

Statistical Analysis of Abilities to Give Consent to Health Data Processing

Antonella Massari¹, Biagio Solarino², Paola Perchinunno¹, Angela Maria D’Uggento³,
Marcello Benevento², Viviana D’Addosio¹, Vittoria Claudia De Nicolò¹, Samuela L’Abbate¹

¹Department of Economics, Management and Business Law, University of Bari Aldo Moro, Bari, Italy

²Institute of Legal Medicine, Department of Interdisciplinary Medicine, University of Bari Aldo Moro, Bari, Italy

³Department of Economics and Finance, University of Bari Aldo Moro, Bari, Italy

Email: antonella.massari@uniba.it, biagio.solarino@uniba.it, paola.perchiunno@uniba.it, angelamaria.duggento@uniba.it, marcello0835@gmail.com, viv.daddo@gmail.com, claudia.denicolo@poliba.it, samuela.labbate@uniba.it

How to cite this paper: Massari, A., Solarino, B., Perchinunno, P., D’Uggento, A.M., Benevento, M., D’Addosio, V., De Nicolò, V.C. and L’Abbate, S. (2024) Statistical Analysis of Abilities to Give Consent to Health Data Processing. *Applied Mathematics*, 15, 508-542.

<https://doi.org/10.4236/am.2024.158032>

Received: June 30, 2024

Accepted: August 16, 2024

Published: August 19, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The recent pandemic crisis has highlighted the importance of the availability and management of health data to respond quickly and effectively to health emergencies, while respecting the fundamental rights of every individual. In this context, it is essential to find a balance between the protection of privacy and the safeguarding of public health, using tools that guarantee transparency and consent to the processing of data by the population. This work, starting from a pilot investigation conducted in the Polyclinic of Bari as part of the Horizon Europe Seeds project entitled “Multidisciplinary analysis of technological tracing models of contagion: the protection of rights in the management of health data”, has the objective of promoting greater patient awareness regarding the processing of their health data and the protection of privacy. The methodology used the PHICAT (Personal Health Information Competence Assessment Tool) as a tool and, through the administration of a questionnaire, the aim was to evaluate the patients’ ability to express their consent to the release and processing of health data. The results that emerged were analyzed in relation to the 4 domains in which the process is divided which allows evaluating the patients’ ability to express a conscious choice and, also, in relation to the socio-demographic and clinical characteristics of the patients themselves. This study can contribute to understanding patients’ ability to give their consent and improve information regarding the management of health data by increasing confidence in granting the use of their data for research and clinical management.

Keywords

Privacy, Health Data, Consent, Cluster Analysis, Logit

1. Introduction

The serious Covid 19 pandemic crisis that has affected the whole world has brought to the fore the important issue of the processing of health data, due to the population's reluctance to release the information necessary for contact tracing to contain the pandemic.

This paper starts from an in-depth and detailed analysis conducted, by the same research team, on the literature review [1]-[79] related to the topic of health data processing and privacy protection in the pandemic period 2020-2022, [80] which highlighted any gaps or unresolved issues in the current literature that this study aims to address.

This study differs from previous research because for the first time, it uses a methodological tool to analyze patients' abilities to give consent to health data processing, developed by the medical area involved in the recalled project.

Although numerous legislative provisions have been introduced to protect the right to privacy, the need to identify new ways to balance the interests of protecting the right to privacy of each individual and the need for government and health authorities to have access to data remains fundamental that can safeguard collective health, especially in cases of widespread infections, allowing a rapid and effective response in resolving health emergency situations.

In the case of the COVID-19 epidemic, the dissemination of health data was crucial for monitoring the spread of the virus, identifying contagion clusters and adopting prevention and control measures, but also a source of concern for the privacy of patients who showed reticence to release of your data.

It is important to find a balance between the need to use health data for public health purposes and protecting patient privacy. In this context, informed consent is an essential tool to ensure that patients are fully informed and aware of the processing of personal health data and can freely express their consent or refusal to such processing.

It is, therefore, necessary that patients are fully informed about how their data is used, and shared and for what purposes.

At the European level, the GDPR was launched to protect health data, which presents itself as an important tool to guarantee the right to privacy and the protection of personal data, balancing the need to protect public health with respect for individual rights, imposing that organizations must guarantee the protection of personal data through appropriate technical and organizational measures, promoting transparency and responsibility in data processing.

It is crucial that the principle of data minimization is respected, which implies the collection and processing of only the data strictly necessary to achieve the purpose for which they were collected, always guaranteeing the security and confidentiality of users' personal data, adopting adequate security measures to protect data from unauthorized access or improper use.

Furthermore, it is necessary to provide users with clear and transparent information on the methods of processing personal data, any purposes of the pro-

cessing and any subjects involved in the processing of the data. It is therefore essential that we operate in accordance with the principles of the GDPR, guaranteeing security, confidentiality and transparency in the processing of users' personal data.

The pilot investigation conducted in the Bari Polyclinic, as part of the Horizon Europe Seeds project entitled "Multidisciplinary analysis of technological contagion tracking models: rights protection in health data management", of which the data are analyzed in this work aims to overcome the distrust that dominates patients in granting the use of their health data and aims to analyze and evaluate the ability to consciously express their consent to the processing of health data.

The methodology used is the PHICAT (Personal Health Information Competence Assessment Tool), a useful tool for assessing the patient's abilities in making decisions regarding the processing of personal health data, developed by the medical area participating in the aforementioned project.

The PHICAT is an instrument for assessing the ability to give consent to the processing of health data that uses a semistructured interview (Attachment 1) that assesses patients' abilities relative to four domains of the ability to give consent to the processing of health data: understanding, evaluation, reasoning and final expression of a choice.

The interview is semistructured in that the interviewer fills in some of the items submitted to the patient in relation to the different purposes of health data processing considered in the interview and uses a Notation Sheet (SN) that reports the information communicated to the respondent and the related responses noted.

The interview involves a preparation phase preparatory to the interview itself and then the scoring by the interviewer and the evaluation of the answers given by the patient.

The PHICAT interview is structured to allow an in-depth assessment of the patient's ability to understand, evaluate, reason and express a choice regarding consent to the processing of personal health data, considering the purposes of the data processing, the benefits and resulting risks. In this interview, patients also consider the effects that their choice to grant or not consent has on their daily lives, on the community and on scientific progress. Each section and sub-section of the interview aims to examine specific aspects of the patient's abilities in relation to four domains: understanding, evaluation, reasoning and final expression of a choice.

The purpose is to evaluate the interviewee's ability to grant or not consent to the processing of data and to what extent his decision is the expression of a conscious choice, analyzing the logical path that leads to granting or denying such consent.

First, the interviewer explains to the patient how the interview will take place, to encourage his active participation, so that the interviewee can ask questions that promote better understanding and continuation of the interview. To achieve the desired purpose, the interviewer's communication must adapt to the patient,

i.e. his cultural level and his emotional conditions, so that the information is clear and produces valid results.

In the first domain, that of understanding, it is verified that the patient has correctly understood the information provided by the interviewer before starting the interview. The patient is then asked to describe and repeat the information acquired. The objective, at this stage, is to ensure that the patient has a complete and correct understanding of the information provided by the interviewer, so that he or she can make informed decisions regarding the processing of personal health data.

The evaluation domain is aimed at ensuring that the patient is fully aware of the risks and benefits of processing health data and that he can express informed consent regarding the different and possible purposes of their use. Attention must be paid to the patient's beliefs which may be influenced by irrational anxieties and fears related to the processing of health data, such as fear of control by institutions or loss of privacy or may be influenced by a poor understanding of the purposes and objectives, benefits of processing health data.

The reasoning domain allows us to analyze the patient's cognitive and decision-making processes in depth, thus providing important information on the motivations and reasons underlying his choices. The interviewer analyzes together with the patient the reasoning that led him to make a certain choice, asking him the reasons and possible consequences of his decision.

The reasoning is evaluated both in qualitative terms, considering the reasons provided by the patient to explain his choice, and quantitatively, evaluating whether the patient can enumerate the rational consequences of his decision.

Finally, it is verified whether the choice made by the patient is unambiguous and aware, furthermore whether the patient understands the effect that the decision made has on the personal situation and that of the community, above all whether he is able to evaluate the consequences of his choices in light of the objectives that they wish to pursue the processing of health data, *i.e.* to feed the health file, or for scientific research purposes.

The investigation also evaluates logical coherence, *i.e.* whether the patient's final choice derives logically from his reasoning.

The methodology used for this analysis involves the attribution of scores by the interviewer which provide, at the end of the interview, a preliminary assessment of the patient's decision-making capacity, but which must not be used as the sole parameter to determine the inability to provide consent to the processing of health data. It is essential that the assessment of decision-making capacity is conducted by qualified personnel considering all clinical and contextual factors that may influence the patient's ability to make informed decisions on the processing of health data.

2. General Information

2.1. Age, Gender and Province of Residence

The respondents were 70% male and 30% female (**Figure 1**). The average age of

female respondents is 34 years: 27% are aged between 18 and 22 and 20% are distributed in the 23 - 27 and 38 - 42 age groups. The average age of male respondents is 37 years: 26% are aged between 23 and 27 and 11% are evenly distributed between four age groups: 28 - 32, 33 - 37, 48 - 52 and 52 - 56 years (Figure 2).

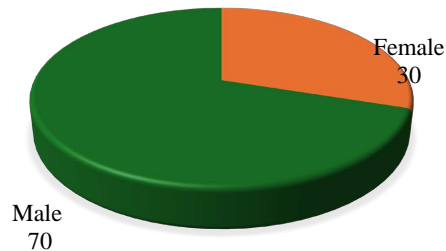


Figure 1. Percentages: gender.

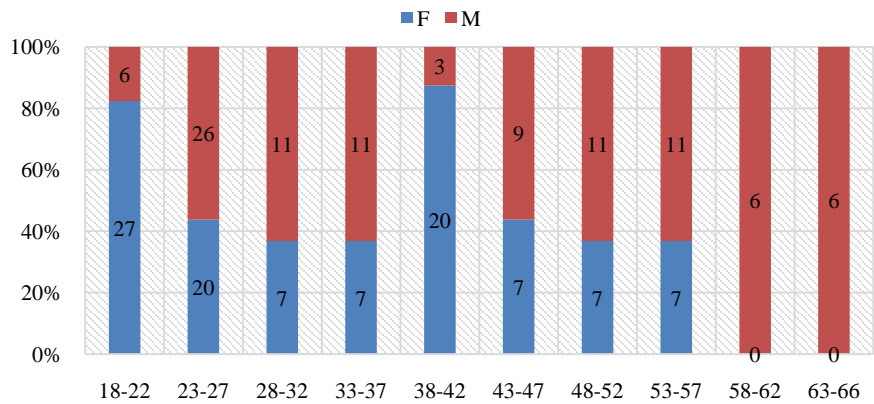


Figure 2. Age in classes and percentage values based on gender.

Almost all (86%) of the interviewees reside in the province of Bari, of which 46% are in the municipality of Bari and the remaining 40% are in different municipalities of the same province (Figure 3).

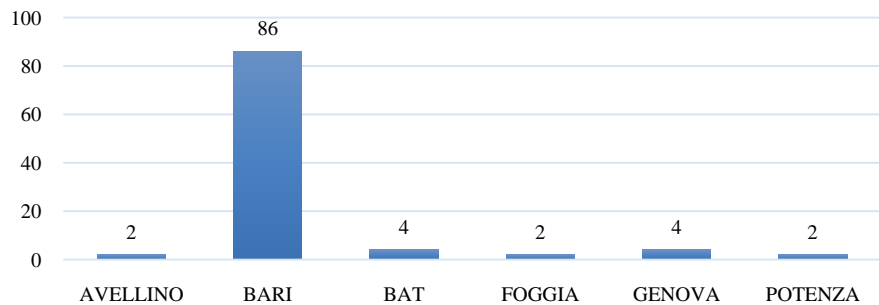


Figure 3. Percentages of provinces of residence.

2.2. Social Information

Education and Information

Just over half of both male and female respondents have a high school diploma,

34% of male respondents and 27% of women have a degree (Figure 4). 52% of those interviewed are employed, 16% are freelancers and 24% are students (Figure 5).

