ORIGINAL ARTICLE

The interface of population-based cancer registries and biobanks in etiological and clinical research – current and future perspectives

KRISTIN ANDERSSON1, FREDDIE BRAY2, MARC ARBYN3, HANS STORM4, ROBERTO ZANETTI3, GÖRAN HALLMANS6, JAN W. COEBERGH7 & JOAKIM DILLNER1,8

1Laboratory Medicine, Lund University, Malmö, Sweden, 2Cancer Registry of Norway, Oslo, Norway, 3Unit of Cancer Epidemiology, Scientific Institute of Public Health, Brussels, Belgium, 4Cancer Prevention and Documentation, Danish Cancer Society, Copenhagen, Denmark, 5Piedmont Cancer Registry – CPO, Torino, Italy, 6Umeå University, Umeå, Sweden, 7Comprehensive Cancer Centre South, Eindhoven, Netherlands and 8Karolinska Institutet, Stockholm, Sweden.

Abstract

Background. The availability of quality assured, population-based cancer registries and biobanks with high quality samples makes it possible to conduct research on large samples sets with long follow-up within a reasonable time frame. Defined quality for both cancer registries and biobanks is essential for enabling high quality biobank-based research. Recent networking projects have brought these infrastructures together to promote the combined use of cancer registries and biobanks in cancer research. Materials and methods. In this report we review the current status and future perspectives of cancer registries and biobanks and how the interface between them should be developed to optimally further cancer research. Results and discussion. Major conclusions for future improvements are that the research exploiting cancer registries and biobanks, and the research that is building and optimising the infrastructure, should evolve together for maximally relevant progress. Population-based and sustainable biobanks that continuously and consecutively store all samples (“Biological registries”) under strict quality control are needed. There is also a need for increased education, information and visibility of the interdisciplinary sciences required for optimal exploitation of these resources.

This report is the result of an international workshop held in Oslo, Norway on May 28, 2009 jointly organised by two European Union networks that both have as major goals the optimisation of the essential cancer research infrastructure that is created by the interaction of cancer registries and biobanks. These networks are the EU FP6 Network of excellence ‘Cancer Control using Population-based Registries and Biobanks’ (CCPRB) and the EU FP7 Era-Net ‘Optimisation of the Use of Registries for Scientific Excellence in Research’ (EUROCOURSE).

The CCPRB project started as a collaboration between cancer registries and biobanks in the Nordic countries that has expanded to include partners in other Central and Southern European Countries. The aims include the strengthening of the collaboration between cancer registries and biobanks from different countries and to improve the work on quality assurance with respect to the current infrastructure and the documentation of its use. The sharing of knowledge and experiences, and conducting collaborative research on principles of cancer control and quality assurance has been important activities. The EUROCOURSE project, which began in April 2009, will tackle and describe the problems and shortcomings in various aspects that ultimately affect cancer research and cancer control including registry practices, privacy protection, data safety, aspects of comparability, quality and timeliness, and also aims to expose inefficiency and diversity in methods for quality control. EUROCOURSE will also work on strengthening best practices and identifying bottle necks in the areas of linkages with other datasets, research staff and funding, clinical evaluation and frameworks for discussion as well as patient contact.
High quality research in the field of cancer control needs an availability of data linked to individuals, be it of biological nature or other personal data. Both biological and other data should be of high quality, correct, comparable and usable in a broader context across borders.

Both in order to be able to compare cancer control strategies and to be able to study rare disease, it is necessary to perform the cancer research as international collaborations. The possibility to perform international studies on cancer prevention, treatment, care and outcome can improve the prospects to avoid development of new cancer cases and to improve the life of cancer patients.

