

# Real-world assessment of healthcare provided by the National Health Service: The network of regional Beaver<sup>®</sup> research platforms

Federico Rea <sup>(1,2)</sup>, Pietro Pugni <sup>(2)</sup>, Dario Pescini <sup>(2)</sup>, Luigi Palmieri <sup>(3)</sup>, Simona Giampaoli <sup>(3)</sup>, Flavia Carle <sup>(1,4)</sup>, Giovanni Corrao <sup>(1,2)</sup>

(1) National Centre for Healthcare Research & Pharmacoepidemiology, at the University of Milano-Bicocca, Milan, Italy

(2) Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy

(3) Department of Cardiovascular, Dysmetabolic and Ageing-Associated Diseases, Istituto Superiore di Sanità, Rome, Italy

(4) Center of Epidemiology and Biostatistics, Polytechnic University of Marche, Ancona, Italy

**CORRESPONDING AUTHOR:** Prof. Giovanni Corrao, Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca, Via Bicocca degli Arcimboldi, 8, Edificio U7, 20126 Milano, Italy. Tel.: +39.02.64485854; E-mail: giovanni.corrao@unimib.it

**DOI:** 10.2427/12862

## ABSTRACT

Real-world evidence can provide answers on healthcare utilization and appropriateness, post-marketing drugs safety and comparative effectiveness, and cost-effectiveness profiles of healthcare pathways. Healthcare utilization databases, possibly integrated with drug and disease registries, electronic medical records, survey and cohort data (i.e. real-world data), allow to trace healthcare 'footprints' left from beneficiaries of National Health Service.

Beaver<sup>®</sup> is a research platform available on demand to Italian regions which we developed for computing indicators of healthcare utilization and clinical outcomes, as well as for generating evidence on effectiveness and cost-effectiveness profile. Two distinct solutions may be adopted. One, the so-called Beaver<sup>®</sup> Light front-end allows to automatically compute health indicators of adherence to official guidelines. Two, the so-called Beaver<sup>®</sup> Full front-end involves an eight-step procedure entirely driven by the study protocol.

In order to fulfil the directives recently issued by the European Parliament and Council and the Italian Authority for the protection of individual data, the platform resides in each region's infrastructure, so limiting the free movement of electronic health data. Indeed, regional authorities should be responsible for data safety and for allowing data accessibility. The use of standardized and validated algorithms enables to obtain regional estimates that, being obtained by employing regional platforms containing data extracted with standardized procedure, may be compared and possibly summarized by using common meta-analytic techniques.

In conclusion, the Beaver<sup>®</sup> regional platform is a promising tool which may contribute to stimulate healthcare research in Italy.

*Key words:* Healthcare utilization databases; Real-world evidence; Protocol-driven data

## BACKGROUND

Real-world evidence (i.e., scientific evidence derived from the clinical ‘footprints’ of real patients as they access their routine medical care) can provide answers on (i) healthcare utilization, in terms of appropriateness with respect to evidence-based guidelines [1,2]; (ii) post-marketing drugs’ safety, particularly in special populations (e.g., pregnant women [3,4], children [5,6] and elderly [7,8]); and (iii) comparative effectiveness [9,10] and cost-effectiveness [11,12] profiles of drug therapies and healthcare pathways.

Real-world data (i.e., data source useful to trace clinical ‘footprints’ so constituting the necessary, even not sufficient, base for generating real-world evidence) can be collected from various types of computerized sources, such as healthcare utilization databases, drug and disease registries, electronic medical records, surveys and cohorts data, etc... Moreover, good quality electronic data is very widespread in Italy, being it at the local, regional or national scale.

Although these electronic health data sources have the potential to improve our understanding of care quality and outcomes [13], concerns about privacy issues, data access and data processing need to be carefully addressed for boosting translational research and generating credible evidence in the framework of real-world investigations.

This paper describes the new opportunity arising from the so-called Beaver<sup>®</sup> regional research platform, a web-based system for interconnecting and processing healthcare data. Beaver<sup>®</sup> is able to automatically compute health indicators in high priority areas according to guidelines from the Italian Health Ministry, as well as to generate reliable evidence addressing healthcare policy by means of a protocol-driven data process.

