Deliverable 3.2

Good practice on data linkages and performance measurement in relation to access to national health care data systems
**Table of content**

1. Introduction ...........................................................................................................................................3
2. Data requirements for performance measures .......................................................................................4
3. Sources of data for health services research .........................................................................................6
4. Health care membership, administrative and clinical data .................................................................8
5. Data Linkage ...........................................................................................................................................9
6. Privacy and confidentiality issues ......................................................................................................11
7. Identifying regional and national data systems ..................................................................................14
8. Mapping "Best Practices" ...................................................................................................................16
9. Access to national healthcare data systems - recommendations ....................................................25

References ..................................................................................................................................................27

Appendix I - EuroREACH Advisors meeting, Israel ...........................................................................31
Appendix II ...............................................................................................................................................33
  Figure 1: Decision tool for selection of relevant databases for inclusion in a national electronic inventory ...........................................................................................................................................33
  Figure 2: Criteria for assessing the coverage and accuracy of a clinical database ..................................34

Appendix III - EuroREACH Database "Passport" ..................................................................................35
Appendix IV - Narrative for each country's data capabilities .................................................................37
Appendix V ...............................................................................................................................................38
  Table 1: Summary of country narratives ............................................................................................38
  Table 2: Mapping the availability of person-level healthcare data sets and related data by country ...........................................................................................................................................44

Appendix VI - Data Linkage: References ...............................................................................................50
Health services and population data are required for a wide range of purposes, including disease surveillance, health systems management, effectiveness research and strategic planning. Within any governmental unit (country, state, province or regional authority) numerous entities may generate and maintain data that could potentially contribute to health services and comparative effectiveness research, quality monitoring and other purposes.

The breadth and level of details available for each data system dictate the research and policy objectives that can be addressed. For example, aggregate data on hospital discharges could facilitate surveillance for certain diagnoses, or monitoring of trends in utilization and service needs, but would be insufficient for investigating the demographic and clinical factors associated with these diagnoses.

The task established for Work Package 3 is to "review national progress with, and good practice examples in establishing comprehensive systems of performance measurement in European countries and in granting researchers access to corresponding patient-oriented data for the study of health services in a disease-based perspective." Thus the Work Package objectives reference two distinct systems of data, 1) systems of performance measurement, and 2) patient-oriented (or person-level, when referring to the general population) data for the study of health services. The distinction between these two types of data is an important one. A number of countries have developed performance measures for health care. While the types of measures and the specific issues to be addressed in each country vary with demographic makeup, morbidity patterns, local health needs, and other factors, most systems of performance measures have in common that they consider multiple factors, including quality of care, cost of care and access to care (1). Examples of existing performance measurement systems include the Healthcare Effectiveness Data and Information Set (HEDIS) (2, 3), U.S. Agency for Healthcare Research and Quality - Quality Indicators (AHRQ QI) (4) UK National Health Service Indicators for Quality Improvement (5), the Israel Ministry of Health’s Quality Indicators for Community Health Care (6), and European Commission’s European Community Health Indicators (ECHI) project (7).
2 Data requirements for performance measures

A performance measure is expressed as a rate or percentage. The measure may reflect processes or outcomes of care in a particular population of interest, for example, the percentage of diabetics having a glycohemoglobin measurement in the preceding year (process), or the percentage of diabetics with LDL under 100 mg/dl (outcome) (3). Other measures reflect system factors, and have as their unit of analysis institutions or providers rather than individuals. An example is the NHS measure of acute units with key characteristics deemed essential to appropriate care (including continuous physiologic monitoring and acute stroke protocols/guidelines) (5). The basic requirement for each measure is the determination of the population or subset of events of interest (the denominator) and the number in the denominator meeting the criteria for the measure of interest (the numerator).

Performance measures may be calculated on the basis of an entire population or on a representative sample. Data required for the calculation of each measure may be gathered from electronic databases, population surveys, chart review or other methods. Performance measures systems store the resulting data in aggregate form, precluding analysis of factors potentially related to performance that were not considered in the process of defining the measure. For example, in the case of a measure defined to track glycohemoglobin testing in diabetes ages 18 years and over, later analysis of performance in subgroups of different ages would not be feasible unless the decision had been made at the outset that data would be collected for these subgroups. Should measures undergo changes in definition over time, it would be difficult or impossible to compare performance in different years without repeating the data gathering and calculation process on the basis of the new measurement criteria.

Research data systems differ from repositories of quality indicators in the following respects:

a. Data are maintained at the level of individual, rather than in aggregate form. A unique identifier common to all data sources provides the simplest solution to linkage of files from multiple sources; in the absence of such an identifier, probabilistic linkage methods must be developed.
b. Component data systems must be automated

c. Component data systems and participating organizations use common coding systems for diagnoses, treatments, pharmaceuticals and other services, provider identity and type or maintain continuously updated dictionaries.

d. Component data systems report a standard set of elements for each type of service (example: date, provider, patient, diagnosis for physician visit)

e. Component data systems are maintained in or can be transferred to a uniform automated format (standard software, platform)

f. Systems consist of databases linked for different purposes, allowing for future development and expansion as new data sources become available and are added to the system.

g. Content is not limited to address specific research topics determined in advance. Thus, new projects can be carried out in relatively short periods of time to address urgent issues.
3 Sources of data for health services research

Data for the calculation of performance measures are in many cases derived from person-level systems maintained for administrative, quality monitoring and research purposes. The following are examples of potentially useful data sources that could be linked to form a comprehensive system:

a. Census data

b. Vital records (birth and death registries)

c. Health care membership and administrative data

i. Eligibility/membership files

ii. Claims, including date and place of service, identity of provider/patient, diagnoses, treatment provided

iii. Pharmacy purchases

iv. Laboratory test results

v. Clinical data from electronic medical records

d. Communicable disease/adverse event reporting systems

e. Chronic disease registries

f. Large-scaled population-based surveys

The European Community Health Indicators (ECHI) Monitoring project surveyed data registries commonly maintained by member countries and discusses the strengths and weaknesses of each, as well as the level to which they may be comparable across countries (7). In compiling data for a region, it is also crucial to define the population covered by the system, in order to allow for calculation of rates of events.

Several useful guidelines exist for the assessment of the utility of data sources for a comprehensive data system. In their survey of health care databases in British Columbia Black et al. (2005), provide a decision tree (Figure 1) for determining which data sources to
include in the system as well as an evaluation questionnaire (8). The British Directory of Clinical Databases uses an extensive evaluation tool to assess the quality of the databases included (9, 10) (Figure 2) (38). Indicators of a high-quality database, according to Black, include the availability of individual-level data, a covered population representative of the geographic unit of interest, completeness of data collection, reliability of coding of conditions and interventions. The use of numeric scores for each of the criteria assessed allows for comparison between databases (10).
4 Health care membership, administrative and clinical data

Data generated in the course of delivery of and reimbursement for health care offer a valuable resource for health services and comparative effectiveness research. The use of electronic medical records and automated claims processing facilitate access to the information required to identify exposures and outcomes of interest in large populations.

Administrative data are by definition collected and maintained to support the operations of the organization, rather than for the purposes of research. This fact can impact adversely on data quality and availability. For example, administrative files will capture utilization of those services that the organization has an interest in tracking for purposes of payment or management; services obtained outside the system or those for which a separate payment is not generated will not be captured. Precision in diagnostic and procedure coding is limited by the coding system and practices of the organizations and providers. Organizations will likely resist additional data collection or reporting mandates that may be required to enrich the value of their files for research purposes, but do not add value to their administrative processes. Finally, data sharing agreements must take into consideration patient privacy as well as the proprietary nature of the data for the organization.

In systems with stable populations, study of long-term outcomes is also possible.