The purposes of the meeting were to a) build a common view on the usefulness of biological registries (biobanks) and how integration of biobanks with cancer registries could improve cancer control, b) summarise the progress of CCPRB during 2004–2009, c) prepare the work plan for EUROCOURSE on biobank-cancer registry interfaces, d) review current success and shortcomings in biobank-registry integration and identify areas where further work is needed to establish European policies aimed at strengthening the EU Research Arena and e) use the presentations and ensuing discussions as a basis for a public report outlining recommended actions to be taken – a key deliverable of the EUROCOURSE project. This paper is the public report corresponding to the latter of the meeting purposes and the major issues and recommendations are listed in Table I.

At the meeting there were representatives from both national cancer registries (e.g. in Norway, Finland, Iceland, Denmark and Belgium), regional cancer registries (e.g. in The Netherlands and Italy), clinical cancer registries (e.g. the rectal, breast and lung cancer clinical registries in Norway), several large-scale biobanks (e.g. the Janus Biobank, the HUNT cohort, the Malmö Biobank Consortium and the Northern Sweden Health and Disease Study), as well as scientists performing cutting-edge research using registry linkages and biobank samples.

**Current status**

The information presented at the meeting displayed a wide range of activities and achievements in both FP5, FP6 and FP7 networks. They covered many different areas involving cancer registries, biobanks and their interaction. The networks have been important in achieving progress and the main areas presented at the meeting are described below.

**Data quality in cancer registries**

The establishment and continued re-assessment of internationally-standardised methods for the evaluation of data quality in cancer registries is of great importance in securing high levels of data quality and comparability. Currently, data quality can be measured and analysed in a number of ways using different criteria. This can uncover different aspects of the relative quality of registry data. But there remains a need to ensure quality indicators are comparable between registries, wherever possible. Four key areas regarding data quality in cancer registries, that have been recommended and used for decades and that are widely accepted and used by registries as tools for their quality control, latest reviewed by Bray and Parkin 2009, are a) comparability, b) validity (accuracy), c) completeness and d) timeliness [1–3].

b) Validity or accuracy is defined as the proportion of cases in a data-set with a given characteristic (e.g. site or age) which truly have that attribute [1,2]. Validity is an important marker of data quality in the international compilation of data from national and regional registries (Cancer Incidence in Five Continents) published by the International Agency for Research on Cancer (IARC) [3]. There are multiple information sources to a well-functioning registry, and an over-reliance on one source or the lack of availability of another can affect the accuracy of the information, as well as its completeness (see below). The information can to some extent be verified if it is retrieved from more than one source. Common methods used for the evaluation involve reabstracting and recoding, methods involving comparisons of histological verification, and proportions of registrations based on death certificates only or missing information [1].

c) Completeness is defined as the extent to which all incident cancers occurring in the catchment population are included in the registry database [2]. There are both quantitative and semi-quantitative methods in its assessment, and completeness from Norway was considered very high, using a combination of these techniques in the evaluation [4].

d) Timeliness is evaluated in terms of the time from diagnosis to registration, and the time from registration to the reporting of incidence via, for instance, the publishing of the annual report [1]. Timeliness has become a bigger priority in recent times as the realisation of timely and accurate data for different purposes has become a bigger issue. In Norway, actions were taken to improve the timeliness, demonstrated by the less-than-one-year turnaround...
Table I. Summary of recommendations for prioritised actions.

<table>
<thead>
<tr>
<th>Issue</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical registries</td>
<td>Development of new clinical-site specific registries Improvement of the existing clinical registries Systematic biobanking of cancer tissue Collaborate with interest organisations and clinical specialists Online registration, expansion of databases, methods of quality assurance Use clinical registries as a basis for building such infrastructures</td>
</tr>
<tr>
<td>Biobanks and data quality</td>
<td>Continuously update data on prospective cases Predicting future number of cases Linkage of biobanks to registries Implement collaborative predictions of future cases across relevant cancers as well as biobanks Evolve together and learn from each other by continued networking Stop discarding samples, use quality assurance appropriate for also for research uses and build capacity in data management and registry linkages</td>
</tr>
<tr>
<td>Ethics and biobanks</td>
<td>Improve contact with, and information flow to, patients Patient-oriented websites, involvement of patient organisations, involve other stakeholders, be clear about policies and usefulness</td>
</tr>
<tr>
<td>Interface of Cancer registries and biobanks</td>
<td>Develop convergence &amp; standardisation Shared resources such as joint electronic quality manuals, standard operating procedures, information management systems. Active networking to share experiences and joint education. Explain why population-based and sustainable biobanks are needed Increasing the interface between cancer registries and biobanks with the public, politicians and others concerned with the work.</td>
</tr>
</tbody>
</table>

