

## 台湾の保健福祉データベース

## —ビッグデータ分析による高品質なリアルワールドエビデンスを得る機会—

The Health and Welfare Database of Taiwan

— Opportunities for High Quality Real-World Evidence from Big Data Analyses —

Cheng Yang Hsieh<sup>\*1,\*2</sup>, Chien-Chou Su<sup>\*1</sup>, Shih-Chieh Shao<sup>\*1,\*3</sup>, Swu-Jane Lin<sup>\*4</sup>,  
Yea-Huei Kao Yang<sup>\*1</sup> and Edward Chia-Cheng Lai<sup>\*1,\*6</sup>

## Summary

台湾の医療保険請求に基づく研究用データベース (National Health Insurance Research Database: NHIRD) は、医療調査のための人口レベルのデータソースの優れた事例であり、臨床的な判断や医療政策の策定を支援するリアルワールドエビデンスを生み出している。最近、台湾の厚生省 (衛生福利部, Ministry of Health and Welfare: MOHW) は、NHIRD 及びその他 70 以上の医療関連データベース (国民健康聞き取り調査など) を中央で集中管理するためのデータリポジトリサイトである保健福祉データ (Health and Welfare Data: HWD) センターを設立した。HWD は、NHIRD と他のデータベースのリンケージによって、研究の幅を広げる機会を作り出している。すなわち研究者は、NHIRD で利用可能な変数よりも多くの変数、例えば検査データ又は身体検査結果などにアクセスすることができる。この情報は、研究の質と能力を大幅に向上させる可能性がある。一方で、データリンケージに当たり、被

験者から得られたデータの機密性とプライバシー保護は、重要な問題である。データの機密性を犠牲にすることなく、データリンケージを行うための方法論的及び統計的アプローチを開発する必要がある。

NHIRD はその代表性と包括的な情報によって、台湾人のリアルワールドエビデンス (RWE) を得るための最も重要な人口ベースのデータ源となっている。更に、NHIRD は、医療や治療及びアウトカムに関する国や人種による違いを比較する国際共同研究の機会も提供する。台湾の HWD センターは、現在のどのデータベースにも含まれていない重要な情報を入手するために、積極的なデータ収集を主張し、設計し、実施するよう努力すべきである。更に、良い研究者を育成するために必要な疫学と統計的知識を備えた医療データベースの使用に関する教育や訓練にも労力を費やす必要がある。NHIRD の過去の成功経験によって、われわれは HWD センターがいつか、単なるデータ置き場から強力な研究エンジンに変身し、さまざまな情報源からのデータを、将来の台湾での大規模データ分析によって裏

\*1 School of Pharmacy, Institute of Clinical Pharmacy and Pharmaceutical Sciences, College of Medicine, National Cheng Kung University, No.1, University Road, Tainan City 701, Taiwan

\*2 Department of Neurology, Tainan Sin Lau Hospital, Tainan, Taiwan

\*3 Department of Pharmacy, Chang Gung Memorial Hospital, Keelung, Taiwan.

\*4 Department of Pharmacy Systems, Outcomes & Policy, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois, USA

\*5 Department of Neurology, Tainan Sin Lau Hospital, Tainan, Taiwan

\*6 Department of Pharmacy, National Cheng Kung University Hospital, Tainan, Taiwan

Chien-Chou Su と Cheng Yang Hsieh は、この論文に等しく貢献した。

付けられた、リアルワールドデータによって医学を導ける“光”へと変換できることを願っている。

## 1. 緒言

高齢化と医療費の増加は、世界中の医療制度にとって大きな負担となり、台湾も例外ではない<sup>1)</sup>。台湾の医療資源には制約があるとはいえ、そのような負担を軽減する機会がある。例えば、医療介入を特定するために利用可能な最良のデータと専門知識を利用することで、結果を改善し、費用の増加を相殺することができる。NHIRDは、医学研究のための人口レベルのデータソースの良い例となり、臨床決定と医療政策の策定を支援するエビデンスを生み出す<sup>2-5)</sup>。データ科学技術の急速な普及と迅速な取り込みにより、NHIRDは台湾においてビッグデータ分析(BDA)を実施し、RWEを得るための基礎を築いている。RWEが支持する費用対効果の高い介入を採用することで、医療システムに対する経済的負担を増大させることなく、患者の成果を改善することが可能になる<sup>6)</sup>。

しかし、無作為化比較試験とは異なり、NHIRDなどの電子的な医療データベースを用いた後ろ向き研究では、妥当性が確認されていない診断や治療のコード化、疾患の重症度に関する情報の欠如、測定されていない交絡因子など、多くの共通の課題に直面する可能性がある<sup>7)</sup>。これらの限界に対処する一つの方法は、主要データベースで利用できない外部データを取得するために、請求データベース、疾患/投薬監視データ及びレジストリなどの複数のデータベースをリンクすることである<sup>8)</sup>。しかしリンケージにあたっては、複数のデータベースのデータ品質と統合の容易さがまた別の懸念となる。

最近、MOHWは、NHIRDと他の医療関連データベース(国民健康聞き取り調査など)を、中央で集中管理するためのデータリポジトリサイトであるHWDセンターを設立し、分析の能力を高めた。このほか、データのプライバシーを損なうことのないいくつかの分析的アプローチと方針が展開された。HWDは、複数のデータベースを利用して研究を実施するための強固な基盤を提供している。

本稿では、NHIRDをその歴史、構造、内容から、それが他のデータベースとともにHWDにどのように集約されたかについて紹介する。また、論文文化されたNHIRD研究を要約し、論文公表が臨床診療に及ぼす影響を評価する。更に、NHIRD研究の強みと限界を検討し、可能な方法論的解決策を提案した。最後に、BDAを促進し、台湾で高品質のRWE BDAを抽出するためのHWDの将来の機会と役割について、われわれの見解を示す。

## 2. 台湾の全国請求データベースの誕生

1995年3月1日、台湾は単一保険制度(“single-payer”)を開始し、国民健康保険(NHI)プログラムが義務付けられた。2014年までに、国民のほぼ99.99%が加入している。したがって、NHIプログラムの行政管理に使用されるデータベースは、台湾の人口のほぼ全てをカバーし、台湾のNHIRDの情報源であった<sup>9)</sup>。

2002年、台湾の国家衛生研究院(NHRI)は、研究目的でNHIRDを利用可能にするために、NHIRDの設立と維持を開始した。NHRIがまとめた情報には、対象者の人口統計学的特性、医療従事者、施設及び入院患者、外来ケア、契約した薬局からのサービス請求が含まれていた。NHIRDの主要なデータファイルの例は、来院ごとの外来診療費、入院ごとの入院費、契約された薬局での処方費、外来診療及び入院患者のオーダーの詳細、調剤の詳細、及び登録期間や保険料などの保険関連情報を含む対象者の基本情報であった。

