

Database Riset Bersumber Data Sekunder BPJS Kesehatan

Guardian Yoki Sanjaya¹, Anis Fuad¹, Lutfan Lazuardi¹, Furqonudin Ramdhani¹,
Wan Aisyiah Baros², Erzan Dhanalvin²

¹Program Studi Ilmu Kesehatan Masyarakat, Fakultas Kedokteran, Universitas Gadjah Mada.

²Grup Penelitian dan Pengembangan BPJS Kesehatan

¹gysanjaya, ²anisfuad@ugm.ac.id, ³lutfan.lazuardi@ugm.ac.id,

ABSTRAK

Latar Belakang: BPJS Kesehatan dalam pencapaian Universal Health Coverage pada tahun 2019 akan memiliki salah satu basis data pelayanan kesehatan terbesar di dunia. Seiring dengan meningkatnya volume data yang tersimpan, semakin cepatnya proses mengumpulkan dan menghasilkan data, keragaman datanya serta kebutuhan terhadap kualitas data yang mencerminkan fakta, maka menjadi penting bagi BPJS Kesehatan untuk mengkaji kondisi manajemen data saat ini serta penyiapan database riset sebagai salah satu output.

Metode Penelitian: Pendekatan kualitatif dilakukan untuk mengembangkan database riset BPJS Kesehatan. Literatur review, diskusi kelompok terarah dan identifikasi data di BPJS Kesehatan dilakukan untuk menggali konsep database riset untuk jaminan kesehatan nasional.

Hasil: Database riset BPJS Kesehatan termasuk dalam konsep big data analytics karena volume, jumlah dan frekuensi yang tinggi serta tipe datanya yang beragam. Terdapat 5 dataset yang disepakati untuk dijadikan database riset BPJS Kesehatan. Jumlah peserta BPJS yang terus bertambah menuntut perlunya tatakelola data yang baik untuk memastikan representatif database riset terhadap pertumbuhan tersebut. Beberapa isu lain seperti keamanan data, teknik query dan infrastruktur, serta monitoring penggunaan database riset untuk penelitian perlu dipersiapkan dalam pengembangan database riset.

Kesimpulan: Terdapat 3 aspek penting yang perlu dipertimbangkan dalam pengembangan database riset BPJS Kesehatan, yaitu konsep big data analytics, representasi database riset yang diambil dari transaksi data di BPJS Kesehatan dan manajemen data yang baik.

Kata Kunci: BPJS Kesehatan, Database Riset, Dataset kohort, big data analytics

ABSTRACT

Background: National Social Health Insurance or BPJS Kesehatan in achieving Universal Health Coverage in 2019 will have one of the largest health database in the world. Along with the increasing volume of data stored, accelerated the process of collecting and producing data, data diversity as well as the need for data quality, it becomes necessary for BPJS Kesehatan to assess the current state of data management and development of research databases to be used broadly.

Methods: A qualitative approach was taken to develop a research database of BPJS Kesehatan. The literature review, focus group discussions and the identification of data from existing information systems in BPJS were undertaken to explore the concept of a research database sourced from the national health insurance.

Results: Research database of BPJS Kesehatan included in the concept of big data analytics because the high volume of data sample, large number and frequency of data transactions and the diverse data types. There are 5 dataset agreed to serve as a research database of BPJS Kesehatan. The growing number of participants, required BPJS Kesehatan to have good data governance to ensure a representative of growth data for research database. Several other issues such as data security, query techniques and infrastructure, as well as monitoring the use of research database for research need to be prepared in the development of research database.

Conclusion: There are three important aspects to consider in the development of a research database, namely the concept of big data analytics, the representative of sample database research and good data management.

