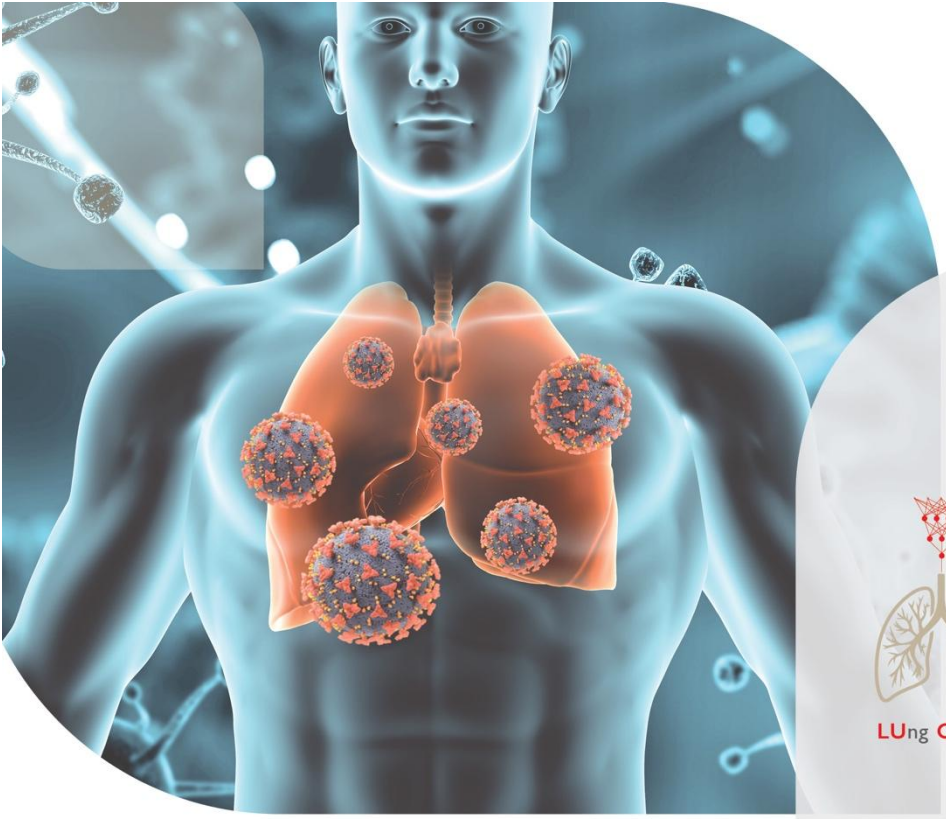

LUCIA
LUng Cancer-related risk factors and their Impact Assessment

LUCIA Understanding Lung Cancer related risk factors and their Impact

Horizon Europe Grant Agreement Number: 101096473

Deliverable Number	D4.9
Deliverable Title	Mid-period report on LUCIA: validation and evaluation
Due date of deliverable	31st May 2025
Actual Submission Date	11th June 2025
Responsible partner	Biobizkaia Health Research Institute (BB)
Contributors	Clinical Partners (SAS, BB, OSA, CHUL, LU), Technological Partners (TECH, NAN, PRO, VICOM, CNAG) and YAG
Revision (draft, 1, 2, …)	1.0
Dissemination Level	PU - Public

Start Date of the project: January 1, 2023
Duration: 48 months

Document information

Status	Revision date	Authors
Draft version 1.0	09/05/2025	Jon Eneko Idoyaga Uribarrena
Draft review	25/05/2025	Eunate Arana Arri
Internal review	29/05/2025	Clara Frick
Final version	30/05/2025	Jon Eneko Idoyaga Uribarrena

Executive Board Document Sign Off

Role	Partner	Signature	Date
Project coordinator	TECH	Yoav Broza	11/06/2025
WP1 Lead	ULSTER	Jonathan Wallace	10/06/2025
WP2 Lead	VICOM	Alba Garin-Muga	30/05/2025
WP3 Lead	TECH	Yoav Broza	03/06/2025
WP4 Lead	BB	Jon Eneko Idoyaga Uribarrena	30/05/2025
WP5 Lead	UHEI	Jonathan Sleeman	30/05/2025
WP6 Lead	HOPE	Marie Nabbe	02/06/2025
WP7 Lead	TECH	Liat Tsuri	03/06/2025

Contents

Executive summary:

This deliverable has been conceived in the frame of T4.3 “General population screening”, T4.4 “Diagnosis and precision follow-up and stratification” and T4.5 “Contextual-empirical investigations to evaluate the realization of identified values”. These tasks are devoted to the recruitment of a prospective cohort of around 3,000 volunteers that will be followed up for 2 years and on the diagnosis, including Indeterminate Pulmonary Nodules (IPN) characterization, with an accentuation on the Never Smokers and Reduced Smokers (NSRS) patients, incorporating Breath Analyzer (BAN), Wide-biomarker-spectrum Multi-Use Sensing Patch (WBSP) and spectrometry-on-card (SPOC) into clinical studies.

Regarding Task 4.5, it examines whether identified socio-technical values (in WP1-3) (e.g., transparency, bias, accountability, explainability) are realized when the technology is used. To achieve this goal, the different contexts of the use of technology are to be analyzed as different contextual variables come into play to impact the way values are understood.

Volunteers have been recruited from different clinical centers (“Servicio Andaluz de Salud” (SAS) and “Osakidetza Servicio Vasco de Salud” (OSA), in Spain; “Centre Hospitalier Universitaire de Liège” (CHUL), in Belgium and “Centre for Tuberculosis and Lung Diseases (CTLD) of Riga East University Hospital (REUH)”, in Latvia). Non-invasive devices such as Breath Analyzer (BAN), Multiomics (MO) and spectrometry-on-card (SPOC) are monitoring these participants.

The entire study cohort is currently being followed up. For 2 years, participants of the study will attend to 4 visits: baseline, month 6, month 12 and month 24. During these visits, the following tests and procedures will be carried out:

- **Baseline visit**: blood test, spirometry, lifestyle questionnaires, sociodemographic data, medical record data, exposure to harmful agents data, physical exploration, LDCT scan and new lung cancer screening devices testing (breath analyzer and spectrometry-on-card)
- **6 months visit:** remote visit where sociodemographic data, medical record data and exposure to harmful agents data will be recorded
- **12 months visit**: spirometry, lifestyle questionnaires, sociodemographic data, medical record data, exposure to harmful agents’ data and physical exploration.
- **24 months visit**: spirometry, lifestyle questionnaires, sociodemographic data, medical record data, exposure to harmful agents data, physical exploration, low-dose computed tomography (LDCT) scan and new lung cancer screening devices testing (breath analyzer and spectrometry-on-card)

Only the Basque Country Clinical Site (Osakidetza Basque Health Service) is performing the LDCT scan in this phase of the study. During visit 1, based on the assessment carried out by the results of the LDCT (only in Osakidetza), and by established risk prediction models (Lung Cancer Risk Assessment Tool (LCRAT) and Liverpool Lung Project (LLP) risk model), subjects will follow one of the following pathways (figure 1):

- **Continuation of Phase 1**: Wide population screening if low to moderate risk of lung cancer is identified.
- **Referral to Phase 2:** Precision screening if high risk of lung cancer is assigned. Those participants referred to phase 2 will undergo the same visit scheme as the ones that remain in phase 1. The only difference is that another lung cancer screening non-invasive device will be added in visits baseline and 24 months: Wide-biomarker-spectrum Multi-Use Sensing Patch (WBSP). In this phase, all clinical sites will perform a LDCT scan.
- **Referral to Phase 3**: Diagnosis if by results of LDCT lung cancer or Indeterminate Pulmonary Nodules (IPN) are found. These participants will undergo the usual treatment for their diagnosis as per usual clinical practice until the end of the study. The non-invasive lung cancer screening devices (breath analyzer, spectrometry-on-card and Wide-biomarker-spectrum Multi-Use Sensing Patch [WBSP]) will also be tested.