NHIRDは、2015年より前の診断記録に対しては国際疾病分類第9版、臨床修正版(ICD-9-CM)、それ以降は第10版(ICD-10)を使用した。特別に開発されたシステムである薬価基準収載医薬品コード(National Health Insurance Drug Codes)を医薬品の記録に使用した。この薬品コードシステムは階層的に構成されていないため、多くの研究者は研究を行うときに、WHO解剖治療化学分類法(ATCコード)などの他の国際コードシステムにコードをマッピングした。

NHIRDは代表性があり、かつ、包括的であったため、台湾人のRWEを得るための最も重要な人口ベースのデータ源となった。論文審査のある専門誌で公開されているNHIRD研究の数は2,000以上に達していることが分かった。更に、NHIRDは、医療や治療、アウトカムに関する国や人種間差を比較する国際的な共同研究の機会を提供している。NHIRDを含むいくつかの国のデータベースを用いた国際的なデータベース研究が発表されている<sup>10-13)</sup>。NHIRDはまた、国際比較研究を実施するためのアジア薬剤疫学ネットワーク(AsPEN)の中核データ資源の一つとなっている。AsPENは、アジアにおける国際研究を支援する国際的なイニシアチブである。AsPENの詳細は他を参照されたい<sup>14)</sup>。

被験者及び被験者から得られたデータを含む全ての研究において、データの機密性及びプライバシーは重要事項である<sup>15)</sup>。患者のプライバシーを保護するために、NHRIは患者及び医療提供者(医療機関を含む)の識別子を暗号化した。しかし、研究上必要であれば、研究者はNHIRD内のファイル結合のために、NHRIによって作成された一

意的(ユニークな)かつ匿名の識別子を使用することもできる。NHRI はまた、他の制約によりデータベースの使用を制限した。例えば、NHIRD を使用する全ての申請者は、台湾の国民でなければならず、データの使用は研究に限ると規定された。全ての申請は、使用の合理性を保証するために専門家により審査された。研究者は、申請中、台湾のコンピューター処理個人データ保護法及び関連する規則に従い、患者又は介護者のプライバシーを潜在的に侵害する可能性のある情報を得ようとする意図がないと宣言する契約書に署名する必要がある。2012年以降、申請者はNHIRD に研究を依頼する前に、施設内治験審査委員会(IRB)の承認を受けなければならない。

### 3. NHIRD へのアクセス可能性

前述のデータ機密性の保護戦略以外に、NHRI は研究者の要求するデータ量を台湾の人口の絶対数の10%以下に制限した。すなわち、研究者は、最大で約230万人の個人データしか要求できなかった。したがって、研究者は、NHIRD のデータ抽出を申請する前に、疾患の有病率及びサンプルサイズを考慮しなければならない。症例数が230万人を超える可能性のある有病率の高い疾病、例えば糖尿病では、研究者は、NHIRD 内の糖尿病患者全員からの無作為標本の選択について検討する必要がある。疾病に特異的なデータベースに加えて、NHRI は、2000年、2005年、及び2010年に、200万人の対象者を無作為にそれぞれ元のNHIRD からサンプリングした三つの時系列健康保険データベース(LHID2000, LHID2005 及び LHID2010)を提供している。LHID には、サンプリングされた個人の1997年から最新の請求データまでが含まれていた。LHID の代表性は、NHRI によって検証された<sup>16)</sup>。

### 4. 台湾におけるビッグデータ分析の機会：保健福祉データベース

NHIRD は行政目的のために作成され、政府省庁によって管理されているが、医療研究に対するNHIRDの使用及びデータの所有権が、最近大きな議論を巻き起こしている。いくつかの市民団体は、データプライバシーの懸念から、MOHW によるNHIRDの使用に対して訴訟を起こした。裁判所の最終判決はNHIRDの使用を禁じていないが、NHIRDの法的側面に関する議論はまだ進行中である。

健康データの保護を強化するために、MOHW は、大規模データレポジトリ、HWD センター、疾病管理センターなどの異なる部門からの健康データベースを集中化した。HWD センターは、より厳重な監視下でNHIRDなどのデー

タベースを管理する。2015年11月以降、研究者はHWD センターを訪れ、MOHW サーバーへのリモート接続を介してHWD にアクセスすることによってオンサイト分析を行う必要がある。電子装置はセンターに持ち込むことはできず、個人レベルのデータは持ち出せない。センターから持ち出される全ての分析結果は、患者の再特定を防止するためにデータ管理者によって審査されなければならない。

オンサイトでの解析の規制によって解析は不便になり、研究の所要時間とコストは増大するが、HWD はNHIRD と他のデータベースを結びつけて研究の幅を広げる機会を作り出している。これは台湾のBDAの基礎を提供する。現在、HWD には70以上のデータベースがあり、三つのカテゴリに分類できる：

- (A) 国民健康聞き取り調査やがんのスクリーニングデータを含む一次的(直接収集)な行政データ。
- (B) 台湾の加齢に関する長期的研究や高齢者状態調査の報告書からのデータなどの二次的な調査データ。
- (C) 人口ベースの請求データ(すなわち、NHIRD)及びその無作為に選択されたデータ(すなわち、LHID)。

HWD で集中管理されているデータベースの詳細な特徴は、HWD センターの公式ウェブサイトで公開されている<sup>17)</sup>。Table 1 に情報を要約した。

NHIRD は全国の住民ベースのデータを提供するため、暗号化されていない個人識別番号(PIN)を用いて他のデータベースとリンクする際には、その中心のデータソースとして採用されることが多い。例えば、PINを用いてNHIRD と国民健康聞き取り調査とをリンクさせ、患者の社会経済的地位や生活習慣などの詳細を取得することができる。更に、医療提供者のPINを使用して、NHIRD と認定専門医のレジストリとをリンクして、医療提供者の特性を取得したり、病院の識別番号を使用してNHIRD と施設データベースをリンクして、追加のデータを得ることができるかもしれない。Fig. 1 は、HWD 内のデータベース間の相互リンクの概略図である。複数のデータベースを組み合わせることで、研究者は、検査データ、身体検査結果、障害のレベル、生活の質、BMI、喫煙、配偶者の有無、教育、家庭収入など、NHIRD で利用可能な変数よりも多くの変数にアクセスすることができる。この情報は、研究の質と能力を大幅に向上させる可能性がある。

HWD センターでの集中管理により、データベース間のリンク能力が高められ、同時にデータ機密性の管理が強化される。BDAの時代に、HWD は、臨床及び政策の意思決定を支援する高品質のRWEを生み出す研究を促進するのに重要な役割を果たすだろう。近い将来、より多くの全国のサーベイランスデータベースと病院ベースの電子カル