## WHAT REAL-WORLD DATA ARE WE TALKING ABOUT?

Beaver<sup>®</sup> is at least powered by data accounting for healthcare services delivered by the Italian regions, and the associated expenditures, of which, since the early 2000s, the National Government has made it mandatory for regional governments to collect. The resulting regional databases, that are the so-called healthcare utilization (HCU) or administrative databases, collect a variety of information, including: (i) demographic and administrative data on residents who receive National Health System (NHS) assistance (beneficiaries of NHS practically coinciding with the whole resident population); (ii) hospital discharge records providing information on primary diagnosis, co-existing conditions and procedures coded according to the International Classification of Diseases, 9th Revision Clinical Modification (ICD-CM-9) classification system performed to inpatients admitted in public and private hospitals; (iii) emergency room accesses,

providing ICD-CM-9 codes of causes of access to general and specialized emergency/acceptance departments of public and private hospitals; (iv) drugs dispensed by territorial pharmacies and medicaments directly administered in the outpatient setting and day-hospital coded according to the Anatomical Therapeutic Chemical (ATC) classification system; (v) data on outpatient services, including specialist visits, laboratory tests and diagnostic imaging (e.g., X-rays, Computerised Tomography-CT, Magnetic Resonance Imaging-MRI, Positron Emission Tomography-PET) coded according to the National Tariff Nomenclator; (vi) Certificates of Delivery Assistance (i.e., the so called CeDAP) providing information on the mother’s socioeconomic traits, as well as medical information on pregnancy, childbirth, and child presentation at delivery; (vii) other data, including those regarding vaccinations, mental health services, and any other form of assistance included in the list of the Essential Assistance Levels (LEA) [14] provided by the NHS. As a unique identification code is used for all databases within each region, their record linkage allows searching out the diagnostic-therapeutic pathway supplied to beneficiaries of NHS.

The above-described data represent the minimum set available in each region and that alone justifies / supports the implementation of a regional platform. However, other data may be integrated within a given Beaver<sup>®</sup> platform, depending on the availability of relevant data, and the specific regional concerns. For example, it is possible to feed a Beaver<sup>®</sup> regional platform with other institutional sources (e.g., drug and therapeutic plan registries instituted by the Drug Agency -AIFA-, national registry of rare diseases instituted by the Health Institute -ISS-, health surveys and population census instituted by the Central Institute of Statistics -ISTAT-), as well as non-institutional ones (e.g., population-based cancer registries, medical records from primary healthcare and specialist clinics), provided that data is recorded using the same identification code already used for recording NHS administrative data (e.g., the national fiscal code).

## PRIVACY AND OTHER ETHICAL ISSUES

In order to ensure a high level of protection of natural persons, rights and rules, ratified by the European Parliament and Council in the recently issued Directive 95/46/EC [15], represented the milestone for designing and managing Beaver<sup>®</sup>. In particular, because there “... *are circumstances under which it may be reasonable and economical ensuring data protection to be broader than a single project...*”, a common processing platform may be established by public authority (in our case regional government), so realizing an environment where data may be processed “... *in a manner that ensures appropriate security and confidentiality of the personal data, including for preventing unauthorized access ...*”.

Some true and false constraints established by EU Directive deserve to be emphasized. In particular, directive faces that “... the processing of special categories of personal data may be necessary for reasons of public interest in the areas of public health without consent of the data subject ...”, where “... all elements related to ... health status ... and the determinants having an effect on that health status, healthcare needs, resources allocated to healthcare, the provision of, and universal access to, healthcare as well as healthcare expenditure and financing, and the causes of mortality...”, are here intended in the context of public health. In other terms, to all those elements which have been listed in the Background section of the current paper, and that justify the Beaver<sup>®</sup> research platform achievement. On the other hand, the Italian Authority for the protection of individual data established that, although the informed consent to the processing of personal data must always be collected when it is possible to provide adequate information to the subjects included, the impossibility of informing them does not however preclude the processing of the data themselves [16]. This occurs when, for example, the large sample size (as in the case of the population residing in a region), as well as the long-time window between data collecting (such as the therapy and/or outcome occurrence) and processing, make it impossible to obtain the consent. On the other hand, obtaining consent only on the survivors of the treatment, and / or the outcome, make it impossible to generate any valid evidence.