In general, a prerequisite for research of a specific issue (e.g. describing outcomes in diabetics receiving different treatment modalities) is a comprehensive data system that allows for the following:

a. Identification and description of the population included in the data system, including demographic characteristics and vital status.

b. Unbiased determination of conditions of interest, exposures to treatments under study and outcomes (through recorded diagnoses, sentinel health events, laboratory results, pharmacy purchases, and other types of encounters).

c. Ability to link demographics, clinical characteristics and service utilization at the patient level.
5  Data Linkage

The wide range of data is most easily accessed for populations who receive the bulk of their medical care within a closed system administered by a single payer and for which detailed encounter data are collected in a standardized format. In such cases, there is reasonable certainty that the data system will offer near-complete, unbiased capture of the variables of interest (British General Practice Research Database, Canadian provinces). In many areas, however, such as in the United States, patients have multiple sources of coverage (dual Medicare-Medicaid or Medicare-Veterans Health Administration in the United States are two examples), and linkage between multiple data sets is required to assemble a complete picture of the health care utilization. Ad hoc projects to access and link files for a specific project are costly and time-consuming. Cooperative efforts to link data held by government entities, insurers and individual researchers offer the possibility of assembling a complete picture of service utilization in the population of interest, easing data access for researchers, reducing costs of research and allowing more timely investigation of pressing issues (11). Preparation of a linked data set involves identifying the sources and quality of the required data elements and establishing a method of combining them to create a more complete picture of the experience of individuals than could be obtained from any single data source. Data linkage requires not only a thorough understanding of the component databases to be linked, but also expertise in statistics and programming in order to establish a methodology for identifying matches between files, while minimizing errors. Linkage of data is simplest when all of the data sources use a common unique key to identify individual members. The ideal identifier is unique, permanent, and applicable to the entire population of interest. Unique identifiers assigned at birth exist in a number of countries, including Sweden, Norway, Denmark, and Israel. In practice, however, numbering systems are not universal, even within health systems. Therefore, other identifying information, such as name, birth date, gender, and residence may be taken into consideration in order to identify matching records. In the event that the purpose of the project is to identify discrete episodes of care, dates of service may also be of importance in linking different services related to a single episode of care.
Two basic methods exist for linkage of disparate data sets, deterministic and probabilistic. Deterministic linkage requires an exact match between linkage variables (identity number, last name/first name, etc.). In cases in which data entry errors, name changes and other factors have resulted in differences between linkage variables in the two files (incorrect coding of identity number, or the appearance of a maiden name in one file and a married name in the other), true matches between records will be missed. In probabilistic linkage, in contrast, less than exact matches may be accepted, according to a predetermined method that assigns a score to the level of the match. Level of acceptable error depends on how crucial identification of a specific person is. Different fields may be given different weight (matched birth date may be more important than matching spelling of last name). In cases in which the goal is to link data collected by multiple agencies, probabilistic matching may allow more complete matching of records related to the same individual, at the expense of an increased risk of incorrect matches.

The following issues must be taken into consideration in linking data from multiple sources for the creation of a research data set:

a. Is use of health care data and linkage to other data sources for the purposes of research done with the express permission of the subjects, or with a waiver of consent?

b. Who will be responsible for creating linked data sets? Where will the linked data set be located? Who will have access and under what circumstances? Will linked data be de-identified?

c. An accurate listing of the underlying population is crucial for the calculation of rates of health events and mortality. Examples of linked national data systems that incorporate a population registry include those of Scandinavian countries, Canadian provinces, and the Western Australian registry, which was built from electoral roll registrations.

d. How will the privacy of individuals in the data sets be protected? One option is to remove identifiers from the final file after linkage has been completed. Other options include single-coding (wherein the researcher holds a key to identify participants separately from the research file) and double-coding (wherein the data custodian, rather than the researcher using the data, holds the key to identifying participants).

A more comprehensive references list can be found in appendix VI
6 Privacy and confidentiality issues

One of the great strengths of administrative data systems is the availability of personal identifiers such as identification numbers, name, gender, date of birth, address and postal code. These identifiers are essential for meaningful research in that they enable linkage between data sources, permit analysis of outcomes according to socio-demographic factors, and allow for characterization of the underlying study population, essential for determining the extent to which the study results are generalizable (13). In addition to personal identifiers, health care records will necessarily include sensitive data regarding medical and family history, personal behaviors and risk factors for disease. The fact that such data are not specifically collected for research raises the question of whether it is in fact ethical and appropriate to use them for secondary purposes. Establishing the circumstances in which personal health data can be used for research requires balancing the good to the public to be achieved through such research with the individual’s right to privacy (14, 15).

In general, it is acceptable for organizations to use their own data for internal purposes, for example for auditing, quality monitoring, utilization review and other activities to support the mission of the organization, without prior consent of the subject. In contrast, research involving human subjects, even when such research is based entirely on patient care data and does not involve patient contact or treatment interventions, requires explicit consideration of the risks and benefits to the individual, and the steps to be taken to protect patient privacy. Distinguishing between such internal monitoring activities and research can be difficult. In fact, the Institute of Medicine (IOM) suggests that it is more appropriate to view these activities as existing on a continuum, explaining:

"Some HSR projects are clear examples of research; applying scientific methods to test hypotheses and produce new, generalizable knowledge. Other projects are certainly clear examples of internal exercises to assess the quality of the operations of the specific organization with no intention of producing generalizable knowledge. At the same time, quality assessment and quality improvement (QA and QI) exercises sometimes reveal interesting and important data that the organization recognizes to be of general interest, and that therefore ought to be published. In addition, both scientific research in health services
and investigations into the internal operation of a health services organization use many of the same methods (e.g., chart review, database analysis and linkage) (16).

Lowrance describes three basic approaches to the issue of secondary use of patient care data, and details the problems inherent in each: 1) Use data only with consent of the subjects, 2) use personal data without consent in cases in which this would be in the public interest and, 3) remove or alter identifiers (whether reversibly or irreversibly) prior to using the data for research purposes (17). Requiring patient consent to use clinical data for research purposes is likely to be impractical, particularly in large health care systems with many years' worth of data. Furthermore, requiring individual consent for use of health care data for research purposes would likely result in biased results, since subjects willing to give consent may differ in many ways from those refusing it (16). Invoking the public interest may be appropriate in situations in which the research is for purpose of responding to and controlling public health threats, but not all studies will meet this threshold (17).

The Canadian Institutes of Health Research (CIHR) Best Practices for Protecting Privacy in Health Research lays out circumstances under which an ethics board may consider a waiver of informed consent:

"(i) The research involves no more than minimal risk to the subjects;
(ii) The waiver or alteration is unlikely to adversely affect the rights and welfare of the subjects;
(iii) The research could not practicably be carried out without the waiver or alteration;
(iv) Whenever possible and appropriate, the subjects will be provided with additional pertinent information after participation; and
(v) The waived or altered consent does not involve a therapeutic intervention (18)."

Additional considerations include the determination that initiating contact with subjects appearing in data sets may in itself be harmful, or may have been prohibited by law or by a previous data-sharing agreement (18).

The US Health Insurance Portability and Accountability Act (HIPAA) (19) distinguishes between two types of health care data. *Protected health information* includes patient
identifiers, or sufficient data to allow an individual to be identified. De-identified data is defined as any file lacking eighteen specific types of identifiers (including name, address, unique identifiers such as SSN, birth, death, admission and discharge dates), or for which the risk of re-identification of an individual has been statistically determined to be very small. Use of protected health information without patient consent is permitted for treatment, payment and operations. However, patient consent, or IRB waiver of consent is required in order to use such data for research purposes. In contrast, de-identified data sets may be used without consent (20). Protected information may be used without consent for public health activities carried out by public health authorities (21). Contrary to the intentions of the HIPAA privacy rule, a 2009 Institute of Medicine report concluded it “does not protect privacy as well as it should, and that, as currently implemented, [it] impedes important health research (15).”