**Table 1** 保健福祉データベースセンターのデータベース

A. 一次行政データ				
出生、死亡及び母性データ	疾患及び損傷	スクリーニング／省庁横断	福祉／社会	調査データ
<ul style="list-style-type: none"> <li>出生証明書</li> <li>死因データ</li> <li>複数の死因データ</li> <li>母子保健データベース</li> </ul>	<ul style="list-style-type: none"> <li>台湾がんレジストリ</li> <li>台湾がんレジストリ年次報告書</li> <li>人工生殖データ</li> <li>希少疾病データ</li> <li>18 件の報告及び調査データ</li> <li>交通事故データ</li> </ul>	<ul style="list-style-type: none"> <li>がんスクリーニングデータ(乳がん, 子宮頸がん, 口腔がん, 大腸がん)</li> <li>台湾先住民(出生証明書, 死因データ, 戸籍)</li> </ul>	<ul style="list-style-type: none"> <li>障害者集団プロフィール</li> <li>低所得者及び中低所得者世帯</li> <li>家庭内暴力データ</li> <li>小児・青少年保護の報告書データ</li> <li>性的暴行の報告書データ</li> </ul>	<ul style="list-style-type: none"> <li>国民健康聞き取り調査</li> <li>高血圧, 高血糖及び高脂血症調査</li> <li>台湾出生コホート試験データ</li> <li>避妊の知識, 避妊に関する態度及び実践</li> <li>在宅ケア助成金利用状況調査の報告書</li> <li>台湾の高齢者福祉団体の必要条件調査の報告書</li> <li>他の 12 の関連調査データ</li> </ul>
B. 二次調査データ				
保健行動		福祉／社会		健康／ライフスタイル
<ul style="list-style-type: none"> <li>全体的な学生健康調査</li> <li>成人喫煙行動サーベイランス調査ファイル</li> <li>全体的な青少年タバコ調査行動</li> <li>行動危険因子サーベイランスシステム</li> </ul>		<ul style="list-style-type: none"> <li>台湾における若年状況調査</li> <li>高齢者状況調査の報告</li> <li>一人親家庭状況調査</li> <li>身体的及び精神的障害者生活及び需要評価調査</li> <li>在宅ケア助成金利用者状況調査の報告書</li> <li>低所得者及び中低所得者家庭生活状況調査</li> <li>女性生活状況調査</li> </ul>		<ul style="list-style-type: none"> <li>台湾の加齢に関する長期的研究</li> </ul>
C. 医療保険請求に基づく研究用データベース (NHIRD)				時系列健康保険データベース (LHID)
<ul style="list-style-type: none"> <li>来院による外来診療費</li> <li>入院による入院費</li> <li>契約した薬局で調剤された処方箋の費用</li> <li>外来診療オーダーの詳細</li> <li>入院患者オーダーの詳細</li> <li>契約した薬局で調剤された処方箋の詳細</li> <li>医療施設の保健サービス利用</li> </ul>		<ul style="list-style-type: none"> <li>契約医療施設のレジストリ</li> <li>有資格専門家のレジストリ</li> <li>医療施設の認定プロフィール</li> <li>医療関係者のレジストリ</li> <li>保険受取人のレジストリ</li> <li>難病の登録</li> </ul>		<ul style="list-style-type: none"> <li>時系列健康保険データベース 2000</li> <li>時系列健康保険データベース 2005</li> <li>時系列健康保険データベース 2010</li> </ul>

注：2000 年，2005 年及び 2010 年の各 200 万人の無作為の受取人データは，それぞれ元の NHIRD からのものである。LHID には，サンプルされた人々の 1997 年から最新の請求データまでが含まれる。

テ情報が，HWD に追加される予定である。データの量と質の継続的な増加と分析ツールの迅速な進歩により，将来，利害関係者と研究者は，データにアクセスして使用する際の倫理的及び法的要件を維持する上でより大きな責任を負う可能性が高くなる。

## 5. NHIRD による「情報漁り」を減らすための警告と対策

NHIRD を用いた観察研究は，診断コードの妥当性や測定されていない交絡因子など，いくつかの課題に直面する可能性がある。しかし，NHIRD 又はその他の大規模なデータベースを医療研究に使用する際の基本的な課題は，相関をみるに値するような仮説，又は生理学的・生態学的メカニズムの欠如，そして臨床的，政策的又は科学的意義であった<sup>18)</sup>。

仮説によらない研究はデータベース関連の研究に限定されないが，コンピューター技術の進歩とデータベース研究が比較的低コストであることが，かなりの量の「情報漁

り」型の研究をもたらしたことは否定できず，これは医療研究の BDA に害を与えている可能性がある。最近，European Journal of Internal Medicine に掲載された短報は，NHIRD を使っておそらく同じモデルに基づいて 150 の論文を発表した 1 人の研究者による研究の妥当性を批判している<sup>19)</sup>。その結果，いくつかのジャーナルは現在，NHIRD に基づく研究を検討の対象としていない。多くの NHIRD ベースの研究は，高品質で影響力の強いエビデンスを提供し，一流のジャーナルに発表されており，われわれはこれを残念な反応と考えている<sup>2-5)</sup>。あらゆる研究は科学的調査の全ての要件を満たすべきであるが，データベース研究は観察研究ではバイアスの可能性が高いことから，更に高い基準で実施されるべきであるということに改めて気づかされる<sup>20,21)</sup>。