of close of year of registration to publication of the annual report, *Cancer in Norway* [4].

Completeness and timeliness of the presented national registries’ data is attributable to the legal premises (mandatory to report cancer and certain pre-cancers) and that data is reported from multiple (clinical, pathological, death certificate, autopsy) sources.

Besides the aspect of the quality of the data of cancer registries, their quality as a whole as a surveillance system is also considered by the Eurocourse project: this implies extending the coverage to uncovered EU countries/areas, and defining the optimal size for regional registries.

**Clinical registries**

As important as the national cancer registries are, the development of clinical, disease-specific registries has also been proven to be a winning concept. In Norway, several clinical registries have been created, starting with the national colorectal cancer project in 1986, on the initiative of patient organisations, clinicians and the national cancer registry because many pathologists and clinicians found that the data collected from the cancer registry was insufficient for all of their specific needs with regards to quality control clinical research, and the population-based evaluation of clinical care. The outcome of the research conducted based on the rectal cancer registry has led to improved diagnostic methods as well as better treatment with fewer recurrences and longer survival for the patient [5–7]. Similar clinical registries have been developed in many countries, for example for colorectal cancer [8,9] in Denmark and Sweden, prostate cancer [10] in Sweden and brain tumours in Austria [11].

**Population-based clinical biobanks and research biobanks**

Large-scale biobanks have been in operation for many decades. There are large clinical biobanks with formal organisations containing well-established policies and procedures for the clinical use of samples and which may also be used for medical research, for example the biobanks with formalin-fixed tissue which exist at all departments of clinical pathology. There are also many large research biobanks, which are collected exclusively for research purposes. Because the need for high-quality specimens for research has increased both for the clinical and the research biobanks, the need for standards to guide
the proper collection, processing, storage and dissemination of the specimens has continuously increased.

Both clinical and research biobanks may be population-based. An example of a clinical population-based biobank is the cervical cytology biobanks, which is based on the population-based cervical cancer screening programs and in many countries contain multiple cellular specimens from almost the entire adult female population. Cervical cytology biobanks that use standardised biobanking procedures and are integrated in a national or regional Screening Registry with possibility to link to other registries are highly useful for cancer research in general [12].

The large-scale research biobanks that have participated in the CCPRB network (e.g. the Janus Biobank, the Malmö Biobank Consortium and the Northern Sweden Health and Disease Study) are good examples of research biobanks which have existed for a long period of time. To a large extent these research biobanks have enrolled using population registries and have had sufficiently high attendance rates to be considered as being population based. These characteristics resemble those of the Italian cohort ORDET established at the beginning of the nineties [13] as well as some of the biobanks in the European EPIC consortium [14].

**Biobanks and data quality**

The work on quality assurance in biobanks has traditionally focused on the biological material rather than the underlying data. A major task for the Networks has been to create guidelines for quality assurance regarding both samples and data, to implement them in the daily work of the biobanks within the Network, and to establish dissemination routes [15].

The FP5 TuBaFrost project was a well-known example of a successful multidisciplinary approach, both to develop routines and methods for biobank activities as well as making them available for researchers outside the consortia. TuBaFrost has published bottom-up work on how to standardise tissue collection and quality control, how to unite local frozen tumour tissue banks into a large-scale Network, as well as how to create a central database [16–18], the access rules and incentives for such biobanks [19] and the use of virtual microscopy [20].

