

Received : 8 Agustus 2025 Revised : 4 Maret 2026 Accepted : 20 April 2026

Narrative Review: Big Data BPJS Kesehatan untuk Penelitian dan Kebijakan Kesehatan Indonesia

✉¹Syarif Rahman Hasibuan, ²Mahar Santoso, ³Ede Surya Darmawan

¹Fakultas Kedokteran, Universitas Pembangunan Nasional Veteran Jakarta, Indonesia

²Pusat Sistem dan Strategi Kesehatan, Kementerian Kesehatan, Indonesia

³Departemen Administrasi dan Kebijakan Kesehatan, Fakultas Kesehatan Masyarakat, Universitas Indonesia, Indonesia

ABSTRAK

Artikel ini menjelaskan struktur data, metodologi pembentukan sampel, dan cakupan variabel yang tersedia. Selain itu, dijabarkan potensi penggunaannya untuk analisis utilisasi layanan kesehatan, tren penyakit, prediksi risiko, evaluasi kebijakan, serta estimasi beban kerja tenaga kesehatan. Data sampel ini juga memiliki peluang besar untuk dikaji dalam konteks kepesertaan ganda dan pengembangan skema coordination of benefits (CoB), terutama dalam integrasi pembiayaan JKN dan asuransi kesehatan swasta di Indonesia. Meskipun telah tersedia sejak 2019, pemanfaatannya dalam publikasi ilmiah masih terbatas dan belum terdokumentasi secara baku. Artikel ini mengidentifikasi tantangan utama dalam pemanfaatan data dan memberikan rekomendasi untuk mendorong penggunaannya melalui pelatihan teknis, penguatan tata kelola data, dan kolaborasi riset lintas sektor. Data ini memiliki potensi besar untuk menjadi sumber utama dalam penguatan sistem kesehatan nasional berbasis data di masa depan.

Kata kunci: BPJS Kesehatan, Data Sampel, JKN, Big Data, Kebijakan Kesehatan Berbasis Bukti

Narrative Review: Big Data from BPJS Kesehatan for Health Research and Policy in Indonesia

ABSTRACT

This article describes the dataset's structure, sampling methodology, and the scope of available variables. It also outlines its potential applications for analyzing healthcare service utilization, disease trends, risk prediction, policy evaluation, and estimating the healthcare workforce burden. Furthermore, the dataset offers significant opportunities for research on dual coverage and the development of a coordination of benefit (CoB) scheme, particularly for integrating JKN and private health insurance financing in Indonesia. Although available since 2019, its use in scientific publications remains limited and lacks standardized documentation. This article identifies key challenges in data utilization and provides recommendations to enhance its adoption through technical training, strengthened data governance, and cross-sector research collaboration. The dataset holds considerable potential to become a cornerstone for strengthening Indonesia's data-driven national health system in the future.

Keywords: BPJS Health, Sample Data, JKN, Big Data, Evidence-Based Health Policy.

PENDAHULUAN

Perkembangan teknologi digital telah membuka peluang besar untuk memanfaatkan data administratif kesehatan sebagai sumber utama penelitian berbasis populasi (Ahn, 2020). Pada riset kesehatan, data administratif merujuk pada data yang dihasilkan sebagai produk sampingan dari proses administrasi program, seperti pencatatan kepesertaan, kunjungan serta klaim layanan kesehatan. Data-data tersebut kemudian dimanfaatkan untuk tujuan penelitian dan evaluasi kebijakan. Sementara itu, istilah *big data* digunakan untuk menggambarkan karakteristik data yang berskala besar, bersifat longitudinal, dan memerlukan kapasitas komputasi serta pendekatan analitik khusus untuk pengolahannya (Batko & Ślęzak, 2022).

Korea Selatan adalah salah satu contoh negara yang berhasil menjadikan data sistem jaminan kesehatan nasionalnya, yaitu *National Health Insurance Database* (NHID), sebagai tulang punggung berbagai studi epidemiologis, kebijakan kesehatan, dan pemodelan klinis berskala nasional (Lee dkk., 2017; Lim & Jang, 2025). Keberhasilan ini dicapai melalui penyediaan data *kohort* berskala besar dan longitudinal, seperti NHIS-NSC dan NHIS-HEALS, yang mewakili populasi dan dilengkapi dengan variabel sosial-demografis, diagnosis, pelayanan kesehatan, hingga hasil skrining berkala (Kim dkk., 2022). Selama kurun waktu 2014–2019, terdapat lebih dari 2.500 publikasi ilmiah yang memanfaatkan data dari NHID dan HIRA Korea Selatan (Kim dkk., 2020).

Perlu dicatat bahwa pemanfaatan data perbandingan dengan Korea Selatan dalam artikel ini dimaksudkan sebagai ilustrasi potensi, bukan sebagai standar yang harus segera dicapai. Terdapat perbedaan signifikan dalam hal kematangan infrastruktur data, ekosistem penelitian, serta integrasi data administratif ke dalam proses kebijakan antara kedua negara. Korea Selatan telah membangun sistem data kesehatan nasional selama lebih dari dua dekade dengan dukungan regulasi yang kuat dan investasi berkelanjutan dalam kapasitas analitik. Indonesia, meskipun memiliki potensi serupa melalui cakupan JKN yang luas, masih dalam tahap awal pengembangan ekosistem penelitian berbasis data administratif.

Indonesia memiliki potensi besar melalui program Jaminan Kesehatan Nasional (JKN) yang diselenggarakan oleh BPJS Kesehatan. Sejak 2014, program ini telah mencatat hampir seluruh interaksi pelayanan kesehatan masyarakat Indonesia. Dalam mendukung pemanfaatan data ini untuk penelitian dan kebijakan, BPJS Kesehatan mulai merilis data sampel peserta JKN sejak tahun 2019 (BPJS Kesehatan, 2020). Data ini kini mencakup informasi kepesertaan dan klaim pelayanan sejak 2015 hingga 2023, serta ditambahkan jenis data kontekstual untuk isu prioritas seperti Diabetes Mellitus, Tuberkulosis, Kesehatan Ibu dan Anak, Kesehatan Mental, serta yang terbaru yaitu layanan pasien Covid-19 (BPJS Kesehatan, 2025). Struktur data ini bersifat longitudinal dan dirancang untuk mewakili seluruh wilayah Indonesia, dengan sistem pengambilan sampel berbasis rumah tangga dan stratifikasi wilayah serta pola pemanfaatan layanan (BPJS Kesehatan RI, 2023).

Sebagai upaya mendorong pemanfaatan data, BPJS Kesehatan telah menerbitkan dokumen teknis yang menjelaskan struktur, metodologi sampling, dan isi variabel secara rinci (BPJS Kesehatan RI, 2023). Namun, hingga saat ini belum tersedia tinjauan yang secara komprehensif mendeskripsikan struktur data ini sekaligus merefleksikan pengalaman penggunaannya dalam praktik penelitian. Sebagian besar informasi yang tersedia bersifat teknis-normatif, sementara tantangan praktis yang dihadapi peneliti dalam mengakses dan menganalisis data belum banyak terdokumentasi.