## 6. NHIRD の検証

NHIRD は保険償還目的で作成された NHI データから作成されているため，研究実施のためのデータの妥当性を

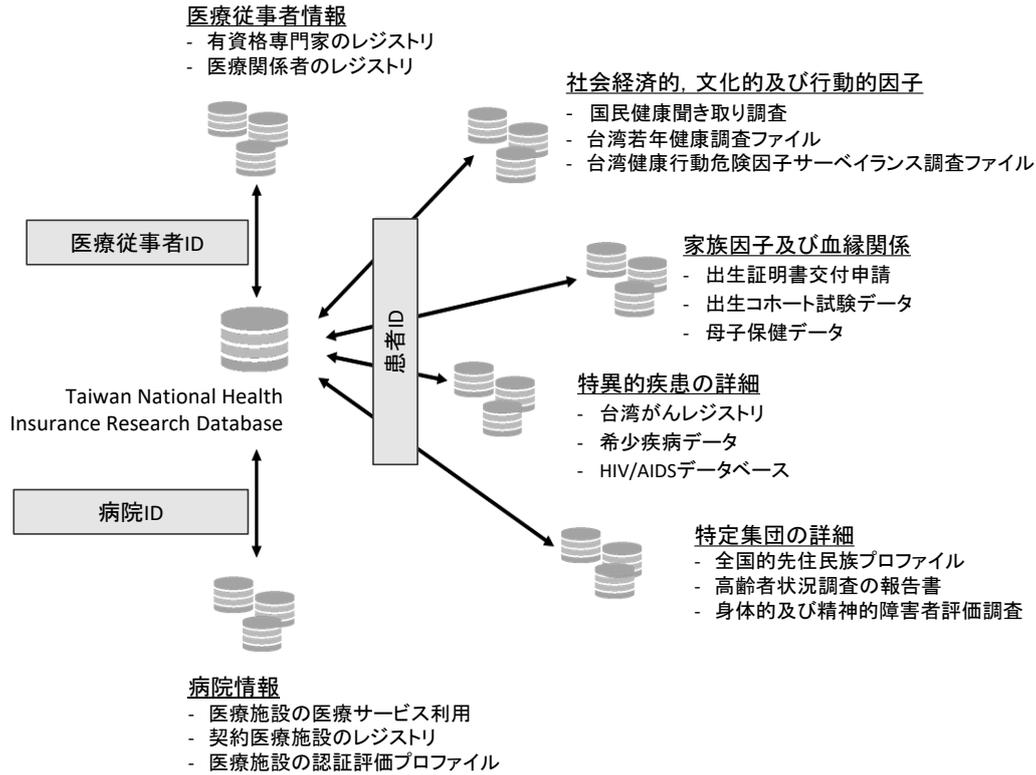


Fig. 1 HWD 内のデータベースの相互リンクの概念説明

評価する必要がある。NHI による保険償還拒否を避けるために、一部の病院や医療提供者は、請求書を提出する際に、診断をより深刻なものに「アップコーディング」することを選択するかもしれない。そのため、NHI データから得られた統計は、臨床の現実とは比較できない可能性がある。例えば、台湾の年あたりの脳卒中イベントは、NHIRD を用いて 80,000 人(人口 1000 人あたり約 3.5 人)であると報告されている<sup>22)</sup>。しかし、この発生率は、実際よりも 2 倍以上高く、過大評価されている可能性がある<sup>23)</sup>。理由の一つとして考えられることとして、脳卒中によく似た症例(例えば、めまい、頭痛又は意識障害)を脳卒中の診断に「アップコーディング」している可能性が挙げられる。

NHIRD における診断コードの妥当性を評価するためのバリデーション研究がいくつか実施されている。これらの研究及び特定の疾患の妥当性が確認された診断コードの一覧を Table 2 に要約した。妥当性が確認されたコードの多くは、比較的重症な疾患のためのものであり、感度及び陽性的中率が高い。これら全ての研究において、NHIRD より多くの臨床情報を備えたデータベースを基準として使用している。例えば、Chen らは、てんかんの診断コードを検証するために、病院のカルテを検討した<sup>24)</sup>。Hsieh らは、台湾脳卒中レジストリを用いて、NHIRD の虚血性脳卒中の診断コードを検証した<sup>25)</sup>。Sung と Hsieh らは、腎機能

障害、高血圧、糖尿病、高脂血症、冠動脈疾患、心房細動などの脳卒中患者のいくつかの血管危険因子に対して一連のバリデーション研究を行った<sup>26)</sup>。NHIRD における急性心筋梗塞及び脳卒中患者の院内死亡率は、医療センターの電子カルテからの死亡記録と相互比較することで検証されている<sup>27)</sup>。これらの研究及び検証されたコードは、NHIRD の患者及び転帰を特定するための重要なツールを提供する。台湾では、重度の疾患を有する患者は、難病の認定を申請することができ、各医療受診の際の一定の NHI の支払い及び自己負担金を免除される。難病の証明のための全ての申請は専門家が審査するため、診断は非常に正確であると考えられ、難病のファイルはそれぞれの研究における症例確認に使用される。難病のリストは、Table 3 に示したように公開されている。

## 7. 測定されていない変数又は利用できない変数を克服する方法

もともと研究目的で作成されていない医療データベースには、研究者が必要とする全てのデータが含まれているわけではなく、NHIRD もこの例外ではない。例えば、疾病重症度の指標は、しばしば NHIRD には欠けている。これは、相関の推定にバイアスを与える可能性がある。ベースライン時の疾患の重症度は、医療研究における最も重要

**Table 2** NHIRD における診断コードの妥当性を検証する研究のまとめ

疾患／状態	ICD-9-CM コード	重症度, %	陽性的中率, PPV %	注釈	参考文献
急性虚血性脳卒中	433.xx, 434.xx	94.5	97.9	参照基準として神経学専門家によるカルテ審査	<i>Pharmacoepidemiol Drug Saf</i> , 2011 <sup>44)</sup>
てんかん	345.xx	81.4	76.8	参照基準として神経学専門家によるカルテ審査, 特異性: 99.8%	<i>Epilepsia</i> , 2012 <sup>24)</sup>
急性心筋梗塞	410.xx	88.0	92.0	参照基準として神経学専門家によるカルテ審査	<i>J Epidemiol</i> , 2014 <sup>45)</sup>
腎機能障害	250.4, 283.11, 403.x, 404.x, 580-589, 753.0, 753.1	38.4	76.0	参照基準として台湾脳卒中レジストリ, 脳卒中患者における検証のみ, 特異性: 94.7%	<i>Int J Stroke</i> , 2015 <sup>46)</sup>
急性虚血性脳卒中	433.xx, 434.xx	97.3	88.4	参照基準として台湾脳卒中レジストリ	<i>Medicine</i> , 2016 <sup>47)</sup>
結核接触	少なくとも1回の胸部X線検査をとまなう V01.1 又は 795.5	98.3	該当なし	参照基準として呼吸器科医によるカルテ審査	<i>Int J Cardiol</i> , 2016 <sup>26)</sup>
高血圧	401.x, 402.x, 403.x, 404.x, 405.x	92.4	88.5	脳卒中患者の状態はほとんどない, 参照基準として台湾脳卒中レジストリ	<i>Int J Cardiol</i> , 2016 <sup>26)</sup>
糖尿病	250.x	90.9	92.0		
高脂血症	272.x	69.1	89.5		
冠動脈疾患	410.x, 411.x, 412.x, 413.x, 414.x	63.7	47.6		
心房細動	427.31	72.8	71.1		
全てのがん	140-208	91.5	93.6	参照基準として台湾の全国がんレジストリ	<i>Pharmacoepidemiol Drug Saf</i> , 2017 <sup>48)</sup>
心不全	428	該当なし	97.6	参照基準として心臓学専門家によるカルテ審査	<i>J Am Heart Assoc</i> , 2017 <sup>49)</sup>
虚血性脳卒中	433-437	該当なし	94.2		
静脈瘤	454	該当なし	98.0	参照基準としてカルテ審査	<i>JAMA</i> , 2018 <sup>50)</sup>

な交絡因子の一つである。「適応による交絡」は、治療はランダムではなく、むしろ病気のベースラインの重症度に基づいて患者に示されているという状況を説明している<sup>28)</sup>。ベースラインの重症度は常に結果にも影響するため、ベースラインの重症度を調整せずに実際の治療効果を分離することは困難である。適応による交絡の例には、シクロオキシゲナーゼ II 選択的の非ステロイド性抗炎症薬 (NSAID) と非選択的 NSAID との間の胃腸出血リスクの比較がある<sup>29)</sup>。選択的 NSAID は胃腸副作用がより少ない可能性があるため、臨床医は、既に胃腸出血のベースラインリスクが高い患者に選択的 NSAID を処方する傾向が強く、結果として選択的 NSAID の投与を受ける群は偏りのある好ましくないリスクプロファイルになる。