High-quality knowledge that can provide the basis to improve the healthcare for a number of people are expected from real-world data. For this reason, we believe that the true ethical challenge is not of protecting institutional HCU data up to the point of making them inaccessible, but rather make them usable to generate solid evidence useful for better caring future patients. This subject raises some questions like “Who can access these data, and under what conditions?”. Justified constraints, that we hope will be homogeneously adopted by all the regional administrations, should be the following:

- public or no-profit agencies, external to the regional administration (e.g., academy, or other research institutions), able to autonomously process data derived from secondary sources and to generate credible evidence (according to scientific experiences developed in this field), should be included into a list of agencies accredited for submitting research protocols requiring access to the platform;
- a detailed protocol including at least the “genuine” research question (addressing an issue of interest for the regional and/or national health policies), the methodology to be adopted (including methods to take into account the sources of systematic vulnerability, so important in the context under consideration) and legal details

about the ownership of, and the commitment to publish the results, should be submitted for obtaining the permission to process the data stored in the regional platform;

- data processing, once authorized, should be performed without transferring data from the platform that holds them, but with specific queries based on and driven by the protocol.

## BEAVER<sup>®</sup> ARCHITECTURE AND FUNCTIONING

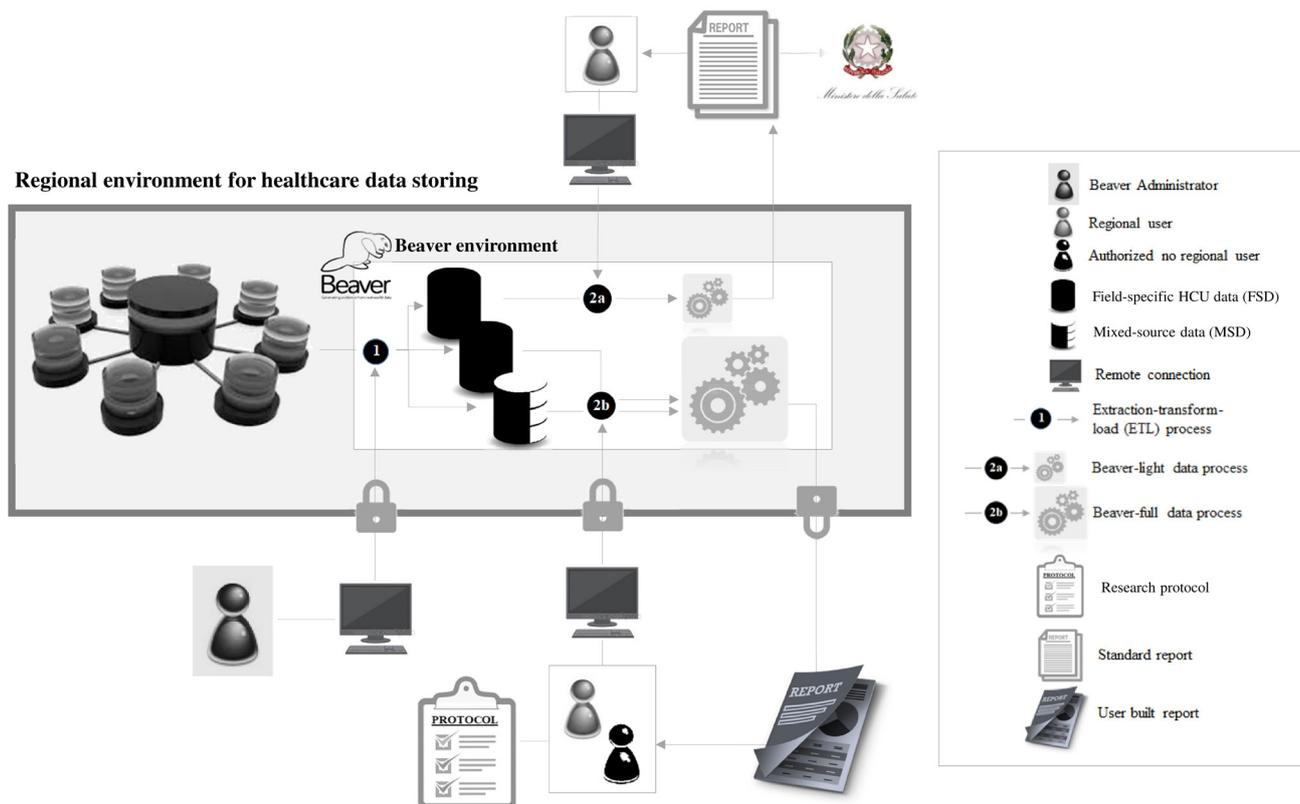
The above-mentioned general principles drove the design of the Beaver<sup>®</sup> platform and led to a two layers architecture: the platform administration layer and the remote user layer (Figure 1). For any technical characteristic, the reader can refer to Table 1.