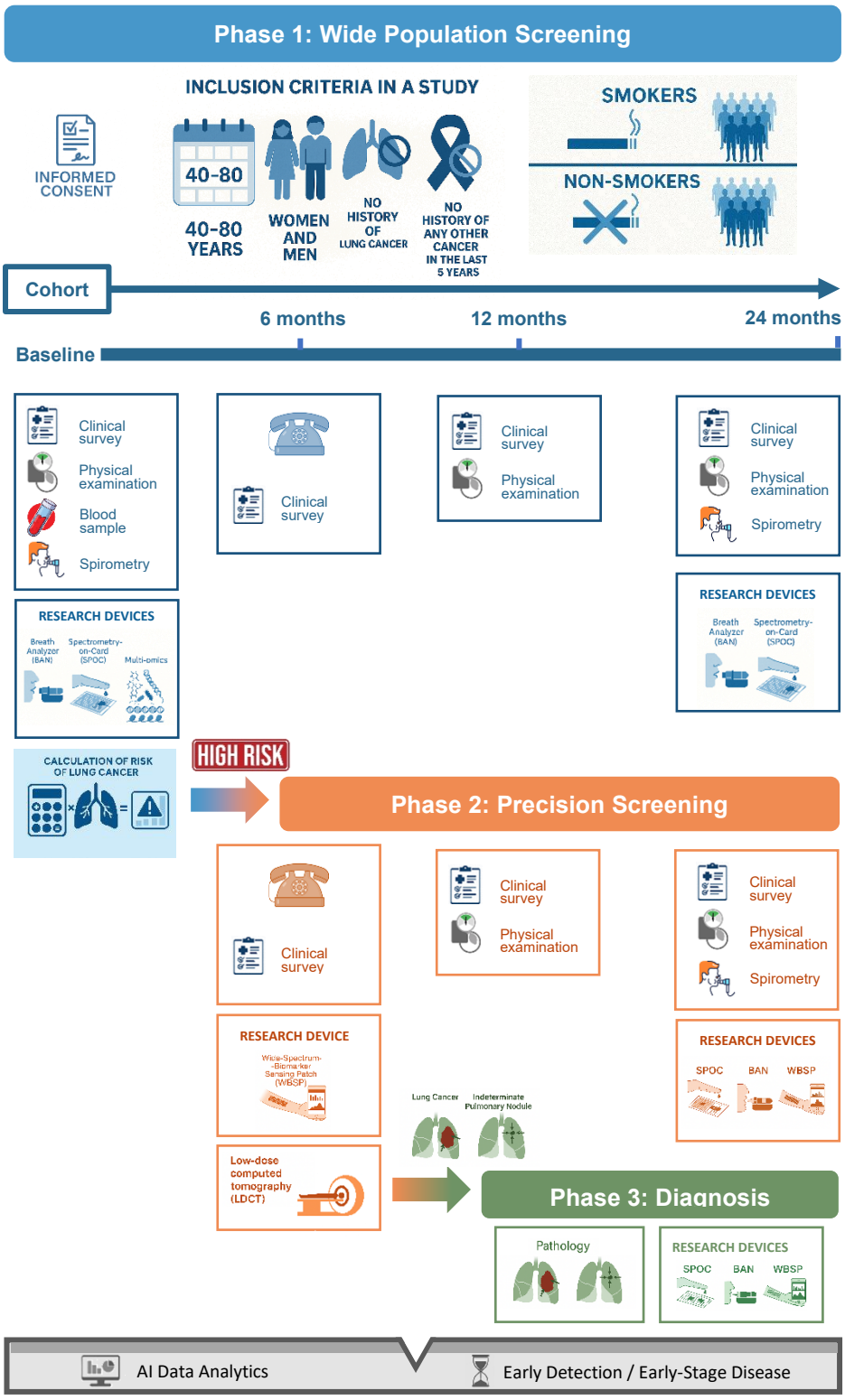

Figure 1: *Flow diagram of the LUCIA prospective study*

From the wide population screening phase (or Phase 1), we estimate that 1,000 high-risk of LUCIA volunteers will be detected by the LCRAT and LLP risk prediction models and be referred to the precision screening phase (or Phase 2). We estimate that around 400 LC or pulmonary nodule cases will arise, considering the prevalence of these pathologies all over Europe. These patients will be followed with the diagnosis of pulmonary nodules or LC and be referred to the diagnosis phase (or Phase 3). Nevertheless, as in the current situation, we have not found as many LC patients or pulmonary nodules as was estimated, we are also recruiting:

- Patients with new diagnosis of pulmonary nodules or lung cancer, prior initiation of treatment, outside the screening phases from pneumology consultations of the clinical partners participating in the prospective study (LU, OSA, CHUL, SAS).
- Patients with lung cancer or pulmonary nodules followed-up in specialized consultations of pneumology or oncology departments.

This deliverable aims to collect information on the recruitment situation at the mid-period of the clinical prospective study and the validation and evaluation of the technologies used in the study. The report includes:

- A report on the status of the prospective clinical study
- A report on deviations and issues during recruitment
- A report on the status of the development of the LUCIA risk prediction model
- A report on the validation and evaluation of the LUCIA technologies (Non-invasive devices such as Breath Analyzer (BAN), Multiomics (MO), spectrometry-on-card (SPOC) and Wide-biomarker-spectrum Multi-Use Sensing Patch (WBSP)
- A report with the planning, execution and results of Task 4.5 “Contextual-empirical investigations to evaluate the realization of identified values”
- A detailed description of implemented and planned measures to compensate for any incurred delays

Status of the Prospective clinical study

In this section, we provide an overview of the number of recruited participants per clinical site. We also include a detailed description of the study population's characteristics regarding sex, age and exposure to tobacco. A great effort has been done by all clinical centers in order to achieve a heterogeneous sample that reflects the diversity of the European population.

The recruitment in the four clinical sites ended on April, 4th 2025 to ensure the follow up of participants included in the study and allow sufficient time to analyze all data.

The recruitment sites included in the LUCIA prospective clinical study are listed below:

- **Spain**:
  - Andalusia:
    - 3 Healthcare Centers of the Aljarafe Sevilla Norte Health District (Carmona, Gerena and los Carteros)
    - Virgen de la Macarena University Hospital
  - Basque Country:
    - Cruces University Hospital
    - Basurto University Hospital
    - San Eloy Hospital
    - Urduliz Hospital
    - Galdakao-Usansolo Hospital
    - Around 40 Primary Health Care Centers: Balmaseda PCU and Mamaria PCU
- **Belgium:**
  - Centre Hospitalier Universitaire de Liège
- **Latvia:**
  - Latvian Oncology Centre
  - Tuberculosis and Lung Disease Centre

The numbers shown in the following pages reflect the recruitment situation and study population characteristics of each clinical site on April 4, 2025:

- **Andalusia:**

  The Andalusian clinical site is located in the Sevilla province and it is composed by 3 primary healthcare centers (Carmona, Gerena and los Carteros) and the Virgen de la Macarena University Hospital. The Andalusian Healthcare Service (Servicio Andaluz de Salud, SAS by its acronym in Spanish) manages this clinical site. The situation in this site is as follows:

Table 1: Recruitment situation in Andalusia

NUMBER OF VOLUNTEERS RECRUITED	694
Number of volunteers included in Phase 1	293
Number of volunteers included in Phase 2	117 (some participants pending risk calculation)
Number of CTs performed on the Phase 2 cohort	15
Number of volunteers included in Phase 3	0
Number of volunteers with BAN performed	612
Number of volunteers in Phase 2/3 with the patch performed	15
Number of volunteers with the SPOC performed	0
Number of samples sent to the CNAG and number of samples with OK quality control at the CNAG	213

SAS

Participants	Number of participants	694	
		Number	%
Sex	Male	297	57
	Female	397	43
Age	40-50 (1974-1984)	187	27
	51-64 (1960-1973)	331	48
	65-80 (1944-1959)	176	25
Relationship with tobacco	Ex smoker	124	17.86
	Smoker	126	18.15
	Never smoker	89	12.82
	Missing Info	355	51.15

	Female		Male	
Relationship with tobacco	N.	%	N.	%
Exsmoker	55	7.9	69	9.9
Smoker	71	10.2	55	7.9
Never smoker	63	9.0	26	3.7
Missing info	208	29.9	147	21.1

	Female		Male	
Age distribution	N.	%	N.	%
40-50 (1974-1984)	116	16,7	71	10,2
51-64 (1960-1973)	198	28,5	133	19,1
65-80 (1944-1959)	83	11,9	93	13,4

The Andalusian clinical site is currently in the process of including the data in the eCRF and calculating the risk for the population they have recruited. Risk calculation is still pending for some participants, including 355 participants whose smoking history still needs to be incorporated. It's an ongoing process that will be soon completed.

