasalazine	Dichotomous	A07EC01
Tetracycline	Dichotomous	J01AA02, J01AA03, J01AA06, J01AA07, J01AA08
Thioridazine	Dichotomous	N05AC02
Trimethoprim/Sulfamethoxazole	Dichotomous	J01EA01 (TMP), J01EC01 (SMX), J01EE01
Voriconazole	Dichotomous	J02AC03

Appendix D Algorithm to identify patient’s residency code.



- ✧ Category 3,5,6: Patient’s NHID identity belongs to nursery, accountant, lawyer, architect, physician, dentist, Chinese medical physician, or family member attached on the insurer (i.e., daughter, wife...)
- ✧ Code for respiratory tract infection: ICD-9: 460-466, 480-487; ICD-10: J00-J22
- ✧ Code for dental care: ICD-9:520-525; ICD-10: K00-K08

Appendix E Codes for skin cancer topography definition

ICD-O-3 (CASITE)	ICD-O-3 Topography
C44-	C44 SKIN
C440	C44.0 Skin of lip, NOS
C441	C44.1 Eyelid
C442	C44.2 External ear
C443	C44.3 Skin of other and unspecified parts of face
C444	C44.4 Skin of scalp and neck
C445	C44.5 Skin of trunk
C446	C44.6 Skin of upper limb and shoulder
C447	C44.7 Skin of lower limb and hip
C448	C44.8 Overlapping lesion of skin
C449	C44.9 Skin, NOS

Appendix F Codes for squamous cell carcinoma morphology

M-code (HISTBEH)	ICD-O-3 Morphology
80502	Papillary carcinoma in situ
80503	Papillary carcinoma, NOS
80513	Verrucous carcinoma, NOS
80513	Verrucous epidermoid carcinoma
80513	Verrucous squamous cell carcinoma
80522	Papillary squamous cell carcinoma, non-invasive
80522	Papillary squamous cell carcinoma in situ
80523	Papillary squamous cell carcinoma
80523	Papillary epidermoid carcinoma
80543	Warty carcinoma
80543	Condylomatous carcinoma
80543	Warty-basaloid carcinoma
80702	Squamous cell carcinoma in situ, NOS
80702	Epidermoid carcinoma in situ, NOS
80702	Intraepidermal carcinoma, NOS
80702	Intraepithelial squamous cell carcinoma
80703	Squamous cell carcinoma, NOS
80703	Squamous cell carcinoma, usual type
80703	Epidermoid carcinoma, NOS
80703	Squamous carcinoma
80703	Squamous cell epithelioma
80713	Squamous cell carcinoma, keratinizing, NOS
80713	Epidermoid carcinoma, keratinizing
80713	Squamous cell carcinoma, large cell, keratinizing
80713	Keratoacanthoma
80723	Squamous cell carcinoma, large cell, nonkeratinizing, NOS
80723	Epidermoid carcinoma, large cell, nonkeratinizing
80723	Squamous cell carcinoma, nonkeratinizing, NOS
80733	Squamous cell carcinoma, small cell, nonkeratinizing
80733	Epidermoid carcinoma, small cell, nonkeratinizing
80743	Squamous cell carcinoma, spindle cell
80743	Epidermoid carcinoma, spindle cell
80743	Squamous cell carcinoma, sarcomatoid
80743	Pseudovascular squamous cell carcinoma
80753	Squamous cell carcinoma, adenoid
80753	Squamous cell carcinoma, acantholytic
80753	Squamous cell carcinoma, pseudoglandular
80762	Squamous cell carcinoma in situ with questionable stromal invasion
80762	Epidermoid carcinoma in situ with questionable stromal invasion
80763	Squamous cell carcinoma, microinvasive
80783	Squamous cell carcinoma with horn formation
80802	Queyrat erythroplasia
80812	Bowen disease
80812	Intraepidermal squamous cell carcinoma, Bowen type
80833	Basaloid squamous cell carcinoma
80833	Papillary-basaloid carcinoma
80843	Squamous cell carcinoma, clear cell type
80853	Squamous cell carcinoma, HPV-positive
80863	Squamous cell carcinoma, HPV-negative

Appendix G Codes for basal cell carcinoma morphology

M-code (HISTBEH)	ICD-O-3 Morphology
80903	Basal cell carcinoma, NOS
80903	Basal cell epithelioma
80903	Rodent ulcer
80903	Pigmented basal cell carcinoma
80903	Basal cell carcinoma with adnexal differentiation
80913	Multifocal superficial basal cell carcinoma
80913	Multicentric basal cell carcinoma
80913	Superficial basal cell carcinoma
80923	Infiltrating basal cell carcinoma, NOS
80923	Infiltrating basal cell carcinoma, non-sclerosing
80923	Infiltrating basal cell carcinoma, sclerosing
80923	Basal cell carcinoma, desmoplastic type
80923	Basal cell carcinoma, morpheic
80923	Basal cell carcinoma, sarcomatoid
80933	Basal cell carcinoma, fibroepithelial
80933	Fibroepithelioma, NOS
80933	Fibroepithelioma of Pinkus type
80933	Fibroepithelial basal cell carcinoma, Pinkus type
80933	Pinkus tumor
80943	Basosquamous carcinoma
80943	Mixed basal-squamous cell carcinoma
80953	Metatypical carcinoma
80973	Basal cell carcinoma, nodular
80973	Basal cell carcinoma, micronodular
80983	Adenoid basal carcinoma

Appendix H Codes for malignant melanoma morphology

M-code (HISTBEH)	ICD-O-3 Morphology
87202	Melanoma in situ
87203	Malignant melanoma, NOS
87203	Melanoma, NOS
87203	Nevoid melanoma
87213	Nodular melanoma
87223	Balloon cell melanoma
87233	Malignant melanoma, regressing
87283	Meningeal melanomatosis
87303	Amelanotic melanoma
87403	Malignant melanoma in junctional nevus
87412	Precancerous melanosis, NOS
87413	Malignant melanoma in precancerous melanosis
87422	Lentigo maligna
87422	Hutchinson melanotic freckle, NOS
87423	Lentigo maligna melanoma
87423	Malignant melanoma in Hutchinson melanotic freckle
87433	Superficial spreading melanoma
87433	Low cumulative sun damage melanoma
87443	Acral lentiginous melanoma, malignant
87443	Acral melanoma
87453	Desmoplastic melanoma, malignant
87453	Desmoplastic melanoma, amelanotic
87453	Neurotropic melanoma, malignant
87463	Mucosal lentiginous melanoma
87613	Malignant melanoma in giant pigmented nevus
87613	Malignant melanoma in congenital melanocytic nevus
87703	Mixed epithelioid and spindle cell melanoma
87703	Malignant Spitz tumor
87703	Spitz melanoma
87703	Spitzoid melanoma
87713	Epithelioid cell melanoma
87723	Spindle cell melanoma, NOS
87733	Spindle cell melanoma, type A
87743	Spindle cell melanoma, type B