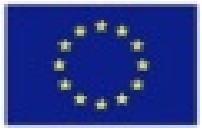


Directorate-General for Health & Consumers

Executive Agency for Health and Consumers

EUBIROD

EUropean Best Information
through Regional Outcomes in Diabetes

WP6: EPIDEMIOLOGICAL ANALYSIS

D6.1: Statistical Materials

August 2010

This report is Deliverable D6.1 of "WP6 Epidemiological Analysis", the European project "European Best Information through Regional Outcomes in Diabetes" (EUBIROD), co-funded by DG-SANCO, European Commission, 2008 (G.A. 2007115)

***Scientific Coordinator**: Prof. Massimo Massi Benedetti*
***Technical Coordinator**: Fabrizio Carinci*

A joint production of the EUBIROD Consortium:

Adelaide and Meath Hospital, Dublin, Ireland
Centre Hospitalier de Luxembourg, Luxemburg
Dutch Institute for Healthcare Improvement, Nertherlands
Havelhöhe, Berlin
Hillerød University Hospital, Hillerød, Denmark
IMABIS Foundation, Malaga, Spain
International Diabetes Federation, Belgium
Inst. Scient. Santé Pub. WIV, Brussels, Belgium
Joanneum Research, Austria
Medical University of Silesia, Katowice, Poland
Ministry of Health, Cyprus
NOKLUS, Norway
Paulescu Institute, Romania
Sahlgrenska Academy, Gothenburg, Sweden
Serectrix snc, Italy
University of Dundee, Scotland
University of Malta, Malta
University of Perugia, Italy
University of Debrecen, Debrecen, Hungary
University Children's Hospital, Ljubljana, Slovenia
Vuk Vrhovac University Clinic for Diabetes, Zagreb, Croatia

***WP Leader:** Serectrix snc, Italy*
***Task Leader:** Serectrix snc, Italy*

Compiled for Serectrix snc by:
Fabrizio Carinci, Senior Statistician, Serectrix snc, ITALY
Luca Rossi, Statistician, University of Perugia, ITALY

The Statistical Engine Core Development Team:
Fabrizio Carinci, Senior Statistician, Serectrix snc, ITALY
Luca Rossi, Statistician, University of Perugia, ITALY

Citation
Fabrizio Carinci, Luca Rossi on behalf of the EUBIROD Consortium, Statistical Materials, EUBIROD Consortium, 2010

Project-Website
http://www.eubirod.eu

Table of Contents

Executive Summary

The EUBIROD project aims to implement a sustainable European Diabetes Register through the coordination of existing national/regional frameworks and the systematic use of the BIRO technology. The project runs between 2008-2011, is co-funded by DG-SANCO, European Commission, and coordinated by the University of Perugia, Italy.

The project is based on the usage of the BIRO System, a tool specifically built to share an "Evidence-Based Diabetes Information System" among seven European countries. The system, developed between 2005-2009, has a structured architecture that involves two data processing steps, corresponding to a local and a global component, linked by a uni-directional flow of information.

The EUBIROD Statistical Materials aim at documenting the construction of the statistical routines required by the BIRO System to deliver European Diabetes Reports on a range of key indicators identified by the Consortium, according to the outputs specified by a standardized report template.

The "Epidemiological Analysis" was planned as a specific work-package of the EUBIROD project to deliver the following:

- a complete set of statistical routines for the automatic delivery of the EU Diabetes report, directly connected to the BIRO system adopted by the EUBIROD Consortium
- development of risk adjustment models to include standardized estimates of diabetes indicators in all EUBIROD reports
- production of local statistical reports, based on the application of a common template across all centres
- production of aggregate tables to be sent to the central EUBIROD server
- production of the European Diabetes Report using the ordered sets of aggregate data sent by EUBIROD users
- production of graphical outputs and user friendly layouts to facilitate the interpretation of all results

Deliverable D6.1 is a report that includes all software specifications, a description of the various steps required to run the software, and all the source code published both in pdf and electronic package.

During the first year, the activity concentrated on the revision of the BIRO statistical engine to make it functional for the immediate use of all partners of the EUBIROD Consortium. An initial release of the software was made available at the 1st EUBIROD Annual Meeting in May 2009.

In the second part of the year, new functionalities were continuously added to the original version. Statistical procedures were updated to include standardization algorithms capable of delivering tables of observed minus expected in risk adjusted BIRO outcome indicators. The underlying statistical model, based on logistic regression, performs standardization using age bands and sex as covariates, implementing the method used by the US Agency AHRQ.

The activity continued with increased efforts during the second year, initially concentrating its attention on the development of the statistical engine to satisfy all requirements of the BIRO report template and to deliver the complete list of updated indicators.

A new release of the software was deployed to all partners in November 2009. The majority of users succeeded in delivering local reports. More improvements were planned and included in a master plan duly listing all priorities for the final production of the software.

A new release of the software was deployed in May 2010. At the Special BIRO Academy Meeting, a total of N=14 reports were independently produced by all partners were presented. The results were extremely satisfactory, although more improvements were requested in the following areas:

- role of missing values in the calculation of all indicators
- cross-tabulation of stratified indicators
- presentation of standardized values
- graphical displays
- log files
- structure of the pdf report
- functioning of the central engine
- usability under BIROX
- lack of explanations in the presentation of results

Continuous development of the statistical and central engine occurred during the following four months. Support for the interpretation of results and to realize further improvements was offered by the University of Dundee, based on their long standing experience in the production of reports for clinical specialists and health policy.

The release of the software was finalized in August 2010.

The present deliverable reports the advancements made by the EUBIROD statistical materials to foster the epidemiological analysis of diabetes data, through the following features:

- multidimensional tables, stratified by two exposure factors and one outcome variable. This feature allows fine comparisons e.g. calculation of relative risks of the outcome across different levels of exposure factors.
- revised structure of the BIRO report allowing stratification of results by centre at the local level. This feature delivers finely stratified reports in which all indicators are displayed by sub source (clinical centre or unit) within each local register (ex.: Austria can benchmark differences in diabetes indicators between centres in the region of Styria).
- for each parameter/indicator, a root table displays the frequencies of missing vs valid values
- stratification of all results by a class variable (Type of Diabetes)
- graphical displays of all stratification levels
- unique coding structure for all outputs delivered as html tables, image files, and CSV data. This feature allows to easily reuse all objects for presentations, to dynamically link results to the BIRO web portal, or to feed other online repositories (e.g. the DG-SANCO health information platform “HEIDI”)
- revised pdf report including cover pages providing information on the EUBIROD Consortium and explanatory figures as help files
- same outputs realized for the statistical engine applied to the central engine
- recursive application of the central engine. This feature enables each user to pool statistical objects obtained from both the analysis of individual and aggregate data (ex.: different centres of Germany can deliver reports to an institution acting as national coordinator. Such entity can compile the national report using the central engine, then send the results to the Coordinating Centre, which can use them again using the central engine to produce the European Report).

- storage of all outputs in a directory selected to the user, compliant with the BIROX distribution.

The statistical routines are made accessible to the average BIRO user through the enhanced version of the BIROBox interface.

A final development phase is required to define strict rules that would assign a unique code to each centre adopting BIRO. This way the BIRO system may be applied in a recursive fashion, expanding the range of its applications at the international level. In fact, the software would be able to recognize the list of sources involved in the calculation, attributing each indicator to a very well defined set of contributors. Such feature, although not initially foreseen, resulted from of an assessment of the practical conditions existing on field, in situations where the direct processing of all individual data can be undertaken by different istitutions, but aggregate data cannot be sent separately to the EUBIROD central server. Therefore, aggregate tables must be amalgamated by one or more national coordinators prior to any transmission, which means that the central engine shall be made available to each BIRO user rather than only to the EU server administrator.

A further, unplanned update of the statistical materials including such improvements is foreseen prior to the end of the EUBIROD project.

In conclusion, the set of routines realized for the EUBIROD project provide a flexible solution to set the basis for continuous monitoring of diabetes across Europe. The statistical reports include basic figures that can be helpful to benchmark quality and outcomes, but can also significantly enhance the average capacity of all centres to increase the quality of their information, and to share more standardized information.

The availability of such targeted set of statistical routines as open source is particularly relevant for those users e.g. diabetes register administrators, who maintain large databases but until now have neither exchanged data with international peers, nor used common tools for epidemiological analysis. The range of outputs delivered by the EUBIROD statistical engines may be exploited to build flexible EU platforms that would automatically tap into regional/national databases to gather and immediately deliver public health information in a standardized format.

1. Introduction

Despite a wide political recognition of the importance of diabetes as a high priority public health issue that needs to be carefully monitored and duly reported, accurate statistical information in this field is still lacking.

Standardized definitions and detailed lists of core diabetes indicators have been finely specified during the last decade. Major international projects, including the EU-funded ECHI, EUDIP, EUCID and the OECD-coordinated Health Care Quality Indicators Project, were followed by efforts to allow the routine calculation of such indicators through extensive data collection. Until today, no initiative has yet succeed to deliver complete information on diabetes on a regular basis.

Between 2007-2009, the European Commission published two major reports involving public health experts in the presentation of the state of the art in different disease areas at high priority: the EU report on "Major and chronic diseases" (MCD)[1] and the "Status of Health in the European Union" (EUGLOREH)[2].

In both cases, diabetes was considered as a key area deserving full chapters, but showing that complete information was hard to find, and even when available, it was largely inconsistent across different data sources.

The MCD report declared that "*European networks of excellence in this field collect extensive data as a by-product of clinical activity and systematic linkage of administrative data...Although diabetes represents almost an ideal model to investigate chronic diseases – as demonstrated by an overwhelming number of epidemiological studies – to report on its state at the population level still represents a major challenge with no obvious solution European-wide*". In the final conclusions, the chapter stated that "*Paradoxically, key indicators that are crucially needed to plan diabetes care, like prevalence of impaired fasting glucose and death with diabetes as primary or secondary cause are still inconsistently available at the moment. Identifying solutions to make all key indicators available at all levels can be highly effective to reduce the burden of diabetes both in economical and clinical terms*".

Results presented in the EUGLOREH report were totally consistent with the above picture: "*For a number of reasons, among which an objective difficulty in measuring and exchanging of data on a large scale in a timely manner, the St.Vincent's objectives are still very relevant 20 years later. Tracking quality of care is paramount to prevent diabetes complications, but it is not an easy matter to realise it Europe-wide...However, collecting standardized and comparable data across countries remains a difficult job, mainly of collaborative nature, for which the support of health professionals is crucial, given their role in providing accurate clinical information*".

Accredited independent sources confirm the existence of the same limitations well beyond the continental boundaries and affecting both the policy field and the scientific literature.

The fourth edition of the IDF Atlas published in 2009[3] stated that: "*A search of published medical literature...covering the five-year period 2004-2008 to look for studies that referred to diabetes quality of care (found that) out of over 1,500 publications there were...only three attempts to compare quality of diabetes care across countries, each of which were limited to comparisons of data from just two countries and only one that compared national data..So, why is it that there is a large number of studies of diabetes care within countries, many based on multiple sites, yet so few international comparisons?* ***The simple answer is lack of consistently applied standards that would enable international comparisons. Standard systems and definitions, applied to comparable populations result in data that can be collected and compared relatively easily. The more unified systems are, the easier these comparisons become***".

The report also highlights that a major global initiative e.g. the OECD Health Care Quality Indicators Project, until now was only able to report on two diabetes indicators out of the nine originally identified: annual eye examination and lower-extremity amputation rate.

The simple answer addressed by the IDF Atlas may require a long time to be resolved before data is harmonized globally – if that would be ever possible – a time that cannot be waited by an exceptionally growing number of diabetes patients exposed to the risk of fatal complications.

In this report, we present a technical solution developed in the framework of a EU project, the EUBIROD "epidemiological analysis", arising from the perspective of the best possible usage of the existing information and the rapid propagation of the model at a minimum overall cost.

The logic underpinning the proposed solution is based on the examination of the main factors hampering the efficient data collection from multiple sources and the resolution of the methodological problems affecting the rigorous epidemiological analysis of routine diabetes information.

1.1 The problem of diabetes information

Diabetes is a growing burden for modern society on a global scale: its impact on the quality of life translates very rapidly into fatalities and disabilities producing a relevant change on the social structure that can undermine the financial stability of health systems.

A total of 285 millions are reported to be currently affected worldwide[3] (IDF 2009), corresponding to a prevalence of 6.6%, 46% of which in the 40-59 age group, with a projection of 438 millions (7.8%) forecast for 2030. Almost 80% of the total diabetic population resides in developing countries.

Worldwide, diabetes registers have been variously used to provide an immediate response to the needs of people with diabetes[4,5]. The stable integration of the available data sources offer an efficient solution for active surveillance at the regional, national and international level[6]. Registers using standardized data definitions allow for any aspect relative to the condition of a single individual to be compared against the average achieved by the general population under the same conditions[7]. Under ideal conditions, it is possible to obtain reliable indicators of quality and outcomes over time, avoiding the organizational burden of complex epidemiological studies that are difficult to repeat on a routine basis.

Previous experiences show that the successful implementation of electronic diabetes registers has been often realized through a long term process gradually evolving within the specific cultural context. For this reason, promoting and linking regional registers has been indicated as a convenient strategy for the progressive creation of national frameworks[4].

So, why is it so difficult to standardize these efforts and deliver systems of international comparisons that can be usefully made accessible for policy making?

The main reason is that diabetes registers have been created at different levels (health care authorities, regional, national) and for different scopes. The "basket" of data sources share a common structure, but includes fruits of different types (**Figure 1**).

Some registries emerge as a collaboration between **general practitioners**, who are usually very keen on recording risk factors for epidemiological purposes, or have contractual obligations which involve the due registration of pharmaceutical prescriptions for the control of health expenditure. In some contries that have undertaken "pay for performance" schemes, these registries are also made compulsory by the National Health System to link payments to the attainment of specific levels of health indicators.

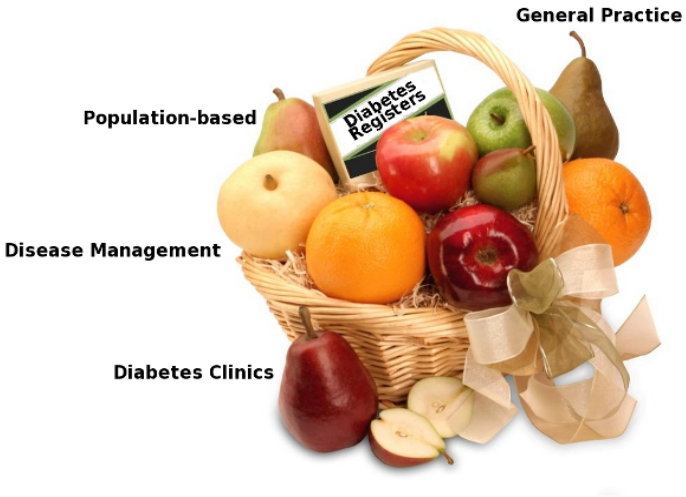


Figure 1. Types of Diabetes Registers

Usually, such registries are very effective in providing precise epidemiological measures, particularly in situations of universal coverage, since each practitioner has an assigned list of individuals who can be monitored over time. Therefore, exact denominators are available for all indicators. However, such sources have the limitation of a more difficult recording of terminal outcomes, as it can be difficult to link internal information to external databases e.g. hospital discharges, ambulatory data, etc.

Another type of diabetes registry that is quite common is that of "**specialist care**", usually involving diabetes clinics. Such registries include very detailed information on processes and intermediate outcomes, as these constitute the "core business" of their routine activity. The usual limitations of registries based on specialist care is that the case-mix bias is very difficult to control, and the catchment area of each data source is almost impossible to define. As a matter of fact, patients can be sistematically selected based on their specific characteristics, so that a relevant fraction of the population may not be observed. Furthermore, the geographical distribution of subjects enrolled in specialist registries can be very sparse and it would be difficult to attribute precise denominators to each data source. This implies that the main value of these registries lie in the average quality of care applied to the population recorded in the databases. However, it would be difficult to infer how representative such population could be, compared to the average population. Such potential bias cannot be controlled efficiently by risk adjustment methods, as we know very little on the population of subjects that has not been observed or is lost to follow up. Therefore, any comparison across centres or regions based on registries of specialist care must be taken with extreme caution in strict epidemiological terms.

Disease management registries share the same limitations of specialist care registries, but they offer the potential advantage of integrated care programs which by definition are amenable to link information from multiple sources. In fact, when coordinated care is performed more intensely, data from different practices can be made available through a more complete patient folder which can be used to popolate the set of characteristics included in the associated registry. This way, the range of characteristics and services applied to the target population may be covered much more efficiently by the information included in the registry, which can be used for the calculation of diabetes indicators. In many cases, disease management registries foresee the routine use of computerized benchmarking systems, which as a side effect can also lead to a significant increase in the quality of information over time.

Population-based diabetes registries, in their common definition, provide the most complete and accurate information for the calculation of diabetes indicators. The key element in this type of registry is the availability of a unique identifier and a set population denominator, which allows attributing each service and any outcome recorded to a specific individual over time. This way, classes of subjects can be easily allocated to the standardized categories involved in the calculation of any diabetes indicator. Population-based registry can automatically involve data linkage across different sources, usually in collaboration with the local government, through which they can have access to massive databases routinely maintained for administrative reasons.

Although optimal, the conditions required to establish population-based registries are not easy to realize under common circumstances. The organizational burden is significant and to be efficiently implemented, they require a substantial critical mass and technical capacity. The cost can be also high, which means that the level of support offered by the local government (both in legislative and economic terms) must be strong - and to be really effective, long term. Furthermore, the range of candidate institutions interested in undertaking this work can be usually large – including the academia, professional associations, industrial partners, etc - and it is normally difficult to assign such activity long term to one or more specific entities, in a highly competitive field e.g. diabetes.