After risk calculation, those participants who are classified as high risk and referred to phase 2, will undergo a LDCT scan. 15 LDCT scans have already been performed and so far no lung cancer or IPN cases have been found.

Regarding the devices, 612 BAN tests have been performed and 15 participants have already used the WBSP. No SPOC tests have been performed yet.

- **Basque Country:**
  In the Basque clinical site, 5 Hospitals (Cruces University Hospital, Basurto University Hospital, San Eloy Hospital, Urduliz Hospital and Galdakao-Usansolo Hospital) and around 40 Primary Health Care Centers are involved in the recruitment. The situation in this site is as follows:

Table 2: Recruitment situation in the Basque Country

NUMBER OF VOLUNTEERS RECRUITED	1,417
Number of volunteers included in Phase 1	1,149
Number of volunteers included in Phase 2	193
Number of CTs performed on the Phase 2 cohort	160*
Number of volunteers included in Phase 3	75
Number of LDCT performed	1,117
Number of volunteers with BAN performed	1,417
Number of volunteers in Phase 2/3 with the patch performed	142
Number of volunteers with the SPOC performed	35
Number of samples sent to the CNAG and number of samples with OK quality control at the CNAG	400

*There are still 300 LDCT to performed, so the distribution of Phases could change

Data from lesions found				
Lung-RADS® 1	**Lung-RADS® 2**	**Lung-RADS® 3**	**Lung-RADS® 4**	**Lung Cancer**
67	253	27	5	4

OSA Osakidetza

Participants — Number of participants: **1,417**

		Number	%
Sex	Male	577	40.71
	Female	840	59.29
Age	40-50 (1974-1984)	535	37.75
	51-64 (1960-1973)	630	44.46
	65-80 (1944-1959)	252	17.78
Relationship with tobacco	Ex smoker	669	47.21
	Smoker	375	26.46
	Never smoker	373	26.32

	Female		Male	
Relationship with tobacco	N.	%	N.	%
Exsmoker	383	27.02	286	20.18
Smoker	216	15.24	159	11.22
Never smoker	241	17.01	132	9.31

	Female		Male	
Age distribution	N.	%	N.	%
40-50 (1974-1984)	315	22.23	220	15.52
51-64 (1960-1973)	389	27.45	241	17.01
65-80 (1944-1959)	136	9.59	116	8.18

In the Basque Country Clinical Site, additionally, **1,117** participants have already completed the Low Dose CT (LDCT) scan as the whole population undergoes a CT scan. The rest of the Basque Country's participants already have an appointment for a LDCT in the upcoming month.

All the participants have completed the BAN test and samples for multiomics analysis have been collected.

142 phase 2 and 3 participants have already used the WBSP and around 400 samples have been sent to the CNAG for its analysis.

Regarding screening encouragement in hard-to-reach communities, around 10% of the currently recruited participants consist of migrants and individuals with low socioeconomic status.

The recruitment of never smokers and reduce smokers is under the percentages needed to assure the diversity of the cohort; as well as the distribution by sex.

In the 3rd plenary meeting, which took place in Seville, the consortium reached the agreement to follow-up the nodules detected in the LDCT scans performed during LUCIA recruitment:

- **Lung-RADS® 1** → PHASE 1 → RISK ESTIMATION
- **Lung-RADS® 2:**
  - **Solid nodule** → PHASE 1 → RISK ESTIMATION
  - **Part solid nodule or GGO** – CT 6M / 12M
    ↳ No change / growth → **PHASE 3**
- **Lung-RADS® 3 or 4 → PHASE 3**

- **Belgium**:
  The Centre Hospitalier Universitaire de Liège is the designated Belgian clinical site. The situation in this site is as follows:

Table 3: Recruitment situation in Belgium

NUMBER OF VOLUNTEERS RECRUITED	507
Number of volunteers included in Phase 1	383
Number of volunteers included in Phase 2	96
Number of CTs performed on the Phase 2 cohort	85
Number of volunteers included in Phase 3	28 (25 directly included in phase 3 and 3 coming from phase 1)
Number of volunteers with BAN performed	146
Number of volunteers in Phase 2/3 with the patch performed	0
Number of volunteers with the SPOC performed	200
Number of samples sent to the CNAG and number of samples with OK quality control at the CNAG	0 (40 samples sent but sequencing data was bad. Currently investigating a contingency plan together with CNAG)

CHUL

Participants — Number of participants: **507**

		Number	%
Sex	Male	243	47.8
	Female	265	52.2
Age	40-50 (1974-1984)	63	12.4
	51-64 (1960-1973)	184	36.22
	65-80 (1944-1959)	260	51.18
Relationship with tobacco	Ex smoker	248	48.8
	Smoker	97	19.1
	Never smoker	143	28.1
	Missing Info	20	3.9

	Female		Male	
Relationship with tobacco	N.	%	N.	%
Exsmoker	109	41.1	139	57.2
Smoker	54	20.4	43	17.7
Never smoker	90	34	53	21.8
Missing Info	12	4.5	8	3.3

	Female		Male	
Age distribution	N.	%	N.	%
40-50 (1974-1984)	30	11.32	33	13.58
51-64 (1960-1973)	106	40	78	32.1
65-80 (1944-1959)	128	48.3	132	54.32

In Belgium, participants who initially enrolled in Phase 1 have transitioned to Phases 2 or 3, while others were directly included into Phases 2 or 3. As a result, the total number of participants in each phase (Phases 1, 2, and 3) is not yet final. Additionally, risk calculations are pending for some of the participants. This is partly due to incomplete information on participants' smoking habits, caused by missing or unclear responses. The Belgian clinical site is actively working to retrieve this information through reminders and follow-up calls to patients, as well as through routine data collection during Phase 2 and 3 visits.

- **Latvia:**
  The Latvian clinical site is located at the Riga East University Hospital. In this are included: Hospital "Gaiļezers" 8 Departments: Pneumonology - 1), Hospital "Oncology Centre of Latvia" (LA), Hospital "Biķernieki" (+ outpatient department of CTLD), Hospital "Latvian Centre of Infectious Diseases" and Hospital "Centre of Tuberculosis and Lung Diseases" (RI) (Departments: Pneumonology - 6 & Thoracic surgery – 1). The situation in this site is as follows:

Table 4: Recruitment situation in Latvia

NUMBER OF VOLUNTEERS RECRUITED	237
Number of volunteers included in Phase 1	168
Number of volunteers included in Phase 2	22
Number of CTs performed on the Phase 2 cohort	21
Number of volunteers included in Phase 3	47
Number of volunteers with BAN performed	168
Number of volunteers in Phase 2/3 with the patch performed	12
Number of volunteers with the SPOC performed	40
Number of samples sent to the CNAG and number of samples with OK quality control at the CNAG	0 (only test samples have been sent to the CNAG)

LU INSTITUTE OF CLINICAL AND PREVENTIVE MEDICINE

Patients	Number of patients	237	
		Number	%
Sex	Male	102	43.04
	Female	135	56.96
Age	40-50 (1974-1984)	62	26.16
	51-64 (1960-1973)	102	43.04
	65-80 (1944-1959)	73	30.80
Relationship with tobacco	Ex smoker	88	37.13
	Smoker	52	21.94
	Never smoker	97	40.93

	Female		Male	
Relationship with tobacco	N.	%	N.	%
Exsmoker	42	31.11	46	45.10
Smoker	17	12.59	35	34.31
Never smoker	76	56.30	21	20.59

	Female		Male	
Age distribution	N.	%	N.	%
40-50 (1974-1984)	41	30.37	21	20.59
51-64 (1960-1973)	58	42.96	44	43.14
65-80 (1944-1959)	36	26.67	37	36.27

Latvian clinical site has recruited a total number of 237 participants, which are distributed as follows:

- 168 Phase 1 participants
- 22 Phase 2 (high risk) participants
- 27 Phase 3 participants

168 participants have undergone the BAN test, while the number of patches and SPOC tests is an ongoing process.