Artikel ini bertujuan untuk menyajikan tinjauan deskriptif dan reflektif terhadap struktur, isi, serta peluang dan tantangan dalam pemanfaatan data ini. Penelusuran literatur yang dilakukan bersifat indikatif untuk menggambarkan kondisi pemanfaatan data saat ini, bukan sebagai evaluasi sistematis terhadap seluruh publikasi yang ada. Dengan menyampaikan kekuatan dan keterbatasan data berdasarkan pengalaman penggunaan nyata, kami berharap dapat memberikan panduan praktis bagi peneliti dan pembuat kebijakan dalam mengoptimalkan potensi data sampel BPJS Kesehatan.

METODE PENELITIAN

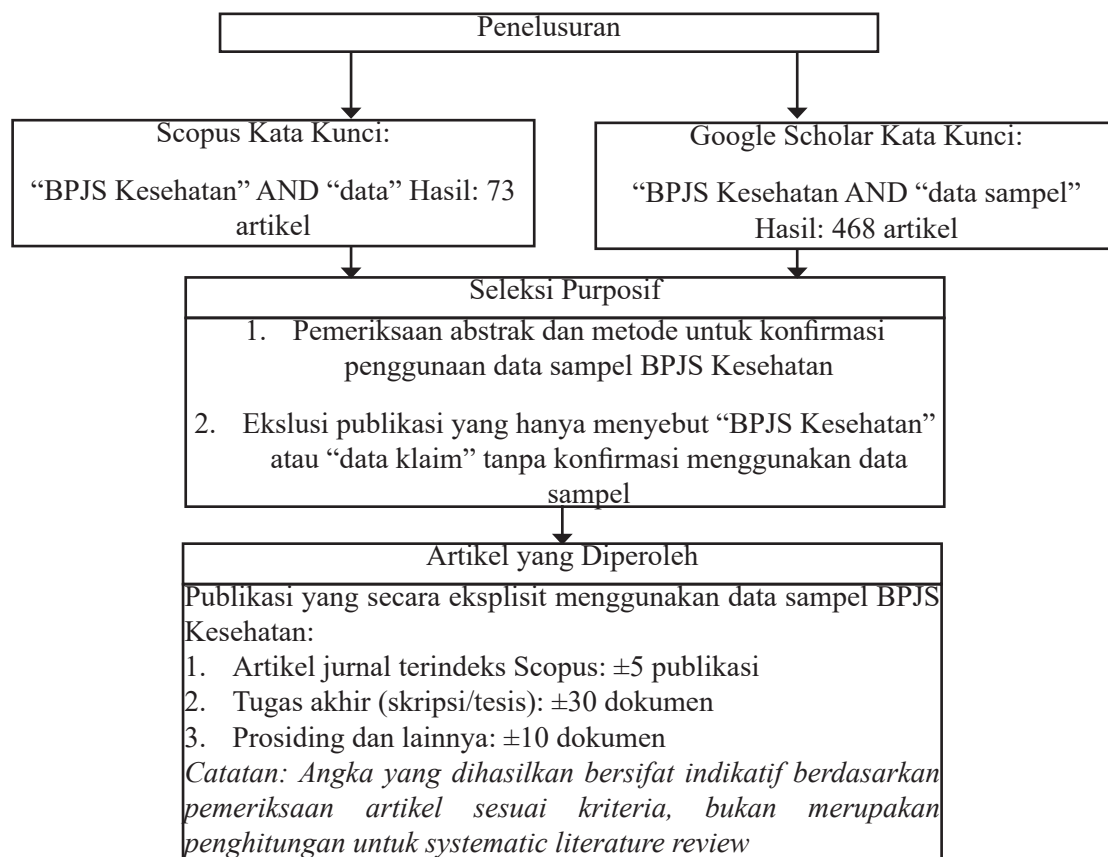
Kajian ini disusun menggunakan pendekatan *narrative* dan *practical review*, yaitu bentuk tinjauan yang menggabungkan dua komponen utama: (1) komponen naratif, berupa sintesis dokumentasi teknis dan penelusuran literatur; serta (2) komponen praktis, berupa refleksi pengalaman langsung penulis dalam mengakses dan menganalisis data sampel BPJS Kesehatan. Komponen praktis berfungsi untuk mengidentifikasi tantangan teknis dan metodologis yang tidak selalu terdokumentasi dalam panduan resmi, namun krusial bagi peneliti yang akan memanfaatkan data ini. Pendekatan serupa telah digunakan dalam berbagai studi untuk menilai kesesuaian dan kesiapan sumber data sekunder sebagai dasar analisis ilmiah (Antonelli dkk., 2020; Shaffril dkk., 2020).

Komponen naratif dalam kajian ini bersumber dari dua jenis referensi. Referensi utama yaitu dokumen teknis dari BPJS Kesehatan yang memuat penjelasan rinci

mengenai metodologi pemilihan sampel, termasuk unit observasi, skema stratifikasi, serta metode pembobotan (BPJS Kesehatan RI, 2023). Berbagai artikel dari penelusuran literatur yang bersifat indikatif menggambarkan kondisi pemanfaatan data saat ini. Penelusuran yang dilakukan bukan sebagai tinjauan sistematis terhadap seluruh publikasi yang ada.

Penelusuran literatur dilakukan pada tanggal 15 April 2025 melalui dua basis data ilmiah dengan rincian sebagai berikut: (a) Basis data *Scopus* dengan menggunakan kata kunci "BPJS Kesehatan" AND "data", tanpa filter tambahan (tahun, bahasa, atau jenis dokumen). Berdasarkan hasil pencarian diperoleh 73 artikel. (b) Basis data *Google Scholar* dengan menggunakan kata kunci "BPJS Kesehatan" AND "data sampel", mencakup seluruh jenis dokumen. Berdasarkan hasil pencarian diperoleh 468 artikel.

Tujuan penelusuran adalah untuk memperoleh gambaran indikatif, sehingga pada pencarian tidak diterapkan kriteria inklusi



Gambar 1
Alur Penelusuran Literatur

dan eksklusif. Pemilihan publikasi untuk pembahasan dilakukan secara purposif dengan meninjau bagian abstrak dan metode untuk mengidentifikasi publikasi yang secara eksplisit menyatakan penggunaan data sampel BPJS Kesehatan. Publikasi yang hanya menyebut "BPJS Kesehatan", "data klaim", atau "JKN" tanpa konfirmasi penggunaan data sampel tidak dimasukkan dalam pembahasan utama, meskipun tetap dihitung dalam total hasil pencarian. Alur penelusuran literatur disajikan pada Gambar 1.

Komponen praktis dalam kajian ini berasal dari pengalaman langsung penulis dalam mengakses dan menganalisis data sampel BPJS Kesehatan untuk berbagai proyek penelitian. Pengalaman ini mencakup identifikasi isu kualitas data (seperti nilai usia negatif akibat kesalahan pencatatan, inkonsistensi status kematian antara data FKTP dan kepesertaan, serta kelengkapan variabel), pemahaman terhadap struktur hierarkis data, dan tantangan teknis dalam penerapan bobot *sampling*. Refleksi praktis ini menjadi pelengkap penting terhadap informasi yang tersedia dalam dokumentasi resmi.