In Europe, the EU Data Protection Directive recognizes in principle the need to achieve balance between “the protection of fundamental rights and freedoms of individuals and in particular the fundamental right to data protection, on the one hand, and the achievement of the internal market – the free flow of personal data in this case – on the other (22).” In practice, implementation of this directive varies widely in EU member states, resulting in a great disparity in the extent to which health data are available for public health needs, such that well-developed data systems exist in some countries, while use of personal health data without individual consent, or of administrative data collected for another purpose is impossible in others (23, 24). A survey of health services researchers and experts in data protection conducted under the Health Information and Knowledge Strand of the EC Directorate General for Health (DG SANCO) identified numerous problems associated with accessing health care data for public health purposes. Areas in which problems were identified included 1) linkage of data, due to the lack of personal identifier or local restrictions on access to data with such identifiers, 2) the need for informed consent for the use of individual health care data (24).
7 Identifying regional and national data systems

We used three methods to map existing data systems and to identify "best practices"

a. Literature search

We searched the medical literature (using the U.S. National Library of Medicine's PubMed site) and the internet (using Google) for articles and internet sites using the following search terms: Distributed data networks, data linkage/health care data linkage, comprehensive national data systems, national health information systems, health services research databases, health care data infrastructure. In addition, we sought information on well-known health care databases, such as the British General Practice Research Database, the U.S. Veterans' Administration and Medicare data systems, and the Diabetes Audit and Research in Tayside, Scotland (DARTS) database. Bibliographic listings provided additional information sources.

b. EuroREACH expert panel

Expert panel members are specialists in the areas of epidemiology, health services and public health research, health care administrative data and data linkage in Australia, Canada, Denmark, England, France and the United States. During the first meeting of the expert panel, held in Tel Aviv, Israel on May 24-25, 2011, we gathered information on issues related to the development of comprehensive regional and national databases for health research including privacy protection, linkage strategies and governance. Furthermore, several members on the panel reported specifics on comprehensive linked data systems existing in their own countries. Agenda and list of participating experts can be found in appendix I.

c. Survey of countries participating in the EuroREACH project

We developed three survey tools in order to assess the data capabilities of each participating country. The primary focus of the survey was the availability of person-level health services utilization data, covering a significant, representative proportion of the national population.
The survey tools consisted of the following:

- Decision tool for evaluating individual databases
- Database passport
- Country narrative
- Diabetes case study

The decision tool is a flow chart for selection of relevant databases for inclusion in our national electronic inventory (for details see appendix II).

After pilot testing of the survey instruments, we incorporated them into an internet-based survey interface and requested that each member of the project team to complete the country narrative and up to five database passports meeting the criteria in the decision tool. The survey was conducted from August-October 2011 on the platform of the EuroREACH website.

The database passport includes information about its custodian, time period covered, population included type and source of data. The availability of the data for research as well as linkage availability and methods used were also part of the passport (see appendix III). The results of the mapping can be found in Table 2.

Information in the country narrative included description of the availability of healthcare data in the country, what are the uses of the data, processes of collection and stewardship and what are the barriers for accessing the data for comparative research. The questioner can be found in appendix IV and the results of the survey are presented in table 1.
8. Mapping "Best Practices"

The literature review, database survey and expert panel input together have enabled us to identify the following list of "best practices" that demonstrate effective use of administrative data for research and quality monitoring.

a. Great Britain - Oxford Record Linkage Study*

The Oxford record linkage study (ORLS) comprises all hospitalizations, all deaths and some birth certificates for the population of the jurisdictions covered by the administrative Health Authority of the Oxford Region (population 2.5 m), for the period from 1963-2009. Hospitalizations and deaths for all of England have been linked for the period from 1998 to 2009, while unlinked hospitalization data are available from 1968 onward.

ORLS began as a joint project between the National Health Service (Oxford Regional Health Authority (RHA)) and academics (University of Oxford) and was funded largely by the NHS. Its ‘home’ was in the Oxford RHA (i.e. NHS). Data collected were mainly those that the NHS routinely collected, along with names and addresses, which were used for linkage until the inclusion of the NHS number as an identifier on hospital records became more widespread. Mortality data were provided by the Office for National Statistics. The rationale behind the ORLS was to maximize the value of existing data by making linkage possible. With the abolition of the Regional Health Authorities in 1995, hospital statistical datasets were deleted in other regions, while ORLS shifted to the University Unit and continued to function, collecting hospital data from individual health authorities within the former Oxford RHA. In 1995, the Department of Health awarded ORLS the contract to perform record linkage of national hospital and mortality data. Permission was received to extract data pertaining to the Oxford Region from the national dataset and incorporate them into the ORLS. Since 2005, the NHS National Information Centre (funded by the Department of Health) has held English national data on a variety of topics and performs linkage. The Oxford group continues to do English national linkage with funding from the National Institute for Health Research, and continues to take the Oxford subset for ORLS.

* Source: Professor Michael Goldacre, presentation to the EuroREACH expert panel meeting, May 24, 2011, Tel Aviv, Israel
Barriers to the creation and maintenance of the ORLS have included privacy and legal concerns and costs. In the early years, available IT resources were inadequate to facilitate collection of all of the data that might have been useful (for example, primary care data are not available in the system).

Specifics on privacy and legal issues:
Until 1998, Medical Research Council guidelines were generally supportive of the use of personal health data, if appropriate clinical and ethical permissions were obtained. In 1998 the Data Protection and Human Rights Acts were passed. At that time, the General Medical Council (physicians’ regulatory body) expressed concern about use of data outside the clinical care setting. During the first decade of the 21st century, the Health and Social Care Act confirmed that use of personal health data for public benefit, outside the clinical care setting, is legal if duly authorized, and established the PIAG (now NIGB) to rule on whether specific uses can be made.

b. General Practice Research Database (GPRD)

The GPRD is a primary care database containing information derived from electronic medicals records for approximately 8% of the population of the United Kingdom. Over 600 practices contribute data to the GPRD, and the system included data for over 11 million persons (26). Participating practices record each episode of illness, new symptoms, patient encounters, diagnoses and abnormal laboratory test results, referrals to outpatient clinics and hospital admissions. The system also captures all prescriptions generated by the GP and other information recorded by practice staff, including vaccinations, weight and blood pressure measurements (27). GPRD data has been linked to national death and hospitalization data, disease registers at the person level, and to socioeconomic and census data at the small area level (26).

The Directory of Clinical Databases (http://www.icapp.nhs.uk/docdat) created in 2001, provides descriptions and independent assessment of multicenter clinical databases containing person-level information in existence in the UK. The directory was compiled through structured interviews with data custodians and data quality was assessed using a standardized instrument (9).
c. Denmark*

The population of Denmark numbers approximately 5 million people. Health care coverage is universal. Many types of health care registries have been developed, and it is likely that no researcher has a complete list of them all. (A summary of the most important registers may be found at [http://sjp.sagepub.com/content/39/7_supp]. No single organization or government agency is explicitly responsible for coordinating the development and maintenance of registers. The National Board of health is responsible for collecting and maintaining most health-related data. Data collected in Denmark include the Civil Registration System (CRS), prescription drug purchases, hospital inpatient, emergency and outpatient encounters, admissions to psychiatric hospitals, a range of disease-specific registries, and cause of death. Collection of primary care data has recently begun, facilitated by the fact that all physicians use a single EMR for recording visit data. Physicians have historically been resistant to report details of a visit, therefore the primary care data are generally limited to cost of visit, and therefore lack details on diagnoses recorded and services provided. Linkage of different registries can be accomplished using the unique identifier assigned by the CRS, the CPR number. The CRS collects information allowing for linkage between spouses and between parents and children. Informed consent for research using health care data is required only if the study involves contacting patients and collecting additional information or samples. Citizens are given the option at the time they update their information in the Civil Registration system to request that they not be contacted for research studies; approximately 25% of the population “opt-out” in this fashion.

A listing of Danish registers may be found at [http://www.ecreph.dk/Database.aspx]

d. Australia: Western Australia Data Linkage System*

WADLS was established in 1995 and is managed by the Data Linkage Branch (DLB) of the Department of Health of Western Australia (DoHWA). The system currently covers approximately 2.3 million people. The functions of the WADLS are:

* Source: Carsten Pedersen, presentation to the EuroREACH expert panel meeting, May 24, 2011, Tel Aviv, Israel
* Source: Emma Fuller, presentation to the EuroREACH expert panel meeting, May 24, 2011, Tel Aviv, Israel
1) Performing linkage within and between core data collections held by the DoHWA,
2) Geocoding of address information on health records, and
3) Providing linked and geocoded data to support health planning, evaluation and research, as well as offering guidance on the use of the data system.