Keywords: BPJS Health, Research Database, Dataset cohort, big data analytics

PENDAHULUAN

Sudah lebih dari satu tahun, BPJS Kesehatan memulai kiprahnya untuk menjalankan amanat menuju tercapai Universal Health Coverage pada tahun 2019. Sampai saat ini, sudah lebih dari 140 juta peserta telah terdaftar di BPJS Kesehatan. Jumlah transaksi pelayanan pada tingkat primer dan sekunder juga terus bertambah. Terkait dengan hal tersebut, basis data administratif pelayanan kesehatan (*administrative health care database*) yang dimiliki oleh BPJS akan menjadi salah satu yang terbesar di dunia. Pengalaman dari berbagai lembaga pembayar (*payor*) di dunia telah menunjukkan bahwa, data administratif yang dimiliki dapat bermanfaat tidak hanya untuk kalangan internal tetapi juga kepada komunitas pengetahuan sehingga berkontribusi langsung kepada pengembangan pengetahuan, sebagaimana yang dihasilkan oleh database *Taiwan National Health Insurance Research Database (NHIRD)* di Taiwan¹ atau *Clinical Practitioner Research Database (CPRD)* di Inggris². Hal yang sama juga terjadi di negara maju seperti Perancis, Amerika Serikat dan Jerman.

Sama halnya dengan NHIRD dan CPRD, Grup Litbang BPJS Kesehatan secara bertahap telah melakukan berbagai langkah fundamental dalam manajemen data sekunder. Dimulai dengan menyusun roadmap manajemen data pada tahun 2013, grup Litbang kemudian melanjutkan dengan menyusun prototipe aplikasi *big data* untuk mengolah data longitudinal yang dapat dimanfaatkan untuk riset. Bekerjasama dengan Pusat Kebijakan dan Manajemen Asuransi Kesehatan UGM, Departemen Litbang BPJS Kesehatan sedang mengembangkan Basis Data Riset (Badari) bersumber data kepesertaan, utilisasi pelayanan primer dan lanjutan seperti halnya NHIRD dan CPRD.

NHIRD memuat berbagai macam variasi sample data yang diekstraksi dari total data klaim yang ada untuk kepentingan penelitian. CPRD memiliki

pengalaman melakukan *data linkage* antara data yang bersumber dari layanan primer dengan sumber data lain yang relevan untuk mendapatkan analisis yang lebih komprehensif. Keduanya memiliki keunikan pengalaman yang dapat dijadikan contoh bagi pengembangan database riset BPJS Kesehatan. Selain itu, baik NHIRD maupun CPRD memiliki mekanisme tertentu untuk mengekstraksi data, melakukan data linkage, penyediaan data untuk penelitian, serta menjaga keamanan data yang penting untuk dipelajari.

BPJS Kesehatan, melalui grup Penelitian dan Pengembangan yang bekerjasama dengan Pusat Kebijakan dan Manajemen Asuransi Kesehatan (KP-MAK), Fakultas Kedokteran Universitas Gadjah Mada melakukan studi yang bertujuan untuk mengembangkan database riset yang bersumber dari data BPJS Kesehatan.

METODE PENELITIAN

Pendekatan kualitatif dilakukan untuk mengembangkan database riset yang bersumber dari data yang dimiliki oleh BPJS Kesehatan. Pengumpulan data dilakukan dengan pendekatan literature review, diskusi kelompok terarah dan identifikasi data yang dimiliki BPJS Kesehatan. Literatur review dengan kata kunci "*research database*" dan "*health insurance claim*" atau "*national health insurance*", dilakukan dengan bantuan google scholar dan PubMed. Artikel yang menunjukkan tersedianya database riset yang bersumber dari klaim asuransi kesehatan dikumpulkan untuk melihat pola pengembangan database riset yang bersumber dari data jaminan kesehatan.

Pola-pola database riset tersebut kemudian didiskusikan melalui diskusi kelompok terarah (FGD) yang melibatkan pihak internal maupun eksternal BPJS Kesehatan. Dari internal organisasi FGD melibatkan grup MPKR, MPKP, Litbang, Kepesertaan, Keuangan, OTI dan PPID. Sedangkan dari eksternal FGD melibatkan TNP2K, BPS, Litbang Kemenkes, BKKBN, WHO Indonesia, Registrasi Kanker dan Perwakilan

rumah sakit. FGD difokuskan untuk menentukan konsep database riset BPJS Kesehatan, dataset yang perlu dipersiapkan serta tatakelola data di BPJS Kesehatan. Sumber data baik primer maupun sekunder, diidentifikasi melalui struktur database sistem informasi yang digunakan di BPJS Kesehatan.