Evaluasi dilakukan secara naratif-reflektif, dan hasilnya dibahas dengan membandingkannya terhadap pemanfaatan data sampel di negara lain yang menerapkan sistem jaminan kesehatan nasional, khususnya Korea Selatan. Perbandingan ini dimaksudkan sebagai ilustrasi potensi, dengan tetap mengakui perbedaan konteks infrastruktur data, ekosistem

penelitian, dan integrasi kebijakan antara kedua negara. Artikel ini tidak melibatkan data primer dari subjek manusia secara langsung, tidak menggunakan informasi identitas pribadi, serta hanya memanfaatkan data sekunder yang telah dianonimkan, maka tidak diperlukan persetujuan etik dari komite etik manapun.

HASIL DAN PEMBAHASAN

Struktur Data dan Variabel

Data sampel BPJS Kesehatan terdiri atas lima jenis data utama yaitu (1) kepesertaan; (2) data kunjungan fasilitas kesehatan tingkat pertama (FKTP) kapitasi; (3) data FKTP non-kapitasi dan tindakan; (4) data kunjungan fasilitas kesehatan rujukan tingkat lanjut (FKRTL); serta (5) data diagnosis sekunder pada kunjungan FKRTL. Seluruh berkas ini bersifat longitudinal, mencatat riwayat pelayanan peserta dari tahun ke tahun, dan dapat digunakan untuk melacak tren maupun membentuk studi kohort retrospektif. Setiap peserta memiliki kode unik yang merupakan nomor peserta yang sudah diacak untuk menjaga kerahasiaan data. Untuk memberikan gambaran lebih sistematis, Tabel 1 menyajikan klasifikasi ringkas variabel yang tersedia dalam data sampel berdasarkan kategori fungsional.

Data kepesertaan mencakup variabel demografi dan administratif seperti tanggal lahir, jenis kelamin, segmentasi peserta, kelas rawat, serta wilayah tempat tinggal. Variabel wilayah dicatat dalam bentuk provinsi dan kabupaten/kota, yang memungkinkan penggabungan

Tabel 1
Struktur Variabel Data Sampel BPJS Kesehatan

Kategori Variabel	Variabel dalam Data Sampel BPJS
Identifikasi peserta	Jenis kelamin, usia, segmentasi peserta (PBI, PBPU, PPU, BP)
Pelayanan kesehatan	Tanggal kunjungan, jenis layanan, INACBG
Biaya dan klaim	Tarif INA-CBG, jenis prosedur
Lokasi	Kota/kabupaten FKTP, domisili peserta
Diagnosis	CMG, ICD utama, sekunder
PBI	: Penerima Bantuan Iuran
PBPU	: Pekerja Bukan Penerima Upah
PPU	: Pekerja Penerima Upah
BP	: Bukan Pekerja
INA-CBG	: Indonesia Case Based Group
CMG	: Casemix Main Group

Sumber: Data Primer Diolah

dengan data makro eksternal dari Badan Pusat Statistik. Meskipun variabel seperti usia dan lama rawat inap tidak tersedia secara langsung, peneliti dapat mengkalkulasikannya dari kombinasi variabel tanggal lahir dan tanggal kunjungan (untuk usia), serta tanggal masuk dan keluar layanan (untuk *length of stay*). Praktik semacam ini juga lazim digunakan dalam analisis data asuransi kesehatan di negara lain, seperti Korea Selatan (Kim dkk., 2022).

Isu Kualitas Data

Struktur variabel dan label dalam data ini umumnya mudah dipahami, namun terdapat beberapa permasalahan teknis yang perlu diperhatikan dalam proses analisis. Misalnya, ditemukan data dengan usia negatif sebagai akibat kesalahan pencatatan tanggal lahir atau tanggal kunjungan, serta variabel status perkawinan yang sering kali tidak terisi. Dalam diagnosis, kode ICD Z kadang muncul tanpa diagnosis sekunder, menyulitkan identifikasi penyakit. Ketidaksesuaian juga ditemukan antara status pulang meninggal di data FKTP dengan status hidup di data kepesertaan. Selain itu, peserta dapat mengalami perubahan segmentasi dan kelas iuran dari tahun ke tahun, misalnya dari PBI ke PBPU atau sebaliknya, serta perpindahan antara kelas 1, 2, dan 3. Pada data FKRTL, ditemukan kasus ketika kode INACBG mengindikasikan adanya prosedur, tetapi kolom jenis tindakan berdasarkan ICD-9-CM tidak terisi, menimbulkan ambiguitas.

Metodologi Pengambilan Sampel dan Pembobotan

Secara metodologis, pemilihan sampel dilakukan melalui *stratified random sampling*, dengan unit *sampling* utama adalah keluarga. Stratifikasi didasarkan pada tiga kategori pemanfaatan layanan (FKTP, FKRTL, dan tidak berkunjung), serta fasilitas layanan kesehatan tempat peserta terdaftar. Desain ini menghasilkan stratifikasi hingga tiga kali jumlah FKTP dan satu keluarga diambil secara acak dari setiap strata. Akibatnya, desain *sampling* bersifat tidak proporsional dan memerlukan bobot untuk menjamin representativitas (BPJS Kesehatan RI, 2023). Tahapan pemilihan sampel disajikan pada Tabel 2.

Pemilihan peserta dalam data sampel BPJS Kesehatan dilakukan secara bertahap berdasarkan tahun bergabungnya peserta ke dalam program JKN. Terdapat dua subpopulasi utama dalam proses ini, yaitu (1) peserta yang telah bergabung sebelum tahun perilisan data sampel; dan (2) peserta baru yang mulai terdaftar sebagai peserta aktif pada tahun perilisan data sampel. Subpopulasi kedua mencerminkan peserta yang sebelumnya belum tercatat dalam sistem dan baru tercakup pada tahun terakhir periode data. Hal ini mencerminkan strategi perluasan cakupan peserta, serta memungkinkan analisis longitudinal dan kohort baru secara lebih komprehensif.

Tabel 2
Tahapan Pemilihan Sampel dalam Dataset BPJS Kesehatan

Tahap Proses	Deskripsi
Pembentukan Strata	Peserta dikelompokkan ke dalam tiga tipe berdasarkan riwayat akses layanan: Strata 1: Pernah akses FKTP Strata 2: Pernah akses FKRTL Strata 3: Belum pernah akses layanan
Unit <i>Sampling</i>	Unit sampling adalah keluarga, dipilih secara acak dari strata yang telah dibentuk.
Unit Pengamatan	Unit observasi adalah individu/ peserta dalam keluarga terpilih.
Pemberian Bobot	Setiap individu diberikan bobot berdasarkan strata wilayah, kelompok usia, dan jenis kelamin, agar mewakili 1% populasi peserta JKN.

