BIFAP DATA ACCESS GOVERNANCE

Version 2.
February 2021

EMAIL

fvigilancia@aemps.es

C/ CAMPEZO, 1 – BUILDING 8
28022 MADRID
TEL: 91 822 50 73
FAX: 91 822 51 61

Content

Abbreviations

AEMPS	Spanish Agency of Medicines and Medical Devices
AEPD	Spanish Data Protection Agency
BIFAP	Pharmacoepidemiological Research in Primary Care Database
CA, CCAA	Autonomous Community, Autonomous Communities
CPD	BIFAP Data Centre
ICD-9-MC	International Classification of Diseases, Ninth Revision, Clinical Modification
ICD-10	International Classification of Diseases, Tenth Revision
ICPC-2	International Classification of Primary Care
CIAS	Healthcare Identification Code for the Autonomous Community
CIPA	Personal Identification Code for the Autonomous Community
Dataset	Pseudonymised Patient Data file used by the researcher for data analysis.
DPIA	Data Protection Impact Assessment
FHC	Medical Records Filter
GIE	Integral Management of Studies
HCE	Electronic Medical Records
HCE-AP	Primary Care Electronic Medical Records
PI	Principal Investigator
LOPDDD	Organic Law 3/2018, of 5 December, on the Personal Data Protection and the Guarantee of Digital Rights
GDPR	General Data Protection Regulation
SNS	National Health System
SNOMED-CT	Systematised Nomenclature of Medicine – Clinical Terms
SFTP	Secure File Transfer Protocol
U.INV	BIFAP Team Research Unit
U.SINF	BIFAP Team Information Systems Unit

1 INTRODUCTION

BIFAP (Pharmacoepidemiological Research in Primary Care Database) is a database of medical records belonging to the National Health System (SNS) aimed at investigating the patterns of use, safety and effectiveness of medicinal products. BIFAP is funded and administered by the Spanish Agency of Medicines and Medical Devices (AEMPS) and draws on the data recorded by family doctors and primary care paediatricians in Electronic Medical Records (HCE-AP) provided by the Autonomous Communities (CCAA) that voluntarily participate through collaboration agreements.

This document describes the flows and the different kinds of processing to which the data is subjected, the management of access to the data, the use of the data by virtue of the user type of the database, and the characteristics of the exploitation of the data. It addresses in particular the requirements and conditions applicable to researchers conducting studies with BIFAP data, from the moment the researcher proposes a study protocol to the moment the results are published, as well as the actions and procedures that are carried out in compliance with the personal data protection regulations.

From 2015, any researcher in the public field in Spain has been authorised to carry out studies using BIFAP data, for non-commercial purposes. The governance document reflecting BIFAP's data access and exploitation management procedures was approved and published on the BIFAP website in June 2017. The adaptation to the new data protection regulations and the incorporation of new sources of information makes the revision of the document relevant.

Document update has taken into account:

- The experience since the start of the project back in 2001, which has involved the exploitation of BIFAP data for the conduct of pharmacoepidemiological studies by the BIFAP team within the AEMPS and, in the last 5 years, also by external public sector researchers in Spain.
- Evaluations of procedures for the pseudonymisation and anonymisation of personal data, including an analysis of residual risk of re-identification.
- Established institutional agreements, the roles and responsibilities of the organisation itself (AEMPS), and discussions and agreements reached within its advisory bodies (BIFAP Scientific Committee and BIFAP Advisory Committee).
- With regard to the personal data protection, the applicable regulations and guides, and in particular:
    - The General Data Protection Regulation (GDPR)[1]
    - Organic Law 3/2018 (LOPDDD)[2]
    - The latest BIFAP personal data protection guidelines published by the Spanish Data Protection Agency (AEPD) and other bodies[3 4 5 6].

[1] Regulation (EU) No 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation, GDPR)
[2] Organic Law 3/2018, of 5 December, on the Personal Data Protection and the Guarantee of Digital Rights (LOPDDD). Seventeenth additional provision (17th AD)
[3] 'Guidelines and guarantees in anonymisation of personal data' procedures. AEPD, 2016. https://www.aepd.es/sites/default/files/2019-09/guia-orientaciones-procedimientos-anonimizacion.pdf
[4] Practical guide to personal data processing risk analysis. AEPD, Feb 2018. https://www.aepd.es/sites/default/files/2019-09/guia-analisis-de-riesgos-rgpd.pdf

- Relevant International Guidelines applicable to BIFAP on the Protection of Personal Data[7 8 9].
- The advice of the Data Protection Officer for the Spanish Agency of Medicines and Medical Devices.
- Risk analysis and data protection impact assessment carried out for the Register of the personal data processing activities of the AEMPS of the Ministry of Health related to BIFAP. (ASSI-GDPR tool: "Security Audit of Information Systems for Compliance with the General Data Protection Regulation.")

The main recipients of this document are:

- The database administrators and managers at AEMPS, who have an obligation to know and apply this document.
- Researchers who intend to carry out studies using BIFAP, and who must take into account the conditions and procedures established for data access.
- The institutions involved, in particular those bodies responsible for pharmaceutical care, primary care and information systems of the Autonomous Communities' health services.
- The healthcare professionals who practice in the health centres of the National Health System of the participating Autonomous Communities.
- Patients and the general population, because the dissemination of the requirements and processes involved in accessing the data contained in BIFAP provides confidence that health data is safely used to conduct research studies that provide relevant knowledge to improve the health of the population.

2 GENERAL TERMS AND CONDITIONS FOR ACCESS TO DATA FOR RESEARCH STUDIES

The BIFAP database is accessible for health research studies, preferably pharmacoepidemiological studies, by researchers affiliated to public bodies and by collaborating primary care physicians. Researchers must first register on the BIFAP Services website, complete the necessary training, and submit a research protocol that will undergo an evaluation and approval procedure.

The procedures for the evaluation, approval, and subsequent monitoring and follow-up of the protocols are intended to ensure that the health data contained in BIFAP is used safely.

[5] Practical guide for impact assessments on the protection of personal data AEPD. Oct 2018 https://www.aepd.es/sites/default/files/2019-09/guia-evaluaciones-de-impacto-rgpd.pdf

[6] Data Protection Impact Assessment Report Model for Public Administrations. AEPD. 9 July 2019 version. https://www.aepd.es/sites/default/files/2019-09/Modelo-informe-EIPD-AAPP%20%284%29.rtf

[7] OECD. (2015) *Health Data Governance: Privacy, Monitoring and Research*, OECD Health Policy Studies, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/9789264244566-en

[8] Opinion of the European Data Protection Group of Article 29 (GT 29) 05/2014 on anonymisation techniques, adopted on 10 April 2014.

[9] Opinion 03/2013 on purpose limitation. Adopted on 2 April 2013. European Union Article 29 Data Protection Working Party, WP203

2.1 Prior registration of researchers

Prior to protocol submission, researchers must be registered with BIFAP. Only those researchers who meet the established conditions are effectively registered. In addition to the collaborating primary care physicians, and the technical staff of the AEMPS and participating Autonomous Communities who are already registered as BIFAP users, public researchers who wish to register for BIFAP research can do so if they meet the following conditions:

1. Being in possession of a health sciences degree; other degrees such as information and communication technologies shall be assessed on a case-by-case basis. Pre-graduate students may be part of the study research team whose principal investigator must be registered.
2. Working in a public institution or body of a healthcare, university or health research nature or, in the case of people without any work activity, being a post-graduate student of a public university.
3. Performing these functions within the Spanish territory.

Once the researcher has the access keys to the BIFAP services platform, they have to carry out **the necessary training** on the BIFAP features as a source of information.

2.2 Data access request

The computer platform for registered researchers allows the request to carry out an investigation protocol. This request is required for projects by any type of researcher, including AEMPS own projects, and must include the following documentation:

1. Structured study **protocol**, detailing the study design and information from the BIFAP database that will need to be extracted for the project.
2. **Identification of all project researchers**, indicating which of them is the Principal Investigator (PI), who will be the ultimate responsible person for the study. All those who have access to data and/or contribute in any way to the project must be included.
3. **Research experience of the team**, including the PI curriculum.
4. **Funding sources for the study.** Only non-profit research projects are accepted.
5. **Conflict of interest statement** by all members of the research team. In particular, if any of the researchers has received funding in the past 2 years from a company that markets any of the medicinal products researched in the submitted project or any other conflict of interest on the part of any of the researchers.

2.3 Scientific Committee Opinion

The study proposal shall be evaluated by the Scientific Committee, which shall deliver an opinion on the research project submitted.

The favourable opinion of the Scientific Committee is necessary to carry out any research studies in BIFAP. It takes into account, to deliver its opinion:

1. The scientific suitability and technical quality of the protocol.
2. The project's interest for public health.
3. The relevance of the project, assessing the compatibility of the project with BIFAP's goals.

4. The competence and potential conflicts of interest of researchers and the means available to them in order to carry out the study.
5. The feasibility of carrying out the project using BIFAP data. Supported by a report from BIFAP team technicians:
    - It assesses the possibilities of obtaining the requested data from the database, the complexity of the study and the human and technical resources from the AEMPS required to carry it out.
    - It identifies the BIFAP data to be extracted and processed by researchers themselves for validation or variable definition, in collaboration with the BIFAP team.
    - It indicates whether conducting the study necessarily requires access to clinical or diagnostic information in unstructured free text, i.e. information that is added to clinical diagnostic codes by the primary care physician in the HCE as free text.

The Scientific Committee's evaluation also includes, where appropriate, if justified by the objectives of the proposed study itself, the access of researchers to aggregate data on the geographical location of patients, which will have to be approved by the Advisory Committee.

2.4 Research team commitments to access data

Once the BIFAP Scientific Committee reaches a favourable opinion for a study, the research team must accept the BIFAP data terms of use and the express commitments it acquires, including confidentiality commitments. Acceptance is a prerequisite for access to BIFAP data.

The researcher's commitment is basically an agreement between the AEMPS as the data manager and transferor and the researcher as the data receiver, whereby the latter acquires a series of commitments for the management and use of data. These commitments must be updated in the case of subsequent amendments to the approved protocol.