Regarding the number of DNA samples, only test samples have been sent to the CNAG.

Deviations and issues during recruitment

At month 28 (April 2025) a total number of **2,855 participants** have been included in Phase 1 (Wide Population Screening) of the study and have already performed baseline visits. As explained previously, the aim is to recruit at least 3,000 volunteers. Given that volunteers with lung cancer or indeterminate lung nodules will continue to be recruited in the coming months from the oncology and pulmonology departments, there is no risk of not reaching the number of volunteers needed to achieve the LUCIA project's objectives.

The four clinical sites are focused on gathering all data and including it in the eCRF for its subsequent analyses.

As data are being entered into the eCRF, risk prediction models (LCRAT and LLP) are applied to the data and utilized to calculate the risk of lung cancer for Phase 1 participants. Based on these models:

- Participants are referred to Phase 2 of the study (Precision Screening): If the score of the models are equal or higher than **3%** (for the **LLP model**) and/or **1.7%** (for the **LCRAT model**). If with either of these models a participant is estimated to have a high risk of suffering from lung cancer, he/she will be included in Phase 2.
- Participants remain in Phase 1 (Wide Population Screening): If the score of the models reflects a low risk of suffering from lung cancer.

Currently, we have included:

- **2,277** Phase 1 participants
- **428** Phase 2 participants
- **150** Phase 3 participants

Nevertheless, these figures are not complete, as the Andalusian clinical site is still awaiting risk calculations for some participants, and the Belgian clinical site is missing smoking habit information for several participants. As a result, both sites are currently unable to calculate the risk scores for some volunteers.

Currently, the focus is on recruiting volunteers diagnosed with lung cancer and IPN to meet the recruitment needs of each phase. Great efforts are being made to include more Phase 3 participants (diagnosed with nodules or lung cancer) from now on so that we can obtain more data from lung cancer and/or Indeterminate Pulmonary Nodules (IPN) patients. These patients are essential so that we can compare the results of the tests performed on their

breath, blood and skin samples to the ones obtained on healthy participants included in phase 1 of the study and to cover all the objectives of LUCIA project.

Clinical partners are also defining the variables to be included in the eCRF for the characterization of lung cancer subtypes, as lung cancer cases are starting to appear as Phase 3 participants are being included in the study.

Status of the development of the LUCIA risk prediction model

The development and validation of Artificial Intelligence (AI)-based models for predicting the one-year incidence of lung cancer using electronic health record (EHR) data is a core objective of Task 5.2. The development of these models, initially described in Deliverable D5.1, has involved extensive preprocessing of retrospective EHR datasets from three clinical partners: Osakidetza (OSA), Servicio Andaluz de Salud (SAS), and University Hospital of Liège (CHU Liège, CHUL). Preprocessing aimed to ensure data quality, consistency, and compatibility for cross-institutional validation.

While model training and internal testing have been conducted exclusively on OSA's data, the datasets from SAS and CHUL—already preprocessed—are now being employed in the validation phase. A diverse set of AI models has been designed and optimized based on the specific characteristics of the OSA dataset. Their performance has been rigorously evaluated to assess predictive accuracy and clinical utility in identifying individuals at high risk for developing lung cancer within one year.

The external validation using SAS and CHUL data is currently underway, and the feasibility of this process is being assessed in light of observed differences between the available patient cohorts.

Additionally, within T5.2, we have developed an improved and simplified lung cancer (LC) risk prediction model based on Lung Cancer Risk Assessment Tool (LCRAT). While LCRAT leverages comprehensive individual data—including smoking history, demographics, personal and family medical background, and environmental exposures—our model focuses on a reduced set of input features, achieving comparable or superior performance with greater simplicity and ease of integration into clinical workflows. Check the published article for more details: Survival Stacking Ensemble Model for Lung Cancer Risk Prediction - PubMed

For external validation, the developed model can be applied to the LUCIA prospective cohort, provided the number of incident lung cancer cases is sufficient to ensure statistical reliability and the required variables are available (only if there is a match between the features from the model and the prospective available data). This validation will enable further assessment of the model's clinical utility and transportability to real-world healthcare settings.

Validation and evaluation of the LUCIA technologies

LUCIA technologies are non-invasive devices that have been included in the Prospective Clinical Study for their validation and evaluation as Lung Cancer early detection technologies. These technologies are being tested over the volunteers and/or their samples.

The devices/technologies included are: Breath Analyzer (BAN), spectrometry-on-card (SPOC), Wide-biomarker-spectrum Multi-Use Sensing Patch (WBSP) and Multiomics (MO).

The status of these devices at this moment is as follows:

- **Breath Analyzer (BAN)**

  Nanose Medical (NAN) supplied the clinical sites BAN devices and test kits as follows (see table 5):

Table 5: Number of BAN devices supplied

		BB	SAS	LU	CHUL	Total
BAN system		2	3	2	1	8
Test kits		1,550	869	410	178	3,007

Until 11th May 2025, 6,981 breath samples were collected from 2,367 subjects at the 4 clinical sites that participate in the prospective clinical study.

Each sensor array contains multiple sensors, which are measured during the exposure to the breath sample.

- <u>Sensors' data pre-processing</u>

  Nanose (NAN) initiated the analysis of sensor responses data from 642 patients recruited at BB. Of those, 27 were verified lung cancer patients and 615 phase 1 patients (that were labeled as 'control'). At this stage, NAN's goal was to differentiate the verified lung cancer patients from any other group. First, data pre-processing engine was developed in order to get a clean, reliable dataset for machine learning (ML) analysis. NAN continues to improve the engine to

exclude the need for combined manual inspection. All the valid sensor traces are then transformed according to the attempted ML method.

- Machine learning data analysis

Given the relatively small number of breath samples from verified lung cancer at this stage and the potential for patient/breath-level data variability, NAN developed a robust algorithmic pipeline designed to minimize bias and enhance generalizability in classification outcomes. Extensive efforts were made in the development of the feature selection process to reduce dimensionality, a critical step when working with small datasets. At this stage, NAN used ensemble learning that can be generalized relatively well and that is inherently robust to noise and outliers. Permutations approach was then added to further mitigate overfitting risks in complex clinical datasets. In this approach, unique permutations of train and test sets are generated. This aims to ensure that the results are not overly reliant on any single train-test, reducing the risk of overfitting split and allows capturing variability in patient/breath-level data for more generalized performance metrics.

Based on initial analysis we achieved the below results (see table 6):

Table 6: Preliminary accuracy, sensitivity and specificity results

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Train	74.9±1.1	77.6±3.5	74.9±1.6
Test	70.6±9.6	66.1±14.3	75.0±14.2

The relatively high accuracy (~74.9%) of the training set indicates that the model effectively learns patterns within the data. Moreover, although having a strongly imbalanced dataset (i.e. 25 LC vs 613 control) the accuracy gap between test and train set is moderate (~4.5%), while the sensitivity gap is more pronounced (~11.5%) and the specificity gap is minimal (~0.8%). This indicates that the model performs consistently well on the majority class but does not yet generalize optimally to the minority class. Nonetheless, the model demonstrates an encouraging ability to detect meaningful signals within a very small sample group (an inherently challenging task). As this is a preliminary analysis of the dataset, we anticipate improving sensitivity as additional data becomes available and through the application of targeted strategies to better represent the minority class.

