HORIZON 2020
The EU Framework Programme for Research and Innovation

CAPABLE

Cancer Patients Better Life Experience

Grant Agreement No. 875052
Start Date: 01/01/2020 (48 Months)

Deliverable No. 5.1

Data ready for modelling and reasoning development, including procedures for anonymisation /pseudo-anonymisation

Due Date: 30/06/2020
Submitted On: 30/06/2020

Coordinator	University of Pavia (UNIPV)
Deliverable Lead Partner	IBM ISRAEL
Contributing Partners	UNIPV, AMC, NKI
Contact	Prof. Silvana Quaglini
Email	silvana.quaglini@unipv.it
Website	www.capable-project.eu

Deliverable Type		
R	Document, report	**X**
DEM	Demonstrator, pilot, prototype	
DEC	Websites, patent fillings, videos etc.	
OTHER		
Dissemination Level		
PU	Public	**X**
CO	Confidential (Consortium members including the Commission Services)	
CI	Classified Information (Commission Decision 2015/444/EC)	

This deliverable is a part of a project receiving funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875052.

Table of Contents

1. Versions History 4
2. Executive Summary 5
3. Data Sources 6
3.1. ICSM Hospital - Renal Cell Carcinoma 6
3.1.1. Data Set Details 6
3.1.2. Data Anonymization Process 6
3.1.3. Longitudinal Data Description 7
3.1.4. Data Dictionary 7
3.1.5. Possible predictions 8
3.2. AVL-NKI Hospital - Melanoma 8
3.2.1. Data Set Details 8
3.2.2. Data Anonymization Process 8
3.2.3. Longitudinal Data Description 9
3.2.4. Data Dictionary 9
3.2.5. Possible Predictions 9
3.3. AIMAC 9
3.3.1. Data Set Details 9
Data from the AIMAC Discussion Forum 9
Data from Regular Questionnaires 20
Data from COVID-related Questionnaire 22
3.3.2. Data Extraction and Anonymization Process 25
Forum Data 25
Regular Questionnaire Data 25
Covid-19 questionnaire data 25
Quality check of imported data 26
3.3.3. Longitudinal Description 26
Forum Data 26
Questionnaires Data 26
4. Glossary 27
5. References 28
6. Annexes 29
6.1. Annex 1 - ICSM Data Dictionary 29
6.2. Annex 2 - NKI Data Dictionary 37

List of Figures

Figure 1. Longitudinal data description of ICSM data 7
Figure 2. Longitudinal data description of NKI data 9
Figure 3. The page of the AIMAC website related to the patients' forum, see text for the full list of topics 10
Figure 4. Distribution of the number of characters per post. The y axis is in logarithmic scale 12

Figure 5. Distribution of the number of words per post. The y axis is in logarithmic scale 13
Figure 6. Distribution of the number of posts in a thread. The y axis is in logarithmic scale 14
Figure 7. Distribution of the number of sentences per message. The y axis is in logarithmic scale 16
Figure 8. Distribution of the number of characters in a sentence. The y axis is in logarithmic scale 17
Figure 9. Multiple answers are stored in the same column. This format must be parsed to allow specific data analysis 24

1. Versions History

Version	Date	Author	Comments
0.1	May 27, 2020	Ella Barkan, Flora Gilboa	Initial creation of draft
0.1	May 31, 2020	Ella Barkan, Flora Gilboa	First draft
0.2	June 3, 2020	Ella Barkan	Final Draft
0.3	June 22, 2020	Matteo Gabetta	Review
0.4	June 24, 2020	Flora Gilboa	Final Version after review

2. Executive Summary

This document describes patient's data coming from the three following data sources:

1. Data set from ICSM Hospital (Istituti Clinici Scientifici Maugeri) of Renal Cell Carcinoma (RCC) patients and their treatments

2. Data set for Melanoma cancer patients treated by immune checkpoint inhibitors from AVL-NKI (Antoni van Leeuwenhoekziekenhuis Netherlands Cancer Institute)

3. Three data sources from AIMAC (Italian Association for Cancer patients, relatives and friends)
    a. data from AIMAC's discussion forum
    b. data from questionnaires that are filled-in when patients contact AIMAC volunteers either by telephone or by accessing the AIMAC headquarters
    c. data from a questionnaire that has been recently put online during the COVID-19 pandemic

Status:

- Data agreements have been signed
- Data from all three sources: NKI, AIMAC and Maugeri was partially shared with CAPABLE partners.

3. Data Sources

3.1. ICSM Hospital - Renal Cell Carcinoma

3.1.1. Data Set Details

- Source: IMDC (International Metastatic Database Consortium) data set for Renal Cell Carcinoma (RCC) in which ICSM participates
- Size: 343 patients, 917 treatment lines
- Time period of reported treatments: 1998-2020
- Description: follow-up on treatments of RCC patients
- Main data elements:
  - Demographic information
  - Surgeries
  - Tumor characteristics
  - Basal Renal Functions, weight and all the hematological parameters at each treatment line
  - Metastases and brain metastases characteristics
  - Treatment details (drugs, adverse effects, dose change) and outcomes

3.1.2. Data Anonymization Process

Anonymization procedure included the 3 following steps. Proposed techniques are from PDPC [1]

1. **Attribute Suppression**

   - ***Attribute suppression*** refers to the removal of an entire part of data (e.g. "column" in databases and spreadsheets) in a dataset. It is used when an attribute is not required in the anonymised dataset, or when the attribute cannot otherwise be suitably anonymised with another technique. This technique has been applied at the start of the anonymisation process, as it is an easy way to decrease identifiability at this point

   - **ICSM** removed all attribute(s) related to Personal Information leaving year of birth and gender

2. **Pseudonymisation**

   - ***Pseudonymisation*** is replacement of identifying data with made up values. Pseudonyms can be irreversible, where the original values are properly disposed and the pseudonymisation was done in a non-repeatable fashion, or reversible (by the owner of the original data), where the original values are securely kept but can be retrieved and linked back to the pseudonym, should the need arises.

   - How to use it: replace the respective attribute values with made up values. One way to do this is to pre-generate a list of made up values, and randomly select from this list to replace each of the original values. The made up values should be unique, and should have no relationship to the original values (such that one can derive the original values from the pseudonyms). For reversible pseudonyms, the identity database cannot be shared with the recipient; it should be securely kept and can only be used by the organisation to resolve any specific queries (however, the number of such queries must be controlled, otherwise they can be used to "decode" the entire pseudonymisation).

- **ICSM** performed reversible pseudonymisation. It is used when data values need to be uniquely distinguished and where no character or any other implied information of the original attribute shall be kept.

3. Data Perturbation

- When ***Perturbation*** is applied, the values from the original dataset are modified to be slightly different. It is used for quasi-identifiers (typically numbers and dates) which may potentially be identified when combined with other data sources, and slight changes in value are acceptable, i.e. they do not affect the clinical interpretation of data. This technique should not be used where data accuracy is crucial.
- How to use it: it depends on the exact data perturbation technique used. These include rounding and adding random noise. The degree of perturbation should be proportionate to the range of values of the attribute. If the base is too small, the anonymisation effect will be weaker; on the other hand, if the base is too large, the end values will be too different from the original and utility of the dataset will likely be reduced.
- **ICSM** proposes to use this technique for patients, shifting all dates including the treatment related dates. The degree of perturbation has been discussed with clinicians in order to preserve the clinical significance of the results.

3.1.3. Longitudinal Data Description

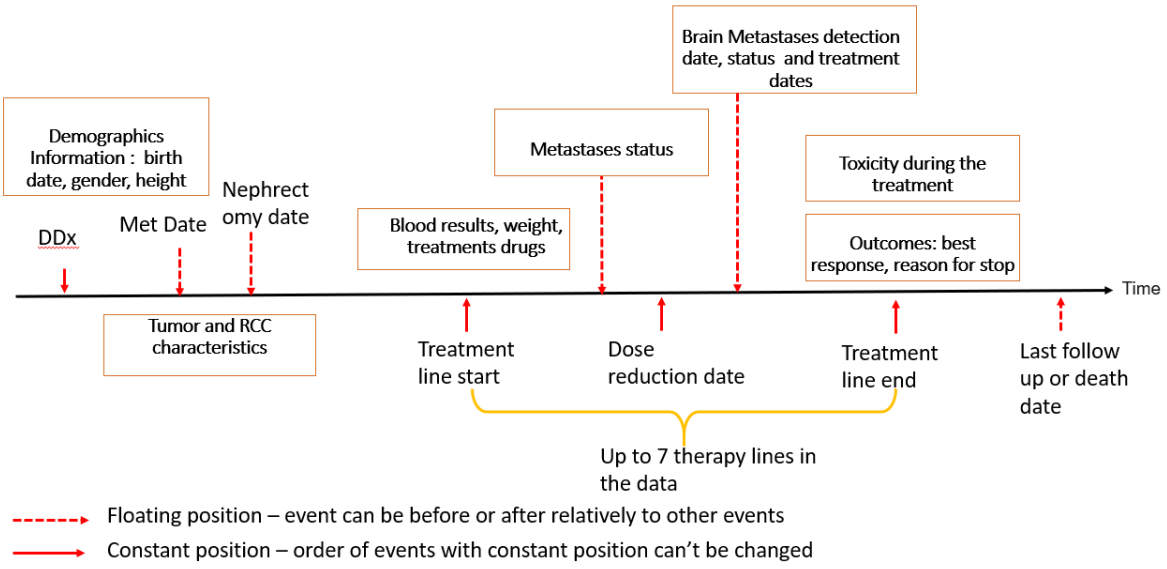

Figure 1. Longitudinal data description of ICSM data

Figure 1 describes the longitudinal perspective of ICSM data. After the diagnosis (DDX=Date of diagnosis) a renal cancer patient typically undergoes nephrectomy, and in general he/she is no more treated until metastases occur. However, metastases may already be present at DDX, and in that case nephrectomy may or may not be performed.

After metastases occur, drug treatment is initiated, and different lines of treatment may follow, according to the patient's response and treatment toxicity. Those are documented by blood examination and imaging.

Particular detail is reported for brain metastasis for their prognostic value.

3.1.4. Data Dictionary

Data dictionary has explanations for each field and element data in the provided data set. See Annex 1 in section 6.1

3.1.5. Possible predictions

(Definite list will be finalized after joined work with oncologists, based on the preliminary data analyses)

- Survival rate
- Best Response for line treatment
- Reason for stop of line treatment
- Toxicity existence and toxicity type during the line treatment

3.2. AVL-NKI Hospital - Melanoma

3.2.1. Data Set Details

- Source: DMTR (Dutch Melanoma Treatment Registry) data subset from patients with melanoma treated in AVL-NKI
- Description: follow-up on patients with stage III, unresectable stage III and stage IV melanoma treated with (adjuvant) immune checkpoint inhibitors
- Size: 500 patients, one treatment line for each patient
- Time period of reported treatments: 2015-2020
- Main data elements
    - Patient Characteristics: Demographic and Social information, Comorbidities, Primary Melanoma details
    - Staging: Pathology Details, Metastases, Blood Tests, Mutations
    - Additional Surgeries
    - Treatment Details: type of treatment, dose, administrations
    - Follow-up during the treatments: toxicity, metastases, blood tests, additional treatments (radiotherapy)
    - Treatments outcomes

3.2.2. Data Anonymization Process

After consulting with NKI Data Protection Officer and Information Security Officer Research/IT to ensure data protection, the following steps were taken:

1. The coordinating investigator in the NKI site received the full anonymous dataset from the data desk and removed all remaining patient identifiers (date of birth is made into categories of 10 years, dates that said anything about start/stop of treatment were regenerated into quarterly years);
2. An independent party reviewed the anonymization process of the coordinating investigator
3. Approval for data sharing was asked at Technology Transfer Office.