HASIL DAN PEMBAHASAN

Dari hasil analisa kualitatif yang dilakukan, terdapat 3 isu penting yang didapatkan yaitu 1). Berbagai database riset bersumber klaim asuransi kesehatan yang telah dikembangkan di beberapa Negara, 2). Konsep database riset untuk BPJS Kesehatan dan 3). Tatakelola database riset untuk pengembangan dan penggunaan data. Keberlangsungan database riset perlu mempertimbangkan beberapa tantangan yang teridentifikasi dari penelitian ini.

1. Database riset dan konsep *big data analytics* jaminan kesehatan nasional

Di dunia, tidak banyak tersedia sistem database riset jaminan kesehatan yang telah berkontribusi banyak bagi pengetahuan. Beberapa yang dikelola secara nasional adalah *Taiwan National Health Insurance Research Database* yang dikembangkan oleh *Taiwan Bureau of National Health Insurance*¹, *GPRD (General Practitioner Research Database)* di Inggris², *Korean national health insurance claims database* di Korea Selatan³ dan *Japan Medical Data Center* yang menyediakan data klaim layanan rawat jalan, rawat inap dan obat⁴. Selebihnya dikelola oleh perusahaan asuransi atau dikelola oleh masing-masing Negara bagian seperti di Amerika, walaupun sudah memulai untuk mengintegrasikan data klaim ke dalam *All-Payer Claims Databases* atau *APCDs*⁵. Berbagai database tersebut telah terbukti memberikan nilai tambah, tidak hanya bagi lembaga yang memiliki data, tetapi juga menghasilkan ratusan pengetahuan yang terpublikasikan dan dimanfaatkan publik. Seiring dengan data yang cukup besar jumlahnya, metode

penggalian data seperti data mining dapat digunakan untuk memperkaya analisis statistik yang sudah rutin dipakai⁶.

Karena banyaknya data sekunder yang diolah, maka konsep *big data analytics* (BDA) dapat diterapkan pada database riset BPJS Kesehatan. BDA memiliki 2 makna, yaitu 1). Data yang besar dengan karakteristik *volume*, *velocity* dan *variety*. *Volume* mengindikasikan data yang jumlahnya besar, *velocity* dikaitkan dengan frekuensi dan kecepatan pertambahan data dan *variety* mencerminkan banyaknya tipe dan format data yang tersedia. 2). *Analytics* berarti proses mengolah data tersebut dengan berbagai pendekatan, mulai dari yang paling sederhana dengan mendeskripsikan dan memvisualisasikan data, sampai pendekatan data mining, *natural language processing*, *artificial intelligence* dan *predictive analytics*⁷.

Kombinasi dari sumber data kesehatan dan penerapan BDA dapat memberikan manfaat bagi pengembangan sistem kesehatan di Indonesia seperti pengembangan sistem pendukung keputusan klinis, pelayanan kesehatan masyarakat, mendukung operasional pelayanan kesehatan, pengembangan kebijakan, keuangan dan administratif serta membantu penelitian klinis dan non-klinis yang terkait kesehatan.