The test set included only 12 breath tests (6 from each group: lung cancer and control). This small sample size leads to high variability in test performance metrics. For example, the accuracy of the test set has STD of ±9.6%, indicating fluctuations across different test splits. Sensitivity (66.1±14.3%) and specificity (75.04±14.2%) show similar variability, reflecting differences in how the model identifies true positives and negatives across different test partitions.

The conclusions regarding generalization challenges were drawn by analyzing the variability in the dataset's performance across multiple splits (Figure 2 below). By repeatedly permuting the test set and re-evaluating the model, it became evident that this variability was primarily driven by the limited number of test samples and the inherent diversity in patient signals. The first observation supporting this is that the test set exhibit bell curve characteristics (Fig 2A), providing valuable insights into the model's consistency and variability.

The scatter plot in Figure 2B illustrates the relationship between sensitivity (true positive rate) and specificity (true negative rate) across various test set splits. Each point represents a single test split, with the red marker indicating the mean sensitivity (0.661) and specificity (0.754). The diagonal dotted line represents where sensitivity and specificity metrics are equal. The plot reveals the anticipated variability in sensitivity and specificity across splits, yet it also demonstrates a solid overall balance. This is evident from the clustering of many points near the mean values and the diagonal line. The aggregation of ROC across all iterations in Figure 2C shows that the mean AUC is 0.752 is well above a random split. These results again suggest that the model performs consistently across a wide range of scenarios, despite challenges such as a small test set size and the inherent heterogeneity of patient signals.

These findings are encouraging, indicating that the model is on the right track and suggest that overfitting is not the primary issue. As data collection efforts continue, increasing the dataset size and diversity will likely further enhance the model's performance and consistency.

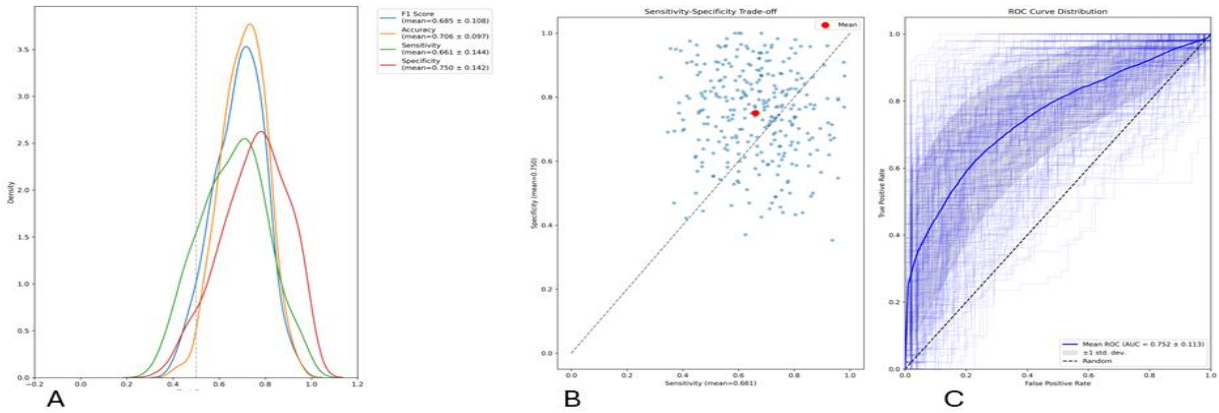

Figure 2: Performance metrics Distribution of (A) test set. (B) Performance Distribution Scatter plot of Sensitivity-Specificity Trade-off. (C) ROC curve distribution.

- **Spectrometry-on-card (SPOC)**

  The first prototype was fully described in previous deliverables. The device was produced and sent to the clinical partners. After a number of initial checks on test samples a problem with the prototype was identified and thus mitigation strategies were taken to secure the continuance of the Volatile Organic Compounds (VOCs) blood sampling.

  For that, a second device that serves as SPOCv2 was adapted and prepared in collaboration within TECH projects. See figures below.

  - <u>Technical specifications of the SPOCv2</u>
    - PC Interface: USB-C connector with USB 2.0 data communication
    - Operating System for PC software: Windows 11, 64bit
    - Data format: JSON
    - Central micro controller: Arm Cortex M4 with 512kB of flash memory
    - Max Power consumption: limited by USB C Power delivery. Either 800mA or 1.5A
    - Dimensions: 119 x 75 x 21.8mm
    - Breath detection: differential pressure sensor
    - Temperature control: 2 heaters with aluminum heat spreaders covering the sensing chamber.
    - Resistance measurement: 6 channel multiplexer per chip with software programmable load resistor per sensor. 2 x 24bit ADC with each 4-channel multiplexer to address all 8 chips.

- Load resistance values: 10k, 100k, 1M, 10M. Load resistance can be set for individually for each of 48 sensors.
- Acquisition time: (default) 1s for all 48 sensors.

- <u>Sensor's fabrication and chip development</u>

Technion has fabricated an electronic chip on a silicon wafer, where each chip contains 6 sensors from the same chemistry (see design in Figure 3). This way the variability is minimized in each chip resulting in decreasing the average variability between the devices.

The basic sensor unit was designed to be very inexpensive and dispensable. In this way, changing the sensor units is very easy, thus, allowing for a fast and convenient change of specific characteristics of the overall device. This principle was also maintained throughout the overall electronic design, ensuring easy replacement of the sensor chips. This feature of the device provides for flexibility and tentative potential modification of the device for future redirection towards different stages of the same disease or even different diseases.

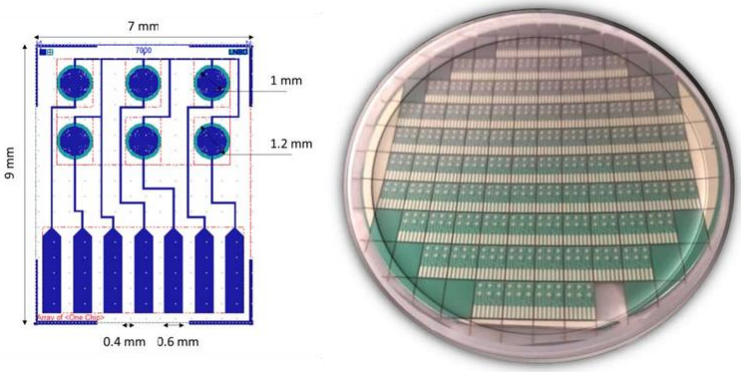

Figure 3: The new chip design contains six sensors (left), Picture of 92 chips on wafer (right)

Eight chips with different chemistries were fabricated. As can be seen in Figure 4, each chemistry had different morphology on the silicon chips.