Therefore, it seems clear that a variety of diversified approaches exist and it would be almost impossible to identify a prevalent condition under which a European framework should be realized.

The most efficient and natural solution would the one designed to capture the best information from all sources, so that standardized data would be pooled into a “melting pot” used to deliver European Diabetes Indicators.

The development model of a European Health Information System based on this approach would rather be inspired by a “bazaar” approach[8] where data, methods and software are progressively put together taking into account all the above differences, rather than a “cathedral” effort aimed at identyfying a unique solution to collect data for the European Diabetes Register. To this end, the adoption of open source tools represents a viable and sustainable solution, as they can be mutually exchanged at virtually no cost and progressively expanded.

Figure 2. Development Model of a European Health Information System

A fundamental reason to avoid restrictive guidelines to the management of diabetes data relates to the **respect of privacy protection principles and legislation across Europe**. This problem is particularly relevant for diabetes as the size of the affected population is extremely large and managing diabetes data would involve access to routine administrative data and/or medical records collected at the national or regional level. Here we briefly report the essential considerations made in the EUBIROD Privacy Impact Assessment[9] for diabetes registries.

In all cases, local data processing is subject to Art. 8(3) of the EU Data Protection Directive[10].

Diabetes registries organized across Europe involve centres collecting information related to an identified or identifiable natural person for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of healthcare services. In this case, the data collector is exempted from requesting consent from the data subject, in consideration of the need to protect the competing and general interests of societies in improved healthcare. The further processing of these data, other than caring for the patient and managing health services, would not be covered by the exemptions of Art. 8(3): in other words, consent would be required for any secondary use of those data.

However, according to Art. 11(2), for research and statistical analysis, even if consent was required in the first instance, the provision of information to the data subject could be waived if it proves impossible or would involve a disproportionate effort. The exemptions provided by the Directive are in line with the principles contained in the Convention on the Protection of Individuals for the Automatic Processing of Personal Data[11], envisaging the possibility of restricting the exercise of the data subject's rights with regard to data processing operations that pose no risk [Art. 9(3)]. Examples of no-risk or minimal-risk operations are therein considered, in particular, the use of data for statistical work, if those data are presented in aggregate form and stripped of their identifiers. Similarly, scientific research is included in this category.

However, while these solutions have been variously implemented at the national level, the problem of running a European health information system of population-based disease registries would be extremely difficult to realize, particularly if the transnational exchange of individual data is involved. To the best of our knowledge, such a system has never been trialled at a European scale for large populations and diseases at high prevalence e.g. diabetes.

Until today, the heterogeneous implementation of the EU Directive across Europe has made difficult to identify straightforward solutions for the transnational exchange of medical data. A robust architecture shall demonstrate its practical validity against the most restrictive interpretations of the Directive, so that its application would be possible across all boundaries.

In fact, the free flow of information, regardless of frontiers, is a principle enshrined in Art.10 of the European Human Rights Convention[12]. Accordingly, Art.12 of the Convention on the Protection of Individuals with regard to Automatic Processing of Personal Data (1981) and Art.25 of the Directive discipline the transborder data flow.

The main rule contained in Art.12(2) of the Convention, is that, in principle, obstacles to trans-border data flows are not permitted between Contracting States in the form of prohibitions or special authorisations of data transfers. The rationale for this provision is that all Contracting States, having subscribed to a common core of data protection provisions set out in Chapter II, offer a certain minimum level of privacy protection. In addition, no restrictions should be placed on the trans-border flow of medical data towards a State that has not ratified the Convention when the protection of medical data can be considered to be in line with the principle of equivalent protection therein laid down.

Therefore, any solid solution against the most restrictive interpretation of the EU Directive would automatically allow the cross border flow of personal data, provided that an adequate level of privacy protection is envisaged in the countries involved in the processing operations.

1.2 The BIRO System

The project "Best Information through Regional Outcomes" (BIRO) was specifically funded by DG-SANCO to trial an innovative solution aimed to overcome the aforementioned problems in the field of diabetes information. Activities started in September 2005 and successfully ended after 40 months in May 2009. Results of the project[13] are extensively described in twenty-two deliverables, summarized in a comprehensive monograph, all publicly available at the official website[14].

The expression "BIRO system" is referred to an overarching technology that has been specifically developed to implement a "Shared Evidence-Based Diabetes Information System" (SEDIS) through a joint effort of seven pioneer institutions from different Member States of the European Union.

The main characteristics of the system architecture have been described in a scientific paper presenting its characteristics of enhanced privacy protection[15].

Briefly, the BIRO system is based upon a structured architecture involving two data processing steps, corresponding to a local and a global component, linked by a uni-directional flow of information (**Figure 3**).

A basic version of the system runs in each single register ("local SEDIS").

As a fist step, a Java-powered "database engine" standardizes the register database to BIRO definitions ("mapping") and creates a local PostgreSQL database fully compliant with specifications provided by a "data dictionary". This way, any potential heterogeneity of the local data is either resolved by the user or discarded at the outset.

The standardized database is directly accessed by R[16] statistical routines ("statistical engine") to produce initial estimates for the local population. Results are included in each local report, made available in .html and .pdf format through the use of Latex.

The statistical engine also produces aggregate results that in the BIRO framework are referred to as "statistical objects", i.e. "elements of a distributed information system carrying essential data in

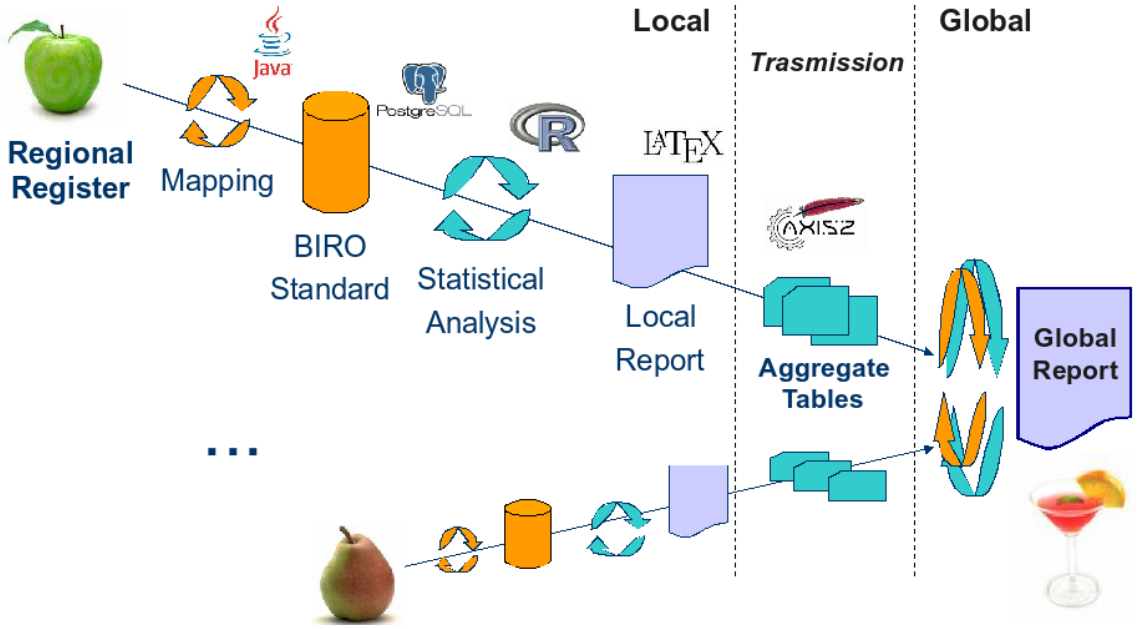


Figure 3. The BIRO System

the form of embedded, partially aggregated components, that can be used to compute a summary measure or relevant parameter for the whole population from multiple sites".

For each indicator included in the data dictionary, one or more statistical objects are created as CSV data files including original variables and summary columns (normally frequencies) that are required to produce overall estimates for the whole group of BIRO participating centres.

Communication software is used to send statistical objects to a central server located in Perugia (Italy) and administered by the BIRO Coordinating Centre (Perugia, Italy). This way aggregate results encrypted bundles can be securely transferred using web services according to the Axis 2 protocol.

The central server runs ad hoc software ("Java CSVImporter") to load all objects into a global PostgreSQL database according to a predefined structure ("central BIRO database").

At the final stage, a second instance of the statistical engine ("central engine") processes the central BIRO database to produce the European diabetes report ("global SEDIS") in both .html and .pdf formats.

All results are made directly available to a Web Portal that has been specifically developed to populate sections and pages of a dedicated website for the European Commission and the public.

The practical advantages of the BIRO system lie in its intrinsic ability to reduce the natural heterogeneity of different databases and to standardize procedures at all levels in a way that is sustainable and minimally invasive on everyday routine (Figure 4):

- **a common data model[17] incentivates compliance towards a set of agreed definitions through the adoption of a European Diabetes Data Dictionary. The database engine ensures that the format is homogenously applied. The dictionary can be further expanded and be continuously updated.**
- **a common report template[18] drives the development of statistical routines towards the delivery of a well defined range of indicators. A report using the same format is created separately for all participating centres and on a global scale for the whole collaboration. The range of indicators can dinamically evolve according to the needs and goals of the community and the statistical engine consequently adapted.**

Figure 4. Heterogeneity in the BIRO System

- **the exchange of aggregate tables ensures the highest level of privacy protection. The further development of BIRO is not endangered by increased restrictions on the usage of individual data that may not be considered "anonymous".**
- **The unique set of statistical routines included in the statistical and central engines allow to standardize all analytical procedures. This way results obtained by all registers are made comparable for any algorithm applied. The progressive development of statistical routines is easier and the application of all procedures is immediately possible for the whole partnership. The computational burden of data analysis is distributed across partners, so that the global analysis for very large populations only require minimum workload at the level of the central server.**
- **The adoption of open source software available at no licensing cost makes possible to rapidly propagate the approach universally.**

Although well defined in its fundamental elements at the stage of the initial proposal, the characteristic features of the BIRO system emerged as a collective response to the problem of diabetes information previously specified.

In particular, the discussion and particular solution arising from the process of privacy impact assessment allowed to understand and share all the separate steps involved in the system, identifying a common solution amenable to all partners that is directly linked to statistical processing.

As explained in detail in the BIRO privacy paper[15], statistical objects are only sent in aggregate form to the central server. For the most sensitive variables, aggregated records are not transmitted if groups contain less than five patients. Statistical objects are sent as tables stored in compressed bundles of flat text comma delimited files (CSV). Hence, there is no possibility, either directly or indirectly, that a patient could be identified with "reasonable means".

A further attention dedicated to the construction of the statistical engine refers to the level of "professional privacy" for what relates to the possible use of the report for performance evaluation and benchmarking. Partners of the BIRO Consortium agreed that such usage, particularly if conducted on an international level, could hamper data quality, completeness, and eventually discourage participation to the project.

The statistical engine has been consequently implemented to protect the ID of participating centres through the use of a pseudonym, optionally sending centre-stratified data saving only percentages in statistical objects, rather than absolute numbers, to avoid that the size of individual centres could indirectly reveal their identity by third parties.

Aggregated statistical objects are sent to the central statistical engine to carry out global analysis.

The security mechanisms implemented for the transfer of aggregate data provide a further shield towards unauthorized access that fully complies with security requirements enshrined in both the EU and international data protection norms. BIRO centres belong to European countries that have fully implemented the EU Data Protection Directive. Hence, an adequate level of privacy protection is fully guaranteed across those countries. It follows that the exchange of data envisaged in the BIRO project is legally viable, considering the system architecture and the composition of the Consortium. The same principle also applies for the transnational exchange of data between such Member States and international entities that apply the same level and rules of data protection.

Such safety measures are extremely restrictive, as the BIRO system has been showed to process only anonymous data; therefore, privacy rules should not limit its implementation. The fact that data processing occurring in BIRO is to be considered legitimate under EU legislation represents a fundamental element that underpins the solid implementation of the approach across a broad range of international partners.

1.3 The EUBIROD Project

"EUropean Best Information through Regional Outcomes in Diabetes" (EUBIROD)[19] is a three year public health project in the field of diabetes started on the 1st September 2008, sponsored by the European Union under the Health Information Strand of the Public Health Program (DG-SANCO).

The project mission is to implement a sustainable European Diabetes Register through the coordination of existing national/regional frameworks and the systematic use of the novel technology developed in the BIRO project by a subset of partners involved in EUBIROD.

The project fosters the objectives of the Conclusions of the EU Council for the systematic data collection and monitoring of diabetes complications and health outcomes across Europe[20]. Such influential document, at point 20, states that Member States are invited to develop and implement "**an evaluation system with measurable targets to track health outcomes and cost-effectiveness, taking into account Member States' organisation and delivery of their respective health services, ethical, legal, cultural and other relevant issues and available resources**".

The need to cover a very large population at the continental level, and the concomitant lack of convincing, sustainable statistical solutions to allow regular updating of diabetes data, makes broad on field testing of the BIRO technology highly strategic for the European Commission.

In fact, the statistical engine realized for BIRO allows to concentrate action on the standardization of existing data, distributing the workload of the analysis across a network of regional registries, which can autonomously deliver finely stratified diabetes indicators to a central server capable of producing European reports.

To demonstrate the solidity of the method, such a solution requires to be applied over a large number of countries, so that the feasibility of a fully automated European Diabetes Register can be evaluated taking into account a range of different procedures, data collection methods and technical skills.

To overcome the above problems, EUBIROD proposes an action to implement, extend, and customise the application of the BIRO technology in 20 States, including EU Member States, Acceding/Candidate Countries, and EFTA Countries.

The project includes nineteen partners managing diabetes registers in different European regions, one technological partner leading the privacy impact assesment presented in this report, one collaborating institution from outside Europe, and a major representative of the needs and expectations of people with diabetes: the International Diabetes Federation.

The statistical engine of the BIRO system aims to produce the "European Diabetes Reports" envisaged in EUBIROD, including a total of N=72 standardized diabetes indicators on top of a reference population of 500,000 subjects.

The project involves the adoption of a two pronged implementation strategy.

On one hand, EUBIROD supports improved information at the micro level, through the dissemination of standardized procedures for data processing statistical reporting, including an increased ability in interpreting results based on complex concepts e.g. stratification, risk adjustment etc. To this end, specific tasks include the conduction of residential courses and developing an e-learning platform under the banner of "BIRO Academy", including specific subjects on statistical methods.

On the other hand, EUBIROD specifically addresses the macro level, by delivering timely reports that will include information for policy to all Member States and the European Commission.

1.4 Statistical and central engine in the BIRO project

The main objectives of the statistical routines implemented for the first time in the BIRO project were:

- to run the same specialised statistical software in each partner region, directly tapping into a standardized PostgreSQL database extracted from local data, formatted according to common definitions specified in a BIRO concept/data dictionary.
- to implement and disseminate use of advanced statistical methods to collect and analyse population-based data stored in diabetes registries through a fully documented repository of open source statistical software that will allow users to replicate and further extend the approach.

In BIRO, R software has been adopted as a development platform for the statistical engine, launched directly by script command files or with the aid of a GUI interface. The engine connects to the local database using R Postgres drivers. The concept of "statistical object" has been introduced as "an element of a distributed information system that carries essential data in the form of embedded, partial aggregate components, required to compute a summary measure or relevant parameter for the whole population from multiple sites".

Objects are created as tables including statistical aggregations of local data (e.g. the arithmetic mean, percentile, variance, etc.), stored as flat text comma delimited files. A taxonomy has been specified to provide details of all objects being implemented. Specifications provided by the report template have been used to process data and deliver objects as small datasets. Graphical functions and Latex are used to produce individual centre outputs and full local reports in the form of .html files and .pdf documents. A compressed folder is created to deliver all statistical objects produced by local runs of the statistical engine, stored in a directory named with datetime/centre id, transmitted to the central server.

The statistical engine has been successfully developed and tested on both Vista and Linux. Average hardware allowed completing a full BIRO report from a test sample of more than 2,000 patients and several thousands episodes in less than 8 minutes. Installation of the software is identical regardless of the hardware, and requires R>1.8, Latex, Java 6.0 and PostgreSQL plus various additional libraries/packages that are included in its distribution. All R functions are released under the GPL license.

The results highlighted that the statistical engine can provide a platform for accurate benchmarking that currently does not exist at the point of health care provision. It may serve multiple users, from the European Union, to provide updated benchmarking of key indicators on a routine basis, and the local physician, to monitor the status of patients in a modern standardized procedure. The system may improve, through a shared infrastructure, the validity and completeness of information available. Existing registers may be optimised on the basis of common standards, and new ones can be created with a fostered structure. Advantages proposed by the system should be part of a progressive approach through which statistical functions are constantly improved. Users, once inducted to using the software, can apply it independently and submit better aggregate data to the central server, at the same time safeguarding privacy at the highest level of protection, as a result of the application of rigorous rules set by the BIRO privacy impact assessment.

The development of the statistical engine provided the basis for an expandable open product that through its availability at no charge can crucially help disseminating the BIRO approach across Europe. All details of the statistical routines realized in the BIRO project can be found at the specific deliverable report, available at:

http://www.biro-project.eu/documents/downloads/D8_1%20Statistical_Engine.pdf

1.5 Statistical and central engine in EUBIROD

The statistical tools developed in EUBIROD have been targeted by a specialized Workpackage, WP6: “Epidemiological Analysis”. The WP leader is Serectrix, with all EUBIROD partners involved in the delivery of tasks envisaged.

The main objective of Epidemiological Analysis was “To deliver the EUBIROD Statistical Engine needed for local reports and international comparisons.”.

The WP has been considered central to the realization of the project, as EUBIROD focuses on the analysis and reporting of diabetes indicators, and was not regarded purely as an Information Technology application. The core business of the project was considered to exploit the valuable features of the BIRO project and concentrate more on the epidemiological aspects, including the implementation of a sound methodology to support policy decisions in diabetes

The major task of the WP is to build up the EU Diabetes report on top of a consolidated statistical engine, including an in depth validation of standardised estimates using risk-adjusted models.