The “core” data linkage project incorporates data from the following sources:
1) DoHWA (hospital discharges, midwives notifications, cancer registrations, mental health contacts, emergency presentations
2) Registry of Births, Deaths and Marriages
3) Electoral Commission, Western Australia. The electoral roll constitutes the primary population listing for the system.

In order to protect the privacy of persons in the system, the linkage and analysis tasks are performed separately. Linkage is accomplished using identifiers only. Linked datasets have identifiers removed before they are made available to researchers. The data linkage process involves probabilistic methods to calculate the likelihood that two records belong to the same entity (person, family, event and location).

Applicants for data must submit an application form specifying the datasets and variables required for the project. Applications must be approved not only by the Department of Health and the WADLS, but also by the custodians of each of the data collections from which information is requested. Data for each linkage project are encrypted. Encryption is different for each project, such that it would be impossible to link datasets produced for two different projects. Applicants must certify that they will not attempt to re-identify individuals in the dataset or to use a dataset created for a specific project for a new project.

Between 1995 and 2011, over 750 research projects have been conducted using WADLS linked data. The establishment of the system has promoted a reduction in the use of “name identified” data for health services research in WA. The system offers training and support to researchers using their data.
e. Population Health Research Network (PHRN)

The Population Health Research Network (PHRN) is a 21st century initiative to collaboratively build nationwide data linkage infrastructure for population health related research in Australia. Bona fide researchers will be able to access linkable de-identified data from a diverse and rich range of health datasets, across jurisdictions and sectors in order to support nationally and internationally significant population based research that will improve health and enhance the delivery of health care services in Australia. Collaborating nodes are located in Western Australia, New South Wales/Australian Capital Territory, Victoria, Queensland, South Australia/Northern Territory and Tasmania. The Australian Institute of Health and Welfare is a network participant. Cross-jurisdiction linkage is undertaken by the Centre for Data Linkage at Curtin University in Perth and the Sax Institute in Sydney is developing a secure remote access laboratory to provide approved researchers with safe access to linked data.

Three proof-of-concept collaborations have been designed to use data sourced and linked across jurisdictional boundaries, to demonstrate that Australia’s rich information resources could be used more strategically to provide a resource for examining rare events and achieving complete case or outcome assessment, which may otherwise be limited by cross-border flows. Each of the projects was chosen against the following criteria: ability to demonstrate linkage of data between two or more state or territory nodes; ability to provide research datasets combining datasets for two or more nodes; demonstrated ability to deliver research outputs based on a combined dataset; ability to complete the project on time and on budget; and focus on an issue of national importance.

For more information see www.phrn.org.au

f. United States

Health care coverage in the US is provided by a large number of players from the public and private sector. Many health plans have significant turnover in membership, and there is a large uninsured population. Thus, assembling data for large, representative portions of the population for research purposes constitutes a considerable challenge. EMRs are not in wide
use in the US and therefore automated clinical data may be difficult to obtain (11). Linkage
between administrative datasets and clinical data from EMRs (vital signs, electronic
laboratory data, etc.) is more feasible within health care organizations. One widely used
source of administrative data for research purposes is the Medicare program, the largest single
provider of health insurance in the US (28). Medicare provides health care coverage to about
97% of the elderly in the US, and maintains a rich system of administrative data for its
covered population (29). Medicare data systems fall into three categories 1) identifiable data
files, 2) limited data set (LDS) files, and 3) non-identifiable data files (30). Identifiable data
files contain person-level demographic and utilization data with provider and patient
 identifiers, and are made available to researchers only after approval of a formal data request.
The availability of a personal identifier in these files allows for the linkage of data from many
sources (hospital inpatient and outpatient, physician visits, demographic data). LDS files
contain information similar to that available in the identifiable files, but lack direct identifiers
as specified in HIPAA rules. De-identified Medicare claims data ("basic files") which have
been altered further to protect beneficiary confidentiality are public use files available online
(31). Medicare data have been linked to data from the Surveillance, Epidemiology, and End
Results (SEER) cancer database, the Medicare Current Beneficiary Survey, the National
Health Interview Survey, the National Health and Nutrition Examination Survey, among other
databases (29).

The US Medicaid program is jointly funded by the states and the federal government to
provide medical care to low-income and disabled persons. While federal law established
guidelines and mandated coverage of some groups, individual states have considerable latitude
in determining eligibility, therefore the covered population varies from state to state (29a,
29c). The Centers for Medicare and Medicaid Services (CMS) maintains person-level data on
Medicaid eligibility, service utilization and payments.

Most US states maintain databases containing details of hospital admissions, although the
specific elements collected vary by state (31a). Recognizing that discharge data alone are not
sufficient to answer many needs, individual US states are developing "all payer claims
databases" (APCD) (32) which are collections of utilization of health care services by state
residents, regardless of insurer or provider. Items of interest include medical, dental and
pharmacy services, as well as information on providers and patients. As more such databases
are built at the state level, there will be greater interest in standardizing the format and coding of common data elements so that comparisons between states will be possible. Sources of data commonly included in APCD include eligibility, private and commercial claims, Medicare and Medicaid claims. Problems with implementation and maintenance of APCDs include 1) The absence in the US of a standardized national patient identifier, which may make linking of data in and between states difficult, 2) assuring compliance in reporting, which is likely to be higher in states in which mandatory reporting has been established by legislation, 3) standardizing data received from multiple sources into a single consistent format, 4) incomplete capture of service utilization by the uninsured (12).

The Mini-Sentinel distributed data system (33), sponsored by the US Food and Drug Administration, serves as an example of the potential for the pooling of data from multiple custodians to answer questions of national interest. Health plans and other organizations collecting automated patient care data (for example, inpatient hospital, pharmacy purchasing, physician visits, clinical data from electronic medical records, laboratory) create research data sets according to a standard format. These data sets remain in the hands of the custodians, but may be queried through the use of programs distributed by the coordinating center. The results of these queries are processed by the coordinating center. The use of a distributed data system and analysis model allows the conduct of population-based health services research without the need for release of sensitive medical information by the custodial organization. Furthermore, the existing infrastructure allows for rapid response to urgent questions.

Currently the system contains information for more than 60 million persons in the US,

The Veterans Health Administration is the largest integrated health care system in the United States, with 162 VA hospitals, 137 nursing homes, 43 domiciliary and more than 850 community and facility-based clinics (34). In 2010, the VHA system had 8.3 million enrollees (35). The VHA maintains extensive clinical and administrative data which support patient care and provide the basis for effectiveness research (36). Person-level data systems that have been used for research purposes include the Medical SAS datasets, patient and practitioner databases, and condition specific databases (cancer, agent orange, substance abuse, spinal injury). Data are made available to researchers via the VA Information Resource Center (VIReC) (34).
At the national level, the National Health Care Utilization and Cost Project (HCUP) contains all discharge data from 1,056 hospitals located in 42 States, approximating a 20-percent stratified sample of U.S. community hospitals (37).

The U.S. National Library of Medicine maintains HSR Information Central, a selection of information sources at the state, federal and international level (38), which includes information on interactive reports, aggregated data, surveys, individual level data, and access to a searchable database. Another on-line data source is the Centers for Disease Control (CDC) Wonder internet site (39).

g. Canada*

Responsibility for health care in Canada is divided between the national and provincial governments. Funding is transferred to the provinces by the national government, and the provinces provide care. Approximately 70% of health care in Canada is publicly funded (lower than the OECD average), the remainder is privately funded.