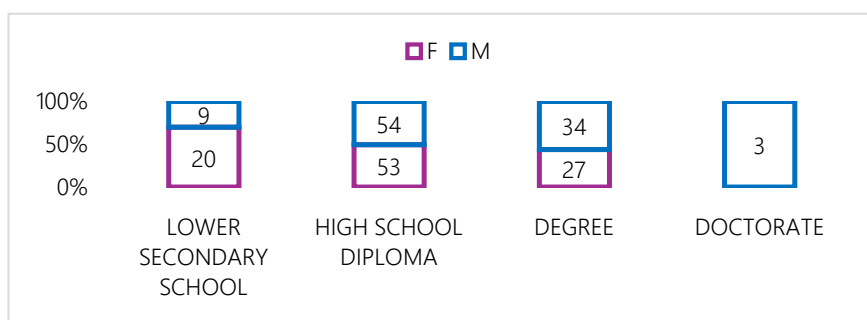


Figure 4. Percentages of educational qualifications by gender.

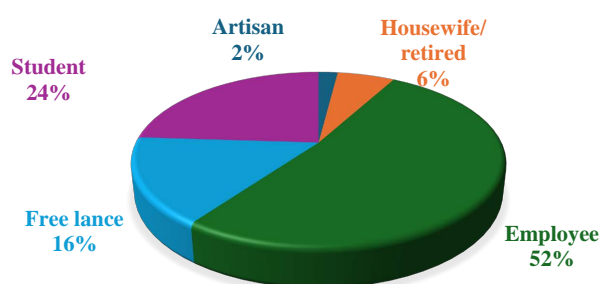


Figure 5. Percentages of employment status.

Analyzing a double-entry table between employment status and educational qualification, it emerges that half of the employees are graduates and half are divided between graduates and those with a lower secondary school diploma; 83% of the students have a high school diploma. 56% of graduates are employees and 38% are freelancers (Table 1).

Table 1. Profession and qualification.

| Qualification | Profession | | | | | Total |
|------------------------|------------|-------------------|----------|------------|---------|-------|
| | Artisan | Housewife/Retired | Employee | Free Lance | Student | |
| Lower secondary school | 0 | 1 | 4 | 0 | 1 | 6 |
| High school Diploma | 1 | 2 | 13 | 1 | 10 | 27 |
| Degree | 0 | 0 | 9 | 6 | 1 | 16 |
| Doctorate | 0 | 0 | 0 | 1 | 0 | 1 |
| Total | 1 | 3 | 26 | 8 | 12 | 50 |

2.3. Clinical Information

The questionnaire containing general information also included two questions to investigate patients' familiarity with hospitals and data management, and two

other questions to find out if the patient is a regular donor and the number of donations made.

2.3.1. Familiarity with Hospital and Data Management

44% of respondents say they are familiar with data management, of which 27% are women and 51% are men (Figure 6). 52% of those interviewed declare that they are familiar with hospitals, of which 33% are women and 60% are men (Figure 7).

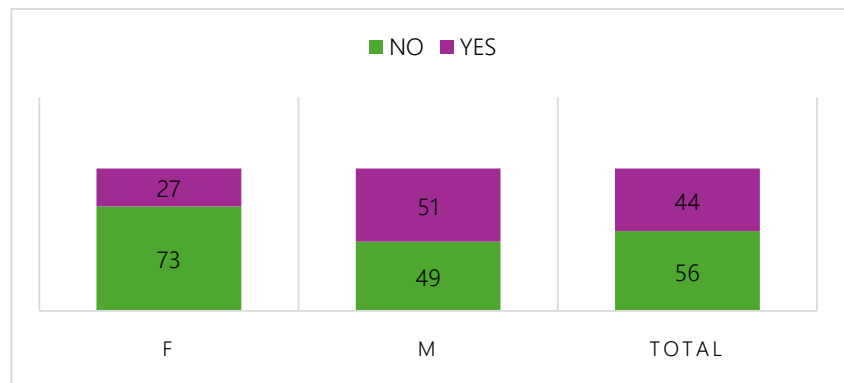


Figure 6. Percentages familiar with data management by gender.

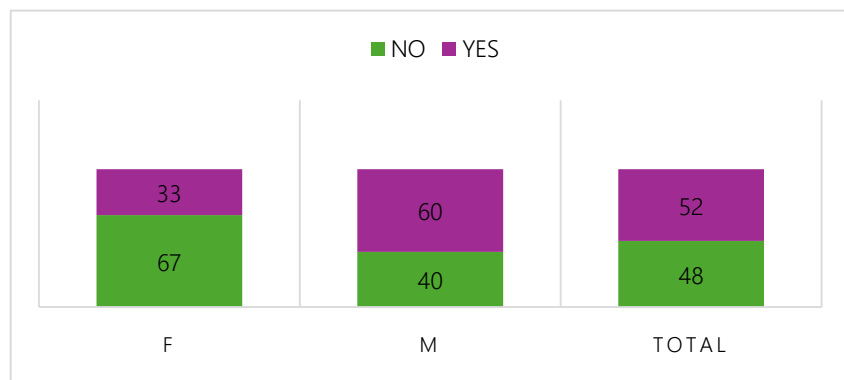


Figure 7. Percentages familiar with hospitals by sex.

2.3.2. Donation

72% of those interviewed are donors and the majority are men; in fact, among the male interviewees, 83% are donors, among the female interviewees only 27% are donors (Figure 8).

There are 13 interviewees who made a number of donations between 1 and 4, 10 between 5 and 8, the others are equally distributed among the other classes which can be seen Figure 9.

3. Information on the Ability to Express Consent to the Processing of Health Data

In the semi-structured interview conducted, the patient's skills were assessed in relation to the four domains of the ability to give consent to the processing of

personal health data (Personal Health Information, PHI). The four domains are described below:

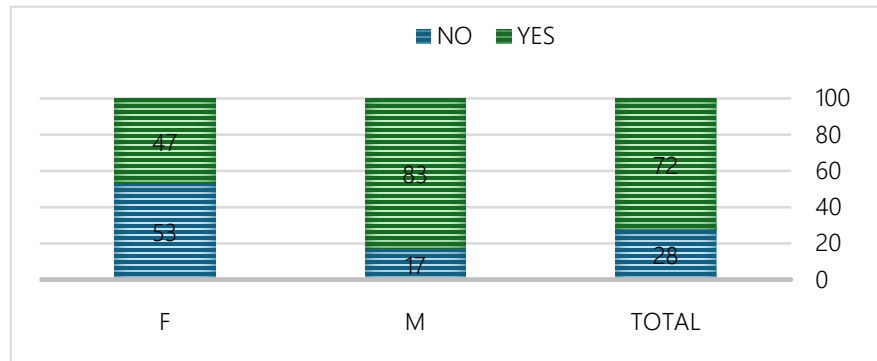


Figure 8. Donor percentages by sex.

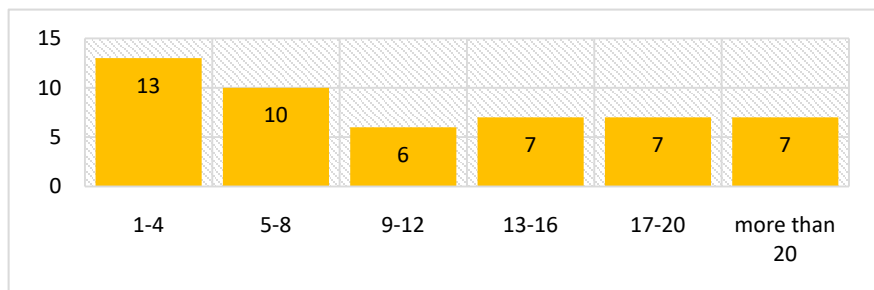


Figure 9. Number of donations.

3.1. Comprehension

This section investigated the understanding of the definition of personal health data, the purposes and limits of the proposed data processing, and the benefits/risks deriving from data processing.

A score between 0 and 2 was assigned for each item in the two sub-sections of Understanding (General and Purpose/Risks) and a global score between 0 and 4, the sum of the scores of the two sub-sections.

A score of 2 (*i.e.* The patient remembers the content of the item and offers a sufficiently clear version of it. A literal repetition of the description of the experimenter in fact is not required, paraphrasing in the patient's own words is preferred. For patients items regarding risks, the patient must provide a reasonably accurate indication of the probability that these will if it has been described in the communication) was given to 42% of respondents with regard to general understanding, to 84% for understanding purpose/risks (**Figure 10**).

In fact, in each of the three items of the two sub-sections, general understanding and understanding of the purposes and risks, score 2 finds higher percentages, see **Table 2** and **Table 3**.

The score awarded for total understanding is the sum of the scores obtained from general understanding and purposes/risks, 40% of respondents receive a maximum score of 4 (**Figure 11**).

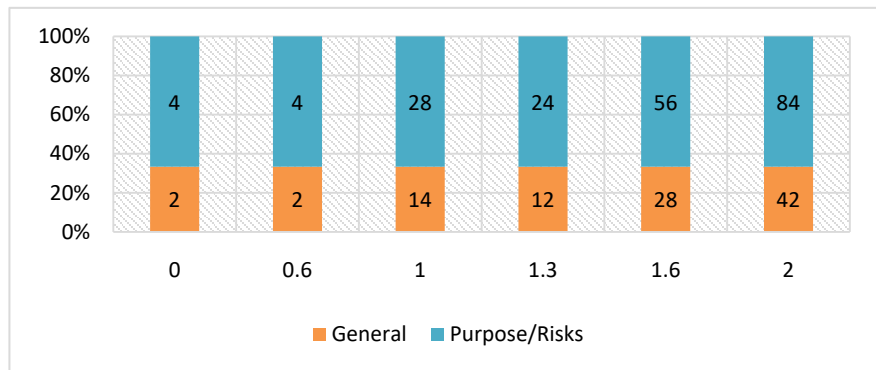


Figure 10. Percentages general understanding and purpose/risks.

Table 2. General understanding: percentages of scores broken down by item general understanding.

| Score | Item | | |
|-------|---------------------------|-----------------|--------------------------------|
| | Healthcare personnel data | Data processing | Electronic Health Record (EHR) |
| 0 | 4 | 4 | 2 |
| 1 | 28 | 40 | 26 |
| 2 | 68 | 56 | 72 |
| Total | 100 | 100 | 100 |

Table 3. Understanding purposes and risks: Percentages of scores divided by item understanding purpose and risks.

| Score | Item | | |
|-------|------------------|----------------------------------|--|
| | Purpose (ESF/DS) | Purpose (statistical/scientific) | Risk (data leak and advertising/profiling) |
| 0 | 8 | 6 | 4 |
| 1 | 14 | 28 | 14 |
| 2 | 78 | 66 | 82 |
| Total | 100 | 100 | 100 |

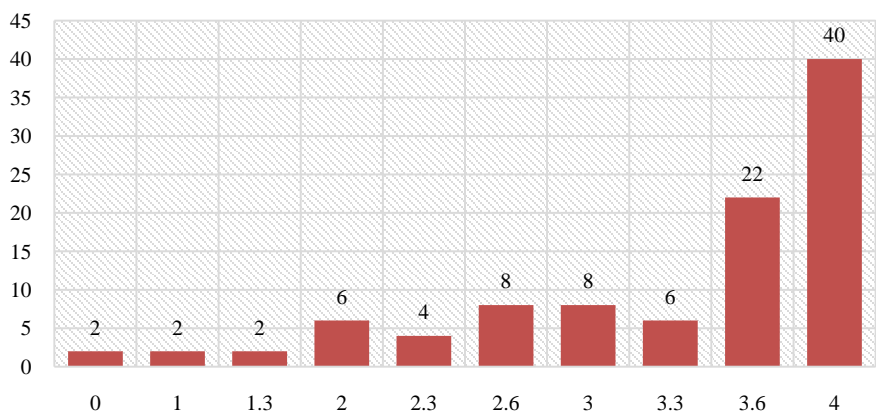


Figure 11. Total comprehension percentages.

3.2. Evaluation

The purpose of this section is to establish whether the patient recognizes the risks and benefits deriving from the processing of health data for the proposed purposes, *i.e.* the score is attributed to the evaluation of the meaning of the information for the patient's specific situation.

A score of 1 or 0 was assigned to each sub-section of the evaluation: that is, depending on whether the patient evaluates the electronic health record as useful or not for the purposes related to "personal health" and for the purposes of "statistics and research". 62% of patients agree with both purposes, while 30% of patients believe that the electronic health record is useful only for their health (Table 4).

Table 4. Evaluation: percentages of scores divided by sub-section.

| Purpose: personal health | Purpose: Statistics and Research | | Total |
|--------------------------|----------------------------------|----|-------|
| | 0 | 1 | |
| 0 | 2 | 6 | 8 |
| 1 | 30 | 62 | 92 |
| Total | 32 | 68 | 100 |

The total evaluation was given a score between 0 and 2, the sum of the scores of the two sub-sections and 62% were given a score of 2 (The patient recognizes the benefits deriving from the processing of data for the purposes communicated, of of which: one relating to daily life, the other concerning the community. Or the patient believes that the processing of his/her health data cannot bring benefits for himself or for the community but offers non-delusional reasons that have a rational basis) (Figure 12).

