360-Degree Delphi: addressing sociotechnical challenges of healthcare IT

Heiko Dominik Waldmueller (heiko.waldmueller@stud-mail.uni-wuerzburg.de)
Universitatsklinikum Aachen https://orcid.org/0000-0002-1745-3060

Cord Spreckelsen
Universitatsklinikum Aachen, Institut für medizinische Informatik

Hannah Rudat
Universitatsklinikum Aachen, Institut für medizinische Informatik

Norbert Krumm
Universitatsklinikum Aachen, Palliativmedizin

Roman Rolke
Universitatsklinikum Aachen, Palliativmedizin

Stephan Michael Jonas
Technische Universitat Munchen

Research article

Keywords: Delphi technique, 360-Degree Delphi, stakeholder, project planning, feedback, medical informatics, consensus, quality improvement, qualitative research, health information management, group processes

Posted Date: August 2nd, 2019

DOI: https://doi.org/10.21203/rs.2.12019/v1

Version of Record: A version of this preprint was published at BMC Medical Informatics and Decision Making on June 5th, 2020. See the published version at https://doi.org/10.1186/s12911-020-1071-x.

Abstract

This approach uses the strong advantages a Delhi study has, identifies the disadvantages of traditional Delphi techniques and aims to introduce and evaluate a modified approach called 360-Degree Delphi. Key aspects of 360-Degree Delphi are tested by applying the approach to the needs and requirements analysis of a system for managing patients' advance directives and living wills. 360-Degree Delphi, as a modified Delphi process, is specified as a structured workflow with the optional use of stakeholder groups. Consensus is created within individual stakeholder groups, but is also communicated between groups, while the iterative structure of the Delphi process remains unchanged. We hypothesize that (1) 360-Degree Delphi yields complementary statements from different stakeholders, which would be lost in classical Delphi; while (2) the variation of statements within individual stakeholder groups is lower than within the total collective. A user study is performed that addresses five stakeholder groups (patients, relatives, medical doctors, nurses and software developers) on the topic of living will communication in an emergency context. About 25% of the invited experts (stakeholders) agreed to take part in the Delphi round 0 (three patients, two relatives, three medical doctors, two qualified nurses and three developers), forming a structured panel of the five stakeholder groups. Two raters created a bottom-up coding, and 238 thematic codes were identified by the qualitative text analysis. Based on a consented coding list, a quantitative online-questionnaire was developed and send to different stakeholder groups. With respect to the hypotheses, Delphi round 0 had the following results: (1) doctors had a completely different focus from all the other stakeholder groups on possible channels of communications with the patient; (2) the dispersion of codes within individual stakeholder groups and within the total collective was approximately 28% higher in the total collective than in the sub-collectives, but without a marked effect size. With respect to the hypotheses, Delphi round 1 had the following results: different stakeholder groups had highly diverging opinions with respect to central questions on IT-development. In comparison, the total group would not be representative of either of these individual stakeholder needs (mean 4,344, stdev +/- 1,870)

1. Background

Healthcare IT projects have critical sociotechnical implications, including challenging ethical, organizational and legal issues. As a consequence, their success depends on multi-disciplinary efforts in exploring project opportunities, eliciting and prioritizing requirements and agreeing preconditions for IT applications. In the past, several systematic approaches such as Delphi studies, focus groups and nominal group processes, have been used to support the requirements specification, consensus, and evaluation of healthcare IT projects [1]. Special sociotechnical implications with many involved parties can limit these approaches as individual needs might not be represented in a group consensus.

Healthcare IT projects tend to accumulate the same problems that are well-known in IT projects in general: the famous (although controversial) CHAOS report and other IT surveys have revealed that many

projects are started but then cancelled (15.52 percent in 2005, 11.54 percent in 2007) or perform poorly [2-5]. A major reason behind these failures is the abundance of requirements and scope changes [3], while stakeholder involvement and user inclusion during the project process have proved to be critical for success [6,7]. This is especially true in clinical systems: a systematic investigation of 40 scientific articles identified the inclusion of all stakeholders in the implementation process as a success factor in 24 out of the 40 articles [8]. Thus, it is necessary to create the possibility of coping with the complexity of multiple stakeholders [9].

The Delphi technique (DT) is a well-established iterative survey technique for gathering information or reaching consensus in a collective [11]. The technique combines qualitative and quantitative data elicitation methods. Developed in 1963 by the American RAND Corporation and named after the oracle of Delphi [12], the DT was originally used for military decision-making. Nowadays it is widely used in medicine and healthcare [1,6]. The main advantage of this method is that it can find a consensus in a situation of uncertainty or when there is a lack of empirical evidence.

During an investigation into the communication of advance directives and last wills through web-based technologies, several shortcomings of the Delphi technique became apparent. Implementation of emergency and end-of-life patient data is a project of potentially national moment that can easily be challenged by any party using the system. It is therefore essential to include all groups of stakeholders and to cover all user needs appropriately. Advance directives and last wills are used or needed by both laypeople and healthcare professionals. Thus, user needs vary from the simplified creation of documents to the assisted interpretation of a living will. It has been reported that the choice of the expert group can have a strong influence on the DT outcome [13]. In a traditional expert group, aspects could be underrepresented because of the selection of the study collective. Stakeholder groups within the total collective might have incompatible needs for which agreement cannot be found, and this could lead to a watered-down version of a consensus [11].

Thus, the aim of this manuscript is to introduce a modification of the traditional Delphi technique to address the issues described above. Instead of one group of pre-selected experts, a structured expert board of system-stakeholders is surveyed. This means that practical and diverse statements will be made and can be incorporated into the project. Additionally, while a single collective is investigated, stakeholder groups with different interests in the project are created and evaluated separately. Specifically, (a) patients, (b) relatives, (c) doctors, (d) nurses and (e) developers are selected as different stakeholders. Using this approach, we are able to include opinions and statements that would not be covered using a single expert collective as is traditionally done. Consequently, the following hypotheses are formulated:

Hypothesis 1: Using the new DT design, complementary statements from different stakeholder groups can be collected that would be lost if experts were chosen from only one or a few stakeholder groups.

Hypothesis 2: The opinion of individual stakeholder groups has a lower dispersion than the opinion of the total collective; that is, the variation of statements within individual stakeholder groups is lower than it is within the total collective.

Both hypotheses will be evaluated on the basis of the users' unbiased statements and opinions. Hypothesis 2 could be proven on a quantitative round. However, a stronger point can be made if the dispersion can be shown within answers to open questions rather than within the fixed scale items of a quantitative questionnaire. Thus, the evaluation is performed on the results of the qualitative Delphi round, and the study participants are not influenced during the consensus phases or required to limit their answers to fixed scales.