In particular, all researchers in the study make the following commitments:

1. To use the data to which they have access solely and exclusively for the purposes of the study, in accordance with the approved protocol version. The conduct of analyses not provided for in the protocol will require the approval of the BIFAP Scientific Committee.
2. To maintain absolute confidentiality and privacy on the pseudonymised data files to which they have access, which they may not copy or use for any purpose other than the study, nor disclose or assign to anyone outside the research team, even for conservation purposes. This commitment shall be maintained upon completion of the study.
3. To avoid performing any activities aimed at re-identification of patients included in the pseudonymised data files to which they have access.
4. To sign a specific commitment to confidentiality in the event that the researcher consults and extracts data from BIFAP on the AEMPS or the Autonomous Community's computer equipment.
5. To implement the necessary measures to prevent any unauthorised or improper use of data.

Failure by the researcher to comply with the commitments and conditions of access to BIFAP may be a cause for refusal to conduct future studies with BIFAP data, regardless of the relevant legal proceedings.

The PI also assumes the following commitments:

6. To avoid initiating the study without having obtained the favourable opinion of a research ethics committee or a research with medicinal products ethics committee, as appropriate, accredited in Spain.
7. To request the BIFAP Scientific Committee to approve any substantial amendments to the protocol, and in particular to include new funding sources, including those for the publication of the study.
8. To notify the BIFAP Scientific Committee of the incorporation of new researchers into the project, including all those who will appear as authors in the publications. All researchers throughout the project must provide the conflict of interest statement and investigator commitment in order to participate.

3 BIFAP DATA LIFESPAN

This section describes the stages in the treatment of health data included in BIFAP, from the extraction or capture of data in the Autonomous Communities' health services, until the publication of the results of the exploitation of the data in the corresponding scientific reports or publications.

Each stage shall describe:

1. **Data processing activities or operations** performed in accordance with BIFAP's purpose.[10]
2. The identification and characteristics of **the data** being processed.
3. The description of **participants**, i.e. the profile and roles of those people or institutions involved in the activities.
4. The **technological elements** (technologies and information systems) used in processing operations.

An outline of the entire process and the reference to the sections describing each stage and lifespan element is summarised in the following chart:

Chart 1: BIFAP DATA LIFESPAN DESCRIPTION

	STAGES					
ITEMS		DATA COLLECTION	DATA CLEANSING AND STRUCTURING	EXPLOITATION: STANDARDISED REPORTING	EXPLOITATION: RESEARCH STUDIES	
	Treatments Types	Removal, adaptation, transfer	Storage, organisation, adaptation, structuring	Query, extract, transmission, use, dissemination	Query, extract, transmission	Use, dissemination

[10] According to the GDPR, "processing" is any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated procedures, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction".

Process activities	Extracting data from the Autonomous Communities (3.1.1)	Data processing to facilitate their exploitation (3.2.1)	Extracting the subset of data and producing standardized reports (3.3.1)	Obtaining structured patient data files (3.4.1)	Building the DataSet Data analysis Dissemination of study results (3.5)
Processed data	Information contained in BIFAP (3.1.2)	Transformations in data in order to debug and structure them (3.2.2)	Subset of pre-calculated data for the BIFAP Express application (3.3.2)	Characteristics of the data being processed (3.4.2)	Research results (aggregated data) (3.5)
Participants	CCAA and AEMPS Information Systems Technicians (3.1.3)	BIFAP Technicians at AEMPS (3.2.3)	AEMPS technicians Technicians from participating Autonomous Communities Collaborating physicians (3.3.3)	Researchers BIFAP Technicians at AEMPS (3.4.3)	Researchers: (3.5)
Technologies	1st pseudonymisation Data transfer (3.1.4)	2nd pseudonymisation Obfuscation procedures. Access controls (3.2.4)	BIFAP Express (3.3.4).	Tools for data management and file transfer (3.4.4)	Publication in the Spanish Register of Clinical Studies and Reference on the BIFAP website (3.5)

3.1 DATA COLLECTION

3.1.1 CCAA data extraction and transfer operations

The BIFAP research database is generated at this stage. The controllers for the computer logs of the electronic medical records of primary care (HCE-AP) in the participating Autonomous Communities extract the information from the primary source, the patients' HCE-AP, for the subsequent transfer to BIFAP of the data necessary for the conduct of pharmacoepidemiological research studies, based on a pre-established data model.

In addition to the HCE-AP, data controllers at the Autonomous Communities also extract data that is part of the BIFAP data model from other sources of information.

The information is extracted at least twice a year from the central information system services at each Autonomous Community.

The **extraction methodology** differs from one Community to another due to differences in their clinical management software, but has the following points in common:

- At least twice a year, a new database update is created, incorporating the new data recorded in the time period since the previous update. Submission is made within two months of data closure (end of calendar semester). An increase in the frequency of updating is envisaged if necessary to increase efficiency and agility in carrying out studies particularly relevant to public health.
- In addition to patient data for the last period, new exports include updated data from previous years. That is, patient information that has been updated, corrected or completed is also obtained from the primary sources of the Autonomous Communities and data from additional sources of information that have been agreed to be extracted for BIFAP under the Partnership Agreement with the AEMPS are incorporated.
- Personal data that identify patients is never extracted, such as: patient identity; medical record number, Social Security number; home address…). The Personal Identification Code for the Autonomous Community (CIPA) shall be used to perform the pseudonymisation procedure detailed below.

3.1.2 *Data captured from primary sources: a description of the information contained in BIFAP*

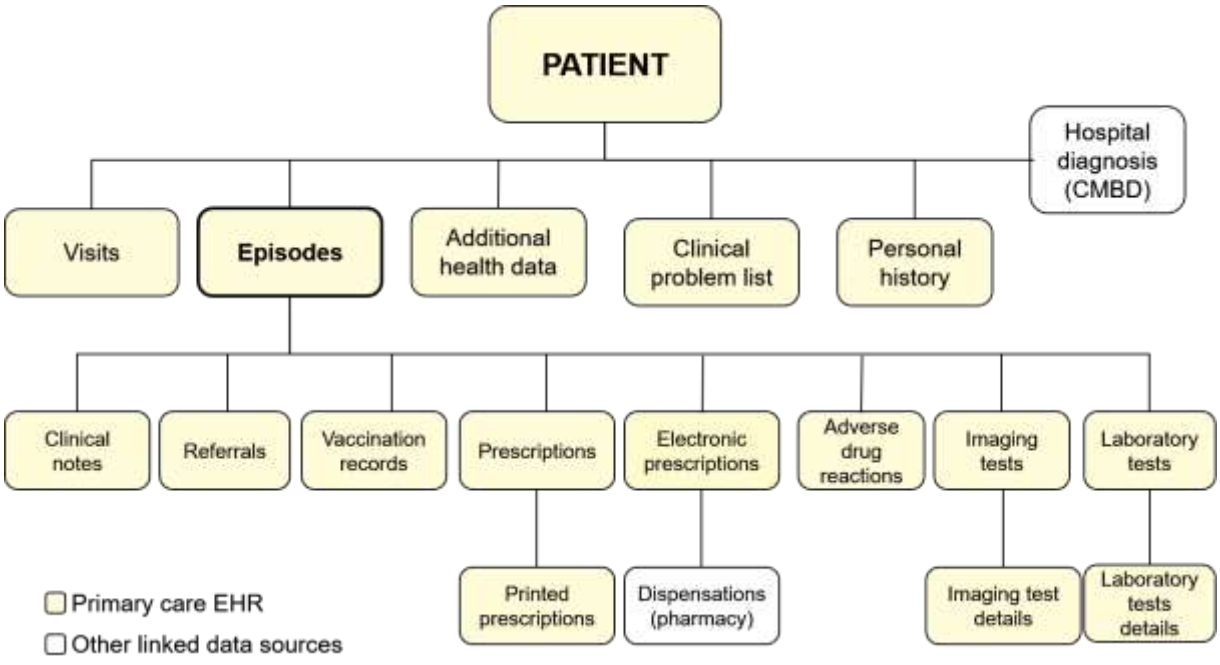
The **population scope** of the HCEs extracted for use in BIFAP includes information from the entire population tended to in the participating Autonomous Community primary care.

As of 2020, in several Autonomous Communities, only patient data have been extracted for the quotas of primary care physicians who had voluntarily requested to collaborate with BIFAP. This was the data extraction model ("collaborative model") with which the BIFAP project was initiated, prior to the centralisation of electronic medical records registry in the Autonomous Communities' computer systems, and that involved an active role for the primary care physician in the extraction of HCE-AP data from his office's own computer equipment.

The BIFAP website (http://bifap.aemps.es) reports the population data included in BIFAP after each update.

To extract the information, the BIFAP **data model** is followed, aimed at organising the data for the research. It takes into account the particularities of the different information systems at the Autonomous Communities.

This model is organised schematically as follows:



1. **Primary Care Electronic Medical Record (HCE-AP) data**

The process of recording a diagnosis in the primary care medical record is generally as follows: at each clinical encounter, the physician identifies the reasons for the patient's consultation and, for each one, creates a new clinical episode if necessary. The conditions and history that are part of the patient's medical record are also recorded.

a. PATIENT DATA: patient pseudonym unique identifier, gender, date of birth, date of entry into the database, date of deletion from database, cause of deletion (includes patient's death), status in database.

b. VISITS: date

c. CONSTRAINTS: diagnosis with dates, code, and descriptor

d. HISTORY: diagnosis with dates, code, and descriptor

e. GENERAL PATIENT DATA: includes data such as tobacco, alcohol, blood pressure, body mass index, etc. The date of collection and type are recorded.

f. EPISODES OR DIAGNOSIS: descriptor with date, code.

g. RADIOLOGY (image tests): dates, type, reason, results

h. COMMENTS ASSOCIATED WITH THE EPISODE: date, observations

i. INTER-VISITS: dates, medical specialty, motivation, results

j. VACCINES: date, code, name, batch

k. ANALYTICS: request and result dates, type, determination, value, units, ranges

l. PRESCRIPTIONS: dates, type, drug code, number of containers, dosage.

m. ADVERSE REACTIONS: data on suspected adverse reactions used for reporting to the pharmacovigilance system.