The BIRO statistical engine will need to be extended to allocate different schemes for risk adjustment and various means to compare observed and expected rates.

In particular, the dedicated WP had the major task of adding recent advancements made by the AHRQ contributing in the construction of the US system of quality indicators. In EUBIROD, the Consortium aimed to replicate the same methodology to align European estimates of diabetes indicators to international gold standards.

The same methodology needed to be separately applied at the level of both the statistical and central engines. Such developments were needed to use both internal and external standards (e.g. multivariate regression models estimated from the regional or national level) and benchmark results obtained at all levels.

In other terms, the statistical engine must enable comparisons of observed rates between centres in the region, along with comparisons of expected rates computed according to different regional/national/European standards. At the European level, the model must encompass standardization of results obtained from different countries under common terms of reference.

All tasks require to be implemented for each indicator and must take into account all stratification criteria originally included in the report template.

A particular attention in the revision of the initial BIRO prototype had to be dedicated to the analysis of data quality and the selection of observations to be included in the statistical analysis.

The further refinement of graphical displays and printed output was targeted as a means to facilitate the uptake of the report and the correct interpretation of its results by end users.

2. Materials and Methods

2.1 General design of the Statistical and Central Engine

The general design of the EUBIROD Statistical and Central Engines builds upon the approach developed in the framework of the BIRO Project.

Briefly, the statistical routines revolve around the application of R software, used to load and transform the original variables included in the BIRO database, then carrying out the whole statistical analysis through various routines.

As for the BIRO prototype, the application of R routines is triggered by the user, either through a script command file or with the aid of the BIROBox. R routines directly connect to the local database using proper Postgres drivers.

The source code implements all specifications given by the original *report template,* including the associated definitions of statistical objects.

The set of R functions realized for the statistical engine already documented in the BIRO deliverable are used to process the Postgres database including all the individual observations, and deliver a range of files in separate folders under a specified working directory (see section 2.2):

- EUBIROD pdf reports, created through the application of the high quality, open source typographical software Latex.
- statistical tables in html format
- graphical outputs, saved in:
  - png format for the html report
  - pdf for the pdf report
  - vectorial format (.svg) for high quality typographical production

  All graphs available for later use and/or inclusion in document, slides, etc.
- EUBIROD html report linked to all tables and graphical images
- statistical objects in the form of small CSV datasets. The taxonomy of statistical objects delivered by the EUBIROD statistical routines is based on the structure included in the original BIRO deliverable. However, this has been expanded in depth to reflect the substantial increase in the range of outputs delivered by EUBIROD software for the different strata. The characteristics of these improvements are reported in section 2.5, “Enhanced Tables and Stratified Outputs”. A folder including all statistical objects as CSV files is created at each run by the statistical engine, to be passed as input to the central engine.

The statistical objects are loaded into a central database through the use of a CSV importer that transfer their contents to an overall Postgres database. Aggregate tables of the same kind are piled up to allow calculations of all indicators for the overall population.

The central engine uses such central database to deliver the global report, mirroring the entire set of results and creating the same range of reports delivered by the statistical engine. Further usage of statistical objects created by the central engine allows a recursive application of the central engine to progressively cover a broader population without using individual data.

2.2 Revised BIRO Statistical and Central Engine Directory Structure

The original directory structure created for the BIRO project has been revised in EUBIROD to allow the neat application of the BIROBox within the BIROX distribution (for more specifications see EUBIROD deliverables D5.3: Database Engine and D7.1: Customized toolbox, available at: http://www.eubirod.eu/deliverables.htm).

In particular, the output directories have been moved from the source code subdirectory to an external working directory, to allow more flexibility in the management of the outputs and safeguard the core software.

Through the new EUBIROD structure, it is now possible to direct all outputs to a different location, particularly one outside the machine running BIROX on the VirtualBox, so that the results can be saved more conveniently and later accessed by the user on own operating system.

The new directory structure is shown in detail in **Box 1**.

```
lib  /maps
     /r/source          /biro
                        /packages                          /linux
                                                           /pdf
                                                           /vignette
                                                           /win
     /templates
_se_ source/r /formats
              /include
              /main
              /scripts
_ce_ source/r /formats
              /include
              /main
              /scripts

<Working Directory>
     _ce_      /data/<datetime>/<year>/<region_id>        <local_comp.csv>
                                                          <cum_comp.csv>
               /output  /data/<datetime>/<region_id>      <statobjects>
                        /reports/<datetime>/<region_id>   /graphs
                                                          /tables
                                                          /html
                                                          /images
                                                          /pdf
                                                          <database>.pdf,.html,.log
     _se_      /data/<datetime>/<year>/<centre_id>        patient.csv
                                                          episode.csv
               /output/data/<datetime>/<year>/<centre_id> <statobjects.csv>
               /reports/<datetime>/<year>/<centre_id>     /graphs
                                                          /tables
                                                          /html
                                                          /images
                                                          /pdf
                                                          <database>.pdf,.html,.log
```

Box 1. Revised EUBIROD structure of the Statistical directory

2.3 Running the Statistical Engine

As for BIRO, the application of the statistical engine is characterised by the processing structure shown in **Box 2** (pseudo-code).

```
Start
1. Setup environment
2. Compute Indicator Statistics
 For each indicator in the Report Template:
 Loop Start
    Reference Indicator
IF i-th statistical procedure is TRUE then
 Apply Statistical Procedure
 Output production
    END
Loop End
3. Compile results
End
```

Box 2. Simplified Structure of the EUBIROD Statistical Engine

The same loop is presented as a flow chart in **Figure 5**.

Briefly, this can be explained as follows. The first step relates to the definition of the workspace, data preparation, and output formatting. The execution starts with a fresh setup of the complete environment, including a check of the local OS version, any required installation of additional R packages, and the definition of global variables. The BIRO database is formatted by applying definitions in the data dictionary: new variables are created using a predefined set of cutoffs, new tables are created by merging and linking the original datasets into a new format amenable to statistical analysis. Finally, html and tex (pdf) outputs are initialized and formatted where required.

A second step is required to compute all indicator statistics. The complete list of BIRO indicators is read from the report template, along with definitions included in the data dictionary. An indicator "cohort" is automatically constructed, based upon the agreed specifications relative to the particular category of patients that must be included in each indicator.

Appropriate database and statistical procedure are executed to reproduce algorithms foreseen for each indicator, until the complete list of tasks is finalised and the set of planned outputs is entirely produced.

The loop ends when the complete list of indicators in the BIRO report template is produced. All results are compiled into an overall report that is produced in PDF and HTML format for the local centre site, including output files that include raw data, text listings (individual html tables) and graphical outputs. Results are stored in a directory with a unique timestamp, whose content is sent by invoking a BIRO routine towards the central server, where they are used by the central engine to produce European results from a part or all BIRO participating centres.

The complete list of functions specifically created to realise the statistical engine, along with their location in storage files, resembles the one presented in detail in Box 4 of the BIRO Deliverable D8.1.

All details can be found in the EUBIROD source code annexed to the present deliverable.

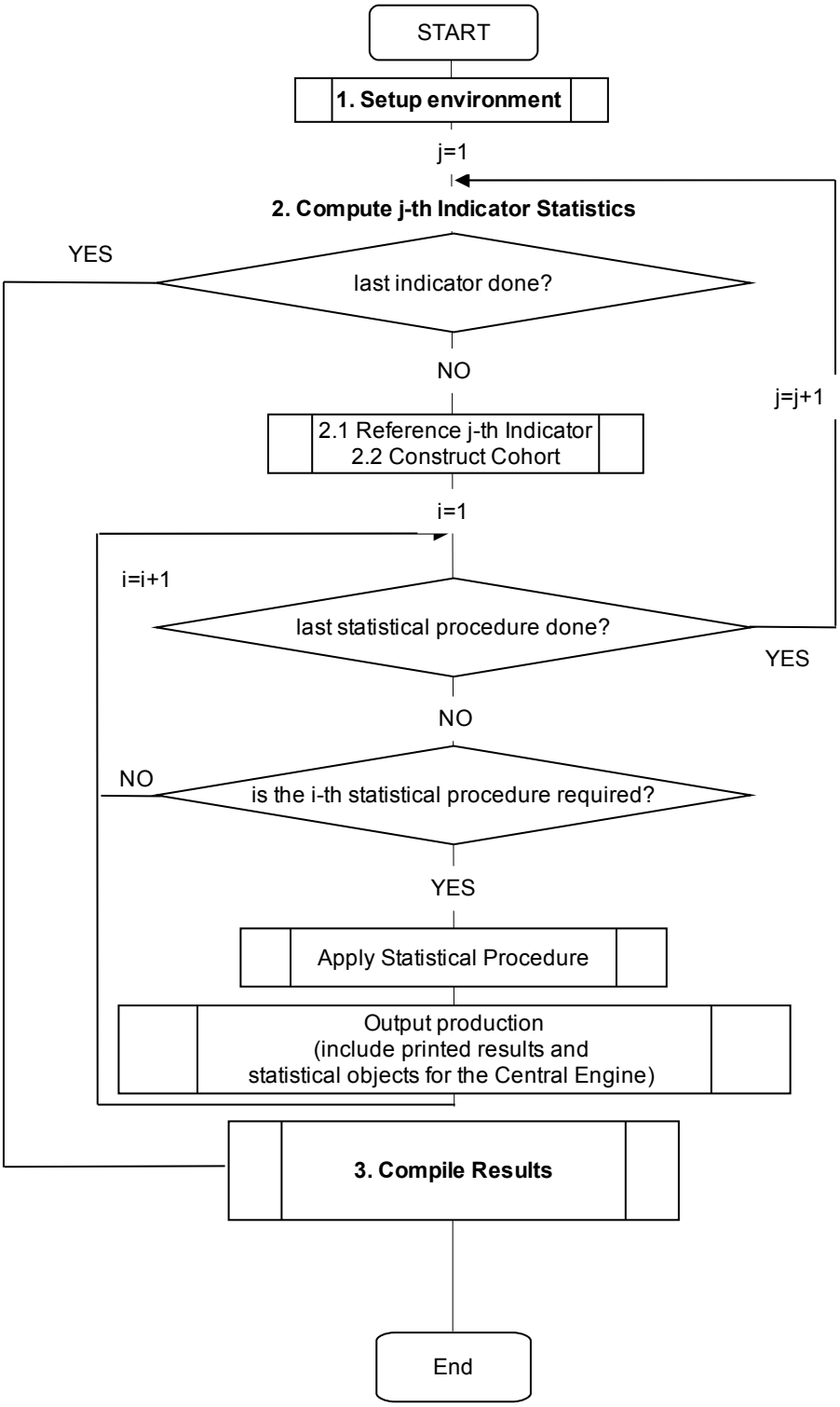


Figure 5. Statistical Engine Flow Chart

The EUBIROD statistical engine is run either directly or from within the BIROBox using the source code included in **Box 3**. These specifications include the list of parameters required by the statistical routines, related to: the usage of the specific database driver; name of the target database and centre ID, population and activity tables; format of the database tables (wide/narrow); size of fonts in graphical display; output directory; time interval and reference year; name of the log file; whether the tex file must be compiled to create the pdf.

```
rm(list=ls())
source("/home/fabrizio/Desktop/testrun-2.0.7/_se_/source/r/main/biro_se_.r")

BIRO_se(dirse="/home/fabrizio/Desktop/testrun-2.0.7/_se_",
     dirout="_se_",
     dbformat="postgres",
     driverClass="org.postgresql.Driver",
     classPath="BIROCommonLibraries/postgresql-8.4-701.jdbc4.jar",
     identifier.quote="`",
     pathdb="jdbc:postgresql://localhost/foligno",
     user="postgres",
     password="postgres",
     dbname="foligno",
     dirdatastore="",
     centre_id="2",
     startdate="2008-01-01",
     enddate="2008-12-31",
     yearnow=2010,
     refanadate="12-31",
     logfile="statisticalEngine.log",
     cex= 1,
     wide=1,
     filepop="",
     filepopdiab="",
     activitytable=0,
     compiletex=1)
```

Box 3. Commands to Run the Statistical Engine

The above parameters are more directly and easily applied using the BIROBox, as shown in **Figure 6**.

Pushing the “Run Statistical Engine” button triggers the application of the routine (via the “Rserve” package).

A “Statistical Engine Status” window is displayed in foreground, progressively showing all major actions performed for each indicator scheduled in the analysis (based on the available data). The log window also includes a progress bar to allow monitoring the percentage of work already performed.

A log file including all messages shown in the status window is saved in the output directory at the end of the procedure, using the same name of the input database with a .log extension. This file can provide fundamental information on the input data used for the analysis, the range of outputs produced, and the execution times.

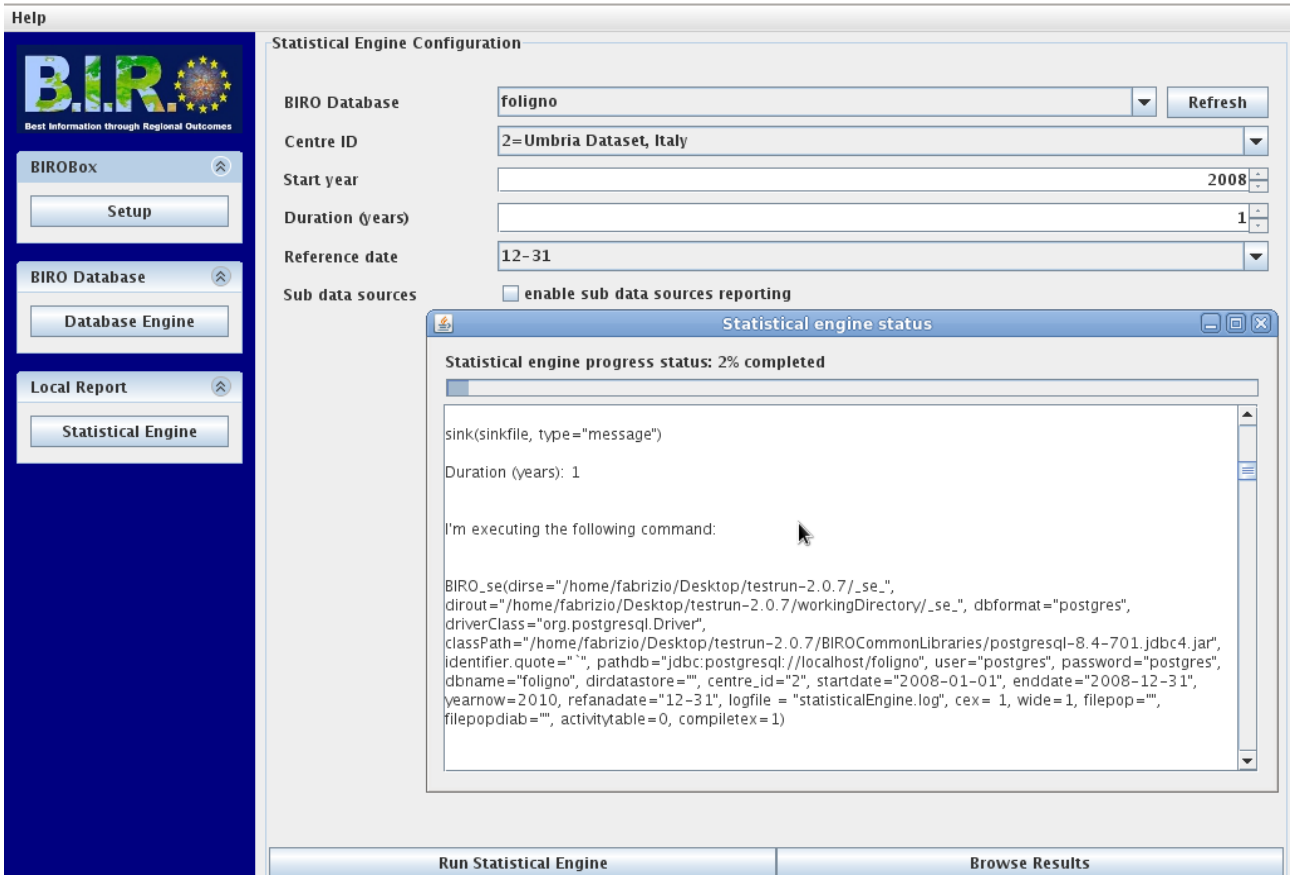


Figure 6. Running the EUBIROD Statistical Engine from the BIROBox

2.4 Running the Central Engine

The application of the central engine requires loading the statistical objects (aggregate tables) created by the statistical engine (eventually for recursive application also by the central engine itself) into a central database.

The operation is performed by the CSV importer, specifically developed to pile up all similar tables from different runs into an overall database including the basic elements required for the calculation of all indicators related to a global target population.

A typical application of the CSV Importer is shown in **Box 4**. Here, a java application is run using the common driver and a configuration file to import an entire directory including all statistical objects. These objects are ordinarily transferred to a central server by ad hoc communication software specifically developed for the scope by the Consortium. However, this operation can occur even on the same computer when both the statistical and the central engine are used on top of multiple data sources to compile an overall report.

```
java -Xmx1024m -cp
 BIROAdaptor2.jar:lib/postgresql-8.2-504.jdbc3.jar
 eu.biro.adaptor.csv.CSVImporterMain
 CSVImporterConfig.conf
 "_se_/output/data/#231010003157/2008/2/"
```

Box 4. Loading Statistical Objects with the CSV Importer

The central engine is triggered by a command file resembling the application of the statistical engine, as shown in **Box 5**.

Further parameters allow to disaggregate the results by a sub source (see section 2.5 on enhanced outputs) and to adjust the denominators of the population files when they do not exactly correspond to the catchment area of each data source (centre).

In the latter case, it is possible to specify a linear approximation by dividing the population in proportions of equal size (e.g. dividing the general population data supplied by user by the total number of centres in the region).