2. Method

The traditional Delphi method consists of five phases: preparation (Phase 1), content development on the relevant topic with an optional so-called round 0 survey (Phase 2), initial content evaluation based on statements in a round 1 survey (Phase 3), re-evaluation of content statements with feedback of all participants in a round 2 survey (Phase 4), and, if necessary, repeats of the pattern of re-evaluation until a consensus is reached (Phase 5) (Fig. 1). The participants in the survey rounds are experts selected by the science or monitoring team. The definition of an expert in a certain area is subjective and finding the consensus can be influenced by the selection of the experts or the number of iterative rounds in the study [13].

2.1. 360-Degree Delphi Design – what's new?

The new approach is called 360-Degree Delphi (360°D for short) to represent the aspect that a collection of opinions is gathered. The name is a reference to 360-Degree feedback, in which the competence or performance of personnel is evaluated not only by their supervisors or managers, but also by their co-workers and subordinates. In a similar way, 360°D evaluates aspects of a topic by considering the opinion of all relevant stakeholders. Another main innovation of the 360°D approach, compared to the conventional DT, is the composition of the expert collective, which follows a practically-oriented stakeholder principle. To differentiate the stakeholder groups, the criteria for inclusion need to be precisely defined. Experts are usually selected according to the number of their publications, their prior

work in the field under investigation or the interest they have shown in the specific topic, although there are other methods. With the 360-Degree approach, people are selected on the basis of their participation in the domain. In other words, they are experts as the result of their daily roles and work in the area that will be influenced by the findings of the study. Another innovation is the mandatory performance of certain DT components that are usually voluntary (Fig. 2).

In a similar way to traditional DT, 360°D is structured in five main phases, but some aspects are partially modified (Fig. 2): (Phase 1) preparation phase, (Phase 2) qualitative round/questionnaire, (Phase 3) first qualitative round, (Phase 4) first consensus round, and (Phase 5) optional repetitions of Phase 4. Only changes in the sub-tasks performed are reported in the following sections: we refer readers to the online supplements for a complete itemized task description.

2.1.1 Phase 1 – preparation phase

The preparation phase consists of six steps, (1.1) the creation of the monitoring team, (1.2) the creation of the advisory board, (1.3) the definition of the scientific problem, (1.4) the definition of the stakeholder groups, (1.5) the creation of the survey infrastructure, and (1.6) the obtaining of ethical approval. Steps 1.1, 1.3 and 1.5 are performed in exactly the same way as in the traditional DT.

Step 1.2. Creation of the advisory board

In 360°D the founding of an advisory board is mandatory. The advisory board does not conduct the study, but it gives scientific assistance to minimize bias by the monitoring team. It is especially important in the 360°D approach to review the definition of the stakeholder groups [14]. The people involved in this board are science-related, not necessarily topic-related, and are experienced in evaluating scientific projects. The advisory board accompanies the study by meeting the monitoring team over the course of the study. Since the advisory board is not directly involved in the work, advantages and disadvantages can be identified and notified to the monitoring team.

Step 1.4. Definition of stakeholder groups

The principal new step in 360°D is the definition of stakeholder groups; the purpose here is to categorize the participants on the basis of their different (possibly corporate) interests in or views on the defined problem. In 360°D the term “participant” is used for any member of the study collective (which is the expert panel in classical DT). In finding a consensus regarding a certain topic, the participants’ current fields of interest and work should be related to that topic.

Step 1.6. Ethical advice and data protection statement

Ethical approval from an institutional review board or ethics committee and a statement or certificate by a privacy officer are recommended for the 360°D approach as part of good scientific practice. Ethical approval will be mandatory if the data acquired are related to individuals, or if there are questions related

to sensitive topics such as infectious state, social status or financial status. Similarly, data protection is critical not only for ensuring data quality but also to increase compliance by the study participants.

2.1.2. Phase 2 – qualitative round

Phase 2 (or round 0) links five sub-steps: (2.1) creation of open questions, (2.2) test-phase, (2.3) collective selection, (2.4) main poll, and (2.5) evaluation.

Step 2.1 Creation of open questions

Since the goal of the first question-phase is the collection of general knowledge and problem statements from a stakeholder perspective, open questions for qualitative data acquisition are formulated by the monitoring team using standard question design methods (e.g. [15,16]).

Using the defined problem and the defined stakeholder groups, the wording of the questions is adjusted to make them understandable by the stakeholder groups. Therefore, questions addressed to particular stakeholder groups might be formulated differently. The level of abstraction and the content should be the same for every stakeholder group. A change of perspective can help stakeholders to provide answers based on their role. For example, a question from a patient's perspective could be "How would you ***provide*** your will in an emergency situation?", while the analog from the perspective of a medical doctor could be "How would you ***gather*** the patient's will in an emergency situation?".

Step 2.2 Test-phase

Before the main poll, the qualitative open questions are tested to see whether they are understood correctly or whether editing is necessary. Participants in this test-phase should not be included in the later rounds. The questions can be tested either in face-to-face interviews or by a pre-survey created for test purposes. A major advantage with oral interviews is that direct and indirect feedback can be obtained from the participants.

Step 2.3. Collective selection

In addition to the creation of the survey, the monitoring team selects the participants. Participants are selected on the basis of a practically oriented stakeholder principle: their practical experience in the field of study, their possible future interaction with the system, or their current usage of similar systems. Each stakeholder group should contain approximately the same number of participants. It is recommended that a bigger pool is contacted, as the response rate is approximately 25% and the number of participants should not be less than ten for a qualitative collective performing a qualitative study [14].

Step 2.4 Main poll

The main quantitative poll is performed in the same way as round 0 with the general DT. However, depending on the questions, appropriately formulated questionnaires might be necessary for each stakeholder group (compare Step 2.1).

Step 2.5 Evaluation

The answers to the open questions consist of qualitative text data. To evaluate the results, a systematic approach is used. Some variants for qualitative data evaluation are available from social studies. Well-known techniques are “Grounded Theory” [18] or “Qualitative Content Analysis” [19,20]. To reduce the subjectivity of the qualitative analysis, multiple analyses by different people from the monitoring team can be performed independently and merged, providing the opportunity to assess inter-rater reliability. Performing the analysis separately for each stakeholder group is important, even if similar or almost identical questionnaires are used for all stakeholder groups. A statement catalog is created using, for example, a coding system based on the statements for each stakeholder group and in total. Grouped by codes derived by bottom-up coding (see [20]) and agreed by the raters, the catalog of statements provides the basis for creating quantitative items for the following rounds.

2.1.3. Phase 3 – First quantitative round

Phase 3 is the starting point for deriving consensus in phases 4 and 5. The participants score the items derived from the qualitative phase. The sub-steps of this round are similar to a traditional DT: (3.1) creation of closed questions for rounds 1 and 2, (3.2) test-phase, (3.3) collective selection (2), (3.4) round 1 main poll and (3.5) evaluation of round 1. Steps 3.2, 3.3, and 3.4 are performed as in traditional DT, with the collective being selected using the same practically oriented stakeholder principle as in Step 2.3. A new collective can be formed or persons can be added to the existing collective.