2. **Data from sources other than HCE-AP:**

a. HOSPITAL DISCHARGE DIAGNOSES: admission reasons recorded and coded by the RAE-CMBD system: includes dates of admission and discharge, type of discharge, primary and secondary diagnoses at hospital discharge.

b. DATA CONCERNING ELECTRONIC DISPENSATIONS OF MEDICINAL PRODUCTS: identification of the dispensed drug, date.

3. **Other data expected to be incorporated in the future.**

The BIFAP programme is open to the incorporation of new sources of information deemed relevant to its purposes. The possibility of such extension is covered by the Partnership Agreements, and is carried out following its assessment by the Advisory Committee. In 2019, the Advisory Committee approved an Improvement Plan which provides for the incorporation of the following data over the coming years:

a. DIAGNOSTIC INFORMATION

- SAR-CMBD system data other than Hospital Diagnosis at Discharge
- Clinical record of specialised care: admitted and ambulatory patients)
- Registration of Causes of Death by the National Institute for Statistics

b. INFORMATION ON DRUG USE

- Medicinal products prescribed in specialised dispensing care in pharmacy office
- Hospital dispensing medicinal products (inpatients and outpatients).

c. DEMOGRAPHIC DATA

- Census section of the patients' home (the home would not be recorded, only the aggregated level information of the census section to which the home belongs)

3.1.3 Participants in data capture

The Collaboration Agreements between the Autonomous Communities and the AEMPS provide for the identification of a health information system interlocutor in each Autonomous Community and a computer interlocutor (U.SINC) in the AEMPS BIFAP team to carry out the data extraction process.

The interlocutor designated by each Autonomous Communities' health information system extracts the necessary information from the information system so that the export of the data model is carried out correctly and to resolve, together with the computer interlocutor of the AEMPS, any problems that may arise.

Each transfer of data to the AEMPS IT Manager is documented and recorded by the participants.

3.1.4 Technological elements in data capture

3.1.4.1 Techniques used during data extraction

Before starting data extraction for BIFAP, the Personal Identification Code for the Autonomous Community (CIPA) is used in the Autonomous Communities for the combination or linking of the different records for the same patient (primary care HCE, electronic dispensing records, hospital discharge diagnoses and other data included in the future).

Subsequently a **pseudonymisation procedure** is carried out: the CIPA is eliminated by the computer technicians of the autonomous community, maintaining a pseudonymised identifier that distinguishes each patient from the others in BIFAP, to have the most up-to-date data, to enable them to be tracked throughout all database updates, and to be able to carry out studies that involve tracking patients over time. The pseudonym thus allows the information received each year from a patient to be linked to the existing information from previous annual extracts of the same patient without the CIPA or any other personally identifiable information being available in BIFAP at any time.

A dissociation procedure is performed to obtain the pseudonym. The dissociation procedure breaks the relationship between the code that identifies the patient at the primary source (the HCE-AP), and the patient ID code that is received at the BIFAP Data Centre (random number).

The private key to such dissociation is known only to the IT managers of each of the Autonomous Communities of origin. It is stored in a file that can only be accessed by the IT manager of each of the Autonomous Communities using their user code and secret password.

This disassociates patient information from clinical data. Neither BIFAP-managed personnel nor researchers will ever be able to access personal patient identification data.

The methodology of dissociation can be applied by the Autonomous Community unilaterally or in agreement with the computer interlocutors (U.SINF) of BIFAP in the AEMPS. The only requirement is that this dissociated identifier cannot be changed at each data submission, to avoid the insertion of duplicate patients for shipments in different years and to ensure continuity of patient follow-up for years.

As a final process, a **transfer file** is generated with relevant data for BIFAP.

3.1.4.2 Computer support structures for data transfer

The transfer of data from the Autonomous Communities to the BIFAP Data Processing Centre (CPD) in the AEMPS is carried out via an SFTP server enabled by the AEMPS for this purpose where each IT manager in the Autonomous Community has a protected folder to store the data.

3.2 DATA DEBUGGING AND STRUCTURING

3.2.1 Data processing activities to facilitate their exploitation

The received data are loaded into the BIFAP CPD by checking the dependency compliance of the data model and ensuring consistency in the received data types.

The initial load is performed in a repository where all data submissions are located.

With each update, a new research database is generated; the process is referred to in BIFAP as "**Step into Production**".

Since the data are received from several Autonomous Communities with different data models, during the transition to database production, each data source is subjected to a harmonisation process integrating into the common BIFAP data model in order to be exploited.

During this stage, the organisation, structuring and adaptation of the extracted data are carried out in order to ensure its quality and usefulness for its purposes, which are described in the following section. All these activities are carried out by the BIFAP team at AEMPS.

In addition, additional technical procedures for data protection are carried out, including a second pseudonymisation procedure and an obfuscation procedure (see section 3.2.4).

3.2.2 Transformations in data in order to debug and structure them

After obtaining the most up-to-date patient information with each database update, the following operations are performed:

3.2.2.1 Reference dates:

- Calculation of the start and end dates of the period in which the clinical record meets quality standards. This defines a follow-up start date and a follow-up end date.

- On the patient's date of death, a re-assignment of the patient's date of death is made based on the administratively registered date of death and clinical events recorded by the physician.

3.2.2.2 Selection of records with minimum quality for research:

For the initial population, a discard of medical records whose information does not meet the minimum quality requirements is performed, in the following cases:

- Invalid or not coded gender
- Invalid date of birth (before 1800 or after follow-up start date)
- Age over 115 years at end of follow-up date
- Tracking start date on or after tracking end date
- No patient clinical records
- HCE-AP with "inactive" record without administrative deletion record.
- Existence of data in the HCE-AP with a date prior to the start of follow-up or after the end of follow-up date.
- Administrative deletion due to transfer to another primary care quota

3.2.2.3 Identifying data:

The data extracted during the capture process (see section 3.1) have no patient identification data. In order to ensure effective anonymisation, the following additional steps are taken:

- On the first pseudonym, a second pseudonymisation procedure (see details of the procedure in paragraph 3.2.4.1)
- Obfuscation procedures are carried out on the non-coded data, infree-text registered, of the HCE-AP (see procedure details in paragraph 3.2.4.2).

3.2.2.4 Data on prescribed or dispensed medicinal products

Data on medical prescriptions and pharmacy dispensing are purged and adapted for research use. For this purpose, with each update of the database, data of the prescribed or dispensed medicinal products registered in BIFAP are extracted, compiled into an anonymous drug data file (BIFAP Drug Dictionary), with the aggregate data, without individualised patient information, of the prescribed and dispensed medicines. In this file:

- A recoding of the codes of the medicinal products that have been prescribed as active pharmaceutical ingredient to a single dictionary is done.
- On the prescription data, a standardised calculation of the duration and dose of prescriptions and dispensations is performed using an algorithm.

After the cleansing of this data, for each prescription or dispensation, there is always the identification of the drug (national code) and the anatomy-therapeutic-chemical group (ATC Classification System), along with the other information associated with the recipe.

3.2.2.5 Data related to clinical and diagnostic events

Different diagnostic coding systems with different levels of granularity currently exist in BIFAP: the International Classification of Primary Care (ICPC-2) in its two versions (ICPC-2

1998) and the International Classification of Diseases, both in its version 9, clinical modification, ICD-9-MC and in its version 10 (ICD-10-MC). In addition, data recorded in the electronic primary care medical record (HCE-AP) include information that includes information written by the physician in free text. In order to enable the exploitation of this information, a structuring of the diagnostic coding is carried out.

For this purpose, in each database update the data recorded in BIFAP are extracted, compiled into anonymous files with the codes and descriptors that already contain aggregate data, without individualised patient information, of the terms used to describe clinical and diagnostic episodes. The following actions are performed thanks to these files:

- A dictionary of medical terms is maintained based on the ICPC (ICPC-BIFAP Dictionary) dictionary which, based on the literals that primary care physicians have recorded in the HCE-AP (not fully structured in the source HCE-AP itself), increases the granularity of the ICPC dictionary.
- The investigation of the different registered diseases and clinical events is facilitated by the maintenance of BIFAP groups of Diagnostics, defined from coding ICPC-BIFAP and ICD-9 and defined by means of search algorithms. This allows limiting the need for free text comments by researchers.

Currently, tools are being developed to relate the terminology used at source in HCE-AP, both in the literals of the codes that physicians record, and later in free text in commentaries, to medical terminology ontologies used internationally, such as SNOMED-CT, through natural language processing (PLN) technologies

3.2.3 Participants in data debugging and structuring

The data are purged by the BIFAP U.SINF controller, on their computer equipment at the AEMPS.

The procedures for standardisation of drug terminology and the creation of standardised diagnostic terminology, as well as the creation and maintenance of the BIFAP Diagnostic Groups are carried out by the BIFAP team staff of the AEMPS, both U.SINF and U.INV. In this step, anonymous data files are processed, without individual patient information, in order to standardise the terminology and facilitate its subsequent exploitation.

3.2.4 Technological elements in data debugging and structuring

3.2.4.1 Second pseudonymisation procedure

In this phase, a new dissociation of the received identifier is performed and a new identifier is randomly generated from the first identifier received from each Autonomous Community (both identifiers are related in a table).

This means that although in BIFAP the patient identifiers are received dissociated from the central services of each Autonomous Community, before the data can be exploited for research studies a new dissociation of the received identifier is performed.

The code of the medical quota received in the transfer file is also dissociated, generating a table that relates the code of the medical quota to a new pseudonymised identifier generated randomly.

3.2.4.2 Obfuscation procedures

Free text information is used in BIFAP for better event characterisation, event validation, or event identification that is not properly encoded.

In order to ensure the effective anonymisation of the information, an obfuscation procedure is carried out for information that may contain sensitive information and may be found in the unstructured text from the electronic medical record.

The procedure is part of the data disturbance techniques and a systematic variation and deletion of data is carried out that prevents the resulting data (words) from providing information on specific cases. In particular, a random exchange of data is carried out, i.e. the introduction of random distortion in the free text dataset, maintaining the detail and structure of the original information.