3.2.3. Longitudinal Data Description

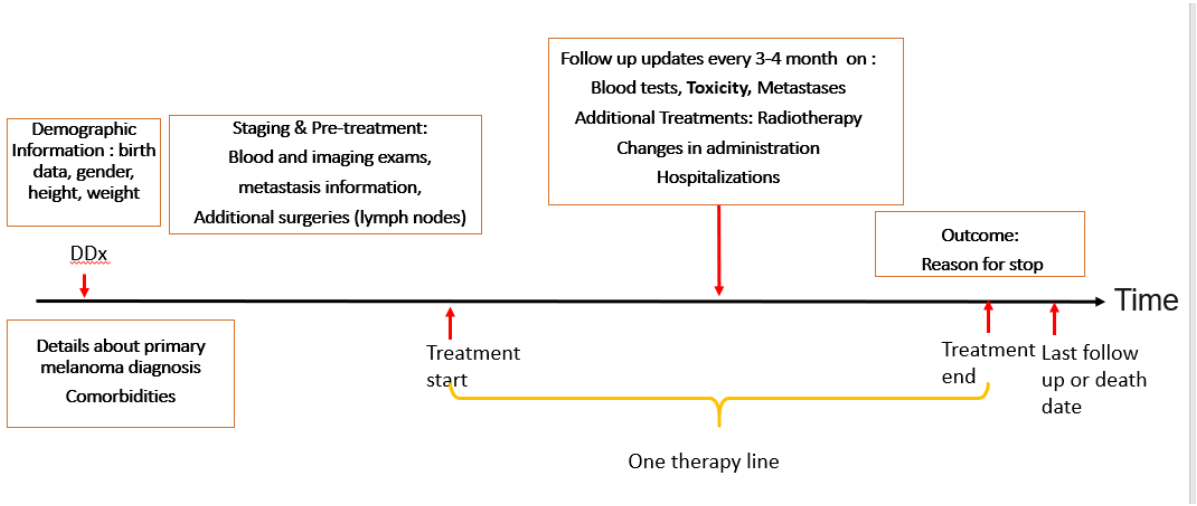

Figure 2. Longitudinal data description of NKI data

3.2.4. Data Dictionary

Data dictionary has explanations for each field and element data in the provided data set. See Annex 2 in section 6.2.

3.2.5. Possible Predictions

(The definite list will be finalized after further discussion with the oncologists, based on preliminary data analysis)

- Survival rate
- Reason for stop of line treatment
- Toxicity existence and toxicity type during the line treatment

3.3. AIMAC

3.3.1. Data Set Details

AIMAC, the cancer patients' association, provided three different kinds of data to the project:

1. data from its discussion forum;
2. data from questionnaires that are filled-in when patients contact AIMAC volunteers either by telephone or by accessing the AIMAC headquarters;
3. data from a questionnaire that has been recently put online during the COVID-19 pandemic.

Those three data sets will be described in the next sections.

Data from the AIMAC Discussion Forum

The forum contains unstructured data, i.e., texts (forum posts) that patients or their relatives share within their community. The forum is hosted by the AIMAC website (Figure 3).

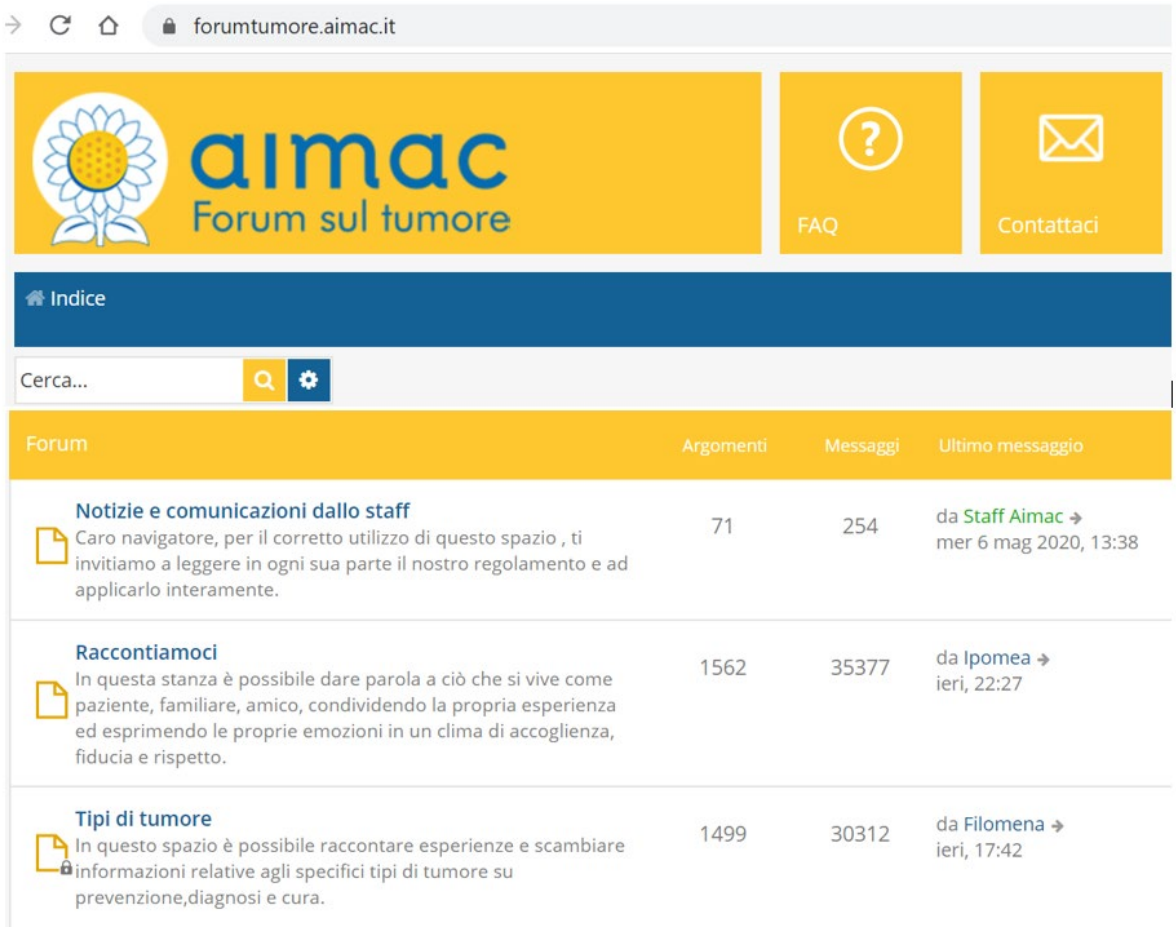

Figure 3. The page of the AIMAC website related to the patients' forum, see text for the full list of topics

The forum is structured in the following sections (under parenthesis the number of messages is reported, data at May 28th, 2020):

- News and communications from the staff (n=254)
- Let's talk about ourselves (n=35377)
- Types of cancer (n=30312)
- Living with and after cancer (n=768)
- Coping with tumors: treatments and rehabilitation (n=3098)
- Cancer and then ... life! (Our stories) (n=668)
- Off topic: scattered thoughts (n=5149)

The AIMAC forum is managed using the phpbb3 software [2]. This software stores messages and all the related information in several database tables. AIMAC provided three of these tables in CSV format. AIMAC filtered all the provided tables in order to remove every sensible information about the users.

The three provided tables are: phpbb_topics, phpbb_forums and phpbb_posts.

In this document, we focus on the fields that are of interest to the CAPABLE project.

phpbb_topics

This table contains data about the thread of messages. A thread is a collection of messages that replies to a particular message. This message is called thread starter and it sets the topic of the thread.

We considered the following information:

Column	Data
topic_id	Identifier of a thread of messages
forum_id	Identifier of a collection of thread on a specified subject

The topic_id is used to identify the messages that belong to the same thread.

phpbb_forums

This table contains data about the forums. A forum in the phpbb terminology is a collection of threads about the same subject. The subject of a forum is chosen by the forum administrator. When a user wants to post a message and to start a thread, it has to choose in which forum to post his message.

We considered the following information:

Column	**Data**
forum_id	Identifier of a collection of thread on a specified subject
forum_name	Name of the forum
forum_desc	Text that describe the subject of the forum

The forum_id is used to identify the messages and the threads that belong to the same forum.

phpbb_posts

This table contains all the information about the messages posted on the forum. In particular it contains the following information:

Column	**Data**
post_id	Identifier of a post
topic_id	Identifier of a thread of messages
forum_id	Identifier of a collection of thread on a specified subject
poster_id	Numerical identifier of the user that posted the message
post_text	Text of the message

The most important field is the post_text. It contains the full text of a message interleaved with some HTML tags used by the phpbb software to correctly format the text in the website.

To access the data the Python programming language [3] was used with the Pandas library [4].

Preliminary dataset analysis

To better understand the data in the AIMAC forum, a preliminary analysis of the dataset was performed to collect some useful insights.

The following statistics and figures are generated with scripts written using the Python language with the Pandas, NumPy [5] and Matplotlib [6] libraries.

Statistics about posts

Note: a word is considered as a group of characters separated by spaces or punctuation characters.

N° of posts	74930
N° of unique users	3955
Average number of character in a post	509 (Standard deviation: 695)
Minimum number of character in a post	1
Maximum number of character in a post	33868
Average number of word in a post[1]	99.3 (Standard deviation: 132.5)
Minimum number of word in a post[1]	0
Maximum number of word in a post[1]	5835

The statistics about posts show a heterogeneous distribution of posts length, both in terms of characters and words. The histograms in Figure 1 and Figure 2 show that a lot of messages are quite lengthy. Moreover, the length of a message can vary significantly from message to message. This may indicate that the useful information is surrounded by text that can be considered as "noise", making the process to retrieve the relevant information more complicated.

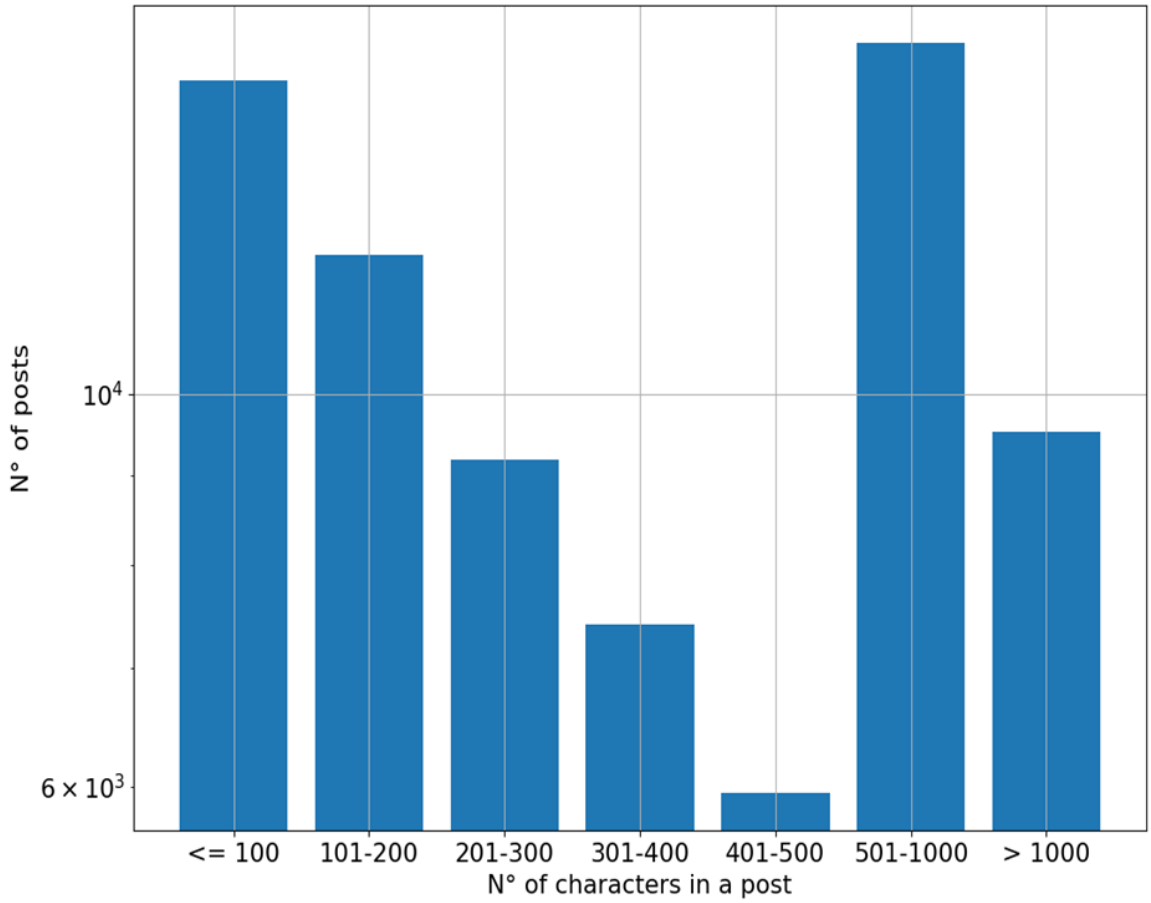

Figure 4. Distribution of the number of characters per post. The y axis is in logarithmic scale

[1] This statistic is calculated after the HTML tag filtering and the Message citation filtering