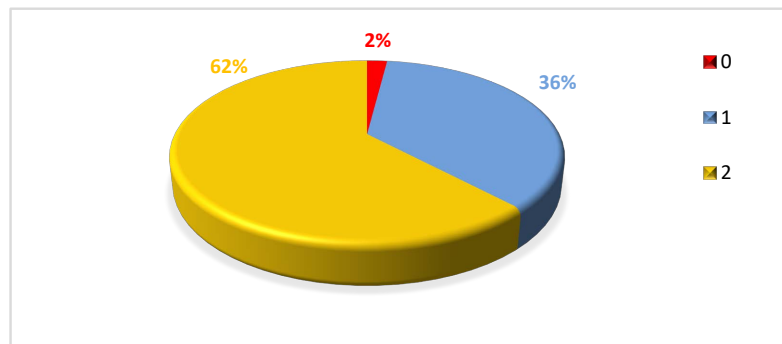


Figure 12. Rating percentages.

3.3. Reasoning

The reasoning sub-section intends to investigate the cognitive processes that underlie the patient's expressed choice. Therefore, we mean reasoning in the decision-making process, focusing attention on the ability to compare the alternatives (allowing or not allowing data processing), knowing how to motivate the

choice in light of the consequences, including the ability to deduce the effect of one’s choice on one’s personal situation, but also on public health and scientific progress.

A score between 0 and 2 was assigned to the three sub-sections of the reasoning (qualitative, quantitative and logical coherence) and an overall score between 0 and 6 was the sum of the scores of the three sub-sections.

In particular, the two items in the qualitative reasoning and quantitative reasoning sub-sections were given a score of 1 or 0, therefore both of these sub-sections report a sum score of the two items. Qualitative reasoning means investigating the cognitive processes that underlie the choice expressed by the patient and therefore the motivations to give consent or not to the processing of health data to feed the electronic health record and for research purposes. The investigator judges the reasons that lead the patient to consent to the processing of health data for research purposes to be consistent in 74% of cases and to feed the electronic health record in 90% of cases (Table 5).

Table 5. Qualitative reasoning: percentages of scores broken down by item.

| Motivations: research | Motivation: Electronic health record | | Total |
|-----------------------|--------------------------------------|----|-------|
| | 0 | 1 | |
| 0 | 2 | 6 | 8 |
| 1 | 30 | 62 | 92 |
| Total | 32 | 68 | 100 |

With quantitative reasoning we intend to establish whether the patient can describe the possible consequences of his choice taking into account the purposes of the proposed health data processing, the possible benefits and risks connected both for himself and for the community. In 72% of cases the experimenter judges the consequences that the patient believes could have a personal nature when giving consent to the processing of health data are coherent, while patients who believe that the consequences could have an impact on the community are divided in half (Table 6, Figure 13).

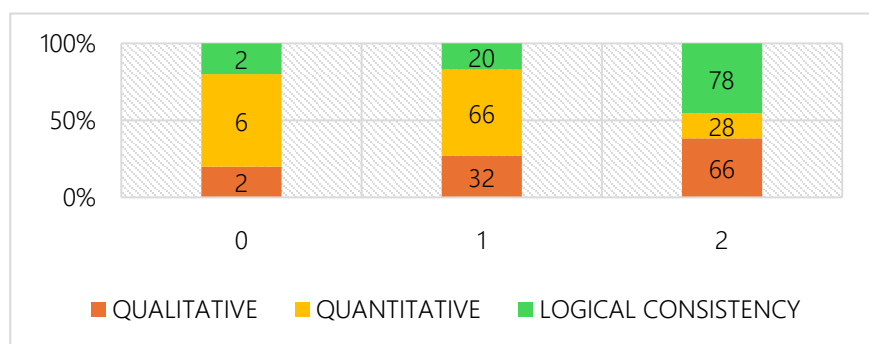
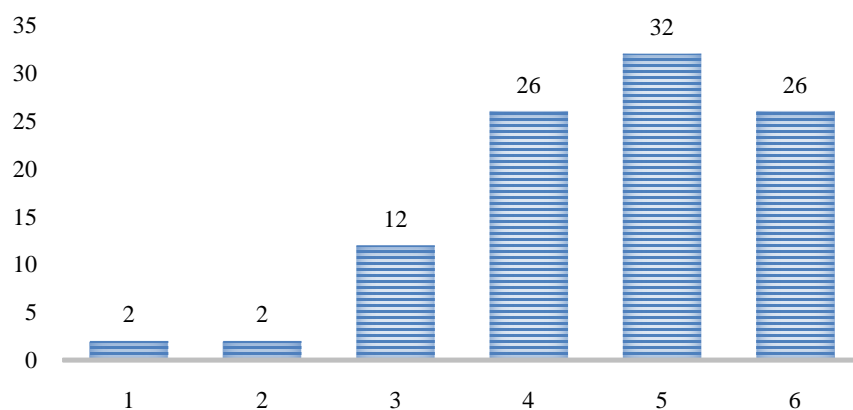


Figure 13. Percentages qualitative reasoning, quantitative reasoning and logical consistency.

Table 6. Quantitative reasoning: percentages of scores broken down by item.

| Consequences: Personal | Consequences: Community | | Total |
|---------------------------|-------------------------|----|-------|
| | 0 | 1 | |
| 0 | 6 | 22 | 28 |
| 1 | 44 | 28 | 72 |
| Total | 50 | 50 | 100 |

Comparing the sum scores of each sub-section of the reasoning, as reported in **Figure 13**, a score of 2 was assigned (The patient cites at least two reasons in explaining the choice, The consequences must be more specific than “I will need it” or “is better”) to 66% of the interviewees with regards to quantitative reasoning, for qualitative reasoning the highest percentage (66%) was given a score of 1 (The patient provides only one rational consequence relating to everyday life or the effects on the community, but not for the other), while 78% of interviewees were given a score of 2 (The patient’s final choice (in Expression of a choice) derives logically from the patient’s reasoning) for logical coherence (**Figure 14**).

**Figure 14.** Total percentage reasoning.

Finally, the score attributed for total reasoning is the sum of the scores obtained from qualitative, quantitative reasoning and logical coherence, 32% of respondents were awarded a score of 5 and 26% of respondents received a maximum score of 6 (**Figure 14**).

3.4. Expression of a Choice

This section aims to record the patient’s final choice to give consent to the processing of health data for the proposed purposes and verify that this is clear and free of ambiguity. A score of 2 or 1 was given depending on whether the patient expressed consent fully consciously or not. 90% of those interviewed expressed their consent clearly and without ambiguity (**Figure 15**).

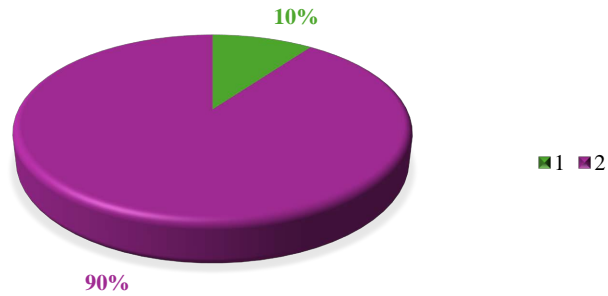


Figure 15. Percentages expressing a choice.

3.5. Demographic Profiles

The purpose of the survey was to evaluate patients’ skills in relation to the four domains of the ability to give consent to the processing of personal health data. The global scores were then calculated by summing the total scores reported by the patients in each of the four domains investigated: total understanding, total reasoning, total evaluation and expression of a choice.

The range of global scores varies from 5 to 14, the mean is 11.4, the mode is 14 and corresponds to the highest global score (Table 7). As shown in Figure 16, the largest percentage of patients, 60%, report a score between 11 and 14, 34% between 8 and 11, only 6% a score between 5 and 8.

Table 7. Global score.

| Global score | N. of patients | Global score | N. of patients |
|--------------|----------------|--------------|----------------|
| 5 | 1 | 10.6 | 2 |
| 6 | 1 | 11 | 3 |
| 8 | 1 | 11.3 | 2 |
| 8.3 | 1 | 11.6 | 1 |
| 8.6 | 2 | 12 | 6 |
| 9 | 4 | 12.3 | 1 |
| 9.3 | 1 | 12.6 | 4 |
| 9.6 | 2 | 13 | 4 |
| 10 | 1 | 13.6 | 4 |
| 10.3 | 1 | 14 | 8 |
| Total | | | 50 |

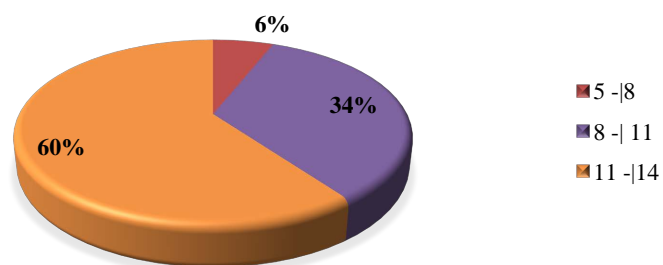


Figure 16. Overall class score percentages.

The scores of the patients interviewed do not directly define the state of legal capacity or inability to provide consent to the processing of health data. A criterion was therefore defined by calculating the average of the global score (**Table 8**).

Table 8. Percentage of patients with global scores below or a above average.

| Criteria | Mean |
|----------|------|
| 0 | 44 |
| 1 | 56 |
| Total | 100 |

In **Table 8**, 0 indicates a lower than average score, 1 indicates a higher global score: 56% of patients receive a higher than average global score.

Patients with “average” or above average global scores in all sections probably have sufficient decision-making capacity to express valid consent to the processing of health data. Conversely, although very low global scores suggest the inability to make decisions regarding the processing of health data, they cannot independently constitute the sufficient basis for expressing a final judgment of inability.

4. Cluster Analysis

4.1. The Twostep Cluster Analysis

Cluster analysis aims to define specific profiles and is very advantageous as it provides clusters that are “relatively distinct” from each other (*i.e.* heterogeneous), each made up of units with a high degree of “natural association”. The different approaches to cluster analysis have in common the need to define a dissimilarity or distance matrix between the n pairs of observations, which represents the point from which each algorithm is generated.

The cluster analysis technique chosen is the one defined as TwoStep, an extension of the distance measures used by Banfield and Raftery [81]; it is a scalar cluster analysis algorithm capable of simultaneously treating continuous and categorical variables or attributes. It has two advantages: it deals with mixed variables and automatically determines the optimal number of clusters, although it allows the researcher to fix the desired number of clusters. It is achieved through two steps: In the first step, defined as pre-cluster, the records are pre-classified into many small sub-clusters; in the second step the sub-clusters (generated in the first step) are grouped into a number of clusters that optimizes the BIC (Bayesian Information Criterion) defined as:

$$\text{BIC}_K = -2l_k + r_k \log n \quad (1)$$

where l_k is the number of independent parameters and:

$$l_k = \sum_{v=1}^k \zeta_v \quad (2)$$

is the log-likelihood function, for the step with k clusters, which can be interpreted as the dispersion within the clusters. Furthermore, it represents the entropy within the k clusters if only categorical variables are considered.

4.2. Identification of Profiles

The cluster analysis made it possible to outline some profiles relating to the interviewees present in the data archive. Two simulations were carried out which led to the identification of clusters with specific profiles depending on the “demographic characteristics” of the respondent, the “familiarity with the hospital environment” and above all the “global score” obtained from completing the questionnaire (obtained by adding the total scores reported in the domains total understanding, total reasoning, total evaluation and expression of a choice).

Patients with “average” or above average PHICAT global scores in all sections probably have sufficient decision-making capacity to express valid consent to the processing of health data.

The first application of the cluster considers demographic aspects and global score; 4 clusters emerged (**Figure 17**), with a higher percentage composition of respondents for cluster 1 (32%) and lower for cluster 3 (16%).