It consists of replacing words with residual sensitive data, by words in other records or by the string **** in the free text fields of the HCE. Using data mining algorithms, all text fields in the HCE are processed for those strings of characters that may contain sensitive data (given name, family name, city, healthcare institutions, demonyms, etc.) and then be randomly replaced.

Depending on each Autonomous Community, this procedure is carried out by the IT managers of the Autonomous Community during the extraction of data from the HCE and before the generation of the transfer file, or by the U.SINF manager of the BIFAP team.

3.2.4.3 Technologies used for the structuring of diagnostic data

The structuring of diagnostic data includes manual and semi-automatic review processes. The following technologies are used as semi-automatic review processes, with manual validation of automated procedures:

- Online tools are available to collaborating physicians so they can review any terms used in diagnostic descriptions of HCE-AP and compare them with reference terms from reference medical terminology dictionaries to confirm or rule out equivalence.
- Language processing techniques for mapping expressions used in the HCE-AP to dictionaries and reference ontologies (UMLS, SNOMED-CT)

3.2.4.4 Computer support structures for data debugging and structuring

Data access throughout the debugging process is physically performed within the AEMPS facility (Calle Campezo, 1 Ed 8. 28022 Madrid). Access to AEMPS facilities is only possible for staff working in the organisation. Other people accessing the AEMPS must have authorisation and check in at their entrance and exit from the premises.

The BIFAP database is restricted to the computer technicians of the BIFAP's U.SINF at the AEMPS headquarters.

Access to the computers where the database is accessed requires a password. Only people who are currently part of the BIFAP's U.SINF can access as password users.

Database structuring tasks (creation and maintenance of our own terminology dictionaries) involving access to unstructured text records of patients' medical records are also performed under these same conditions.

3.3 EXPLOITATION OF DATA FOR STANDARDISED REPORTING

3.3.1 Standardised reports with pre-calculated data

Tracking usage patterns and general monitoring of the safety of medicines under actual conditions of use are part of the pharmacovigilance activities of AEMPS (See section 4.1.1 on legal regulations for health data processing at BIFAP). To do this, tools are developed to provide standardised information from BIFAP.

The application "BIFAP EXPRESS" is currently in production for this purpose. BIFAP EXPRESS is an application accessible to certain users of the BIFAP Services website that explores the characteristics of the use of prescribed medicinal products in primary care, based on pre-calculated data extracted from the Electronic Medical Record (HCE) and prescription-dispensing records from the electronic prescription in the BIFAP patient population. The development of the BIFAP EXPRESS tool is also part of the objective of providing return information to the Autonomous Communities and collaborating physicians.

The data generated in BIFAP Express consultations are not intended to be used to conduct research studies, but to know the pattern of use of medicinal products as a complement to other activities; users are therefore warned that the results of queries should not be published.

BIFAP Express provides access to information previously extracted from BIFAP, through pre-determined consultations, and includes only aggregate anonymous information about the use of medicinal products and associated diagnoses of BIFAP patients.

For the same purposes, in addition to the BIFAP Express application, standardised reports are generated necessary for drug safety vigilance.

3.3.2 Data included in standardised reports

The data source that BIFAP EXPRESS feeds is a subset of BIFAP that is extracted after the database upgrade. This is done by collecting information from the "treatment courses" generated throughout the database for each drug, including age and gender of patients and diagnostic codes of episodes associated with prescriptions.

Pre-calculated data allows the user instant access to aggregated data of each active pharmaceutical ingredient, by calendar year periods and in calendar quarters, of the last 6-7 years available and stratified by age group and by sex:

- Usage prevalence parameters
- The relative frequency of 11 ICPC-2 codes and 11 ICD-9 codes most frequently recorded as an "associated episode". Associated Diagnostic Episode: an approximation to indications (or diagnostics that are compatible with the indication) of prescriptions made by primary care physicians, based on the information they record in the "Associated Episode" field of the Electronic Medical Record (HCE).
- The relative frequency of the most frequently prescribed doses for each drug.

The user interacts to choose information of interest from pre-set criteria and display it on screen, print it, or export it in spreadsheet format.

The scope of the information obtained in queries to BIFAP Express can be:

- Global (includes data for the entire BIFAP population)

- By Autonomous Community (includes only the BIFAP population of the Autonomous Community corresponding to the user)
- By quota or CIAS (patient population tended to by a primary care physician, identified by a CIAS).

3.3.3 Participants in the extraction and usage of data for standardised reporting

Access to the Standardised Reports based on the user profile is summarised in the following table:

	Global BIFAP Population	By Autonomous Community with data in BIFAP
AEMPS	YES	YES, from all Autonomous Communities
Administration of participating Autonomous Communities and collaborating physicians*	YES	YES, from its Autonomous Community
All other users	NO	NO

*: BIFAP Express data for users who are collaborating physicians and CIAS holders also include data on the population scope of their CIAS, so that it shows the aggregated data of the patient population served by that cooperating physician corresponding to its primary care quota. AEMPS users who are database administrators also have this information.

The use of Standardised Reporting data for external reports, publications, etc. requires express authorisation from the AEMPS to ensure that the data intended to be published has been interpreted correctly.

3.3.4 Technological elements for standardised reporting

With each database update, pre-calculated data is generated by accessing the BIFAP data subset.

Standardised reports, such as BIFAP Express reports, are offered on the BIFAP Services web site through a web application, which allows the user to interactively select the data contained in those reports.

3.4 EXPLOITATION OF DATA FOR RESEARCH STUDIES

Epidemiological studies in BIFAP include several distinct stages:

1. **Generating structured patient data files:** Includes procedures for obtaining data files according to the specifications detailed in the study protocol. Requires the participation of the investigator and the BIFAP team.

   For the generation of the data file(s), the specific tools already available can be used, or obtained through direct programming tasks (see section 3.4.4). Different patient data files,

as well as summary tables with aggregate data, are obtained as a result depending on the requirements of the protocol (see section 3.4.2)

2. **Building the DataSet for analysis**

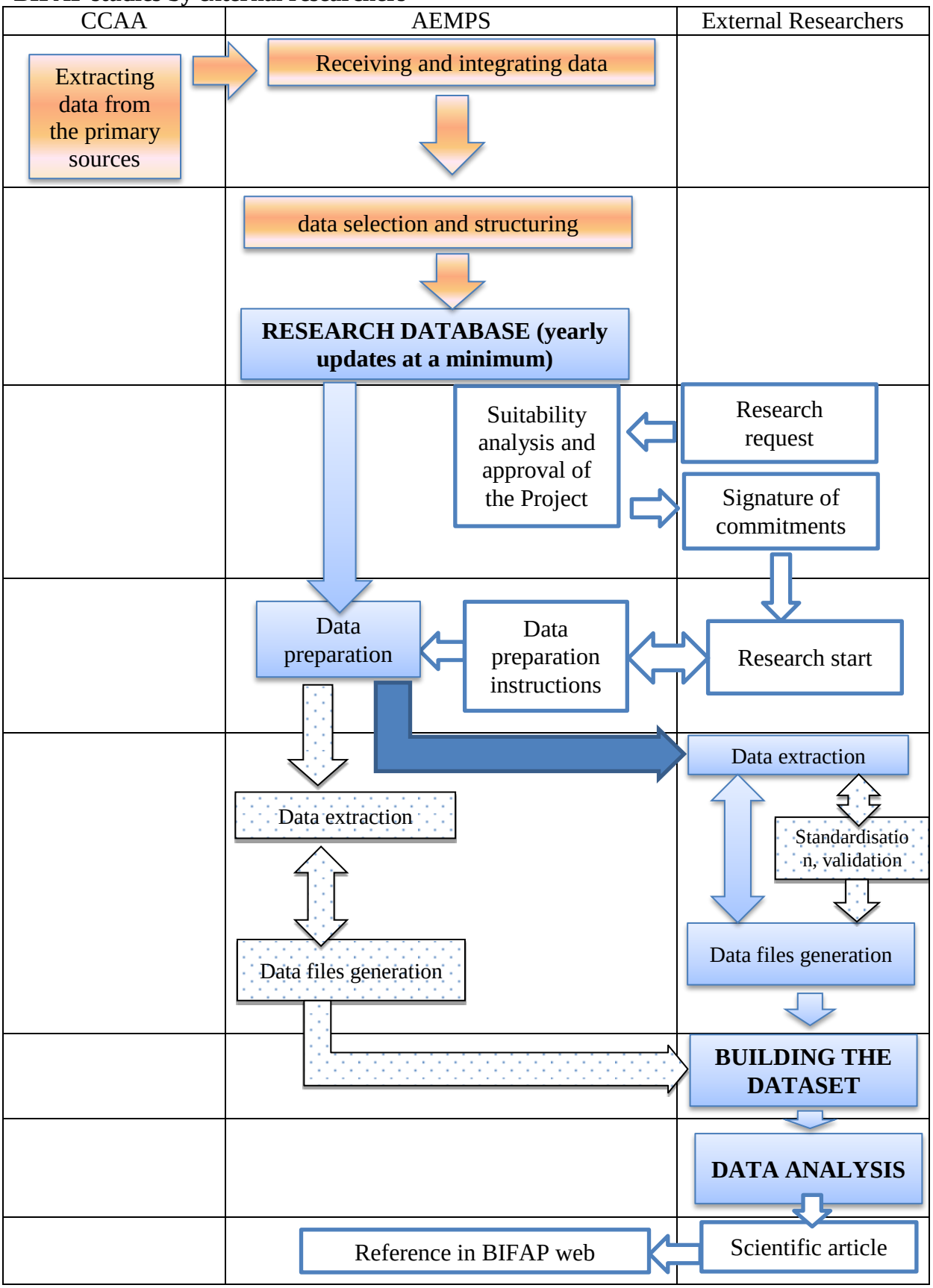
   This action is the responsibility of the research team. The data files generated, provided by the BIFAP team or obtained by the researcher, represent tables with individual and gross data, and the researcher will carry out the necessary processes for building its DataSet for the statistical analysis and obtaining the results of the study (section 3.5).

3. **Statistical analysis and publication of the results of the study:** these actions, a responsibility of the research team, are described in section 3.5.