Commands in Box 5 are executed following a procedures whose flow chart closely resembles that of the statistical engine (see **Figure 7**).

```
rm(list=ls())
source("_ce_/source/r/main/biro_ce_.r")

BIRO_ce(dirce="_ce_",
        dirout="workingDirectory/_ce_",
        dbformat="postgres",
        driverClass="org.postgresql.Driver",
        classPath="CSVImporter/lib/postgresql-8.2-504.jdbc3.jar",
        identifier.quote="`",
        pathdb="jdbc:postgresql://localhost/central",
        user="postgres",
        password="postgres",
        dbname="central",
        dirdatastore="",
        centre_id="umbria",
        where="",
        logfile="ce.log",
        cex=1,
        report_list_id=c("centre_id","dbname"),
        disaggregation_by="",
        standardization_by="dbname",
        dividedprev=TRUE,
        divisornumber=6)
```

Box 5. Commands to Run the Central Engine

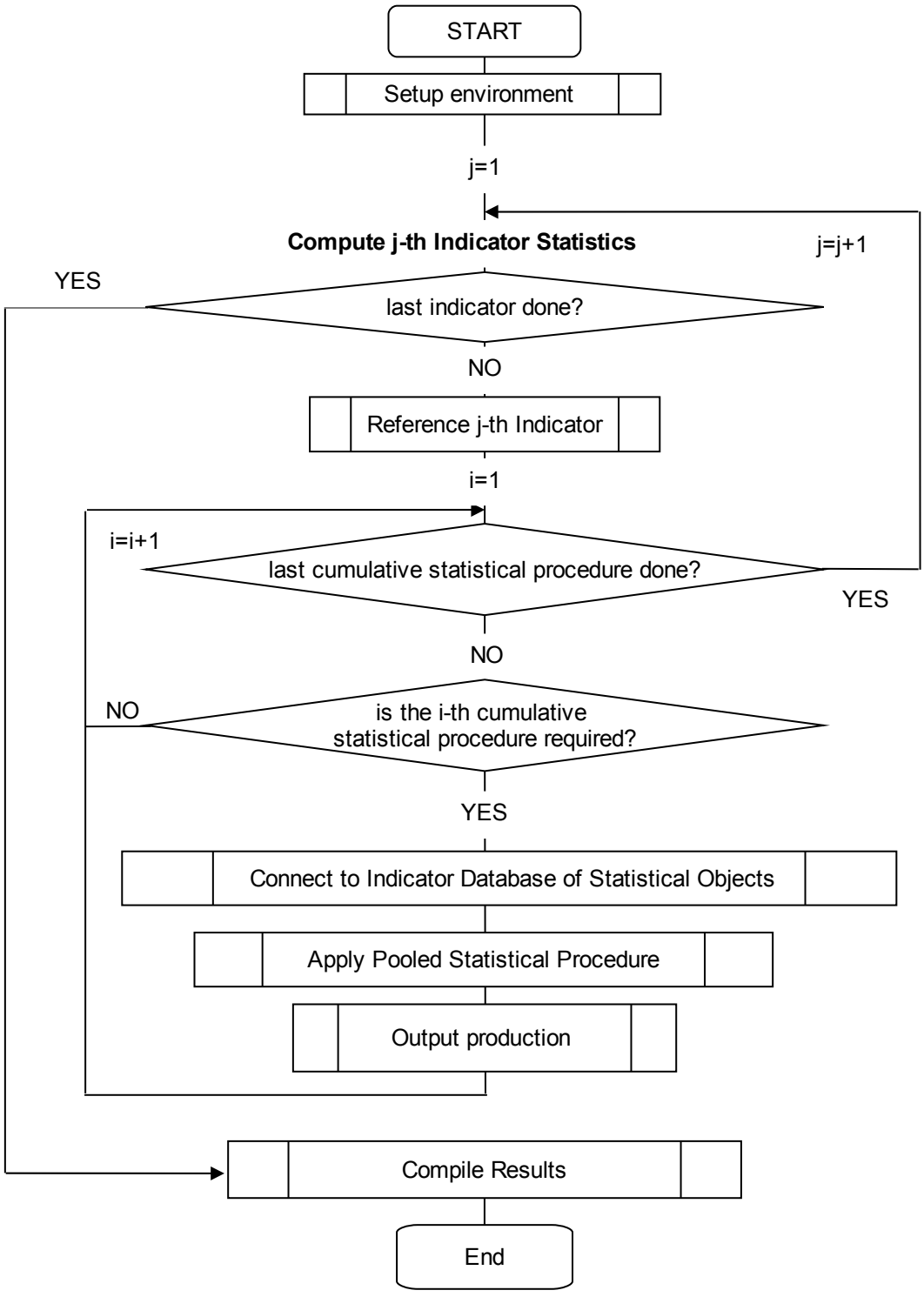


Figure 7. EUBIROD Central Engine Flow Chart

2.5 Enhanced Tables and Stratified Outputs

The EUBIROD statistical routines produce more sophisticated outputs compared to the previous BIRO release. The design is the result of several rounds of discussions and presentations of test results to all EUBIROD partners. The suggestions made were evaluated by the EUBIROD core development team following each meeting. The University of Dundee (R.McAlpine) provided an essential contribution to identify the best format, based on their long standing experience in the field of diabetes reporting in Scotland.

A general epidemiological perspective has been applied to plan all tables and graphical outputs to be produced in EUBIROD.

A first improvement relates to the introduction of a “root table” and a “body table” for each tabular output. This was required to assess the impact of missing values on all tables. In fact, unless missing values are reported as a separate class for each reported table (unusual), patterns of observations included in frequency tables may hide the influence of values discarded from the tabulation due to missing values. Furthermore, tables reported for separate indicators would usually show different total number of observations, which can appear very confusing for the reader.

The “root table” reports the same total number of observations for all indicators. Observations are then stratified by valid/non valid values for each variable taken in account in the separate tables. Since only observations with valid values for all variables used in the tables are duly reported in frequency tables, only one cell in the “root table” would include the total number of observations passed to the “body table”. This way, the possible influence of the composition of the “root table” on the reported “body table” is clearly displayed to the user.

A second cluster of improvements was made in the general construction of all tables, summarized as follows:

- all frequency tables produced for each indicator now include one target response (outcome) and up to two cross tabulated exposure factors.

  If the target response has i categories, exposure 1 has j levels and exposure 2 has k levels, the resulting table has a total of (i*j*k) cells. For example, a binary indicator of low/high level of HbA1c (two categories), stratified by four different age bands and sex (two levels), implies the construction of a table with 2*4*2=16 cells.

- The table is constructed with the outcome in the rows and the exposure factors in the columns. A one way table includes only one exposure factor, while a two way table includes two of them. The columns are built by nesting levels of exposure factor 1 within each level of exposure 2. In the previous example, for each level of Hba1c, there will be four columns for males and four for females.

- each (i,j,k) cell in the table presents the absolute frequency of observations and the column percentage relative to the specific cross tabulation of exposure factors. In the previous example, the percentage of males/females with a specific level of Hba1c in each age band out of all males/females in the same age band.

  This representation allows the direct computation of the relative risk across different levels of the exposure factors, by dividing a certain percentage for the percentage shown in a different cell.

- The column marginals present the total number of observations for the specific exposure (exposure 1 for tables with two exposure factors), with row percentages computed over the grand total.

- The row marginals present the total number of observations for the specific level of the target response, with column percentages computed over the grand total.
- All tables include the calculation of the Chi Square Test (value and associated p>Chi-Square), to test the association between the exposure and the outcome of interest. When two exposures are included, the Chi-Square can be used to test the association between exposure 1 and the outcome of interest, stratified by levels of exposure 2.

A third fundamental improvement relates to the introduction of a general “Class variable”, which triggers the creation of n tables for the target response and associated exposures, one for each level of the class variable.

In diabetes indicators, the usual “class variable” of choice is “Type of diabetes”. Thus this option allowed us to replicate the production of all tables for all indicators for levels: Type 1, Type 2, Other Type.

All graphical displays are created according to a common structure for all strata of response, exposure and levels of the class variable.

Finally, a very powerful option that has been introduced in EUBIROD relates to the “sub_source_id”. Through this parameter, it is possible to display all outputs comparing levels of a certain variable, usually the centre ID. W

hen data from multiple centres are present in the same data source, graphs may be used to benchmark results obtained by different centres against the overall average (ex: regions from different parts of Europe or centres within a region).

This way some of the best results available only through the central engine can also be produced at the level of the statistical engine, leading to a more relevant use of the EUBIROD software at the local level.

2.6 Risk Adjustment Methodology

Standardized estimators allow to rigorously compare quality of care and outcomes across different centres, regions or countries taking into account the possible imbalance in the case-mix, which can be systematically associated to systems performance. For example, a centre with older and sicker patients would normally experience higher rates of diabetic complications compared to the average population.

Risk adjustment methods allow standardizing all results against an ideal population usually corresponding to the total population target of the analysis. In the case of EUBIROD, the best comparison would be made against the European population.

The most advanced application of risk adjustment involves the use of multivariate models to assign weights for each risk factor of interest (exposure variable) on the rate of outcomes observed for a specific indicator.

Since all risk adjusted indicators in EUBIROD are expressed in terms of binary outcomes (yes/no, low/high, etc), a natural candidate for the multivariate modelling approach is that of logistic regression.

The EUBIROD statistical routines implement the method adopted by the US Agency AHRQ[21] for the calculation of standardized quality health care indicators.

Briefly, this work as follows:

- a multivariate model is run on top of the overall population based upon a specified outcome and a set of target covariates intended as potential risk factors (confounders). In quality of care, these can be assumed to be observed components of the case-mix that are potentially associated to the outcome of interest. Their effect shall be isolated by that potentially related to the quality of care delivered by a specific centre or region, which we may want to monitor or benchmark across a group of providers.
- weights extracted from the multivariate model are applied to each subject in the sample, applying the logistic model to compute an estimated probability of the outcome for that specific subject.
- the sum of the estimated probability across each centre or region is computed as the average “expected rate of events” (as specified for each indicator) for the particular centre.
- the quantity (observed rate/expected rate) is used as a multiplier (penalty if >1, premium if <1) of the average population rate to compute the “standardized rate” for each centre in the overall sample
- the percentage of observed minus expected over the expected number of cases for each centre is used as a measure of the excess/reduction of cases in each centre, compared to the average level
- all risk adjusted measures are published along with 95% confidence intervals, based upon a precise formula of the variance of the estimates.

Graphical display including barplots of standardized rates against the average, and forest plot of O-E/E%, with the related 95% confidence intervals, may offer an immediate representation of the variability of results across the whole sample of centres included in a report.

The technical details used for the calculation of EUBIROD risk adjusted indicators, directly obtained from the AHRQ, are included in **Box 6**.

Box 6. EUBIROD Risk Adjustment Method based on AHRQ Quality Indicators (modified from direct communication received from AHRQ Online Suport, Version 3.0, 23/5/2006)

The EUBIROD Statistical and Central Engine compute risk adjusted indicators using the method implemented by the AHRQ for quality of care indicators.

All of the AHRQ Quality Indicator routines begin with estimating a logit model of a 0/1 outcome variable and a set of subject-level covariates as dependent variables, and using the results to form the expected outcome for each subject (e.g. P=pr(outcome=1)).

I. Notation:

Y_{ij} = 0 or 1, outcome for patient j in centre i.
X_{ij} = covariates (e.g., gender, age, DRG, comorbidity)
P_{ij} = predicted probability from logit of Y on X
= $\exp(X_{ij}\beta)/[1+ \exp(X_{ij}\beta)]$
where β is estimated from logit on entire sample.
e_{ij} = $Y_{ij} - P_{ij}$ = logit residual (difference between actual and expected).
n_i = number of patients in sample at centre i.
α = average outcome in the entire sample[1] (e.g. Ybar).

II. Estimating the Risk Adjusted Rate (RAR) and SE using the *Ratio* Method[2] of Indirect Standardization for each Centre:

1. Estimating RAR:

let $O_i = (1/n_i)\sum(Y_{ij})$ be the observed rate at centre i
let $E_i = (1/n_i)\sum(P_{ij})$ be the expected rate at centre i

RAR_i
$= \alpha(O_i/E_i) = \alpha\ [(1/n_i)\sum(Y_{ij})]/\ [(1/n_i)\sum(P_{ij})]$ (where sum is for j = 1 to j = n_i)
= **population rate * observed/expected at centre i.**

2. Estimating Variance of RAR (SE is the square root):

$Var(RAR_i)$
$= Var[\alpha(O_i/E_i)]$
$= (\alpha/E_i)^2 Var[O_i]$ (since $var(aX) = a^2var(X)$ for any constant a)
$= (\alpha/E_i)^2 Var[(1/n_i)\sum(Y_{ij})]$ (by the definition of O_i)
$= (\alpha/E_i)^2 (1/n_i)^2 Var[\sum(Y_{ij})]$ (since $var(aX) = a^2var(X)$ for any constant a)
$= (\alpha/E_i)^2 (1/n_i)^2 [\sum Var(Y_{ij})]$ (since $var(\sum X_i)=\sum var(X_i)$ if Xi are independent)
$= \mathbf{(\alpha/E_i)^2(1/n_i)^2\sum [P_{ij}(1-P_{ij})]}$ (since Y is 0/1, so var(Y) = P(1-P))

[1] For the AHRQ QI, the sample is the entire reference population consisting of the discharges in the States Inpatient Database for the participating states pooled over three years (2001-2003). Therefore, the "average outcome for the entire sample" is the population rate.

[2] Risk-adjusted rate = (Observed rate / Expected Rate) * Population Rate

3. Results

The statistical routines planned for WP6 of the EUBIROD project have been all successfully developed and practically tested on data from participating centres during the first two years of the project.

Additional files attached to the present deliverable have been made available to document all the details of the work undertaken and to present the range of results that can be obtained through the EUBIROD reports:

- Help files included in each EUBIROD Report are included in Appendix 1, available at: http://www.eubirod.eu/documents/downloads/D6_1_Statistical_Materials_Appendix1.pdf
- A sample of the EUBIROD Statistical Engine Report is included in Annex 1 (1,058 pages), available at: http://www.eubirod.eu/documents/downloads/D6_1_Statistical_Materials_Annex1.pdf
- A sample of the EUBIROD Central Engine Report is included in Annex 2 (530 pages), available at: http://www.eubirod.eu/documents/downloads/D6_1_Statistical_Materials_Annex2.pdf
- All the EUBIROD source code for the Statistical Engines is included in Annex 3 (292 pages), available at: http://www.eubirod.eu/documents/downloads/D6_1_Statistical_Materials_Annex3.pdf

3.1 Structure of the EUBIROD Report

The structure of all EUBIROD reports is shown in **Figure 8**.

The pdf version is the most complete and user friendly, as it includes bookmarks, the list of all BIRO/EUBIROD contributors, a set of basic Help files to help the reader in the interpretation of results, and the parameters used by the statistical routines to produce the report (see Annex 1, 2).

For each indicator, the statistical engine produces a root and a body table. In the most complex case where two exposures and the class variable are present, the outputs include a separate table for each exposure and both exposures for each level of the class variable.

For instance, if the two exposures are Age, Gender, and the class variable as usual is Type of Diabetes, the report will include a root and a body table for Age, Gender, Age*Gender for Type 1, Type 2, and Other Type of Diabetes.

In each report, and for each indicator, the table section is followed by a list of graphs for all the variables included in the process. These include barplots for categorical variables, and boxplots, trellis plots for continuous variables. All graphs are stratified by sub source (usually centres or regions) if the option “enabl sub data source reporting” is selected.

For risk adjusted indicators, the report includes additional tables of standardized rates and observed minus expected excess/reduction along with 95% confidence intervals.

Graphs included in these outputs are barplots and forest plots.

Maps and longitudinal trends have been planned but not yet implemented due to the need of identifying a common solution for geographical coding and unique IDs to be used by the central engine (see Discussion).