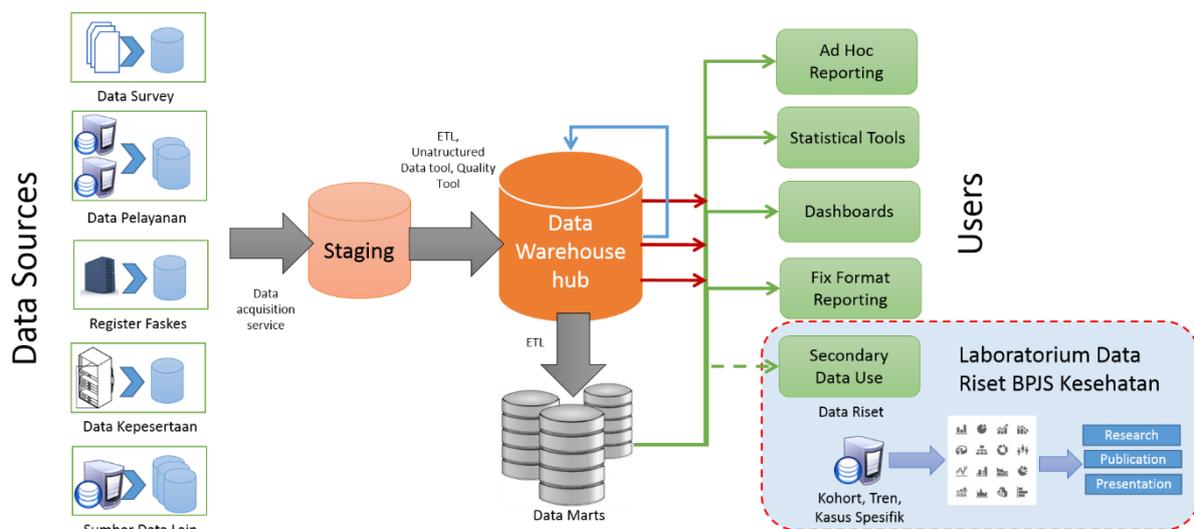
2. Pengembangan database riset BPJS Kesehatan

Berbagai macam database riset bersumber dari klaim asuransi kesehatan di dunia dikembangkan dengan mempertimbangkan: 1). Pendekatan dataset, 2). Sampel yang representatif dan 3). Data linkage. Dataset mencerminkan isu spesifik yang diprioritaskan seperti dataset untuk kohort individual, dataset penyakit diabetes, dataset penyakit psikiatri, dataset untuk penyakit gigi dan mulut dan sebagainya. Sampel data merupakan bagian dari total data yang representatif terhadap dataset yang telah ditetapkan tersebut. Setiap dataset memungkinkan untuk memiliki jumlah sampel yang berbeda, tergantung dari kriteria inklusi dan

jumlah total data yang dimiliki. Sedangkan data linkage merupakan penggabungan database riset dengan sumber data lain, yang berhubungan dengan entitas individu atau wilayah. Berbagai sumber data lain seperti data layanan rumah sakit, data survey kesehatan, data statistik vital dan registrasi penyakit (kanker), memungkinkan untuk digabungkan dengan data klaim asuransi kesehatan. Data linkage meningkatkan efektivitas dan efisiensi penelitian kesehatan, memungkinkan penggunaan kembali dataset yang ada dan menambah nilai penelitian melalui peningkatan kualitas data dan analisis, serta meningkatkan kerjasama dan *outcome* penelitian kesehatan⁸.

Database riset BPJS Kesehatan dikembangkan mengikuti ketiga pola diatas. Sumber data berasal dari transaksi data yang terjadi di beberapa sistem informasi BPJS Kesehatan. Sistem informasi tersebut antara lain

sistem informasi kepesertaan, registrasi fasilitas kesehatan, data pengumpulan premi, data pelayanan primer dan rumah sakit, data pelayanan prolanis dan data survey. Setiap sistem informasi sudah memiliki dokumentasi teknis yang jelas, terutama struktur dan kamus data. Walaupun data tersebut tersimpan di dalam database yang terpisah, namun demikian BPJS Kesehatan sudah memiliki datawarehouse yang digunakan untuk kepentingan internal. Datawarehouse digunakan untuk membuat laporan rutin, melakukan statistik sederhana dan membuat dashboard organisasi. Database riset dikembangkan dengan mengambil sebagian data tersebut (sampel yang representatif) sesuai dengan dataset yang diinginkan dan disimpan dalam satu server database yang terpisah, sehingga tidak mengganggu transaksi data rutin. Gambar 1 menunjukkan bagaimana database riset dikembangkan.



Gambar 1. Skema pengembangan database riset BPJS Kesehatan

Database riset merupakan suatu sistem yang terpisah dari keseluruhan data BPJS Kesehatan. Database riset terdiri dari beberapa dataset dengan sampel yang representatif. Dataset yang saat ini ditetapkan adalah:

1. Dataset kohort individual (sampling 5% dari data kepesertaan yang representative dan pelayanan primer dan lanjutan yang telah diterima peserta).

2. Dataset kohort keluarga (sampling 5% dari data keluarga yang terdiri dari data individu kepala keluarga dan anggota keluarga inti)
3. Dataset khusus seperti kohort penyakit pada anak, penyakit yang jarang ditemukan (rare disease), penyakit kronis dan penyakit katastrofik. Dataset ini mengambil semua data dengan kriteria inklusi sesuai dengan kekhususannya.
4. Dataset trend kunjungan fasilitas kesehatan (sampling 20% fasilitas kesehatan yang representative berdasarkan wilayah, level, kepemilikan dan kelas fasilitas kesehatan)
5. Data linkage yang menggabungkan database riset BPJS dengan sumber data lain yang berhubungan dengan wilayah, seperti data Riskesdas.

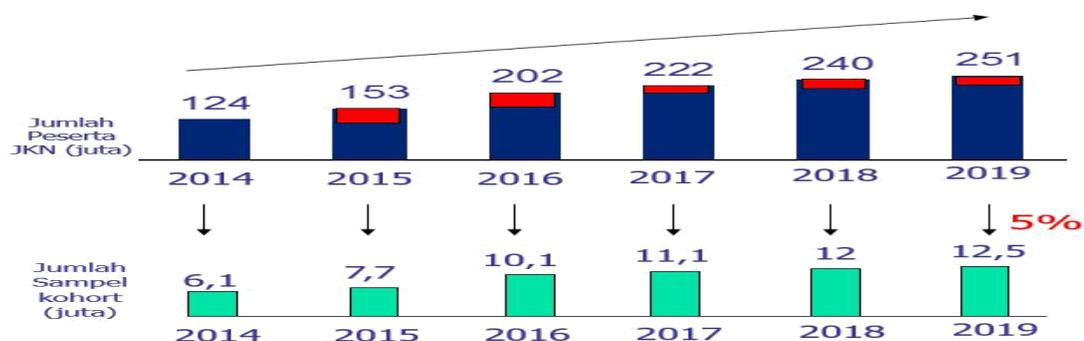
Untuk tahap awal, BPJS Kesehatan memprioritaskan dataset kohort individual dengan menggabungkan 5% data kepesertaan yang representatif, dengan data pelayanan primer dan layanan lanjutan dari peserta yang dijadikan sampel kohort.

1) Dataset Kohort Individual

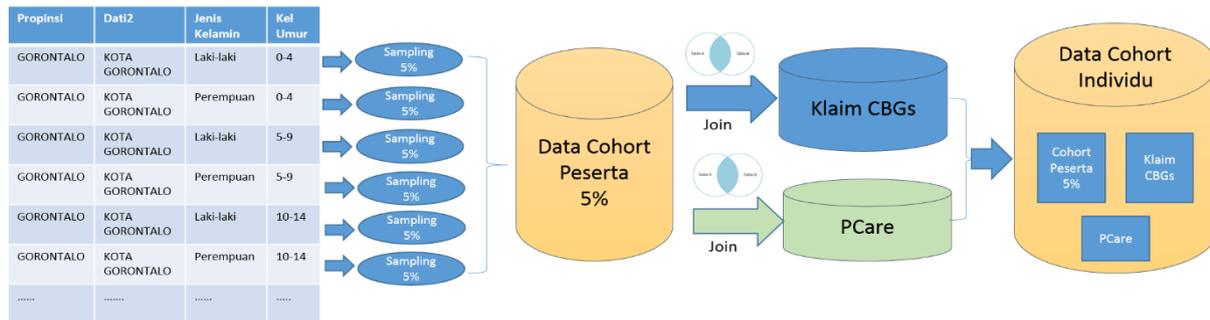
Dataset kohort individual merupakan sampel basis data peserta BPJS Kesehatan yang representatif terhadap keseluruhan peserta BPJS Kesehatan yang akan dipantau secara longitudinal. Data set akan dibuat

setiap tahun mengikuti perkembangan jumlah kepesertaan BPJS Kesehatan. Pada akhir tahun, akan dievaluasi sejauh mana proporsi sampel masih tetap representatif terhadap keseluruhan data peserta. Data riset BPJS Kesehatan diperoleh dari sampling 5% seluruh peserta yang kemudian dide-identifikasi sehingga tidak mencederai privasi dan kerahasiaan pasien. Sampel tersebut diambil secara proporsional berdasar lokasi Kabupaten/Kota peserta dan secara proporsional berdasarkan jenis kelamin dan kelompok umur. Kategori kelompok umur berdasarkan kategori Badan Pusat Statistik (BPS). Data riset tersebut selanjutnya dapat digunakan untuk menilai profil penggunaan layanan kesehatan, epidemiologi sampai dengan gambaran pembiayaan layanan kesehatan baik di tingkat layanan primer maupun sekunder.

Dataset pertama untuk tahun 2014 akan memuat sekitar 6,2 juta sampel peserta. Pada tahun-tahun berikutnya, seiring dengan jumlah peserta BPJS Kesehatan yang akan terus meningkat, maka penyesuaian sampel harus terus dilakukan sehingga tetap representatif. Penyesuaian ini termasuk dinamika pertumbuhan penduduk baik kelahiran dan kematian. Diperkirakan mulai tahun 2019, akan tersedia sekitar 12,5 juta individu terpilih yang akan dipantau secara terus menerus sebagai sampel data kohort individu. Upaya updating ini diharapkan akan berjalan secara rutin setidaknya setahun sekali.



Gambar 2. Pertumbuhan kepesertaan BPJS Kesehatan dan jumlah individu yang tercakup dalam database riset BPJS Kesehatan



Gambar 3. Alur pengambilan dataset kohort individual

2) Database Kohort Individual

Pengambilan dataset kohort individual dilakukan dengan 3 tahapan. Pertama melakukan klasifikasi grup peserta berdasarkan propinsi, kabupaten, jenis kelamin dan kelompok umur. Secara keseluruhan terdapat 15.210 grup dari 34 provinsi dan 1 kelompok peserta luar negeri. Tiap-tiap grup diambil sampel 5% yang dilakukan secara random query. Berdasarkan hasil query ke dalam database seluruh peserta BPJS Kesehatan tahun 2014, diperoleh 6,1 juta individu. Sampel individu tersebut selanjutnya digabungkan dengan data kunjungan ke FKTP yang diambil dari database PCare dan kunjungan FKTL yang diambil dari klaim INA-CBGs sehingga menjadi satu kesatuan data riset BPJS Kesehatan seperti tergambar pada skema berikut

Data sampel 5% kepesertaan yang digabungkan dengan data pelayanan (PCare dan Klaim CBGs) tersebut, selanjutnya disusun menjadi dataset kohort individu. Teknik sampling yang dilakukan dibuatkan prosedur menggunakan aplikasi berbasis Java sehingga dapat dilakukan *query* secara otomatis. Proses pengembangan dataset individu ini selanjutnya dapat direplikasi untuk pengembangan dataset keluarga, dataset fasilitas kesehatan, maupun dataset penyakit khusus dengan prinsip *query* yang disesuaikan. Pengembangan database riset tersebut selanjutnya dapat dijadikan salah satu fungsi utama departemen Litbang BPJS Kesehatan untuk memperkuat penggunaan data

BPJS Kesehatan ke dalam berbagai aktivitas riset, publikasi, presentasi maupun kegiatan ilmiah lainnya.

3. Tatakelola Database Riset BPJS Kesehatan

Kegiatan pengembangan database riset BPJS Kesehatan sangat terkait dengan tatakelola data yang ada di lembaga ini. Tatakelola data yang jelas akan membantu dalam menentukan 1). Data yang disepakati secara internal sebagai database riset, yang merupakan representatif dari data BPJS Kesehatan, 2). Bagaimana pengelolaan database riset secara rutin baik dari segi pengambilan sampel, penambahan data, penyediaan data untuk publik dan pemantauan penggunaannya.

Di BPJS Kesehatan terdapat beberapa sistem informasi yang menghasilkan data transaksi rutin. Data kepesertaan, data fasilitas kesehatan, data iuran premi, baik individu maupun perusahaan atau pemerintah daerah, data layanan primer yang berasal dari PCare, data klaim rumah sakit, data klaim obat, surat eligibilitas peserta (SEP), data verifikator dan lain sebagainya, merupakan sumber data yang perlu kesepakatan terhadap pemanfaatannya sebagai bagian dari database riset BPJS Kesehatan. Sebagian besar, sumber data untuk database riset di BPJS Kesehatan berasal dari berbagai aktivitas yang dilakukan oleh penyedia layanan kesehatan yang bekerjasama dengan BPJS Kesehatan. Kualitas dan akuntabilitas data berkaitan juga dengan sumber data primer yang berasal

dari penyedia layanan kesehatan tersebut. Penyediaan database riset dilakukan seefisien mungkin dengan mengelola resiko yang mungkin terjadi dan sedapat mungkin memenuhi aspek legal berdasarkan regulasi serta peraturan yang ada.⁶

Selain sumber data, tatakelola data berkaitan secara fundamental dengan misi dari pengembangan database riset BPJS Kesehatan. Misi ini yang kemudian dijabarkan oleh BPJS Kesehatan dalam fokus kegiatan tertentu dalam bentuk kegiatan mempersiapkan database riset, *query* data sampel, de-identifikasi data, penyediaan infrastruktur, penyediaan metadata database riset dan prosedur penggunaan database riset. Penggunaan database riset kemudian akan diperjelas dengan aturan spesifik tentang data yang meliputi siapa yang bisa mengakses, bagaimana mengakses, menggunakan, akuntabilitas penggunaan database riset serta proses pengendaliannya. Transparansi dan keterlibatan berbagai stakeholder internal dan eksternal organisasi dalam mempersiapkan database riset BPJS Kesehatan sangat penting. Berbagai tersebut juga semestinya diatur.

Aspek tata kelola ini akan sangat menentukan sejauh mana data riset yang tersedia nantinya dapat dimanfaatkan secara optimal oleh komunitas ilmiah baik lembaga penelitian, universitas maupun lembaga lain yang relevan dan berkepentingan dalam melakukan analisis data BPJS Kesehatan.

4. Tantangan pengembangan database riset BPJS Kesehatan

Data yang tersimpan dalam database riset perlu memperhatikan mengantisipasi aspek keamanan, privasi dan kerahasiaan data. De-identifikasi data dapat dilakukan dengan menghilangkan variabel nama, alamat dan wilayah tempat tinggal, nomor identitas, nomor telpon atau fax, alamat email dan tanggal dan bulan lahir (kecuali tahun lahir)⁹. Namun demikian, nomor identitas atau wilayah akan diperlukan jika akan dilakukan data linkage dengan sumber data lain seperti

survey dan registrasi penyakit. Selain itu, untuk melakukan data linkage perlu melakukan kerjasama lintas organisasi dimana isu kepemilikan data menjadi sangat penting. Regulasi diperlukan untuk mengatur penggunaan data untuk penelitian, data linkage dan kepemilikan data. Selain tersedianya database riset untuk kepentingan penelitian, permintaan data dan output penelitian dengan bersumber database riset perlu dikelola untuk menghindari duplikasi penelitian, penyalahgunaan data dan hasil penelitian untuk pengembangan keilmuan.

Besarnya data yang diperoleh dari query yang dilakukan memerlukan infrastruktur server dan data storage yang memadai. Proses query data dari sumber database, memerlukan waktu yang lama dan membutuhkan kemampuan komputasi yang besar, terutama untuk mengakomodasi 5 dataset yang telah ditentukan. Selain itu, informasi dataset, struktur data dan relasi tabel di dalam database riset perlu didokumentasikan dengan definisi operasional yang jelas. Informasi yang jelas akan mengurangi kesalahan interpretasi data dan dapat memberikan gambaran kepada calon pengguna terhadap pemanfaatan database riset BPJS Kesehatan.

Pengelolaan penelitian yang bersumber dari database riset menjadi beban kegiatan dalam menjamin keberlangsungan pengembangan database riset. Selain infrastruktur, diperlukan sumber daya manusia untuk melakukan pendokumentasian metadata variabel database riset, melakukan review dan persetujuan terhadap permintaan data yang bersumber database riset untuk penelitian, melakukan monitoring output penelitian dengan indikator-indikator yang jelas.

KESIMPULAN DAN SARAN

Konsep big data analytics sangat relevan dengan tersedianya database riset BPJS Kesehatan yang syarat akan volume data yang besar, frekuensi dan kecepatan pengumpulannya yang terus meningkat serta variasi data dan format yang tersedia. Terlebih lagi

ditetapkannya 5 dataset yang dikumpulkan secara kohort. Pengelolaan database riset BPJS Kesehatan menjadi penting karena potensinya dalam mendukung pelayanan, manajemen, pengembangan kebijakan dan penelitian terkait kesehatan.

Meskipun sudah ada langkah maju dalam penyediaan database riset di BPJS Kesehatan, namun demikian masih banyak yang perlu diperbaiki, diantaranya adalah pengelolaan data yang terpadu antara data di tingkat primer dan sekunder, menggabungkan data kepuasan peserta dengan data pelayanan, masih terbatasnya integrasi data dengan data program kesehatan yang dimiliki oleh sektor kesehatan, tatakelola penggunaan database riset dan output publikasi bersumber database riset.

KEPUSTAKAAN

1. Chen Y, Yeh H, Wu J, Haschler I, Chen T, Wetter T. Taiwan's National Health Insurance Research Database: administrative health care database as study object in bibliometrics. *Scientometrics*. 2011;365-380.
2. Chen Y, Wu J, Haschler I, Majeed A, Chen T, Wetter T. Academic impact of a public electronic health database: bibliometric analysis of studies using the general practice research database. *PLoS One*. 2011;6(6).
3. Lee J, Lee JS, Park S-H, Shin SA, Kim K. Cohort Profile: The National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea. *Int J Epidemiol*. January 2016:dyy319. doi:10.1093/ije/dyv319.
4. Japan Medical Data Center. JMDC Claims Data Base. 2016. http://www.jmdc.co.jp/en/srv_pharma/jdm.html. Accessed May 2, 2016.
5. Miller BPB, Love D, Sullivan E, Porter J, Costello A. *All-Payer Claims Databases An Overview for Policymakers*.; 2010.
6. Cottle M, Hoover W. Transforming Health Care Through Big Data. 2013:1-24.
7. Canada Health Infoway. *Big Data Analytics in Health*.; 2013.
8. Bradley CJ, Penberthy L, Devers KJ, Holden DJ. Health services research and data linkages: Issues, methods, and directions for the future. *Health Serv Res*. 2010;45(5 PART 2):1468-1488. doi:10.1111/j.1475-6773.2010.01142.x.
9. El Emam K. Methods for the de-identification of electronic health records for genomic research. *Genome Med*. 2011;3(4):25. doi:10.1186/gm239.