CIHI is a not-for-profit corporation that maintains 27 health databases and registries (40, 41). The CIHI mandate is to coordinate, develop, maintain and disseminate health information in Canada, with less emphasis on providing data for to health services researchers. CIHI data are drawn from a wide range of sources, including hospitals, regional health authorities, professional organizations, regulatory authorities, and provincial and territorial government. The extent to which information is collected and can be identified varies with the source. For example, while some provinces report data to CIHI with the health card number, others report using a scrambled identification number that cannot be re-identified by CIHI. Person-level databases include the following:

i. Acute hospital care, including hospital admissions, emergency department visits, day surgery and outpatient clinic visits. CIHI currently receives data on approximately 95% of hospital services in Canada

ii. Specialized services, including mental health, rehabilitation, trauma and injuries, knee and hip replacement, and organ transplantation

iii. Home care

* Source: Kira Leeb, presentation to the EuroREACH expert panel meeting, May 24, 2011, Tel Aviv, Israel
iv. Continuing and long-term care

v. Pharmaceutical use

In addition, CIHI collects data on health spending. The Canadian Management Information Systems (MIS) Database includes information on staffing, costs, workload and provision of services at the level of hospital department or health region. Data are reported by hospitals and health regions, generally via respective Ministries of Health. Some information on socioeconomic status is available at the level of postal code.

CIHI staff and all outside researchers must sign confidentiality and non-disclosure agreements in order to access CIHI data. Data-sharing and bilateral agreements between CIHI and other agencies and organizations, provinces and territories govern the types of data that may be transmitted and released and how they may be used.

CIHI conducts studies involving linkage across databases in order to assemble a full record of care provided for specific conditions. For example, a study of care provided to stroke patients in Ontario required linkage of emergency department, inpatient, rehabilitation and home care data. In another example, linkage of hospital and pharmacy purchasing data allowed assessment of the extent to which patient hospitalized for acute myocardial infarction received appropriate drug treatment after discharge.

Statistics Canada maintains vital statistics and survey data, and the potential exists for linking CIHI to Statistics Canada data using the national Health Card ID number. Health Canada collects health care utilization data on the populations to whom it is responsible for providing health care (First Nations, federal prisoners, the military and the Royal Canadian Mounted Police). Examples of organizations collecting data at the provincial level are the Manitoba Health Research Council, Health Quality Council of Saskatchewan, Health Quality Council of Alberta and the Institut de la Statistique du Québec. The health authorities of British Columbia and Ontario maintain sophisticated linked data systems which in some cases contain more detailed and extensive information than that held by CIHI. For example, while most provinces collect pharmacy purchasing data only for the population aged 65 and over, British Columbia collects this information for the entire population. Cancer registries are maintained at the provincial level. A system of provider identifiers exists, primarily for hospitals.
9 Access to national healthcare data systems - recommendations

Our review of national progress with, and good practice examples in establishing comprehensive systems of performance measurements and granting researchers access to patient oriented data for the study of the healthcare systems raises the following recommendations that should be addressed:

a. Proactive efforts should be invested in implementing comprehensive data systems rather than ad hoc efforts to respond to a specific research need (8, 11).

b. Sponsorship and governance—Development of a national, comprehensive linked data system requires support at high levels of government. Not only does such a project entail a considerable investment in manpower and information technology, it requires coordination and cooperation between the many data custodians in the private and public sector. Central funding of activities required for the development and maintenance of linked datasets will lessen the burden on organizations collecting data, thereby improving cooperation, and make the cost of research less prohibitive. The process of identifying “champions” and funding sources would have to occur separately in every country considering cooperation in multi-national research activities.

c. Harmonizing legal and privacy issues related to the use of person-level health care data for research purposes. This entails establishing legal authority for collecting, maintaining and sharing data systems, including determining situations in which informed consent is or is not required, and enacting minimum standards for safeguarding patient confidentiality.

d. Developing standardized data structures, coding systems and metadata

e. Developing sophisticated methods of data linkage between files from different sources which take into account the available identifiers.
f. Choosing a model for data consolidation at the national and international levels—a) central maintenance and processing of national data, b) distributed network, c) creation of linked de-identified files for research purposes

g. Establishing and maintaining an inventory of potentially useful databases

h. Extending access to the widest possible group of appropriate users, while implementing safeguards to ensure the privacy of subjects. Safeguards to privacy include strict review and data use agreements, limitations on the individuals who may access identified data, preparation of de-identified/encrypted datasets for outside researchers, prohibitions on publications of identified data by researchers.

i. Developing methods to limit access to personal data, such as "distributed health data networks" which allow remote analysis of data sets maintained by different organizations, offering the possibility of multi-center collaborations while obviating the need for sharing of sensitive data outside the custodial organization (25).

Many of the European countries already use the data generated in the course of delivery of and reimbursement for health care services. This data offer a valuable resource for health services and comparative effectiveness research. Looking into the near future the use of electronic health records (HER) will facilitate access to more comprehensive and accurate information required to identify exposures and outcomes of interest in large populations. In the OECD survey of 25 countries, that was presented recently, 22 countries have national plans to implement EHR that will include data for monitoring and research (42).

However, the use of personal health information will make the data protection barriers even harder to overcome and present a challenge to policy makers.
References


**EuroREACH WP3 Advisors meeting**

**May 23-25 – Tel Aviv, Israel**

**Monday May 23, 2011**
18:30-19:30 - Meeting kick-off

**Tuesday May 24, 2011**
9:00- 9:30 - Introductions
9:30-9:45 - The EuroREACH project—Frederique Hoffmann
9:45-10:00 - Work Package 3 and meeting goals—Nurit Friedman
10:00-13:30 - Building data systems for health services research
   - Oxford Record Linkage Project, English national record linkage, UK TECH collaboration—Michael Goldacre
   - Western Australia linkage project—Emma Fuller
   - US State all-payer databases—Patrick Miller
13:00- 14:00 - Lunch
14:00- 16:00 - Building data systems (cont.)
   - Canadian Institute for Health Information—Kira Leeb
   - FDA Mini-Sentinel—Richard Platt
16:00- 16:30 - Wrap-up discussion
19:00 - Dinner

**Wednesday May 25, 2011**
9:00- 10:30 - Linkage methods—Moderator, Emma Fuller
10:30- 12:00 - Standardizing coding systems—Moderator, Michael De Looper
12:00- 13:30 - Balancing access and privacy—Moderator, Patrick Miller
13:30- 14:30 - Lunch
14:30- 16:00 - Minimum requirements for component data sets and establishing standard reporting formats—Richard Platt
16:00- 16:30 Wrap up & next steps
EuroREACH  WP3- Participating experts

Michael Goldacre, UK, Oxford Record Linkage Project, English national record linkage

Emma Fuller, Australia, Western Australia linkage project

Patrick Miller, USA, US State all-payer databases

Kira Leeb, Canada, Canadian Institute for Health Information

Richard Platt, USA, FDA Mini-Sentinel

Michael de Looper, OECD, Harmonizing data for international comparisons

Carsten B. Pedersen, Denmark, National center for register based research
Appendix II

Figure 1: Decision tool for selection of relevant databases for inclusion in a national electronic inventory

### Figure 2: Criteria for assessing the coverage and accuracy of a clinical database

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Criteria for assessing the coverage and accuracy of a clinical database</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level 1</td>
</tr>
<tr>
<td>A. Extent to which the eligible population is representative of the country</td>
<td>No evidence or unlikely to be representative</td>
</tr>
<tr>
<td>Specify country:</td>
<td></td>
</tr>
<tr>
<td>B. Completeness of recruitment of eligible population</td>
<td>Few (&lt;80%) or unknown</td>
</tr>
<tr>
<td>State when and how completeness was determined:</td>
<td></td>
</tr>
<tr>
<td>C. Variables included in the database</td>
<td>• Identifier</td>
</tr>
<tr>
<td></td>
<td>• Admin info</td>
</tr>
<tr>
<td></td>
<td>• Condition</td>
</tr>
<tr>
<td></td>
<td>• Intervention</td>
</tr>
<tr>
<td>D. Completeness of data (percentage variables at least 95% complete)</td>
<td>Few (&lt;50%) or unknown</td>
</tr>
<tr>
<td>State when completeness was last determined:</td>
<td></td>
</tr>
<tr>
<td>E. Form in which continuous data (excluding dates) are collected (percentage collected as raw data)</td>
<td>Few (&lt;70%) or unknown</td>
</tr>
<tr>
<td>F. Use of explicit definitions for variables</td>
<td>None</td>
</tr>
<tr>
<td>G. Use of explicit rules for deciding how variables are recorded*</td>
<td>None</td>
</tr>
<tr>
<td>H. Reliability of coding of conditions and interventions</td>
<td>Not tested</td>
</tr>
<tr>
<td>State when and how it was most recently tested:</td>
<td></td>
</tr>
<tr>
<td>I. Independence of observations of primary outcome</td>
<td>Outcome not included or independence unknown</td>
</tr>
<tr>
<td>J. Extent to which data are validated</td>
<td>No validation</td>
</tr>
<tr>
<td>State when and how it was last determined:</td>
<td></td>
</tr>
</tbody>
</table>

*For example, timing of physiological measures or distinguishing primary from secondary diagnoses.