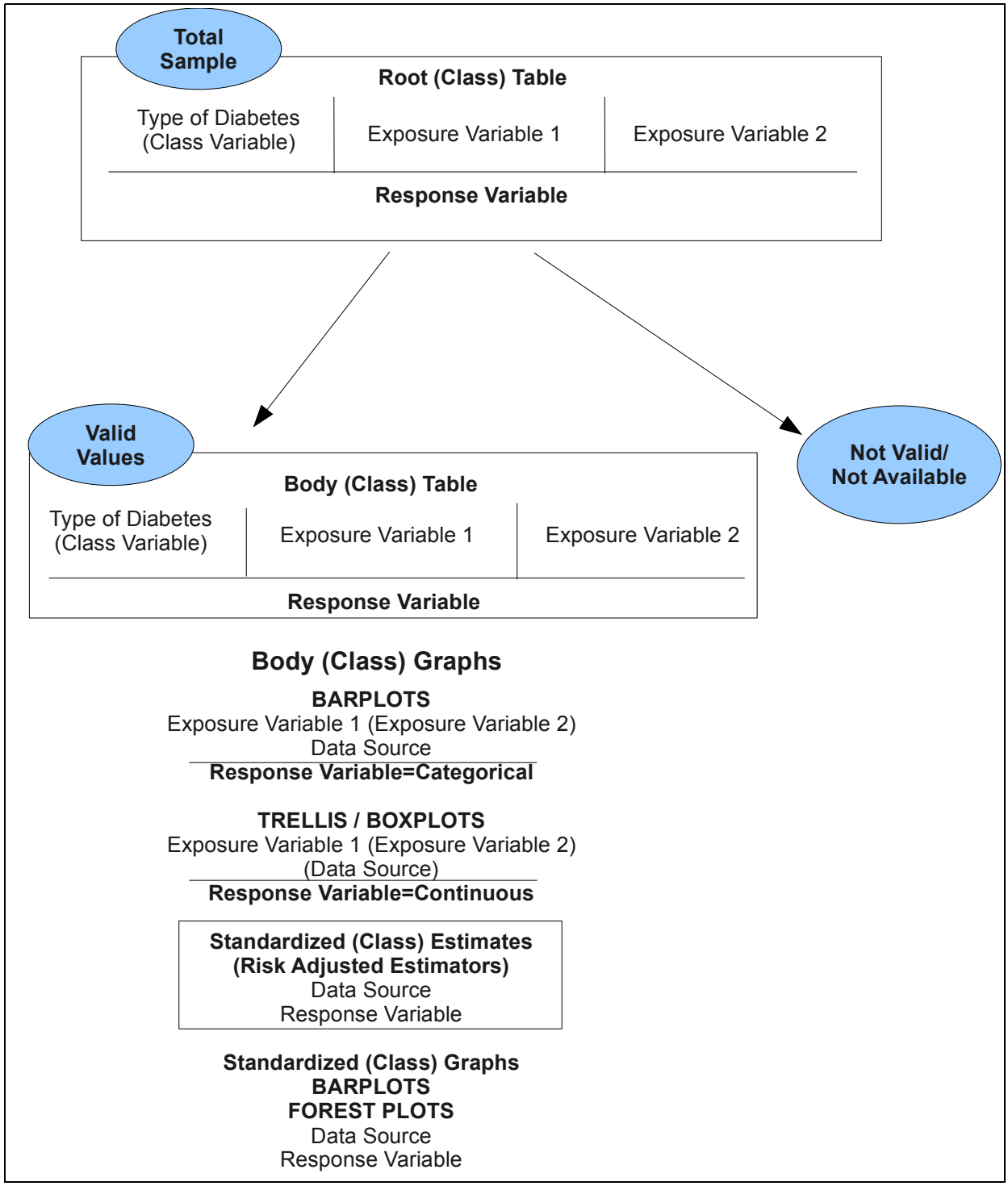


Figure 8. Structure of the EUBIROD Report

3.2 Tabular Outputs

The range of tabular outputs that can be obtained by the EUBIROD statistical routines are shown in **Figure 9-10**.

Briefly, the output for each indicator is clearly labelled in the header, with a unique identification code and the associated description. The particular level of the class variable (Type of Diabetes), if relevant, is also included at the top of each page for the convenience of the reader.

The explanations provided in the figures, also included in the help pages, are self explaining.

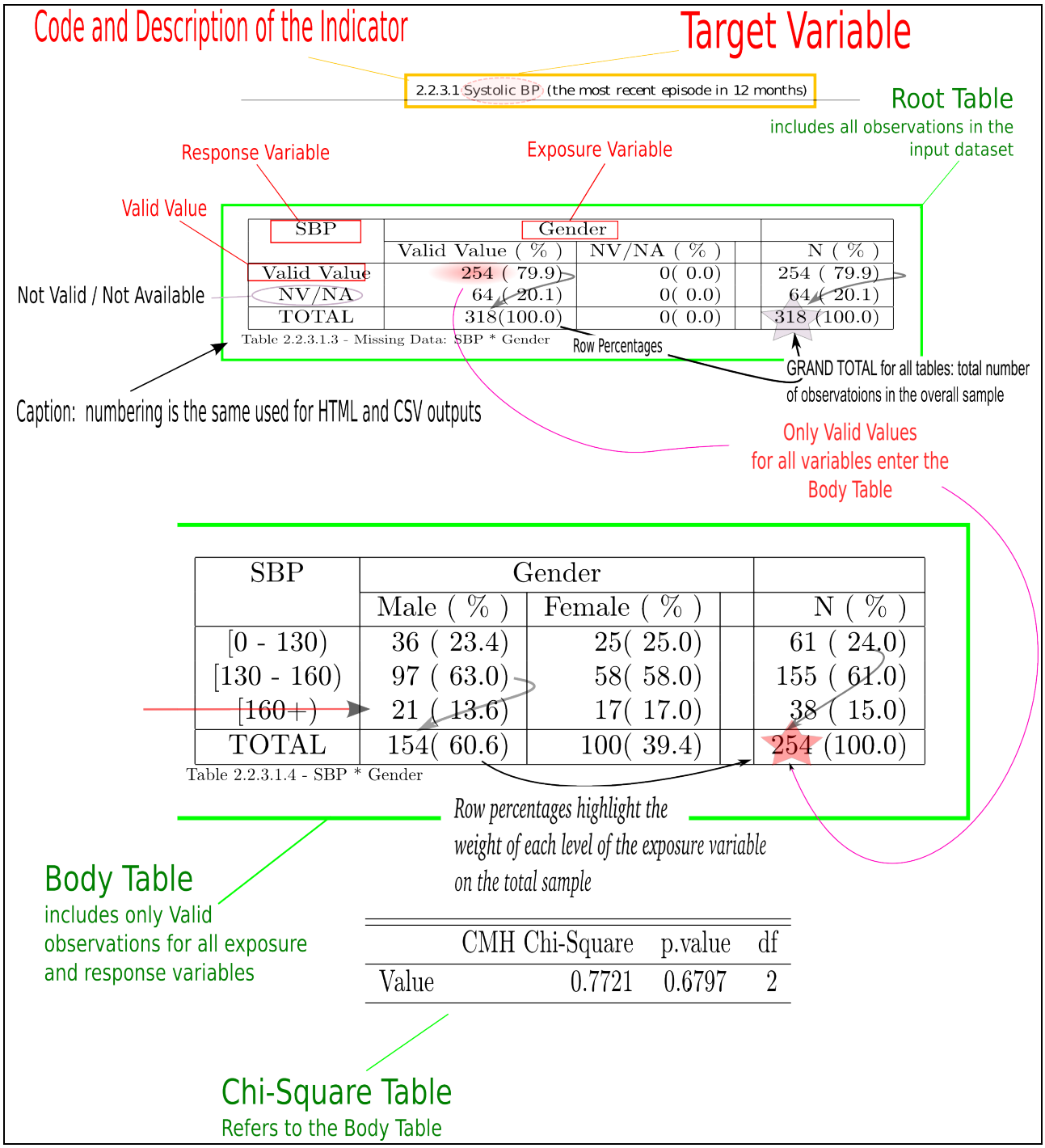


Figure 9. EUBIROD Report One Way Table

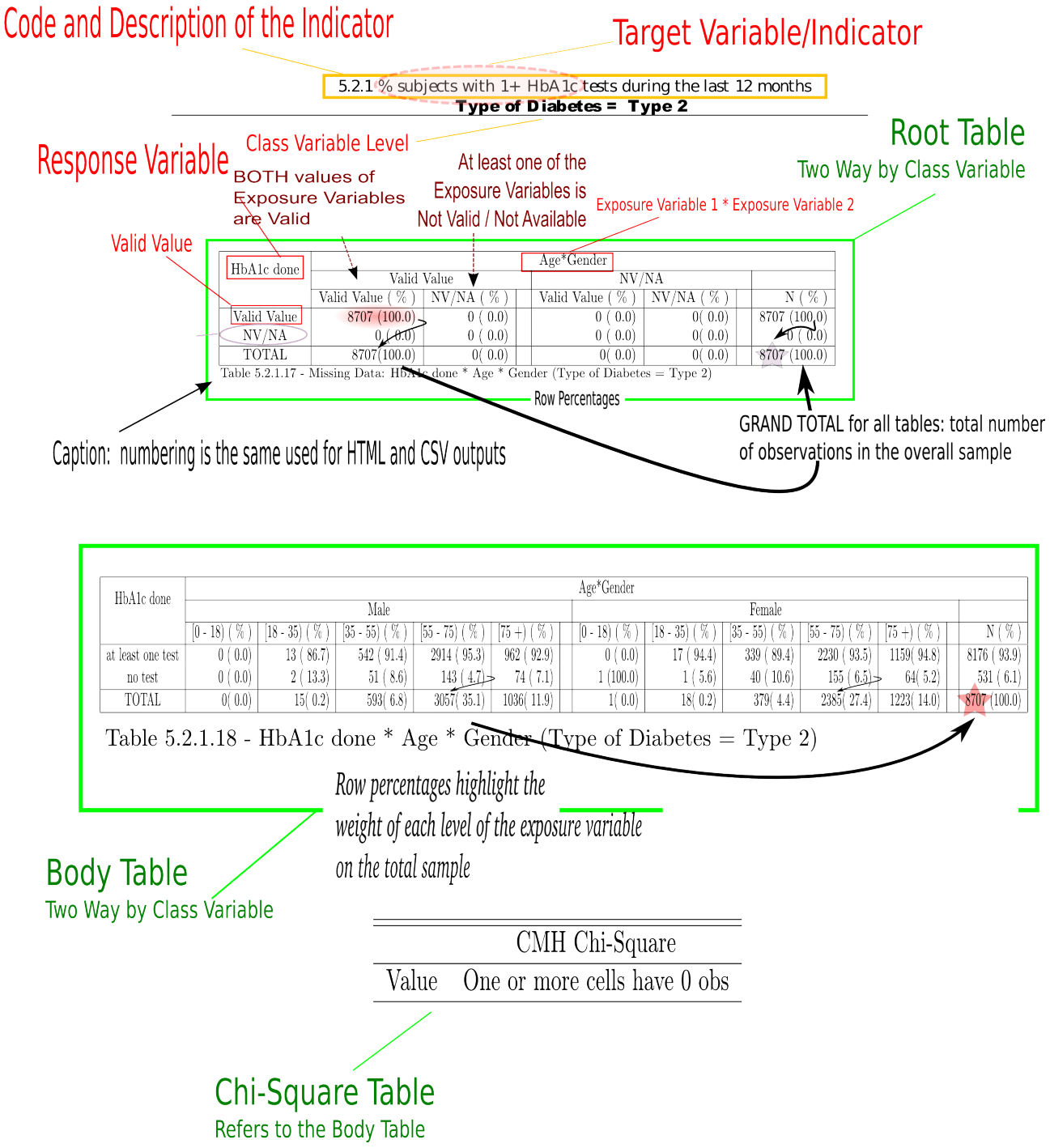


Figure 10. EUBIROD Report Two Way Table

3.3 Graphical Outputs

The range of graphical outputs that are currently produced by the EUBIROD statistical routines are shown in **Figure 11-15**.

As for the tabular outputs, the target indicator is clearly labelled in the header of each page, with a unique identification code and the associated description. The particular level of the class variable (Type of Diabetes), if relevant, is also included at the top of each page for the convenience of the reader.

Graphs included in the outputs are barplots, optionally stratified by levels of the sub source, boxplots, and trellis plots.

Barplots provide an efficient graphical display of all the frequencies presented in the tabular outputs, for all combination of exposure factors and each level of the class variable and/or sub source unit.

Boxplots and trellis plots are only produced for continuous outcome. Boxplots can be conveniently used to assess the variability of the distribution of a continuous outcome (e.g. weight, BMI, systolic blood pressure) across different levels of the exposure factors, class variable, or sub source unit.

Trellis plots include histograms and boxplots and are produced only for indicators presenting continuous outcomes and two exposure factors.

The former present the distribution of categories of continuous outcomes (automatically created by the program) through evenly spaced histograms, whose top value is joined to show an approximated density function. The graph is split into different panels, one for each combination of the level of the class and exposure variables.

Trellis boxplots are created for each panel to display the distribution of the outcome variable for different levels of exposure 1 at each combination of the levels of the class variable and exposure 2.

Standardized estimates are present only for risk adjusted indicators, when sub data sources are included in the analysis.

Graphical outputs include barplots, showing the distribution of standardized values against the overall average, and forest plots, used to represent the percentage of excess/reduction of (O-E)/E % along with their confidence intervals.

Forest plots in particular offer an immediate display of the most significant deviations of any standardized rate from the population average.

Here, only those segments that do not intercept the horizontal central axis (valued zero) have observed rates that are statistically significant from the expected ones. In this case, if the left side of the segment is located at the right of the horizontal axis, then the number of observed outcomes significantly exceeds the expected number, and the related unit is considered “at increased risk” of experiencing a higher rate of outcomes, independently from case-mix factors included in the multivariate model.

On the other hand, if the right end of the segment is located at the left of the horizontal axis, then the number of expected outcomes is significantly lower, and the associated unit is “at reduced risk” of experiencing higher rates of outcomes, compared to the average of the overall population, taking into account case-mix factors included in the logistic model.

Further explanations provided in the figures, also included in the help pages, are self explaining.

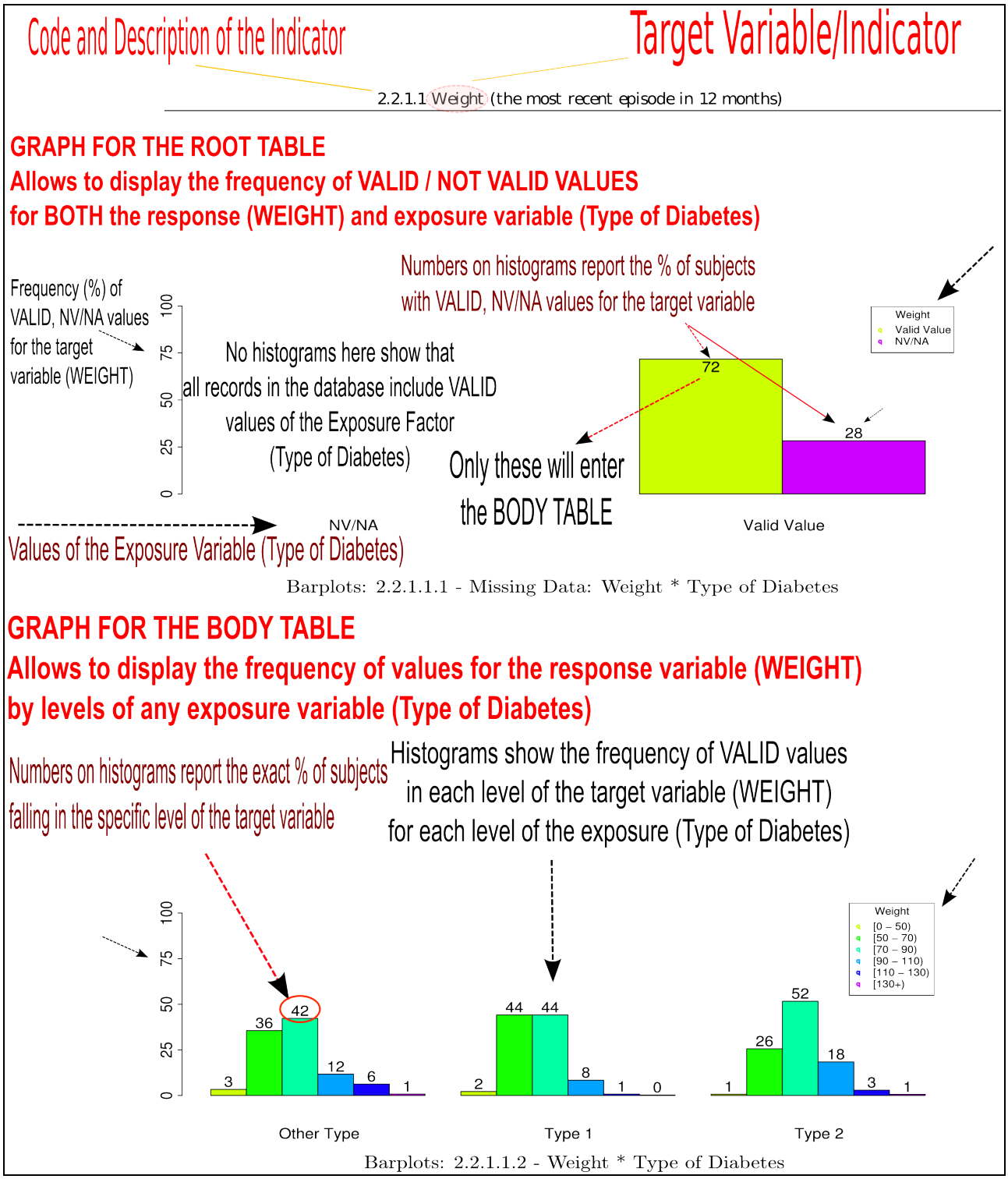


Figure 11. EUBIROD Report One Way Graphs

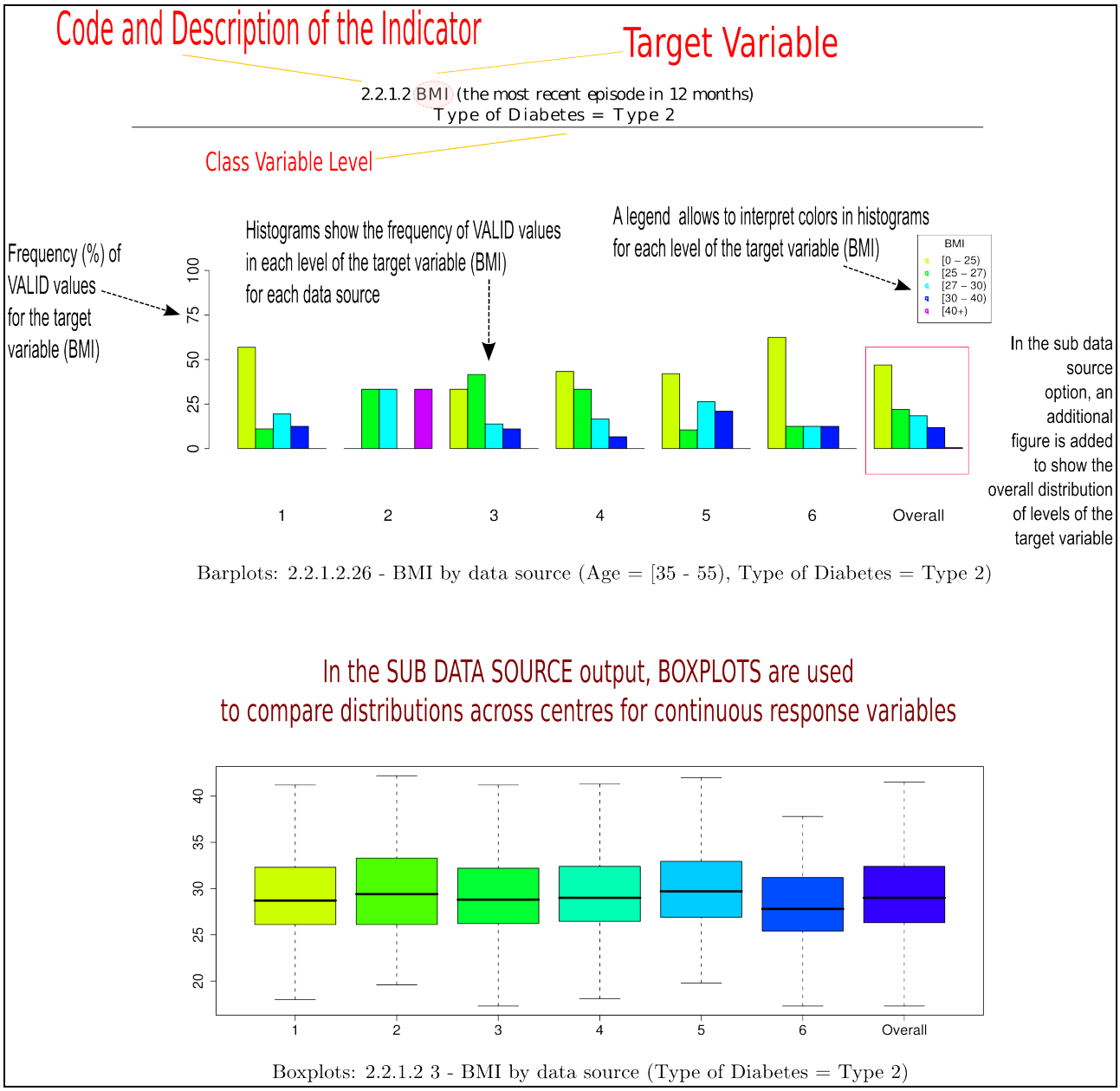


Figure 12. EUBIROD Report Graphs with the option "Sub Data Source"