The first stage, generating structured data files that are made available to researchers, involves the participation of the BIFAP team in the studies of external researchers.

The processes that require processing of BIFAP data as well as the interactions between AEMPS and researchers are summarised in the following outline:

Flow of data processing operations (data processing and data transfer) to perform BIFAP studies by external researchers



3.4.1 Obtaining structured patient data files

This process involves the technical staff of the BIFAP Team and the principal study investigator, and involves data processing and data transfer activities:

a) The **processing** of BIFAP data in order to generate the data file(s) necessary for the conduct of the study and its statistical analysis. All this in the context of a research study, the ultimate purpose of the operation of BIFAP.

b) The **transfer** of data to the research team that must access them, when the team not belong to the AEMPS.

It is carried out in several phases, with different participants (see section 3.4.3):

1. **Data preparation instructions:** In the event that the details specified in the study protocol approved by the Scientific Committee are not sufficient, the following shall be required at least and in accordance with the protocol:
    a. an operational definition of the study population,
    b. the time interval of the data, and
    c. the description of the variables that will be needed for the analysis,.
2. **Data preparation:** It is decided which procedures will be required for obtaining the data files (see section **3.4.4**), which will later be used by the researcher to build the DataSet.
3. **Data extraction:** Selection of the study patient population, including time interval, extraction of study variables, and eventual creation of definitions of diagnostic events, if those required are not already structured. It involves checking the patient information contained in BIFAP, either through certain Study Management Tools or through direct programming.

    Depending on the characteristics of the protocol, there may be a phase in which the investigator must perform **standardisation and data validation** tasks, which are necessary only in some studies, and if so foreseen by the approved protocol. It may include creating diagnostic event definitions, if those required are not among the definitions already created, and/or validating clinical or diagnostic events by examining the free text detailed information contained in the medical record. For this purpose, an initial data transfer is carried out to the researchers, who are always the ones who perform this task.
4. Final **data files generation**, once confirmed that the data is complete and the data necessary for the research team to construct the DataSet and perform the necessary analyses is collected. It involves the completion of the data transfer procedure to the research team.

3.4.2 Characteristics of the data that are processed for the collection of the study's data files

The tools used to extract the data needed in a particular research study as well as direct programming of queries to the database (see section 3.4.4) can generate three types of data:

- **Aggregate Data**: Information that contains only anonymous, summarised, and aggregate data for a set of patients. These data may be stratified by age, sex, and follow-up time, and their purpose is to describe the population under study, including measures of frequency of research events of interest. They can already be part of the statistical analysis of the study. Some of these data are generated as "Results Reports" by BIFAP query tools while others are generated by the researcher from previously extracted data files. They are available in TXT or Excel format.
- **Structured patient data files**: Information with data from individual patient records and that includes only variables with structured and pseudonymised information (standardised microdata), generated for a specific research study, and that, once processed by the research team to build the DataSet, it enables its analysis to respond to the study's goals. Patient data files are available in TXT or Excel format. The data tables can be linked to each other by the BIFAP patient identifier (pseudonym), and refer to the different elements of the BIFAP data model (see section 3.1.2).
- **Unstructured Microdata:** Information with data from individual patient records and including free text information. Depending on the design and characteristics of the study, different queries and data files are generated in certain studies, the purpose of which is to check the validity or reliability of the results. Includes primary care medical record fields in the form of free-text comments, entered by the doctor in the Electronic Medical Record. The visualisation, verification and, lastly, the classification of the data are performed by the researcher using specific tools. The results of the tool's queries are displayed in the tool itself, and do not allow the generation of files (such as text files) that can be exported to other platforms.

3.4.3 Participants in obtaining structured patient data files

In the case of studies whose Principal Investigator belongs to the organisation itself (AEMPS), the whole process of data extraction, analysis of the results and the subsequent publication of the results is carried out in the AEMPS ; and data exploitation management does not involve data transfer procedures to external researchers.

In projects where the research team is external to AEMPS, obtaining structured data files requires the participation of the AEMPS BIFAP team and the Principal Investigator. Depending on the protocol characteristics and previous experience with BIFAP from the research team, this participation varies. In these cases, a transfer of the data controller's data (AEMPS) to the external research team should be performed. The participants in each of the phases are described below:

1. **Data preparation instructions:** A person from the BIFAP technical team (IT technician or epidemiologist) contacts the main study researcher to plan how to generate (extract) and transfer the data files that will allow the analyses foreseen in the protocol, once all requirements for study approval have been met.
2. **Data preparation:** The technical staff of the BIFAP team establishes the procedure for carrying out the data transfer, depending on the characteristics of the protocol approved by the Scientific Committee and the experience and type of researcher.
3. **Data extraction**. Carried out by:
    a. **The researcher**, using the study management tools to extract the data and obtain the data files and the final DataSet. If researchers are required to access

unstructured microdata that includes free HCE text, they must sign a specific document for confidentiality and acceptance of the conditions of access and use of the information. This requirement applies to both members of the research teams who access BIFAP tools on AEMPS equipment and those who have been authorised to remote access from equipment in participating Autonomous Communities' Health Administrations.

b. **The BIFAP team**. Depending on protocol characteristics (design, type of variables, need for validation, etc.) and previous BIFAP experience from the team of external researchers, obtaining the data files will require the participation of the BIFAP team in some of the tasks and whenever it is necessary to conduct direct queries to the database through programming. The AEMPS BIFAP team signs a specific confidentiality document for its access to BIFAP data for its own database management, maintenance and operational tasks.

4. **Data files generation**.

   a. **The researcher**, by extracting the data using the study management tools.

   b. **The BIFAP team**, by extracting data using the study management tools if the extraction was done in collaboration with the researcher, or by extracting data by direct query to the database through programming. In this case, the BIFAP team technician will then transfer the data files to the Principal Investigator.

   c. The BIFAP team technician and the Principal Investigator find that the data files obtained are those necessary for the researchers to build the DataSet for analysis

3.4.4 Tools for data extraction and management in research studies

Tools that facilitate the extraction and exploitation of BIFAP information for research studies allow access and analysis of data according to the research protocol.

They enable the extraction of the information, the direct review of the generated data, or the validation of the information on certain diagnoses, and generate the data files necessary for the intended analysis.

3.4.4.1 Comprehensive Study Management Tools at BIFAP (GIE)

Using the Comprehensive Study Management (GIE) software, different operating modules are made available to the researcher:

- Medical records filter,
- Variables generator,
- Medical records viewer,
- Case-Follow-up module,
- Issues module,
- Medical records viewer on a timeline,
- Treatment course generator,
- Cumulative dose generator.

- Exploitation of incidence of drug use for time series analysis

The different modules are managed integrally around the study through a Web application developed with J2EE technology that integrates the computer procedures related to the extraction of the information and generation of the data file for analysis. To facilitate the management of files and the storage of the different queries of a study, the application is structured in Studies/Folders.

In addition, support procedures are available that refer to:

- Age group management.
- Patient subgroup management
- Identifying clinical events: Diagnostic BIFAP groups (see section 3.2.2.5) and Diagnostic user groups.
- Identification of medicinal products of interest

These tools can only be used on equipment located on the AEMPS. Remote access may be authorised from equipment placed in the Collaborating Autonomous Communities' Health Administrations for the carrying out of specific studies under the following conditions:

- Data processing through the use of tools is limited to the procedures and time required to perform the study.
- The user must have the approval of the AEMPS that will keep the Advisory Committee informed.
- The user must be in compliance with the relevant Autonomous Community, by means of a document signed by the person responsible for access control at its site.

The conditions for using BIFAP GIE tools by any user profile are as follows:

- The user must have additional access permission to the tools that will detail which tools they want to access and for what purpose, which shall be requested from the Administrator (responsible for the BIFAP information systems unit). In the case of external researchers, these are given the profile of "Visiting Researcher". To do this, they must belong to the research team of an approved and running study.
- Access to the tools is only in the environment (IP-Internet Protocol- addresses) of the computer equipment at the AEMPS headquarters or, in the case of authorised remote access, at the headquarters of the Public Administrations of the collaborating Autonomous Communities.
- All activity performed by users accessing GIE tools is monitored by generating Programming Reports: with each output file the application keeps a report on the activity carried out by the user, in order to track and make it possible to trace the queries made in BIFAP with the tools.
- From the tools it is not possible to access the pseudonym of the geographic location variable of the primary care quota, which was pseudonymised in the database purge phase.
- A document on the terms of use and the maintenance of confidentiality is required.

3.4.4.2 Query and extract data by direct programming

In cases where, because existing GIE tools are insufficient, information is extracted from BIFAP through direct programming against the database, access to BIFAP is carried out exclusively in the AEMPS site by the BIFAP U.SINF. The data extracted using this procedure are of the Structured Patient Data Files type.

3.4.4.3 Technological elements to carry out the transfer of data files to the researcher and to ensure confidentiality

The data files are generated from GIE tools or by direct programming and are made available to the researcher. Depending on the characteristics of the generated data (see section 3.4.2), an environment with a certain level of security is established in order to transfer the data.

- **Aggregate patient data reports** (results reports): the researcher is required to be a BIFAP user with an "active researcher" role. Reports are delivered to the researcher by hand, by encrypted email, or made available to her/him from a server in a secure SFTP format.
- **Structured patient data files**: The researcher is required to be a BIFAP user with an "active researcher" role. Files are delivered to the researcher by hand, by encrypted email, or made available to her/him from a server in a secure SFTP format.

  Dates of birth and death, which can serve as the source of indirect re-identification, even though they are standardised data, cannot be part of structured patient data files. This is done by aggregating the data, so that only the year of birth or death or the age in years are obtained. The CA to which the patient data has been recorded are also not available in the structured data files. Exceptions to this may be made in specific situations, if justified by the study protocol, and approved in advance by the Scientific Committee and the Advisory Committee (see paragraph 2.3).
- **Unstructured Microdata:** It is necessary that in addition to being an "active researcher", the researcher is enabled as a "GIE tool user" from the equipment located in the AEMPS or, in the case of authorised remote access, on computer equipment at the headquarters of collaborating Autonomous Communities' Public Administrations. Each member of the research team signs a specific confidentiality document for access to queries and files (see previous section). The results of the tool's queries (Medical records viewer) are displayed in the own tool screen, and do not generate files (such as text files) that can be exported to other platforms.