Appendix III - EuroREACH Database "Passport"

Please complete a separate form for each of the databases found to be eligible for inclusion in the EuroREACH inventory according to the "Decision tool" flow chart

1. Database Name: ________________________________________________________________
2. General description: ______________________________________________________________
   ________________________________________________________________
   ________________________________________________________________
3. Name of Sponsor/ Collector/ Custodian: ____________________________________________

   Is the data custodian (check all that apply):
   ☐ government entity   ☐ health care insurer/payer
   ☐ research facility   ☐ private company
   ☐ other (specify)________________________________________________________
4. Contact person (name, title, address, email, telephone, web site)________________________

   __________________________________________________________________________
5. What time period is covered by this database? ______________________________________
6. What population is included in this database (examples: children ages 0-18, adults ages 65 and older, persons with cancer diagnoses)? ____________________________
7. Do the data constitute a representative sample of the covered population, or a complete census?
   ☐ Representative sample
   ☐ Complete census
8. What is the approximate size of the covered population at present? ______
9. Approximately what percentage of the covered population is included in the database?_______
10. Unit of observation
   ☐ Person level
   ☐ Aggregated by (specify unit as: administrative/governmental unit, postal code, census tract, Other)
11. Data type / content
   a. Select the ONE data type category that best describes this database from the list below:
      ☐ Health care administrative data, including any of the following (complete section 9.b, below)
• Listing and demographic data for the covered population
• Physician/primary care visits/diagnoses/treatments
• Hospital admissions/diagnoses
• Medication purchases
• Laboratory
• Costs of care

☐ Electronic Health Record/clinical parameters (such as blood pressure, weight, height)
☐ Registry
☐ Survey (including perception on quality, quality–of-life)
☐ Public Health/Environmental Surveillance data

b. For health care administrative data, which of the following are included in the database?
☐ Listing and demographic data for the covered population
☐ Physician/primary care visits/diagnoses/treatments
☐ Hospital admissions/diagnoses
☐ Medication purchases
☐ Laboratory
☐ Costs of care

12. Which of the following best describes the availability of this dataset for research purposes:
☐ Not accessible to persons outside the custodial organization
☐ Accessible outside the custodial organization after approval of the proposed project
☐ Public use files with identifiers removed available to the public
☐ Other (describe)_________________________________________________________________

13. Linkage:
a. Can this database be linked to other to databases within the custodial organization using a common identifier or group of identifying variables?  ☐ Yes  ☐ No

b. Can this database be linked to databases maintained by other organizations using a common identifier or group of identifying variables?  ☐ Yes  ☐ No

c. Can the information in this database be linked at the aggregate level to other databases (for example, by postal code, governmental unit, census tract)?  ☐ Yes  ☐ No
Appendix IV - Narrative for each country's data capabilities

The purpose of narrative is to provide a picture of the availability of person-level health care data in your country, the ways in which they are currently used for management and/or research and the existing barriers to data access.

In your response please refer to the topics listed and consider the current situation as well as future plans (up to 5 years from now).

• Health care data availability

Are health care data collected routinely? Health care data include any of the following: inpatient/hospital, physician visits/community care, pharmacy purchasing, clinical parameters (blood pressure, weight, height, etc.), vital statistics, cancer or other chronic disease registries and health care costs. Are the data collected representative of the national population or of a significant portion of the population?

Are population or area level health-related data (such as representative survey data, socioeconomic parameters) routinely collected?

• Collection and stewardship of health care data

Which principal agencies or organizations (government/ hospitals/insurance companies/..) capture person-level health care data?

• Data linkage

Is it possible to link data from multiple sources (for example, hospital with primary care, pharmacy, laboratory, vital records) in order to create a complete picture of an individual's healthcare experience? Is there a unique identifier common to all data sources or are other methods of linkage possible?

• Use of health care data for quality monitoring, operations and research

Are the data collected used for monitoring quality/efficiency at the organizational, national or regional level? Are they used for International comparisons?

• Patient perspective on care quality

Does your country collect data on patient perceptions of care quality on an ongoing basis? Are there any data sources (for example, surveys) with information on responsiveness, patient centeredness, patient care ratings, adverse events, and care effectiveness?

• Barriers to accessing health care data for comparative research

Describe any barriers (legal/political/technical/other) that limit the potential for creating and/or accessing comprehensive data for comparative research.
### Appendix V