NHIRD で測定されていない変数や利用できない変数は、HWD の外部データベースとの結合によって得られる可能性があるが、結合が不可能であったり、データが単にデジタル形式で保存されていなかったりする。操作変数法など、この問題に取り組むための方法論的及び統計的アプローチも開発されている<sup>30)</sup>。もう一つの方法は、セルフコントロールのクロスオーバーデザインを用いて、時間によって変動しない交絡因子を制御することである<sup>31)</sup>。プラ

セボ対照ではなく実薬と比較するデザインは、測定されていない交絡因子が群間でバランスがとれている場合、交絡因子の影響を排除するのに役立つかもしれない。傾向スコア校正、2段階校正、及び高次元傾向スコアなど傾向スコア (PS) を用いる方法もまた、潜在的交絡作用を減少させるために使用されている<sup>32-34)</sup>。しかし、どの方法も全ての状況に適用可能でも適切でもない。例えば、高品質の操作変数を作成したり理想的な実薬対照を置いたりすることが困難な場合もあるし、傾向スコア法の妥当性はデータベース中での変数が得られるか、そのデータベースに代表性があるかによる。

われわれのグループは、過去 10 年間の脳卒中研究に焦点を当てており、NHIRD の脳卒中重症度の一般的に用いられる指標の欠如が、データベースに基づく研究をいかに制限する可能性があるかを十分に認識している。脳卒中症候群の不均一性は、患者間のさまざまな重症度をもたらし、これは誤分類、適用による交絡及びアウトカム評価に影響を与える可能性のある他のバイアスを引き起こす可能性がある。残念なことに、NHIRD を含む行政上の目的で作成されたデータベースでは、脳卒中の重症度指標が欠けていることが多い。

**Table 3** NHIRD によって定義された難病の一覧

ICD-10-CM/PCS コード	高額医療費を要する疾病
C73	甲状腺の悪性新生物
C00.0-C06.9, C09.0-C10.9, C12-C14.8	口腔, 中咽頭及び下咽頭の悪性新生物ステージ I
C50.011-C50.929	乳房の悪性新生物ステージ I
C53.0-C53.9, C55	子宮頸部の悪性新生物ステージ I
C00.0-C96.9 (exclude C73, C94.4, C94.6)	その他の悪性新生物
D66	遺伝性第 VIII 因子欠乏症
D67	遺伝性第 IX 因子欠乏症
D68.1	遺伝性第 XI 因子欠乏症
D68.2	その他の凝固因子の遺伝性欠乏症
D55.0-D58.9	遺伝性溶血性貧血
D59.0-D59.9	後天性溶血性貧血
D46.4, D60.0-D60.9, D61.01-D61.9	無形成性貧血
N18.5, N18.6	慢性腎不全
I12.0	ステージ 5 の腎不全又は末期腎不全を伴う高血圧性腎疾患
I13.11, I13.2	心不全及びステージ 5 腎不全又は末期腎不全の合併を伴う高血圧性心腎疾患
M32.0-M32.9	全身性エリテマトーデス (SLE)
M34.0- M34.9	全身性硬化症
M05.70-M06.09, M06.20-M06.39, M06.80-M06.89, M06.9, M08.00-M08.99	関節リウマチ (若年性関節リウマチ)
M33.20-M33.29	多発性筋炎
M33.00-M33.19, M33.90-M33.99, M36.0	皮膚 (多発性) 筋炎
M30.0, M30.2, M30.8	結節性多発動脈炎
M31.0	過敏性血管炎
M31.30, M31.31	ウェゲナー肉芽腫症
M31.5, M31.6	巨細胞性動脈炎
I73.1	閉塞性血栓性血管炎 (ビュルガー<バージャー>病)
M31.4	大動脈弓症候群 (高安病)
M30.3	川崎病
M35.2	ベーチェット病
L10.0-L10.9	天疱瘡
M35.00-M35.09	乾燥症候群
K50.00-K50.919	クローン病
K51.00-K51.919	潰瘍性大腸炎
F01.50, F01.51, F03.90, F03.91	認知症, 詳細不明
F05	既知の生理的状態によるせん妄
F02.80, F02.81, F06.0, F06.1, F06.8	既知の生理的状態によるその他の精神障害
F20.0-F20.9, F25.0-F25.9	統合失調症
F30.10-F30.13, F30.2-F30.9, F31.0-F31.9, F32.2-F32.9, F33.2-F33.9	気分障害
F22	妄想性障害
F84.0	自閉症
F84.3	その他の小児期崩壊性障害
F84.5, F84.8	その他の広汎性発達障害 (アスペルガー症候群)
F84.9	広汎性発達障害, 詳細不明
E00.0-E00.9, E03.0, E03.1	先天性ヨード欠乏症候群 (先天性甲状腺機能低下症)
E10.10-E10.9	1 型糖尿病
E23.2	尿崩症
E25.0-E25.9	副腎性器障害
E70.0-E71.2, E72.00-E72.51, E72.59, E72.8, E72.9	アミノ酸代謝障害
E74.00-E74.09	糖原病
E74.20-E74.29	ガラクトース代謝障害
E78.1	純型高グリセリド血症
E88.1	リポジストロフィー
E75.21-E75.22, E75.240-E75.249, E75.3, E77.0-E77.9	スフィンゴリピド代謝障害
E75.6, E78.70, E78.9	脂質蓄積障害
E83.00-E83.09	銅代謝障害