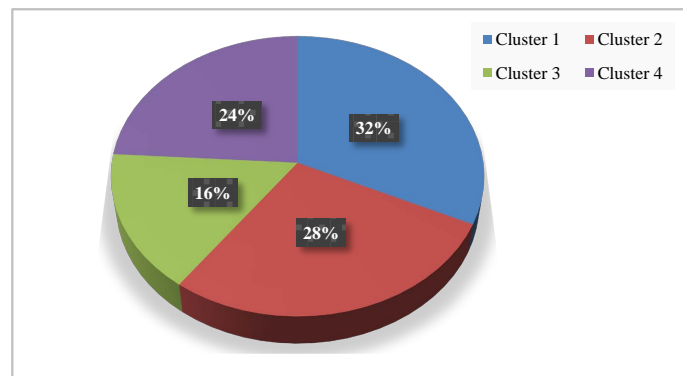


Figure 17. Percentage distribution of interviewees in the different clusters.

The most important variable in defining the profiles is therefore the global score, on the basis of which we can classify our clusters by analyzing the average value of each cluster as indicated in **Table 9**.

Table 9. Mean value and standard deviation of global scores for clusters.

| Cluster | Mean | Standard deviation |
|-----------|-------|--------------------|
| Cluster 1 | 12.33 | 1.84 |
| Cluster 2 | 11.93 | 1.34 |
| Cluster 3 | 11.80 | 1.91 |
| Cluster 4 | 9.34 | 2.50 |

Each cluster will therefore be characterized by the following profile (**Figures 18-21**):

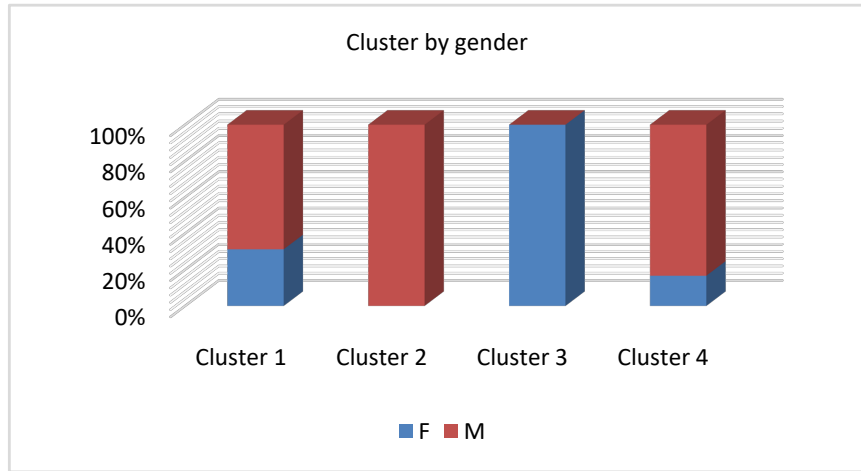


Figure 18. Percentage distribution of interviewees in the different clusters by demographic characteristics—Cluster by gender.



Figure 19. Percentage distribution of interviewees in the different clusters by demographic characteristics—Cluster by age classes.

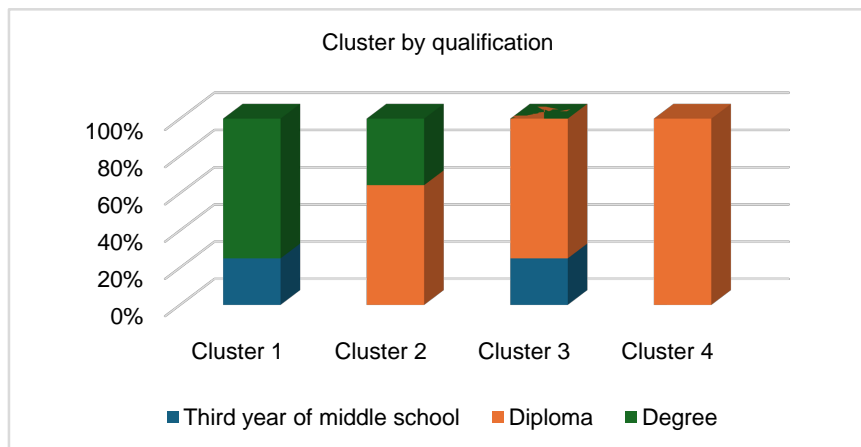


Figure 20. Percentage distribution of interviewees in the different clusters by demographic characteristics—Cluster by qualifications.

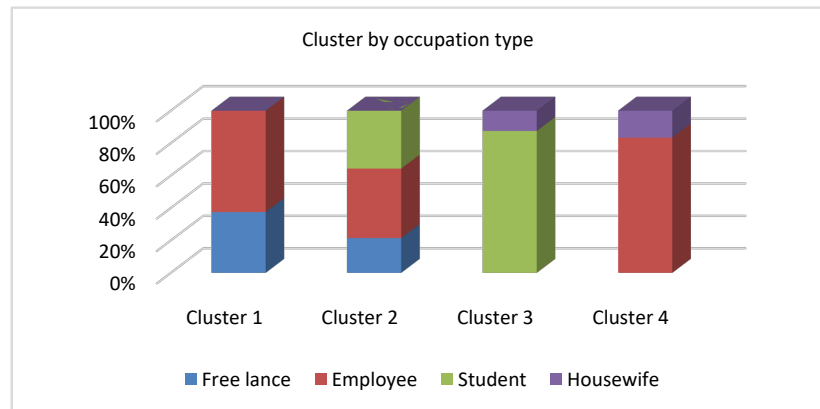


Figure 21. Percentage distribution of interviewees in the different clusters by demographic characteristics—Cluster by occupation type.

- **Cluster 1** “*subject with good decision-making skills such as to be able to express valid consent to the processing of health data*”: these are those who have an average global score of 12.33. As regards demographic characteristics, they are characterized by 31% females and 69% males. These are people over 30 years old, employed (63% independent and 37% employed), whose educational qualification is 75% a university degree and 25% a middle school diploma.

- **Cluster 2** “*subject with sufficient decision-making capabilities to be able to express valid consent to the processing of health data*”: these are those who have an average global score of 11.93. As far as demographic characteristics are concerned, they are all male, aged between 18 and 29. These are employed young people (21% independent and 43% employed) or students (36%), whose educational qualification is 64% a high school diploma and 36% a degree.

- **Cluster 3** “*subject with almost sufficient decision-making capacity to be able to express valid consent to the processing of health data*”: these are those who have an average global score of 11.80. These are students (88% of the total) and housewives (12%) aged between 18 and 29, 75% of whose educational qualifications are high school diplomas and 25% have a middle school diploma.

- **Cluster 4** “*subject with low decision-making abilities such as to be able to express valid consent to the processing of health data*”: these are those who have an average global score of 9.34. As regards demographic characteristics, 83% of them are males over the age of 30, whose educational qualification is a high school diploma and who are almost all employed, with an employee employment contract. The remaining 17% of females are characterized by housewives.

The second application of the cluster considers only some demographic aspects (age and educational qualification), familiarity with data management, being a regular donor or not and the global score; 4 clusters always emerged (Figure 22), with a higher percentage composition of respondents for cluster 1 (38%) and lower for cluster4 (18%).

The most important variable in defining the profiles is therefore the global score, on the basis of which we can classify our clusters by analyzing the average value of each cluster as indicated in Table 10.

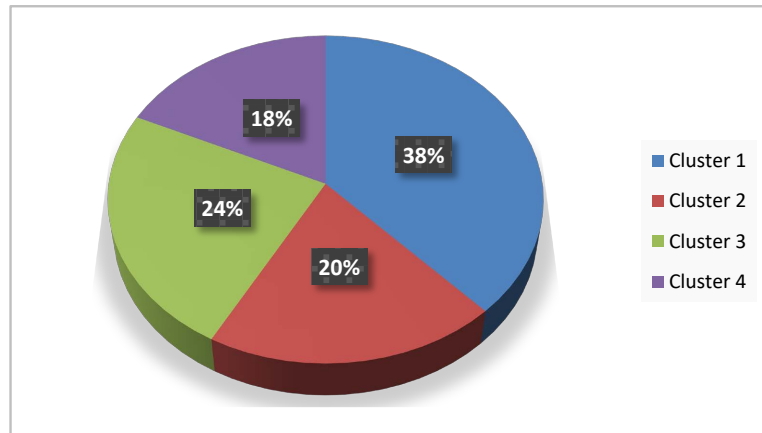


Figure 22. Percentage distribution of interviewees in the different clusters.

Table 10. Mean value and standard deviation of global scores for clusters.

| Cluster | Mean | Standard Deviation |
|-----------|-------|--------------------|
| Cluster 1 | 11.99 | 1.40 |
| Cluster 2 | 10.89 | 2.93 |
| Cluster 3 | 9.64 | 1.81 |
| Cluster 4 | 13.16 | 1.41 |

Each cluster will therefore be characterized by the following profile (Figures 23-26):

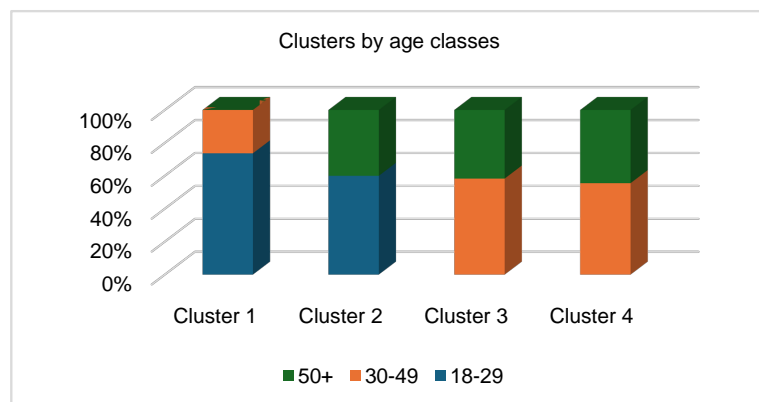


Figure 23. Percentage distribution of interviewees in the different clusters by demographic characteristics and familiarity with the hospital environment – Cluster by age classes.

- **Cluster 1** “subject with sufficient decision-making skills to be able to express valid consent to the processing of health data”: these are those who have an average global score of 11.99. As regards demographic characteristics, they are mostly characterized by young people under 29 years of age (74%) whose educational qualifications are 50% high school diploma and 30% degree. All those who are not familiar with the management of health data but are 50% regular donors are part of this cluster.

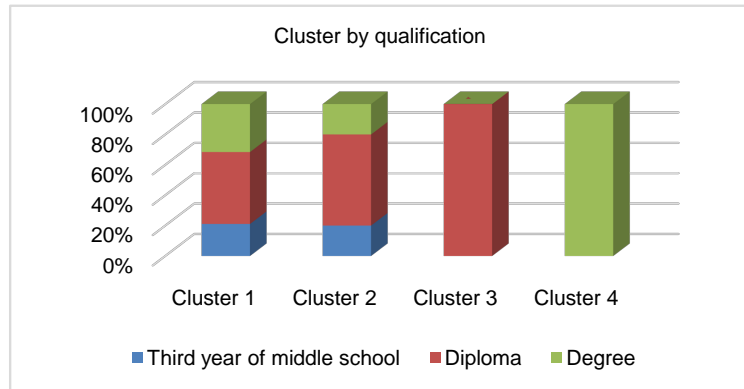


Figure 24. Percentage distribution of interviewees in the different clusters by demographic characteristics and familiarity with the hospital environment – Cluster by qualification.

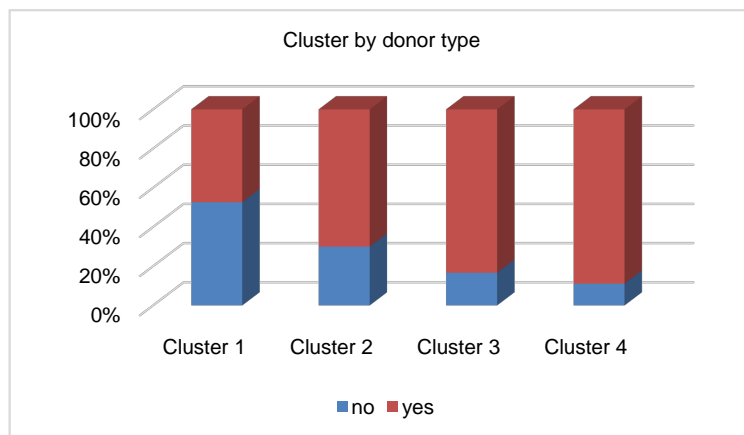


Figure 25. Percentage distribution of interviewees in the different clusters by demographic characteristics and familiarity with the hospital environment – Cluster by donor type.