Step 3.1. Creation of closed questions for rounds 1 and 2

In this step the statement catalog from Step 2.5 is used to generate quantitative questionnaires by, for example, systematically transforming statements into Likert-scale items or priority rankings. Statements are selected according to their importance within the total collective but also within individual stakeholder groups. This guarantees that statements that are important to only one stakeholder group will still be reflected in the total consensus, even if this group is small or would have a low impact otherwise. The remaining process is performed as in traditional DT.

Step 3.5 Evaluation

The data collected in round 1 is quantitative data, so it can be analyzed by descriptive statistics, as in ordinary DT studies or any other survey. The choice of the analytical approach depends on the level of measurement chosen. A new aspect of 360°D is that data are aggregated for each stakeholder group separately instead of overall figures being generated as in traditional Delphi studies. The results are documented accordingly.

2.1.4. Phase 4 – first consensus round

In Phase 4 a consensus may appear. The following steps are conducted to perform phase 4: (4.1) adding of feedback, (4.2) round 2 main poll and (4.3) evaluation of round 2. Steps 4.2 and 4.3 are performed in

an identical way to Steps 3.4 and 3.5. If possible, the collective is not altered.

Step 4.1 Addition of feedback

The evaluation results of each statement for each stakeholder group are communicated in an understandable way to the total collective. For example, the average evaluation and dispersion of each stakeholder group for each question is displayed [14,21]. Feedback influences the responses of the participants towards a consensus. Displaying the means and dispersion of ratings for each stakeholder group instead of an overall mean and dispersion allows a faster consensus within stakeholder groups. This may be a benefit of 360°D over traditional DT, as the latter would not reach overall consensus if two stakeholder groups cannot find a common consensus on a statement.

2.1.5. Phase 5 – 2nd to *n*th consensus round

All further phases are performed in an identical way to Phase 4.

2.2. Evaluation techniques

The evaluation of Phase 1 of 360°D is performed by looking at several aspects: (1) inter-rater reliability is calculated in order to create a solid coding base, (2) Hypothesis 1 is evaluated using excerpts of the coding catalog, and (3) Hypothesis 2 is evaluated by calculating the dispersion of statements within stakeholder groups. Similarly, statistics from Phase 3 are reported alongside to support the initial findings for Hypotheses 2.

2.2.1. Inter-rater reliability

Two coders receive the raw collected data and independently create a coding structure and base. Data are read and the coding base and corresponding statement coding are created in a bottom-up procedure resulting in codes associated with parts of the raw statements. The coding versions by the two coders are compared to check if the bottom-up procedures for each coder show similarities. Code matches between the codes are created based on position in the statement texts and the semantic meaning of the code label. If multiple codings occur at a text position for one coder, but not for the other one, only one code match is created and other codings are labeled as unmatched. Since the code structure and code base are created independently, the semantic meaning of individual codes could vary. Each matched code was therefore presented to three evaluators who rate the code matching either as a semantic match or as a mismatch. The rate of semantically matched codes based on a majority vote scheme (at least two out of three) and single vote scheme (one out of three) is reported in the results.

2.2.2. Evaluation of Hypothesis 1

To investigate Hypothesis 1, we assume that any of the five stakeholder groups would be eligible to stand as an expert collective for the traditional Delphi technique by themselves. Thus, we analyze whether any of the given statements (represented by their codings) are made exclusively or predominantly by one of the stakeholder groups, thereby being complementary to the other groups.

2.2.3. Evaluation of Hypothesis 2

The opinion dispersion of the stakeholders is measured using the code repository derived by the raters. Dispersion D is defined as the average standard deviation across the different coding occurrences. (see Eq. 1 in Supplemental Files)

with C being the different codes, the number of codes and the population standard deviation of occurrences within the code. (see Eq. 2 in Supplemental Files)

with S being the different stakeholders within the group, the number of times the code C is associated with stakeholder S, the average number of times code C is associated with stakeholders in the group, and the number of stakeholders.

Dispersion is reported per stakeholder group and for the total collective using box plots.

In addition, mean and standard deviation are reported and compared for the Phase 3 quantitative questionnaire.

2.3. Software Tools

The web-based tool LimeSurvey (Version 2.05, LimeSurvey GmbH, Hamburg, Germany) was used for online questionnaires. The coding and further analysis was supported by MaxQDA (Version 10, VERBI GmbH, Berlin, Germany), Microsoft Excel (Version 2011, Microsoft Corporation, Redmond, WA, USA) and Matlab (Version R2014b, The Mathworks, Natick, MA, USA).

3. Results

3.1. Study implementation

In order to verify the hypotheses, a 360°D process was performed at the University Hospital of RWTH Aachen University in the context of advance directives and living will communication. Here, 360°D Phases 1 through 3 are reported to indicate differences of opinion in stakeholder groups.

3.1.1. Phase 1 – Qualitative phase

Step 1.1. Creation of the monitoring team. The monitoring team (HW, SMJ) was assembled and tasks were distributed.

Step 1.2. Creation of advisory board. An advisory board (CS, RR) was assembled by the monitoring team. During the first two phases of the study, monthly meetings were held to control the study and verify decisions by the monitoring team.

Step 1.3. Problem definition. Little is known about the actual functionality and problems arising in the creation, distribution and interpretation of advance directives. Also, options for the improvement of communications about advance directives between people affected by an emergency situation and helpers are rare. The scientific problem was therefore defined as: “What are the current perceptions AND proposals about exchange of information in non-responsive situations?”

Step 1.4. Definition of stakeholders. Criteria for the stakeholder groups were defined. The monitoring team identified five stakeholder groups:

1. **Medical doctors**, who are assumed to act in the patient’s interest;
2. **Qualified nurses or caregivers**, because they have a close relationship with patients and might respond first in cases of emergency;
3. **Patients**, who need to formulate their will appropriately and select how it should be transmitted if they are unable to do so themselves;
4. **Relatives**, because they have to make medical decisions in absence of an advance directive or lack of direct patient-doctor communication; and
5. **Developers**, who develop and maintain the system.

Intersections between “patient” and “relative” or any of the others can occur in specific cases, such as a relative of a person with a severe health condition who is also undergoing medical treatment, or a patient who is also a medical professional. Therefore, each participant was instructed to answer as a member of the **most applicable** stakeholder group.

Step 1.5. Infrastructure. The contact channel was email, social networks, phone or personal, with a focus on email. The survey infrastructure was primarily online, to allow remote participation and limit interviewer bias.

Step 1.6. Ethical advice. Data security was defined by the data protection officer and confirmed: to allow anonymity the monitoring team agreed not to save the IP-addresses, timestamps, or any personal data of the participants. Ethical approval was obtained from the ethics committee of the Uniklinik RWTH Aachen (EK 363-15).

3.1.2. Phase 2 – Qualitative round

To investigate the difference between the current DT and the novel 360°D, a qualitative survey was performed. Round 0 statements were not set by the monitoring team, but were collected using a qualitative questionnaire to reduce bias.

Step 2.1. Creation of open questions. The monitoring team created open questions on different aspects related to the topic (see Addenda A and B). A scenario-based question form was developed to compensate for informational gaps. Additionally, technical terms were either avoided or were defined in the survey. An online survey was created for three different categories: (i) *medical professionals* for doctors and nurses, (ii) *care recipients* for patients and relatives and (iii) *developers* for developers. The main difference between the questionnaires was formulation and introduction to the topic.

Step 2.2. Test-phase. The questions were tested using semi-structured interviews. The test respondents were three medically qualified individuals (one doctor and two nurses) who were asked to select and then modify the questions for a better outcome.

Step 2.3. Collective definition. In order to find contacts appropriate to the collective, a list of medical institutions was created using the practically oriented stakeholder principle. To create a heterogeneous collective, there was an equal number of institutions with each of the following characteristics: level of care (maximum care institution, specialist care institution, primary healthcare, doctors office, retirement home), institution sponsor (church-related, public, private), speed of emergency reaction (fast, slow), medical area (oncological, internist, geriatric, emergency services), geographical area (rural, urban). Stakeholders were collected associated with the primary contact person (mostly doctors) such as patients of the doctor, the relatives of the primary contact persons patient. Software developers where identified by projects associated with the field of medical informatics. There was no personal relationship between monitoring team and collective.

Step 2.4 Round 0 main poll. The institutions were contacted by phone or email. A follow-up email, including information on the topic and a printable flyer with access to the online surveys (Short-URL, URL and QR-Code), was sent. Information of the project consisted out of the reason why the study is conducted, to establish an emergency data set for will communication, and a hypothetical use case. The contact person in the identified institutions was asked to print out and hand the flyer to: one patient, one doctor, one nurse and one relative of a patient. The survey was closed in January 2016 with the following numbers of answers: 30 in total, of which ten were complete. There were three in each of the medical doctor and patient stakeholder groups, and two in each of the qualified nurse and relative stakeholder groups. The response rate of the first contact by phone was 80%, and the response rate for completed answers was 25%. The stakeholder label was selected by the participant during the survey and saved in the survey database. The average time for the participants to complete the survey was 25.87 minutes. In addition, three developers with experience in medical informatics but not related to the project were surveyed.

Step 2.5 Evaluation of round 0. The database was converted into a single document for each participant so that data analysis could be performed (HW). The documents were named with a database ID and grouped according to the indicated type of stakeholder. Qualitative coding was performed for the full collective and the stakeholder groups with an inductive category development (bottom-up coding) [20]. Two coders performed the analysis independently using two different systems and compared their results (HW, NK). Both the coding and the coding structure were created independently. Rules for creating the categories were defined to evaluate similar content structures. The coding bases were subsequently consented and brought into a joined format.

Inter-rater reliability. Codings were matched and inter-rater reliability was calculated. In total, 238 codes were identified, of which 158 could be matched by their location. Out of the 158 matched codes, 107 (44.95% of total codes, 67.72% of matched codes) were evaluated as semantically similar in a majority-vote scheme, and 132 (55.46% and 83.54% respectively) in a single-vote scheme by three evaluators. In other words, 44.95% of the codes were created at the same position in the raw text, and two out of the three independent evaluators rated the codes given to these positions as semantically similar. Major sources of differences were variability in coded text passages, consistent coding of suggestions that occurred on multiple occasions, and level of abstraction. The coding and coding structure were then unified and agreed upon by the two coders to form a coding base for future use.

3.1.3. Phase 3 – First quantitative round (Traditional Delphi: Round 1)

Step 3.1 Creation of closed questions for Round 1. Based on the coding base, questions were identified and transformed into a quantitative survey (HR) (see Addendum C). The questions were divided into A “simple questions” and B “matrix questions”. Matrix questions are used to weigh different, potentially conflicting aspects against each other. In this case, privacy and data security, and ease of accessibility and change of data was weighted against each other, as they are often mutually exclusive in a system. For example, if personal data is stored encrypted using a password it is less prone to breach of privacy, but also not easily accessible in case of emergency. Especially in these topics the sub-groups did not match in the Phase 1 and 2 questionnaires.

Step 3.2 Test-phase. The questions were tested by scientists associated and non-associated to the 360 degree project. Markers of the test were (1) understanding questions, (2) duration filling out the survey, (3) selection among ordinal and other scales (SJ/HR/HW).

Step 3.3 Collective selection (2). Sub-groups were contacted on different channels: patients and relative associations via social platforms and support groups, doctors and caregivers via hospitals, IT-developers over direct contact and phone calls. (HR, HW). To access the survey a link for access over the internet was used.

Step 3.4 Round 1 main poll. The questions were converted into a single document for each participant so that data analysis could be performed (HR). The online survey was opened from Febuary 1st 2018 till October 10th in 2018. In total 62 persons participated with 60 persons (11 doctors, 6 caregivers, 9 relatives, 28 patients and 6 IT-developers) completing the questionnaire. The 2 uncompleted persons stopped filling out the form because of connection loss to the data server.

3.3. Hypothesis 1

Several instances were found in which statements were made exclusively or predominantly by just one of the stakeholder groups. The analysis regarding the channels for will communication in an emergency situation (Questions 1 through 4, Addenda A/B) show that the stakeholder groups of patients, relatives and qualified nurses named speech and gesture (i.e. direct contact) as possible options, in 12 statements. In contrast, doctors only once named direct contact with the patient as a possible channel (Table 1).

Indirect contact shows similar characteristics: doctors value relatives as a source of information, and were the only stakeholder group with seven out of a total of eight statements with this coding (Table 2).

General knowledge of advanced directives exists in the doctor, patient, developers and caregiver stakeholder groups, whereas relatives do not mention advance directives. In multiple responses by relatives, the term organ donation card appeared as a named form of will communication, whereas this term was not present in the other groups. Developers did not mention speech, gesture or direct contact at all. In terms of indirect will communication text, patient's will, organ donor card and advance directives were mentioned.

Any collective made of just one stakeholder group would therefore have missed at least one aspect brought up by another stakeholder group. This proves Hypothesis 1.

Table 1.

Table 2.

3.4. Hypothesis 2

Four Phase 1 and 2, the standard deviation across the population of the occurrences of codings is calculated as a measure of dispersion. It is calculated for each code and for each stakeholder group and the total collective (Fig. 4). Two questions were removed because they were unspecific (Question 7 "Do you have any other aspects to mention?" and Question 8 "What is your opinion of the project?"). The average standard deviation as a measure of dispersion is highest in the total collective. On average, the dispersion is 27.8% higher in the total collective than in the individual stakeholder groups but without a marked effect size.

Similarly, statistical analysis was performed for each question of the Round 1/Phase 3 questionnaire (SJ, HW). Results are reported as Box-plots (Figure 3). Due to the small sample size and study setup, no inferential statistics (T-Test, etc.) were performed but descriptive statistics are reported. While most sub-groups overlapped in their replies, several questions were answered almost mutually exclusive between sub-groups (Figure 3, Table 3). For example, caregivers and patients were in strong favor of easy access to all data in case of emergency (mean 5,667, Stdev: +/- 0,816 and mean 4,862, Stdev +/- 1,866 respectively), Likert scale, 1 means restrictive data access, 6 easy access to all data), while relatives preferred a more restrictive data access (mean 2,778, stdev +/- 1,093). Since the standard deviation is low in the sub-groups, this question could already be accepted as consented. In comparison, the total

distribution across all groups is not consented (mean 4,344, stdev +/- 1,870). Similar effects can be seen in most matrix questions (see Table 3), which supports Hypothesis 2.

Table 3.

4. Discussion

An implementation of the first three phases of the 360°D approach for the elicitation of the requirements for a medical IT project (a system for the communication of advance directives or last wills) was performed. The total collective was composed of five stakeholder groups, each having its own requirements and expectations. The first and second questionnaire rounds (traditionally known as Round 0 and Round 1) were completed and the evaluations were made. Two people, working independently, encoded the qualitative answers, creating an independent coding structure. Given the vast potential variety of the categories (codes) that were developed freely and independently by the raters, a high inter-rater reliability was achieved with a matching rate of semantically similar codes of 44.95% based on a majority vote from three independent evaluators.

The investigation yielded observations supporting the two hypotheses regarding the complementarity and focus of the different stakeholder groups. In the first case, it could be observed that none of the individual stakeholder groups was able to mention all the aspects named by the total collective of all stakeholder groups. In particular, traditional expert collectives such as medical doctors had a unique viewpoint with regards to the direct and indirect communication of the patient's will.

The approach did not allow inferential statistics to be applied, since this would have required a randomized quantitative study design. Thus, statistical significance cannot be attributed to the results. Nonetheless, the descriptive statistics revealed a trend concerning the dispersion of the statements, which turned out to be lower within each individual stakeholder group than in the total collective. In combination, the two hypotheses might even indicate larger problems with the traditional DT. Statements made by small or underrepresented stakeholder groups might not get attention in expert panels, as the focus might be dominated by the strongest stakeholder group in the panel. These results might be an explanation for the influence that the selection of the experts plays in Delphi studies [13]. The main limitation of this study is the small number of participants. However, a recent study in Nature Human Behaviour by Navajas et. al. showed, that combining as few as four consensus choices outperformed the wisdom of thousands of individuals [24]. Only very few patients and relatives could be found, despite the high number of invitations to participate in Phase 2. One explanation might be the intimacy of the topic of advance directive, as it is closely linked to the topic of death and loss. Another limitation of the 360°D method is the inherent consensus per stakeholder group. One has to redefine the concept of consensus for the analysis of the user needs, as no single consensus will be found that can guide development priorities: one consensus for each stakeholder group has to be considered. This shifts the burden of finding a final consensus to the study team, but gives them more information and might lead to solutions

that are better tailored towards the target audience and not only towards the expert panel. Partitioning the expert board into stakeholder groups decreases the heterogeneity of each group, which needs careful consideration: past findings regarding the assembling of the expert collective show that heterogeneous collectives are more likely to find highly acceptable solutions than homogeneous ones [22,23]. 360°D can be assumed to compensate for this effect by informing the (relatively homogeneous) stakeholder groups about the consensus reached by other groups during the Delphi process. Nonetheless, this aspect needs further investigation. Also, since the process of data base technology, world wide web and fast computing we think, that there are no limitations considering the processing of big data amounts as it used to be in the past. Printed lists and paper war are history and new possibilities for data evaluation should be applied to techniques like the Delphi method.

Conclusion

Our results support the assumption that the new 360°D approach has the potential to overcome certain drawbacks of the traditional DT by defining the expert collective more clearly and aiming for more heterogeneous stakeholder groups. By evaluating the results not only for the total collective but also for each stakeholder group, important aspects can be detected that might not be represented in larger expert collectives. By further adapting the questionnaires to the individual stakeholders' vocabulary or viewpoint, understanding and compliance might be increased and might result in more focused and useful answers. The next steps will be the continuation of the Delphi rounds towards an inter-stakeholder consensus, but more importantly an intra-stakeholder consensus and a step-by-step guide in "good Delphi-method practice". Further investigations might also help with understanding consensus building in general, for example if some stakeholder groups change their opinions in reaction to other groups.

Abbreviations

DT – Delphi Technique

360°D – 360-Degree Delphi Technique

RWTH Aachen University – Rheinisch Westfälische Technische Hochschule Aachen

SD, stdev – standard deviation

Declarations

(1) Ethics approval and consent to participate. Ethical approval was obtained from the ethics committee of the Uniklinik RWTH Aachen (EK 363-15)..

(2) Consent for publication. Not applicable.

(3) Availability of data and material. Not applicable.

(4) Competing Interests. None to declare.

(5) Funding. No external funding.

(6) Authors' contributions. The paper was written by Heiko Waldmüller. Stephan Jonas, Hannah Rudat, Norbert Krumm, Cord Spreckelsen and Roman Rolke revised and corrected the manuscript. Each step of the performed 360-Degree Delphi study is indicated with HW for Heiko Waldmüller, SJ for Prof. Stephan Jonas, HR for Hannah Rudat, NK for Norbert Krumm, CS for PD Dr. Cord Spreckelsen and/or RR for Prof. Roman Rolke.

(7) Acknowledgements. We acknowledge the help of Ralf Bettermann, Ina Fink, Marko Jovanovic, and Zafer Shishkov as evaluators of the similarity between the two coding systems by HW and NK.

References

1. Paré G, Sicotte C, Jaana M, Girouard D. Prioritizing the risk factors influencing the success of clinical information system projects. A Delphi study in Canada. Methods Inf Med. 2008;47(3):251–9.
2. The Standish Group. CHAOS Report 2015.
3. Emam KE, Koru AG. A replicated survey of IT software project failures. IEEE Software. 2008 Sep;25(5):84–90.
4. Berg M. Implementing information systems in health care organizations: myths and challenges. Int J Med Inform. 2001 Dec;64(2-3):143–56.
5. Heeks R. Health information systems: Failure, success and improvisation. Int J Med Inform. 2006 Feb;75(2):125–37.
6. Brender J, Ammenwerth E, Nykänen P, Talmon J. Factors influencing success and failure of health informatics systems – a pilot Delphi study. Methods Inf Med. 2006;45(1):125–36.
7. Hübner U. What are complex eHealth innovations and how do you measure them? Position Paper. Methods Inf Med. 2015;54(4):319–27.
8. Hoerbst A, Schweitzer M. A systematic investigation on barriers and critical success factors for clinical information systems in integrated care settings. Yearb Med Inform. 2015 Aug 13;10(1):79–89.
9. Southon G, Sauer C, Dampney K. Lessons from a failed information systems initiative: issues for complex organisations. Int J Med Inform. 1999 Jul;55(1):33–46.
10. Akkaya C, Wolf P, Krcmar H. Factors influencing citizen adoption of E-Government services: a cross-cultural comparison (research in progress). In: 2012 45th Hawaii International Conference on System Science (HICSS). 2012. p. 2531–40.
11. Powell C. The Delphi technique: myths and realities. J Adv Nurs. 2003 Feb 1;41(4):376–82.
12. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs. 2000 Oct 1;32(4):1008–15.

13. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health. 1984 Sep;74(9):979–83.
14. Häder M. Delphi-Befragungen: Ein Arbeitsbuch. 3. Aufl. 2014. Wiesbaden: Springer VS; 2013.
15. Agee J. Developing qualitative research questions: a reflective process. Int J Qual Stud Educ. 2009 Jul 1;22(4):431–47.
16. Rossi PH, Wright JD, Anderson AB. Handbook of survey research. San Diego, CA, USA: Academic Press; 2013.
17. Hoddinott SN, Bass MJ. The Dillman total design survey method. Can Fam Physician. 1986 Nov;32:2366–8.
18. Glaser BG, Strauss AL. The discovery of grounded theory: strategies for qualitative research. Piscataway, NJ, USA: Transaction Publishers; 2009.
19. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today. 2004 Feb;24(2):105–12.
20. Mayring P. Qualitative Inhaltsanalyse: Grundlagen und Techniken. Beltz GmbH, Julius; 2015.
21. Keeney S, Hasson F, McKenna H. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. J Adv Nurs. 2006 Jan 1;53(2):205–12.
22. Delbecq AL, Ven AHV de. A group process model for problem identification and program planning. J Appl Behav Sci. 1971 Jul 1;7(4):466–92.
23. Rowe EJ. Enhancing judgement and decision making : a critique and empirical investigation of the Delphi technique. [Internet] [PhD]. University of the West of England at Bristol; 1994 [cited 2016 Mar 21]. Available from: http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.261601.
24. Navajas J, Niella T, Garbulsky G, Bahrami B, Sigman M. Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. Nature Human Behaviour. 2018 Feb;2(2):126–32.

Tables

Table 1. Numbers of options for direct will communication in emergency situation per stakeholder group.

Stakeholder Group	Code		
	Speech	*Gesture*	*Direct Contact*
Patients	Count: 3 *"…if responsive by conversation."*	Count: 1 *"…by gesticulation."*	Count: 0
Relatives	Count: 2 *"speaking"*	Count: 3 *"…by alternative communication."*	Count: 0
Doctors	Count: 0	Count: 0	Count: 1 *"check, if possible to contact patient."*
Caregivers	Count: 2 *"Verbally."*	Count: 1 *"…nonverbal communication…"*	Count: 0
Developers	Count: 0	Count: 0	Count: 0

Table 2. Options for indirect will communication in emergency situation.

	Patients	**Relatives**	**Doctors**	**Caregivers**	**Developers**
Code	Amount	Amount	Amount	Amount	Amount
Text	2	0	0	0	1
Patient's will	3	1	5	3	2
Organ donor card	0	2	0	0	2
Advance directive	1	0	2	1	1
Presumption	0	0	3	0	0
Relative	0	0	7	1	0
Other	0	0	2	0	0

Table 3. Metrics for questions

	Metric (mean and standard deviation)					
Question	**Relatives**	**Doctors**	**Patients**	**Caregiver**	**Developers**	**Total**
Q1	Mean: 3,1	Mean: 4,4	Mean: 3,3	Mean: 4,3	Mean: 2,7	Mean: 3,5
	SD: 1,453	SD: 1,362	SD: 2,234	SD: 2,066	SD: 1,033	SD: 1,920
Q2	Mean: 2,8	Mean: 4,3	Mean: 4,9	Mean: 5,7	Mean: 3	Mean: 4,3
	SD: 1,093	SD: 1,793	SD: 1,866	SD: 0,817	SD: 1,789	SD: 1,870
Q3	Mean: 2,5	Mean: 4,4	Mean: 4,4	Mean: 4,4	Mean: 5,2	Mean: 4,0
	SD: 1,414	SD: 1,713	SD: 2,266	SD: 1,817	SD: 0,837	SD: 1,966
Q4	Mean: 3,1	Mean: 3,7	Mean: 4	Mean: 3,7	Mean: 2,7	Mean: 3,7
	SD: 1,453	SD: 1,421	SD: 2,187	SD: 2,066	SD: 1,033	SD: 1,870
Q5	Mean: 3,3	Mean: 3,8	Mean: 4,3	Mean: 3,2	Mean: 4,3	Mean: 4,0
	SD: 1,732	SD: 1,601	SD: 2,176	SD: 1,722	SD: 1,966	SD: 1,954

Additional Files

Additional File A. Questionnaire Round 0 for *medical professionals*

The following is translated from German. Details of the questionnaire are omitted for readability.

>>Situation description<<

Question 1: How is the will of the patient determined?

Question 2: What problems occur regarding the patient's will when selecting the next treatment steps?

Question 3: How do you proceed regarding the patient's will if the patient is not responsive?

Question 4: Can misunderstandings occur regarding the patient's will?

Question 5: How could this situation be improved from your perspective?

Question 6: What is the basis for the choice of medical therapy?

Question 7: Do you have any other aspects to mention?

>> project description <<

Question 8: What is your opinion of the project?

Question 9: What documents / information would help you?

Question 10: How and where would our project improve the previous situation for you?

Question 11: How and where would our project worsen the previous situation for you?

Question 12: Would you create a data set for yourself or a relative?

Question 13: Which group describes you most precisely?

>> selection of stakeholder groups <<

Additional File B. Questionnaire Round 0 for *patients*

The following is translated from German. Details of the questionnaire are omitted for readability.

>>Situation description<<

Question 1: How can you communicate your will?

Question 2: How is the described situation a problem for you regarding your will?

Question 3: What problems could occur, if you are not responsive?

Question 4: What misunderstandings can occur regarding your will?

Question 5: How could this situation be improved?

Question 6: What is the basis for the choice of medical therapy?

Question 7: Do you have any other aspects to mention?

>> project description <<

Question 8: What is your opinion of the project?

Question 9: What documents / information would help you?

Question 10: How and where would our project improve the previous situation for you?

Question 11: How and where would our project worsen the previous situation for you?

Question 12: Would you create a data set for yourself or a relative?

Question 13: Which group describes you most precisely?

>> selection of stakeholder groups <<

Additional File C. Questionnaire Round 1 (exemplaric for *patients*)

The following is translated from German. Details of the questionnaire are omitted for readability. In total 20 questions (sub-questions excluded) were asked, 15 “Simple questions” and 5 matrix questions.

