Poster Title: Artificial Intelligence-Powered Identification, Access and Utility Mapping of Real-World Data Sources for Lung cancer in Asia Pacific

RWD245

JW WONG[1], SJ JOIS[1], L HOGG[1], K TAN[1], KW L[1], V JAIN[2], V ESCALANTE[2]

1. APAC Healthcare Market Access & HEOR, Ipsos Pte Ltd, Singapore
2. Data Science and Advanced Analytics, Ipsos, London, United Kingdom

INTRODUCTION

Lung cancer (LC) poses a major public health challenge in the Asia-Pacific (AP) region, accounting for approximately 60% of the cases globally[1], with more than 1 million deaths in the AP region in 2022 alone[2]. This burden is driven by the region's rapidly aging population, ongoing industrialization, high smoking prevalence and persistent air pollution, which collectively contribute to increasing lung cancer incidence and mortality.

Clinical trials often lack the diversity to fully assess the impact of lung cancer across heterogeneous populations. This makes Real-World Evidence (RWE) critical for capturing the full range of patient experiences and outcomes. As RWE becomes increasingly accepted in healthcare decision-making, it is essential that the evidence generated is both timely and robust; these qualities must be balanced rather than viewed as competing priorities. While vast amounts of Real-World Data (RWD) is collected in current databases, the process of identifying, evaluating and accessing robust sources remains complex, time consuming, and susceptible to bias.

OBJECTIVE

Our research is focused on identifying and evaluating RWD sources for LC in the AP region. We aim to understand data access, coverage, and utility for evidence generation. The insights gained will help to develop targeted evidence generation plans and guide the selection of suitable data sources for RWE study designs. By identifying and utilizing RWD from diverse sources, we can address a wider array of research questions and achieve a comprehensive understanding of the clinical, economic, and societal impacts of LC in the region.

METHOD

An artificial intelligence (AI)-powered systematic literature review of academic publications (2014-2024, PubMed) was conducted to identify RWD sources for LC in AP. Leveraging Large Language Models (gemini-1.5-pro with enhanced prompt), our proprietary AI system employed a semantic search protocol to identify relevant data sources and extract key information including database type, coverage, demographics, treatments, clinical, humanistic, and economic data. Results were manually validated by two independent reviewers. Human oversight at all stages ensured integrity of AI outputs, while addressing potential biases. Identified data sources were categorised into four key types: cross-sectional studies[3] (including population surveys), patient registries[4], insurance/claims databases[5] and Electronic Hospital Records (EHRs)/ Electronic[6] Medical Records (EMRs). Data sources cited more than twice were prioritized for an in-depth assessment of data utility, focusing on variable coverage and availability across key RWD categories.

RESULTS

A total of 360 citations were retrieved, with 152 unique data sources identified after manual validation. The data sources spanned 5 AP countries. More than one-third of the data sources originated from China (36%), followed by Japan (30%), Australia (15%), South Korea (14%) and Taiwan (9%), with some data sources comprising of data from more than one country. Patient registries (43%) account for the largest proportion, followed by EHR/EMRs (41%), cross-sectional studies (including population surveys; 9%) and insurance/claims databases (7%) (Table 1).

Table 1: Summary table of all real-world data source types by country

Country	Cross-sectional study*	EMR/EHR $	Insurance/claims databases #	Patient Registry ^	Total
Australia	1	3	0	19	23
China	3	36	1	14	54
Japan	2	14	4	25	45
South Korea	6	5	3	7	21
Taiwan	1	4	3	7	14
Total	13	62	11	71	

Abbreviations: EHR: Electronic Hospital Records; EMR: Electronic Medical Records

** Utilizing methodologies including national surveys to collect data at a given point in time*
$ Including standardized databases containing results of clinical and administrative encounters between patients and healthcare professionals
 # Defined as large-scale repositories where data, including health insurance claims, administrative details, or health information, was compiled on a national level, typically under the governance of a national authority or organisation
^Organized system of database with collected clinical data meant to evaluate specific outcomes for a population. It includes cancer, international, national, pharmacovigilance and regional registries and clinical data warehouse
Note: The total number of databases does not sum up to 152 as some patient registries are span across multiple countries