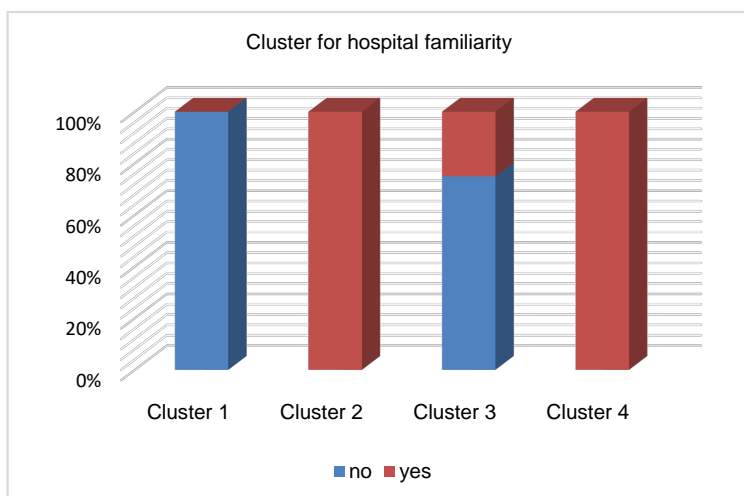


Figure 26. Percentage distribution of interviewees in the different clusters by demographic characteristics and familiarity with the hospital environment – Cluster for hospital familiarity.

- **Cluster 2** “*subject with almost sufficient decision-making capacity to be able to express valid consent to the processing of health data*”: these are those who have an average global score of 10.89. As far as demographic characteristics are concerned, 60% are young people aged between 18 and 29, whose educational qualification is a high school diploma. The remaining 40% are adults over 50 years old. This cluster includes all those who are familiar with the management of health data (100%) and 60% are regular donors.

- **Cluster 3** “*subject with low decision-making abilities such as to be able to express valid consent to the processing of health data*”: these are those who have an average global score of 9.64. These are people over the age of 30, whose educational qualifications are 100% high school diplomas. 75% of those interviewed are not familiar with health data management but 80% are regular donors.

- **Cluster 4** “*subject with good decision-making skills such as to be able to express valid consent to the processing of health data*”: these are those who have an average global score of 13.16. As regards demographic characteristics, the interviewees are over the age of 30, all of whom have a degree. It is interesting to note that all members of this cluster are familiar with health data management and are regular donors (100%).

5. Analysis of the Relationships between the Variables

5.1. Correlation Analysis

The variables detected through the questionnaire, the subject of this analysis, can be classified into 3 groups: socio-demographic variables (age, gender, educational qualification, profession), clinical variables (familiarity with data management, familiarity with the hospital, being a regular donor and number of donations) and variables relating to the four domains that allow evaluating the patient's ability to express consent for the processing of personal health data (understanding, evaluation, reasoning and expressing a choice).

The statistical method used for the analysis of the relationships between variables were correlation and logistic regression. The analyses were carried out using IBM SPSS software (IBM, 2017).

The first phase of the study concerned the relationships between the answers provided by each of the 50 interviewees, with reference to the entire set of questionnaire variables, measured by the Spearman correlation index ρ , particularly suitable in the case of categorical variables. The results obtained are illustrated in the correlation matrices (Attachment 2 e 3), which also report the relative levels of significance.

Regarding the relationships between the variables of the 4 domains and the socio-demographic and clinical ones (Attachment 2), the most important correlations were found between the educational qualification and total understanding ($\rho = 0.315$; p-value = 0.026) and between familiarity in data management and total evaluation ($\rho = 0.286$; p-value = 0.044).

From the analysis of the correlations between the variables of the 4 domains, it

can be seen that the expression of a choice is correlated with total understanding ($\rho = 0.440$; p-value = 0.001) and total reasoning ($\rho = 0.426$; p-value = 0.002).

In turn, the total reasoning score is significantly correlated with that of total understanding ($\rho = 0.404$; p-value = 0.004) and total evaluation ($\rho = 0.496$; p-value ≤ 0.001).

By limiting the analysis to the relationships between the socio-demographic and clinical variables, and neglecting those relating to the 4 domains, significant correlations are highlighted between age and being a habitual donor ($\rho = 0.407$; p-value = 0.003) and between age and the number of donations made ($\rho = 0.573$; p-value ≤ 0.001). The variable familiarity with the hospital is correlated with the educational qualification ($\rho = 0.341$; p-value = 0.015), while the number of donations is positively correlated with being a regular donor ($\rho = 0.692$; p-value ≤ 0.001).

5.2. Logistic Regression Models

To explore the dependency relationships between the variables under study, mainly qualitative nominal and ordinal, the logistic model was used [82] [83].

Logistic regression analysis estimates the regression function that expresses the best relationship between a set of explanatory variables (regressors) and the probability of the occurrence of a modality of a qualitative dichotomous character taken as the dependent variable [84].

In this model, the dependent variable can already be detected as dichotomous or dichotomized for the purposes of the analysis.

Logistic regression, unlike linear regression, considers the distribution of the dependent variable as binomial. The estimate of Y varies between 0 and 1 and takes on the meaning of probability that Y is equal to 1: $P(Y = 1|x) = \pi(x)$.

The logistic regression function is as follows:

$$\log it(\pi(x)) = \beta_0 + \sum_i^q \beta_i x_i = X\beta \quad (3),$$

in which

- $\log it(\pi(x))$ is the natural logarithm of the probability of Y given the vector

x of q explanatory variables: $\log it(\pi(x)) = \ln \left[\frac{\pi x}{1 - \pi x} \right]$,

- $\pi(x)$ is the probability that Y takes on the value 1 as a function of the explanatory variables x .

The probability of Y can also be written as a logistic function:

$$\pi(x) = \frac{e^{x\beta}}{1 + e^{x\beta}} \quad (4).$$

In the various models implemented, the response variable and the regressors were identified based on the results obtained through the correlation analysis.

For the selection of regressors, a significance level of p ≤ 0.05 was assumed. The goodness of fit of the model is evaluated with the Nagelkerke R^2 index.

Numerous logit models were implemented; however, it was deemed useful to

report only those that highlighted significant relationships between the variables.

5.2.1. Model 1—Target Variable Expression of a Choice

The target variable Expression of a choice (Table 11) was found to depend on the Total Reasoning, obtained as the sum of the scores of the responses of Qualitative Reasoning, Quantitative Reasoning and Logical Coherence (p-value = 0.029).

Table 11. Estimates of the Beta coefficients of the domains on the Total Expression of a choice.

| Regressors | B | s.e. | Wald | d.f. | p-value | Exp(B) |
|---------------------|--------|-------|-------|------|---------|--------|
| Total understanding | 0.876 | 0.575 | 2.321 | 1 | 0.128 | 2.402 |
| Total reasoning | 1.675 | 0.766 | 4.787 | 1 | 0.029 | 5.340 |
| Constant | -6.741 | 3.328 | 4.103 | 1 | 0.043 | 0.001 |

R² Nagelkerke: 0.552.

Subsequently, it was deemed appropriate to analyze the response variable Expression of a choice separately in relation to the different types of qualitative reasoning (motivations relating to the health record and motivations relating to research) and quantitative reasoning (personal consequences and consequences on the community). The results obtained (Table 12) show how Personal Consequences are the only significant regressor capable of influencing the dependent variable considered (p-value = 0.027).

Table 12. Estimates of the Beta coefficients of the domain components on the Total Expression of a choice.

| Regressors | B | s.e. | Wald | d.f. | p-value | Exp(B) |
|--------------------------------|---------|-----------|-------|------|---------|--------------|
| Personal consequences | 3.494 | 1.584 | 4.869 | 1 | 0.027 | 32.923 |
| Consequences for the community | 2.300 | 1.399 | 2.704 | 1 | 0.100 | 9.974 |
| FSE qualitative reasoning | -18.055 | 17412.846 | 0.000 | 1 | 0.999 | 0.000 |
| Qualitative reasoning research | 0.511 | 1.184 | 0.186 | 1 | 0.666 | 1.667 |
| Constant | 17.086 | 17412.846 | 0.000 | 1 | 0.999 | 26336399.218 |

R² Nagelkerke: 0.399.

If, in the aforementioned model, the individual components of the Understanding domain (General Understanding and Understanding of Purpose and Risks) and Logical Coherence are included in the set of explanatory variables, the model loses its significance.

It is therefore possible to state that reasoning plays an important role in explaining patients' decision to give consent and, in particular, the ability to enumerate what consequences may arise in one's daily life (personal health, work, family life, etc.).

5.2.2. Model 2—Target Variable Qualitative Reasoning

In this model, the patient’s qualitative reasoning is considered as a response variable in adequately motivating their choice to give consent to the processing of health data both to feed the electronic health record (EHR) and for purposes relating to scientific research.

If Qualitative Reasoning for FSE purposes is set as the response variable, no significant relationship emerges with any of the selected predictors.

On the contrary, if Qualitative Reasoning for research purposes is set as the response variable, the Total Understanding variable is an important regressor (p-value = 0.033). The Total Evaluation variable (Table 13), however, has a modest contribution based on the pre-established threshold (p-value = 0.071).

Table 13. Estimates of the Beta coefficients of the domains on the Total Qualitative reasoning for research purposes.

| Regressors | B | s.e. | Wald | d.f. | p-value | Exp(B) |
|---------------------|--------|-------|-------|------|---------|--------|
| Total understanding | 0.856 | 0.403 | 4.522 | 1 | 0.033 | 2.354 |
| Total rating | 1.194 | 0.662 | 3.252 | 1 | 0.071 | 3.299 |
| Constant | -3.467 | 1.588 | 4.765 | 1 | 0.029 | 0.031 |

R² Nagelkerke: 0.272.

However, by breaking down the Total Understanding and Total Evaluation variables into their respective dichotomized components (Table 14), it is clear that, among the components of the Evaluation for purposes relating to personal health and for statistical and research purposes, only the latter is significantly influencing (p-value = 0.044) the patient’s ability to provide adequate reasons for the expressed choice.

Table 14. Estimates of the Beta coefficients of the components of the domains on Qualitative Reasoning.

| Regressors | B | s.e. | Wald | d.f. | p-value | Exp(B) |
|---|---------|-----------|-------|------|---------|----------------|
| General understanding of personal data | 1.418 | 1.149 | 1.525 | 1 | 0.217 | 4.130 |
| General understanding of data processing | -4.469 | 2.326 | 3.692 | 1 | 0.055 | 0.011 |
| General understanding FSE | -18.949 | 16344.359 | 0.000 | 1 | 0.999 | 0.000 |
| Understanding ESF objectives | 0.819 | 2.102 | 0.152 | 1 | 0.697 | 2.268 |
| Understanding statistical/scientific purposes | 2.738 | 1.448 | 3.574 | 1 | 0.059 | 15.454 |
| Understanding risk | 22.254 | 16344.359 | 0.000 | 1 | 0.999 | 4622867796.567 |
| ESF evaluation | 2.486 | 1.971 | 1.591 | 1 | 0.207 | 12.015 |
| Research evaluation | 2.210 | 1.099 | 4.045 | 1 | 0.044 | 9.116 |
| Constant | -5.456 | 2.389 | 5.214 | 1 | 0.022 | 0.004 |

R² Nagelkerke: 0.660.

Among the 6 dichotomized components of the Understanding domain, those relating to Understanding of data processing (p-value = 0.055) and Understanding of the use of data for scientific and research activities show a modest influence, with values close to the significance threshold. (Table 14).

5.2.3. Model 3—Target Variable Quantitative Reasoning

The following model takes as response variables first the patient's ability to enumerate the consequences on their daily life (personal health, work, family life) and then the consequences on the community (scientific, educational and academic research) deriving from the granting of consent to treatment of their health data.

Considering, in the first analysis, the effects of the explanatory variables Total Understanding and Total Evaluation on the Consequences in daily life (Table 15), only the estimate of the coefficient of the Total Evaluation is significant (p-value = 0.030), while the influence of the other regressor is negligible.

Table 15. Estimates of the Beta coefficients of the domains on Quantitative Reasoning-Consequences in daily life.

| Regressors | B | s.e. | Wald | d.f. | p-value | Exp(B) |
|---------------------|--------|-------|-------|------|---------|--------|
| Total Understanding | 0.565 | 0.378 | 2.230 | 1 | 0.135 | 1.759 |
| Total Rating | 1.409 | 0.651 | 4.685 | 1 | 0.030 | 4.091 |
| Constant | -2.999 | 1.503 | 3.981 | 1 | 0.046 | 0.050 |

R² Nagelkerke: 0.238.

By breaking down the domains of Understanding and Evaluation into the various dichotomized components (Table 16), we find an appreciable relevance of the Understanding variable for the purpose of using data for scientific research purposes (p-value = 0.013) and of Evaluation for statistical and research purposes (p-value = 0.003).