Sumber: Data Primer Dioah

Pemberian bobot pada data sampel BPJS Kesehatan dilakukan untuk mengatasi potensi bias akibat peluang pemilihan sampel yang tidak merata antar kelompok. Hal ini disebabkan oleh metode sampling yang digunakan, yaitu *stratified random sampling*. Langkah pertama adalah menghitung bobot pada tingkat keluarga. Dalam satu strata tertentu, misalnya terdapat 6.200 keluarga sebagai populasi, dan hanya satu keluarga diambil sebagai sampel. Maka peluang keluarga terpilih adalah 1 dibagi 6.200 (sekitar 0,00016), dan bobot yang diberikan kepada keluarga tersebut adalah kebalikannya, yaitu 6.200. Dengan kata lain, satu keluarga sampel mewakili 6.200 keluarga dalam strata tersebut. Proses ini diulang untuk setiap strata, sehingga bobot yang diberikan bervariasi tergantung pada ukuran populasi dalam strata tersebut.

Setelah bobot keluarga dihitung, tahap berikutnya adalah pemberian bobot individu, yaitu dengan mendistribusikan bobot keluarga kepada setiap anggota individu di dalam keluarga tersebut. Misalnya, jika satu keluarga beranggotakan lima orang dan memiliki bobot keluarga 6.200, maka masing-masing individu mewakili 6.200 peserta (dengan bobot individu awal yang identik). Langkah terakhir adalah penyesuaian bobot (*adjustment*) agar distribusi karakteristik individu dalam sampel

lebih merepresentasikan struktur populasi peserta JKN secara nasional. Penyesuaian dilakukan berdasarkan variabel penting seperti jenis kelamin, kelompok usia, dan segmen kepesertaan (misalnya PBI, PBPU, atau PPU). Proses ini menghasilkan bobot akhir per individu yang digunakan dalam analisis untuk memperoleh estimasi yang representatif terhadap populasi. Tahapan penentuan bobot sampel secara ringkas disajikan pada Tabel 3.

Sampai dengan periode rilis data terakhir, jumlah peserta dalam data sampel telah mencapai 2.501.251 orang pada tahun 2023 (Tabel 4). Setelah data dibersihkan dan menyisakan peserta yang hidup sampai 31 Desember 2023, jumlah peserta dalam data sampel tahun 2023 tercatat sebesar 2.424.785 orang. Setelah bobot diterapkan, estimasi total peserta JKN hidup berdasarkan data sampel menjadi 282.460.909 orang. Angka ini sedikit lebih tinggi dibandingkan jumlah peserta JKN yang tercatat secara resmi oleh DJSN pada 31 Desember 2023, yaitu sebesar 267.311.566 orang (Dewan Jaminan Sosial Nasional, 2024). Perbedaan ini mengindikasikan bahwa proses pemberian bobot belum sepenuhnya dikalibrasi terhadap populasi rujukan terkini. Hal ini tidak selalu menunjukkan kesalahan, tetapi menegaskan bahwa bobot dalam data sampel BPJS Kesehatan lebih cocok digunakan

Tabel 3
Tahapan Pembobotan dalam Data Sampel BPJS Kesehatan

Langkah	Deskripsi
1. Pembentukan strata	Data dikelompokkan berdasarkan kombinasi FKTP dan kategori keluarga menurut riwayat akses layanan (akses FKTP, FKRTL, atau tidak pernah).
2. Perhitungan bobot keluarga	Bobot dihitung sebagai kebalikan dari peluang terpilihnya satu keluarga dalam strata tertentu. Contoh: jika 1 dari 6.200 keluarga dipilih, bobotnya adalah 6.200.
3. Distribusi bobot ke individu	Bobot keluarga didistribusikan secara merata ke semua anggota keluarga yang terpilih. Jika satu keluarga berisi lima orang dan bobot keluarga 6.200, maka masing-masing individu dianggap mewakili 6.200 peserta.
4. Penyesuaian bobot individu	Dilakukan penyesuaian agar distribusi sampel individu mencerminkan populasi nasional berdasarkan jenis kelamin, kelompok usia, dan segmen kepesertaan (PBI, PBPU, PPU).
5. Penggunaan bobot akhir	Bobot akhir individu digunakan dalam analisis untuk menghasilkan estimasi yang representatif terhadap populasi peserta JKN.

Sumber: Data Primer Diolah

untuk estimasi proporsional dan inferensial, bukan untuk menghasilkan angka absolut populasi secara langsung tanpa penyesuaian tambahan. Harapannya pada perilisan tahun-tahun berikutnya, penyesuaian bobot berbasis *post-stratification* terhadap sumber populasi resmi seperti BPJS Kesehatan, DJSN atau Kementerian Dalam Negeri dapat meningkatkan akurasi representasi total. Perbandingan jumlah peserta JKN dalam data sampel, estimasi berbobot, dan populasi riil disajikan pada Tabel 4.

Bobot dalam data sampel BPJS Kesehatan dihitung untuk memastikan bahwa peserta sampel mewakili struktur populasi peserta JKN berdasarkan strata wilayah, usia, jenis kelamin, dan segmen kepesertaan. Namun, hingga saat ini belum tersedia penjelasan resmi atau validasi yang menunjukkan bahwa frekuensi kunjungan atau klaim yang dilakukan oleh peserta sampel juga merepresentasikan total utilisasi layanan pada tingkat populasi. Oleh karena itu, analisis pada tingkat layanan (kunjungan atau klaim) perlu dilakukan dengan kehati-hatian dalam interpretasi, dan sebaiknya disertai perbandingan terhadap data agregat nasional jika tersedia. Hal ini penting untuk menghindari generalisasi yang berlebihan terhadap pola pemanfaatan layanan berdasarkan data sampel (Lohr, 2021).

Pemanfaatan Data Sampel BPJS Kesehatan untuk Publikasi Ilmiah

Topik yang diangkat dalam publikasi yang secara eksplisit menggunakan data sampel BPJS Kesehatan umumnya berkaitan dengan analisis utilisasi layanan kesehatan berdasarkan diagnosis tertentu seperti tuberkulosis, diabetes melitus, dan hipertensi; evaluasi dampak kebijakan seperti kapitasi berbasis kinerja; serta studi mengenai perbedaan akses dan

penggunaan layanan antar wilayah atau antar jenis fasilitas kesehatan tingkat pertama (FKTP). Selain itu, terdapat pula studi yang mengeksplorasi penggunaan data untuk evaluasi layanan promotif dan preventif, serta analisis hubungan antara status kepesertaan dan pemanfaatan layanan rujukan atau rawat inap tingkat lanjut.

Beberapa publikasi yang terindeks di Scopus secara eksplisit mencantumkan penggunaan data sampel JKN, antara lain studi mengenai mortalitas rumah sakit, analisis clinical pathway menggunakan pendekatan *process mining*, dan evaluasi kelangsungan hidup pasien dengan diabetes melitus. Judul-judul seperti “*Behind the Hospital Ward: In-Hospital Mortality of Type 2 Diabetes Mellitus Patients in Indonesia (Analysis of National Health Insurance Claim Sample Data)*” (Darmawan *et al.*, 2024) dan “*Potentials of Clinical Pathway Analysis Using Process Mining on the Indonesia National Health Insurance Data Samples*” (Kurniati dkk., 2022) mengindikasikan bahwa pemanfaatan data ini telah mulai merambah pada metodologi yang lebih kompleks dan bersifat prediktif.