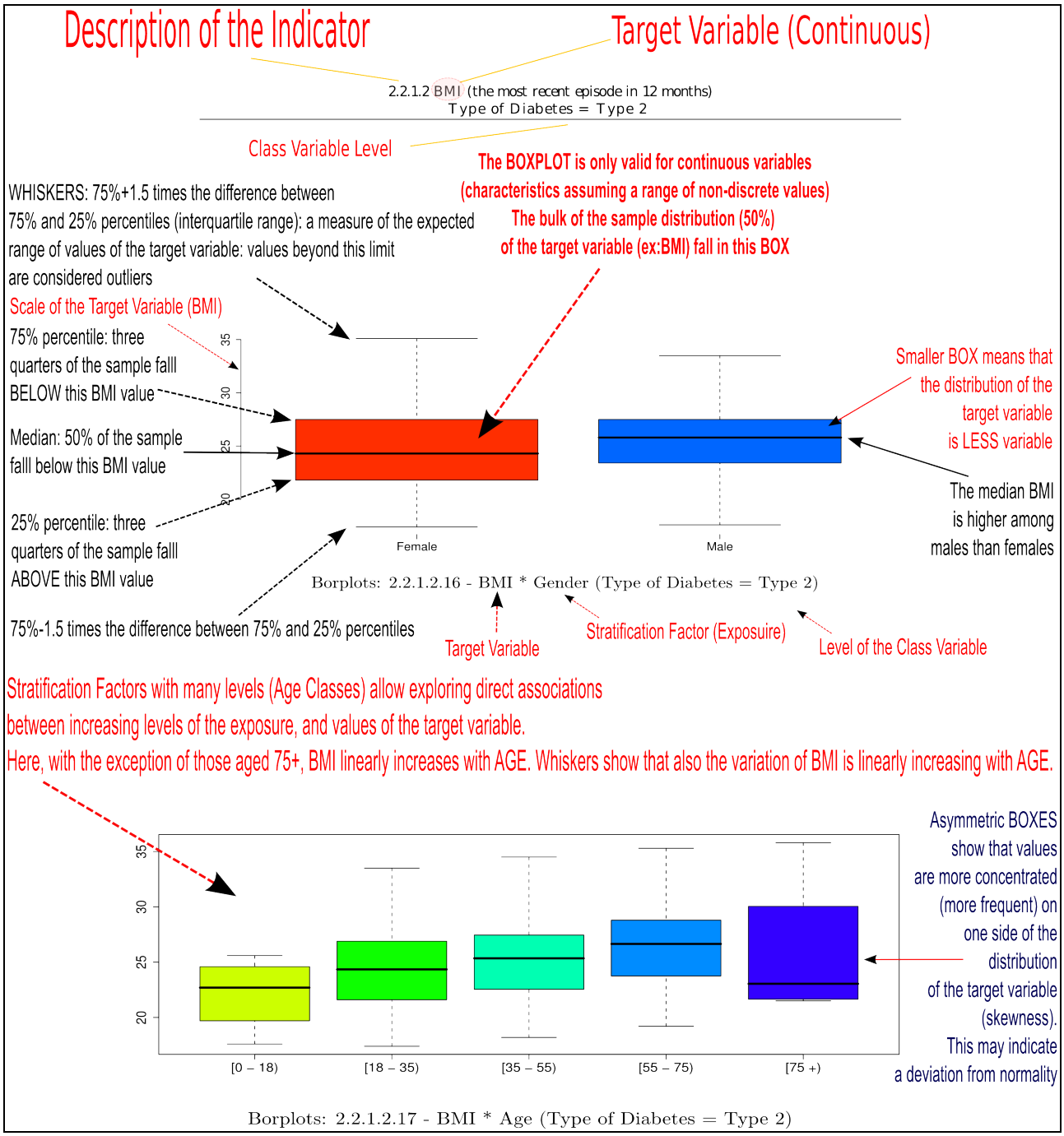


Figure 13. EUBIROD Report Boxplots

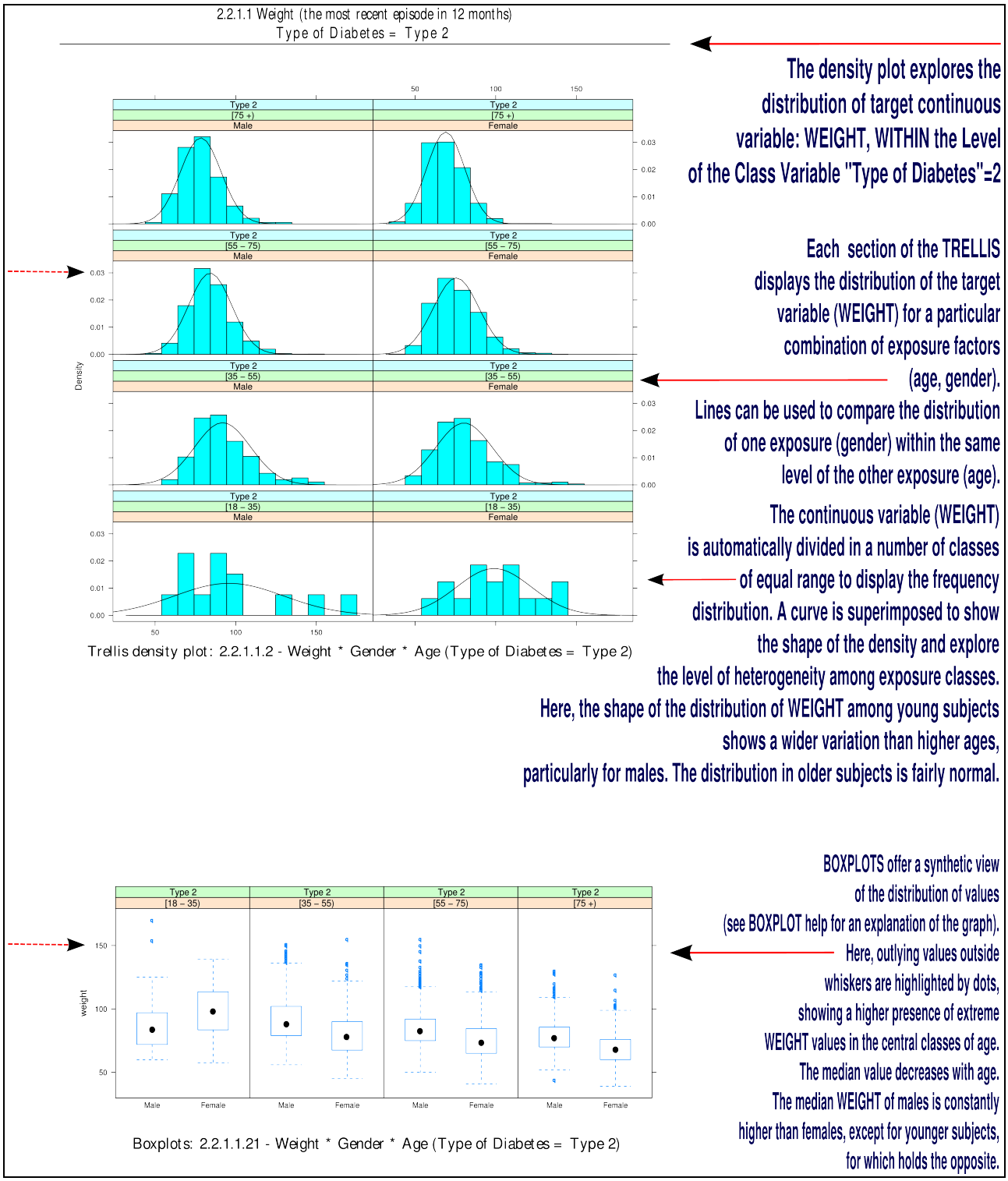
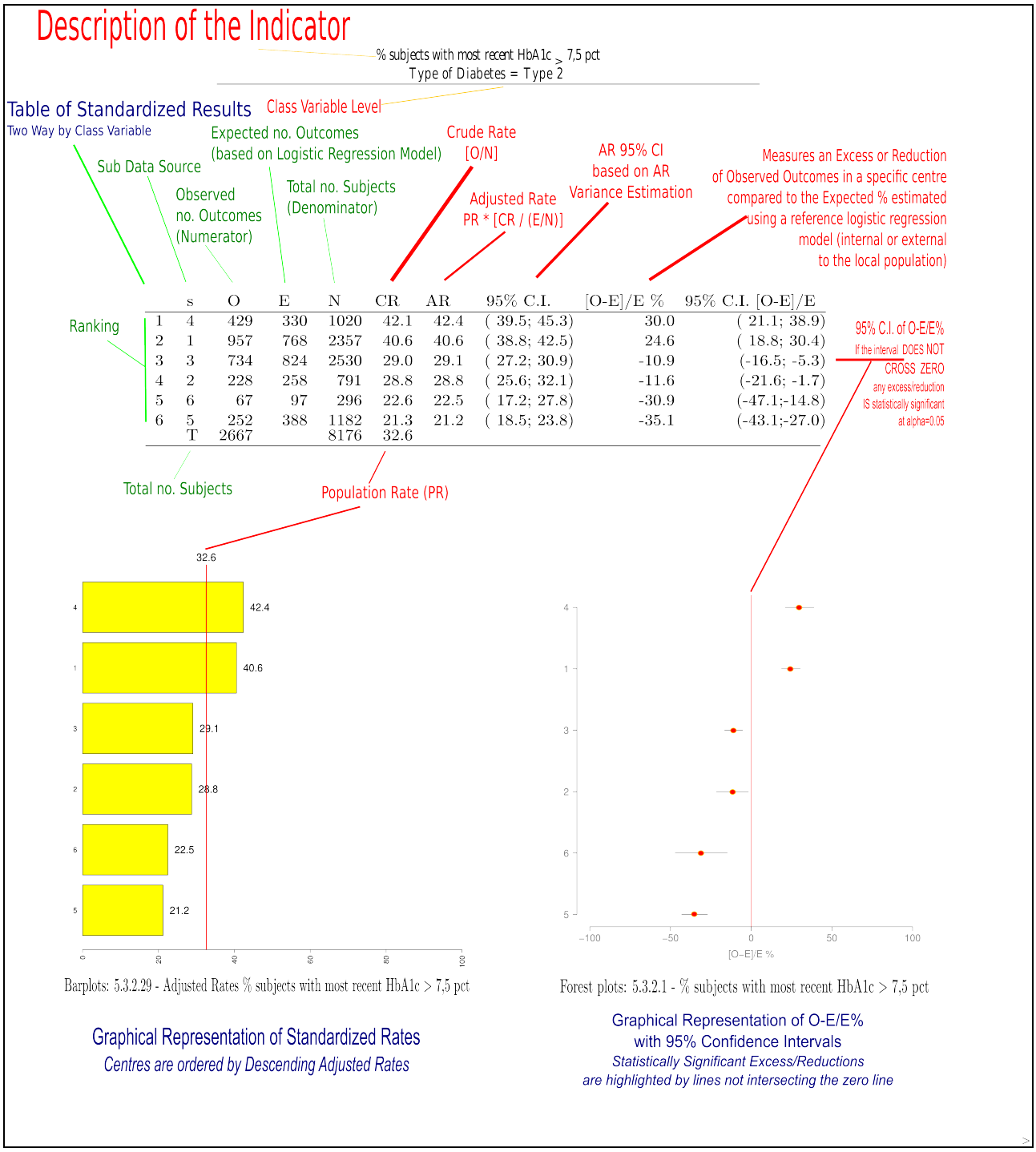


Figure 14. EUBIROD Report Trellis Graphs



	s	O	E	N	CR	AR	95% C.I.	[O-E]/E %	95% C.I. [O-E]/E
1	4	429	330	1020	42.1	42.4	(39.5; 45.3)	30.0	(21.1; 38.9)
2	1	957	768	2357	40.6	40.6	(38.8; 42.5)	24.6	(18.8; 30.4)
3	3	734	824	2530	29.0	29.1	(27.2; 30.9)	-10.9	(-16.5; -5.3)
4	2	228	258	791	28.8	28.8	(25.6; 32.1)	-11.6	(-21.6; -1.7)
5	6	67	97	296	22.6	22.5	(17.2; 27.8)	-30.9	(-47.1;-14.8)
6	5	252	388	1182	21.3	21.2	(18.5; 23.8)	-35.1	(-43.1;-27.0)
	T	2667		8176	32.6				

Figure 15. EUBIROD Report Risk Adjusted Outputs

3.4 Format of the main reports and additional outputs

The main reports produced at each run of the statistical and central engine are directly accessible in a sub-folder of the working directory located under _se_/output/reports, where a directory named with the date and time of the start of execution is created (timestamp). The pdf and html reports are saved each time under the relative timestamp directory and subdirectories with the reference year and centre code for the statistical engine, region code for the central engine.

Figure 16 presents a practical case in which the report for Regione Umbria is included under the specified chain of subfolders.

The html report can be directly accessed by double clicking on the main .html file, named with the database name. The browser will display the table of contents listing all indicators. For those indicators that can be computed by the statistical/central engines (based on the availability of the basic variables originally included in the mapping), links will be active and can be opened by clicking on the specific code/description.

An example of an indicator subpage directly accessible through the main html report is displayed in **Figure 17**. This includes a long list of html tables that are also saved for use in the web portal under the “tables” subdirectory.

The above html files can also be particularly useful to find a particular image that can be included in slides/presentations or high quality typographical outputs.

A quick tip to identify a graph of interest for a specific indicator is shown in **Figure 18** through the use of the Firefox browser: by right clicking on a page, and selecting “Page Info”, the user can access a form that includes a “Media” tab. Selecting it would display a list of all graphical files (png, svg) included in the page. When the user clicks on a specific files, a preview will be available. The location will be printed in the form, from which it can be cut/paste in the browser window, or simply used to access the file. A pdf version of the same file (not visualizable in Firefox) will be also available in the same directory.

Figure 19 displays the contents of the graphs directory, which can be quickly navigated using default image viewers as an additional resource to select the most convenient outputs.

An example of a pdf report is displayed in **Figure 20**. This is directly accessible by clicking on the main page.

Statistical objects are all saved in the data directory, also located under the above timestamp sub directory, in the branch _se_/output/data. The aggregate data saved in CSV format can be directly displayed using an ordinary text editor, as in the case shown in **Figure 21**. The CSV files include a transparent definition of all the variables in the first row, using the normal conventions adopted for this type of files.

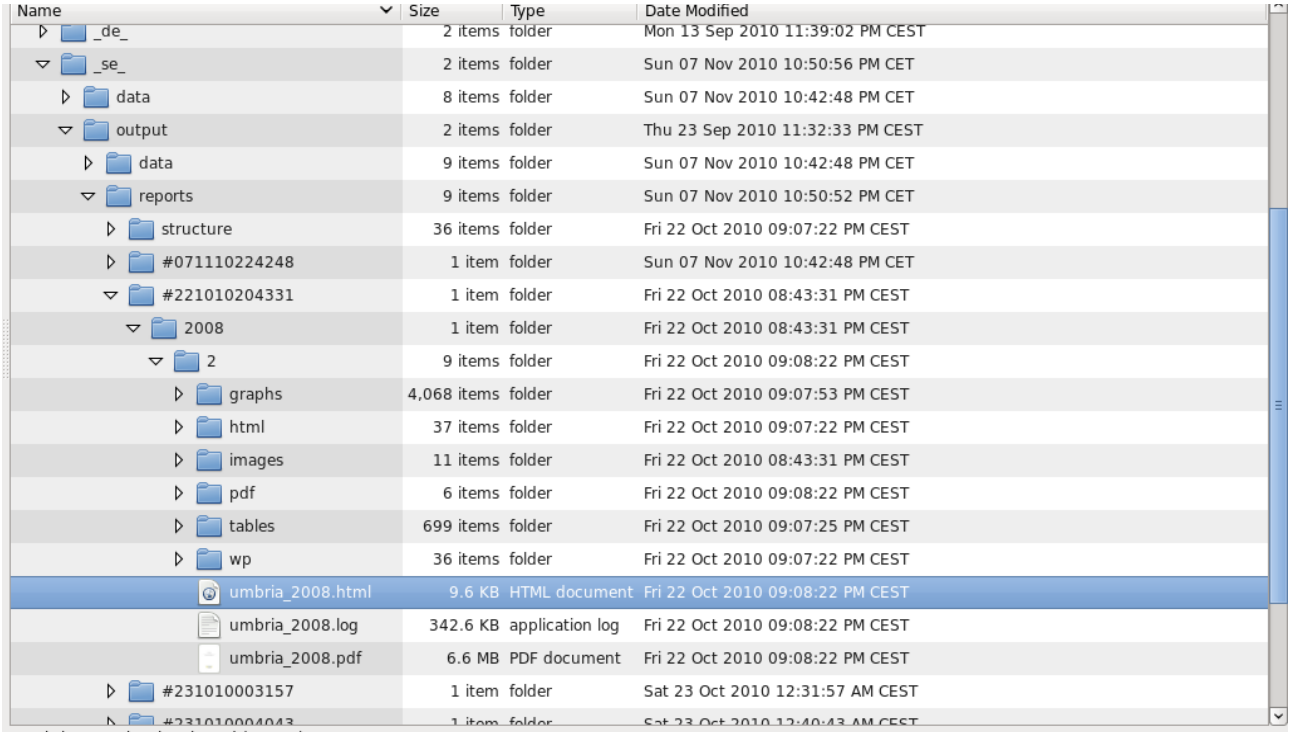
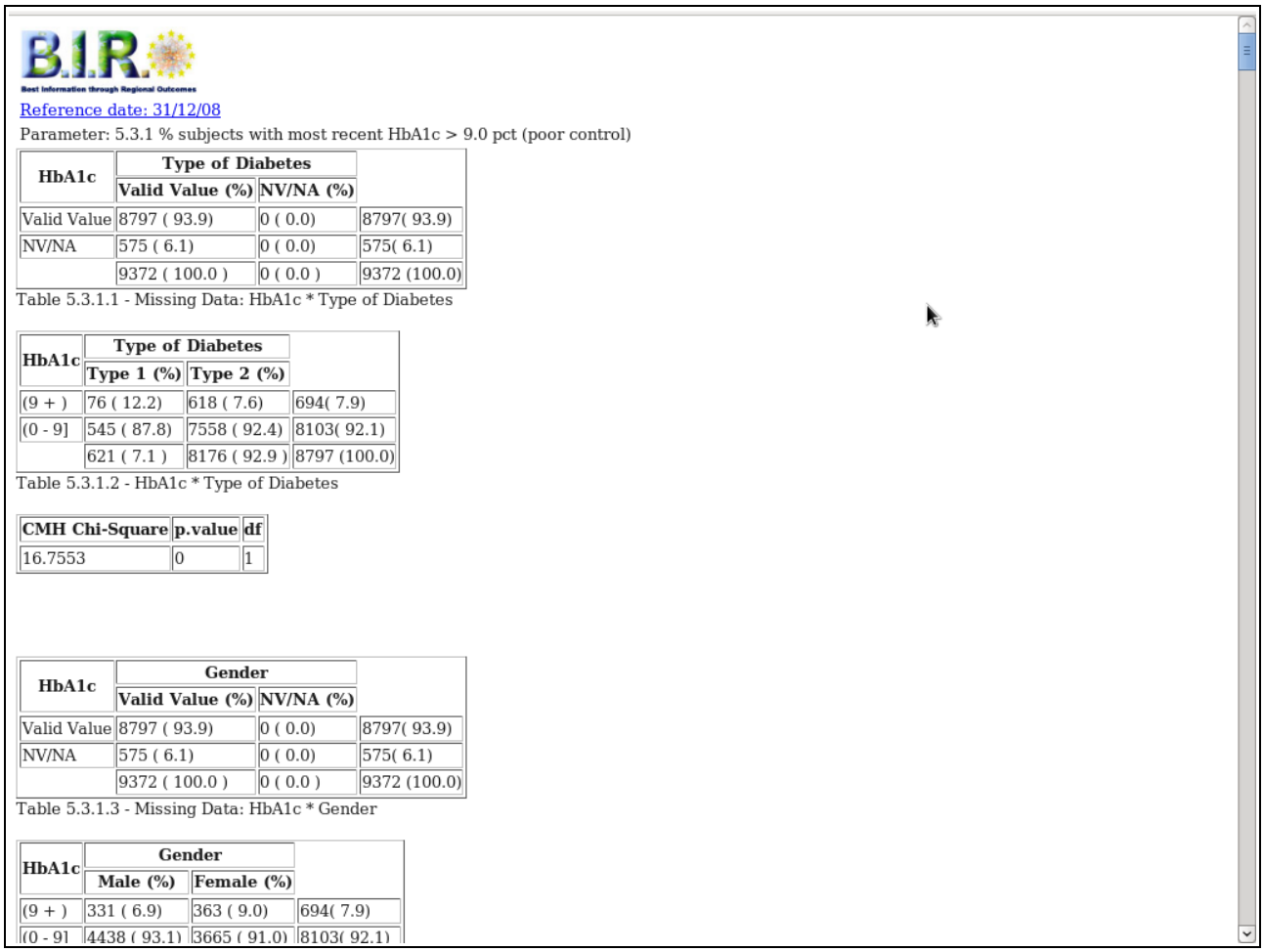


Figure 16. Selecting Outputs from the EUBIROD Output Directory



B.I.R.
Best Information through Regional Outcomes

Reference date: 31/12/08

Parameter: 5.3.1 % subjects with most recent HbA1c > 9.0 pct (poor control)

HbA1c	Type of Diabetes		
	Valid Value (%)	NV/NA (%)	
Valid Value	8797 (93.9)	0 (0.0)	8797(93.9)
NV/NA	575 (6.1)	0 (0.0)	575(6.1)
	9372 (100.0)	0 (0.0)	9372 (100.0)

Table 5.3.1.1 - Missing Data: HbA1c * Type of Diabetes

HbA1c	Type of Diabetes		
	Type 1 (%)	Type 2 (%)	
(9 +)	76 (12.2)	618 (7.6)	694(7.9)
(0 - 9]	545 (87.8)	7558 (92.4)	8103(92.1)
	621 (7.1)	8176 (92.9)	8797 (100.0)

Table 5.3.1.2 - HbA1c * Type of Diabetes

CMH Chi-Square	p.value	df
16.7553	0	1

HbA1c	Gender		
	Valid Value (%)	NV/NA (%)	
Valid Value	8797 (93.9)	0 (0.0)	8797(93.9)
NV/NA	575 (6.1)	0 (0.0)	575(6.1)
	9372 (100.0)	0 (0.0)	9372 (100.0)

Table 5.3.1.3 - Missing Data: HbA1c * Gender

HbA1c	Gender		
	Male (%)	Female (%)	
(9 +)	331 (6.9)	363 (9.0)	694(7.9)
(0 - 9]	4438 (93.1)	3665 (91.0)	8103(92.1)

Figure 17. Opening the EUBIROD HTML Report

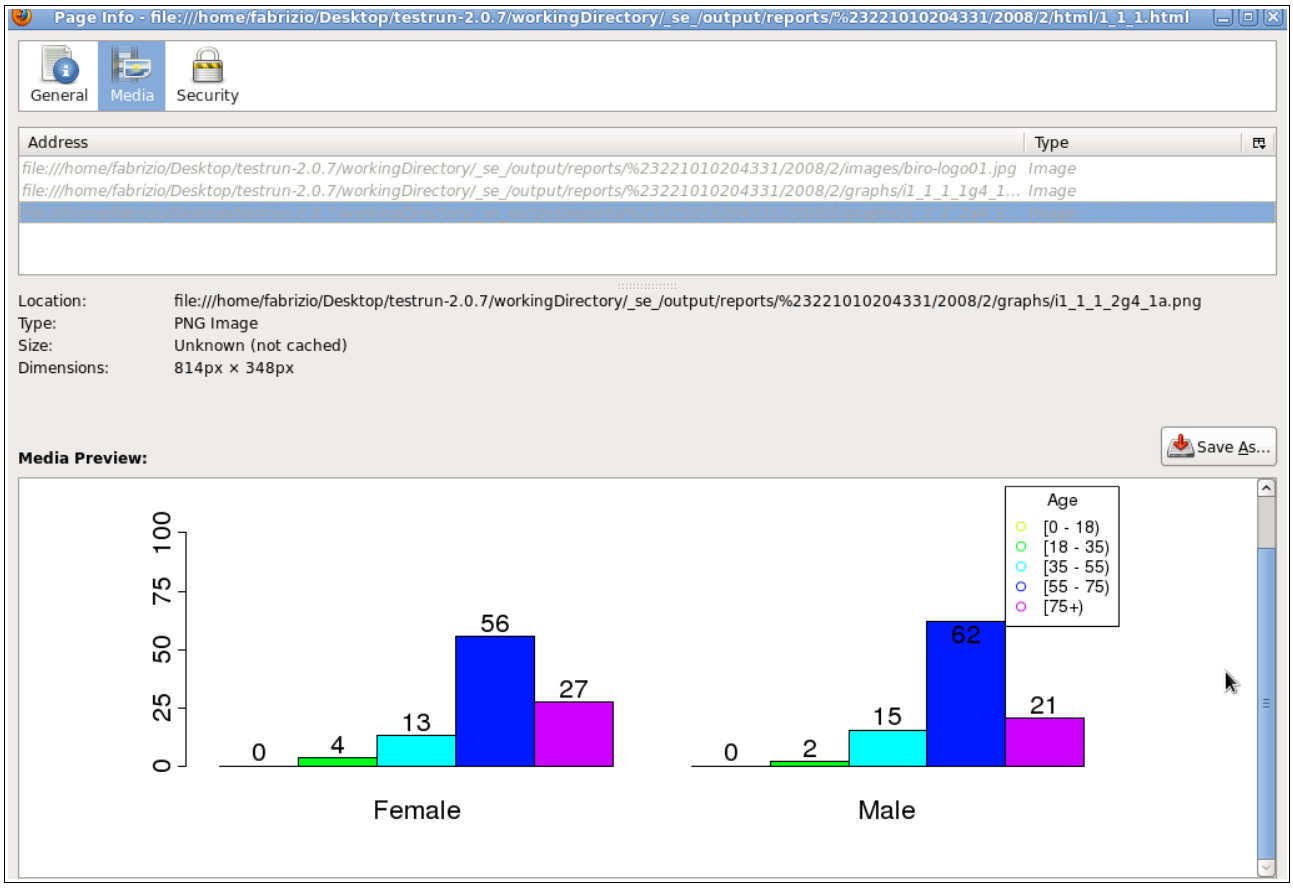


Figure 18. Using Firefox to Browse EUBIROD graphs in the HTML Report

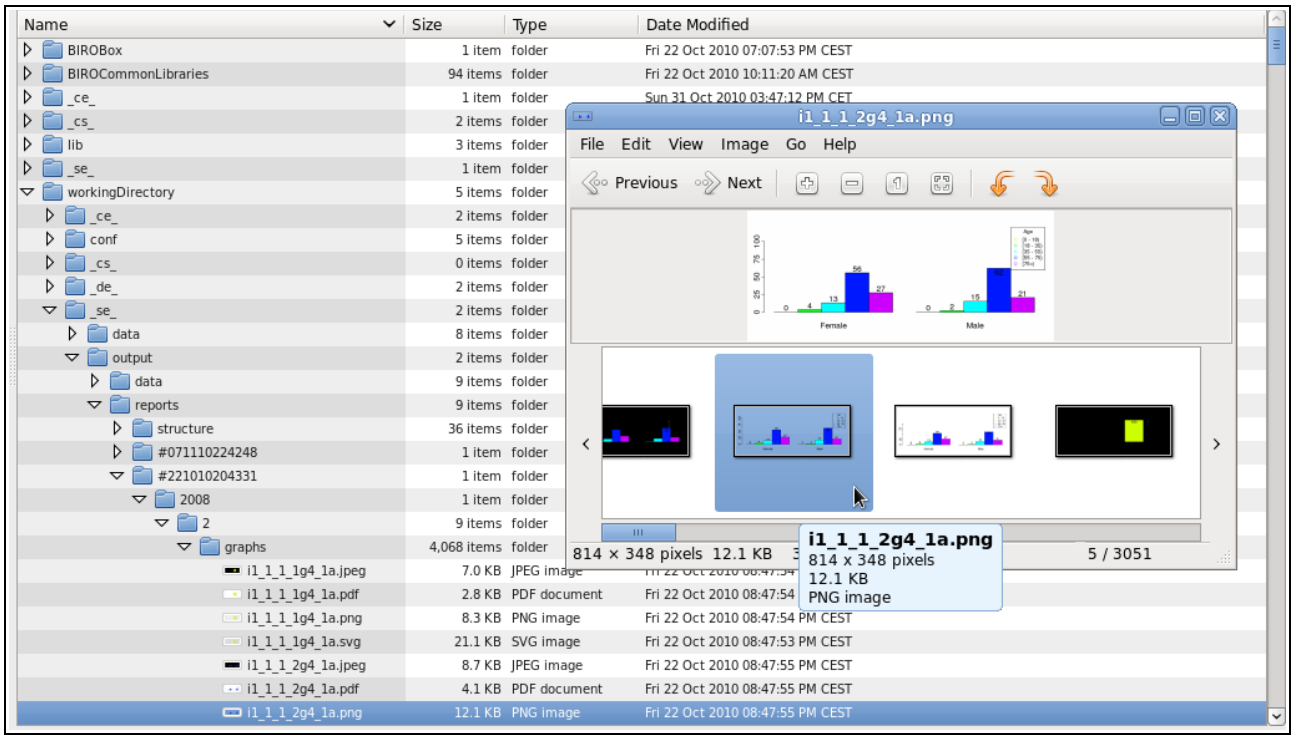


Figure 19. Selecting Images firectly from the EUBIROD Graphs Directory

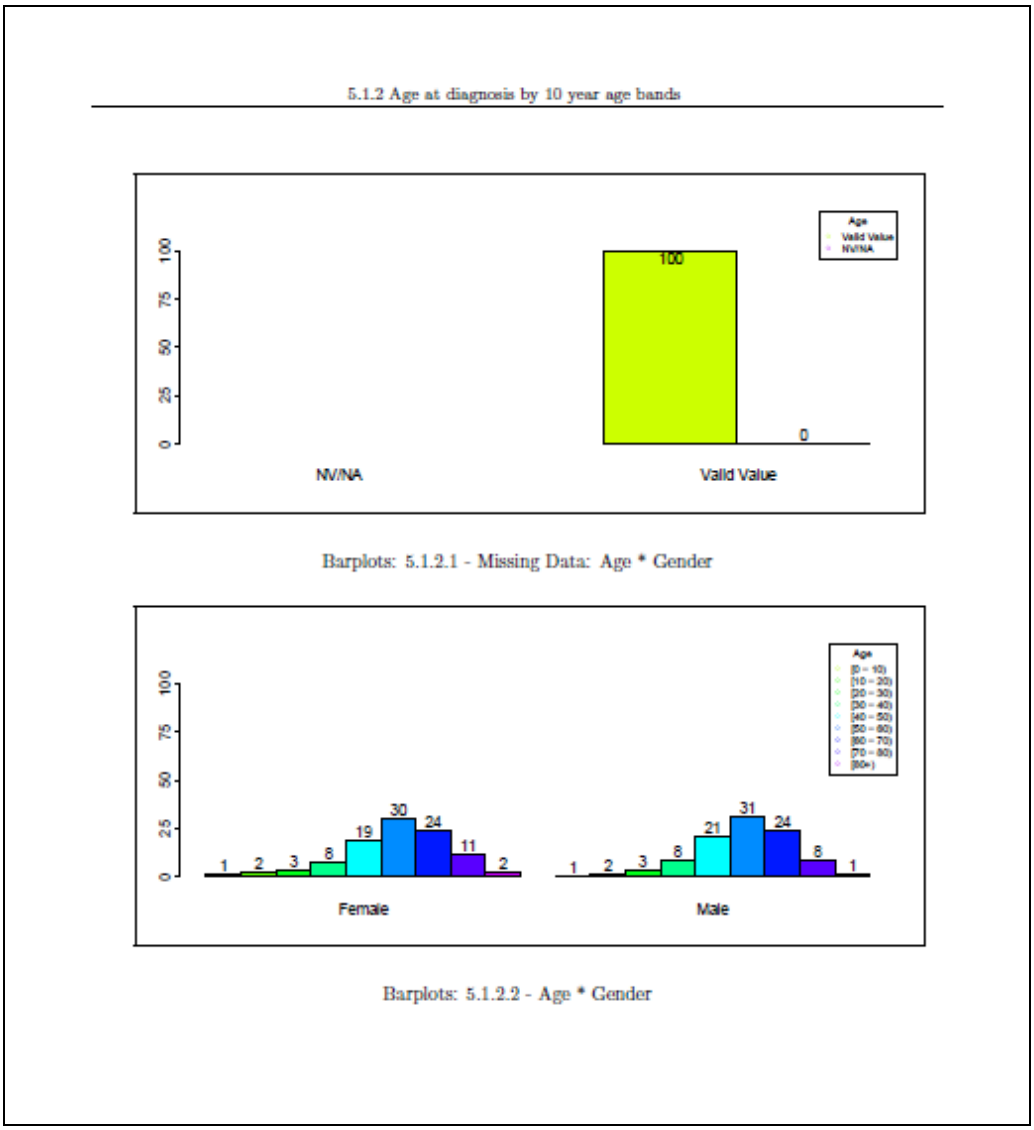


Figure 20. Opening the EUBIROD PDF Report

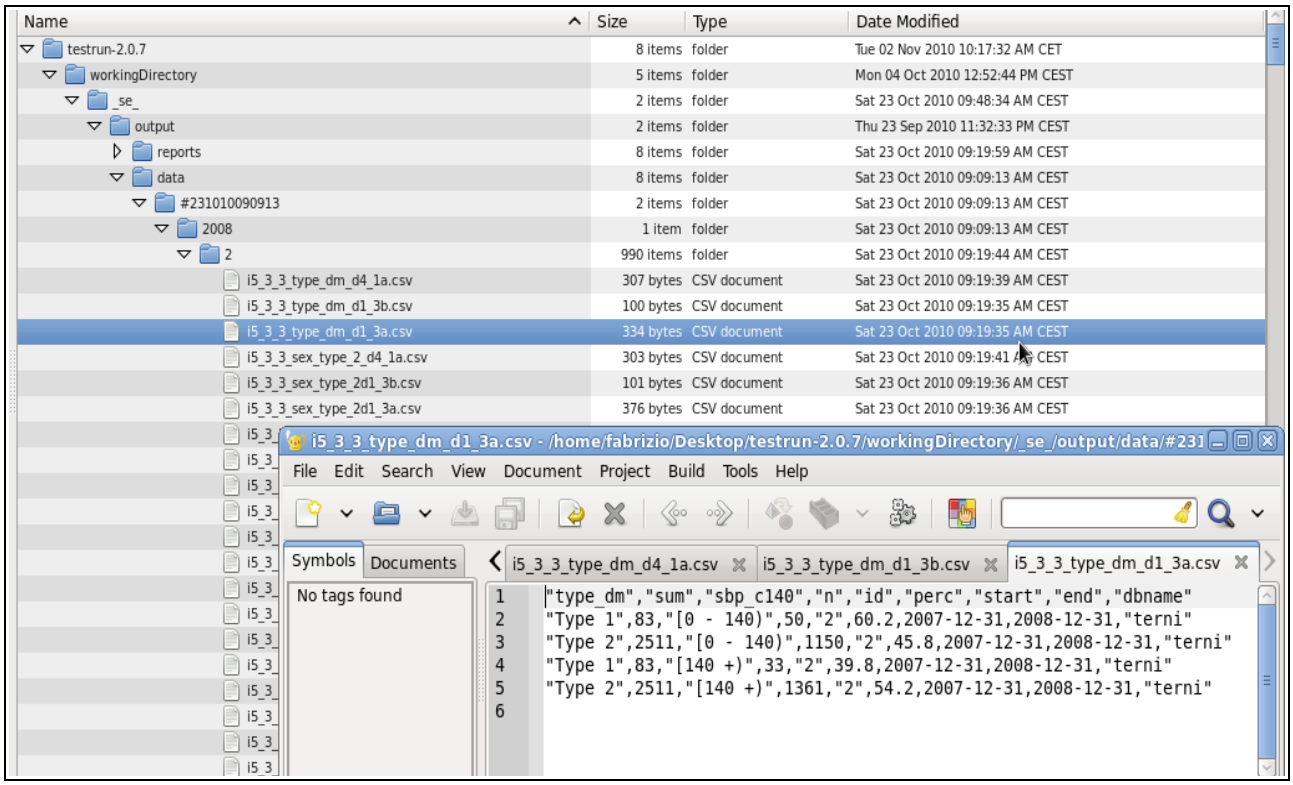


Figure 21. Browsing the EUBIROD Output Data Directory

3.5 Software/Hardware specifications and performance

Statistical engine has been successfully developed without noticeable deviations from the original plan and has been successfully tested on Fedora Linux 13 running on Oracle Virtual Box.

Hardware consisted of average Intel-based PCs/Notebooks, the least powerful with the following specifications: CPU speed 2.0GhZ, hard disk capacity of 8Gb + 1Gb RAM dedicated to the VirtualBox.

Excution times from a test run on data from the Umbria register for the production of the 2008 *local report* are shown in Box 7.

Centre	N Patients	N episodes	Elapsed Time
1	2,842	9.097	10' 46″
2	3,202	8,316	9' 23″
3	1,115	1,948	8' 26″
4	1,268	1,456	8' 17″
5	994	1,329	8' 02″
6	318	438	8' 19″
Overall (Statistical Engine)	9,739	22,584	24' 52″
Overall (Central Engine)	9,739	22,584	15' 30″

Box 7. Execution Times for different EUBIRO test analyses

Execution times show to be reasonable compared to the amount of information that is now contained in each report. The substantial increase in execution times compared to the initial BIRO version (almost double) can be easily explained by the additional options made available, including the fine stratification of tables and graphs, and the application of standardization models in risk adjusted indicators.

As a matter of fact, outputs occupy an average storage space of about 70Mb for the overall analysis, including data to be transmitted to the central server, an amount that is more than double of the previous output delivered by the BIRO project.

The above timings confirm the superiority of the distributed analysis over the centralized approach: delivering the overall report using the central engine takes half of the time of a “traditional” running the engine on top of the whole dataset.

By the way, to realize such potential, it is necessary to process all data sources independently as a first step. In real life conditions, this step is performed by different computers, so that the total time required to carry out the operation is equal to the highest execution time recorded among the different sites, instead of the sum of the different runs on the same machine. This approach has also the advantage of automatically and autonomously producing all reports for each site, simultaneously creating the objects required by the central engine to produce the overall report.

Installation of the software is identical regardless of the hardware, and requires Rv.2.8, Latex, Java 6.0 and PostgreSQL, plus various additional libraries/packages that are included in the distribution.

Software is released using the GPL license and is authored by F.Carinci and L.Rossi.

4. Discussion

The evolution of health services research, statistical methods, and information technology allows policy makers to be constantly supported by the information contained in massive administrative databases and disease registries[22]. Epidemiological analysis is made possible on a routine basis, accessing sources that are in constant evolution. Such databases are usually enforced by national/regional legislation for reasons of disease surveillance, quality of care evaluation and control of health expenditure.

The data quality of these sources must be carefully controlled[23]. However, it is undeniable that administrative data and disease registries are rapidly becoming the major information platform for the advanced analysis of health systems. In many countries, they are used as the building blocks of multidimensional frameworks for performance evaluation and provide the key evidence-based recommendations to policy makers for health reform[24].

Access to routine databases allows linking datasets at the subject level, ordinarily without explicit patient consent, making possible to carefully control for data quality. This approach allows checking for double counts and excluding those who have died or emigrated from the denominators of indicators. Therefore, more precise and unbiased results can be obtained[25]. Data linkage involves access to an updated list of personal identifiers, which can lead to the identification of high spending and high-risk groups, allowing analysts to look at repeated services and to improve the precision of all estimates at population level.

In many areas of interest to the European Commission, target information to compute health indicators is still not standardized across Europe and dispersed across different classes of users and different data administrators. By definition, some of these hurdles may not be overcome, as databases are naturally gathered as a result of the provision of services taking place at different settings.

However, there is a need for solutions allowing to connect all efforts and linking those databases by automatic means, so that the analyst may avoid to construct an “ad hoc” database to perform the statistical analysis and can directly use the data that are already collected at the level of each single source.

The real innovation achieved by the EUBIROD statistical materials, together with all the other tools realized by the EUBIROD Consortium, is to prove that it is possible to conduct a rigorous epidemiological analysis at the European level, by disseminating a common set of routines that can be used by many partners simultaneously and independently.

In fact, the range of statistical outputs that can be produced by the BIRO system is not innovative *per se*, as any user can acquire commercial software that is widely available and can deliver outputs well beyond the capabilities of BIRO, in a very efficient and reliable manner.

The advantage offered by the BIRO statistical engines is that they have been specifically designed to deliver European Health Indicators and they can be distributed without licensing bindings to an unlimited group of users who can deliberately agree to share a common standard. These fundamental routines form a standardized platform that can be further modified to comply with the user needs, including the possibility to deal with different problems and extend the approach to other diseases. Most importantly, the processing workload is distributed by definition across a multitude of users, rather than concentrated at a unique institution, which may pose serious problems of sustainability and capacity with an increasing number of users. The open source approach allows for the core statistical routines used for the calculation of accurate indicators to be adopted by

institutions with limited resources and poor technical skills. Such opportunity offers also the key added value of enhancing the capacity of training and providing education in the field of information for policy, an area that the EUBIROD project aims to further develop through the activity of the "BIRO Academy".

The EUBIROD statistical engines has been progressively developed after several rounds of tests conducted in real life conditions on databases maintained in 20 different countries. The authors have collected all suggestions and subsequently adapted the software to respond to all indications, in close collaboration with the EUBIROD core development team. This way the statistical routines have been efficiently integrated with all the other tools that are now part of the BIRO system.

The advancements embedded in the statistical and central engines allow fostering the epidemiological analysis of **diabetes data**, by adding the following features:

- multidimensional tables, stratified by two exposure factors and one outcome variable. This feature allows fine comparisons e.g. calculation of relative risks of the outcome across different levels of exposure factors.
- revised structure of the BIRO report allowing stratification of results by centre at the local level. This feature delivers finely stratified reports in which all indicators are displayed by sub source (clinical centre or unit) within each local register (ex.: Austria can benchmark differences in diabetes indicators between centres in the region of Styria).
- for each parameter/indicator, a root table displays the frequencies of missing vs valid values
- stratification of all results by a class variable (Type of Diabetes)
- graphical displays of all stratification levels
- unique coding structure for all outputs delivered as html tables, image files, and CSV data. This feature allows to easily reuse all objects for presentations, to dynamically link results to the BIRO web portal, or to feed other online repositories (e.g. the DG-SANCO health information platform "HEIDI")
- revised pdf report including cover pages providing information on the EUBIROD Consortium and explanatory figures as help files
- same outputs realized for the statistical engine applied to the central engine
- recursive application of the central engine. This feature enables each user to pool statistical objects obtained from both the analysis of individual and aggregate data (ex.: different centres of Germany can deliver reports to an institution acting as national coordinator. Such entity can compile the national report using the central engine, then send the results to the Coordinating Centre, which can use them again using the central engine to produce the European Report).
- storage of all outputs in a directory selected to the user, compliant with the BIROX distribution.

The statistical routines are made easily accessible to the average BIRO user through an enhanced version of the BIROBox interface.

Further developments are needed to regulate the flow of information across a network of users who may independently analyse data using both the statstical and the central engine.

In particular, precise specifications are required to define strict rules that would assign a unique code to each centre adopting BIRO. This way the BIRO system may be applied in a recursive fashion, expanding the range of its applications.

Currently the statistical routines may be used totally independently according to the user needs. However, triggering the process of distribution/exchange of all aggregate tables across the network may generate uncontrollable errors in the way the global report is compiled. If not adequately organized, it would be almost impossible to keep track of all data sources involved in the international exchange.

To this end, it is necessary to identify a solution that would allow building a general register of sources and establish a hierarchy in the distribution of the statistical objects.

This way, the software would be able to recognize the list of sources involved in the calculation, attributing each indicator to a very well defined set of contributors. Such feature, although not initially foreseen, has been recognized as an essential element to ensure the integrity of the global report.

A second reason to implement the unique coding and regulate the flow of analysis is related to the fact that the central engine can be now applied recursively, so that analyses performed on top of individual data by the statistical engine can be further compiled by the central engine to generate aggregate results. The need of such possibility has become clear by an assessment of the practical conditions found on field. In fact, there are situations where the direct processing of all individual data can be undertaken by different istitutions, but the aggregate data are not allowed to be sent separately directly to the European level because the participating institutions would not allow to do so for internal policy.

In such situations, aggregate tables can be amalgamated by one or more national coordinators prior to the transmission to the European level, where the global report would continue to be built using an additional instance of the central engine.

Therefore, it is definitely crucial that the central engine is made available to each BIRO user rather than only to the EU server administrator. To allow this by maintaining an internal consistency, it will be necessary to develop an appropriate strategy that would most likely involved geographical referencing and an explicit registration of the hierarchy of relationships. An additional challenge relates to the precise regulation of the time interval in which these relations occur, so that the global report would have a unique reference.

For the above reasons, the development of geographical maps and longitudinal trends has been postponed, as these are strictly related to a precise and unique reference of the data sources and the time involved in the analysis.

A further, unplanned update of the statistical materials including such improvements is foreseen in the last year of the EUBIROD project.

5. Conclusions

The set of routines realized for the EUBIROD project provide a flexible solution to set the basis for continuous monitoring of diabetes across Europe. The statistical reports include basic figures that can be helpful to benchmark quality and outcomes and can also significantly enhance the average capacity of all centres to increase the quality of their information, sharing the analysis of standardized information.

The open source availability of a targeted set of statistical routines can be particularly relevant for those users e.g. diabetes register administrators, who maintain large databases but until now have neither exchanged data with international peers, nor used common tools for epidemiological analysis.

The range of outputs delivered by the EUBIROD statistical engines may be exploited to build flexible EU platforms that would automatically tap into regional/national databases to gather and immediately deliver public health information according to a standardized format.

References

1 DG SANCO Task Force of Major and Chronic Diseases, Major and Chronic diseases in the European Union - Report 2007, European Commission, Luxembourg, 2008, available at: http://ec.europa.eu/health/ph_threats/non_com/docs/mcd_report_en.pdf
2 The EUGLOREH Consortium, The status of health in the European Union: towards a healthier Europe, Chapter: “Diabetes”; available at: http://www.intratext.com/ixt/_EXT-rep/_INDEX.HTM
3 International Diabetes Federation, IDF Diabetes Atlas 4th edition, 2009.
4 Khan L, Mincemoyer S, Gabbay RA. (2009), Diabetes registries: where we are and where are we headed?, Diabetes Technol Ther. 2009 Apr;11(4):255-62.
5 Joshy G, Simmons D. (2006) Diabetes information systems: a rapidly emerging support for diabetes surveillance and care, Diabetes Technol Ther. 2006 Oct;8(5):587-97.
6 Carinci F, Orsini Federici M, Massi Benedetti M (2006), Diabetes registers and prevention strategies: towards an active use of health information, Diab Res Clin Pract 2006; 74:S215-S219.
7 Morris AD, Boyle DIR, MacAlpine R et al (1997), The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. BMJ 315:524–528.
8 Raymond ES, The Cathedral and the Bazaar, O'Reilly Publications, 1999.
9 Di Iorio CT et al, EUBIROD Privacy Impact Assessment, available at: http://www.eubirod.eu/documents/downloads/D5_2_Privacy_Impact_Assessment.pdf
10 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal of the European Communities No. L 281/31; available at: http://ec.europa.eu/justice_home/fsj/privacy/law/index_en.htm
11 Council of Europe. Convention for the protection of individuals with regard to automatic processing of personal data. Strasbourg: The Council, 1981; available at: http://conventions.coe.int/Treaty/en/Treaties/Html/108.htm
12 Concil of Europe, Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocol No. 11, Rome (1950); available at: http://www.echr.coe.int/nr/rdonlyres/d5cc24a7-dc13-4318-b457-5c9014916d7a/0/englishanglais.pdf
13 Best Information through Regional Outcomes (BIRO) official website, available at: http://www.biro-project.eu
14 BIRO Consortium (2009), Best information through regional outcomes: a shared European diabetes information system for policy and practice, Università di Perugia, Perugia, Italia, available at: http://www.eubirod.eu/documents/downloads/BIRO_Monograph.pdf
15 Di Iorio CT, Carinci F, Azzopardi J, Baglioni V, Beck P, Cunningham S, Evripidou A, Leese G, Loevaas KF, Olympios G, Orsini Federici M, Pruna S, Palladino P, Skeie S, Taverner P, Traynor V, Massi Benedetti M (2009) Privacy impact assessment in the design of transnational public health information systems: the BIRO project, Journal of Medical Ethics, Dec;35(12):753-61.
16 R Development Core Team (2010), R: A Language and Environment for Statistical Computing, available at: http://cran.r-project.org/doc/manuals/refman.pdf
17 BIRO Common Dataset, available at: http://www.biro-project.eu/documents/downloads/D3_1_Common_Dataset_v1_7.pdf http://www.biro-project.eu/documents/downloads/D4_3_Dictionary_XML_Update_v1_0.pdf
18 BIRO Report Template, available at: http://www.biro-project.eu/documents/downloads/D7_1%20Reports%20Template.pdf http://www.biro-project.eu/documents/downloads/D7_2_Reports_Template_Update_XML_Metadata_Reports.pdf

19 EUropean Best Information through Regional Outcomes in Diabetes (EUBIROD); available at: http://www.eubirod.eu/
20 Council conclusions on promotion of healthy lifestyles and prevention of Type 2 diabetes (2006/C 147/01); available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2006:147:0001:0004:EN:PDF
21 AHRQ Quality Indicators, Guide to Inpatient Quality Indicators, Quality of Care in Hospitals. Volume, Mortality, and Utilization, Department of Health and Human Services, Agency for Healthcare Research and Quality, June 2002, Version 3.1, March 12, 2007, available at: http://qualityindicators.ahrq.gov/downloads/iqi/iqi_guide_v31.pdf
22 Roos LL, Menec V, Currie RJ. Policy analysis in an information-rich environment, Soc Sci Med. 2004 Jun;58(11):2231-41
23 Roos LL, Gupta S, Soodeen RA, Jebamani L. Data quality in an information-rich environment: Canada as an example, Can J Aging. 2005 Spring;24 Suppl 1:153-70
24 Holman CD, Bass AJ, Rosman DL, Smith MB, Semmens JB, Glasson EJ, Brook EL, Trutwein B, Rouse IL, Watson CR, de Klerk NH, Stanley FJ, A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system, Aust Health Rev. 2008 Nov;32(4):766-77.
25 Ingelfinger J, Drazen J. Registry research and Medical Privacy. N Engl J Med, 2004;350:1452–53