The administrator is the physical person authorized by the Regional Authority for accessing to the regional environment where healthcare data is stored and protected and is in charge of the first layer. Other than for setting and managing Beaver<sup>®</sup> into a dedicate fully secured system within the regional environment, the administrator has the tasks of (i) extracting data relating to specific fields of interest, (ii) harmonizing them according to definite protocols, and (iii) allocating each of them in a dedicated database within the Beaver<sup>®</sup> environment. Two types of databases may be installed within the platform.

The first type, which we call field-specific database (FSD), concerns a specific diagnostic-therapeutic area and provides only data extraction from each administrative regional database of the information that is common to all regions. For example, for achieving the “diabetes FSD”, a two-step procedure of data extraction is carried out. The NHS beneficiaries who leave their ‘footprints’ suggestive of diabetes through specific services (i.e., at least one antidiabetic drug prescription, one hospital admission with primary or secondary diagnosis of diabetes, and/or co-payment exemption for diabetes) provided in a definite time-window, are identified in the first step. Their identification code represents the key for grabbing services provided to patients likely affected by diabetes as recorded from administrative archives in the second step. FSD covering areas of oncology, cardiovascular and respiratory diseases, mental disorders, and pregnancy-child health, are designed and achieved with analogous procedures. Data extract-transform-load (the so-called ETL procedure) for each FSD is designed uniformly through all the Italian regions participating to the Beaver<sup>®</sup> network. Differences regarding the time-window depth covered by administrative recording may depend on data availability.

The second type of database, which we call mixed-source database (MSD), concerns the covering of specific data available in a given region. Experiences of linking administrative data with information from population-based sampling surveys (e.g., the health examination survey

**FIGURE 1. General architecture for the regional Beaver® research platform**



managed for Italy by the National Institute of Health) and hospital-based disease registries (e.g., cancer registry from the National Cancer Institute of Milan) are ongoing in Lombardy. Analogously to FSD, the identification code of subjects included into the survey/registry (first step) serves for linking with administrative data (second step).

The second layer of the Beaver® is devoted to the users who are the physical persons belonging to either the regional administration or an accredited agency who obtained the credentials for accessing a database (FSD or MSD) after the Regional Authority approved the protocol. Two distinct solutions may be adopted.

The first solution (the so-called Beaver® Light front-end, attainable by regional administration personnel only) allows to automatically compute the set of process and outcome indicators defined by the Health Ministry. Indicators are those reported in the official manual for LEAs monitoring through the assessment of pathways experienced by the NHS beneficiaries suffering from chronic diseases (e.g., diabetes, heart failure, chronic obstructive pulmonary disease, breast, colon or rectum cancer, selected mental disorders, ...), experiencing acute episodes (e.g., myocardial infarction, haemorrhagic stroke, ischaemic stroke, ...), or who are going through a physiologic experience (e.g., pregnancy). By choosing a given FSD (e.g., the diabetes FSD), and a reference year (e.g., 2016), a standard report is generated containing

size and rates of the prevalent cohort (e.g., all diabetic patients), incident cohort (e.g., patients newly taken in care for diabetes), process indicators (e.g., prevalent cohort members who adhered to selected recommendations, such as assessments of glycated haemoglobin, lipid profile, urine albumin excretion, serum creatinine and dilated eye exams) and outcome indicators (e.g., incident cohort members who experienced at least one hospital admission for brief-term diabetes complications, uncontrolled diabetes, long-term vascular outcomes, and no traumatic lower limb amputation). Findings stratified for gender, age class, and possibly geographical area of residence (for example local health unit if present), may be obtained using Beaver® Light.