#### Table 1: Summary of country narratives

<table>
<thead>
<tr>
<th>Country</th>
<th>Health care data availability</th>
<th>Collection and stewardship of health care data</th>
<th>Data linkage</th>
<th>Use of health care data for quality monitoring, operations and research</th>
<th>Patient perspective on care quality</th>
<th>Barriers to accessing health care data for comparative research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>Primarily hospital utilization data</td>
<td><strong>Statistics Austria (<a href="http://www.statistik.at">www.statistik.at</a>)</strong>, collects person level health care data (e.g. on mortality); <strong>hospitals</strong> maintain the “minimum basic data set” (MBDS) for inpatient stays which includes patient information, diagnoses and treatments and provides the basis for reimbursement; <strong>Hauptverband (main association of social insurance funds)</strong>.</td>
<td>Pilot project (ELGA) in the federal states of Vienna (21st/22nd districts), Upper Austria (Wels-Grieskirchen), and the Tyrol (West), makes health data accessible to pharmacists and medical doctors, with linkage to medication information. Otherwise, Statistics Austria is able to link mortality, hospital routine data, and epidemiological data using several identifiers (e.g. postal codes, dates of birth and place of residence). There is no unique identifier for all data.</td>
<td>ICD10 codes recently introduced to hospital data collection, facilitating international comparisons. Expenditure data from the System of Health Accounts (SHA) is used internationally to compare health care expenditures. Health reports at federal, regional and even provincial level are published using primarily aggregated data. Health Austria Ltd, <a href="http://www.goeg.at">www.goeg.at</a> is a semi-public institution responsible for national and international exchange regarding health care reporting, and using the information for quality initiatives.</td>
<td><strong>Patient satisfaction survey 2010/2011</strong> carried out by the Federal Institute for quality of health care (BfIQG), Health Austria Ltd. -<strong>Critical Incident Reporting System</strong> (CIRSmedical.at) was launched in November allows medical and non-medical respondents to anonymously report adverse events.</td>
<td>Privacy concerns, legal barriers</td>
</tr>
<tr>
<td></td>
<td>Collection and stewardship of health care data</td>
<td>Data linkage</td>
<td>Use of health care data for quality monitoring, operations and research</td>
<td>Patient perspective on care quality</td>
<td>Barriers to accessing health care data for comparative research</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------------------</td>
<td>--------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>Wide range of person level data sets maintained at the national level</td>
<td>By national identifier (CPR number)</td>
<td>European Centre for Register-Based Health-Related Population Research – Public Health, Major Diseases and Welfare (ECREPH), <a href="http://www.ecreph.org">www.ecreph.org</a> University of Southern Denmark. promotes research based on registers, facilitating collaboration and offering training courses and education within register-based research</td>
<td>Use of data for research is accepted, subject to review procedures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estonia</td>
<td>The Estonian Health Insurance Fund administrative database includes billing data from 2000 onwards for the insured population (1.3 million, approximately 95% of the population) of the country.</td>
<td>The data base includes a national ID and can theoretically be linked to other registries and databases. To date linkage has not been carried out due to restrictive data protection regulations.</td>
<td>The use of data for these purposes has been initiated only recently.</td>
<td>Annual patient survey carried out on a representative sample of 2000 population to study satisfaction with and use of health care services provided (since 2001).</td>
<td>Data protection regulations, technical difficulties in managing inquiries from database of large datasets.</td>
<td></td>
</tr>
<tr>
<td>Health care data availability</td>
<td>Collection and stewardship of health care data</td>
<td>Data linkage</td>
<td>Use of health care data for quality monitoring, operations and research</td>
<td>Patient perspective on care quality</td>
<td>Barriers to accessing health care data for comparative research</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------------------</td>
<td>--------------</td>
<td>---------------------------------------------------------------</td>
<td>----------------------------------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td>England</td>
<td>Health care and health-related data are collected routinely.</td>
<td>National Health Service (NHS), <a href="http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937">http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937</a> General Practice Research Database (GPRD), <a href="http://www.gprd.com/academia/primarycare.asp">http://www.gprd.com/academia/primarycare.asp</a>. Clinical registries at the National and Regional levels.</td>
<td>For some databases linkages have been accomplished. There are unique identifiers in databases but they are not common across databases. Yes</td>
<td>Yes, the Picker Institute regularly collects information on these areas nationwide, but also at community levels.</td>
<td>Often access to data is costly, and requires ethical approval to access. Moreover accessing data can take months.</td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>All major health care providers (hospitals, primary care units, pharmacies) annually deliver person-level data collected and abstracted from their patient administration systems to government agencies who are by law the responsible register-keepers.</td>
<td>The National Institute for Welfare and Health (THL), The Social Insurance institution of Finland, Statistics Finland, Finnish Institute of Occupational Health.</td>
<td>By national identification code. Data linkage is allowed provided researcher has received permission from the authorities. Yes, the national data are widely used for quality evaluation, regional and operational planning and research.</td>
<td>Not collected at the national level. Some hospital districts and health centers periodically collect such data. National surveys sometimes include questions regarding general satisfaction with national health care.</td>
<td>Use of data for research is accepted, subject to review procedures.</td>
<td></td>
</tr>
<tr>
<td>Health care data availability</td>
<td>Collection and stewardship of health care data</td>
<td>Data linkage</td>
<td>Use of health care data for quality monitoring, operations and research</td>
<td>Patient perspective on care quality</td>
<td>Barriers to accessing health care data for comparative research</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------</td>
<td>-------------------------------------------------</td>
<td>-----------------------------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| **France**                  | The sickness funds, which cover over 90% of the population, collect information on the expenditure and procedures performed by health professionals but not diagnosis or health outcomes. | At the national level:  
- AMELI (Sickness Insurance) maintains the SNIIR-AM database, which holds de-identified information from all sickness funds (www.ameli.fr)  
- National Institute of Statistics and Economic Studies (Insee) holds data on Health Status, Health Service Utilization, Non-medical Determinants of Health, Quality of Life  
- PMSI hospital database holds patient level data from the acute healthcare facilities, patient level data for rehabilitative and psychiatric is available since 2009.  
National Institute of Health Surveillance | Possible at the postal code level, not at the individual level  
Since 2005, data for a representative sample of individuals in SNIIR-AM (medical consumption) are linked to ESPS (income, education, life style, etc.)  
Since 2011, this database is linked to PMSI hospital database (diagnostics, procedures, etc.)  
There is a unique patient identifier for these three databases. |  | Privacy and codes of confidentiality in the use of health information, lack of identified data at the national level, not all datasets use same identifier. |
| **Germany**                 | Individual sickness funds covering the population under the GKV (statutory health insurance) or their associations (the largest one being the AOK with around third of the population (24 million people)) have data linked on a patient level. These data are not compiled at the national level | Individual sickness funds | Within sickness funds, but not at the national level, due to the lack of a national patient identifier | Yes (GKV data) in form of secondary data analysis. | German Bureau of Statistics Microcensus  
Lack of a patient identifier to link multiple data sources; large number of sickness covering the insured population; split between private and compulsory health insurance funds excludes about 10% of the population. |
<table>
<thead>
<tr>
<th>Health care data availability</th>
<th>Collection and stewardship of health care data</th>
<th>Data linkage</th>
<th>Use of health care data for quality monitoring, operations and research</th>
<th>Patient perspective on care quality</th>
<th>Barriers to accessing health care data for comparative research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Israel</td>
<td>Four HMOs covering the population collect utilization and clinical data and report certain elements to the national government</td>
<td>At the national level, the Ministry of Health (<a href="http://www.health.gov.il">www.health.gov.il</a>), Israel Bureau of Statistics (<a href="http://www.cbs.gov.il">www.cbs.gov.il</a>), Ministry of Interior; HMOs maintain detailed utilization data for their member population</td>
<td>By national identifier (mispardet zehut)</td>
<td>At the national level, the Ministry of Health reports regularly on utilization of health services and causes of death. HMOs report aggregate data to the national quality measures project. Israel participates in the OECD quality measures project. National cancer registry published periodic reports and collaborates with other government agencies and outside organizations</td>
<td>Brookdale Institute conducts biennial surveys</td>
</tr>
<tr>
<td>Health care data availability</td>
<td>Collection and stewardship of health care data</td>
<td>Data linkage</td>
<td>Use of health care data for quality monitoring, operations and research</td>
<td>Patient perspective on care quality</td>
<td>Barriers to accessing health care data for comparative research</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------</td>
<td>--------------</td>
<td>-------------------------------------------------</td>
<td>-----------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>Person-level health care data exist for the entire population and population, except for EU employees (20,000 of a total 500,000 population). Area level health-related data (such as representative survey data, socioeconomic parameters) are routinely collected</td>
<td>CEPS/INSTEAD⁠¹ - conducts longitudinal surveys, specifically social surveys which have some health questions; Survey of Health, Ageing and Retirement in Europe (will begin in 2012); Luxembourg Bureau of Statistics - National Census last conducted 2011 for the entire population. Luxembourg Ministry of Health - CRP-Sante Public Research Center on Health - Conducts cross-sectional surveys for specific research topics; IGSS General Inspectorate of Social Security - Administrative data. IBBL Biobank of Luxembourg - in the planning stage, will collect genetic data</td>
<td>National ID number of administrative data IGSS is allowed to link demographic, employment and health care data, provided data protection measures are in place</td>
<td>The General Inspectorate of Social Security uses data for monitoring quality and efficiency of the health system in Luxembourg and for international comparisons (OECD and Eurostat).</td>
<td>Survey of patient satisfaction with long-term care</td>
</tr>
</tbody>
</table>

¹ Centre d’Etudes de Populations, de Pauvreté et de Politiques Socio Economiques/International Networks for Studies in Technology, Environment, Alternatives, Development
Appendix V

Table 2: Mapping the availability of person-level healthcare data sets and related data by country

(symbol key: ▲ =data source provides information on health status in the population, ▼ =provides information on non-health care determinants of health, ▼▼ = provides information on health system performance)