The scientific publication of standard operating procedures, that have been created and validated within an international collaboration, has resulted in their use in several laboratories outside the TuBaFrost [21]. Actions in the direction of creating an ISO certification/accreditation for biobanks has also been taken by experts involved in the daily quality work of biobanks [22].

Quality assurance tools for databases on prospective cancer cases have been created with the main objective to identify errors. Areas were such errors can occur are accuracy of personal identification numbers, accuracy of endpoint status, accuracy of data linkages, completeness of follow-up for vital status and population representativeness. Today, our understanding of these factors and their relative quality are both good. The linkage work behind a publication with updated information on occurring cancer cases in the participating Nordic serum banks has been a very useful tool for identifying errors in the participating biobanks [23]. For example only a few cases with incorrect identification can cause large errors in studies, since those individuals will seemingly never get ill or die if the incorrect identifier is not found and corrected.

A system for quality assurance must cover all activities in the biobank. Much work has been put into the development of a ‘Good Biobanking Practice’ (GBP) system which documents the important principles on how to run biobanks. GBP covers seven major areas: 1) Quality assurance, 2) Expert evaluations of research proposals, 3) Transfer of samples, data and results, 4) Data sharing policy, 5) Steering, 6) Legal issues and 7) Guidelines for evaluating biobank value and quality. The guidelines have been developed as collaboration between the biobanks and the scientists using biobanks for cutting-edge research. An important rule for the quality work process is never to make the same mistake again in a biobank [24].

After implementing quality assurance work in a biobank and solving the problems found, it is important to be able to evaluate the results. The practical use of the biobank material for real research is the most natural, relevant and effective method of quality control, revealing the most important weaknesses that requires attention. To be able to document the usefulness of a biobank for society and for the scientific community, methods for evaluation of the quality and value of the biobank are needed. Quality Indices that have been proposed include calculating a Biobank Impact Factor (BIF) as well as a Biobank Impact factor for Public Health (BIPH) [24]. BIPH is based on estimating the impact the research performed using the biobank has for the general population. The BIPH system evaluates biobanks based on their purpose (scientific output). As biobanks need to accumulate prospectively occurring cases of disease before that are really useful, BIPH favours established biobanks that are extensively used in research, whereas recently-established biobanks that have not yet been used for research have, as yet, little impact and can not be compared.
Clinical biobanks are an immense resource for research, particularly when linked to cancer registries and other data sources, but their usefulness could be greatly expanded if the clinical biobanks were systematically optimised for ability to also be used as dedicated research infrastructures. The main measures that need to be taken are: a) to ensure that older clinical samples are not discarded. Usefulness for research is in general increasing with increasing storage time, whereas the opposite is true for usefulness for the patient’s own clinical diagnosis, b) to ensure that the quality assurance system used also covers the biobank and its scientific use, in particular ensuring that the personal identifiers are correct, adequately stored and readily linkable to cancer registries. In this process, an experienced quality assurance officer with multidisciplinary training is invaluable and c) build competence in registry linkages and/or build collaborative networks with institutions experienced in registry management, such as cancer registries.

Ethics and biobanks

Researchers in science and ethics have published together in a number of prominent papers that have researched the scientific, ethical and legal basis behind essential practises such as approval from ethical review boards and preferred procedures of consent and feedback of study results to sample providers [16, 25, 26]. Key concepts include comprehensive use of broad informed consent, respect for integrity and ability to opt out, defined conditions and permissions (data inspection etc) for being able to store and link data, increased emphasis on the need for researchers to explain research in a way everyone can comprehend and the obligation to demonstrate the relevance and benefit of the research. The opinion that there is no need for further regulations and that common principles and nomenclature are already in place is starting to gain widespread acceptance [27–29].