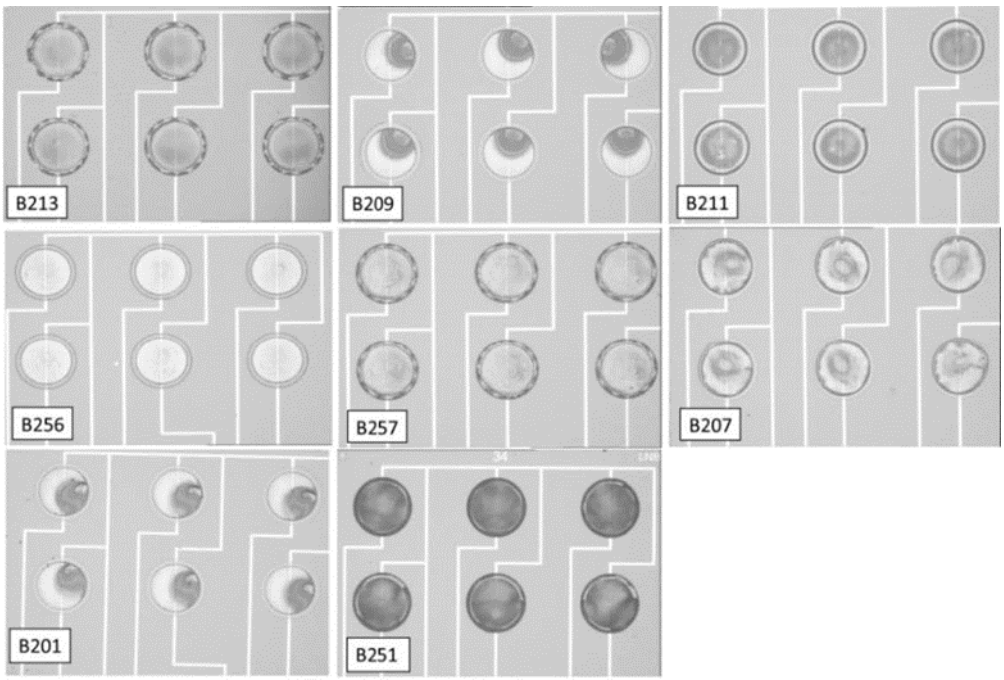

Figure 4: Morphological examination of the different printed sensor chips

- Device analyser performance

The device was sent to clinical sites by Dec 2024. Each device contains eight different chips with different functional groups (Organic molecules bound to the gold nanoparticles). Each chip contains six sensors from the same chemistry. After fabrication and curing of the chips, we measured them to evaluate their response. The measurements were done using a customed measurement chamber. (See Figure 5)

Figure 6 shows the measurement system in the lab. This system was used to test the sensors response before choosing the ones to populate the device. The measurement system contains three small chambers, each holding four chips. This system is connected to a gas generator which can generate VOCs in different known concentrations in N2. During the measurement we first insert pure N2 to determine the baseline response and later we insert different VOCs at escalating concentrations. Pure N2 was inserted between each VOC exposure.

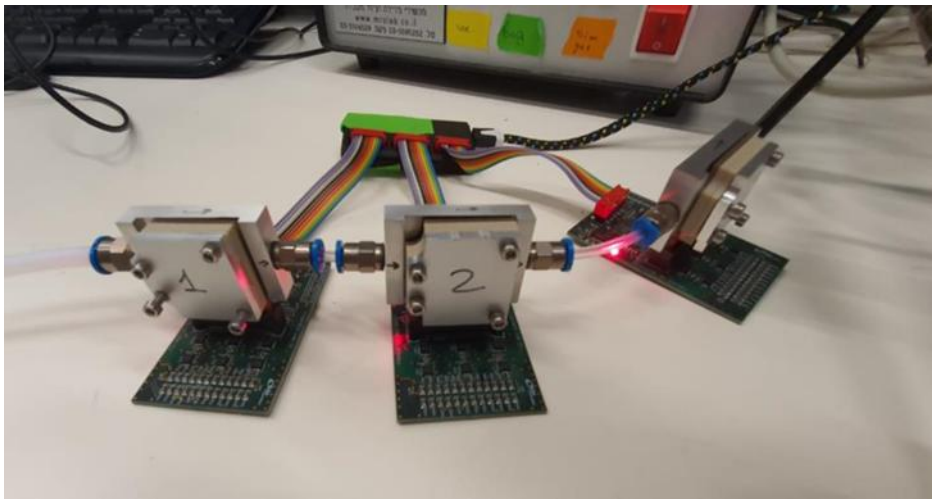

Figure 5: Lab measurement system.

Figure 6(a) represents an example of six chips of B207 (Hexanethiol). As can be seen, the variability index (VI or RSD) of the chips is very low and the average VI is 2.1%. This means that the variability between different sensor replicas of the same chemistry at the same chip, exposed to a single VOC is minimized.

Figure 6(b) depicts the response rate of deferent sensors of the same chemistry (B207) to escalating doses of octane. The linear trend of the sensor's response is evident.

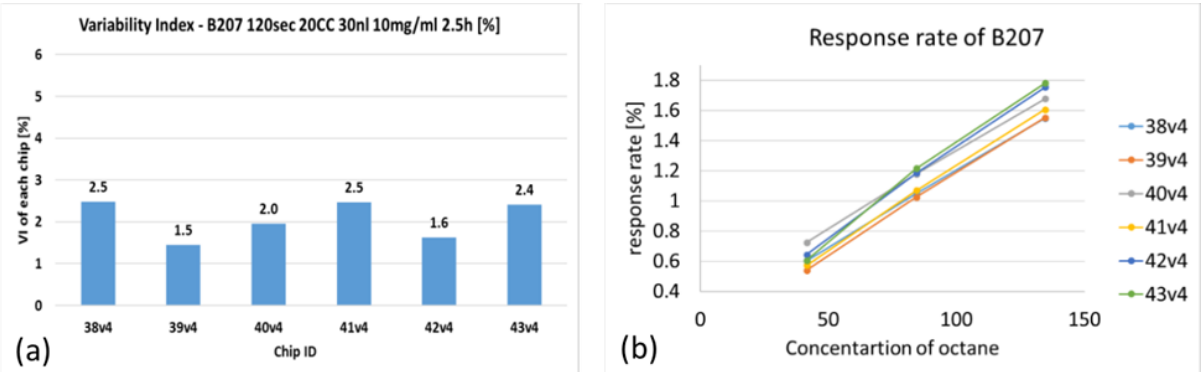

Figure 6: (a) VI of B207 in different chips. (b) Average response rate of six different chips

Figure 7 shows the responses of all the different chemistries to three escalating concentrations of octane (41.4, 84.2 and 134.4 ppm). Each graph contains six different sensors of the same chemistry on the same chip.

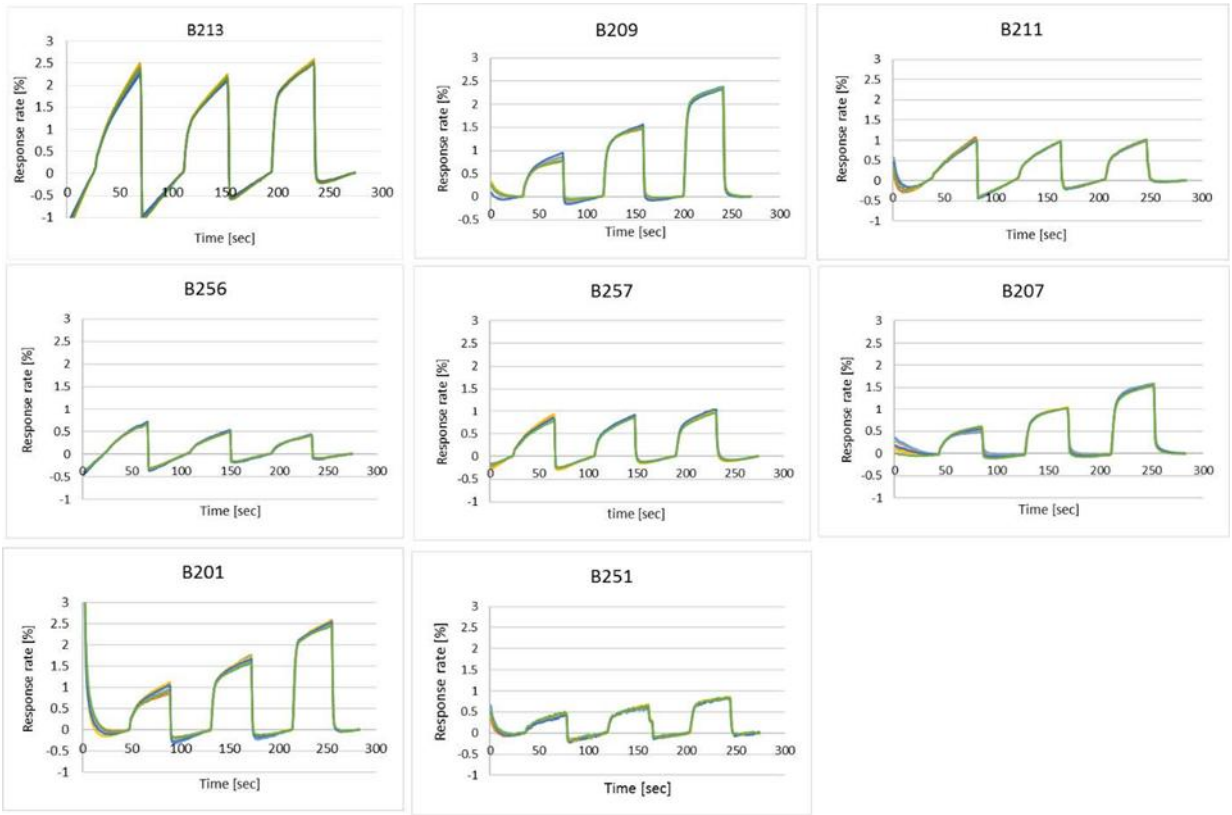

Figure 7: Sensors' response to octane at different concentrations (41.4, 84.2 and 134.4 ppm). Each graph represents the response of six sensors on chip from the same chemistry.