Meskipun belum dilakukan tinjauan sistematis terhadap isi penuh dari artikel-artikel tersebut, indikasi awal menunjukkan bahwa belum banyak publikasi yang membahas keterbatasan teknis atau metodologis dari data sampel BPJS Kesehatan. Berdasarkan pengalaman pengguna dan pengamatan terhadap bagian metode beberapa artikel, tantangan yang umum ditemukan meliputi validitas kode diagnosis, nilai kosong pada variabel tertentu, serta keterbatasan variabel klinis dan perilaku. Permasalahan-permasalahan ini sejalan dengan temuan dari literatur internasional terkait pemanfaatan data

Tabel 4
Perbandingan Jumlah Peserta JKN dalam Data Sampel, Estimasi Berbobot, dan Populasi Riil per 31 Desember 2023

Tahap	Jumlah
Total sampel (mentah) 2023	2.501.251
Setelah eksklusi peserta wafat sebelum 31 Des 2023	2.424.785
Estimasi populasi berbobot (hidup per 31 Des 2023)	282.460.909
Populasi JKN riil (DJSN, 31 Des 2023)	267.311.566

Sumber: Data Primer Diolah

sampel kesehatan, sebagaimana dilaporkan dalam studi-studi mengenai database NHIS Korea Selatan (Song dkk., 2014).

Hasil kajian ini menunjukkan bahwa meskipun data sampel BPJS Kesehatan memiliki struktur longitudinal, cakupan populasi yang luas, serta potensi yang besar untuk penelitian kesehatan publik, dokumentasi pemanfaatannya dalam literatur ilmiah nasional masih terbatas. Sejauh ini, sebagian besar publikasi yang menyebutkan data sampel BPJS Kesehatan berfokus pada topik penyakit tertentu atau studi deskriptif utilisasi atau pemanfaatan layanan, namun belum banyak yang menjangkau metode prediktif, analisis spasial, maupun evaluasi kebijakan kesehatan berbasis data (Fatimah dkk., 2025).

Potensi pemanfaatan data sampel BPJS Kesehatan dalam penelitian kesehatan berbasis data sekunder sangat luas dan belum tergarap optimal. Topik-topik yang direkomendasikan secara resmi oleh BPJS Kesehatan antara lain (1) utilisasi layanan; (2) pola penyakit; (3) prediksi risiko; (4) angka readmisi, (5) jenis pengobatan; dan (6) efisiensi biaya klaim (BPJS Kesehatan, 2025).

Ruang eksplorasi lainnya yang belum banyak dikembangkan di Indonesia meliputi analisis survival berbasis klaim, pengembangan model segmentasi risiko populasi JKN, evaluasi efektivitas biaya berbasis episode penyakit (*episode-based costing*), dan identifikasi rawat inap yang dapat dicegah (*preventable hospitalization*). Selain itu, struktur data yang mencakup tanggal kunjungan secara spesifik meliputi hari, bulan, dan tahun memungkinkan pengembangan analisis berbasis waktu (*day-level analysis*). Hal ini dapat dimanfaatkan, misalnya, untuk mengidentifikasi tren harian kejadian penyakit tertentu seperti demam berdarah *dengue* (DBD), influenza, atau penyakit kronis lainnya, menganalisis pola musiman penyakit, hingga melihat dampak langsung intervensi kebijakan terhadap volume kunjungan secara *real-time*. Contohnya, analisis dapat menunjukkan lonjakan klaim DBD pada bulan-bulan tertentu yang konsisten dari tahun ke tahun, atau adanya penurunan kunjungan selama periode kebijakan pembatasan mobilitas.

Data ini juga dapat digunakan untuk menghitung beban kerja tenaga kesehatan di FKTP dan FKRTL, memproyeksikan kebutuhan SDM berdasarkan klaim, mengevaluasi kesetaraan akses layanan secara geografis melalui integrasi dengan data spasial, serta menilai dampak reformasi kebijakan terhadap pola pelayanan secara kuasi-eksperimental. Beberapa studi juga menunjukkan potensi integrasi data klaim dengan *machine learning* untuk membangun sistem prediksi pasien risiko tinggi atau pembiayaan katastrofik. Dengan cakupan data longitudinal, tersedia pula peluang besar untuk membangun kohort JKN yang memungkinkan analisis perubahan status peserta, *switching segmen*, atau transisi diagnosis kronis. Keragaman potensi ini memperkuat posisi data sampel BPJS Kesehatan sebagai sumber utama dalam transformasi sistem kesehatan berbasis bukti di Indonesia, serta sebagai fondasi bagi penelitian populasi jangka panjang yang dapat diakui di tingkat internasional.

Pemanfaatan data sampel dari sistem jaminan sosial kesehatan di berbagai negara telah berkembang secara signifikan dan mencerminkan arah yang lebih strategis. Di Taiwan, misalnya, *National Health Insurance Research Database* (NHIRD) telah dimanfaatkan secara ekstensif untuk menghasilkan bukti dunia nyata dalam pengambilan keputusan klinis dan perumusan kebijakan publik. Meskipun validitas kode diagnosis dan keterbatasan variabel non-klinis menjadi tantangan, upaya validasi metodologis dan integrasi dengan data rumah sakit telah meningkatkan ketepatan analisis dan cakupan temuan (Hsieh dkk., 2019). Model ini diperkuat dengan kebijakan regulasi akses yang ketat namun terpusat, seperti *Health and Welfare Data Center* (HWDC), yang memungkinkan koneksi aman dan integrasi dengan lebih dari 70 basis data kesehatan lainnya.

Berbagai studi internasional telah menunjukkan bahwa pemanfaatan data klaim asuransi kesehatan nasional memerlukan dukungan kapasitas teknis yang memadai dalam hal perangkat lunak, pendekatan statistik, serta validasi data. Di Korea Selatan, analisis berbasis data NHIS banyak meng-

gunakan teknik lanjutan seperti *propensity score matching* (PSM), *inverse probability of treatment weighting* (IPTW), analisis *survival Cox*, dan pendekatan kausal berbasis *machine learning* yang lebih fleksibel dalam mengatasi bias observasional (Kim, 2025). Sementara itu, di Taiwan, pemanfaatan data NHIRD ditingkatkan melalui validasi kode diagnosis dan integrasi data rumah sakit dengan teknik *natural language processing* dan kecerdasan buatan, yang membuka peluang studi multidimensi berbasis data besar (Hsieh dkk., 2019). Di Jepang, studi menggunakan *National Database* (NDB) berfokus pada pemodelan pola pengobatan dan tren pelayanan kesehatan, meskipun keterbatasan dalam variabel klinis dan identifikasi pasien tetap menjadi tantangan yang diakui secara luas (Suto dkk., 2023). Pengalaman dari negara-negara ini menunjukkan bahwa pengembangan pemanfaatan data sampel BPJS Kesehatan di Indonesia juga memerlukan penguatan kapasitas analisis statistik lanjutan dan strategi validasi data yang sistematis.