<table>
<thead>
<tr>
<th>Population register</th>
<th>Austria</th>
<th>Finland</th>
<th>Estonia</th>
<th>Denmark</th>
<th>France</th>
<th>Luxembourg</th>
<th>Germany</th>
<th>Israel</th>
<th>England</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics Austria maintains records of birth, marriages, causes of death, and mortality data</td>
<td>▲</td>
<td>The Finnish population register is hosted by the Ministry of Finance and contains basic information for all citizens (e.g. the unique and universal personal ID)</td>
<td></td>
<td>Danish Civil Registration system. Personal id (CPR number) enables linkage of spouses to one another, parents to children</td>
<td>Census data Insee (National Institute of Statistics and Economic studies)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health care utilization data (general)</th>
<th>Austria</th>
<th>Finland</th>
<th>Estonia</th>
<th>Denmark</th>
<th>France</th>
<th>Luxembourg</th>
<th>Germany</th>
<th>Israel</th>
<th>England</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social insurance-funded hospitals</td>
<td>▼▼</td>
<td>Collected by municipalities and reported to governmental authorities (THL) and Statistics Finland</td>
<td>Estonian Health Insurance Fund Database <a href="http://www.hajekassa.ee/eng/">www.hajekassa.ee/eng/</a> 2000-onward</td>
<td>Utilization databases maintained by Statistics Denmark</td>
<td>SNIIR-AM database,—de-identified data from all sickness funds. Includes Health Service Utilization, Reimbursement, Health care Expenditures,</td>
<td></td>
<td></td>
<td></td>
<td>General Practice Research Database, Oxford Data Linkage Study</td>
</tr>
<tr>
<td>Hospital</td>
<td>Austria</td>
<td>Finland</td>
<td>Estonia</td>
<td>Denmark</td>
<td>France</td>
<td>Luxembourg</td>
<td>Germany</td>
<td>Israel</td>
<td>England</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>--------</td>
<td>------------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>Minimum Basic data set (MBDS) – person-level data for overnight stays in hospitals funded by social insurance (date of birth, gender, nationality, main place of residence, date and type of discharge), main/additional (ICD 10 diagnoses and selected treatments)</td>
<td>National Discharge Register, National Institute for Welfare and Health (THL). Includes diagnoses and procedures</td>
<td>Diagnostic tests, investigations, surgical procedures and other treatments, Includes primary and secondary diagnoses (ICD-10), as assigned at the end of the treatment episode</td>
<td>Danish National Patient Register (NPR) includes services in Danish hospitals, including inpatient, outpatient, emergency room and psychiatric units</td>
<td>PMSI hospital database-- administrative data (patient level, diagnostics, procedures, etc.) from healthcare facilities, de-identified. It is possible to link patients across facilities. Since 2011, a sample of patients from SNIIR-AM is linked with PMSI and ESPS (income data, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician visits (diagnoses and procedures)</td>
<td>Yes, excluding private care. Includes diagnoses, not procedures</td>
<td>100% of visits to primary care (family physician), close to 95% of visits to specialists visits if reimbursed by the national insurance. Visits to private physicians excluded. Includes recorded diagnoses</td>
<td>The National Health Service Register for Primary Care (NHSPR) contains information about the activities providers contracted with the public healthcare system. Documents primary health care activities, minimal clinical data available</td>
<td>General Medical Observatory (GP visits <a href="http://www.omg.sfmg.org">www.omg.sfmg.org</a>) Small sample of doctors (not representative at the national level)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmacy purchasing</td>
<td>Reimbursed medications</td>
<td>INN name and ATC code, amount purchased and cost</td>
<td>Danish National Prescription Registry-- Individual-level data on all prescription drugs sold in Danish community since 1994</td>
<td>SNIR-AM : data on the volume and expenditure IMS Health (privately owned) Prescriptions by disease, providers, etc. (survey of representative sample of doctors)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory tests</td>
<td>Only in local databases (hospitals, health centers), not in national registers</td>
<td>Tests, without results</td>
<td>Some laboratory data included in NHSR</td>
<td>SNIR-AM (volume and expenditure)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registry</td>
<td>Austria</td>
<td>Finland</td>
<td>Estonia</td>
<td>Denmark</td>
<td>France</td>
<td>Luxembourg</td>
<td>Germany</td>
<td>England</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>--------</td>
<td>------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Births</td>
<td>National Cancer Registry (data available from 1983.)</td>
<td>The National Birth Registry maintained by THL</td>
<td>Danish Civil Registration system</td>
<td>INSEE, population census Linkable at postal code</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer registry</td>
<td>The National Cancer Registry includes all malignant</td>
<td>Contains records of all incidences of precancerous and benign lesions in the Danish population from 1943 onward. Reporting mandatory since 1987</td>
<td>There are no mandatory registries, but a number of</td>
<td>There are no mandatory cancer registries</td>
<td>Israel National Cancer Registry, Israel Ministry of Health</td>
<td>National Cancer Intelligence Network, Office of National Registries</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data uses key:

- Health status
- Non health care determinants of health
- Health care system performance
<table>
<thead>
<tr>
<th>Country</th>
<th>Communicable diseases registry</th>
<th>Census data</th>
<th>Clinical measures</th>
<th>Quality of life/patient satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>Maintained by</td>
<td>INSEE</td>
<td>Austrian Health Survey collects respondents</td>
<td>Austrian Health Survey -- representative survey (n=15474) for the Austrian population aged 15 years or over (6.9 million people), including but slightly underrepresenting people in old-age homes or residential care. National survey data collected periodically. No data on patient satisfaction, except two question in ESPS on visits to GP/specialists (did you understand the doctor and did you have enough time?)</td>
</tr>
<tr>
<td>Finland</td>
<td>THL hosts a database for a defined set of communicable disease. Reporting is mandatory for health care professionals</td>
<td>Health interview Surveys (HIS, Insee)</td>
<td>Health status</td>
<td>Some data on quality of life (irregular) in Barometre Santé (<a href="http://www.inpes.sante.fr">www.inpes.sante.fr</a>). No data on patient satisfaction, except two question in ESPS on visits to GP/specialists (did you understand the doctor and did you have enough time?)</td>
</tr>
<tr>
<td>Estonia</td>
<td>Declaration is obligatory for a number of diseases such as Botulism, , Tuberculosis, HIV/AIDS Legionella, etc. (mandatory declaration of each new diagnosis) at National Institute of Health Surveillance -- Ministry of Health</td>
<td>Micro-census, covering 1% of the population every 4 years</td>
<td>Health care system performance</td>
<td>European Union Survey on Income and Life conditions, CEPS/INSTEAD²; European Social Survey, CEPS/INSTEAD; European Value Survey, CEPS/INSTEAD and STATEC³; Panel Socio-Economique Liewen zu Letzeburg 2</td>
</tr>
<tr>
<td>Denmark</td>
<td></td>
<td></td>
<td></td>
<td>German Bureau of Statistics Microcensus is conducted every year and health related questions are asked once in 4 years</td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td></td>
<td>Biennial surveys, Brookdale Institute</td>
</tr>
<tr>
<td>Luxembourg</td>
<td></td>
<td></td>
<td></td>
<td>British Cohort Study, University of London, Institute of Education, Centre for Longitudinal Studies; British Household Panel Survey</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td>[2] Centre d’Études de Populations, de Pauvreté et de Politiques Socio Economiques/International Networks for Studies in Technology, Environment, Alternatives, Development</td>
</tr>
<tr>
<td>Israel</td>
<td></td>
<td></td>
<td></td>
<td>[3] Institut National de la Statistique et des Études Économiques du Grand-Duché du Luxembourg</td>
</tr>
<tr>
<td>England</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[2\] Centre d’Études de Populations, de Pauvreté et de Politiques Socio Economiques/International Networks for Studies in Technology, Environment, Alternatives, Development

\[3\] Institut National de la Statistique et des Études Économiques du Grand-Duché du Luxembourg
<table>
<thead>
<tr>
<th>Register database</th>
<th>Austria</th>
<th>Finland</th>
<th>Estonia</th>
<th>Denmark</th>
<th>France</th>
<th>Luxembourg</th>
<th>Germany</th>
<th>Israel</th>
<th>England</th>
</tr>
</thead>
</table>

Data uses key:

- ▲ Health status
- ▶ Health
- ○ Non health care determinants of health
- □ Health care system performance
Appendix VI - Data Linkage: References


24. Establishing a plan for standardizing all-payer claims data collection, meeting minutes, May 6, 2009.
   National Association of Health Data Organizations Regional All-payer Health Information Council.
33. Inventory of main differences between national health data protection systems in Europe and of bottlenecks experienced in daily practice in the context of Public Health Monitoring. Summary report by the Work Group on Data Protection and Confidentiality of the Network of Competent Authorities of the Health Information Strand of DG SANCO. [undated]. [Internet]