The second solution (the so-called Beaver® Full front-end, available for every authorized user), involves an eight-step procedure entirely driven by the protocol approved by the Regional Authority. An easy-to-use interface articulated in a sequence of queries and drop down menus for specifying (i) FSD (or MSD) from which data must to be processed, (ii) inclusion and (iii) exclusion criteria, defining for (iv) exposure(s), (v) covariates, (vi) outcome(s) and (vii) follow-up, allows of choosing time window(s), demographics (age and gender), diagnostics (ICD-9-CM codes), therapeutics (ATC codes), outpatient services (National Tariff Nomenclator), and other data useful for carrying out the study. The output of these first seven steps

**TABLE 1. Architecture component descriptions, implementation and operability purpose**

Graphical User Interface (GUI)	Being a web-platform, Beaver <sup>®</sup> ships with three modern frontend written in HyperText Markup Language (HTML), JavaScript <sup>™</sup> (JS) and Cascading Style Sheets (CSS): Beaver <sup>®</sup> Light, Beaver <sup>®</sup> Full and an administrative application. Beaver <sup>®</sup> Full provides a set of sophisticated input controllers designed to easily follow a protocol and to conduct a study, while Beaver <sup>®</sup> Light fully automates those processes.
Security	The three Beaver <sup>®</sup> front-ends provide a classical authentication system based on a username and a password. Beaver <sup>®</sup> never exposes records and/or tables to the final users and is designed to prevent any Standard Query Language (SQL) injection attack.
Engine	The core of Beaver <sup>®</sup> is written using "PHP: Hypertext Preprocessor" (PHP) language and is programmed for managing multiple users, receiving and decoding instructions, preparing the transactions for the data manipulation jobs, launching and monitoring the entire jobs life.
Database	Regional data and system data is stored in a local PostgreSQL cluster, a well-known robust and high performance open source Relational Database Management System (RDBMS).
ETL (Extract-Transform-Load)	From the administrative front-end it is possible to populate the database using the automated ETL functionalities developed specifically for Beaver <sup>®</sup> using the Python <sup>™</sup> language. The ETL component uses Apache Spark <sup>™</sup> and is designed to take into account each Region data availability, data structures and encodings divergences. It connects to the Regional data source, extracts the data, then cleans, normalizes and loads it into the local database. The loaded data undergoes an optimization process in the RDBMS environment and is made accessible from Beaver <sup>®</sup> Light and Beaver <sup>®</sup> Full.
Database transactions	When users start an elaboration, the engine produces one or more SQL transactions. These are composed of several highly optimized queries and are executed in a sequential order. Some queries are proved to work on mutually exclusive relations and for this reason can be parallelized. Beaver <sup>®</sup> provides an innovative module that easily distributes specific queries even if the adopted RDBMS (PostgreSQL) is supposed to work only with non-parallel queries. Furthermore, some queries, especially the ones involved in time-dependent analysis, are sent to Spark <sup>™</sup> in order to maximize the overall performance and for this reason are written in SparkSQL. Beaver <sup>®</sup> interacts with Spark <sup>™</sup> using a Scala module written specifically for this purpose.
Statistical analysis and reports	In order to run analysis and generate reports, the engine instructs the database to produce meaningful data. This process is very simple for certain models and gets more complex for time-dependent analysis. R gets called and receives instructions about what kind of analysis needs to be run on the data previously produced. The results are stored as JavaScript <sup>™</sup> Object Notation (JSON) strings into the database, while the GUI provides a human-readable version of them. Users have the ability to generate a fully printable report on a Adobe <sup>®</sup> Portable Document Format (PDF) file that uses a standard A4 paper layout.
External Tools (Apache Zeppelin, R, SAS <sup>™</sup> , etc.)	Advanced users with knowledge of the underlying Beaver <sup>®</sup> data structure, can operate using external tools in order to run custom and non-standard analysis. The most versatile and powerful tool is Apache Zeppelin, which provides a front-end for interacting with some interpreters used in data analysis (Python <sup>™</sup> , R, generic JDBC SQL drivers, Scala, Apache Spark <sup>™</sup> , Apache Hive <sup>™</sup> , Apache Pig, etc.). Assuming the license availability and a direct access to Beaver <sup>®</sup> database, advanced users can adopt any commercial statistical software like SAS <sup>™</sup> or Stata <sup>®</sup> .
Batch execution and error handling	A section of the Beaver <sup>®</sup> platform grants to users the possibility to monitor the jobs execution status in real-time. The engine has the ability to catch errors whenever they occur and logs each job activity. When a job execution ends, the user has full access to its detailed logs, the SQL transactions generated by the engine and some profiling information that can be useful to reveal bottlenecks and system performance degradation.