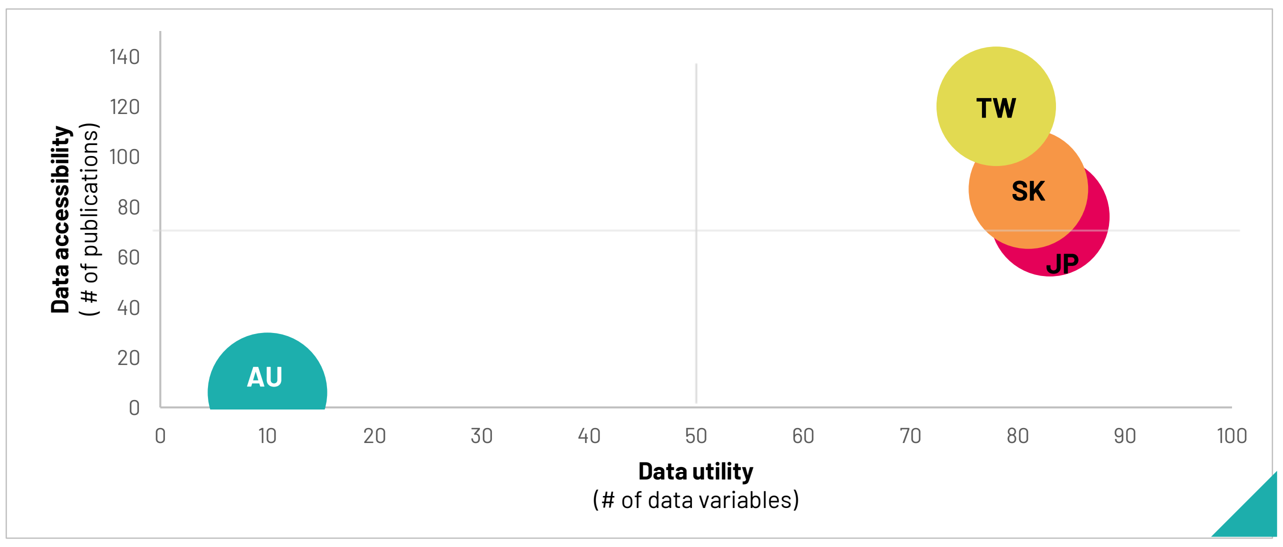
Data sources were prioritized based on number of publications, as a proxy for data access, and the top 25 data sources underwent an in-depth utility assessment. Data utility, measured by the breadth of variables available, was the most substantial in Japan, followed by South Korea , Taiwan and Australia (Figure 1).

The reported variables were classified into 12 categories:

1. Demographics data
2. Clinical data
3. Healthcare Resource Utilization (HCRU)
4. Treatment-related data
5. Lifestyle data
6. Quality of life (QoL) data
7. Direct cost data
8. Indirect cost data
9. LC-specific outcomes
10. Insurance claims data
11. Economic evaluation data
12. Epidemiology

The most widely reported variables were demographic, clinical, treatment, lifestyle and epidemiology data, while indirect costs and economic evaluation data, were less frequently captured.

Figure 1: RWD accessibility vs. utility by market



**None of the 25 prioritized data sources for an in-depth utility assessment comprise Chinese data sources.*

To contextualize the utility of the available data, RWD sources were evaluated against key requirements for RWE and health economic and outcomes research (HEOR) studies. The corresponding studies that can be conducted with the available data demonstrate the potential for evidence generation to support access and value demonstration for LC interventions in each market (Table 2).

Across the region, our analysis indicates that while conducting HEOR studies based on RWD sources is moderately feasible, limited variable requires further primary research to supplement with additional data.

