**Responses**

**Comments from Editor:**

**Editor: Janusic, Tania**

**Comments to the Author:**

**Thank you for your submission. We apologise for the delay in getting back to you. We would like to progress your paper pending completion of the items below:**

**1. General**

**Please remove the numbering of the headings and subheadings.**

Thank you. The numbering has been removed.

**Is this data collection unique to this health service?**

Although the different regions of Spain have a similar organization, this data collection is unique for the Navarre Health System.

**2. Key Features**

**This requires inclusion of the overall dates of the data collection.**

The starting date of BARDENA and the frequency of the information updates has been added into the “key features” section.

**Please add a brief point on data access.**

A brief point on data access has been added as a key feature. To accommodate this additional point, the original second and third key features have been merged.

**3. Data resource basics**

**You state “The data warehouse is continuously monitored and updated”. What do you mean by continuously monitored? By whom? Do you mean audits?**

The reviewer is right pointing out that the terms “continuously monitored” can be misleading. In fact, the data follows and automatic quality check process that includes format and range conformance checking that ultimatelly results in the generation of a quality report. Then, before the data is integrated in the wharehouse, all reports are reviewed by database administrator. The following information has now been added into the “data resource basics” section of the manuscript:

“*During the process of data gathering, transformation and integration, quality reports are automatically generated. These reports are revised and managed by the administrators of the database in order to detect and correct inconsistencies before entering data in BARDENA*”

**4. Data collected**

**“… where the different sources are stored in the backend and then linked all together in a star schema design”. How are they linked? It would flow better if you briefly describe how data are linked here. Is the same personal ID number (from LAKORA) used in the other datasets? (You describe data linkage later. It can be moved here).**

**The reviewer is right again. The primary key value by which all the information is linked is the patients ID number, gathered firstly in the LAKORA database present in the rest of databases allowing patient level data integration.**

We have moved the information regarding the identifier and data linkage to the “data collected” section.

**What do you mean by “main data”? Do you mean patient demographics? How is it different to location and diagnosis dimensions? It is not clear how these 5 dimensions relate to the data collected as listed in Table 1.**

Thank you, main data was misleading. We have integrated the information regarding the dimensions in table 1, in which a detailed description of the data included in each dimension is provided.

**Please expand on what is the Analytic Countability Component (ACC)?**

This module includes cost data of all the resources used and the interventions carried out during the care process, as well as information on the pharmaceutical co-payment assigned to each individual according to the income level. This new information has been added in table 1.

**Please include references or website links for the various sources of data (LAKORA, ATENEA, etc)**

These are software modules accessible from the internal information network of the Navarre Health Service. They are not available as webpage

**You state that BARDENA can be linked to specific registries. This is not shown in Figure 2. You mention acute stroke and diabetes registries. Please provide the full name for these registries and references/webpage links. Are there additional registries that can be linked? Please provide a list. Has consent been provided for linkage?**

Explicar que hay registro dentro de la información de hospital. Añadir en la información de hospital en la Figura 2 -> y Javier.

**Please move the text under “2.3 Access to BARDENA information…” to the “Data resource access” heading.**

We thank the reviewer for the suggestion. We have moved this information to “data resource access” section.

**Under “Ethics clearance” it is not clear what you mean by “unless informed consent” from patients” “allowing access to their data”. Do you mean non-deidentified data can be accessed also, if the patient gives consent? Can patients opt out of having their data anonymised and collected by BARDENA? Does BARDENA provide ethics clearance? Who are the data custodians?**

The data extracted from BARDENA is always de-identified to protect patients´ privacy. We apologize for any confusion caused. We have modified the text accordingly to avoid further misunderstanding.

Regarding ethics clearance, it is important to note that BARDENA does not provide ethics clearance. As stated in the text, approval of the study protocol by an accredited ethical research committee is required to conduct research involving BARDENA data..

Lastly, the Navarre Health Service serves as the data custodian for BARDENA, ensuring the proper management, safeguarding, and responsible use of the collected data. It is committed to upholding privacy, security, and ethical standards in handling the data.

The information about the data custodian has been added in the revised text.

**5. Data Resource Use**

**Please provide more detail on what research questions were addressed, with a couple of examples with details on what was found.**

Añadir en texto sobre las preguntas de investigación y los resultados obtenidos –> (COVID-19, Diabetes, SURBCAN, etc.) -> Ibai

**“BARDENA also contributes to Spanish national networks, such as the Atlas of Variations in Medical Practice Variation”. It is not clear what this involves. Please provide some more detail on how this resource is used by these networks.**

Desarrollar la explicación del Atlas (añadir URL). -> Ibai

Atlas of Variations in Medical Practice of Spain is an initiative aimed at describing systematic and unwarranted variations in medical practice at geographic level-building upon the seminal experience of the Dartmouth Atlas of Health Care. In order to make posible this effort, from navarre Bardena provides information of the hospital admissions , regarding demographic, diagnoses, procedures, dates and types of admission and discharge. As a result, geographical veriations in several dimensions can be analised, such as the small area rate of avoidable complications of diabetes, among others.

**Please expand what you are describing under section 4.2 in the first paragraph. The first sentence does not inform the reader on what is actually being done (ie “At this moment BARDENA information from the hospital setting is integrated with that of other regions in the MBDS, which is piloted at national level”.)**

Thank you for the suggestion. We have modified the text for a better understanding:

“Individual patient data corresponding to hospital health care (demographics, medical diagnoses, procedures, in-hospital mortality, etc.) included in BARDENA is transmitted to the Ministry of Health at the national level. This information is integrated with the data of the rest of the Spanish regions in the MBDS database, which is the database of reference of the Spanish national health system. This allows for interoperability of information and analysis of variability between the Spanish hospitals”

**Please expand on what is being integrated in the EHDEN and for what purpose.**

Thank you. We have included more details on EHDEN in the revised version of the manuscript:

“*Moreover, at this moment BARDENA is in the process of being integrated into the European European Health Data & Evidence Network (EHDEN), a European consortium of data sources from 12 countries funded by the European Union´s Horizon 2020. EHDEN was launched in 2018 and aimed at building a standardized large-scale network to reduce the time needed to provide answers in real world health research. This involves transforming data into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), an open community data standard designed to standardize the structure of the data from the different sources. This allows to other organizations to analyze and consult our data warehouse through the Observational Health Data Sciences and Informatics (OHDSI) program*”.

**Please confirm that the studies in Table 2 were performed by accessing the datasets through the BARDENA platform/warehouse, and were not accessed separate to BARDENA.**

Thank you. Yes, we confirm that these studies were conducted based on data extracted from BARDENA.**6. Strengths and weaknesses**

**This section is OK.**

**7. Data resource access**

**Note that at least one of the authors of a Data Resource Profile must be a researcher involved with the data resource presented and be available to readers as a contact person.**

**Please provide the name and contact details of a co-author to whom enquiries can be submitted.**

Thank you. The contact person is Javier Gorricho, leading author of the manuscript and Head of the Service of Evaluation and Dissemination of Health Outcomes of the Navarre Health Service. Javier Gorricho is the responsible for the management of BARDENA and who handles the data requests. He will be the contact person to whom enquiries should be directed.**8. Figures and Tables**

**Could you please confirm that the figures and tables used in this paper are original and have not been copied or modified/adapted from those on the study website or published in other articles? If any figure or table is being reproduced or adapted, you need to receive written permission from the copyright owner, and this needs to be appropriately acknowledged in the figure/table caption.**

Thank you. We confirm that figures and tables are original and have been generated specifically by the authors for this manuscript. They have not been copied or published elsewhere.

**9. Table 1.**

**Is lifestyle data limited to only smoking, alcohol, physical activity?**

Yes. Right now, these are the only lifestyle parameters integrated in BARDENA that have been validated and are available for use. However, many otherther lifestyle variables are also available such as, performance in activities of daily living (e.g. Barthel), knowledge and understanding of the patients own diseases, awareness of signs of risk factors,or diet. However, this information is only available in a nonstructured format, and the proper parametrization and validation process are still under review.in free text,

**Are these data captured from primary healthcare data? How is physical activity measured?**

**Exactly, the data is captured from primary health care. Similar to aforementioned lyfestyle data, the data is originally filled by the general practitioner as nonstructured text. Then, through natural languaje processing algorithms, the activity is classified in three categories: Active patient, inactive patient and partially active patient.**

**Under Diagnostic procedure, what do you mean by “analytical”?**

We apologyze for the term. Instead of “analytical” we should have said “laboratory tests”. We refer to laboratory data. We have modified the trem accordingly in table 1.

**What “vital signs” are included under nursing interventions?**

We refer to blood pressure, heart rate, respiratory rate and oxygen saturation, skin coloration, hydration state, and awareness level. We have listed them in the table 1 of this new version of the manuscript

**What do you mean by interventions of social workers? Which variables are collected?**

Examples of interventions of social workers are providing support to patients and caregivers, assisting patients in health care transitions, managing social health centers, processing the applications for financial and social assistance, building community connections, and patients´ functional capacity assessment. We have specified them in the revised version of the manuscript.

BARDENA includes information on the type of intervention done, the result of the intervention, and of the beginning and end of the intervention, among others.

**Under “contact with other services” what do you mean by “other services”? Community centres? Aged care centres? Homeless shelters?**

Son sanitarios -> Hospitales de día, hemodiálisis, UCA, etc. -> Javier

Thank you. We refer to palliative and home care hospitalization units and to public social care centers. We have cited them in table 1.

**10. Table 2**

**Please include another column in Table 2 and include which specific datasets were used on the relevant thematic areas.**

Añadir la columna.

**11. Figure 2.**

**It is not clear what the arrows refer to. If they indicate each of the datasets that contribute data, then there should be 7 arrows clearly linking the datasets to BARDENA in the centre.**

**It is not clear why there are from 1-3 coloured cylinders and what the number of these represents. (eg Pharma data includes LAMIA and FARHO, but there are 3 cylinders in the diagram).**

**In the text you mention that BARDENA may link to disease registries, but this is not shown in the diagram.**

**It is not clear what the lock refers to.**Originally lhe lock represented the secure storage and pesudoanonimization process the data goes through, however we decided to get rid of it in the final figure. Thanks for pointing out its unclear meaning.