is a master table, still not accessible to users, which may be used for data analysis (i.e., the eighth and last step of the sequence). Usually, data analysis is made through another sequence of queries specifying variables of interest (possible transformed with respect to their original form) which must be used for obtaining descriptive statistics, as well as for fitting selected models (e.g., logit, log-binomial, Poisson, or Cox ones). However, in order to make Beaver<sup>®</sup>Full as flexible as possible and to apply other and more tailored models and algorithms, data may be directly processed through R or SAS<sup>™</sup> (the latter admitted that the licence is available in the regional environment). Finally, a report including results of data processing is obtained in the form of lists, tables and figures are returned to the user, admitted that the regional administrator approves it.

## DISCUSSION

In light of the increasing demands for low-cost real-world healthcare data and evidence, the new opportunity

arising from the Beaver<sup>®</sup> regional research platform, a web-based system for integrating and processing healthcare data, is described in the current paper.

Beaver<sup>®</sup> has several strengths. First, because the platform was designed and achieved by means of grants from institutional public agencies (please see Acknowledgments section), it is free of charge to the regions concerned. The contractual form defining intellectual property and the rules for installation, updating and maintenance of Beaver<sup>®</sup> is currently being studied, but it is beyond the competence of the authors of this paper.

Second, the rules established by the new European regulation for protection of natural persons in relation to the processing of personal data [15] limiting the free movement of electronic health data, are fully complied. In other terms, because each Italian region must be considered the owner of data on healthcare provided to NHS beneficiaries of that region, data should be stored and processed within a regional secured environment and its movements should be limited to few and exceptional needs (such as those of Health Ministry or for legal questions). Accordingly,

Beaver<sup>®</sup> was designed in order to (i) locally process data, (ii) prevent regional data from leaving the platform that hold it and (iii) allow users to see only the aggregate results of the analysis without any possibility to override the citizens' right to privacy.

Third, because evidence appropriately addressing knowledge-based policy improving effectiveness of healthcare and efficiency of health services are expected for processing real-world data, Beaver<sup>®</sup> has been designed as a technologic tool (i.e., the Beaver<sup>®</sup> platform) attendant the rules for scientific research, at which the Beaver<sup>®</sup> network is required to comply. For example, the platform implements algorithms for automatically computing adherence to recommended clinical examinations of NHS beneficiaries suffering from selected disease and conditions. It should be emphasized that the algorithms implemented in the platform were designed according to the Health Ministry official manual for LEA indicators, the latter being validated for their relationship with measurable clinical outcomes. Other than for monitoring healthcare, the Beaver<sup>®</sup> platform has been designed for allowing users, external to the regional administration, to generate solid evidence. However, physical persons may be authorized to access platform whether (i) they belong to a public agency (e.g., academy, or other research institutions), (ii) they have documented experience on generating evidence from secondary data, and (iii) a detailed study protocol complying with good practice of epidemiological research using secondary data [17-20] have been submitted to and approved from the Regional Authority. In other words, the reports which the Beaver<sup>®</sup> platform is able to produce, i.e., both the standard report with official data on process and outcome indicators of LEA monitoring, as well as the report expressly built for answering a specific research question, are generated accordingly with a predefined, approved protocol.

Finally, as regions where the platform is up and running are connected to the Beaver<sup>®</sup> network, and a common protocol always drives the data process, aggregated data generated within each region may be compared and/or summarized by means meta-analytic techniques [21,22]. This is very important because assessments of between-regions healthcare homogeneity (equity), as well as summarized evidence of healthcare implications, may be obtained with comparable data and methods.

## CONCLUSIONS AND FURTHER RESEARCH

In conclusion, the Beaver<sup>®</sup> regional research platform is a more than promising tool for stimulating healthcare research in Italy. It is currently implemented, or its implementation is ongoing, in some regions (Lombardy, Sicily, Sardinia, Friuli Venice Giulia, Marche and Abruzzo). A cross-validation study involving other regions (Molise and Toscana), other than those above reported, is currently under design.