Berbagai studi di Korea Selatan menggunakan data dari *National Health Insurance Claims Database* untuk menganalisis ketimpangan pelayanan kesehatan dan hasil klinis berdasarkan jenis kelamin, jenis asuransi, serta kondisi sosial-ekonomi. Penelitian menunjukkan bagaimana perbedaan akses layanan kesehatan berdampak pada mortalitas rumah sakit akibat infark miokard (Hong & Kang, 2015). Di sisi lain, penelitian lain memanfaatkan data nasional untuk mengevaluasi tren penyakit kronis seperti migrain, beserta pilihan terapi dan implikasi biayanya dalam kebijakan kesehatan (Lee dkk., 2023).

Berbagai artikel di jurnal internasional menunjukkan bahwa integrasi data lintas instansi, merupakan fondasi penting bagi ekosistem riset dan kebijakan berbasis bukti. Australia melalui *Multi-Agency Data Integration Project* (MADIP) telah menggabungkan data populasi berskala nasional dari berbagai sumber seperti perpajakan, layanan sosial, dan kesehatan, yang secara aktif digunakan dalam perumusan kebijakan publik berbasis data longitudinal (Kindermann dkk., 2017). Selandia Baru melalui *Integrated Data Infrastructure* (IDI)

bahkan mengintegrasikan data dari sektor kesehatan, pendidikan, perpajakan, hingga migrasi, sehingga memungkinkan evaluasi intervensi lintas sektor secara komprehensif (Atkinson & Blakely, 2017). Integrasi data tersebut dapat dijadikan contoh oleh Indonesia yang saat ini memiliki potensi besar melalui data JKN, namun belum sepenuhnya terintegrasi dengan sumber data lainnya.

Jika sistem informasi JKN dapat terhubung secara terstruktur dengan sumber data lainnya seperti data kependudukan (Dukcapil), data kematian, sistem surveilans penyakit, serta rekam medis digital dari sektor swasta dan penyedia asuransi kesehatan swasta, maka ekosistem riset akan menjadi jauh lebih kuat dan menyeluruh. Selain itu, pengembangan skema *coordination of benefit* (CoB) antara JKN dan asuransi swasta tidak hanya berpotensi meningkatkan efisiensi pembiayaan dan keadilan akses, tetapi juga memperkaya sumber data untuk mengevaluasi segmentasi risiko, duplikasi layanan, dan efektivitas intervensi lintas sistem pembiayaan. Hal ini akan membuka jalan bagi kolaborasi data publik-swasta yang aman, etis, dan berorientasi pada kepentingan sistem kesehatan nasional.

Negara-negara dengan sistem jaminan kesehatan nasional menunjukkan bahwa data klaim asuransi kesehatan digunakan tidak hanya untuk penelitian, tetapi juga secara langsung menjadi dasar pembuatan kebijakan publik. Sebuah studi, menggunakan data dari NHIRD untuk mengidentifikasi tren penyakit dan beban ekonomi dari berbagai intervensi, termasuk terapi statin dan metformin (Shih dkk., 2020). Salah satu hasilnya adalah kebijakan pengendalian biaya melalui pembatasan indikasi klinis dan negosiasi ulang harga untuk obat-obatan dengan volume tinggi. Selain itu, hasil kajian terhadap *drug reimbursement lag* juga mendorong pemerintah Taiwan untuk mempercepat proses kompensasi terhadap obat antikanker dan *imunomodulator* agar tidak tertinggal dari negara maju lain seperti Jepang dan Inggris (Shih dkk., 2020).

Data dari NHIS Korea Selatan, digunakan untuk mengidentifikasi variasi layanan berdasarkan karakteristik rumah sakit dan pasien (Kim dkk., 2022). Sebuah studi yang menganalisis perawatan rawat inap pasca

fraktur pinggul di Jepang, Korea, dan Taiwan menunjukkan disparitas signifikan dalam durasi rawat dan biaya perawatan, mendorong evaluasi ulang efisiensi rumah sakit dan penggunaan prosedur ortopedi invasif minimal di Korea (Kim dkk., 2021). Studi-studi semacam ini juga menjadi landasan dalam pengembangan sistem evaluasi rumah sakit nasional yang lebih berbasis luaran klinis.

Sementara itu di Jepang, data dari *National Database of Health Insurance Claims and Specific Health Checkups* (NDB) telah digunakan untuk memetakan variasi praktik klinis urologi antar prefektur. Hasil temuan, yang menunjukkan ketimpangan akses terhadap operasi invasif minimal, seperti laparoskopi, telah mendorong Kementerian Kesehatan untuk menetapkan standar nasional dan mengembangkan sistem pelatihan terstruktur untuk dokter spesialis di wilayah dengan keterbatasan praktik (Sugihara dkk., 2018).

Pengalaman negara-negara seperti Taiwan, Korea Selatan, dan Jepang menunjukkan bahwa data administratif dari sistem jaminan kesehatan dapat menjadi fondasi penting dalam proses perumusan kebijakan publik yang berbasis bukti. Refleksi terhadap konteks Indonesia menunjukkan bahwa data sampel BPJS Kesehatan memiliki potensi serupa, karena mencakup informasi longitudinal tentang kepesertaan dan pemanfaatan layanan pada skala populasi nasional. Pemanfaatan data sampel BPJS Kesehatan yang optimal dapat mendukung pembuatan kebijakan untuk berbagai hal seperti, perbaikan alokasi pembiayaan, evaluasi efektivitas kebijakan rujukan, penghitungan kebutuhan tenaga kesehatan, hingga pemetaan ketimpangan layanan antarwilayah. Untuk itu, penguatan kapasitas analitik, integrasi lintas sektor, dan penyusunan standar penggunaan data menjadi langkah penting agar data sampel tidak hanya menjadi sarana penelitian akademik, tetapi juga alat utama dalam transformasi sistem kesehatan Indonesia menuju arah yang lebih adil dan efisien.

Pembahasan ini juga perlu mencermati aspek representasi dalam struktur data sampel BPJS Kesehatan. Meskipun unit analisis peserta telah dirancang untuk merepresentasikan

sekitar 1% populasi peserta JKN melalui pendekatan stratifikasi dan pembobotan, masih terdapat keterbatasan dalam menjamin bahwa frekuensi dan intensitas kunjungan yang tercatat mencerminkan pola pemanfaatan layanan di populasi secara proporsional. Hal ini menjadi krusial mengingat Indonesia memiliki karakteristik geografis yang sangat beragam serta ketimpangan distribusi fasilitas kesehatan yang dapat memengaruhi perilaku kunjungan peserta. Oleh karena itu, evaluasi lebih lanjut terhadap validitas representasi utilisasi layanan termasuk pada tingkat kunjungan atau klaim perlu menjadi agenda prioritas dalam pengembangan metodologi data sampel, agar analisis yang dihasilkan semakin akurat dan aplikatif bagi kebijakan.