Table 16. Estimates of the Beta coefficients of the components of the domains on Quantitative Reasoning-Consequences in daily life.

| Regressors | B | s.e. | Wald | d.f. | p-value | Exp(B) |
|---|---------|-----------|-------|------|---------|-----------------|
| General understanding given | -1.067 | 1.313 | 0.661 | 1 | 0.416 | 0.344 |
| General understanding of data processing | -2.816 | 1.752 | 2.585 | 1 | 0.108 | 0.060 |
| General understanding FSE | 0.017 | 2.163 | 0.000 | 1 | 0.994 | 1.017 |
| Understanding ESF objectives | -23.260 | 16047.413 | 0.000 | 1 | 0.999 | 0.000 |
| Understanding statistical/research purposes | 4.059 | 1.638 | 6.140 | 1 | 0.013 | 57.945 |
| Understanding the purpose of risks | 24.778 | 16047.413 | 0.000 | 1 | 0.999 | 57647430079.378 |
| ESF evaluation | 1.915 | 1.771 | 1.169 | 1 | 0.280 | 6.785 |
| Research evaluation | 3.472 | 1.176 | 8.718 | 1 | 0.003 | 32.188 |
| Constant | -4.327 | 2.110 | 4.205 | 1 | 0.040 | 0.013 |

R² Nagelkerke: 0.646.

Moving on to consider the Quantitative Reasoning relating to the Consequences on the community as the response variable, the model does not highlight the presence of any predictor with an acceptable level of significance, both with respect to the variables of the 4 domains and with respect to the sociodemographic and clinical variables.

5.2.4. Model 4—Logical Coherence

Considering the dichotomized Logical Coherence as the response variable and those of the domains of Understanding, Qualitative and Quantitative Reasoning, Evaluation and Expression of a choice as explanatory variables, none of the regressors appears to exert any influence on the dependent variable. However, if we observe the matrix of correlations between pairs of variables, Logical Coherence is significantly correlated with Personal Consequences, Qualitative Reasoning with research purposes and the Expression of a final choice; the same variable is, however, inversely correlated with Age.

5.2.5. Model 5—Evaluation of Personal Health and Research Purposes

The latest models developed take into consideration the possible relationship between the dependent variable Evaluation, in its components for personal health purposes or for research purposes, and the predictors relating to all the components of the other domains of Understanding, Reasoning and Expression of a final choice.

The results of the model implemented to estimate the relationship between the variable Evaluation for personal health purposes and the various predictors used so far are not significant.

On the contrary, the model that takes into consideration the variable Evaluation for research purposes shows the importance of Total Reasoning (p-value = 0.014) as an explanatory variable compared to those of Total Understanding and Total Expression of final choice. (Table 17)

Table 17. Estimates of the Beta coefficients of Understanding, Reasoning and Expression of the final choice on the Evaluation for research purposes.

| Regressor | B | s.e. | Wald | d.f. | p-value | Exp(B) |
|--------------------------------|--------|-------|-------|------|---------|--------|
| Total Understanding | 0.237 | 0.377 | 0.394 | 1 | 0.530 | 1.267 |
| Total Reasoning | 0.876 | 0.358 | 5.996 | 1 | 0.014 | 2.402 |
| Totale Final chosen expression | -1.600 | 1.320 | 1.469 | 1 | 0.225 | 0.202 |
| Constant | -0.899 | 2.150 | 0.175 | 1 | 0.676 | 0.407 |

R² Nagelkerke: 0.218.

By breaking down the three domains into their respective individual components, it can be seen that the aforementioned result seems to be due to quantitative reasoning regarding personal consequences (p-value = 0.006). Furthermore, in this last model the one relating to General understanding of data processing also appears among the relevant predictors (p-value = 0.031). (Table 18)

Table 18. Estimates of the Beta coefficients of the components of the domains on Evaluation for research purposes.

| Regressors | B | s.e. | Wald | d.f. | p-value | Exp(B) |
|--|--------|-------|-------|------|---------|---------|
| General understanding given | 1.504 | 1.182 | 1.620 | 1 | 0.203 | 4.499 |
| General understanding of data processing | 2.931 | 1.358 | 4.662 | 1 | 0.031 | 18.747 |
| General understanding FSE | 2.386 | 2.242 | 1.133 | 1 | 0.287 | 10.871 |
| Understanding ESF objectives | 0.387 | 1.686 | 0.053 | 1 | 0.819 | 1.472 |
| Understanding research purposes | -4.138 | 2.101 | 3.879 | 1 | 0.049 | 0.016 |
| Understanding the purpose of risks | -3.262 | 2.508 | 1.691 | 1 | 0.193 | 0.038 |
| FSE qualitative reasoning | 2.033 | 1.342 | 2.295 | 1 | 0.130 | 7.636 |
| Qualitative reasoning research | 3.490 | 1.843 | 3.585 | 1 | 0.058 | 32.771 |
| Personal consequences | 6.687 | 2.437 | 7.530 | 1 | 0.006 | 801.903 |
| Consequences for the community | 2.894 | 1.596 | 3.287 | 1 | 0.070 | 18.058 |
| Logical coherence interviewer | -3.478 | 2.168 | 2.573 | 1 | 0.109 | 0.031 |
| Total final choice expression | -0.614 | 2.338 | 0.069 | 1 | 0.793 | 0.541 |
| Constant | -1.078 | 3.242 | 0.111 | 1 | 0.739 | 0.340 |

R² Nagelkerke: 0.218.

6. Results of Analysis of the Relationships between the Variables

From the analysis of the sample data, useful for understanding and evaluating the abilities of the patients interviewed in relation to the four domains of the ability to give consent to the processing of personal health data, it appears that the final expression of the choice seems to essentially depend on reasoning and, in particular, from the patient's ability to enumerate what consequences his choice may have on his daily life.

Moving on to analyze in detail the two components of qualitative and quantitative reasoning, only the relationship between qualitative reasoning, for research purposes, with the domain of understanding and with the components of understanding relating to data processing emerges as significant and their use for research activities. Evaluation, however, has less importance in explaining qualitative reasoning and becomes relevant only when the evaluation is made relative to the research purposes.

Quantitative reasoning, if we consider the consequences in everyday life, is influenced by the domain of evaluation and by evaluation for research purposes. The domain of understanding, however, has a modest impact on evaluation and only for the component relating to research purposes. The quantitative reasoning, relating to the consequences for the community, does not appear to have significant relationships with respect to the other variables examined in the analyzes carried out. Furthermore, the logistic regression model did not provide noteworthy results for the logical coherence component.

Finally, considering the domain of evaluation, in the case of goals relating to personal health, no significant relationships are found with other variables; in the case of evaluation for research purposes, a marked relationship emerges between evaluation and reasoning and with quantitative reasoning relating to personal consequences.

Albeit to a less relevant extent, the relationship of the evaluation with the general understanding of data processing must be considered.

The results emerging from the analysis of the dependency relationships between the variables under study must be interpreted considering the small size of the sample, used solely for the purpose of conducting the pilot survey, and, also, considering the low variability in the patients' responses interviewed.

7. Conclusions

The 2020 pandemic crisis highlighted the importance of the availability and management of health data to respond quickly to health emergencies while respecting the fundamental rights of individuals. This work, based on a pilot survey at the Polyclinic of Bari as part of the "Horizon Europe Seeds" project, aims to promote patients' awareness and trust regarding the processing of their health data, ensuring the protection of their rights and privacy.

Using the Personal Health Information Competence Assessment Tool (PHICAT), developed by the medical area of the project, the survey assessed patients' competence to consent to the processing of health data through a questionnaire. The data collected were used to assess patients' ability to understand the implications of using their data and to express informed consent. The results were analyzed in relation to the socio-demographic characteristics of the patients, which influence their propensity to give consent.

Analysis of the results leads, therefore, to new insights and a better understanding of the process that leads patients to the expression of a choice by also taking into consideration socio-demographic and clinical characteristics of the respondents themselves.

Patients' global scores in the four key domains (comprehension, reasoning, evaluation, and expression of a choice) ranged from 5 to 14, with a mean of 11.4 and a mode of 14. The 60% of patients scored between 11 and 14, 34% between 8 and 11, and only 6% between 5 and 8. Although scores do not directly determine legal capacity to give consent, an above-average score generally indicates good decision-making ability. However, very low scores, while suggesting decision-making difficulties, are not sufficient on their own to establish total incapacity.

The cluster analysis identified four main profiles of respondents, based on demographic characteristics, familiarity with the hospital environment and global scores, highlighting homogeneous groups by educational qualification and age that show different decision-making skills.

A second application of cluster analysis considered demographic variables,

familiarity with data management and global scoring and confirmed the presence of four clusters with similar compositions.

The analysis of the data showed that patients' ability to make informed decisions depends above all on their ability to reason about the consequences of their choices in daily life. Qualitative reasoning is significant when related to understanding data processing for research purposes, while quantitative reasoning is influenced by assessing personal consequences and understanding data processing. The results indicate the need to interpret the data with caution, due to the limited sample size and variability of the responses, as this is a pilot study.

The PHICAT methodology used allows, therefore, for the assessment of the capacity of patients' consent to data processing, through a semi-structured interview, reported in Attachment 1, and represents a fundamental starting point, including from a legal point of view, for understanding how to intervene in improving and incentivizing patients' giving of consent to their own health data.

A more in-depth discussion of the results and a closer exploration of the potential limitations concern a subsequent research paper of the Horizon Europe Seeds project and future developments of the methodology used will consider a larger sample than the pilot survey used in this paper.

This study can help improve patients' communication and information about the handling of their health data and foster greater confidence in granting consent for research and clinical management purposes.

A further consideration to be made on the topic under study is that of technological advances to improve patient consent and health data management; in particular, Blockchain technology [85]-[87], through a decentralized structure and advanced encryption methods, offer an opportunity to ensure data security, transparency and fairness.

Due to the guaranteed security and anonymization of one's personal data provided by such technology, there is a greater willingness for people to share and release data. This can increase the contribution to the use of sensitive data for the advancement of scientific research and the welfare of the community. Blockchain technology can, therefore, represent a great opportunity to increase individuals' consent to the processing of their sensitive data.

Funding

This study was carried out within Project ID S16 "Multidisciplinary Analysis of Technological Tracking Models of Contagion: The Protection of Rights in the Management of Health Data" - funded by the European Union - NextgenerationEU MUR Program - Promotion and Development Fund - DM 737/2021 CUP H91I21001610006 - Horizon Europe Seeds Notice issued by R.D. 1940 dated 04/06/2021. This manuscript reflects only the authors' view and opinions, neither the European Union nor the European Commission can be considered Responsible for them.

The contribution is the result of joint reflections by the authors.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Naous, D., Bonner, M., Humbert, M. and Legner, C. (2022) Learning from the Past to Improve the Future. *Business & Information Systems Engineering*, **64**, 597-614. <https://doi.org/10.1007/s12599-022-00742-2>
- [2] Sheikh, A., Anderson, M., Albala, S., Casadei, B., Franklin, B.D., Richards, M., *et al.* (2021) Health Information Technology and Digital Innovation for National Learning Health and Care Systems. *The Lancet Digital Health*, **3**, e383-e396. [https://doi.org/10.1016/s2589-7500\(21\)00005-4](https://doi.org/10.1016/s2589-7500(21)00005-4)
- [3] Pititto, G. (2020) La sfida delle app contro il Covid-19. *GEOmedia*, **24**, 12-19.
- [4] Alkhalifah, A. and Bukar, U.A. (2022) Examining the Prediction of COVID-19 Contact-Tracing App Adoption Using an Integrated Model and Hybrid Approach Analysis. *Frontiers in Public Health*, **10**, Article 847184. <https://doi.org/10.3389/fpubh.2022.847184>
- [5] Ong, A.K.S., Prasetyo, Y.T., Yuduang, N., Nadlifatin, R., Persada, S.F., Robas, K.P.E., *et al.* (2022) Utilization of Random Forest Classifier and Artificial Neural Network for Predicting Factors Influencing the Perceived Usability of COVID-19 Contact Tracing “Morchana” in Thailand. *International Journal of Environmental Research and Public Health*, **19**, Article 7979. <https://doi.org/10.3390/ijerph19137979>
- [6] große Deters, F., Meier, T., Milek, A. and Horn, A.B. (2021) Self-focused and Other-Focused Health Concerns as Predictors of the Uptake of Corona Contact Tracing Apps: Empirical Study. *Journal of Medical Internet Research*, **23**, e29268. <https://doi.org/10.2196/29268>
- [7] Ho, S.M., Liu, X., Seraj, M.S. and Dickey, S. (2022) Social Distance “Nudge:” A Context Aware Mhealth Intervention in Response to COVID Pandemics. *Computational and Mathematical Organization Theory*, **29**, 391-414. <https://doi.org/10.1007/s10588-022-09365-0>
- [8] Nishiyama, Y., Murakami, H., Suzuki, R., Oko, K., Sukeda, I., Sezaki, K., *et al.* (2022) MOCHA: Mobile Check-in Application for University Campuses beyond COVID-19. *Proceedings of the Twenty-Third International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, Seoul, 17-20 October 2022, 253-258.
- [9] Verma, R., Yabe, T. and Ukkusuri, S.V. (2021) Spatiotemporal Contact Density Explains the Disparity of COVID-19 Spread in Urban Neighborhoods. *Scientific Reports*, **11**, Article No. 10952. <https://doi.org/10.1038/s41598-021-90483-1>
- [10] O’Callaghan, M.E., Abbas, M., Buckley, J., Fitzgerald, B., Johnson, K., Laffey, J., *et al.* (2022) Public Opinion of the Irish “COVID Tracker” Digital Contact Tracing App: A National Survey. *Digital Health*, **8**, 1-12. <https://doi.org/10.1177/20552076221085065>
- [11] Kozyreva, A., Lorenz-Spreen, P., Lewandowsky, S., Garrett, P.M., Herzog, S.M., Pachur, T., *et al.* (2021) Psychological Factors Shaping Public Responses to COVID-19 Digital Contact Tracing Technologies in Germany. *Scientific Reports*, **11**, Article No. 18716. <https://doi.org/10.1038/s41598-021-98249-5>
- [12] Akinbi, A., Forshaw, M. and Blinkhorn, V. (2021) Contact Tracing Apps for the