>>project description<<

E-Mail for newsletter? (voluntary)

Question 1: Which stakeholder group do you belong to? (selection of stakeholder groups)

Question 2: Please indicate which documents you know?

Question 3: Please indicate which documents you filled out?

Question 4: What misunderstandings can occur regarding your will?

Question 5: Which of these documents should be online accessible for authorized persons?

Question 6: I do know what happens in an emergency if I'm not able to communicate anymore. (1-4)

Question 7: I have precise wishes should one of these situations happen (y/n)

Matrix questions:

Question 8: Privacy is important to me (1) à Fast data access is important to me (6)

Question 9: Data security is important to me (1) à Fast data inspection is important to me (6)

Question 10: I wouldn't release all medical data (1) à medical professionals should see data as fast as possible (6)

Question 11: I want lost data to be recovered fast (1) à I want to make sure non-authorized people have no data access (6)

Question 12: I want users to easily edit and see the data (1) à I don't want non-authorized people to edit or see the data (6)

Question 13: To create an advance directive I want to: Fill out a form, write a free text.

Question 14: I wish there was an easier way to communicate my will in an emergency situation

Question 15: I would create an emergency data set for me. (y/n)

Question 16: Who should be allowed to see my data? (recommendation answers)

Question 17: Who should be able to edit my data? (recommendation answers)

Question 18: How often should the data be updated? (recommendation answers)

Question 19: How much time would you invest in such a data set? (recommendation answers/other)

Question 20: Anything to mention?

Figures

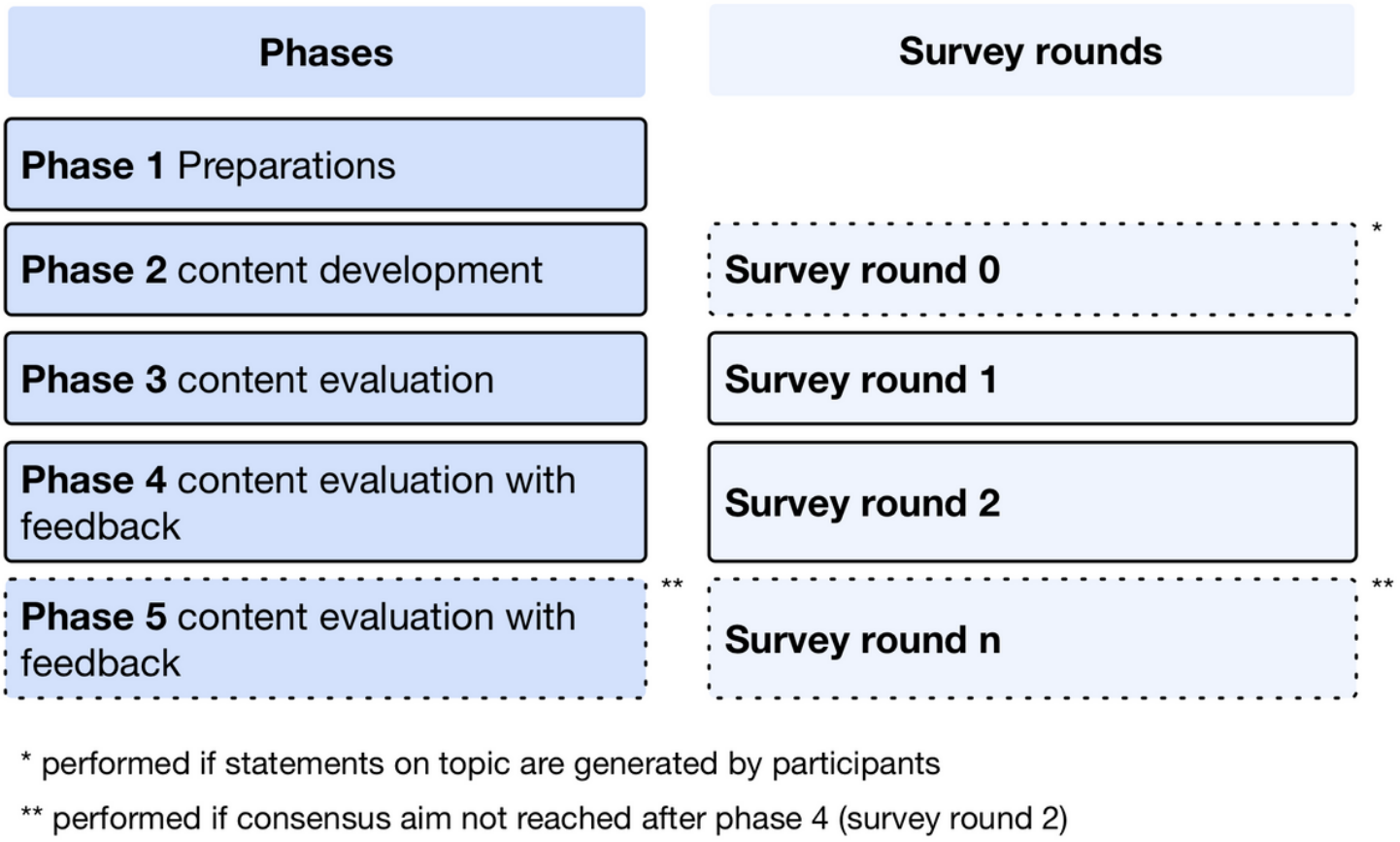

Figure 1

Phases and corresponding survey rounds

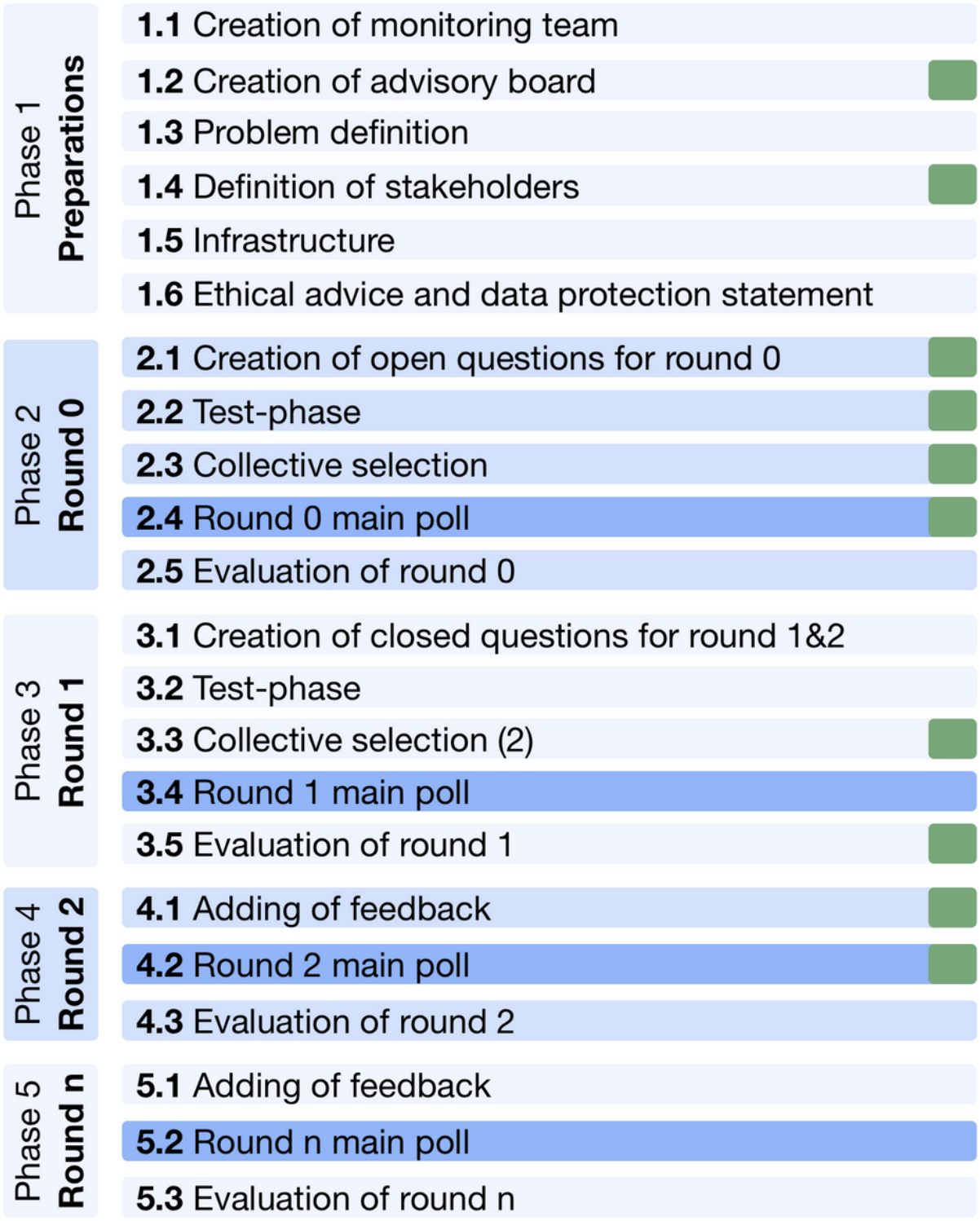

Figure 2

Structure of the 360-Degree Delphi (360°D) approach. Changes to the traditional Delphi technique and items that are now mandatory are marked with a green mark.

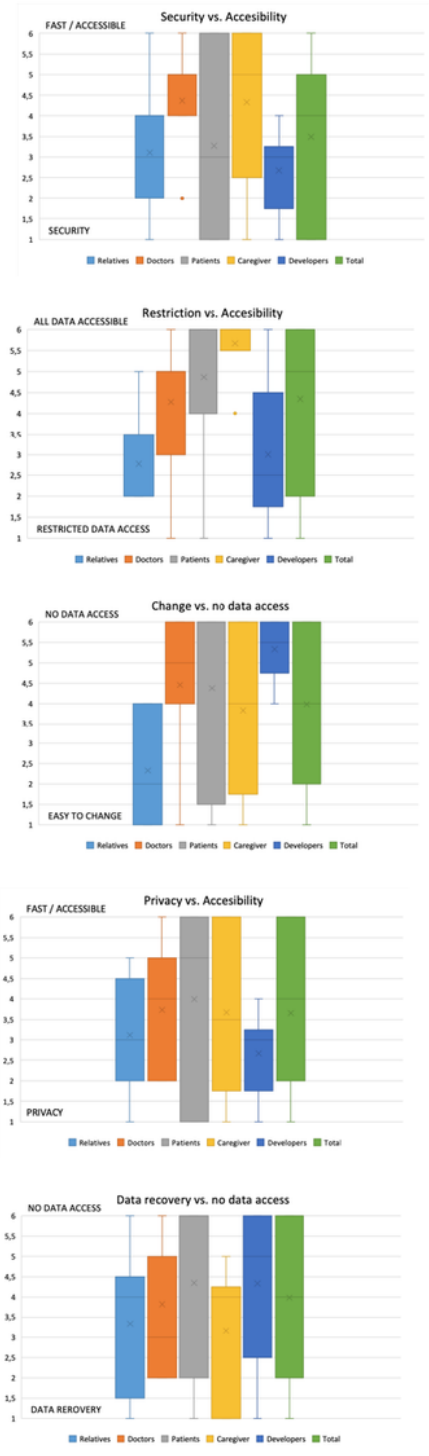

Figure 3

A) Security vs. Accessibility B) Restrictions vs. Accessibility C) Changes vs. no data access D) Privacy vs. Accessibility D) Privacy vs. Accessibility E) Data recovery vs. no data access

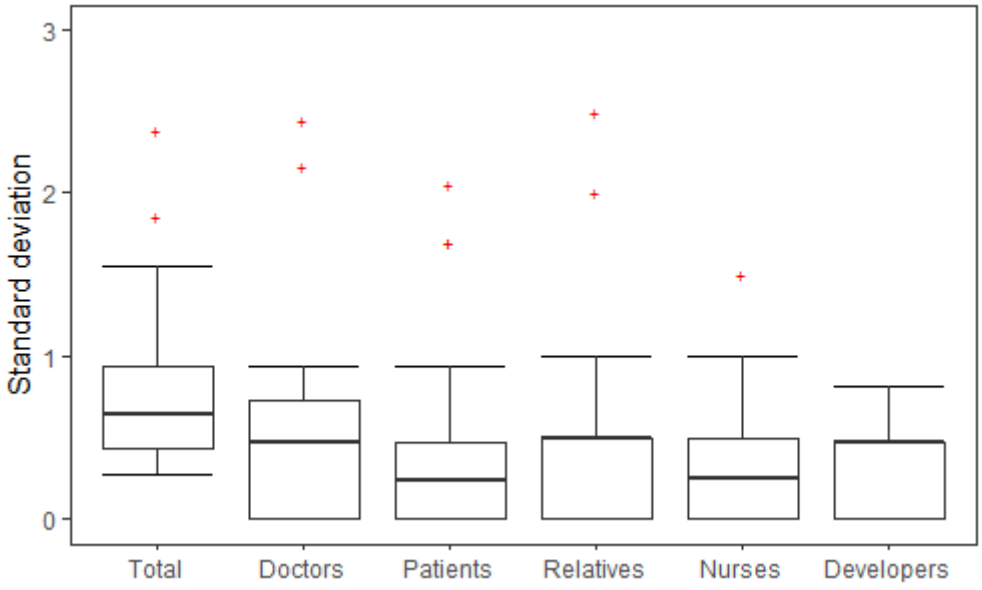

Figure 4

Comparison of dispersion within total collective and individual stakeholder groups in Round 0.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.jpg