Sebagai upaya untuk meningkatkan transparansi ilmiah dan memperkuat pengakuan terhadap sumber data nasional, setiap publikasi yang menggunakan data sampel BPJS Kesehatan sebaiknya secara eksplisit men-cantumkan pernyataan *acknowledgement* kepada BPJS Kesehatan sebagai penyedia data. Selain itu, diperlukan standar penulisan metodologi atau data *citation* yang konsisten, sehingga setiap artikel yang memanfaatkan data ini dapat dilacak secara sistematis, baik dalam pangkalan data jurnal maupun sistem bibliometrik. Praktik semacam ini telah diterapkan secara luas di Korea Selatan dan Taiwan, di mana penulis menyertakan kode akses atau nama resmi *database* seperti “NHIS-NSC” atau “NHIRD” dalam bagian metode dan ucapan terima kasih. Standarisasi ini tidak hanya membantu dalam mengakui kontribusi lembaga penyedia data, tetapi juga berperan penting dalam membangun reputasi sistem data nasional sebagai infrastruktur penelitian berbasis populasi.

Pengembangan sistem pendukung pengguna data, dokumentasi teknis yang lebih rinci, serta integrasi dengan sumber data eksternal seperti survei nasional atau rekam medis elektronik akan sangat membantu memperluas cakupan penelitian yang dapat dilakukan dengan data sampel BPJS Kesehatan. Selain itu, strategi nasional untuk membangun ekosistem penelitian data besar di bidang kesehatan, termasuk pelatihan peneliti

dan insentif publikasi berbasis data JKN, akan menjadi krusial dalam mendorong transformasi data menjadi kebijakan.

Sebagai bagian dari refleksi atas proses penulisan artikel ini, terdapat beberapa keterbatasan penting yang perlu disampaikan. Penelusuran literatur dilakukan dengan kata kunci terbatas dan hanya menggunakan dua basis data ilmiah. Pemeriksaan terhadap isi artikel dilakukan secara selektif, terutama pada bagian abstrak dan metode, tanpa evaluasi sistematis terhadap keseluruhan naskah. Hasil pemetaan pemanfaatan data sampel BPJS Kesehatan yang disajikan bersifat indikatif, dan belum mencerminkan kajian bibliometrik yang komprehensif.

SIMPULAN

Data sampel BPJS Kesehatan merupakan aset strategis dalam mendukung pengembangan riset kesehatan berbasis populasi dan penyusunan kebijakan publik berbasis bukti di Indonesia. Kajian ini menunjukkan bahwa meskipun data ini telah tersedia sejak 2019 dan dilengkapi dengan pedoman teknis yang memadai, pemanfaatannya dalam publikasi ilmiah dan proses kebijakan masih terbatas serta belum terdokumentasi secara sistematis. Tantangan yang diidentifikasi meliputi keterbatasan kapasitas analitik, belum adanya standardisasi pelaporan pemanfaatan data dalam publikasi ilmiah, serta minimnya forum ilmiah yang secara khusus mendorong diseminasi hasil penelitian berbasis data ini. Selain itu, masih terdapat isu teknis terkait representasi frekuensi kunjungan, kelengkapan variabel penting, dan ketepatan pembobotan terhadap populasi terkini. Kajian ini juga menyoroti peluang pemanfaatan data sampel untuk mengeksplorasi isu cakupan ganda kepesertaan asuransi (*double coverage*) dan skema *coordination of benefit* (CoB) antara BPJS Kesehatan dengan asuransi kesehatan swasta, yang menjadi isu strategis dalam tata kelola sistem pembiayaan kesehatan ke depan. Potensi data ini dapat dioptimalkan melalui langkah konkret seperti peningkatan sosialisasi dan pelatihan teknis di kalangan akademisi, peneliti, dan pembuat kebijakan. BPJS Kesehatan bersama Kementerian Kesehatan dapat memfasilitasi terbentuknya forum nasional atau konferensi

tahunan untuk menampung pertukaran gagasan dan hasil riset berbasis data sampel. Perguruan tinggi juga memiliki peran penting dalam mengintegrasikan analisis data administratif ke dalam kurikulum pendidikan kesehatan masyarakat dan kebijakan kesehatan, serta mendorong penggunaannya dalam tugas akhir dan disertasi mahasiswa. Di samping itu, pengembangan sistem identifikasi publikasi yang menggunakan data ini, serta perluasan integrasi data lintas lembaga dan sektor, sangat diperlukan. Dengan langkah-langkah ini, data sampel BPJS Kesehatan dapat berkembang menjadi fondasi utama bagi transformasi sistem kesehatan nasional yang berbasis data, kolaboratif, dan responsif terhadap tantangan pembiayaan jangka panjang.

PUSTAKA ACUAN

- Ahn, E. (2020). Introducing Big Data Analysis Using Data from National Health Insurance Service. *Korean J Anesthesiol*, 73(3), 205–211.
- Antonelli, M., Firenzuoli, F., Salvarani, C., Gensini, G. F., & Donelli, D. (2020). Reading and Interpreting Reviews for health Professionals: A Practical Review. *Internal and Emergency Medicine*, 15(6), 945-955.
- BPJS Kesehatan RI. (2023). *Data Sampel BPJS Kesehatan Tahun 2015-2022*. BPJS Kesehatan.
- Atkinson, J., & Blakely, T. (2017). New Zealand's Integrated Data Infrastructure (IDI): Value to Date and Future Opportunities. *International Journal of Population Data Science*, 1(1), 124.
- Batko, K., & Slezak, A. (2022). The Use of Big Data Analytics in Healthcare. *Journal of Big Data*, 9(1), 3.
- BPJS Kesehatan. (2020). *Peluncuran Data Sampel BPJS Kesehatan 2015-2018*. BPJS Kesehatan. <https://data.bpjs-kesehatan.go.id/bpjs-portal/action/blog-detail.cbi?id=da1c3acd-42bc-11eb-a0ed-3134c80c87ab>. (Accessed: 23 April 2025).