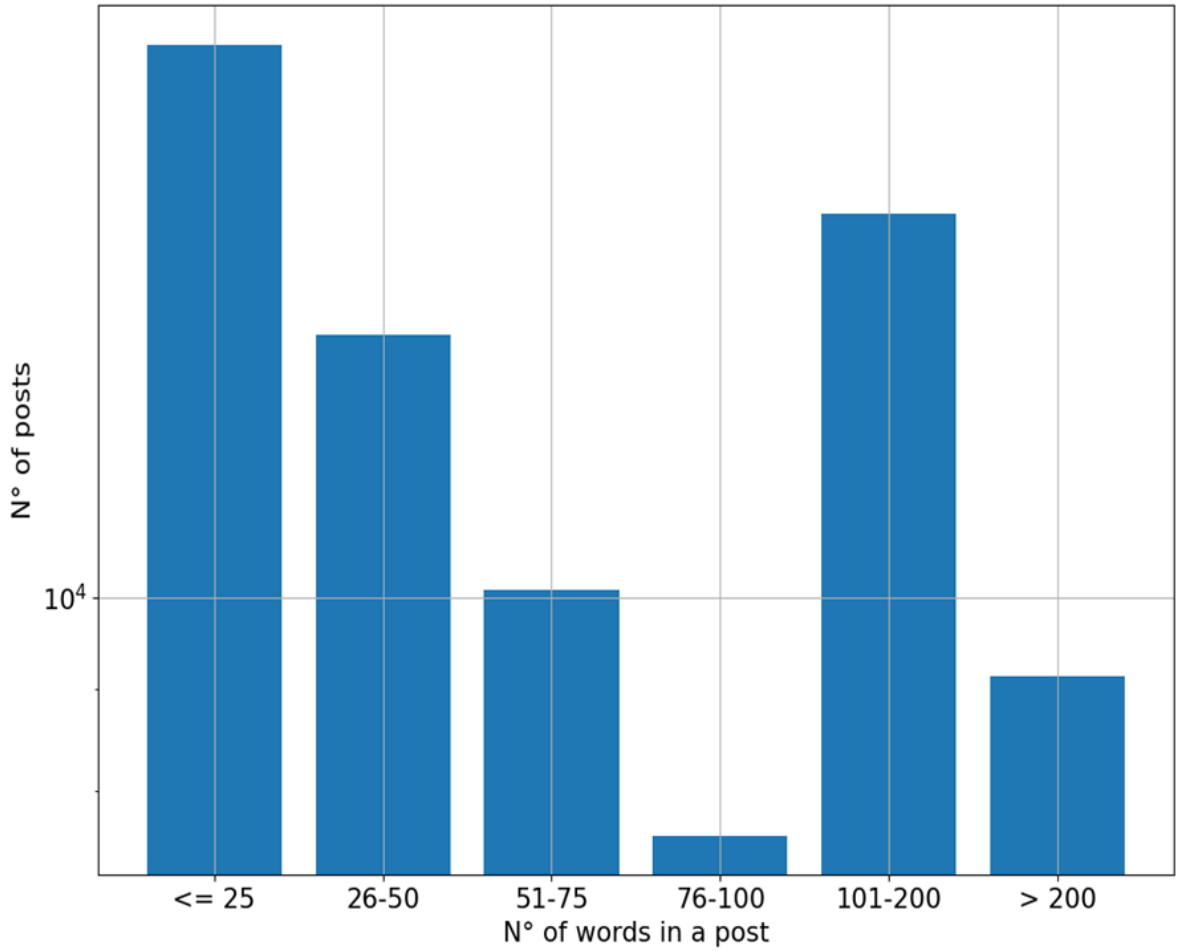

Figure 5. Distribution of the number of words per post. The y axis is in logarithmic scale

Statistics about forums

N° of forums	35
Average number of threads in a forum	117 (Standard deviation: 271,86)
Minimum number of threads in a forum	2
Maximum number of threads in a forum	1535
Average number of posts in a forum	2140.86 (Standard deviation: 5997.97)
Minimum number of posts in a forum	6
Maximum number of posts in a forum	35095

The statistics above show a high participation in each forum both in terms of posts and in terms of threads. This suggests that almost every subject proposed by the administrators is well discussed, and thus interesting for the audience.

Statistics about threads

N° of threads	4093
Average number of posts in a thread	18.31 (Standard deviation: 126.43)
Minimum number of posts in a thread	1
Maximum number of posts in a thread	3907

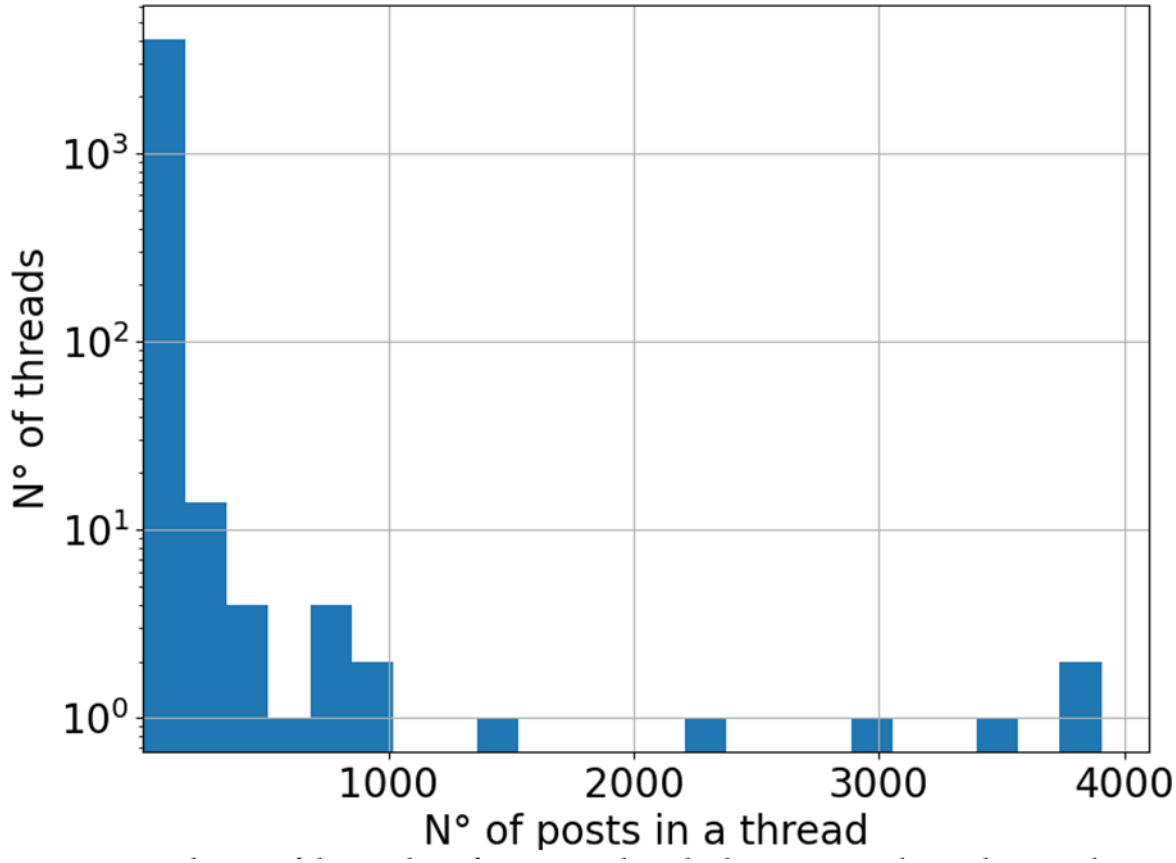

Figure 6. Distribution of the number of posts in a thread. The y axis is in logarithmic scale

Statistics about threads show that only few of them have a lot of messages. An average of 18 messages may indicate a good participation of the users in the discussions.

Data pre-processing

HTML tags filtering

As said in the Forum data section, the field post_text of the phpbb_posts table, contains many HTML tags interleaved in the text. These tags are used by the phpbb software to correctly format the text shown to the user in the web interface. Another use of tags is to keep the information about the URLs posted by the users in their messages.

Except for the URLs tags, all the other tags do not contain any valuable information for our purposes. Moreover the tag itself adds spurious text to the messages. For these reasons, an HTML tags filtering procedure was designed to remove all the tags from the text. If a URL tag is found in the text, this tag is removed, but the URL is kept. In this way we can retrieve the information of the URL at any time.

The HTML tags filtering procedure was developed in Python language, with the Pandas and NumPy libraries.

Message citation filtering

If a forum user cites text from another message, the phpbb software adds the cited text to post_text field of the message enclosed by a couple of particular tags. This cited text is useless for our purposes since it is a copy of the one in the original message. Therefore, it should be removed from the dataset.

For this reason a procedure to filter this kind of text was designed. This procedure looks for the citations tags and removes them along with the enclosed text.

This procedure was developed in Python language, with the Pandas and NumPy libraries.

Part-of-speech (POS) tagging

Part-of-speech (POS) tagging is the process to identify the part of speech (noun, verb, etc.) of a word in text using its definition or its context. This process is useful to determine the structure of a sentence and the role of a word in a text.

To perform the POS tagging, a tool called TreeTagger [7] is used. This tool is language independent and comes complete with parameter files for various languages, Italian included. This tool also provides an "end of sentence" tag that is helpful to split the messages into sentences.

Lemmatisation

Lemmatisation is the process of determining the lemma (also called *dictionary form*) of a word based on its intended meaning. Using only the dictionary form of a word instead of one of its many inflected forms, helps to better understand the meaning of a text and it also helps to find similarity in different messages.

The lemmatisation process is carried by the TreeTegger tool.

Sentence splitting

As shown in the "Statistics about posts" table, the messages have an heterogeneous distribution of length in terms of characters and also in terms of words. For this reason a sentence splitting procedure was designed. This subdivision allows us to work on a smaller unit of text, which is the sentence, making it easier to develop further processing algorithms.

This procedure uses the "end of sentence" POS tag to determine the possible end of a section. However the POS tagging procedure is often confused by the dot character after an abbreviation. To address this problem, the sentence splitting algorithm applies a set of rules to determine if the "end of sentence" POS tag is applied after an abbreviation and,in case, it handles the splitting accordingly.

Due to grammatical errors and mistakes in the use of punctuation in the messages, this algorithm is not able to correctly split all the sentences. Further development is needed to improve this pre-processing phase, although the performance of the current implementation are deemed as sufficient for providing enough text with the necessary quality to the next processing stages.

This procedure was developed in Python language, with the Pandas and NumPy libraries.

Statistics about sentences

The following statistics and figure are realized with the Python language using the Pandas, NumPy and Matplotlib libraries.

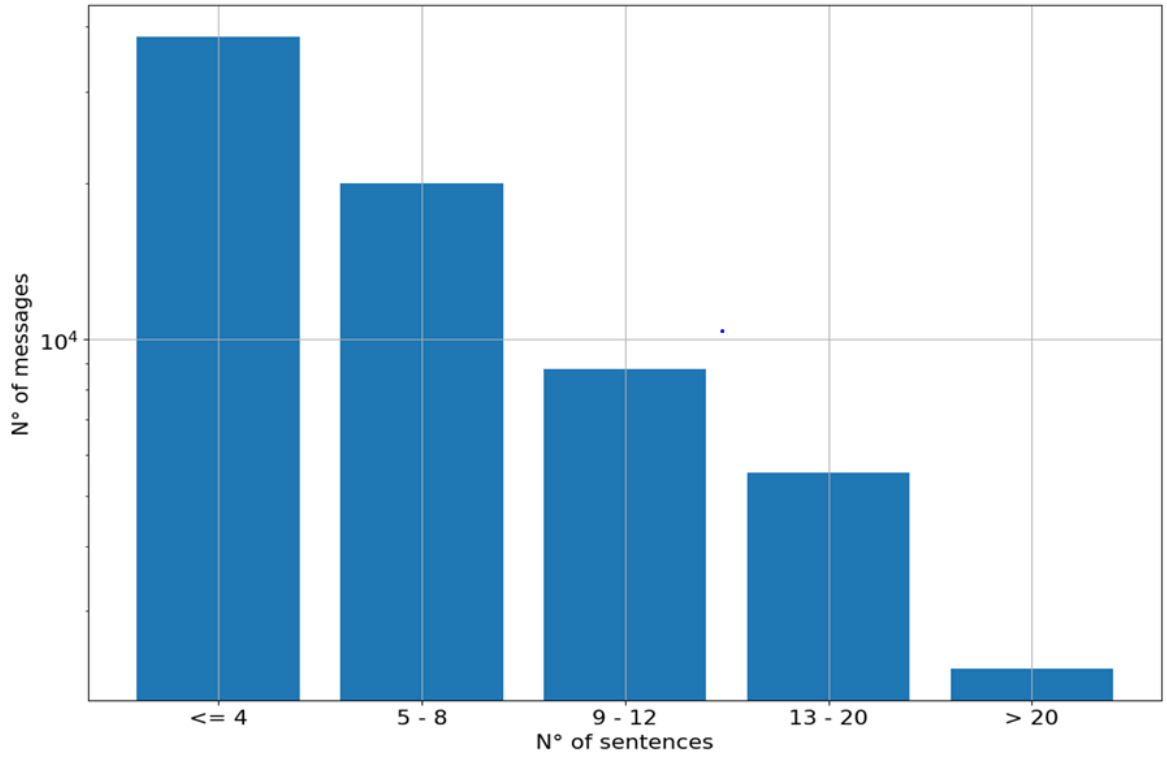

Figure 7. Distribution of the number of sentences per message. The y axis is in log-arithmic scale

Average number of sentences per message	6.18 (Standard deviation: 6.47)
Minimum number of sentences per message	1
Maximum number of sentences per message	226

The distribution in Figure 7 shows that roughly a half of the messages have 4 or less sentences and in general, messages are short in terms of number of sentences. However, there are some long messages in the dataset with more than 12 sentences, that can be hard to elaborate.