- COVID-19 Pandemic: A Systematic Literature Review of Challenges and Future Directions for Neo-Liberal Societies. *Health Information Science and Systems*, **9**, Article No. 18. <https://doi.org/10.1007/s13755-021-00147-7>
- [13] Alarabi, L., Basalamah, S., Hendawi, A. and Abdalla, M. (2021) Traceall: A Real-Time Processing for Contact Tracing Using Indoor Trajectories. *Information*, **12**, Article 202. <https://doi.org/10.3390/info12050202>
- [14] Almalki, M. and Giannicchi, A. (2021) Health Apps for Combating COVID-19: Descriptive Review and Taxonomy. *JMIR mHealth and uHealth*, **9**, e24322. <https://doi.org/10.2196/24322>
- [15] Ami, J., Ishii, K., Sekimoto, Y., Masui, H., Ohmukai, I., Yamamoto, Y., et al. (2021) Computation of Infection Risk via Confidential Locational Entries: A Precedent Approach for Contact Tracing with Privacy Protection. *IEEE Access*, **9**, 87420-87433. <https://doi.org/10.1109/access.2021.3087478>
- [16] Ang, V. and Shar, L.K. (2021) COVID-19 One Year on—Security and Privacy Review of Contact Tracing Mobile Apps. *IEEE Pervasive Computing*, **20**, 61-70. <https://doi.org/10.1109/mprv.2021.3115478>
- [17] Avraam, D., Obradovich, N., Pescetelli, N., Cebrian, M. and Rutherford, A. (2021) The Network Limits of Infectious Disease Control via Occupation-Based Targeting. *Scientific Reports*, **11**, Article No. 22855. <https://doi.org/10.1038/s41598-021-02226-x>
- [18] Barthe, G., Viti, R.D., Druschel, P., Garg, D., Gomez-Rodriguez, M., Ingo, P., et al. (2022) Listening to Bluetooth Beacons for Epidemic Risk Mitigation. *Scientific Reports*, **12**, Article No. 5558. <https://doi.org/10.1038/s41598-022-09440-1>
- [19] Blom, A.G., Wenz, A., Cornesse, C., Rettig, T., Fikel, M., Friedel, S., et al. (2021) Barriers to the Large-Scale Adoption of a COVID-19 Contact Tracing App in Germany: Survey Study. *Journal of Medical Internet Research*, **23**, e23362. <https://doi.org/10.2196/23362>
- [20] Bodini, A., Pasquali, S., Pievatolo, A. and Ruggeri, F. (2021) Underdetection in a Stochastic SIR Model for the Analysis of the COVID-19 Italian Epidemic. *Stochastic Environmental Research and Risk Assessment*, **36**, 137-155. <https://doi.org/10.1007/s00477-021-02081-2>
- [21] Boldrini, C., Passarella, A. and Conti, M. (2022) Models for Digitally Contact-Traced Epidemics. *IEEE Access*, **10**, 106180-106190. <https://doi.org/10.1109/access.2022.3211425>
- [22] Caserotti, M., Girardi, P., Tasso, A., Rubaltelli, E., Lotto, L. and Gavaruzzi, T. (2022) Joint Analysis of the Intention to Vaccinate and to Use Contact Tracing App during the COVID-19 Pandemic. *Scientific Reports*, **12**, Article No. 793. <https://doi.org/10.1038/s41598-021-04765-9>
- [23] Cicala, F., Wang, W., Wang, T., Li, N., Bertino, E., Liang, F., et al. (2021) PURE: A Framework for Analyzing Proximity-Based Contact Tracing Protocols. *ACM Computing Surveys*, **55**, 1-36. <https://doi.org/10.1145/3485131>
- [24] Daniore, P., Ballouz, T., Menges, D. and Von Wyl, V. (2021) The Swisscovid Digital Proximity Tracing App After One Year: Were Expectations Fulfilled? *Swiss Medical Weekly*, **151**, w30031. <https://doi.org/10.4414/smw.2021.w30031>
- [25] Dave, R. and Gupta, R. (2021) Data Quality and Network Considerations for Mobile Contact Tracing and Health Monitoring. *Frontiers in Digital Health*, **3**, Article 590194. <https://doi.org/10.3389/fdgth.2021.590194>
- [26] Dowthwaite, L., Fischer, J., Perez Vallejos, E., Portillo, V., Nichele, E., Goulden, M., et al. (2021) Public Adoption of and Trust in the NHS COVID-19 Contact Tracing

- App in the United Kingdom: Quantitative Online Survey Study. *Journal of Medical Internet Research*, **23**, e29085. <https://doi.org/10.2196/29085>
- [27] Hassandoust, F., Akhlaghpour, S. and Johnston, A.C. (2020) Individuals' Privacy Concerns and Adoption of Contact Tracing Mobile Applications in a Pandemic: A Situational Privacy Calculus Perspective. *Journal of the American Medical Informatics Association*, **28**, 463-471. <https://doi.org/10.1093/jamia/ocaa240>
- [28] Fetzer, T. and Graeber, T. (2021) Measuring the Scientific Effectiveness of Contact Tracing: Evidence from a Natural Experiment. *Proceedings of the National Academy of Sciences of the United States of America*, **118**, e2100814118. <https://doi.org/10.1073/pnas.2100814118>
- [29] Frimpong, J.A. and Helleringer, S. (2021) Strategies to Increase Downloads of COVID-19 Exposure Notification Apps: A Discrete Choice Experiment. *PLOS ONE*, **16**, e0258945. <https://doi.org/10.1371/journal.pone.0258945>
- [30] Giustiniano, D., Bianchi, G., Conti, A., Bartoletti, S. and Melazzi, N.B. (2021) 5G and Beyond for Contact Tracing. *IEEE Communications Magazine*, **59**, 36-41. <https://doi.org/10.1109/mcom.010.2001250>
- [31] Goenka, R., Cao, S., Wong, C., Rajwade, A. and Baron, D. (2021) Contact Tracing Enhances the Efficiency of Covid-19 Group Testing. *ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, 6-11 June 2021, 8168-8172. <https://doi.org/10.1109/icassp39728.2021.9414034>
- [32] Haggag, O., Haggag, S., Grundy, J. and Abdelrazek, M. (2021) COVID-19 vs Social Media Apps: Does Privacy Really Matter? 2021 *IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*, Madrid, 25-28 May 2021, 48-57. <https://doi.org/10.1109/icse-seis52602.2021.00014>
- [33] Huang, J., Kwan, M. and Kim, J. (2021) How Culture and Sociopolitical Tensions Might Influence People's Acceptance of COVID-19 Control Measures That Use Individual-Level Georeferenced Data. *ISPRS International Journal of Geo-Information*, **10**, Article 490. <https://doi.org/10.3390/ijgi10070490>
- [34] Jiang, T., Zhang, Y., Zhang, M.H., Yu, T., Chen, Y.Z., Lu, C.H., Zhang, J., et al. (2022) A Survey on Contact Tracing: The Latest Advancements and Challenges. *ACM Transactions on Spatial Algorithms and Systems*, **8**, 1-35.
- [35] Jones, K. and Thompson, R. (2021) To Use or Not to Use a COVID-19 Contact Tracing App: Mixed Methods Survey in Wales. *JMIR mHealth and uHealth*, **9**, e29181. <https://doi.org/10.2196/29181>
- [36] Kahnbach, L., Lehr, D., Brandenburger, J., Mallwitz, T., Jent, S., Hannibal, S., et al. (2021) Quality and Adoption of COVID-19 Tracing Apps and Recommendations for Development: Systematic Interdisciplinary Review of European Apps. *Journal of Medical Internet Research*, **23**, e27989. <https://doi.org/10.2196/27989>
- [37] Kaptchuk, G., Goldstein, D.G., Hargittai, E., Hofman, J.M. and Redmiles, E.M. (2022) How Good Is Good Enough? Quantifying the Impact of Benefits, Accuracy, and Privacy on Willingness to Adopt COVID-19 Decision Aids. *Digital Threats: Research and Practice*, **3**, 1-18. <https://doi.org/10.1145/3488307>
- [38] Krüger, N., Behne, A., Beinke, J.H., Stibe, A. and Teuteberg, F. (2021) Exploring User Acceptance Determinants of Covid-19-Tracing Apps to Manage the Pandemic. *International Journal of Technology and Human Interaction*, **18**, 1-27. <https://doi.org/10.4018/ijthi.293197>
- [39] Liccardi, I., Alekseyev, J., Woltz, V.L.A., McLean, J.E. and Zurko, M.E. (2022) Public Willingness to Engage with COVID-19 Contact Tracing, Quarantine, and Expo-

- sure Notification. *Public Health Reports*, **137**, 90S-95S.
<https://doi.org/10.1177/00333549221125891>
- [40] Li, T., Yin, S., Yu, R., Feng, Y., Jiao, L., Shen, Y., et al. (2022) CoAvoid: Secure, Privacy-Preserved Tracing of Contacts for Infectious Diseases. *IEEE Journal on Selected Areas in Communications*, **40**, 3191-3206.
<https://doi.org/10.1109/jsac.2022.3211547>
- [41] Liu, M., Zhou, S., Jin, Q., Nishimura, S. and Ogihara, A. (2022) Effectiveness, Policy, and User Acceptance of COVID-19 Contact-Tracing Apps in the Post-Covid-19 Pandemic Era: Experience and Comparative Study. *JMIR Public Health and Surveillance*, **8**, e40233. <https://doi.org/10.2196/40233>
- [42] Maccari, L. and Cagno, V. (2021) Do We Need a Contact Tracing App? *Computer Communications*, **166**, 9-18. <https://doi.org/10.1016/j.comcom.2020.11.007>
- [43] Mahmood, M., Mateu, J. and Hernández-Orallo, E. (2021) Contextual Contact Tracing Based on Stochastic Compartment Modeling and Spatial Risk Assessment. *Stochastic Environmental Research and Risk Assessment*, **36**, 893-917.
<https://doi.org/10.21203/rs.3.rs-455884/v1>
- [44] Mahmood, M., Amaral, A.V.R., Mateu, J. and Moraga, P. (2022) Modeling Infectious Disease Dynamics: Integrating Contact Tracing-Based Stochastic Compartment and Spatio-Temporal Risk Models. *Spatial Statistics*, **51**, Article ID: 100691.
<https://doi.org/10.1016/j.spasta.2022.100691>
- [45] Majeed, A. and Hwang, S.O. (2021) A Comprehensive Analysis of Privacy Protection Techniques Developed for COVID-19 Pandemic. *IEEE Access*, **9**, 164159-164187. <https://doi.org/10.1109/access.2021.3130610>
- [46] Martignoni, M.M., Renault, J., Baafi, J. and Hurford, A. (2022) Downsizing of COVID-19 Contact Tracing in Highly Immune Populations. *PLOS ONE*, **17**, e0268586. <https://doi.org/10.1371/journal.pone.0268586>
- [47] Montagni, I., Roussel, N., Thiébaud, R. and Tzourio, C. (2021) Health Care Students' Knowledge of and Attitudes, Beliefs, and Practices toward the French COVID-19 App: Cross-Sectional Questionnaire Study. *Journal of Medical Internet Research*, **23**, e26399. <https://doi.org/10.2196/26399>
- [48] Moreno López, J.A., Arregui García, B., Bentkowski, P., Bioglio, L., Pinotti, F., Boëlle, P., et al. (2021) Anatomy of Digital Contact Tracing: Role of Age, Transmission Setting, Adoption, and Case Detection. *Science Advances*, **7**, eabd8750.
<https://doi.org/10.1126/sciadv.abd8750>
- [49] Nanni, M., Andrienko, G., Barabási, A., Boldrini, C., Bonchi, F., Cattuto, C., et al. (2021) Give More Data, Awareness and Control to Individual Citizens, and They Will Help COVID-19 Containment. *Ethics and Information Technology*, **23**, 1-6.
<https://doi.org/10.1007/s10676-020-09572-w>
- [50] Narayanan, H.T.S. (2021) Contact Tracing Solution for Global Community. *The Computer Journal*, **64**, 1565-1574. <https://doi.org/10.1093/comjnl/bxab099>
- [51] O'Callaghan, M.E., Buckley, J., Fitzgerald, B., Johnson, K., et al. (2020) A National Survey of Attitudes to COVID-19 Digital Contact Tracing in the Republic of Ireland. *Irish Journal of Medical Science*, **190**, 863-887.
<https://doi.org/10.21203/rs.3.rs-40778/v1>
- [52] O'Connell, J., Abbas, M., Beecham, S., Buckley, J., Chochlov, M., Fitzgerald, B., et al. (2021) Best Practice Guidance for Digital Contact Tracing Apps: A Cross-Disciplinary Review of the Literature. *JMIR mHealth and uHealth*, **9**, e27753.
<https://doi.org/10.2196/27753>
- [53] Parlamento Europeo e Consiglio (2016) GDPR-Regolamento. (Ue) 2016/679. 4.5.2016