3.5 DATA ANALYSIS AND DISSEMINATION OF STUDY RESULTS

On the basis of the information available in the data files, and in accordance with the specifications of the analysis plan, additional transformations are carried out as necessary for the analysis of the data and subsequently the production of results reports and corresponding scientific publications. These actions are the responsibility of the research team.

From the data files obtained, the research team performs additional standardisation of certain data, as well as linking the same patient data between the different data files previously generated, using each patient's unique pseudonymised identifier.

For the purpose of transparency, BIFAP publishes the studies approved by the Scientific Committee, as well as the name of the IP on its website. All publications are also published on the BIFAP website after they have been notified by the researchers. The scientific responsibility for these publications rests with the authors.

AEMPS may grant other researchers access to data files that have already been submitted for specific research for the purpose of replicating or completing the analysis of these data files and contributing to the quality of the research in general. In this regard, the Principal Investigator's express commitment document indicates the following:

- Making the files resulting from case validation or any other algorithm or procedure developed for the study by the research team to improve the quality of the data extracted from BIFAP accessible in the AEMPS, in order to be able to apply such improvement to other studies.
- To know that the AEMPS may assign to other authorised researchers the data files extracted from BIFAP, although, for a period of 2 years from the delivery of the Patient Data Files, the AEMPS may not, without the approval of the Principal Investigator, transfer data for other studies that share the same data and scientific goals, in the opinion of the Scientific Committee. This period is intended for the researcher to have a time of exclusive use of the data for making scientific publications. This will not apply to data already published.

As regards the communication of results, the IP undertakes to:

- Publish information on the study in the Spanish Registry of Clinical Studies (REec) at the beginning of the study and to contribute, once completed, to the publication of the results obtained, in accordance with the procedures published by the AEMPS.
- When the resulting information involves changing the benefit-risk ratio of a drug, report the results of the previous study to the AEMPS, previous and regardless of their publication in a scientific journal or in the REec.
- As ultimate responsible for the full integrity of the investigation from its inception to its publication, ensure that the guidelines of the International Committee of Medical Journal Editors (http://www.icmje.org) on the policy of authorship of scientific communications are followed, and that all authors of the study with BIFAP have stated that they have no conflicts of interest related to the medicinal products under study.
- The following statements should be included in the results publications:
    - “This study is based on data from the Pharmacoepidemiological Research in Primary Care Database (BIFAP). BIFAP is a public programme for independent research funded by the Spanish Agency of Medicines and Medical Devices (AEMPS).”
    - “The results, discussion and conclusions of this study are those considered by the authors only and do not in any way represent the position of the AEMPS.”
    - Acknowledgements: “The authors acknowledge the excellent work of primary care physicians (family doctors/paediatricians) as well as the support of the autonomous communities participating in BIFAP.”

The information with the results of the studies contained in the publications, in accordance with the quality standards of scientific publications, may only contain aggregate, already anonymous, information of the population that is part of the study.

Finally, the BIFAP website publishes a reference to all scientific articles publications that have been made using BIFAP data.

4 DATA PROTECTION REGULATIONS APPLICATION

As of May 2018, Regulation (EU) No 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation, GDPR) is fully applicable in Spain. This regulation applies to any fully or partially automated personal data processing, as well as to the non-automated personal data processing contained or intended to be included in a file.

In addition, in December 2018, Organic Law 3/2018, of 5 December, on the Protection of Personal Data and the Guarantee of Digital Rights (LOPDDD) came into force. The seventeenth additional provision (DA17th) includes specific provisions, applicable in Spain and complementary to the GDPR, on health data treatment.

The participants in the different BIFAP data processing have been described in section 3 and are summarised in CHART 1: BIFAP DATA LIFESPAN DESCRIPTION. With regard to data processing subject to personal data protection regulations, once data collection has been finalised, already pseudonymised, from the Autonomous Communities' information sources (paragraph **3.1.3**), and until delivery is made the structured data files to researchers (paragraph **3.4.3**), the AEMPS acts as data controller, without the intervention of any data processor or co-supervisor. Thus, the following sections describe how the AEMPS, as data controller, applies the different conditions and principles of the data protection regulations in force in Spain.

4.1 LAWFULNESS OF PROCESSING

The possibility of processing of health data is set out in Article 9.2 of the GDPR (circumstances in which processing of health data is not prohibited). In turn, the legal bases or foundations for the processing of data are set out in Article 6 of the GDPR: Lawfulness of processing.

The following sections detail the legal regulations for the processing of BIFAP health data based on these provisions.

4.1.1 Data processing is carried out in the exercise of public authorities and is necessary in the field of public health

Among the situations provided for in the GDPR among the possible legal regulations for the processing of human health data to be lawful, the following are applicable to BIFAP data processing carried out by the AEMPS:

- Treatment of health data is not prohibited if (Art. 9.2.i of the GDPR) treatment is necessary in the field of public health, and in order to ensure high levels of drugs safety and is covered by a Spanish law or European Union law.
- The legal regulations for the processing of health data in BIFAP would be that (Art. 6.1.e of the GDPR) processing is necessary for the fulfilment of a mission carried out in the public interest or in the exercise of public powers conferred on the data controller.

This is justified for BIFAP data processing on the basis that:

- The BIFAP data controller is the Spanish Agency of Medicines and Medical Devices. Royal Decree 1275/2011, of 16 September 2011, approves the Statute of the AEMPS, and establishes powers in the evaluation, authorisation and registration of medicinal products. It also gives it the competence to plan, evaluate and develop the Spanish

Pharmacovigilance System and to carry out pharmacoepidemiological studies to assess the safety of medicinal products for human use (Article 7.17).

- Royal Legislative Decree 1/2015, of 24 July, approving the consolidated text of the Law on the Guarantees and Rational Use of Medicines and Medical Devices, in its Article 55, assigns AEMPS the role of promoting those pharmacoepidemiological studies necessary to assess the safety of licensed and enrolled medicinal products under actual conditions of use.
- As a regulatory development of this standard, Royal Decree 577/2013, of 26th of July, regulating the pharmacovigilance of drugs for human use (Article 4.1.j), it assigns AEMPS the role of promoting the creation and facilitating the use of computerised health databases that serve as a source of information for the conduct of pharmacoepidemiological studies with the participation of the health administrations of the autonomous communities and healthcare professionals and to stimulate the creation and maintenance of a unified record of the available databases.
- In addition, Royal Decree 577/2013 acknowledges the power of the AEMPS (Article 3) to establish the necessary agreements with the competent bodies of the Autonomous Communities for the sharing of the sources of information that depend on them, in particular with automated databases. The signed Agreements are published in the Official State Gazette and the updated list is available on the BIFAP website (http://bifap.aemps.es). These agreements detail the commitments of each Autonomous Community and the Spanish Agency of Medicines and Medical Devices, as well as other operational aspects.
- Paragraph 1 of the DA17th of the LOPDDD indicates laws (and their implementing provisions) in Spain which are covered by Article 9.2(g), (h), (i) and (j) of Regulation (EU) 2016/679 with regard to health-related data processing. In this list of rules, RDL 1/2015 (and therefore its provision for development, in this case Royal Decree 577/2013) is included.

4.1.2 Processing is performed for scientific research purposes

In addition to the legal regulations indicated in the preceding paragraph, Art.9.2.j of the GDPR indicates as one of the exceptions to the general prohibition on the processing of health data that processing is necessary for scientific research or statistical purposes.

The processing of health data included in BIFAP is for scientific research purposes, whether in research carried out by the AEMPS itself or by other health researchers.

For this purpose, the conditions set out in Art. 89.1 of the GDPR and the requirements referred to in paragraph 2.d of the 17th additional provision of the LOPDDD, since the GDPR itself indicates that these conditions must be established on the basis of the law of the Union or of the Member States. The way in which these conditions have been entered into BIFAP is detailed in section 4.3.

4.2 PRINCIPLES APPLICABLE TO DATA PROCESSING

Article 5 of the GDPR sets out the general principles for data processing. The application of these principles to the processing of health data that are part of BIFAP has been carried out in accordance with the following.

4.2.1 Lawfulness, fairness and transparency (art.5.1.a GDPR)

- Legal basis for legitimising treatment: see paragraph 4.1
- Fairness and transparency: see paragraph 4.3.10.

4.2.2 Purpose limitation (art.5.1.b GDPR)

- Selection, registration and training of potential researchers: see section 2.1
- Approval process for study protocols: see sections 2.2 and 2.3.
- Explicit commitments of all researchers: see section 2.4.

4.2.3 Data minimisation (art.5.1.c GDPR)

- Variables pseudonymisation and reduction: see sections 4.3.1 and 4.3.2

4.2.4 Accuracy (art.5.1.d GDPR)

- Periodic updating of the database: see section 3.1.1
- Data debugging and structuring, see section 3.2.2

4.2.5 Limitation of the conservation period (art.5.1.e GDPR)

In the case of BIFAP, it is not possible to set a time limit for the erasing of the different data categories. The principle that data conservation shall only be lawful while the purposes that motivated the assignment remain is assumed. However, the purpose of processing BIFAP data is scientific research, and a predefined limitation of the conservation period could prevent research relevant to public health, as the need to access patients' medical record data that occurred in previous periods of time for this purpose cannot be excluded. It must therefore be admitted that data can be retained for long periods, otherwise scientific research itself would not be feasible. In this case (Art. 89.1 of the GDPR) guarantees for stakeholders are based on the principle of data minimisation and in particular through pseudonymisation, as described above.

4.2.6 Integrity and confidentiality (art.5.1.f GDPR)

- Personal data security: see section 4.3.4

4.2.7 Proactive responsibility of the data controller (art.5.2. GDPR)

Since the creation of the first BIFAP data files (year 2002) and to date, measures have been taken to ensure compliance with the principles of data protection and documentation. Based on the experience gained, and in particular on the availability of BIFAP to external researchers in 2015, the personal data protection procedures carried out at BIFAP have been updated, in particular:

- **Re-identification residual risk analysis**: A review of the characteristics of the information contained in the HCE variables of the BIFAP data model was carried out. The free text information entered by the physician, on the other hand, essential in some cases for the validation of diagnostic information, is recognized as the most important potential source of data sensitive to the privacy of individuals. Specific measures were therefore taken to limit and control access of researchers to this type of information.

- **Data Protection Impact Assessment**: In addition to the actions carried out prior to the entry into force of the GDPR, a Data Protection Impact Assessment (DPIA) has subsequently been carried out using the ASSI-GDPR tool (Annex document).
- **Documentation of the actions** is reflected in the working procedures of the BIFAP team, the quarterly Activity Reports, the Minutes of the meetings of the Advisory Committee and the Scientific Committee, mainly. Other documents that are generated are:
    - Documentation of decisions on particular situations arising, not provided for in this governance document.
    - Documentation of any security breaches.
- **Periodic review of actions**: with each update of the database personal data anonymising procedures are reviewed, understanding it as a process that goes beyond the pseudonymisation of the identification data of persons and whose purpose is to eliminate or minimise the risks of re-identification of the data, while maintaining the truthfulness of the results of the treatment of the data.

4.3 APPLICATION OF THE CONDITIONS FOR THE USE OF DATA FOR HEALTH RESEARCH PURPOSES

The general conditions for the use of health data for scientific research purposes are set out in Article 89.1 of the GDPR.

Sections 2.f) and 2.g) of the 17th DA of the LOPDDD set out certain requirements for the processing of data for health research purposes, and section 2.d indicates specific requirements for the use of pseudonymised personal data, to eliminate or minimise the risk of re-identification of subjects.

These conditions have been completed in BIFAP using different procedures, as described below.

4.3.1 Pseudonymisation

Article 89.1 of the GDPR states that technical and organisational measures should be established to ensure the principle of minimisation of personal data, including pseudonymisation and anonymisation of such data.

A pseudonymisation procedure is performed for this purpose, as a patient identifier is required for BIFAP purposes in order to link a patient's information received from Autonomous Communities' information systems with existing information from previous extracts of the same patient (see section 3.1.4.1). That pseudonymised unique identifier remains in successive database updates. To this end, two pseudonymisation procedures are carried out. The first, at source, at the Autonomous Communities (paragraph 3.1.43.1.4.1). The second is carried out at the AEMPS, by means of a new dissociation of the identifier received from the CCAA (paragraph 3.2.4.1). The geographic location variable of the primary care quota is also pseudonymised and is not accessible to researchers, except in cases duly justified by the study's goal.

This system, which includes a physical and functional separation of both pseudonymisation procedures, makes it impossible for any participant to carry out a re-identification of patients from their pseudonyms.

4.3.2 Deleting/Reducing Variables:

In application of the principle of data minimisation**,** the following measures are also carried out:

a) Among all the information contained in the HCE, only information corresponding to the BIFAP data model, which is necessary for the conduct of research studies, shall be extracted in the participating Autonomous Communities (paragraphs 3.1.1 and 3.1.2).

b) Researchers shall only access data contained in those variables and patient records that are necessary for the proposed study (paragraph 3.4.1).

c) The structure of the Data Files that are delivered to the Principal Investigators shall contain the variables necessary for the analysis, which shall include only structured information (paragraph 3.4.2).

d) Certain variables that may contain indirect identification data, such as the date of birth or death, are subject to data aggregation (paragraph 3.4.4).

4.3.3 Restricting the users who handle data

Access to the data recorded in the HCEs included in BIFAP is restricted to certain users who must also meet specific conditions.

The use of BIFAP is restricted to database administrator users at AEMPS, AEMPS researchers, and researchers belonging to public institutions who want to conduct research without commercial interest (paragraph 2.1). A protocol approval process (paragraphs 2.2 and 2.3) is carried out and access to commitments by researchers, confidentiality and restriction of their activities are conditioned to the exclusive goals of the approved research protocol (paragraph 2.4)

4.3.4 Data security measures

Physical and logical data security measures to prevent re-identification and access by unauthorized third parties are summarised below:

a) The BIFAP database is subject to all physical access controls applied by the Ministry of Health under the National Security Scheme

b) As an additional measure, data storage is performed on encrypted hard disks in order to avoid any opening of the files in the event of disk theft.

c) Access to the computers where the database is accessed requires a password. Only 4 people who are currently part of the BIFAP's IT Unit can access as password users.

d) Access to equipment is only possible at AEMPS headquarters. Access to the AEMPS headquarters is recorded at check-in and check-out.

e) The computer security measures applied at the software level are subject to the security criteria applied by the Ministry of Health in all its information systems.

f) All activity performed by users accessing BIFAP data is monitored by generating Programming Reports: with each output file the application keeps a report on the activity carried out by the user, in order to track and make it possible to trace the queries made in BIFAP with the tools, as well as the delivery of Structured Data Files for statistical analysis to external researchers.

4.3.5 Quality control of research

To submit scientific research to quality standards and, where appropriate, to international guidelines on good clinical practice (17th DA 2.f) 2nd):

BIFAP's Scientific Committee assesses the scientific suitability, technical quality and competence of protocol researchers prior to implementation (paragraph 2.3)

The BIFAP investigators' express commitment document (paragraph 2.4) incorporates the obligation to comply with applicable regulations, including the obligation to publish results in accordance with international standards in this field.

4.3.6 Previous report of the Research Ethics Committee

Paragraph 2 g) of the 17th DA stipulates that the use of pseudonymised personal data for public health research purposes, and in particular biomedical research, shall be submitted to the previous report of the Research Ethics Committee provided for in the sectoral regulations. This is one of the express commitments of researchers using BIFAP (paragraph 2.4).

4.3.7 Technical and functional separation between the research team and those who perform pseudonymisation

Technical and functional separation between the research team and those who perform pseudonymisation and store the information that makes re-identification possible (17th DA 2d) 1st):

The research database is generated in two distinct physical and organisational environments, with different stakeholders: computer managers of the participating Autonomous Communities and BIFAP team at AEMPS headquarters (paragraphs 3.1.3 and 3.2.3). The only identifying variable, the CIPA, is subjected to a layer pseudonymisation procedure, carried out by the respective treatment managers of the different organizations: Autonomous Communities (paragraph 3.1.4) and AEMPS (3.2.4). Finally, there is a technical and functional separation between the data exploder (the research team, either external to the AEMPS or from the AEMPS itself) and the personnel who perform the two pseudonymisation procedures (paragraph 3.4.3).

4.3.8 Express commitment to confidentiality and not to carry out any re-identification activities

The pseudonymised data are only accessible to the research team when there is an express commitment to confidentiality and no re-identification activity shall be carried out (17th DA 2d) 2nd i)):

All researchers have made prior commitments to any access to patient pseudonymised data, including, but not limited to, confidentiality and non-reidentification activities (paragraph 2.4)

4.3.9 Measures to prevent re-identification and access of unauthorized third parties

Where appropriate, measures should be taken to ensure that researchers do not access stakeholder identification data (17th DA 2.f) 3rd). In particular, pseudonymised data should only be accessible to the research team when specific security measures are taken to prevent the re-identification and access of unauthorized third parties (17th DA 2d) 2nd ii)):

In database generation, an obfuscation procedure is carried out that results in the deletion of potential personal data (paragraph 3.2.4.2) and the AEMPS maintains an Event Catalogue that reduces the need for access to unstructured information (paragraph 3.2.2.5). For the transfer of data to researchers, certain variables are subject to data aggregation (paragraph 3.4.4.3) and unstructured HCE data access, as well as the use of information selection and management tools, are carried out on identified and restricted computer equipment and all the activity carried out by the investigators is monitored (paragraph 3.4.4).

4.3.10 Rights of stakeholders

In accordance with Art. 9.2.j of the GDPR, the processing of health data for scientific research purposes should be proportional to the goal pursued, respect in essence the right to data protection and establish appropriate and specific measures to protect the fundamental interests and rights of the person concerned.

Chapter III (Articles 12 to 22) of the GDPR contains the rights of stakeholders that must be taken into account in the processing of personal data. The following refers to the manner in which these rights are exercised and the exceptions that are applicable in the exercise of these rights in BIFAP.

4.3.10.1 Right of information and transparency: publication of information through the website

Articles 12 to 14 of the GDPR (duty to inform the interested parties):

The duty to inform the interested parties, in accordance with Art. 12 of the GDPR, falls upon the data controller, who shall take the necessary measures to provide the interested party with the necessary information. Article 14 of the GDPR specifies the type of information to be provided where personal data have not been obtained from the data subject, as in this case.

BIFAP contains only pseudonymised data, without personal identifiers, of a large number (millions) of patient medical records from the year 2002 to the present, not obtained directly from the interested party, and whose purpose is scientific research.

The Spanish Data Protection Agency has published a Guide on the right to information, in which it is considered that it will not be necessary to inform the interested parties individually when the communication becomes impossible or involves a disproportionate effort. The appropriate means of reporting treatment activities is by making public the information (Art.14.5.b) of the GDPR).

To this end, the following information is available and updated on the BIFAP website (http://bifap.aemps.es):

- Relevant information updated on BIFAP's rationale, organisation and purpose.
- Reference to each of the investigations carried out using their data. References to the research carried out include the name of the mail researcher for each project, and, where appropriate, the publications derived from the research carried out.
- Conditions for access to the data, as well as the exercise of rights of the interested parties (or their justified exception) are detailed in this public governance document, which is accessed through the same website.

- Inclusion of an "Information Clause on Personal Data Processing" (document in Annex) summarising relevant information about BIFAP, and drafted following the guidelines of the ASSI-GDPR program in the DPIA carried out.

4.3.10.2 Stakeholders rights excepted in the data pseudonymised treatment in health research

The following GDPR articles refer to rights of the interested parties which must, in general, be guaranteed in personal data processing and which are granted an exception, in accordance with the rules themselves, in the case of data processing in BIFAP:

- **Article 15 (**right of access by the data subject**), Article 16 (**right to rectification**), Article 18 (**right to restriction of processing**) and Article 21 (**right to object**)**

Paragraph 2 e) of the 17th DA of the LOPDDD provides that the rights of the affected persons referred to in the articles referred to above may be excepted (point 2) when said rights are exercised directly before researchers or research centres using anonymised or pseudonymised data. This is the case with AEMPS as responsible for the processing of BIFAP data and the researchers to whom the data is transferred for research purposes.

- **Article 17 of the GDPR (**right to erasure ('right to be forgotten')**):**

In accordance with Sections 3 c) and d) from Article 17 of the GDPR, the right of deletion shall not apply, respectively, in cases where data processing is necessary for reasons of public interest in the field of public health. This coincides with the legal basis for the processing of data in BIFAP by the AEMPS (see section 4.1.1) and in cases where it is carried out for scientific research purposes (see section 4.1.2), to the extent that this right could make it impossible or seriously hamper the achievement of the objectives of such processing. The latter condition is fulfilled in BIFAP since it contains only pseudonymised data, without identifiers that allow access to the identity of the interested parties (paragraph 4.3.1).

- **Article 19 of the GDPR (**notification obligation regarding rectification or erasure of personal data or restriction of processing**)**

This obligation of the data processor refers to the rights of Articles 16, 17 and 18 of the GDPR, except in BIFAP.

- **Article 20 of the GDPR (**right to data portability**)**

This right is exercised only when processing is based on consent or a contract (Art. 20.1.a)), so it does not apply to data processing in BIFAP, which have different legal regulations from these.

- **Article 22 of the GDPR (**automated individual decision-making, including profiling**)**

Since BIFAP only contains pseudonymised data, without identifiers that allow for access to the identity of the stakeholders (paragraph 4.3.1), no individual decisions, whether automated or not, on the stakeholders are feasible.

5 MANAGEMENT OF UNFORESEEN EXCEPTIONAL SITUATIONS

Certain situations in relation to specific research projects may involve treatment of BIFAP data or transfer of data to third parties not provided for in this document.

In cases not provided for in this document, BIFAP's conditions of access to and transfer of information for that particular study shall:

- Seek advice from the AEMPS Data Protection Officer
- Be authorised by the Director of AEMPS, on the basis of a report from the person who is responsible for the management of the BIFAP Programme. Receive approval from the BIFAP Advisory Committee.
- Have a favourable report justifying the need by the BIFAP Scientific Committee.

In addition, in cases where the AEMPS considers it necessary, the Spanish Data Protection Agency may be consulted.

In the event of discrepancies between the researchers and the AEMPS in the interpretation of the researcher's commitments or in any other terms of access set forth herein, the AEMPS Management shall make a decision on this matter after consultation with the BIFAP Advisory Committee.

6 UPDATE AND REVISIONS

In order to adapt to technological developments, increase the efficiency and quality of research studies with BIFAP and adapt to the current data protection regulations, AEMPS shall annually review the need to update the document, taking into account at least any modifications relating to the following:

- Changes to the regulations in force, in particular those relating to personal data protection.
- Proposals made by the BIFAP Advisory Committee.
- Recommendations arising from international collaboration on personal data protection in health databases, in particular those based on HCE.
- Incorporation of other data sources into BIFAP.
- Changes to BIFAP data management procedures that affect document content.

7 ANNEXES

- Data Protection Impact Assessment Summary (DPIA)
- Information Clause on Personal Data Processing

BIFAP Data Protection Impact Assessment (DPIA)

January 2021

Introduction

An DPIA has been carried out using the ASSI-GDPR Tool (Information Systems Security Audit for compliance with the General Data Protection Regulation).

The following link provides a descriptive report of the application: https://www.aepd.es/sites/default/files/2020-02/premio-2019-buenas-practicas-rgpd-modb-mitramiss.pdf

This document summarises the procedure followed, as well as the conclusions and recommended actions already implemented, presented in the current version of the Access to BIFAP Data Governance Document. Complete detailed information on the methodology and results of the evaluation carried out with ASSI-GDPR is available to Research Ethics Committees reporting on research studies with BIFAP, as indicated in points (g) and (h) Section 2 of the 17th Additional Provision of Organic Law 3/2018, of 5 December, on the Personal Data Protection and the Guarantee of Digital Rights.

The AEMPS continues to carry out additional evaluation procedures, which, in accordance with the specific characteristics of BIFAP, are agreed with the Spanish Data Protection Agency in order to complement the evaluation already carried out with ASSI-GDPR.

Procedure

Paragraph 2. (f) 1 of the seventeenth Additional Provision of LOPDDD provides that for the processing of health data for research purposes, an impact assessment is required to determine the risks of processing that shall specifically include the risks of re-identification associated with data anonymisation or pseudonymisation. The adaptation of BIFAP to this requirement, by carrying out an Impact Assessment on Data Protection (DPIA), has been carried out in accordance with the guidelines of the Spanish Data Protection Agency by the following actions:

- On 4 December 2019, the BIFAP Advisory Committee approved the AEMPS proposal to carry out the DPIA.
- In December 2019, the AEMPS appointed the Data Protection Officer (DPO), a consultant necessary to carry out the DPIA.
- An DPIA was carried out, between January and November of 2020, using the ASSI-GDPR Tool (Information Systems Security Audit for compliance with the General Data Protection Regulation). Technicians from the BIFAP team and from the Coordination, Research and Information Systems Units, advised by the Data Protection Officer of AEMPS, participated in its work.

Conclusions

The conclusions of the DPIA using THE ASSI-GDPR tool are summarised below:

- BIFAP deals with health-related data from a vulnerable group, such as patients, that affects a large number of natural persons and is used in some technical data mining projects in the exploitation of data for public health research purposes, consisting of the use of automatic and semi-automatic algorithms to identify the medical diagnoses that are the subject of research studies.
- An analysis was made of whether the loss of confidentiality, integrity, availability, authenticity or traceability of personal data could lead to certain situations that are detrimental to the parties concerned. The impact of the harm was assessed first and then the likelihood of it occurring if safety measures were not applied. In the assessment carried out, the following were described:
    - Possible harmful situations: to give rise to discrimination, social harm, or damage to reputation.
    - Number of people is "high" (millions).
    - Relevance of the impacts of the damage is "medium" (serious fornatural persons, difficult to repair)
    - Probability of occurrence is "low".
- As in the analysis carried out previously, it was identified that the non-standardised free text information introduced by the physician is the most important potential source of data sensitive to the privacy of subjects, and in particular residual risks of re-identification. The need for access to this data to involve additional control was therefore confirmed, namely that access should be carried out exclusively on computer equipment located at the AEMPS headquarters or at the headquarters of the participating Autonomous Communities' health administrations (see section 3.4.4 of the Access to BIFAP data Governance document).
- Considering both the current design and the data protection measures at BIFAP, the Risk Analysis and Impact Assessment carried out indicated a favourable benefit-risk of the processing activity and certain measures were indicated.

Recommendations and measures implemented

- ENS type II measures should be implemented (Annex II to Royal Decree 3/2010, of 8 January, regulating the National Security Scheme in the field of Electronic Administration.)
    - A commitment is made to carry out the four procedures that are specific to BIFAP and to report to the AEMPS Information Systems area.
- It is advised that a protocol is defined that applies from the initial design of any processing: update of this document, pre-existing:
    - The BIFAP Data Access Governance document has been updated as a planning protocol that describes the data life cycle and the processing that is performed (section 3).
- It is advised that procedures are developed to address the different rights of stakeholders:
    - Once the applicable legal requirements have been assessed, including those stipulated in the LOPDDD in relation to the processing of pseudonymised data for

health research purposes, a “Information Clause on Personal Data Processing” is published on the BIFAP website.

- It is advised that a procedure is developed for managing data security breaches, indicating how to detect and document them, and how to report them to both the AEPD and stakeholders when necessary.
  - The AEMPS shall develop procedures for the management of data security breaches.

BIFAP research database

INFORMATION CLAUSE ON PERSONAL DATA PROCESSING

In accordance with REGULATION (EU) No 2016/679 of the EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and the free movement of such data, repealing Directive 95/46/EC (General Data Protection Regulation), you are INFORMED that the personal data collected will become part of the data treatment, the characteristics of which are as follows:

- **Treatment Name**: BIFAP research database
- **Treatment Description**: BIFAP is an electronic database of medical records of patients tended to in the Primary Care are of the National Health System, which aims to carry out pharmacoepidemiological studies aimed primarily at evaluating the use and safety of drugs.
- **Data Controller**: Medicines for Human Use Department (AEMPS)
- **Data Protection Officer**: delegado_protecciondatos@aemps.es
- **Purpose of treatment**: Conduct population-based studies on the use, safety and effectiveness of drugs, as well as estimates of the prevalence/incidence of certain diseases and medical diagnoses. The results are representative of actual clinical practice.
- **Personal Data Categories**: Health data recorded in the primary care medical record, electronic medicinal products dispensations in pharmacy, and diagnostic records of hospital admissions. BIFAP does not contain patient personal identifiers.
- **Stakeholders Categories**: Patients cared for by family physicians and primary care paediatricians of the National Health System of the participating autonomous communities.
- **Deadlines for the erasing of the different categories of data**: They shall be retained for as long as necessary to fulfil the purpose for which they were collected and to determine possible responsibilities that might arise from that purpose and from the processing of the data. The provisions of the Archives and Documentation Regulations shall apply.
- **Legal basis:**
    - GDPR.- Article 6. 1. e) Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller
    - GDPR.- Article 89. Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes
- **Consignees:**
    - Researchers in Spain assigned to public bodies (health care, university, research) and primary care physicians of the National Health System collaborating with BIFAP. Researchers should register (Treatment Name: BIFAP services user management). In order to receive the data, project researchers must undergo training, submit a protocol that must be approved, and sign commitments.
    - No assignments or transfers of individual patient data records to third countries are foreseen.
- **Source from which the recorded data comes:** The data have been recorded by healthcare professionals who care for patients and incorporated into the health services information systems of the participating autonomous communities and are included in BIFAP already anonymized.
- **Right to information and transparency**: Detailed information on the purpose, the type of data it includes, information flows, access management, participants, data exploitation and information privacy protection procedures, in compliance with current regulations, is available at http://bifap.aemps.es.