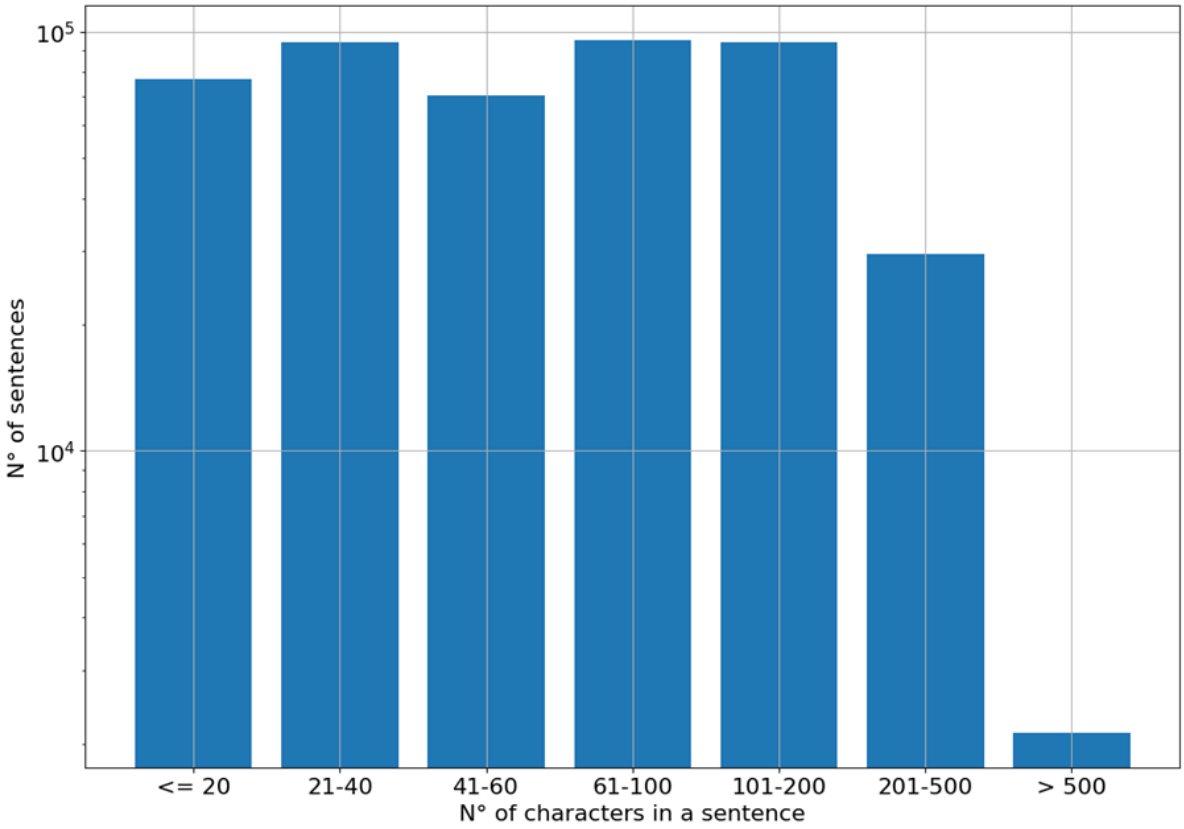

Figure 8. Distribution of the number of characters in a sentence. The y axis is in logarithmic scale

Average number of characters per sentence	80.94 (Standard deviation: 83.31)
Minimum number of characters per sentence	1
Maximum number of characters per sentence	3037

The distribution in Figure 8 shows that, in general, sentences are relatively short. This feature was expected and it is the reason for the introduction of the sentence splitting procedure. However, there are many sentences that have a length of more than 200 characters. This result suggests that those sentences may not be split adequately due to mistakes in punctuation usage.

Manual annotation of the messages

After getting the data, first of all, UNIPV set up a focus group of four persons, who firstly read a set of messages (about 200 each) separately, and then met for discussing:

1. the style of the messages;
2. the problems that could arise with NLP due to the message style, length of the messages, used words, etc.;
3. the motivations of patients and caregivers for writing messages;
4. the possibility of classifying the messages according to some attributes

The attributes that have been considered as most important are:

- the author of the message (caregiver/patient/other); the caregiver could also be detailed as daughter, son, friend, etc
- type of the message (outburst, sharing experience, asking an information, asking a suggestion, complain, rhetorical question, etc)
  In turn, asking for information may be detailed according to the type of information required: diagnostic test interpretation, centers of excellence for a given type of

cancer, information about possible diagnostic tests, information about medical experts, etc.

To prepare data for the analysis the following steps have been performed:

1. Every thread has been given an identifier
2. Every post has been given an identifier
3. Every post has been split in its sentences
4. Every post sentence has been given an identifier

As already mentioned, Step 3 was necessary because a lot of messages were very long (more than 500 words) and contained more than one argument of discussion. Thus, it is much more efficient to tag every sentence instead of the entire post.

In addition to the first 800 messages, the focus group participants are now reading another set of messages prepared in this way. Manual inspection is necessary in order to tag messages and create a repository of annotated messages that could be used as a training set for a supervised NLP technique that should be able to classify new messages.

The schema used for analyzing the messages is shown in the following, after reporting two real messages randomly sampled from the repository to make a practical example. The messages are already reported with their splitting in sentences.

ORIGINAL MESSAGE IN ITALIAN

id: 63505
Buonasera, scrivo x chiedere consiglio su una possibile alternativa alla urotac x indagare un eventuale carcinoma al rene sx. Attualmente l'indagine ha evidenziato formazione solida di 22x 26 mm che però è scarsamente impregnata, quindi, ciò non spiega perfettamente la sua natura… in attesa di consulto presso urologo di fiducia qualcuno saprebbe darmi info in merito? Grazie in anticipo. Antonia

Sentence splitting + ENGLISH TRANSLATION

id: 63505
0: Good evening, I write to ask for advice on a possible alternative to urotac for investigating a possible carcinoma in the left kidney.
1: Currently the investigation has revealed a solid formation of 22x 26 mm which is however poorly impregnated, therefore, this does not perfectly explain its nature.
2: Pending consultation with a trusted urologist, would anyone know how to give me info?
3: Thanks in advance.
4: Antonia

ORIGINAL MESSAGE IN ITALIAN

id: 65013
Buongiorno a tutti stamane ho ritirato il referto delle analisi del sangue di mia madre prescritte dall'oncologa come follow-up a seguito di k colon operato un anno esatto fa e successivo trattamento adiuvante chemioterapico di 6 mesi finito nel marzo scorso. C'erano alcuni valori sballati come globuli bianchi e rossi al di sotto della norma(questi ultimi di parecchio), lieve incremento della bilirubina, transaminasi e gamma GT. Marcatori tumorali negativi, anche se il Cea era sui 19(il limite mi pare fosse 35). Colesterolo molto alto, ma mia madre ha sempre sofferto di ipercolesterolemia ! Devo pensare che siano gli effetti della chemio(oxaliplatino + capecitabina) o potrebbe esserci qualcos'altro? La Tac fatta ad aprile era negativa.
Grazie a chi potrà o vorrà rispondere.

Sentence splitting + ENGLISH TRANSLATION

id: 65013
0: Good morning everyone this morning I collected the report of my mother's blood tests prescribed by the oncologist as a follow-up following a colon colon operated exactly one year ago and subsequent 6-month adjuvant chemotherapy treatment finished last March.
1: There were some bumped values like white and red blood cells below the norm (the latter of a lot), slight increase in bilirubin, transaminase and GT gamma.
2: Negative tumor markers, even if the Cea was around 19 (the limit seems to be 35).
3: Very high cholesterol, but my mother has always suffered from hypercholesterolemia.
4: Do I have to think that they are the effects of chemo (oxaliplatin + capecitabine) or could there be something else?
5: The CT scan done in April was negative.
6: Thanks to those who can or will answer

PostID	Sentence ID	Author	Containing a patient's need	Type	Info/Suggestion required	Contains question marks
63505	0	Patient	Y	Information request	Diagnostic test	N
63505	1	Patient	N			
63505	2	Patient	Y	Information request	Diagnostic test	Y
63505	3	Patient	N			
63505	4	Patient	N			
65013	0	Caregiver* (D)	N			
65013	1		N			
65013	2		N			
65013	3		N			
65013	4		Y	Test result interpretation	Follow-up test	Y
65013	5		N			
65013	6		N			

* D daughter, S son, F father, M mother, B brother, S sister, N nephew, G granma/pa, FR friend, O other

One thousand and 500 hundred messages will be manually annotated in order to build the training set for the supervised NLP techniques. According to the results obtained, additional batches of messages could be annotated

Data from Regular Questionnaires

Since 2012, AIMAC volunteers interview patients and caregivers when they call the association by telephone or directly go to the association headquarters for asking information or obtaining suggestions. These are structured interviews, so data will be easier to analyse with respect to the forum data.

In the following we report the guidelines that interviewers have to follow, because they are important to guarantee the good quality of the collected data.

The basic idea that guided the work of structuring this survey is to gather **more accurate information** in order to:

better understand the situation of the person who approaches us;

guide the person who approaches us by providing *a personalized answer*

better understand and interpret the information needs of the people who approach our Information Service, thus filling an important gap;
respond to research needs because it allows us to have and provide **indicators of effectiveness and efficiency** of our daily work which consist in **listening and supporting people with personalized information***;*
At the center of our work, in fact, there must always be a professional listening and recover the subjectivity of the person which is in front of us.

This requires **"a trained interviewer"**, and for this reason it is necessary to

have a good knowledge of the survey form, a knowledge that then allows us to explore as much as possible all the areas of the form

use **methods similar to the clinical interview** in order to establish a climate of trust and mutual respect in order for the volunteer to gather information such as:

sex and age; place of origin; possible reconstruction of the ecomap of the user who contacts us and/or on behalf of the sick person (who is around, with what role, with what tasks); type of tumor, current stage of the disease in relation to the information received (prediagnostic, diagnostic, therapeutic, progressing, disease-free, relapse, advanced-metastatic, terminal, follow-up, unknown to the user);therapeutic phase (no treatment, awaiting diagnosis, in treatment, post-treatment);type of treatment (surgery, chemotherapy, radiotherapy, hormone therapy, biological therapy, transplant ...);data indicating the relationship with the Services (general practitioner, hospital facilities, social services, voluntary associations ...).

IMPORTANT: the data must be collected together with the interview and for this, it is necessary that operators:

- explain to the patient that they are collecting data, asking to respond anonymously to some simple questions;

- emphasize that their answers are useful not only to be able to provide an adequate response to their personal "question" but also to improve our service for people with cancer and their families;

- point out that the data collected will be treated in accordance with GDPR

- request and obtain their consent

Data has been provided with a codebook, in such a way they are easily interpretable.

The questionnaire items are reported in the following.

Form N° : (automatically generated by the system) Date:
Information point: (generated automatically by the numerical code system of the information desk)

Request mode:

in person/ phone/ Email/ social network/ others (mail/fax)

The user benefits from the service

for the first time/ 2-3 times/ 3-4 times/ 5 times or more

The person requesting the information is:

patient / spouse-partner/ parent / son -daughter / other family member/ 6. Friend-acquaintance/ doctor-psychologist/ nurse / Volunteer/ nd

Patient Information

Gender of the patient

Age of the patient (in completed years or age class)

Patient Nationality: Italian / European Union country / Non-EU country / Asian Country / African country / Latin American country / nd

Province of residence of the patient

Patient's qualification: less than compulsory education / up to compulsory education / diploma degree and over /nd

At present the tumor is:

Primary (first tumor) / In recurrence (resumption of illness after a phase of complete or partial response to therapy) / With metastasis / In remission (no signs indicating that the disease is in progress).

