

別紙① 台湾データ利活用に関するヒアリング調査

研究協力者

加藤 源太（京都大学）

牧戸 香詠子（東京大学）

明神 大也（奈良県立医科大学）

## 【台湾の概要】

台湾（中華民国）の面積は日本の10分の1弱、人口は日本の5分の1弱の約2300万人である。九州7県と比べると面積は2割ほど小さく、人口は6割ほど多い。

台湾の医療保険制度は、1950年代から実施された職業別の総合保険（日本にも船員保険が存在）に端を発している。その後、対象者の拡大に伴い、保険制度が乱立・複雑化するとともに給付の格差などの問題も発生した。そこで1995年に保険者を一本化し、台湾政府の衛生福利部中央健康保険署が所掌する「全民健康保険」制度が確立した。

日本と同様に診療報酬制度となっており、外来の自己負担額は定額・入院の自己負担額は定率で、残りは保険料と税金で賄っている。また、保険診療と自費診療を同時に実施する混合診療が認められている。

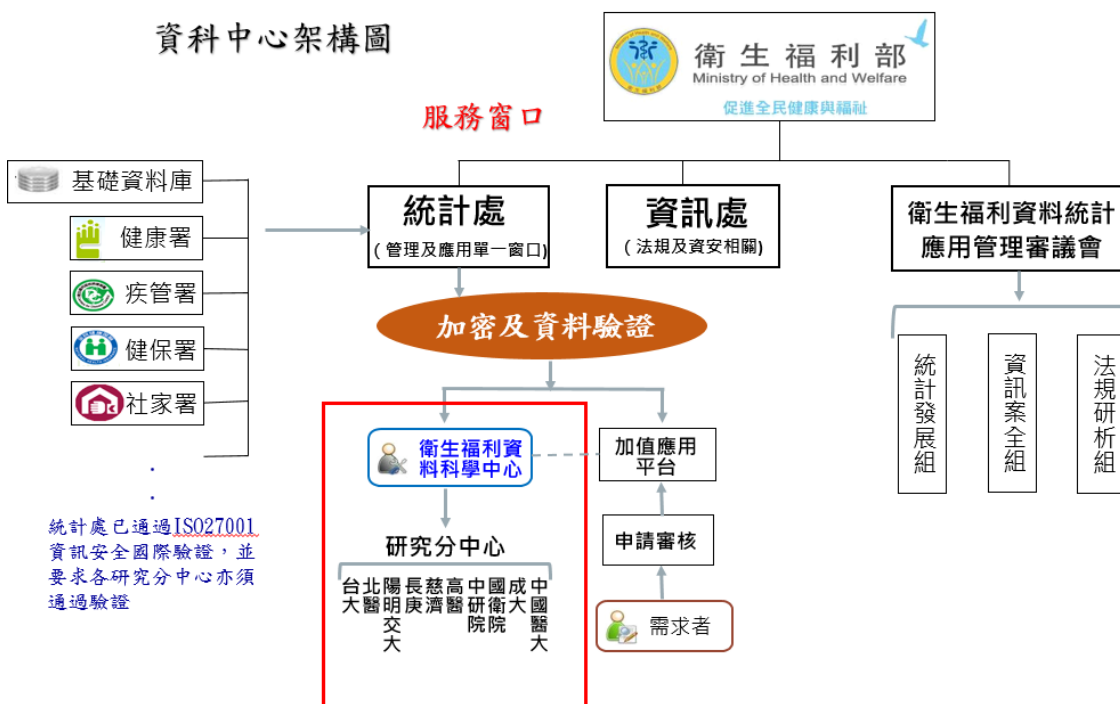
保険診療分については、保険者である中央健康保険局にレセプト請求するが、オンライン請求率はほぼ100%となっている。

台湾の健康保険証にはICカードチップが埋め込まれている。これに診療等の情報がリンクされており、過去の診療・投薬情報は患者の同意があれば確認できる。国民の約9割がこの同意をしているとされている。

## 【台湾のレセプト情報等の提供体制】

2002年に衛生福利部の関連機関である国家衛生研究院 National Health Research Institutes (NHRI)が、医療保険請求に基づく研究用データベースとして National Health Insurance Research Database (NHIRD) を設立した。その後、下記に示す、「衛生福利資料科学中心」で、他の統計情報等も含め、国が収集した医療関係情報を研究者等に提供している。

資料中心架構圖



本報告書では、衛生福利資料科学中心の URL (<https://dep.mohw.gov.tw/DOS/cp-5119-59201-113.html>) に示された内容に加え、当該データを利用している高雄医学大学、国立成功大学、国立台湾大学に所属する研究者 3 グループにそれぞれヒアリングを行った情報を記載する。

【データの内容】

各種データ定義書は <https://dep.mohw.gov.tw/DOS/lp-2503-113.html> に公開されている。

(いわゆる日本でいう) レセプトデータだけでなく、下表に示す出生票や死亡票、がん登録情報、国民健康聞き取り調査などがあり、匿名化済の ID で連結可能となっている。

出生、死亡 母子保健のデータ	疾患及び損傷	保健	健康／ライフスタイル
出生証明書 死因データ 複数の死因データ 母子保健データベース	がんレジストリ 人工授精に関するデータ 希少疾病データ 18 件の報告及び調査データ 交通事故データ	学生健康調査 成人喫煙サーベイランス 青少年喫煙調査 行動危険因子サーベイランス	台湾の加齢に関する長期的研究
調査データ	福祉／社会	福祉／社会	スクリーニング等
国民健康聞き取り調査 高血圧、高血糖及び高脂血症調査 台湾出生コホートデータ 避妊に関する知識・行動のデータ 在宅ケア助成金利用状況調査 高齢者福祉団体の必要条件調査の報告書 他の 12 件の調査データ	台湾における若年状況調査 高齢者状況調査の報告 ひとり親家族の状況調査 身体的及び精神的障害者生活調査 在宅ケア助成金使用者状況調査 低所得者及び中低所得者の生活状況調査 女性の生活状況調査	障害者プロフィール 低所得者及び中低所得者世帯 家庭内暴力データ 小児・青少年保護の報告書データ 性的暴行の報告書データ	がんスクリーニング(乳、子宮頸部、口腔、大腸) 台湾先住民族(出生、死因データ、戸籍)

レセプトデータについては 2001 年以降、約 20 年分のデータが利用可能となっている。データクリーニングのために約 2 年要しており、直近 2 年間のデータは利用できない。サンプリングされていないフルデータと、200 万人サンプリングデータの 2 種類のデータが存在

する。原則として「日」の情報は提供されない、しかしフルデータに関しては、必要に応じて日付を加工して提供を受けることができる。例えば「ステロイド→生検」の順番を見たい場合は、申請時にその旨を指定する。

#### 【申請方法】

申請及び利用できる者は台湾国籍保有者に限られており、留学生含めて外国人の利用はできない。利用申請は Web 上から入力となっており、ここには示すことができないが、申請画面を見せてもらったところ、申請内容は NDB の様式 1 とほぼ同様の内容であった。そのため、申請に慣れた研究者でも、申請フォームの入力に数日間は必要であり、煩雑であるとのことであった。

民間事業者だけの申請は不可能で、アカデミアとの連携が必要とされている。

フルデータを申請する場合は IRB 必須で、基本的に利用機関で IRB 申請することになっている。その一方、日本のような専門委員会の審査は実施されていない。IRB では 20 歳以下の抽出や希少疾患は個人同定のリスクが高いため慎重となっている。

また、日本では IRB は施設間によって審査レベルが異なると言われているが、ヒアリングを行う限り、台湾における IRB は病院機能の評価にも影響するとみられ、審査の質は保たれているとのことであった。また、IRB では個人同定につながらないかを確認しており、ヒアリングした研究者はいずれも個人情報の同定やそれによる国民感情への影響を理解している様子であった。

申請後、利用開始まで要した期間は数か月～約 8 か月であるが、切り出しのデータ量が多いと最大 1 年近くかかっているとのことである。

#### 【利用方法】

データの利用期間は 2 年間で、台湾全土にあるオンサイトセンターのみの利用に限定されている。なお、行政利用においては抽出データが提供されている。

オンサイトセンターは 10 か所、各施設 10 端末前後またはそれ以上ある。

利用者は 1 か月前に予約する必要があるが、人気があり、予約困難な状況である。

オンサイトセンターはガラス張りで外部から様子をうかがえるが、窃視は不可となっている。日本のオンサイトセンターのように（データセンターも同様であるが）PC やタブレット、スマートフォンなどの持ち込みは不可である。しかし端末間のパーティション等はなく、利用者同士の会話は他の利用者の邪魔にならない程度であれば禁止されていない。

最小集計単位は 3 とされており、日本の NDB の 10 より緩和されている。

2 年間の利用期間中、2 回は利用場所を変更できることになっている。

#### 【利用料】

利用料はデータ抽出料と端末利用料、データ処理・分析料、データ保管料の4つで構成され、データ抽出料と端末利用料は必須である。また、行政利用を含めて減免規定は存在しない。日本では国費の還流を防止するための免除規定が存在するが、台湾とは国家予算に関する考え方が違うのかもしれない。

データ抽出料は <https://dep.mohw.gov.tw/DOS/lp-2503-113.html> に示されたデータ1列につき NT\$180、12列以上の場合は1列につき NT\$200（年度ごと）発生する。さらに、抽出データのファイルサイズが1GB未満で NT\$5,000、1~10GBで NT\$10,000、10~100GBで NT\$20,000、100~500GBで NT\$30,000、500GB~で NT\$50,000 発生する。仮に全データを利用すると、70万台湾ドル程度必要となるとのことであった。

端末利用料は4時間ごとに NT\$700、夜間利用の場合は追加で1日 NT\$1,400 発生するが、少なくとも国立成功大学・高雄医学大学は夜間利用を実施していなかった。

データ処理・分析料は、研究者の代わりにセンターの担当者が実施する際に発生する料金で、4人時間あたり NT\$7,500 となっている。

データ保管料は、2年の利用期間を延長した場合に半年ごとに発生するもので、データ量が1TB未満で NT\$5,000、1~2TBで NT\$10,000、2~5TBで NT\$20,000、5TB~で NT\$30,000 となっている。

#### 【考察】

2年間のデータクリーニング期間があることもあり、利用開始時点で ETL 処理が一定程度できていて、日本の商用 DB と同じくらいの処理が施されている印象を受けた。

また、データ提供側が研究者に対して開示している情報は日本よりずっと多いと感じた。それでも、初心者が申請から分析まで一人で実施するのは難しく、有識者や経験者による研究チームを構成するなどして、先人頼っている現状も見られた。

日本では利用料の減免措置があるが、台湾では減免措置がないため利用料の工面は困っているのは実情であるものの、直近(<https://www.mohw.gov.tw/dl-88128-2cd8d699-50a7-43ad-9b89-3c5b695d798c.html>) では、年間約 300 件の提供があった。日本の人口規模に直すと年間 1500 件だが、この中には、日本でいう調査票情報の提供案件も含まれているため、NDB と直接比較はできない。しかし web からの申請に骨が折れる、直近2年間のデータは利用できない、オンサイトの端末の予約が争奪戦になり困難である、利用料の確保が難しい、データ提供まで長い場合は1年近くを要する、などと利用者の視点からは様々な制約の指摘はみられたものの、概ね受け入れられている印象であった。これは研究者側のデータ利用環境への信頼性や、データ提供側が極力情報開示に努めているからではないかと、ヒアリングを通して感じられた。

また、個人同定のリスクや倫理的課題は研究者としてもよく理解されていた印象で、この点は日本においては研究者側も十分理解すべきだと感じた。

このように、国民感情の違いはあるものの、データ提供側・利用側の歩み寄りがなされてい

る印象であった。

日本では今後 NDB の利活用が進むと見込まれる。NDB の利活用の目的は1億人の悉皆データが国民保健の向上、ひいては世界の公衆衛生の向上に資することである。

データ提供側はただ必要に迫られたからデータ提供を拡大するというのではなく、どうすれば利用者が使いやすくなり、社会の役に立つ知見をデータから引き出していけるのかについて、前向きな検討を行っていく必要があるだろう。またデータ利用側も、営利を目的としたデータ提供が前提となっているわけではないことに配慮しつつ、悉皆データに含まれる個人同定のリスクを十分理解した上で、既存の利用環境やルールを踏まえて目的を達成できるよう取り組んでいく必要があるだろう。

#### 【謝辞】

今回ヒアリングを受けてくださった、高雄醫學大學公共衛生學系 莊弘毅 教授・杜鴻賓 教授、高雄市立小港醫院家庭醫學科 楊鎮誠先生、國立成功大學藥學系 賴嘉鎮教授、國立臺灣大學公共衛生學院 杜裕康先生・周佩萱先生・陳建煒先生に御礼申し上げます。