Table 3 つづき

ICD-10-CM/PCS コード	高額医療費を要する疾病
E20.1, E83.50-E83.59, E83.81 D81.3, D81.5, E79.1-E79.9 E76.01-E76.9 E71.310-E71.548, E80.3, E88.40-E88.89, H49.811-H49.819 E88.9 Q00.0-Q00.2 G90.1, Q01.0-Q04.9, Q06.0-Q06.9, Q07.8, Q07.9 Q20.0-Q24.9 Q25.0-Q28.9 Q33.0 Q33.3, Q33.6 Q33.8, Q33.9 Q41.0-Q45.9 Q60.0-Q60.6 Q61.00-Q61.9 Q62.0-Q62.39 Q63.0-Q63.9 Q77.0-Q77.2, Q77.4, Q77.5, Q77.7-Q77.9, Q78.4 Q90.0-Q99.1, Q99.8, Q99.9 Q35.1-Q35.7, Q36.0-Q37.9 T31.20-T31.99, T32.20-T32.99 T26.00XA-T26.92XA T20.30XA-T20.39XA, T20.70XA-T20.79XA  Z94.0 Z94.1 Z94.2 Z94.4	カルシウム代謝障害 プリン及びピリミジン代謝障害 グリコサミノグリカン代謝障害 その他の代謝障害  代謝障害, 詳細不明 無脳症及び類似先天奇形 神経系のその他の先天奇形 心球先天奇形及び心臓中隔の先天奇形又は他の心臓の先天奇形 循環器系のその他の先天奇形 先天性嚢胞肺 肺の形成不全, 低形成及び異形成 肺のその他の先天奇形 消化器系のその他の先天奇形 腎の無発生及びその他の減形成 嚢胞性腎疾患 腎盂及び尿管の先天性閉塞性欠損 腎のその他の先天奇形 骨軟骨異形成症, 長管骨及び脊椎の成長障害を伴うもの 染色体異常 先天性口蓋裂及び口唇裂 体表面積 > 20% の熱傷 眼及び付属器に局限する熱傷 頭部及び頸部の熱傷, 体の一部の損失を伴う下層組織の深部壊死 (深部第3度) 腎移植状態 心臓移植状態 肺移植状態 肝移植状態
Z94.81, Z94.84 Z94.83 Z94.82 T86.10-T86.19 T86.40-T86.49 T86.20-T86.23, T86.290-T86.298 T86.810-T86.819 T86.00-T86.09 T86.890-T86.899  T86.850-T86.859 A80.0-A80.2, A80.30-A80.39 G80.0-G80.2, G80.4-G80.9 (G82.20-G82.54, G83.0-G83.9) + (B91, G14) T07  Z99.11	骨移植状態 膵臓移植状態 腸移植状態 腎移植の合併症 肝移植の合併症 心臓移植の合併症 肺移植の合併症 骨髄移植の合併症 膵臓移植の合併症  腸移植の合併症 その他の麻痺を伴う急性灰白髄炎 脳性麻痺 その他の麻痺性症候群 (遅発性の急性灰白髄炎) 重症度スケールが 16 以上の多発性損傷 (外傷重症度スコア $\geq 16$ ) 長期機械的人工換気, 以下のいずれかとして定義: 1. 21 日以上の上侵襲的機械的人工換気 2. 侵襲的機械的人工換気に続く非侵襲的換気, 全期間 21 日以上 3. 侵襲的機械的人工換気に続く陰圧換気, 全期間 21 日以上 4. 特異的疾患, 例えば末期心不全, 慢性肺疾患, 一次神経筋疾患, 慢性低換気症候群, 21 日以上の上侵襲的換気が必要とする.
E41  E43  T70.3XXA T79.0XXA	主要な腸切除, 腸管不全による重度の栄養失調で 30 日間の完全静脈内 栄養状態にあり, 経口食餌によって十分な栄養を得ることができない 患者  他の慢性疾患による重度の栄養失調で 30 日間の完全静脈内栄養状態に あり, 経口食餌によって十分な栄養を得ることができない患者 減圧症 空気塞栓症

**Table 3** つづき

ICD-10-CM/PCS コード	高額医療費を要する疾病
G70.00, G70.01	重症筋無力症
D80.1, D80.6, D80.8, D80.9	主として抗体欠乏を伴う免疫不全症
D81.0-D81.2, D81.4, D81.6, D81.7, D81.89, D81.9	複合免疫不全症
D82.0-D82.9	その他の大きな欠損に関連する免疫不全症
D83.0-D83.9	分類不能型免疫不全症
D84.0-D84.9	その他の免疫不全症
S12.000A-S12.9XXA+S14.101A-S14.159A, S24.101A-S24.159A, S34.101A-S34.139A	脊髄損傷を伴う脊柱の骨折
S14.101A-S14.159A, S24.101A-S24.159A, S34.101A-S34.139A	脊柱損傷のエビデンスのない脊髄損傷
G32.0, G95.0, G95.11-G95.89, G95.9, G99.2	その他の脊髄疾患
J60	炭鉱夫塵肺症
J61	石綿肺症
J62.0, J62.8	その他のケイ素又はケイ酸による塵肺症
J63.0-J63.6	その他の無機塵による塵肺症
J64, J65	塵肺症
I60.00-I60.9, I61.0-I62.9	脳血管疾患（急性期） くも膜下出血 脳内出血
I63.00-I63.9	脳梗塞
G45.0-G45.2, G45.4-G46.8, I67.0-I67.2, I67.4-I67.7, I67.81, I67.82, I67.841-I67.848, I67.89, I67.9, I68.0, I68.8	その他の脳血管疾患
G35	多発性硬化症
G71.0, G71.2	先天性筋ジストロフィー 外皮の先天異常
Q81.0-Q81.9, Q82.8, Q82.9	先天性表皮水疱症
Q84.9	外皮の先天奇形，詳細不明
Q80.0-Q80.9	先天性魚鱗癬
A30.0-A30.9	らい（ハンセン病）
K70.2-K70.31, K74.1-K74.69	合併症を伴う肝硬変 コントロール不良の腹水 食道又は胃静脈瘤出血 肝性昏睡又は肝代償不全
P07.10	出生から3か月以内に入院治療を受けた早産児に起因する神経学的，筋肉的，骨格的，心臓又は肺の合併症.
P07.20	生後3か月で中等度の障害を有すると認定された早産児に起因する神経学的，筋肉的，骨格的，心臓又は肺の合併症.
T57.0X1A, T57.0X2A, T57.0X3A, T57.0X4A	ヒ素及びその化合物の毒作用（鳥脚病）
G12.20-G12.29	運動ニューロン疾患
A81.00-A81.09	クロイツフェルト・ヤコブ病

この問題を解決するために、われわれは、脳卒中患者に重症度スコアを割り当てた国立衛生研究所脳卒中スケール (National Institutes of Health Stroke Scale; NIHSS) の代わりになる新規7項目のNHIRDベースの脳卒中重症度指数 (SSI) を、データマイニング技術と台湾脳卒中レジストリデータを用いて開発した<sup>35)</sup>。SSIは、以下の式を用いてNHIRDで利用可能な7変数に基づいて計算される。例えば、一般病棟に滞在し、滞在中に菌感受性検査及び尿路カテーテル法を受けた患者のSSIは、7.13(すなわち、1.3642 + 1.6569 - 5.5761 + 9.6804)となる。

$$\text{脳卒中重症度指数} = + 3.5083 (\text{気道吸引})$$

- + 1.3642 (菌感受性検査)
- + 4.1770 (集中治療室滞在)
- + 4.5809 (経鼻胃管挿管)
- + 2.1448 (マンニトール又はグリセロールのような浸透圧療法)
- + 1.6569 (尿路カテーテル法)
- 5.5761 (一般病棟滞在)
- + 9.6804 (定数)

その後、SSIは、脳卒中の重症度を測定する際のSSIの有効性を改善するために、更新された脳卒中レジストリデータ及び脳卒中の修正Rankinスケールで調整することによって、「推定NIHSS」(eNIHSS)に変換することができる。eNIHSSは、急性虚血性脳卒中<sup>36)</sup>及び脳内出血<sup>37)</sup>の

両方で有効性が検証されており、脳卒中転帰研究のためのより良い case-mix adjustment を提供するために使用されている<sup>38-40</sup>。SSI は、虚血性脳卒中及び脳内出血を有する患者のために、以下の式を用いて eNIHSS に変換することができる：

$$\text{虚血性脳卒中：eNIHSS} = 1.1722 \times \text{SSI} - 0.7533$$

$$\text{脳内出血：eNIHSS} = 1.3894 \times \text{SSI} - 3.6788$$

NHIRD の脳卒中研究に対する eNIHSS の使用を容易にするため、標準重症度データが入手できない場合に、研究者が脳卒中の重篤度を容易に評価するためにオンラインの使いやすいツールを作成した<sup>41</sup>。われわれは、脳卒中研究における妥当性と代表性について eNIHSS の評価と改善を続けている。較正と検証のための外部データを必要とする開発段階を除いて、そのような代理尺度は、いったん開発されると自己完結型である。代理尺度のユーザーは、重症度スコアを提供する場合もあればしない場合もあるような複数のデータベースを相互リンクする必要はない。したがって、この方法はデータのプライバシーに関する懸念を軽減することができる。もちろん、全ての指標やスコアと同様に、代理尺度もより多くのより良いデータが利用できるようになると、常に絶えず評価され、最適化される必要がある。しかし、これらの代理尺度の本質的な有用性は、それが将来の研究の結果を予測したり、説明したりするのにどれだけ効果的かによって決まる。

## 8. 展望

われわれは、医療研究におけるビッグデータ分析の改善についていくつかの展望を示した。第 1 に、NHIRD データの入手可能性に約 2 年のタイムラグがある。このようなタイムラグは、政策策定や臨床決定を支援するために NHIRD から得られた分析の適時評価及び効率の可能性を低下させる可能性がある。規制当局と研究者は、新たなリスクを迅速に特定するために、よりリアルタイムなデータを生成するためのより良い方法を見つけ出さなければならない。台湾の HWD センターも、現在のデータベースでは利用できない重要な情報を入手するために、データ収集努力(調査など)を積極的に主張、設計、実施しなければならない。われわれは全ての行政手続を簡素化し、効率的な分析方法を提供して、研究者にとってより良い環境を提供しなければならない。更に、より多くの研究者を育成するために必要な疫学と統計的知識を備えたデータベースの使用に関する教育や訓練プログラムにも労力を費やす必要がある。HWD センターのデータは大量ではあったが、ほとんど全てが構造化されたデータであった。

人工知能の時代に、台湾の BDA の次のステップは、NHIRD 及び他の構造化データベースと非構造化データを連携することである。日常的な臨床診療から生まれたデータは、電子カルテ、医療機器、電気生理学的モニター及び放射線画像など、大部分は非構造的に存在する。機械学習と自然言語処理を含む人工知能技術を使用して、非構造化データを処理し、臨床診療を支援するための結論を生むことができる<sup>42</sup>。例えば、自然言語処理技術を適用することで、電子カルテの情報を活用し、急性虚血性脳卒中のための経静脈的血栓溶解療法において正しく適な意思決定を支援することができる<sup>43</sup>。NHIRD の過去の成功経験によって、われわれは HWD センターがいつかデータレポジトリ(置き場)から強力な研究エンジンに変身し、さまざまな情報源からのデータを、将来の台湾での大規模データ分析によって裏付けられたリアルワールドデータによって医学を導ける“光”へと変換できることを願っている。

## 文 献

- 1) Lin YY.; Huang CS. Aging in Taiwan: Building a Society for Active Aging and Aging in Place. *Gerontologist*. 2016, 56 (2), p.176-183.
- 2) Wu CY.; Kuo KN.; Wu MS. *et al.* Early Helicobacter pylori eradication decreases risk of gastric cancer in patients with peptic ulcer disease. *Gastroenterology*. 2009, 137 (5), p.1641-1648.e1641-1642.
- 3) Yeh YC.; Chen PJ.; Lai MS. Antiviral therapy and hepatitis B virus-related hepatocellular carcinoma recurrence. *JAMA*. 2013, 309 (8), p.765-766.
- 4) Chien HC.; Kao Yang YH.; Bai JP. Trastuzumab-Related Cardiotoxic Effects in Taiwanese Women: A Nationwide Cohort Study. *JAMA Oncol*. 2016, 2 (10), p.1317-1325.
- 5) Chang SH.; Chou IJ.; Yeh YH. *et al.* Association Between Use of Non-Vitamin K Oral Anticoagulants With and Without Concurrent Medications and Risk of Major Bleeding in Nonvalvular Atrial Fibrillation. *JAMA*. 2017, 318 (13), p.1250-1259.
- 6) Schneeweiss S. Learning from big health care data. *N Engl J Med*. 2014, 370 (23), p.2161-2163.
- 7) Hsiao FY.; Yang CL.; Huang YT. *et al.* Using Taiwan's national health insurance research databases for pharmacoepidemiology research. *J Food Drug Anal*. 2007, 15, p. 99-108.
- 8) Lee JT.; Huang N.; Majeed A. The need for better linkage between administrative data and clinical datasets. *BMJ*. 2015, 351, h5816.
- 9) Hsing AW.; Ioannidis JP. Nationwide Population Science: Lessons From the Taiwan National Health Insurance Research Database. *JAMA Intern Med*. 2015, 175 (9), p.1527-1529.
- 10) Pratt N.; Andersen M.; Bergman U. *et al.* Multi-country rapid adverse drug event assessment: the Asian Pharmacoepidemiology Network (AsPEN) antipsychotic and acute hyperglycaemia study. *Pharmacoepidemiol Drug Saf*. 2013, 22 (9), p.915-924.
- 11) Roughead EE.; Chan EW.; Choi NK. *et al.* Proton pump