How many years have passed since the first diagnosis

Phase of the diagnostic-therapeutic procedure

diagnosis waiting/Diagnosis/ surgery/ clinical trial / waiting for treatment/ medical treatment (chemo/radiotherapy, hormone therapy, etc.)/ palliative care / pain therapy / follow-up / follow-up visits/ long term survivor (free from disease and treatment for at least five years)/Other/ nd

The patient is

hospital admission/ in day hospital/ in home care/at home/ in hospice /nd

Site of the primary tumor:

colon/ Skin/ Esophagus/ Pharynx/ Liver/ Breast/ Bones/ Ovary/ Pancreas/ Penis/ pleura/ Lung/ prostate / kidney/ Rectum / hematopoietic system (eg. Lymphoma, leukemia, myeloma, etc.)/ neuroendocrine system/ SNC/ Stomach/soft tissue/ Testicles/ Thyme/ Thyroid/ Uterus/ vagina / Bladder/ biliary tract /Vulva/ other / nd

How did you find out about the information service:

family member or acquaintance/ another patient/ mass media / social network-internet/ brochure-poster-AIMAC volunteers/ health workers (doctor, psych., nurse, social worker) / other voluntary associations / randomly (I saw the information desk and entered…)/ nd

Information requested by the user

In the section **Information requested by the user**, data relating to the research topics covered by the request for information are entered. It is possible to mark more fields because the user can ask for more information.

There are four sections: A) information about the disease and the diagnostic-therapeutic pathway; B) Information on rights; C) Information about quality of life; and D) Request for educatioal/informative material

A) Information on the pathology and / or diagnostic-therapeutic process

Pathology in general / Standard treatment/ Side effects / Medical terms/ Complementary Treatment* / Fatigue / Drugs /Prevention/ Clinical Trial / Pain therapy / Follow-up/Rehabilitation /Palliative care /Diagnosis/Prognosis / Risk factors / Genetic factors / Voluntary Associations /Oncologic structures /Body / office of the information point/Instrumenat diagnostics Centers /Accomodations/Patronage/Rehabilitation Facilities/House Assistance/ …/Other Specify___

*The item ***Complementary treatment*** includes the set of different systems, practices and therapeutic products not covered by conventional medicine: acupuncture, phytotherapy, Steiner anthroposophic medicine, homeopathy etc.

B) Information on patient rights

allowed working leave/ disability/ accompainement/ Reimbursements / Handicap / Ticket exemption / Allowance / Part-time/ Insurances/ fiscal facilitations/ job location change / …/ Other specify

C) Information related to Quality of Life

Nutrition / Hair loss/ Lifestyle and wellbeing (physical act, cigarettes, etc.) / Sexuality/ How to communicate with the patient and the family /Anti-cancer treatments and infertility/ Techniques to preserve fertility / pregnancy after the pathology/ Tumor in pregnancy / PMA /Individual support / Support in group/ Public service psychologists /Other specify_____

D) If informative material has been provided, indicate the number of the booklets

Material provided: provide a specific entry for the AIMAC booklets (1 – 34)

Other publications provided

Dvd chemo / Dvd radio/ DST Profiles/The PSA test / Father after cancer /Neoplasia and weight loss – what to do?/ Material of other associations / Pharmacological profiles/Occupational cancer prevention /When a child is sick /Breast reconstruction /Cervix cancer

Data from COVID-related Questionnaire

Just after the beginning of the lock-down in taly, AIMAC put online a survey to assess the impact of the Covid19 pandemic on the needs of cancer patients. After asking for patient consent to store the data, the following questions are done:

Gender Male/ Female

In which region do you live right now? Abruzzo /Basilicata/ Calabria/ Campania/ Emilia-Romagna/ Friuli-Venezia Giulia/ …/ Veneto

Indicate your age 0-20 years / 21-40 years / 41-50 years / 51-60 years / >=61

Where do you live? Center of a large city / Suburb of a large city / Small or medium inland center / Small or medium coastal center / IIsolated house

Indicate your qualification No qualification/ Elementary level /Lower school level /High school level /Bachelor/Master

Do you currently have a partner? Y/N

Do you have children No/Yes, at least 1 child under 10 years old/Yes, all over 10 years old

Who are you living with at home these days? (multiple answers are allowed)

Alone/with your partner/ with child (s)/ with parent (s)/ with other relatives or cohabitants/ with a caregiver/ Other_____

Right now

Do you have an adequate internet connection? Yes /I don't know / No

Do you have adequate digital devices (pc, tablet, smartphone, etc.)? Yes/ I don't know/ No

Do you have enough digital skills? Yes / I don't know / No

Now, you are predominantly:

Unemployed/ Student/ Part-time worker /Full-time worker/ Retired/Housewife or Household

What is your job position?

Manager, entrepreneur, freelancer/ Employee, self-employed/Worker, home worker, cooperative member/ Other

In these days I don't work, because I'm sick or on leave / I am in layoff/ Work from home (agile or smart working)/ I normally go to my workplace/ Work both at home and in the workplace / I was fired / I resigned/ Other_____

Can you indicate the site of your tumor? Lung / Kidney /Breast / Uterus / Prostate/ Leukemia / Colorectal/ Pancreas/ Melanoma/ Thyroid/ lymphoma/ Liver/ Bladder / Other: _____

Can you indicate the therapeutic phase of your illness?

First therapeutic phase after the first diagnosis / Local tumor recurrence treatment (recurrence after response to therapy, followed by free interval from illness) / Treatment tumor progression (progression = aggravation / extension of the disease, without there being a complete response to therapy) / Remote metastasis treatment /Pain therapy / Other _____

How much free time do you have these days?

More than usual / Much more than usual /Less than usual /As before

How do you spend your free time these days? (Multiple answers are possible)

I listen and read the news / I pray or attend online religious services / I practice sport at home / I surf the web to pass the time / I do not do anything /I speak on the phone or in a video call with friends and relatives / I read, cook or dedicate myself to other hobbies

What are the emotions / moods you feel most because of this situation? Rate the intensity from 1 (minimum) to 7 (maximum)

Anger/ irritation/ Fear / Anxiety / Sadness / Happiness /Relaxation / Indifference

Thinking at this moment, how much do you agree with the following statements:

(very much in agreement, In agreement, Very much in disagreement, In disagreement)

I am agitated / I tend to avoid conflicts / I am good at solving the problems that arise in relation to the disease/ I feel little monitored and accompanied by my oncologists /I feel poorly monitored and accompanied by my general practitioner /I am afraid of my health /I suffer the situation /I organize the time of my day /This situation also has positive aspects /People are more collaborative with each other

In this period you are satisfied with the communications you have / have had with your care team (oncologist / oncological dh)

very much/quite/ little/ no

Thinking about managing your disease right now how much you agree with the following statements: (very much in agreement, In agreement, Very much in disagreement, In disagreement)

It would help me communicate via email with my medical team /It would help me to have a dedicated hotline / It would help me to be able to communicate by chat (whatsapp, etc.) / It would help me to be able to receive a remote cancer video consultation if necessary /It would help me to be able to share side effects, clinical data, etc. through apps.

Have you been contacted by the hospital that is treating you?

yes often / yes occasionally / no

You have been contacted to? (Multiple answers are possible)

Change an appointment /Submit reports /Request any oncological problems/ Check your general state of health at the moment (presence of flu symptoms, etc.)/ Other_____

How are you communicating / managing any side effects of the therapy? (Multiple answers are possible) I asked my family doctor /I'm writing them down in a notebook waiting to hear from my oncologist/ I asked the pharmacist / attending physician for suggestions on how to manage them/ I went to the emergency room or the hospital where I am being treated/ I asked help to patients associations /I searched information on the internet/ Other_____

What could help you to follow drug therapy correctly? (Multiple answers are possible)

An app on a device (Smartphone, Tablet) that reminds me to take medicines at the right time / An app that summarizes all the drugs taken daily/ An alarm /The collaboration of a family member or my caregiver / Other_____

How useful would an app for the recognition of vital signs (pressure, fever, etc.) would have been useful in this moment of forced home permanence? (One reply only)

very much / quite / little / no

How are you managing your fears and anxieties? (Multiple answers are possible)

Phone a friend/ I contact a patient in my same condition/ Remote psychological support/ I ask help to patients associations/ I downloaded a dedicated app/ I write on forums, face-book groups or similar/ Other_____

Have you felt the need for nutritional advice / support? (One reply only)

yes, often / yes, occasionally / no

How much would it help you to take care of your nutrition right now (multiple answers are possible) very much /quite / little / no

have an app where you can keep your own nutrition diary / get advice through an app from your medical team / being able to monitor changes in weight, lean mass and fat mass through an app / have a chat to ask the medical team questions

Are you exercising? (One reply only)

every day / 3 days a week /occasionally /never

What helps you exercise regularly at this time of home closure? (One reply only) very much /quite / little /no

a tutorial on youtube / a group I meet online (zoom, meet, skype, etc.) / a dedicated app / I do some exercises independently

Do you feel that your autonomy skills in managing the disease have increased in these days? (One reply only) very much /quite/ little / no

Due to this experience, do you feel more comfortable in managing daily problems related to the disease? (One reply only) very much /quite/ little / no

The file has been imported in R and MATLAB and some pre-processing has been done in order to prepare the data for the analyses. For example, a pre-processing is necessary for all the multiple choice questions, because of the format used to store them, which does not allow a simple data analysis, as shown in Figure 9. Each item of this type has been transformed in *n* binary items where *n* is the number of possible answers.

BJ
paure_ansie
["1"]
["2"]
["1","6"]
["3"]
["99"]
["1"]
["4","6"]
["1","3"]
["1","2","3"]

Figure 9. Multiple answers are stored in the same column. This format must be parsed to allow specific data analysis

3.3.2. Data Extraction and Anonymization Process

Forum Data

Forum data were provided at Jan 28th 2020. As described above, AIMAC provided the data to UNIPV, as a set of tables that allowed it to reconstruct the conversations without containing any reference to the IP addresses.

These data are anonymous at the source. No personal data is stored that could allow us to identify the patient. The users are invited to use a unique nickname for all the posted texts. Usernames of forum's users are publicly available, so they were left like that. The forum has a moderator, who is in charge of removing any non-allowed information, including personal data that could identify users, their relatives or doctors.

Regular Questionnaire Data

AIMAC provided UNIPV with two CSV files, one with data collected from 2012 to 2016 and another one with data collected from 2016 to 2020. The reason is that the database for the data collection has changed over time. The two data schema are similar but not identical, thus a data pre-processing has been necessary in order to harmonise data from the two sources and obtain a unique database for the subsequent analyses, also managing the missing values due to some attributes that are not present in both data sources. Files have been imported in R and MATLAB.

A total of 74.726 records, each one corresponding to an interview, are available.

The first file "DB_SION_AIMAC_2012-2016" is structured in this way:

- for single-answer questions, the columns are organized into double, datetime and categorical data.
- for some questions to which it is possible to provide more than one answer, there are several associated columns, as many as there are possible answers. The idea of building this file is based on the fact that, if a maximum of n answers can be selected for question A, there will be n columns in the file. However, the latter do not have any unequivocal correspondence with the answer, i.e. the i-th column (A_i) does not necessarily correspond to the answer given for the i-th option. In other words, in the i-th column we can find any code from 1 to the maximum number of options available for the answer.

The second file "DB_SION_AIMAC_2016-2020" is structured in this way:

- for single-answer questions, the columns are organized into double, datetime and categorical data (in the same way as in the previous file).
- To manage multiple answers to certain questions in the questionnaire, the file has only one column for each question to contain the answers, inserted in string arrays.

Due to the different setting of the files it was necessary to create two different reading functions that would create a homogeneous structure in Matlab.

It was decided to treat the multiple-answer questions (A, B, C, D, info_requests) (which can also coincide with all the boxes present) with a vector of elements equal to the number of maximum possible answers (i.e. all the boxes in this case); in this vector, each element in the i-th position corresponds to the i-th box: it assumes the value 0 if that box has not been filled in and 1 if it has been

About anonymisation, questionnaires are anonymous at the source, i.e., no personal data are collected that could lead to the patient's identification.

Covid-19 questionnaire data

The database extraction at May27th 2020 contains 493 filled-in questionnaires. The file has been provided by AIMAC to UNIPV as a csv file and has been imported in R and MATLAB. The file "covid19_270520" is structured in this way:

- for single-answer questions, the columns are organized into double, datetime and categorical data (in the same way as in the previous questionnaire);
- to manage the multiple answers of the questionnaire, the file has only one column for each question which contains the answers inserted in a string array.