Table 2: RWE and HEOR study feasibility based on available data in each market

Type of study	Variables required	Australia	Japan	South Korea	Taiwan
Cost of illness	2, 7, 8, 12	🟨	🟨	🟨	🟨
Burden of illness	2, 6, 7, 8, 9, 12	🟨	🟨	🟨	🟨
Utility/ HRQoL study	6	🟥	🟨	🟨	🟨
HCRU	2, 3, 4	🟥	🟨	🟩	🟩
Patient segmentation	1, 4, 12	🟨	🟩	🟩	🟩
Clinical effectiveness/ safety/ adverse events	2, 4	🟥	🟩	🟩	🟩
Treatment pathway/ compliance study	4	🟨	🟨	🟨	🟨
Budget impact analysis	3, 7, 10,12	🟨	🟨	🟨	🟨
Cost-effectiveness analysis	1,2,4,6, 7, 8, 12	🟨	🟨	🟨	🟨

Abbreviations: HCRU: healthcare resource utilisation; HRQoL: Health-Related Quality of Life
**None of the 25 prioritized data sources for an in-depth utility assessment comprise Chinese data sources.*

🟩 *High feasibility with comprehensive RWD (defined as data sources reporting information on all required variables i.e., robust coverage)*

🟨 *Moderate feasibility with some RWD (defined as data sources reporting information on some of the required variables i.e., less robust coverage and requires further primary research)*

🟥 *Not feasible (defined as data sources reporting information on none of the required variables i.e., poor coverage)*

CONCLUSIONS

Mapping LC RWD holds significant promise for evaluating the feasibility of RWE and Health HEOR studies. These activities are crucial for guiding subsequent evidence generation efforts (refer to Table 2: Market HEOR Map). China is a significant generator of RWE with the largest number of RWD, but access is a key challenge. Conversely, Australia, Japan, South Korea and Taiwan provide diverse datasets with greater access and utility that enhance understanding and management of LC.

Despite the substantial amount of real-world data, we identified two key challenges in its application. Firstly, the variability in data access often leaves critical gaps in evidence, highlighting the need for collaborative evidence generation initiatives. Secondly, data variable availability varies across markets. Australia, notably lacks utility/QoL, clinical, HCRU, and treatment-related data, and indirect cost data is scarce across all markets. This necessitates market-specific data collection (e.g., patient surveys, micro-costing studies, HRQoL questionnaires) to bridge these gaps and enhance the evidence base for assessing LC burden.

Our findings illuminate both the opportunities and complexities involved in applying RWD to LC research in the AP region. Pharmaceutical companies must adeptly navigate this intricate RWD landscape to formulate effective evidence generation strategies and choose robust data sources for retrospective, prospective, or hybrid RWE study designs. The expanding RWD ecosystem increases potential yet heightens the complexity, investment, and expertise required for evidence generation. Collaboration is imperative to harmonize data collection, broaden access to underrepresented data categories, and amplify the utility of RWD. These concerted efforts will elevate the quality of RWE studies and ultimately foster better decision-making and improve patient outcomes across the region.

REFERENCES

1. Asia Pacific Policy Review & Engagement for Lung Cancer (ASPIRE). Lung Cancer in APAC: Urgent call for action & comprehensive management. ASPIRE Lung Cancer. (2024). Accessed on 12 August 2025. Available at: https://aspirelungcancer.com/publications/Aspire_Infographic_v2.0.pdf
2. FP Analytics. Combating Lung Cancer Across the Asia Pacific. (2024). Accessed on 12 August 2025. Available at: https://combatinglungcancerasiapacific.com/
3. Kesmodel US. Cross-sectional studies – what are they good for? Acta Obstetricia Et Gynecologica Scandinavica. 2018;97(4):388-393. doi:10.1111/aogs.13331
4. R.E. Gliklich and N.A. Dreyer, "Registries for Evaluating Patient Outcomes: A User's Guide, 4th Edition" Agency for Healthcare Research and Quality, publication No. 19(20)-EHC020 (AHRQ, Rockville, MD, 2020).
5. Konrad R, Zhang W, Bjarndóttir M, Proaño R. Key considerations when using health insurance claims data in advanced data analyses: an experience report. Health Syst (Basingstoke). 2019 Mar 1;9(4):317-325. doi: 10.1080/20476965.2019.1581433. PMID: 33354323; PMCID: PMC7738306.
6. Ambinder EP. Electronic Health records. Journal of Oncology Practice. 2005;1(2):57-63. doi:10.1200/jop.2005.1.2.57

CONTACT INFORMATION

Louise Hogg – Director, RWE, Healthcare, Ipsos
Louise.Hogg@ipsos.com
www.Ipsos.com/en/healthcare