Interface of cancer registries and biobanks

A major goal with the EU networks, besides quality improvement, was to enhance and encourage the usage of cancer registries and biobanks in research projects. From the early phases of implementing a research project, the basic characteristics of the biobank and registry data (including the number of specific cases available) and their quality needs to be known [15]. Ways of spreading such knowledge for biobanks include the submission of databases for joint analysis and standardised description of cohorts, improvement of the technical accuracy of databases, estimation of population representativeness, and estimation and prediction of the number of expected cases of specific cancers. Results from such efforts have been published [23], resulting in, for example, rapid answers as to how many cases a particular biobank can contribute to a given study. This publication has been found to be highly useful and it is highly recommended that similar efforts should be undertaken in the future.

Future perspectives

There are two categories of science in this area that mutually depend on each other: ‘The Science of Registries and Biobanks’ is the science behind setting up, optimising and managing a registry or a biobank and ‘The Registry and Biobank-based Science’ which is the research based on information and samples held by the registry or biobank. A major lesson learned from the EU networks on registries and biobanks is that research exploiting the infrastructure and research building and optimising the infrastructure should evolve together: the building of infrastructure benefits from input from innovative users, while the exploitation of the infrastructure is suboptimal unless there is close contact with the scientists behind the building of the infrastructure.

It is important that biobanks are not be considered a static activity, but as a dynamic and scientific activity. Biobanking needs to continuously evolve according to the ongoing development of new techniques and new scientific goals, as pointed out by an international expert group in biobanking [30].

Data quality in cancer registries

As with all quality assurance, the evaluation of data quality in cancer registries is a continuous process. The issues of comparability, validity, completeness and timeliness require regular assessment in each registry. This can be time-consuming, but is very important for the future usefulness of the registries in cancer control and cancer research. Further standardisation of existing methods for quality control, as well as the promotion of their use, are high priority areas of activity, as is the further development and evaluation of novel methods for quality assurance of registry data.

Clinical registries

The development of cancer-site-specific clinical registries, as a complement to the population-based registries, has proven to be of great benefit for evidence-based patient care, with evidence of improved survival and better care as a direct result [5–7, 31]. Additional clinical registries should be
developed and the existing clinical registries further improved by the likes of online registration and regular reconsideration and expansion of the datasets, where appropriate. Clinical registries are capable of providing front-line research on the biology of the neoplastic diseases, by combining high quality, and clinically-relevant data, with a high degree of research competence in epidemiology, biostatistics and molecular biology.

The possibility to use the Clinical registries as a basis for systematic biobanking of cancer tissue has only been partially explored, but would fit very well with the overall concept that the building of infrastructure needs close contact with the professionals that are most dedicated and competent in envisaging their most important uses.

Population-based clinical biobanks and link to research biobanks

The cervical cytology biobanks, when linked to a comprehensive cervical cancer screening registry and a cancer registry, constitute a huge resource for public health related and fundamental research such as auditing the effectiveness of screening, surveillance of HPV vaccination effects, evaluation of new biomarkers, study of molecular pathways of carcinogenesis reducing the dependency on expensive and long-lasting prospective studies [32–34].

Research on the cytology and pathology biobanks is to an increasing extent combined with projects performed within the research biobanks, in particular in cancer research. This parallel utilisation of the two forms of biobanks to obtain as many informative samples and data as possible is expected to increase in the future. New technologies like the various forms of “omics” technologies are expected to greatly increase to power of all biobanks, especially when the biobanks are combined to enable life-course studies with multiple samples taken during the life-time of study subjects.

Biobanks and data quality

The workshop found that there has been substantial progress in this area. In the future it will be important to maintain the quality work as well as to continue to provide new instruments to identify errors in case-control studies, to continuously update the data on number of prospective cases in the biobanks, and to predict future numbers of cancer cases. As time is a critical factor in research, it will be crucial to avoid spending extra time on separate linkages for every new study, commonly only on one cancer site, when a single linkage can, with a similar effort, be done across all relevant cancer sites, for many different biobanks, as a basis for many different studies in a variety of study designs.

The quality systems included in ‘Good Biobank Practice’ (GBP), should be under constant evaluation and improvement in order to ensure that they will always be up-to-date. New lessons are learned continuously, and as new molecular technologies are developed, so the utility and demands on biobank samples change. Tools such as internet-based Quality Manual Systems may be helpful in facilitating the sharing of experiences in this respect.

Quality needs to consider the entire process of the science of biobanking, from ensuring that optimally useful samples are collected and stored to ensuring that the samples and data are put to best possible use for the benefit of cancer patients. The ‘Biobank Impact factor for Public Health’ (BIPH) biobank usefulness evaluation that includes an evaluation of the scientific, published output of biobanks is an important concept in this regard.

Ethics and biobanks

The workshop found several important areas where immediate actions are needed to promote harmonisation of the ethical framework regarding cancer registration and biobanking. For example the sample donors’ expectations, the ideas of ownership, the issue of benefit-sharing and worries about privacy are areas that need particular attention. The goal of good research governance should be to achieve a fair and transparent balance between the interests of stakeholders, a regard for the public benefit, to manage expectations from different groups, and to conduct research that is in the interest of both patients and healthy sample donors. Different ways to achieve this should be considered, for example via the creation of patient-oriented websites describing research projects, involving patient organisations and other stakeholders. Policies need to be clearly documented including how a project is approved, how privacy is protected, how data is shared, and how results are disseminated. The main recommendation can be summarised as to let the public know how the projects using registries and biobanks will benefit the public good.

Interface of cancer registries and biobanks

Although major progress has been done in the work of collaboration between cancer registries and biobanks there are issues that require further evaluation and improvement. For example, most steps in the biobank-based research process will benefit from interaction with experienced cancer registry scientists and this should be promoted. Our
recommendation that a multidisciplinary team of experts is an invaluable resource when building up and using a biobank is in line with previous recommendations [35].

For clinical population-based biobanks (e.g. cervical screening archives and clinical serology biobanks) there are often no personnel with experience of data cleaning and data management, resulting in impaired accessibility and sometimes even the simplest linkages cannot be managed. This problem can easily be solved by collaborations where the databases are cleaned and registry linkages are performed at experienced centres, such as cancer registries or major research biobanks with data management experience, which is a model explored by CCPRB. The workshop made several recommendations for future work in this area. Continued networking will be important for the sharing of experiences (to avoid making the same mistake twice) and in guaranteeing sufficient interdisciplinary competence via the interaction of scientists in several disciplines. It will also be the most effective way to achieve a process towards convergence and standardisation that can be used in practice. Shared resources, such as shared quality manual systems and information management systems, could facilitate networking and harmonisation and convergence in practical terms. The joint education and the availability of educational material across the biobanking sciences will also be essential.

An important task for the future will be to explain the issues as to why population-based and sustainable biobanks that continuously store all samples are needed and to ensure that the synergies are obtained by increasing the interface between cancer registries and biobanks with the public, politicians and others concerned with the work of cancer registries and biobanks. Finally, there is an overwhelming need to further promote biobanking sciences as an advanced interdisciplinary science, recognising its key strategic importance in progress in cancer control via the utilisation of cutting-edge cancer research.

Acknowledgements

Supported by the EU FP-6-2002-LifeSciHealth contract Cancer Control using Population-bases Registries and Biobanks (CCPRB) and the EU FP-7-ERA-NET contract Optimisation of the Use of Registries for Scientific Excellence in Research (EUROCOURSE). The authors would like to thank all the speakers at the workshop: Max Parkin, Laufey Tryggvadóttir, Bjorn Møller, Eero Puikkala, Froydis Langmark, Trond-Eirik Strand, Rolf Kåresen, Kristian Hveem, Joyce Carlson, Kari Hemminki, Matti Lehtinen, Mats G. Hansson and Evert-Ben van Veen. We are also grateful to Hilde Langseth and Matti Hakama for useful comments on the manuscript.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References