To read the data we used a Matlab tool ("import data") that allows to view the file by opening it through an editor and deciding on any changes to the structure, transcoding data and other import criteria. A code was then automatically generated to import the file, which can be used directly for subsequent imports, which therefore will no longer require the aforementioned tool. We found only one problem: as regards multiple answers, in some cases some information was not read adequately. More precisely, when the first row contained only one compiled option, the whole column was interpreted as "numeric" and therefore only the first number that appeared in the array was imported. To solve this, we slightly modify the code created automatically by Matlab by changing the data type for some variables from numeric to categorical.

Questionnaires are anonymous at the source, i.e., no personal data are collected that could lead to the patient's identification.

Quality check of imported data

For all the imported questionnaire data sets, we looked for any errors by checking if there were any values outside the allowable range. To do this we followed this logic: since all the answer data are categorical, we extracted all the data collected as an answer to a question and the unique elements were taken from this vector (each different element repeated at most once). Then we intersected this vector with the valid values and we checked its length with the length of the vector of the unique elements: if the two lengths are equal then it means that the process of intersection with the valid values has obtained a positive value for each line.

For the questionnaire compilation date, we checked for the allowed interval 2012-2020.

3.3.3. Longitudinal Description

Forum Data

The aim of the analysis of forum data is to detect, through the patients' messages, their needs and their expectations. The information extracted will be used to inform the CAPABLE system developers about some important app functionalities, i.e., those that help patients and their caregivers to face the daily routine against cancer. Since the messages are anonymous, we cannot reconstruct longitudinal patterns. However, since messages come with a date, we could detect possible changes in the patients' needs over time.

Questionnaires Data

Concerning the regular questionnaires, users can have more than one interview, so different records could be from the same user, but this information is not available because questionnaires are completely anonymous at the source, i.e., no personal data are collected that could lead to the patient's identification.

Concerning the Covid-19 questionnaire, a user is supposed to fill-in the questionnaire only once.

4. Glossary

IMDC	International Metastatic Database Consortium
AIMAC	Italian Association for Cancer patients, relatives and friends (Associazione Italiana Malati di Cancro, parenti ed amici)
ICSM	Istituti Clinici Scientifici Maugeri
RCC	Renal Cell Carcinoma
NKI	Netherlands Cancer Institute
AVL	Antoni van Leeuwenhoekziekenhuis
DMTR	Dutch Melanoma Treatment Registry
PDPC	Personal Data Protection Commission
REDCap	Research Electronic Data Capture

5. References

[1] Guide to Basic Data Anonymisation Techniques – PDPC. https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Other-Guides/Guide-to-Anonymisation_v1-(250118).pdf

[2] The phpbb forum software: https://www.phpbb.com/

[3] Python programming language: https://www.python.org/

[4] Pandas Python Data Analysis Library: https://pandas.pydata.org/

[5] NumPy The fundamental package for scientific computing with Python. https://numpy.org/

[6] Matplotlib: Visualization with Python: https://matplotlib.org

[7] TreeTagger - a part-of-speech tagger for many languages. https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/

6. Annexes

6.1. Annex 1 - ICSM Data Dictionary

This annex presents a Data Dictionary for ICSM data set extracted from REDCap data management tool.

	#	Variable / Field Name	Field Label *Field Note*	Field Attributes (Field Type, Validation, Choices, Calculations, etc.)
Instrument: **Demographic** (demographic)				Collapse
	1	record_id	Record ID	text
	2	id	ID	text
	3	gender	Gender	radio 1, M 2, F
	4	race	Race	dropdown 1, White 2, Black
	5	dob	Date of birth	text (date_dmy)
	6	height	Height *cm*	text (number)
	7	status	Status	radio 1, Alive 2, Dead
	8	death_date Show the field ONLY if: [status] = '2'	Death date	text (date_dmy)
	9	last_fu_date Show the field ONLY if: [status] = '1'	Last Follow Up date	text (date_dmy)
	10	demographic_complete	Section Header: *Form Status* Complete?	dropdown 0, Incomplete 1, Unverified 2, Complete
Instrument: **Tumor Characteristics** (tumor_characteristics)				Collapse
	11	ddx	Date of initial diagnosis	text (date_dmy)
	12	dmet	Date of diagnosis of metastatic disease	text (date_dmy)
	13	nephrectomy	Nephrectomy	yesno 1, Yes 0, No
	14	nephrectomy_type Show the field ONLY if: [nephrectomy] = '1'	Nephrectomy Type	radio 1, Partial 2, Radical
	15	nephrectomy_site Show the field ONLY if: [nephrectomy] = '1'	Nephrectomy Site	radio 1, Right Kidney 2, Left Kidney 3, Both Kdney
	16	nephrectomy_date Show the field ONLY if: [nephrectomy] = '1'	Date of nephrectomy	text (date_dmy)
	17	biopsy_for_diagnosis	Biopsy for diagnosis	yesno 1, Yes 0, No

	18	biopsy_date Show the field ONLY if: [biopsy_for_diagnosis] = '1'	Biopsy date	text (date_dmy)
	19	biopsy_site Show the field ONLY if: [biopsy_for_diagnosis] = '1'	Biopsy site	dropdown 1 Right Kidney 2 Left Kidney 3 Both Kidney 4 Lung 5 Liver 6 Pancreas 7 Bone 8 Soft tissue 9 Lymph nodes 10 Brain 11 Other
	20	biopsy_site_specify Show the field ONLY if: [biopsy_site] = '11'	Biopsy site specify	text
	21	surgery_for_diagnosis	Surgery for diagnosis	yesno 1 Yes 0 No
	22	surgery_date Show the field ONLY if: [surgery_for_diagnosis] = '1'	Surgery date	text (date_dmy)
	23	surgery_site Show the field ONLY if: [surgery_for_diagnosis] = '1'	Surgery site	dropdown 1 Right Kidney 2 Left Kidney 3 Both Kidney 4 Lung 5 Liver 6 Pancreas 7 Bone 8 Soft tissue 9 Lymph nodes 10 Brain 12 Peritoneum 11 Other
	24	surgery_site_specify Show the field ONLY if: [surgery_site] = '11'	Surgery site specify	text
	25	t_stage	T stage	dropdown 1 1 2 2 3 3 4 4 5 1a 6 1b 7 2a 8 2b 9 3a 10 3b 11 3c

	26	n_stage	N stage	dropdown 0, 0 1, 1 2, 2 3, 3 4, x
	27	m_stage	M stage	dropdown 0, 0 1, 1 4, x
	28	tumor_size	Tumor size cm	text (number)
	29	necrosis	Necrosis	yesno 1, Yes 0, No
	30	mvi	Microvascular Invasion (MVI)	yesno 1, Yes 0, No
	31	fuhrman_grade	Fuhrman Grade	dropdown 1, 1 2, 2 3, 3 4, 4
	32	non_clear_cell	Non clear Cell	dropdown 0, ccRCC 1, non ccRCC
	33	ccrcc	ccRCC	yesno 1, Yes 0, No
	34	paprcc	papRCC	yesno 1, Yes 0, No
	35	chrrcc	chrRCC	yesno 1, Yes 0, No
	36	unclassified_rcc	unclassified RCC	yesno 1, Yes 0, No
	37	collect_duct	collect Duct	yesno 1, Yes 0, No
	38	xp_translocation	XP Translocation	yesno 1, Yes 0, No
	39	sarcomatoid	Sarcomatoid	yesno 1, Yes 0, No
	40	tumor_characteristics_complete	Section Header: *Form Status* Complete?	dropdown 0, Incomplete 1, Unverified 2, Complete
Instrument: **Therapy Start** (therapy_start)				Collapse

	41	previous_treatment	Previous Treatment *with Interferons or Interleukina*	yesno 1 Yes 0 No
	42	line_therapy_start_date	Line Therapy Start Date	text (date_dmy)
	43	line_therapy_drug	Line therapy Drug	text
	44	line_therapy_drug_1	Line therapy Drug Classified 1	dropdown 1 Atezolizumab 2 Axitinib 3 Bevacizumab 4 Cabozantinib 5 Capecitabina 6 Carboplatino 7 Cisplatino 8 Epirubicina 9 Everolimus 10 Farmaci sperimentali 11 Fluorouracile 12 Gemcitabina 13 Ifosfamide 14 Interferone alfa 15 Interleukina 2 16 Ipilimumab 17 Irinotecan 18 Linfociti citotossici 19 Metotrexate 20 Nivolumab 21 Oxaliplatino 22 Paclitaxel 23 Pazopanib 24 Sorafenib 25 Sunitinib 26 Temsirolimus 27 Tivozanib 28 Vimblastina 29 Vinblastina 30 Vinorelbina

	45	line_therapy_drug_2	Line therapy Drug Classified 2	dropdown

1	Atezolizumab
2	Axitinib
3	Bevacizumab
4	Cabozantinib
5	Capecitabina
6	Carboplatino
7	Cisplatino
8	Epirubicina
9	Everolimus
10	Farmaci sperimentali
11	Fluorouracile
12	Gemcitabina
13	Ifosfamide
14	Interferone alfa
15	Interleukina 2
16	Ipilimumab
17	Irinotecan
18	Linfociti citotossici
19	Metotrexate
20	Nivolumab
21	Oxaliplatino
22	Paclitaxel
23	Pazopanib
24	Sorafenib
25	Sunitinb
26	Temsirolimus
27	Tivozanib
28	Vimblastina
29	Vinblastina
30	Vinorelbina

	46	line_therapy_drug_3	Line therapy Drug Classified 3	dropdown 1 Atezolizumab 2 Axitinib 3 Bevacizumab 4 Cabozantinib 5 Capecitabina 6 Carboplatino 7 Cisplatino 8 Epirubicina 9 Everolimus 10 Farmaci sperimentali 11 Fluorouracile 12 Gemcitabina 13 Ifosfamide 14 Interferone alfa 15 Interleukina 2 16 Ipilimumab 17 Irinotecan 18 Linfociti citotossici 19 Metotrexate 20 Nivolumab 21 Oxaliplatino 22 Paclitaxel 23 Pazopanib 24 Sorafenib 25 Sunitinib 26 Temsirolimus 27 Tivozanib 28 Vimblastina 29 Vinblastina 30 Vinorelbina
	47	experimental_drug	Experimental drug	text
	48	weight	Weight *kg*	text (number)
	49	hemoglobin	Hemoglobin *g/dl*	text
	50	hemoglobin_lower_limit	Hemoglobin Lower Limit *g/dl*	dropdown 11.7 11.7 13.2 13.2 12 12
	51	corrca	Serum Corrected Calcium *mg/dl*	text
	52	high_ca	High Ca?	yesno 1 Yes 0 No
	53	ldh	LDH *mU/ml*	text (number)
	54	ldh_upper_limit	LDH Upper Limit *mU/ml*	dropdown 220 220
	55	neutrophil_absolute_count	Neutrophil absolute count *10^9/L*	text (number)
	56	lymphocyte_absolute_count	Lymphocyte absolute count *10^9/L*	text (number)
	57	platelet_count	Platelet count *10^9/L*	text (number)
	58	na_serum_sodium	Serum Sodium (Na) *mmol/l*	text (number)

	59	karnofsky_performace_kps	Karnofsky performace %	text (number, Min: 0, Max: 100)
	60	creatinine	Creatinine	text (number)
	61	therapy_start_complete	Section Header: *Form Status* Complete?	dropdown 0, Incomplete 1, Unverified 2, Complete
Instrument: **Therapy Stop** (therapy_stop)				Collapse
	62	line_therapy_stop_date	Line Therapy Stop Date	text (date_dmy)
	63	best_response	Best response	dropdown PD, Progressive Disease (PD) SD, Stable Disease (SD) PR, Partial Response (PR) CR, Complete Response (CR)
	64	max_change_shrinkage	Max change (% shrinkage) %	text (number)
	65	dose_reduction	Dose Reduction?	yesno 1, Yes 0, No
	66	dose_reduction_date Show the field ONLY if: [dose_reduction] = '1'	Dose Reduction Date	text (date_dmy)
	67	final_dose_level Show the field ONLY if: [dose_reduction] = '1'	Final Dose Level	text (number)
	68	toxicity_leading_dose_red Show the field ONLY if: [dose_reduction] = '1'	Toxicity Leading to Dose Reduction	text
	69	therapy_stop_reason	Reason for Therapy STOPPING	dropdown P, Progression T, Toxicity PT, Progression or Toxicity D, Death O, Other CR, Complete Response
	70	reason_for_stopping_specif Show the field ONLY if: [therapy_stop_reason] = 'O'	Reason for stopping - Specify	text
	71	toxicity_leading_stop_therapy Show the field ONLY if: [therapy_stop_reason] = 'PT' or [therapy_stop_reason] = 'T'	Toxicity Leading to Stop Therapy	text