IT Gazzetta ufficiale dell'Unione europea L 119/1.

- [54] Pepper, C., Reyes-Cruz, G., Pena, A.R., Dowthwaite, L., Babbage, C.M., Wagner, H., *et al.* (2022) Understanding Trust and Changes in Use After a Year with the NHS COVID-19 Contact Tracing App in the United Kingdom: Longitudinal Mixed Methods Study. *Journal of Medical Internet Research*, **24**, e40558. <https://doi.org/10.2196/40558>
- [55] Reid, M., Enanoria, W., Stoltey, J., Philip, S., Fuchs, J., Lockwood, A., *et al.* (2021) The Sars-Cov-2 Pandemic: The Race to Trace: Contact Tracing Scale-Up in San Francisco—Early Lessons Learned. *Journal of Public Health Policy*, **42**, 211-221. <https://doi.org/10.1057/s41271-021-00285-y>
- [56] Ritsema, F., Bosdriesz, J.R., Leenstra, T., Petrignani, M.W.F., Coyer, L., Schreijer, A.J.M., *et al.* (2022) Factors Associated with Using the COVID-19 Mobile Contact-Tracing App among Individuals Diagnosed with Sars-Cov-2 in Amsterdam, the Netherlands: Observational Study. *JMIR mHealth and uHealth*, **10**, e31099. <https://doi.org/10.2196/31099>
- [57] Riza, E., Kakalou, E., Nitsa, E., Hodges-Mameletzis, I., Goggolidou, P., Terzidis, A., *et al.* (2021) Appraisal of a Contact Tracing Training Program for COVID-19 in Greece Focusing on Vulnerable Populations. *International Journal of Environmental Research and Public Health*, **18**, Article 9257. <https://doi.org/10.3390/ijerph18179257>
- [58] Rugani, G. (2020) Le condizioni ricavabili dal Regolamento generale sulla protezione dei dati per le applicazioni nazionali di tracciamento dei contatti: Alcune considerazioni. *European Papers Carnets Européens-Cuadernos Europeos-Quaderni Europei*.
- [59] Samhi, J., Allix, K., Bissyandé, T.F. and Klein, J. (2021) A First Look at Android Applications in Google Play Related to Covid-19. *Empirical Software Engineering*, **26**, Article No. 57. <https://doi.org/10.1007/s10664-021-09943-x>
- [60] Saw, Y.E., Tan, E.Y., Liu, J.S. and Liu, J.C. (2021) Predicting Public Uptake of Digital Contact Tracing during the COVID-19 Pandemic: Results from a Nationwide Survey in Singapore. *Journal of Medical Internet Research*, **23**, e24730. <https://doi.org/10.2196/24730>
- [61] Schultz, É., Touzani, R., Mancini, J. and Ward, J.K. (2022) From Contact Tracing to COVID-19 Pass Holder; the Tortured Journey of the French Tousanticovid Contact Tracing App. *Public Health*, **206**, 5-7. <https://doi.org/10.1016/j.puhe.2022.02.009>
- [62] Seberger, J.S. and Patil, S. (2021) Us and Them (and It): Social Orientation, Privacy Concerns, and Expected Use of Pandemic-Tracking Apps in the United States. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama, 8-13 May 2021, 1-19. <https://doi.org/10.1145/3411764.3445485>
- [63] Shahroz, M., Ahmad, F., Younis, M.S., Ahmad, N., Kamel Boulos, M.N., Vinuesa, R., *et al.* (2021) COVID-19 Digital Contact Tracing Applications and Techniques: A Review Post Initial Deployments. *Transportation Engineering*, **5**, Article ID: 100072. <https://doi.org/10.1016/j.treng.2021.100072>
- [64] Shoji, M., Ito, A., Cato, S., Iida, T., Ishida, K., Katsumata, H., *et al.* (2021) Prosociality and the Uptake of COVID-19 Contact Tracing Apps: Survey Analysis of Inter-generational Differences in Japan. *JMIR mHealth and uHealth*, **9**, e29923. <https://doi.org/10.2196/29923>
- [65] Suver, C. and Kuwana, E. (2021) Mhealth Wearables and Smartphone Health Tracking Apps: A Changing Privacy Landscape. *Information Services & Use*, **41**, 71-79. <https://doi.org/10.3233/isu-210114>

- [66] Szocska, M., Pollner, P., Schiszler, I., Joo, T., Palicz, T., McKee, M., et al. (2021) Countrywide Population Movement Monitoring Using Mobile Devices Generated (big) Data during the COVID-19 Crisis. *Scientific Reports*, **11**, Article No. 5943. <https://doi.org/10.1038/s41598-021-81873-6>
- [67] Tretiakov, A. and Hunter, I. (2021) User Experiences of the NZ COVID Tracer App in New Zealand: Thematic Analysis of Interviews. *JMIR mHealth and uHealth*, **9**, e26318. <https://doi.org/10.2196/26318>
- [68] Tsinaraki, C., Mitton, I., Minghini, M., Micheli, M., Kotsev, A., Hernandez Quiros, L., et al. (2021) Mobile Apps to Fight the COVID-19 Crisis. *Data*, **6**, Article 106. <https://doi.org/10.3390/data6100106>
- [69] Utz, C., Becker, S., Schnitzler, T., Farke, F.M., Herbert, F., Schaewitz, L., et al. (2021) Apps against the Spread: Privacy Implications and User Acceptance of Covid-19-Related Smartphone Apps on Three Continents. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama, 8-13 May 2021, 1-22. <https://doi.org/10.1145/3411764.3445517>
- [70] Wilson, A.M., Aviles, N., Petrie, J.I., Beamer, P.I., Szabo, Z., Xie, M., et al. (2021) Quantifying Sars-CoV-2 Infection Risk within the Google/Apple Exposure Notification Framework to Inform Quarantine Recommendations. *Risk Analysis*, **42**, 162-176. <https://doi.org/10.1111/risa.13768>
- [71] Wu, S., Neill, R., De Foo, C., Chua, A.Q., Jung, A., Haldane, V., et al. (2021) Aggressive Containment, Suppression, and Mitigation of Covid-19: Lessons Learnt from Eight Countries. *BMJ*, **375**, e067508. <https://doi.org/10.1136/bmj-2021-067508>
- [72] Wu, Y., Ye, Q., Shen, F., Zhang, Z. and Jiang, C.L. (2022) Country- and App-Level Factors Affecting the Adoption and Evaluation of COVID-19 Mobile Apps. *BMC Public Health*, **22**, Article No. 2457. <https://doi.org/10.1186/s12889-022-14918-8>
- [73] Wymant, C., Ferretti, L., Tsallis, D., Charalambides, M., Abeler-Dörner, L., Bonsall, D., et al. (2021) The Epidemiological Impact of the NHS COVID-19 App. *Nature*, **594**, 408-412. <https://doi.org/10.1038/s41586-021-03606-z>
- [74] Yuan, Y., Ding, J., Wang, H., Jin, D. and Li, Y. (2022) Activity Trajectory Generation via Modeling Spatiotemporal Dynamics. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Washington DC, 14-18 August 2022, 4752-4762. <https://doi.org/10.1145/3534678.3542671>
- [75] Zaidi, A., Ahuja, R. and Shahabi, C. (2022) Differentially Private Occupancy Monitoring from Wifi Access Points. *2022 23rd IEEE International Conference on Mobile Data Management (MDM)*, Paphos, 6-9 June 2022, 361-366. <https://doi.org/10.1109/mdm55031.2022.00081>
- [76] Bradford, L., Aboy, M. and Liddell, K. (2020) COVID-19 Contact Tracing Apps: A Stress Test for Privacy, the GDPR, and Data Protection Regimes. *Journal of Law and the Biosciences*, **7**, Isaa034. <https://doi.org/10.1093/jlb/Isaa034>
- [77] Wang, T., Zhao, L., Cao, Y., Qu, Z. and Li, P. (2018) Medical Data Visualization Analysis and Processing Based on Machine Learning. *Journal of Computer and Communications*, **6**, 299-310. <https://doi.org/10.4236/jcc.2018.611027>
- [78] Ba'abbad, I., Althubiti, T., Alharbi, A., Alfarsi, K. and Rasheed, S. (2021) A Short Review of Classification Algorithms Accuracy for Data Prediction in Data Mining Applications. *Journal of Data Analysis and Information Processing*, **9**, 162-174. <https://doi.org/10.4236/jdaip.2021.93011>
- [79] El-Hadj Imorou, S. (2020) Socio-Economic and Health Determinants of Rural Households Consent to Prepay for Their Health Care in N'dali (North of Benin). *Open Journal of Social Sciences*, **8**, 348-360. <https://doi.org/10.4236/jss.2020.85024>

- [80] Massari, A., D'Addosio, V., De Nicolò, V.C. and L'Abbate, S. (2024) Privacy Protection in COVID Data Tracking: Textual Analysis of the Literature. *Applied Mathematics*, **15**, 235-255. <https://doi.org/10.4236/am.2024.153013>
- [81] Banfield, J.D. and Raftery, A.E. (1993) Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics*, **49**, 803-821. <https://doi.org/10.2307/2532201>
- [82] Agresti, A. (2018) An Introduction to Categorical Data Analysis. 3rd Edition, Wiley.
- [83] Hosmer, D.W. and Lemeshow, S. (2000) Applied Logistic Regression. Wiley. <https://doi.org/10.1002/0471722146>
- [84] Fabbri (1997) Statistica multivariata. Analisi esplorativa dei dati. McGraw-Hill Companies.
- [85] Sestino, A. and Merlo, V. (2021) Livello percepito di protezione dei dati personali e bisogno di sicurezza dei consumatori come predatori dell'intenzione all'uso di applicazioni mobile per il monitoraggio dei contagi da Covid-19. Uno studio preliminare. 287-300. <https://hdl.handle.net/10807/265176>
- [86] Merlo, V., Pio, G., Giusto, F. and Bilancia, M. (2023) On the Exploitation of the Blockchain Technology in the Healthcare Sector: A Systematic Review. *Expert Systems with Applications*, **213**, Article ID: 118897. <https://doi.org/10.1016/j.eswa.2022.118897>
- [87] Perchinunno, P., Massari, A., L'Abbate, S. and Crocetta, C. (2023) Blockchain and the General Data Protection Regulation: Healthcare Data Processing. In: Gervasi, O., et al. Eds., *Computational Science and Its Applications—ICCSA 2023 Workshops*, Springer, 377-388.