- BPJS Kesehatan. (2025). *Peluncuran Data Sampel BPJS Kesehatan 2015-2023*. BPJS Kesehatan. <https://data.bpjs-kesehatan.go.id/bpjs-portal/action/blog-detail.cbi?id=779e0f02-bd13-11efa976-9b1ff87869bd> (Accessed: 23 April 2025).
- Darmawan, E. S., Permanasari, V. Y., Nisrina, L. V., Kusuma, D., Hasibuan, S. R., & Widayanti, N. (2024). Behind the Hospital Ward: In-Hospital Mortality of Type 2 Diabetes Mellitus Patients in Indonesia (Analysis of National Health Insurance Claim Sample Data). *International Journal of Environmental Research and Public Health*, 21(5), 1-19.
- Dewan Jaminan Sosial Nasional. (2024). *Sistem Monitoring dan Evaluasi Jaminan Sosial, Sistem Monitoring dan Evaluasi Jaminan Sosial*. <https://kesehatan.djsn.go.id/kesehatan/> (Accessed: 23 November 2024).
- Fatimah, A., Anselmi, L., Gibson, J., Maharani, A., Irmansyah, I., Idaiani, S., Susilo, D., Wilkinson, J., Sutton, M., Susanti, H., Brooks, H., Bee, P., & Thabrany, H. (2025). Indonesian National Health Insurance Scheme Longitudinal Sample Data 2015–2020: Overview and Potential Uses for Health Policy Analysis. *BMC Health Services Research*, 25(1), 1-12.
- Hong, J. S., & Kang, H. C. (2015). Sex Difference in the Treatment and Outcome of Korean Patients with Acute Myocardial Infarction Using the Korean National Health Insurance Claims Database. *Medicine*, 94(35), 1-8.
- Hsieh, C. Y., Su, C. C., Shao, S. C., Sung, S. F., Lin, S. J., Kao Yang, Y. H., & Lai, E. C. C. (2019). Taiwan's National Health Insurance Research Database: Past And Future. *Clinical Epidemiology*, 349-358.
- Kim, D. W. (2025). Statistical Methods for Baseline Adjustment and Cohort Analysis in Korean National Health Insurance Claims Data: A Review of PSM, IPTW, and Survival Analysis with Future Directions. *Journal of Korean Medical Science*, 40(8), 1-10.
- Kim, H. K., Song, S. O., Noh, J., Jeong, I. K., & Lee, B. W. (2020). Data Configuration and Publication Trends for the Korean National Health Insurance and Health Insurance Review & Assessment Database. *Diabetes & Metabolism Journal*, 44(5), 671-678.
- Kim, H., Cheng, S. H., Yamana, H., Lee, S., Yoon, N. H., Lin, Y. C., Fushim, K., & Yasunaga, H. (2021). Variations in Hip Fracture Inpatient Care in Japan, Korea, and Taiwan: An Analysis of Health Administrative Data. *BMC Health Services Research*, 21(1), 1-11.
- Kim, M. K., Han, K., & Lee, S. H. (2022). Current Trends of Big Data Research Using the Korean National Health Information Database. *Diabetes & Metabolism Journal*, 46(4), 552-563.
- Kindermann, B., Hinde, S., & Abbondante, M. (2017). Multi-Agency Data Integration Project-Linking Data-Improving Lives. *International Journal of Population Data Science*, 1(1), 196.
- Kurniati, A. P., Wisudiawan, G. A. A., & Kusuma, G. P. (2022). Potentials of Clinical Pathway Analysis Using Process Mining on the Indonesia National Health Insurance Data Samples: An Exploratory Data Analysis. In *2022 International Conference on Data Science and Its Applications (ICoDSA)*, 294-299.
- Lee, J., Lee, J. S., Park, S. H., Shin, S. A., & Kim, K. (2017). Cohort Profile: The National Health Insurance Service–National Sample Cohort (NHIS-NSC), South Korea. *International Journal of Epidemiology*, 46(2), e15-e15.

- Lee, S., Yeo, J., Lee, S. H., Lee, Y. J., Lee, S., & Ha, I. H. (2023). Trends in Healthcare Utilisation of Patients with Migraine in South Korea: a Retrospective Observational Study Using Health Insurance Review and Assessment Service National Patient Sample data from 2010 to 2018. *BMJ open*, 13(3), 1-12.
- Lim, S. J., & Jang, S. I. (2025). Leveraging National Health Insurance Service Data for Public Health Research in Korea: Structure, Applications, and future Directions. *Journal of Korean Medical Science*, 40(8), 1-8.
- Lohr, S. L. (2021). *Sampling: Design and Analysis*. Chapman and Hall/CRC.
- Shaffril, H. A. M., Samsuddin, S. F., & Samah, A. A. (2021). The ABC of Systematic Literature Review: the Basic Methodological Guidance for Beginners *Quality & quantity*, 55(4), 1319-1346.
- Shih, Y. R., Liao, K. H., Chen, Y. H., Lin, F. J., & Hsiao, F. Y. (2020). Reimbursement Lag of New Drugs Under Taiwan's National Health Insurance System Compared with United Kingdom, Canada, Australia, Japan, and South Korea. *Clinical and Translational Science*, 13(5), 916-922.
- Song, S. O., Jung, C. H., Song, Y. D., Park, C. Y., Kwon, H. S., Cha, B. S., Park, J. Y., Lee, K. U., Ko, K. S., & Lee, B. W. (2014). Background and Data Configuration Process of a Nationwide Population-Based Study Using the Korean National Health Insurance System. *Diabetes & Metabolism Journal*, 38(5), 395-403.
- Sugihara, T., Fujimura, T., Yasunaga, H., Matsui, H., Kamei, J., & Kume, H. (2019). Regional Clinical Practice Variation in Urology: Usage Example of the Open Data of the National Database of Health Insurance Claims and Specific Health Checkups of Japan. *International Journal of Urology*, 26(2).
- Suto, M., Iba, A., Sugiyama, T., Kodama, T., Takegami, M., Taguchi, R., ... & Iso, H. (2024). Literature Review of studies Using the National Database of the Health Insurance Claims of Japan (NDB): limitations and Strategies in Using the NDB for Research. *JMA Journal*, 7(1), 10-20.

Lampiran 1. Topik-Topik Potensial untuk Penelitian Berbasis Data Sampel BPJS Kesehatan

Bidang Fokus	Topik Penelitian	Potensi Analisis
Analisis beban penyakit dan layanan	Pola penyakit populasi, analisis harian (<i>day-level</i>), proyeksi kebutuhan layanan, proyeksi kebutuhan obat	Diagnosis dominan, tren musiman, estimasi volume kunjungan, kebutuhan farmasi berbasis diagnosis
Prediksi risiko dan segmentasi populasi	Segmentasi peserta JKN, prediksi risiko biaya tinggi, integrasi dengan <i>machine learning</i> / <i>Artificial Intelligence</i>	Model prediktif berbasis klaim, stratifikasi risiko, sistem peringatan dini (<i>early warning</i>)
Evaluasi efektivitas dan efisiensi	Efektivitas biaya per episode penyakit, preventable hospitalization, analisis readmisi	<i>Episode-based costing</i> , identifikasi layanan yang tidak efisien, evaluasi readmisi yang dapat dicegah
Analisis longitudinal dan survival	Analisis survival pasien kronis, evaluasi luaran jangka panjang	Kelangsungan hidup berdasarkan klaim, transisi antar diagnosis, outcome pasien kronis
Perencanaan SDM dan beban kerja	Estimasi beban kerja tenaga kesehatan, proyeksi kebutuhan SDM di FKTP/FKRTL	Analisis volume layanan per fasilitas, prediksi kebutuhan tenaga medis berdasarkan klaim
Evaluasi akses dan kesetaraan	Kesetaraan akses layanan kesehatan, variasi antar wilayah, integrasi data spasial	Analisis spasial dengan data BPS, disparitas pemanfaatan layanan, indeks keadilan geografis
Evaluasi kebijakan dan reformasi	Dampak kebijakan berbasis klaim, <i>pre-post policy evaluation</i>	<i>Quasi-experimental design</i> , <i>natural experiment</i> berbasis kebijakan, analisis perubahan pola layanan
Mutu layanan dan validasi sistem	Validasi rujukan, evaluasi indikator mutu berbasis data sekunder	Validitas pola rujukan, estimasi indikator mutu seperti LOS, frekuensi kunjungan, kepatuhan protokol
Dinamika kepesertaan dan <i>switching</i>	Perubahan status peserta, transisi segmen (PBI, PBPU, PPU), keberlanjutan kepesertaan	Analisis switching segmen, dampaknya terhadap utilisasi, efek jangka panjang terhadap klaim dan layanan

Sumber: Data Primer Diolah