	80	brain_mets_present	Brain mets present?	yesno 1 Yes 0 No
	81	brain_mets_date	Brain mets date	text (date_dmy)
	82	brain_mets_number	Brain mets number *99 if not numerable*	text (integer, Min: 1, Max: 99)
	83	size_of_largest_mets	Size of largest mets *cm*	text (number)
	84	cerebral_met_supratentoria	Cerebral met (Supratentorial)	yesno 1 Yes 0 No
	85	cerebellar_area_infratento	Cerebellar Area (infratentorial)	yesno 1 Yes 0 No
	86	symptoms_at_presentation	Symptoms at presentation	yesno 1 Yes 0 No
	87	xrt_for_brain_mets	XRT for brain mets *even if asymptomatic*	yesno 1 Yes 0 No
	88	xrt_date Show the field ONLY if: [xrt_for_brain_mets] = '1'	XRT date	text (date_dmy)
	89	xrt_date_2 Show the field ONLY if: [xrt_for_brain_mets] = '1'	XRT date 2	text (date_dmy)
	90	stereotactic_radiosurgery	Stereotactic Radiosurgery (SRS)?	yesno 1 Yes 0 No
	91	srs_date Show the field ONLY if: [stereotactic_radiosurgery] = '1'	SRS date	text (date_dmy)
	92	srs_date_2 Show the field ONLY if: [stereotactic_radiosurgery] = '1'	SRS date 2	text (date_dmy)
	93	neurosurgery_brain_met	Neurosurgery for brain mets?	yesno 1 Yes 0 No
	94	neurosurgery_date Show the field ONLY if: [neurosurgery_brain_met] = '1'	Neurosurgery date	text (date_dmy)
	95	best_response_of_tki	Best response of TKI	dropdown 1 Progressive Disease (PD) 2 Stable Disease (SD) 3 Partial Response (PR) 4 Complete Response (CR)
	96	shrinkage_brain_metastases	Shrinkage of brain metastases %	text (number)
	97	comment_of_brain_lesions	Comment of brain lesions	notes
	98	brain_mets_complete	Section Header: *Form Status* Complete?	dropdown 0 Incomplete 1 Unverified 2 Complete

6.2. Annex 2 - NKI Data Dictionary

Variable	Description (English)
age	Categories of 10 years
sex	Sex
datovl	Date of death
ptloc	Location primary tumor
typmel	Type of melanoma
breslow	Breslow thickness in millimeters (unknown = 999)
ulcer	Ulceration
dermit	Dermal mitoses
sattel	Satellitosis / in-transits
kliermet	Lymph node metastases (macroscopic)
afstmet	Distance metastases
locmetprim1	(Sub)cutan [C44]
locmetprim2	Lymph nodes [C77]
locmetprim3	Lungs [C34]
locmetprim4	Liver [C22.0]
locmetprim5	Brain [C70, C71, C72, C75.1-3]
locmetprim6	Gastro-intestinal [C15-20]
locmetprim7	Bone [C40, C41]
locmetprim8	Otherwise [C76]
behadj	Has the patient been treated with (neo)adjuvant therapy? If so, which one?
behadjand	Other (neo)adjuvant systemic therapy, namely:
stageadj	Stage in adjuvant treatment:
behadjdat	Start date of (neo)adjuvant treatment:
episode	Episode
who	WHO score
comorb	Comorbidity?
comorbcar	Cardial
comorbvas	Vascular
comorbdia	Diabetes
comorbpul	Pulmonal

Variable	Description (English)
comorbneu	Neurological/Psychiatric
comorbmda	Gastrointestinal/liver
comorburo	Urogenital
comorbtro	Thrombotic
comorbmus	Muscle and joints
comorbend	Endocrine disorders
comorbinf	Infectious diseases
comorbmal	Malignancy
comorbauto	Autoimmune
comorbove	Other comorbiditeit
comcar00	Angina Pectoris
comcar01	Myocard infarct
comcar02	PTCA
comcar03	CABG
comcar04	Valve suffering
comcar05	Valve replacement
comcar06	Atriumfibrilleren / -flutter
comcar07	Arrhythmias
comcar08	Decompensatio cordis
comcar09	Cardiomyopathy
comcar10	Post-heart transplant status
comcarxx	Other
comvas1	Hypertension
comvas2	Peripheral vascular suffering
comvas3	Carotis stenosis
comvas4	Aneurysm Aorta
comvasx	Other
comdia1	Not insulin dependent
comdia2	Insulin dependent
comdia3	Diabetes with organ failure
comdiax	Other
compul1	COPD / CARA / emphysema / chron.bronch.

Variable	Description (English)
compul2	Pulmonary fibrosis
compul3	status after lung resection/transplant
compulx	Other
comneu6	Schizophrenia/severe depression/psychosis
comneu1	TIA
comneu2	CVA
comneu3	Muscle diseases
comneu4	Hemiplegia / Spinal cord injury
comneu5	Parkinsonism / Dementia
comneux	Other
commda1	Ulcer suffering or reflux ophages
commda2	Symptomatic gallstone suffering
commda3	Pancreatitis (acute or chronic)
commda4	IBD
commda5	Diverticulitis
commda6	Liver disease/failure (Cirrhosis/hepatitis)
commdax	Other
comuro1	Chron. renal insufficiency (Creat>110)
comuro2	Dialysis dependent kidney failure
comuro3	Post-kidney surgery/transplant status
comuro4	Pregnant when surgery.
comurox	Other
comtro1	DVT
comtro2	Pulmonary embolism
comtro3	Reduced clotting
comtrox	Other
commus2	Rheumatoide disorders / SLE / sclerodermy
commus1	Sarcoidose / Besnier Boeck
commus3	Vasculitis
commusx	Other
comend1	Hypo- / hyperthyreoidism
comend2	Hypo- / hyperparathyreoidism

Variable	Description (English)
comend3	Adrenal gland disease (Addison,Cushing,Conn)
comendx	Other
cominf1	HIV / AIDS
cominf2	Tuberculosis
cominf3	Malaria
cominfx	Other
commalsolide	Solid malignancy
commalhemato	Hematological malignancy
commalbehandel	Treatment malignities?
commus22	Rheumatoide disorders/SLE/sclerodermy
graves	Morbus Graves
commda42	IBD
medic	Medication use
immumod1	Corticosteroids
immumod2	Imuran (Azathioprin)
immumod3	Interferon
immumodand	Other
stage	Staging of metastases?
epistage	What is the tumor stage before the start of the (neo)adjuvant treatment
stagect	Staging CT thorax and/or abdomen
stagepet	Staging PET-CT
stageher	Staging MRI of CT van de Brain
labs100	Lab determination: S100
labs100w	Value S100 (ug/l)
labldh	Lab determination: LDH
labldhw	LDH Value (U/L)
lokrec	Location(s) current tumor presentation
locrelym	Have the metastases of the current tumor presentation been proven?
loclymet1	Neck
loclymet2	Axilla
loclymet3	Inginual superficial
loclymet4	Inginual deep

Variable	Description (English)
loclymet5	Popliteal
loclymet6	Other
locreitm	In-transit metastases
locafmetLungs	Lungs [C34]
locafmetLiver	Liver [C22.0]
locafmethersen	Brain [C70, C71, C72, C75.1-3]
locafmetdarm	Gastrointestinal [C15-20]
locafmetBone	Bone [C40, C41]
locafmetlklier	Lymphe nodes [C77]
locafmetcutis	Cutis/subcutis [C49]
locafmetand	Other [C76]
hersenmet	Brainmetastases
totnmet	Total number of metastases
hispritum	Histology primary tumor investigated
revhispri	Revision of histology primary tumor
hisrevaan	Histology and/or cytology current tumor presentation present
genmutrecbraf	Gene mutation BRAF known/determined
genmutrecnras	Gene mutation NRAS known/determined
genmutreckit	Gene mutation KIT known/determined
genmutrecgnaq	Gene mutation GNAQ known/determined
genmutrecgna11	Gene mutation GNA11 known/determined
genmutrecander	Other genes known/determined
genmutrecandvrij	What other gene mutations are known/determined?
genmutsang	Sanger sequencing
genmutnext	Next generation sequencing
genmutsequ	Sequenom analysis
genmutpcr	Real-time PCR
genmutcobas	Cobas-BRAF test
genmutmelting	High resolution Melting
genmutand	Other technology
typbraf0n	BRAF mutation
typbraf1	c.1799T>A (p.(Val600Glu))

Variable	Description (English)
typbraf2	c.1798_1799delinsAA (p.(Val600Lys))
typbraf3	c.1798_1799delinsAG (p.(Val600Arg))
typbraf4	c.1799_1800delinsAT (p.(Val600Asp))
typbraf5	c.1799_1800delinsAA (p.(Val600Glu)), (=E2 variant)
typbraf6	c.1781A>G (p.(Asp594Gly)
typbraf7	c.1794_1796dup (p.(Thr599dup))
typbraf8	c.1795_1797dup (p.(Thr599dup))
typbraf9	c.1799_1802delinsAAAT (p.(Val600_Lys601delinsGluIle))
typbraf10	Other
typbrafand	BRAF mutation that is different than previously mentioned choices, namely:
typnras0n	NRAS mutation?
typnras1	c.181C>A (p.(Gln61Lys))
typnras2	c.182A>G (p.(Gln61Arg))
typnras3	c.182A>T (p.(Gln61Leu))
typnras4	c.183A>T (p.(Gln61His))
typnras5	c.180_181delinsTA (p.(Gln61Lys))
typnras6	c.34G>T (p.(Gly12Cys))
typnras7	c.35G>A (p.(Gly12Asp))
typnras8	c.35G>C (p.(Gly12Ala))
typnras9	c.37G>T (p.(Gly13Cys)
typnras10	c.37G>A (p.(Gly13Ser))
typnras11	c.38G>A (p.(Gly13Asp))
typnras12	c.44G>A (p.(Gly15Glu))
typnras13	Other
typnrasand	NRAS mutation that is different than previously mentioned choices, namely:
typkit0n	KIT mutation
typkit1	c.1671G>C (p.(Trp557Cys))
typkit2	c.1672A>G (p.(Lys558Glu))
typkit3	c.1676T>A (p.(Val559Asp))
typkit4	c.1679T>A (p.(Val560Asp))
typkit5	c.1727T>C (p.(Leu576Pro))

Variable	Description (English)
typkit6	c.1922T>A (p.(Leu641His))
typkit7	c.1924A>G (p.Lys642Glu))
typkit8	c.2591C>T (p.(Ser864Phe))
typkit9	Other
typkitand	KIT mutation that Is different than previously mentioned choices, namely:
typgnaq0n	GNAQ mutation
typgnaq1	c.626A>T (p.(Gln209Leu))
typgnaq2	c.548G>A (p.(Arg183Gln))
typgnaq3	Other
typgnaqand	GNAQ mutation that Is different than previously mentioned choices, namely:
typgnamut0n	GNA11 mutation
typgnamut1	c.626A>T (p.(Gln209Leu))
typgnamut2	c.547C>T (p.(Arg183Gln))
typgnamut3	Other
typgnamutand	GN11 mutation that is different than previously mentioned choices, namely:
sequen	Sequence
redstopseq	Reason for declining (systemic) treatment
typbeh1	Surgery
typbeh2rt	Radiotherapy
typbeh2ht	Hyperthermia
typbeh3	RFA / microwave
typbeh4	Systemic therapy
typbehand	Other
sysadj	Purpose of the surgery and systemic therapy combination ?
chirurgie	Surgery
tarchir1	Skin / subcutis
tarchir2	Lymph node station(s)
tarchir3	Soft tissue / Bone
tarchir4	Lung / thorax
tarchir5	Liver
tarchir6	Stomach / pancreas / spleen