- inhibitors and risk of *Clostridium difficile* infection: a multi-country study using sequence symmetry analysis. *Expert Opin Drug Saf.* 2016, **15** (12), p.1589-1595.
- 12) Roughead EE.; Chan EW.; Choi NK. *et al.* Variation in Association Between Thiazolidinediones and Heart Failure Across Ethnic Groups: Retrospective analysis of Large Healthcare Claims Databases in Six Countries. *Drug Saf.* 2015, **38** (9), p.823-831.
  - 13) Pratt N.; Chan EW.; Choi NK. *et al.* Prescription sequence symmetry analysis: assessing risk, temporality, and consistency for adverse drug reactions across datasets in five countries. *Pharmacoepidemiol Drug Saf.* 2015, **24** (8), p.858-864.
  - 14) Andersen M, Bergman U.; Choi NK. *et al.* The Asian Pharmacoepidemiology Network (AsPEN): promoting multi-national collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf.* 2013, **22** (7), p.700-704.
  - 15) Salerno J.; Knoppers BM.; Lee LM.; Hlaing WM.; Goodman KW. Ethics, big data and computing in epidemiology and public health. *Ann Epidemiol.* 2017, **27** (5), p.297-301.
  - 16) National Health Insurance Research Database. Data subsets. Available from [https://nhird.nhri.org.tw/en/Data\\_Subsets.html](https://nhird.nhri.org.tw/en/Data_Subsets.html). Cited on March 1, 2018.
  - 17) Ministry of Health and Welfare. Health and Welfare Data Science Center. Available from <https://dep.mohw.gov.tw/DOS/np-2497-113.html>. Cited on March 1, 2018.
  - 18) Wu YT.; Lee HY. National Health Insurance database in Taiwan: A resource or obstacle for health research? *Eur J Intern Med.* 2016, **31**, e9-e10.
  - 19) Hampson NB.; Weaver LK. Carbon monoxide poisoning and risk for ischemic stroke. *Eur J Intern Med.* 2016, **31**, e7.
  - 20) van Walraven C.; Austin P. Administrative database research has unique characteristics that can risk biased results. *J Clin Epidemiol.* 2012, **65** (2), p.126-131.
  - 21) Hall GC.; Sauer B.; Bourke A. *et al.* Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf.* 2012, **21** (1), p.1-10.
  - 22) Hsieh FI.; Lien LM, Chen ST. *et al.* Get With the Guidelines-Stroke performance indicators: surveillance of stroke care in the Taiwan Stroke Registry: Get With the Guidelines-Stroke in Taiwan. *Circulation.* 2010, **122** (11), p.1116-1123.
  - 23) Hsieh CY.; Tsao WC.; Lin RT. *et al.* Three years of the nationwide post-acute stroke care program in Taiwan. *J Chin Med Assoc.* 2018, **81** (1), p.87-88.
  - 24) Chen CC.; Chen LS.; Yen MF. *et al.* Geographic variation in the age- and gender-specific prevalence and incidence of epilepsy: analysis of Taiwanese National Health Insurance-based data. *Epilepsia.* 2012, **53** (2), p.283-290.
  - 25) Hsieh CY.; Chen CH.; Li CY. *et al.* Validating the diagnosis of acute ischemic stroke in a National Health Insurance claims database. *J Formos Med Assoc.* 2015, **114** (3), p.254-259.
  - 26) Sung SF.; Hsieh CY.; Lin HJ. *et al.* Validation of algorithms to identify stroke risk factors in patients with acute ischemic stroke, transient ischemic attack, or intracerebral hemorrhage in an administrative claims database. *Int J Cardiol.* 2016, **215** p.277-282.
  - 27) Cheng CL.; Chien HC.; Lee CH. *et al.* Validity of in-hospital mortality data among patients with acute myocardial infarction or stroke in National Health Insurance Research Database in Taiwan. *Int J Cardiol.* 2015, **201**, p.96-101.
  - 28) Salas M.; Hofman A.; Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am Journal Epidemiol.* 1999, **149** (11), p.981-983.
  - 29) Wolfe F.; Flowers N.; Burke TA. *et al.* Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. *J Rheumatol.* 2002, **29** (5), p.1015-1022.
  - 30) Davies NM.; Smith GD.; Windmeijer F. *et al.* COX-2 selective nonsteroidal anti-inflammatory drugs and risk of gastrointestinal tract complications and myocardial infarction: an instrumental variable analysis. *Epidemiology.* 2013, **24** (3) p.352-362.
  - 31) Petersen I.; Douglas I.; Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. *BMJ.* 2016, **354**, i4515.
  - 32) Brookhart MA.; Wyss R.; Layton JB. *et al.* Propensity score methods for confounding control in nonexperimental research. *Circ Cardiovasc Qual Outcomes.* 2013, **6** (5), p.604-611.
  - 33) Sturmer T.; Schneeweiss S.; Rothman KJ. *et al.* Performance of propensity score calibration--a simulation study. *Am J Epidemiol.* 2007, **165** (10), p.1110-1118.
  - 34) Schneeweiss S.; Rassen JA.; Glynn RJ. *et al.* High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology.* 2009, **20** (4), p.512-522.
  - 35) Sung SF.; Hsieh CY.; Kao Yang YH. *et al.* Developing a stroke severity index based on administrative data was feasible using data mining techniques. *J Clin Epidemiol.* 2015, **68** (11), p.1292-1300.
  - 36) Sung SF.; Hsieh CY.; Lin HJ. *et al.* Validity of a stroke severity index for administrative claims data research: a retrospective cohort study. *BMC Health Serv Res.* 2016, **16** (1), p.509.
  - 37) Hung LC.; Sung SF.; Hsieh CY. *et al.* Validation of a novel claims-based stroke severity index in patients with intracerebral hemorrhage. *J Epidemiol.* 2017, **27** (1), p.24-29.
  - 38) Hsieh CY.; Lin HJ.; Hu YH.; Sung SF. Stroke severity may predict causes of readmission within one year in patients with first ischemic stroke event. *J Neurol Sci.* 2017, **372**, p.21-27.
  - 39) Hsieh CY.; Wu DP.; Sung SF. Trends in vascular risk factors, stroke performance measures, and outcomes in patients with first-ever ischemic stroke in Taiwan between 2000 and 2012. *J Neurol Sci.* 2017, **378**, p.80-84.
  - 40) Hsieh CY.; Lin HJ.; Chen CH. *et al.* "Weekend effect" on stroke mortality revisited: Application of a claims-based stroke severity index in a population-based cohort study. *Medicine.* 2016, **95** (25), e4046.
  - 41) Stroke Severity Index Calculator. <http://140.123.174.220:508/SSI/hdmlab/ssi2.jsp>, (accessed 2018-3-1).

- 42) Jiang F.; Jiang Y.; Zhi H. *et al.* Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*. 2017, 2, e000101.
- 43) Sung SF CK.; Wu DP. *et al.* Applying natural language processing techniques to develop a task-specific EMR interface for timely stroke thrombolysis: A feasibility study. *Int J Med Inform*. 2018, 112, p.149-157.
- 44) Cheng CL.; Kao YH.; Lin SJ. *et al.* Validation of the National Health Insurance Research Database with ischemic stroke cases in Taiwan. *Pharmacoepidemiol Drug Saf*. 2011, 20 (3) , p.236-242.
- 45) Cheng CL.; Lee CH.; Chen PS. *et al.* Validation of acute myocardial infarction cases in the national health insurance research database in taiwan. *J epidemiol*. 2014, 24 (6) , p.500-507.
- 46) Hsieh CY.; Cheng CL.; Lai EC. *et al.* Identifying renal dysfunction in stroke patients using diagnostic codes in the Taiwan National Health Insurance Research Database. *Int J Stroke*. 2015, 10 (1) , E5.
- 47) Su VY.; Yen YF.; Pan SW. *et al.* Latent Tuberculosis Infection and the Risk of Subsequent Cancer. *Medicine*. 2016, 95 (4) , e2352.
- 48) Kao WH.; Hong JH.; See LC. *et al.* Validity of cancer diagnosis in the National Health Insurance database compared with the linked National Cancer Registry in Taiwan. *Pharmacoepidemiol Drug Saf*. 2017.
- 49) Lin YS.; Chen TH.; Chi CC. *et al.* Different Implications of Heart Failure, Ischemic Stroke, and Mortality Between Nonvalvular Atrial Fibrillation and Atrial Flutter—a View From a National Cohort Study. *J Am Heart Assoc*. 2017, 6 (7) , e006406.
- 50) Chang SL.; Huang YL.; Lee MC. *et al.* Association of Varicose Veins With Incident Venous Thromboembolism and Peripheral Artery Disease. *JAMA*. 2018, 319 (8) p.807-817.