We propose two priority directions for further research

[13]. First, there is an urgent need to 'unlock' more detailed data concerning health surveys, drug and disease registries, and large-scale, disease agnostic DNA and biological collections, so expanding and making it sustainable the range of record linkages in order to deliver precision medicine and innovate discovery of new drug targets or repurpose existing drugs. Second, there should be major expansion in the underlying methods and applications of innovative approaches using real-world data to design and carry out real-world studies. This is important for delivering learning health systems and patient benefit.

The recently established Center of Healthcare Research and Pharmacoepidemiology [<http://www.chrp.it/>], a consortium of sixteen public universities (to which seven other universities are being added), working with the close cooperation of public institutions (e.g., the National Institute of Health), many regions, research and treatment institutes, and scientific societies, has this as its main mission: to promote and spread methods and experiences on the appropriate, correct and safe use of real world data for generating robust evidence useful for addressing health policy.

## Acknowledgments

Support come from Italian (i) Ministry of Health within the FRAME (Flussi Regionali Automatizzati per il Monitoraggio dell'assistenza e la generazione di Evidenze scientifiche di indirizzo per le politiche sanitarie) project, grant number 2014-CONV-0001 and (ii) the National Institute of Health within the BRIDGE Health ("Bridging Information and Data Generation for Evidence based Health Policy and Research") European project, grant number 2017-CONV-0017; the Programme CCM 2015 – Area Azioni Centrali of the Center for Disease Control of the Italian Ministry of Health "Creazione e sviluppo del network italiano a supporto del progetto europeo BRIDGE-Health (BRidging Information and Data Generation for Evidence-based Health Policy and Research) finalizzato a dare strutturazione e sostenibilità alle attività europee nel campo della 'Health Information' ". Further supports will come from the Italian Ministry of Health within the "Metodologia per il monitoraggio e la valutazione dei percorsi diagnostico-terapeutico assistenziali ... nell'ambito del Nuovo Sistema di Garanzia dell'assistenza sanitaria" project (grant number 2017-CONV25-0095).

## Funding

The activities described in this paper were conducted in the framework of the Project 'Creazione e sviluppo del Network Italiano a supporto del progetto europeo BRIDGE-Health finalizzato a dare strutturazione e sostenibilità

alle attività europee nel campo della Health Information (HI)' (Creation and development of the Italian Network to support the European project BRIDGE-Health aimed at structuring and providing sustainability to European activities in the field of Health Information), funded by the Italian Ministry of Health, Centre of Disease Control (CCM), and in the framework of the European project 'BRIDGE Health - bridging Information and Data Generation for Evidence-based Health Policy and Research' funded by the European Commission/DG Santé (Agreement no-664691 – BRIDGE Health).

## Statement

The findings and conclusions provided in this paper are those of the authors, who are responsible for their contents; the findings and conclusions do not necessarily represent the views of the European Commission/DG Santé or the Italian Ministry of Health or the National Institute of Health. Therefore, no statement present in this report should be considered as official position of the European Commission/DG Santé or the Italian Ministry of Health or the National Institute of Health.