Variable	Description (English)
tarchir7	Colon
tarchir8	Brain
tarchir9	Other
tarchirand	Targethirurgie other than previously mentioned choices, namely:
transit	Have in-transit metastases been removed?
transverw	How many in-transit metastases have been removed?
transbase	How many in-transit metastases did the patient initially have?
lymfstverw	How many lymph nodes have been removed?
lymfproc	What procedure has been done to remove the lymph node stations?
lymfklverw	Hoeveel lymphe nodes zijn er in totaal verwijderd?
lymfklpos	How many of these removed lymphe nodes were considered tumorpositive?
lymfkldia	What is the diameter (mm) of the largest tumor metastases in the removed lymph node(s)?
lymfklrad	What was the radicality of the resection?
typchir	Type and type of surgery, namely:
compchir	Has there been a surgical complication?
radiotherapie	Radiotherapy
tarradtype	Type of radiotherapy
tarradtypeand	Other types of radiotherapy
tarrt1	Skin / subcutis
tarrt1frac	Number of fractions of radiotherapy (skin/subcutis) (unknown = 99)
tarrt1dos	Dose of radiotherapy per fraction in Gy (skin/subcutis) (unknown = 999.9)
tarrt2	Lymph node station(s)
tarrt2frac	Number of fractions of radiotherapy (lymphe node station(s)) (unknown = 99)
tarrt2dos	Dose of radiotherapy per fraction in Gy (lymphe node station(s)) (unknown = 999.9)
tarrt3	Soft tissue / bone
tarrt3frac	Number of fractions of radiotherapy (Soft tissue / bone) (unknown = 99)
tarrt3dos	Dose of radiotherapy per fraction in Gy (Soft tissue / bone) (unknown = 999.9)
tarrt4	Lungs(en) / thorax
tarrt4frac	Number of fractions of radiotherapy (Lung/thorax) (unknown = 99)

Variable	Description (English)
tarrt4dos	Dose of radiotherapy per fraction in Gy (Lung/thorax) (unknown = 999.9)
tarrt5	Liver
tarrt5frac	Number of fractions of radiotherapy (Liver) (unknown = 99)
tarrt5dos	Dose of radiotherapy per fraction in Gy (Liver) (unknown = 999.9)
tarrt6	Intra-abdominal (Other dan liver)
tarrt6frac	Number of fractions of radiotherapy (intra-abdominal) (unknown = 99)
tarrt6dos	Dose of radiotherapy per fraction in Gy (intra-abdominal) (unknown = 999.9)
tarrt7	Brain
tarrt7frac	Number of fractions of radiotherapy (Brain) (unknown = 99)
tarrt7dos	Dose of radiotherapy per fraction in Gy (Brain) (unknown = 999.9)
tarrt8	Other
tarrt8frac	Number of fractions of radiotherapy (Other) (unknown = 99)
tarrt8dos	Dose of radiotherapy per fraction in Gy (Other) (unknown = 999.9)
tarrt9	Unknown
hyperthermie	Hyperthermia
tarht1	Skin / subcutis
tarht2	lymph node station(s)
tarht3	Soft tissue / bone
tarht4	Lung / thorax
tarht5	Liver
tarht6	Intra-abdominal (Other dan Liver)
tarht7	Brain
tarht8	Other
tarht9	unknown
nhyp	Number of times hyperthermia (unknown = 99)
rfa	RFA
tarrfa1	Soft tissue / bone
tarrfa2	Lung / thorax
tarrfa3	Liver
tarrfa4	Intra-abdominal (Other dan Liver)
tarrfa5	Other

Variable	Description (English)
tarrfaand	Other, namelijk:
klinsyst	Kliniek systemische therapie
typsyth1	Type of systemic therapy
typchem	Type of chemotherapy
andchem	Other type of chemotherapy, out of study, namely:
doschem	Dose of chemotherapy (unknown = 9999)
nkurchem	Number of chemotherapy administrations (unknown = 99)
dosredchem	Dose reduction
reddoschem	Reduced dose
nkurred	Number of reduced administrations (unknown = 99)
redreduchem	Reason dose reduction
toxgraad	Toxicity chemotherapy
typtox1	Bone marrow suppression
typtox2	Severe infection/sepsis
typtox3	Other
typtoxand	Type of toxicity: Other
gevtox2	short-term drug use
gevtox3	long-term drug use
gevtox4	day treatment without hospitalization
gevtox5	Hospitalization
gevtox6	IC admission
gevtox7	Operation
gevtox8	permanent damage
gevtox9	Death
gevtox9com	Explain the cause of death here:
naambraf	Name BRAF inhibitor
naambrafand	Other name BRAF inhibitor, namely:
dosbraf	Dose of BRAF inhibitor (unknown = 9999)
braf-do-sisaanpassing	Braf dose adjustment
redredubraf1	Reason dose adjustment
reddosbraf1	Adjusted dose of BRAF inhibitor

Variable	Description (English)
naammek	Name MEK inhibitor
naammekand	Other name MEK inhibitor, namely:
dosmek	Dose of MEK inhibitor (unknown = 99.9)
mek-do-sisaanpassing	MEK dose adjustment
redredumek1	Reason dose adjustment
reddosmek1	Adjusted dose of MEK inhibitor (unknown = 99.9)
redstopbraf	Reason stop treatment
toxbraf	Toxicity BRAF inhibitor
toxbrafhuid	Skin malignancy
toxbrafphoto	Photosensitivity
toxbrafvoet	Hand-foot syndrome
toxbrafhuidtox	Skin toxicity
toxbrafpyr	Pyrexia
toxbrafLiver	Liver failure
toxbrafartral	Artralgies
toxbrafalopecia	Complete baldness, (ir)reversible
toxmeklivent	Decrease left ventricle ejection fraction
toxmekvisus	Visus changes as consequence of retinopathy/occlusive retinavene
toxbrafOther	Other
toxhuidm	Toxicity of skin malignancy
typtoxbrafand	Type of toxicity: Other, namely:
gevtoxb2	short-term drug use
gevtoxb3	long-term drug use
gevtoxb4	day treatment without hospitalization
gevtoxb5	Hospitalization
gevtoxb6	IC admission
gevtoxb7	Operation
gevtoxb8	permanent damage
gevtoxb9	Death
gevtoxb9com	Explain the cause of death here:
dosipi	Dose of ipilimumab (unknown = 999)

Variable	Description (English)
startipi	Start date ipilimumab
nkuripi	Number of ipilimumab administrations (unknown = 9)
redstopipi	Reason stop treatment
toxipi	Toxicity ipilimumab
typtoxipileuc	Leucopenia, trombopenia, anaemia
typtoxipineuro	Neuropathy
typtoxipicolitis	Colitis
typtoxipidarm	Intestinal perforation
typtoxipihuid	Skin toxicity
typtoxipiuveit	Uveitis
typtoxipibijnier	Adrenal insufficiency
typtoxipihypofyse	Hypopituitarism
typtoxipischild	Thyroid insufficiency
typtoxipiLiver	Hepatitis/Liverfailure
typtoxipiOther	Other
typtoxiand	Type of toxicity: Other, namely:
gevtoxi2	short-term drug use
gevtoxi3	long-term drug use
gevtoxi4	day treatment without hospitalization
gevtoxi5	Hospitalization
gevtoxi6	IC admission
gevtoxi7	Operation
gevtoxi8	permanent damage
gevtoxi9	Death
gevtoxi9com	Explain the cause of death here:
medtoxcortipi	Coricosteroids
medtoximipi	Other immune modulating agents
testpd	Has a PD-L1 test been carried out?
testpdvrij	What test has been done?
testpdper	Percentage positive (unknown = 99)
startpd	Start date anti PD 1 antibodies
naampd	Name anti PD 1 antibody

Variable	Description (English)
naampdand	Other name anti-PD 1 antibody, namely:
dospd	Dose of anti-PD 1 antibody (unknown = 9999)
nkurpd	Number of administrations anti-PD 1 antibody (unknown = 999)
anti-pd-1-do-sisaanpassing	Anti-PD-1 dose adjustment
redredupd	Reason postponement of the administration
redstoppd	Reason stop treatment
toxpd	Toxicity anti-PD 1 antibody
typtoxpdleuc	Aplastic anemia (leucopenia, trombopenia, anemia)
typtoxpdneuro	Neuropathy
typtoxpdcolitis	Colitis
typtoxpddarm	Diarrhea
typtoxpdnierf	Decline in renal function
typtoxpdnefr	Nefritis
typtoxpddys	Dyspnoe
typtoxpdpneumi	Pneumonitis
typtoxpdbijnier	Adrenal insufficiency
typtoxpdhypofyse	Hypopituitarism
typtoxpdschild	Thyroid insufficiency
typtoxpdmoe	Fatigue
typtoxpdrash	Rash
typtoxpdpru	Pruritis
typtoxpdvit	Vitiligo
typtoxpdhepa	Hepatitis
typtoxpdOther	Other
typtoxpdand	Type of toxicity: Other, namely:
gevtoxpd2	short-term drug use
gevtoxpd3	long-term drug use
gevtoxpd4	day treatment without hospitalization
gevtoxpd5	Hospitalization
gevtoxpd6	IC admission
gevtoxpd7	Operation

Variable	Description (English)
gevtoxpd8	permanent damage
gevtoxpd9	Death
gevtoxpd9com	Explain the cause of death here:
medtoxcortpd	Coricosteroids
medtoxtnfpd	TNF alpha blocker
medtoximipd	Other immune modulating agents
medtoximipdand	Other, namely:
ipnitestpd	Has a PD-L1 test been carried out?
ipnitestpdvrij	What test has been done?
ipnitestpdper	Percentage positive (unknown = 99)
startipnicomb	Startdatum combinatiefase met ipilimumab en nivolumab
startipniond	Startdatum nivolumab als onderhoudsbehandeling
dosipnicombi	Dose of ipilimumab (unknown = 9999)
nkuripnicombi	Number of administrations ipilimumab (unknown = 999)
dosipnicombn	Dose of nivolumab (unknown = 9999)
nkuripnicombn	Number of admissions nivolumab (unknown = 999)
ipnicomb-uitstel	Ipi nivo dose adjustment
reduitipnicomb	Reason to postpone the combination phase
dosipniond	Dose of nivolumab (unknown = 9999)
nkuripniond	Number of administrations nivolumab (unknown = 999)
ipniond-uitstel	Postponement maintenance treatment with nivolumab
reduitipniond	Reason to delay maintenance treatment
redstopipni	Reason stop treatment (during combination phase or during maintenance treatment)
toxipni	Toxicity ipilimumab and nivolumab
typtoxipnileuc	Aplastic anemia (leucopenia, trombopenia, anemia)
typtoxipnineuro	Neuropathy
typtoxipnicolitis	Colitis
typtoxipnidarm	Diarrhea
typtoxipnihepa	Hepatitis
typtoxipninefr	Nefritis
typtoxipnipneumi	Pneumonitis

Variable	Description (English)
typtoxipnibijnier	Adrenal insufficiency
typtoxipnihypofyse	Hypopituitarism
typtoxipnischild	Thyroid insufficiency
typtoxipnimoe	Fatigue
typtoxipnirash	Rash
typtoxipnipru	Pruritis
typtoxipnivit	Vitiligo
typtoxipniOther	Other
typtoxipni	Type toxiciteit: Other, namelijk:
gevtoxipni2	short-term drug use
gevtoxipni3	long-term drug use
gevtoxipni4	day treatment without hospitalization
gevtoxipni5	Hospitalization
gevtoxipni6	IC admission
gevtoxipni7	Operation
gevtoxipni8	permanent damage
gevtoxipni9	Death
medtoxcortipni	Explain the cause of death here:
medtoxalphaipni	TNF alpha blocker
medtoximiipni	Other immune modulating agents
typandst	Type of other systemic therapy, namely;
toxandst	Toxcity other systemic therapy
typtoxandst	Type of toxicity other systemic therapy, namely
gevtoxast2	short-term drug use
gevtoxast3	long-term drug use
gevtoxast4	day treatment without hospitalization
gevtoxast5	Hospitalization
gevtoxast6	IC admission
gevtoxast7	Operation
gevtoxast8	permanent damage
gevtoxast9	Death
gevtoxast9com	Explain the cause of death here:

Variable	Description (English)
opname	Hospital admission
redopna	Reason admission
icdg	Number of days in intensive care (unknown = 99)
chirurgie-ii	Other surgery
tarchir1ov	Skin / subcutis
tarchir2ov	Lymph node station(s)
tarchir3ov	Soft tissue / bone
tarchir4ov	Lung / thorax
tarchir5ov	Liver
tarchir6ov	Stomach / pancreas / spleen
tarchir7ov	Colon
tarchir8ov	Brain
tarchir9ov	Other
tarchirandov	Targetsurgery other than previously mentioned choices, namely:
typchirov	Type of surgery, namely:
compchirov	Has there been a surgical complication?
fup	Follow up
datlcont	Date last contact
behepi	Treatment episode
pedkg	Weight in kg:
pedcm	Length in cm:
pedtannerp	Tanner's stage for pubic hair (P stadium):
pedtannerm	Tanner's stage for breast development (M stage):
pedtannerg	Tanner's stage for genitalia (G stage):
statlcontpal	Has palliative treatment been performed during this follow-up period of the current treatment episode?
statlcont	Status with last contact
doodoorz	What was the cause of death
oorovand	Cause death, other namely: