Building Infrastructure to Exploit Evidence from Patient Preference Information (PPI) Studies: A Conceptual Blueprint

Luca Giordano 1,2,*, Andrea Francavilla 1, Corrado Lanera 1, Sara Urru 1, Paola Berchialla 3 and Ileana Baldi 1, *

1 Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac Thoracic Vascular Sciences and Public Health, University of Padova, Via Loredan 18, 35121 Padova, Italy; luca.giordano.1@phd.unipd.it (L.G.); andrea.francavilla@studenti.unipd.it (A.F.); corrado.lanera@unipd.it (C.L.); sara.urr@phd.unipd.it (S.U.)
2 ClinOpsHub s.r.l., 72023 Mesagne, Italy
3 Department of Clinical and Biological Sciences, University of Turin, 10124 Torino, Italy; paola.berchialla@unito.it
* Correspondence: ileana.baldi@unipd.it

Abstract: Patients are the most important actors in clinical research. Therefore, patient preference information (PPI) could support the decision-making process, being indisputable for research value, quality, and integrity. However, there is a lack of clear guidance or consensus on the search for preference studies. In this blueprint, an openly available and regularly updated patient preference management system for an integrated database (PPMSDB) that contains the minimal set of data sufficient to provide detailed information for each study (the so-called evidence tables in systematic reviews) and a high-level overview of the findings of a review (summary tables) is described. These tables could help determine which studies, if any, are eligible for quantitative synthesis. Finally, a web platform would provide a graphical and user-friendly interface. On the other hand, a set of APIs (application programming interfaces) would also be developed and provided. The PPMSDB, aims to collect preference measures, characteristics, and meta-data, and allow researchers to obtain a quick overview of a research field, use the latest evidence, and identify research gaps. In conjunction with proper statistical analysis of quantitative preference measures, these aspects can facilitate formal evidence-based decisions and adequate consideration when conducting a structured decision-making process. Our objective is to outline the conceptual infrastructure necessary to build and maintain a successful network that can monitor the currentness and validity of evidence.

Keywords: PPI; PPMSDB; systematic reviews; analytic infrastructure

1. Introduction

Research has always been interested in factors that promote the use of health services and influence patients’ attitudes to health [1]. Help-seeking behaviour (HSB) is a way of finding a medical solution through appropriate interaction with medically trained professionals [2]. There are two dominant approaches to HSB: developing a model of the HSB pathway to describe the individual’s actions and research into factors determining behaviour and identifying factors influencing this path [1]. Treatment-seeking behaviour, instead, is the sequence of actions and an integral part of the identity of an individual, family or community that patients and caregivers take to resolve their health problems [3]. Various treatments exist for several pathologies [4–6]. It is interesting, therefore, to understand the treatments that patients (or cohorts of patients) prefer to undertake and their real needs. In recent years, patients’ voices are becoming more critical for companies developing new medical products and for the authorities assessing, regulating, and deciding which products are effective, safe, well-tolerated, and cost-effective [7–13].
Aligning health care policy with patient preferences could improve the effectiveness of health care interventions by implementing the adoption of, satisfaction with, and adherence to clinical treatments or public health programs [14,15]. Therefore, it is fundamental to know the preferences of a treatment expressed by the patients [16]. This type of data is referred to as patient preference information (PPI). PPI is defined by the Food and Drug Administration (FDA) as “qualitative or quantitative assessments of the relative desirability or acceptability to patients of specified alternatives or choices among outcomes or other attributes that differ among the alternative health interventions” [7]. PPI aims to capture patients’ needs and provide a significant opportunity for patients to express their preferences [17]. PPI can be assessed through patient preference studies (PPSs) using preference exploration or elicitation methods [16]. The former are qualitative studies, such as patient interviews or focus groups, that examine the patient’s subjective experiences and decisions [18]. The latter are quantitative methods, mostly adopting approaches developed in health economics, collecting quantifiable data for statistical analysis. Therefore, PPI aims to capture patients’ needs and provide a significant opportunity for patients to express their preferences; in summary, PPI enhances the possibility of integrating patient preferences into the decision-making process [17].

There is an emerging consensus that the patient perspective should be incorporated into decisions in the medical product lifecycle (MPLC) [19–24]. Furthermore, PPI can be used in every phase of MPLC to identify unmet medical needs, to notify the selections of endpoints [25], and to inform about benefit-risk assessments [13]; in fact, PPI can give insights into the trade-offs that patients make between benefits and risks and show the relative importance of outcomes for patients [12,26,27]. The evidence also demonstrates that integrating PPI into clinical practice can optimise symptoms management, supportive therapy, and patient-centered care and ultimately benefit survival during oncological treatment [28–32]. Furthermore, patient-centric decision-making results in better transparency and accountability in the development of medical products. It may also improve the quality of research and make study outcomes more relevant to the patients, with more products developed in line with patients’ needs [17,22,33]. A recent systematic review of attempts and initiatives about using PPI in benefit–risk assessment, published in late October 2020, concludes that patient preference elicitation tools are largely understood, and researchers and experts perform their use better. Unfortunately, despite the efforts identified and the initiatives undertaken, the pace of progress remains slow [34].

Moreover, evidence of proper use of these data in policy decision-making is lacking as PPI remains poorly implemented: therefore, many questions about PPI validity, representativeness, and robustness remain unanswered. Finally, there is a lack of guidance on conducting PPSs to inform decision making, possibly explaining the limited use of these studies [24]. However, most stakeholders consider PPI essential in informing future decision-making across the MPLC [35]. In addition, including PPI could improve the transparency and acceptability of regulatory or reimbursement decisions [36]. Given these considerations, there is a need for infrastructures that can monitor the contemporaneity and validity of the evidence.

This study aims to provide a conceptual blueprint for an openly available and regularly updated patient preference management system for an integrated database (PPMSDB), collecting preference measures, characteristics, and meta-data, which could allow researchers to quickly overview a research field using the latest evidence and data to identify research gaps. Different initiatives [37,38] address the methodological issues and provide recommendations and guidance on the design and conduct of PPI studies. Soekhai et al. [18] have identified 32 methods for PPSs: 10 exploration and 22 elicitation methods. According to the type of data collection and analysis, these study methods were assorted to different groups. The exploration methods can be divided into three distinct groups—individual methods, group methods, and individual/group methods—while the elicitation methods can be classified into four distinct groups—discrete—choice based methods, ranking methods, indifference methods, and rating methods [18]. In conjunction with proper
statistical analysis of quantitative preference measures, these aspects may facilitate formal evidence-based decisions and adequate consideration when conducting a structured decision-making process.

The PPMSDB would enable the evaluation of previous studies and the continuous inclusion of new studies. As such, it would respond to the need for cumulative data aggregation approaches in addition to a retrospective summary of past studies.

2. Materials and Methods

The PPMSDB should fulfil the following:

1. Define a structural architecture of hierarchical databases suitable to represent the complex interplay between outcomes of interest, attributes investigated, and levels considered. So that, on the side and in synergy with information such as the variables considered, preferences metrics’ measures and weights, and all the relevant meta-data regarding the PPS linked.

2. Collect and harmonise data from PPS based on quantitative methods (elicitation preference).

3. Allow researchers to insert their PPS data in autonomy (based on quantitative methods) manually and programmatically [39].

4. Allow for peer-review of the data inserted, with a “not-checked” label/meta-information for PPS inserted from public resources that still miss an external and independent confirmation of the record’s contents [39].

5. Permit the extraction and download of preferences' weights based on categorical and textual criteria and filters selected or inputted by the users. Textual search should allow at least for exact word matching and the more flexible regular expression searching on the PPS textual fields, such as title, abstract/description, inclusion/exclusion criteria, attributes, outcomes, etcetera. Users should refine the final selection provided by criteria and filters, including the possibility of manually excluding single studies of interest [40].

6. Permit the computation and reporting of pooled PPI of interest on the PPS selected, such as a simple list of weights and outcomes, their summary statistics of localisation and dispersion (i.e., minimum, maximum, mean, median, standard deviation, and inter-quartile range), and regression analyses on them (possibly allowing interactions and non-linear effects) [41].

7. In the input phase, permit to compute derived metrics of interest unreported in the original PPS [41].

8. Allow the download of a report, including the selections to obtain the information retrieved, giving a formal context to the data extracted [42].

9. Possibly, track the number of filtering explored up to the exported data to permit correction for multiple tests in selecting proper data [43].

10. Designed to allow for horizontal (number of instances) easily and vertical (computational power) scalability [40].

11. Adhere to the four open science principles: open access, open methodology, open data, and open source [44].

The aim of the PPMSDB is to enable the evaluation of previous studies and the continuous inclusion of new studies. As such, it will respond to the need for cumulative approaches for data aggregation in addition to a retrospective summary of past studies.

Nowadays, several databases are dedicated to patients and others to clinical trials. However, none of these collect information on patient preference studies, although the use of these data can be beneficial throughout the MPLC (Figure 1).
Figure 1. Summary of the main milestones and benefits of the PPMSDB.

Structure of the Collector/Database

Data collectors are computer tools, components of the structured query language (SQL) that allow collecting different data sets in a relational database. Since our goal is to obtain digital evidence from the database, we deal with data structure extraction. Some data structures can be used for the table relationship search technique. The data structure extraction consists of three processes. The first process, called attribute extraction, compares the field name and the task to be performed. The second process is the key extraction: after obtaining the primary and external keys, it confirms table relationships by checking the vertical relationships between them. The third process is constraint extraction. It obtains cardinality between the primary key and the external key.

Finally, data collectors check the validation of the relationship between two tables by comparing 1 to 1 or 1 to many relationships in the related two tables.

The tools to achieve the intended objectives are the following:

MySQL relational database: every entity should be defined with its properties and relations. In this context, the same attributes could be defined with multiple names or have multiple sets of levels considered by different PPS. Treatment can be investigated against various sets of attributes by different PPS, but multiple PPS can consider the same attribute. Each PPS included in the PPMSDB has its own set of variables collected and meta-data information, which can overlap with others.

Source-standard: for every type of PPS, a template is defined for explicitly inputting the PPI and all the linked variables and meta-data to link with a corresponding PPMSDB standard meta-information (i.e., table and field of the database). In this phase, the user can ask additional information (e.g., unit of measure or conversion factors for non-standard units).
Standard-PPMSDB: once the system collects the information linked to the standard meta information, they are processed and incorporated in the PPMSDB in a uniform and standard internal format, keeping a back-trace to the original data and computation to allow to retrieve and use of the information both with the standard uniform and the source original shapes and format.

Interactive web interfaces, APIs, downloadable templates (JSON schemas or Excel spreadsheets), and tutorials will be developed to allow the user to populate the PPMSDB quickly, manually, and programmatically.

The system can be authorised with three distinct levels of privileges: as a user, they can explore the PPMSDB and all its active functionalities; as an author, they can add or update PPS into the PPMSDB; and as a reviewer, they can check PPS content in the PPMSDB for consistency and purpose changes accordingly. Simple tags are attached to every record reporting review status. Moreover, a machine learning model could be created and trained on the reviewers’ activity to produce an artificial intelligence system that can support the reviewer’s job and produce automatic reviews (with a dedicated tag).

Incorporating a dedicated R back-end behind the web interface to the PPMSDB (which will be developed as an R-Shiny application) can activate the regular-expression engine. We chose to rely on R because, beyond the coding language, it is supported by an active community that covers various areas of interest, such as the development of the database management engine, the creation of the web-interface, the definition of the API, the creation of high-level graphic plots, and the implementation of state-of-the-art analysis systems and models. Moreover, the result from the user query would be managed directly by R software, so any further refinement (i.e., filtering) can be easily implemented. For adding a specific single report to the resulting query, an ad hoc template can be automatically produced based on the query result and provided to the user (JSON schema or excel spreadsheet). Each new addition would be processed and included in a report, describing the overall process leading to the query result (including all PPS and pre-processing). At the same time, R could provide the possibility to customise the interface, expanding a specific PPS to be part of the PPMSDB.

R software will be the back end of the whole system of the web interface and API to the PPMSDB. It could be directly used to compute the required statistics. A dedicated R-package will be developed, tested, and validated for any specific task. Empowering the capabilities of the R packages [knitr] (to parse text and code), [rmarkdown] (to permit high-quality and flexible text personalisation), [pander] (to create outputs of the format of choice from the computed code interleaved with the processed text), and [targets] (to keep track of the entire pipeline followed by every piece of data, function or code, from the first input to the output), within the interactivity of Shiny, a report can be produced reporting all the meta-information of choice.

Thanks to the login credential system (possibly activated automatically by implementing the open-source ShinyProxy platform), every user would be unique. Assigning each (set of) queries to a user-specific project, the number of tentative queries can be tracked as well as the variations within them.

Docker could be used to separately encapsulate the PPMSDB, the API engine, the Shiny interface, and the ShinyProxy itself: horizontal scaling would be straightforward, while for vertical scaling, the entire system could exploit cloud computational services such as Microsoft Azure or Amazon AWS (both accept Docker containers).

Moreover, to avoid downtime, the whole Docker ecosystem will be deployed and managed under Docker Swarm or Kubernetes.

The development of all the code could track changes, updates, issues, and contributions by using git, hosting the repository on GitHub and the containers on Docker Hub. To facilitate the experience of use, such a methodology will be described as internal documentatation of all the functions developed and as a wiki page and tutorials. Raw and processed data will be explored and retrieved from the web interface. Everything should be free of charge for the end-user.
All the mentioned software and tools are open-source and free for non-commercial uses.

3. Discussion

A recent survey [45] of the medical literature shows that there has been an increase in systematic reviews on patient preferences, suggesting increasing interest in conducting preference-eliciting studies and systematic reviews of these studies. Therefore, PPMSDB infrastructure could be helpful to monitor the currentness and validity of the evidence.

Open access to PPMSDB data and metadata would provide pre-existing research with a usable and sustainable future. For example, extracted metadata and coding could be used to update a meta-analysis or even conduct another meta-analysis on a similar subject with an overlap in the relevant literature. This is crucial for several reasons. First, accumulating science and keeping evidence updated is a cooperative task, and participation in this task must be supported and incentivised, in addition to the single classical publication. Second, an open-access PPMSDB would enable the application of state-of-the-art statistical procedures to existing data and test these effects on the meta-analytic results. Third, open access to existing meta-data provides other researchers with research questions the opportunity to use subsets of the pre-existing meta-analytic data.

The ongoing accumulation of evidence would inform the researchers about the latest findings in a specific research area; for example, they could estimate when the results are robust enough to justify further research investment. When multiple preference-eliciting studies are done, a systematic review of these studies may then be needed to synthesise and summarise the study findings.

Some limitations must be acknowledged. This project requires continuous funding and support, not only in the development and implementation phase but also afterwards, to ensure maintenance and updating.

Another limiting factor could be the wide variability of PPSs. In a preliminary phase, precise guidelines should be drafted for the requirements for the inclusion of PPSs into the PPMSDB. In drawing up these guidelines, however, we can refer to those issued by the Innovative Medicines Initiative–Patient Preferences (IMI-PREFER) and the Health Preference Research–International Health Economics Association [37,38] for the design and conduct of PPSs. Moreover, having guidelines that can uniquely unite PPSs may increase the quality standards of such studies.

4. Conclusions

PPI could support decision-making throughout the MPLC [35], from industry to health technology assessment (HTA) through regulatory authorities. The rationale for PPI importance is indisputable for research value, quality, and integrity. Thanks to the above-mentioned process, researchers/sponsors/stakeholders could elaborate on some critical PPI that can be useful to design better study protocols and take go/no-go development decisions during the MPLC. Furthermore, PPI could help define the study’s endpoints during the clinical development phases and make better decisions [35]. This project could also be helpful for the Competent Authority and Ethics Committees to decide if a study could fit the patients’ needs or supply scientific advisors. Competent Authorities could consider the PPI provided by pharmaceutical companies, for example, during the submission and validation phase, for scientific opinion or during commission decisions [35]. Finally, the patients could benefit from this project to acquire greater awareness of the therapeutic possibilities related to their disease and make more informed decisions [34].

In conclusion, implementing a PPI could enhance many unexplored possibilities from different points of view and help reach a point where the patient is at the centre of the decision-making process.

Author Contributions: Conceptualisation, I.B. and P.B.; methodology, L.G. and C.L.; writing—original draft preparation, A.F. and L.G.; writing—review and editing, C.L., A.F., L.G. and S.U.; supervision, I.B. and P.B. All authors have read and agreed to the published version of the manuscript.
**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