## References

1. Massa I, Balzi W, Burattini C, et al. The challenge of sustainability in healthcare systems: Frequency and cost of inappropriate patterns of breast cancer care (the E.Pic.A study). 2017;34:103-7
2. Prat G, Lefèvre M, Nowak E, et al. Intensive Care Med 2009;35:1047-53
3. Cantarutti A, Merlino L, Giaquinto C, Corrao G. Use of antidepressant medication in pregnancy and adverse neonatal outcomes: A population-based investigation. *Pharmacoepidemiol Drug Saf* 2017;26:1100-8
4. Salvatore S, Domanska D, Wood M, et al. *PLoS One* 2017;12:e0190101
5. Palanca-Maresca I, Ruiz-Antorán B, Centeno-Soto GA, et al. Prevalence and risk factors of prolonged corrected QT interval among children and adolescents treated with antipsychotic medications: A Long-Term Follow-Up in a Real-World Population. *J Clin Psychopharmacol* 2017;37:78-83
6. Sturkenboom M, Nicolosi A, Cantarutti L, et al. Incidence of mucocutaneous reactions in children treated with niflumic acid, other nonsteroidal antiinflammatory drugs, or nonopioid analgesics. *Pediatrics* 2005;116:e26-33
7. Deitelzweig S, Luo X, Gupta K, et al. Comparison of effectiveness and safety of treatment with apixaban vs. other oral anticoagulants among elderly nonvalvular atrial fibrillation patients. *Curr Med Res Opin* 2017;33:1745-4
8. Corrao G, Mazzola P, Monzio et al. Antihypertensive Medications, Loop Diuretics, and Risk of Hip Fracture in the Elderly: A Population-Based Cohort Study of 81,617 Italian Patients Newly Treated Between 2005 and 2009. *Drugs Aging* 2015;32:927-36
9. Hanna N, Leow JJ, Sun M, et al. Comparative effectiveness of robot-assisted vs. open radical cystectomy. *Urol Oncol* 2017. pii: S1078-1439(17)30481-7. doi: 10.1016/j.urolonc.2017.09.018
10. Taipale H, Mehtälä J, Tanskanen A, Tiihonen J. Comparative effectiveness of antipsychotic drugs for rehospitalization in schizophrenia. A nationwide study with 20-year follow-up. *Schizophr Bull* 2017. doi: 10.1093/schbul/sbx176
11. Lee D, Gladwell D, Batty AJ, et al. The cost effectiveness of licensed oromucosal midazolam (Buccolam®) for the treatment of children experiencing acute epileptic seizures: an approach when trial evidence is limited. *Paediatr Drugs* 2013;15:151-62
12. Scotti L, Arfè A, Zambon A, et al. Cost-effectiveness of enhancing adherence with oral bisphosphonates treatment in osteoporotic women: an empirical approach based on healthcare utilisation databases. *BMJ Open* 2014;4:e003758
13. Hemingway H, Feder GS, Fitzpatrick NK, et al. Using nationwide 'big data' from linked electronic health records to help improve outcomes in cardiovascular diseases: 33 studies using methods from epidemiology, informatics, economics and social science in the ClinicAl disease research using Linked Bespoke studies and Electronic health Records (CALIBER) programme. Southampton (UK): NIHR Journals Library; 2017 Jan. [Programme Grants for Applied Research, No. 5.4.] Available from <https://www.ncbi.nlm.nih.gov/books/NBK414778/> doi: 10.3310/pgfar05040
14. Senato della Repubblica, Camera dei Deputati. Definizione e aggiornamento dei livelli essenziali di assistenza (LEA). Novembre 2016. Available at <https://www.senato.it/service/PDF/PDFServer/BGT/00994183.pdf> (accessed December 30.12.2017)
15. European Parliament and Council. Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, 4.5.2016. Available from <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN> (accessed December 30, 2017)
16. Garante per la protezione dei dati personali. Autorizzazione n. 9/2016 - Autorizzazione generale al trattamento dei dati personali effettuato per scopi di ricerca scientifica - 15 dicembre 2016 [5805552]. Available at <http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/5805552> (accessed December 30.12.2017)
17. Berger ML, Mamdani M, Atkins D, et al. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part I. *Value in Health* 2002;12:1044-52
18. Corrao G. Building reliable evidence from real-world data: methods, cautiousness and recommendations. *Epidemiology Biostatistics & Public Health* 2013;10:e8981
19. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, eds. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. AHRQ Publication No. 12(13)-EHC099. Rockville, MD: Agency for Healthcare Research and Quality; January 2013
20. ENCePP Guide on Methodological Standards in Pharmacoepidemiology. Available from <http://www.encepp.eu/>

- standards\_and\_guidances/methodologicalGuide.shtml (accessed December 30, 2017)
21. La Gamba F, Corrao G, Romio S, et al. Combining evidence from multiple electronic health care databases: performances of one-stage and two-stage meta-analysis in matched case-control studies. *Pharmacoepidemiol Drug Saf* 2017;26:1213-9
  22. Scotti L, Rea F, Corrao G. One-stage and two-stage meta-analysis of individual participant data led to consistent summarized evidence: lessons learned from combining multiple databases. *J Clin Epidemiol* 2017 Nov 29. pii: S0895-4356(17)30113-0