As can be seen in Figure 8, the VI in all of the chemistries are no more than 5% and the average VI of all the chemistries within six chips is 3.2%. Again, reiterating the low predicted device to device differences.

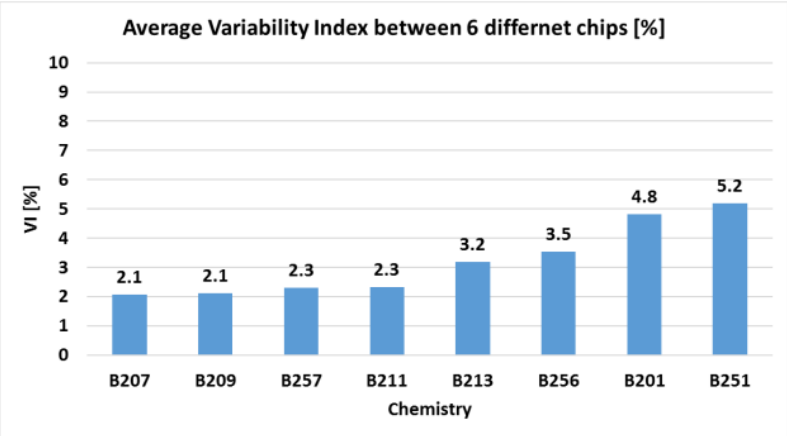

Figure 8: Average variability index of each chemistry across six different chips

Before populating the devices with the sensor chips, these chips were exposed to cancer related biomarkers. Although individual biomarkers reactivity is not as good indicator for nanoarray classification capability, TECH still went ahead and exposed the chips to these biomarkers. This was done in order to exclude the possibility of having the choosing chemistries of the nano chip totally blind to cancer related biomarkers. The tested biomarkers were chosen since they were identified in previous study in our lab and reported in the literature.

Figure 9 and Figure 10 show the responses of different chemistries of the selected nanoarray to increasing concentrations of Ethyl acetate and Acetone. Figure 13 shows the final device sent to the clinical sites.

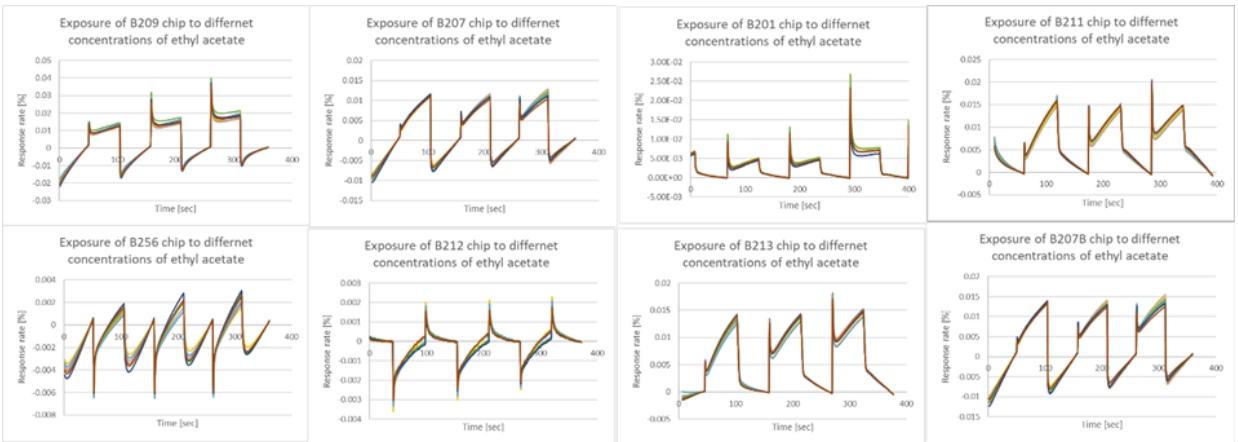

Figure 9: Sensors response to Ethyl acetate at different concentrations (0.1ppm, 0.5 and 2.5 ppm). Each graph represents the response of six sensors on the same chip from the same chemistry

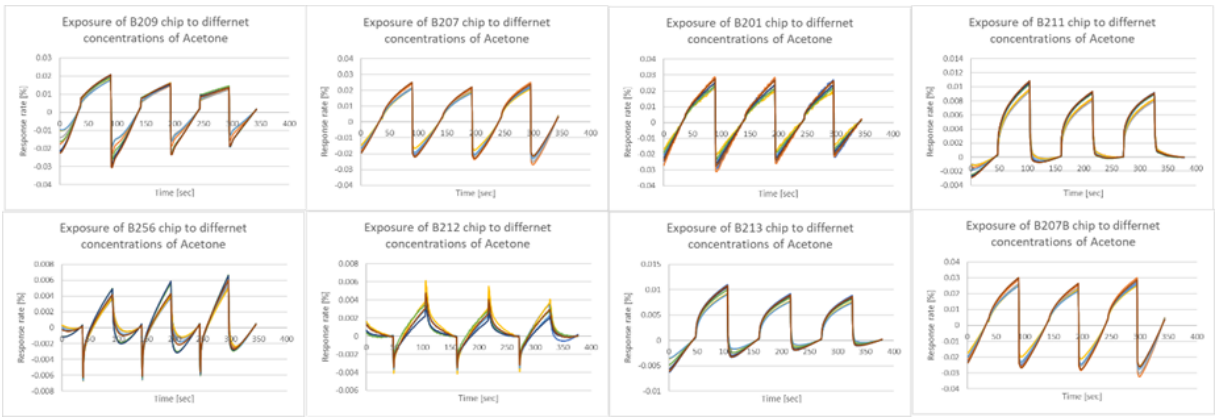

Figure 10: Sensors response to Acetone at different concentrations (0.5, 1.5 and 4.5 ppm). Each graph represents the response of six sensors on the same chip from the same chemistry

Figure 11: picture of a device sent to partners

- **Wide-biomarker-spectrum Multi-Use Sensing Patch (WBSP)**

  The detailed information regarding the WBSP development and construction was previously described in D3.1 and M18 periodic report. Briefly see figure 12:

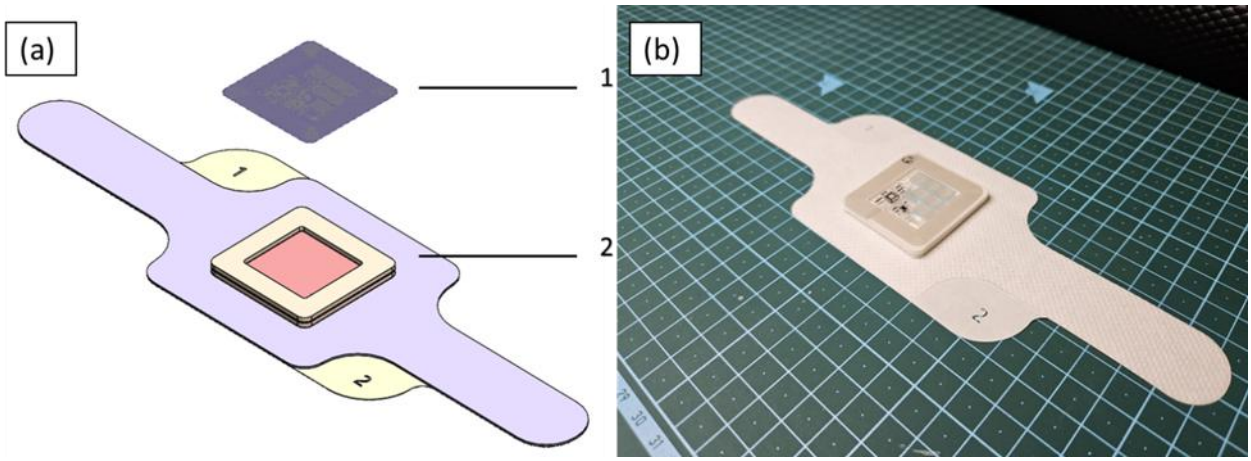

Figure 12: (a) Exploded view of the WBSP; 1 Sensor Electronics (TECH); 2 Skin Interface Patch (PRON) and (b) Complete WBSP unit.

During the past months TECH concentrated on the preparation and production of hundreds of patches to be provided to the clinical partners. Initially 50-100 patched were sent to each of the four clinical sites, during last month additional 300-350 patches were added and sent. In addition, during this time TECH have worked closely with EMODA to finalize the app for sampling the patch measurements as can be seen in the figure below:

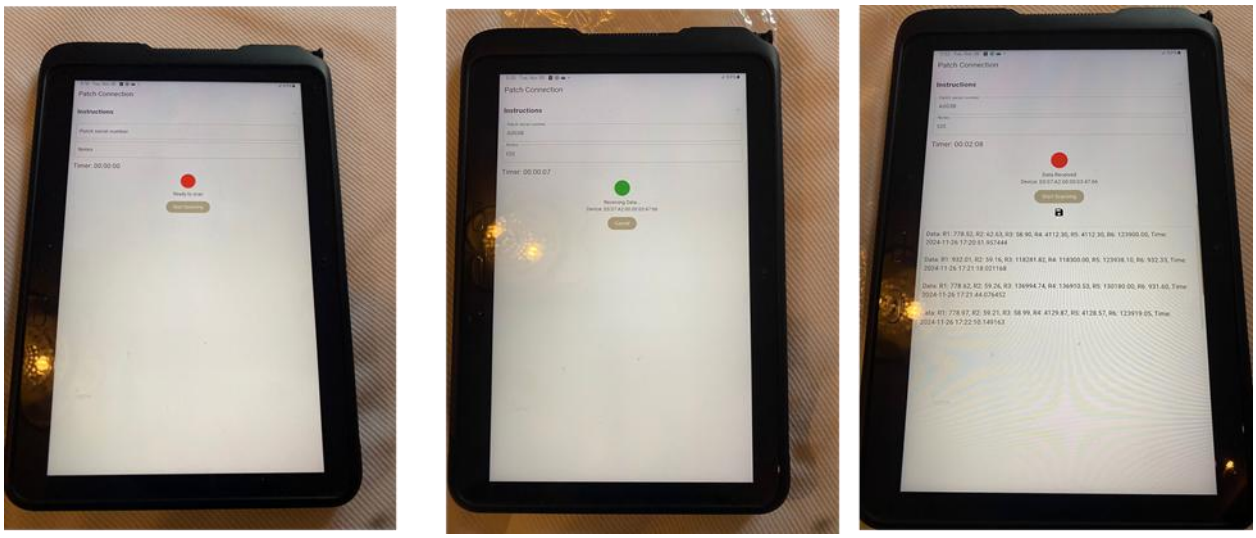

Figure 13: the GUI interface of the developed app. allowing adding details and starting measurement via NFC by introducing the smartphone\tablet to the patch. Red\ green color indicate status of operation. Sensor measurement appear on bottom (right picture) and are uploaded automatically to the health data platform.

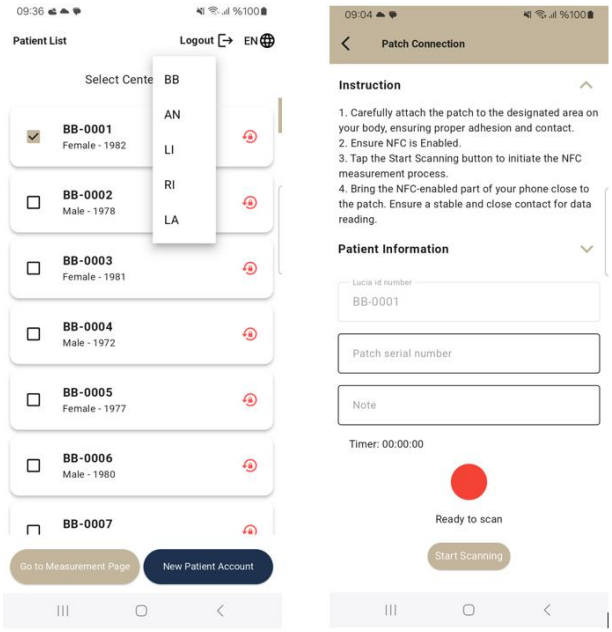

Figure 14: WBSP app screen. On the left image, patient list is displayed, while on the left image the patch measurement screen is shown.

Figure 15 shows the patient list screen on the left, where clinicians can select a clinical center and view the participants associated with that site. Each participant is identified by a unique code, their birth year and sex, ensuring privacy by not displaying personal details such as names or surnames. Clinicians can select a participant and proceed to measure patch data using their phone via NFC, as shown in the image on the right. The selected participant is displayed along with their Lucia ID number.

The image on the right shows the patch measurement screen, where clinicians can enter the patch's serial number, add notes, and begin the scanning process. By placing the back of the phone on the patch and waiting briefly, they can perform the measurement. The screen also displays the duration of the measurement. Once the phone is removed from the patch, the results and elapsed time are shown, allowing the clinician to save the data.

To interact with the patch, Near Field Communication (NFC) is used, a wireless technology that enables short-range data exchange between devices. Specifically, the patch communicates using the NFC-V protocol, which is one of the standardized types of NFC communication based on ISO 15693.

The WBSP is intended to be used in phase 2 and 3, which have just started and thus no results are available yet. One main advantage of the developed set-up (patch and app) is that it allows the sample data taken by the physician to be directly uploaded to the LUCIA Health Data Platform, that stores data from patients from different clinical site and countries. Then after data can be accessed online by authorized staff to perform data analysis. Figure below shows a picture of the output format on the HDP of samples of the patch via app.

Code	CreatedAt	Note	PatchId	PatientId	R1	R2	R3	R4	R5	R6	Time
RI-0033	00:23.0	t00	a014a-01ri0033		314.37	47463.64	47470.91	45037.93	286.81	524.43	59:28.3
RI-0033	22:07.0	t20	a014a-01ri0033		446.34	147.36	132.39	390.52	390.88	931.02	21:13.2
RI-0033	44:29.0	t40	a014a-01ri0033		422.5	113.46	116.17	446.03	364.87	859.88	43:24.1
RI-0033	06:46.0	t60	a014a-01ri0033		407.36	95.75	99.72	144.75	358.39	789.25	05:36.6
BB-0648	02:33.0	T0	A020A-LUCIA-BB-0648		286.43	59.7	59.68	112.42	11898.17	73.9	59:49.8
BB-0648	23:10.0	T20	A020A-LUCIA-BB-0648		287.89	59.73	59.75	112.93	11363.16	94.6	20:27.3
BB-0648	46:38.0	T40	A020A-LUCIA-BB-0648		289.43	61.07	61.31	114.25	11110.17	104.24	43:53.3
BB-0648	03:49.0	T60	A020A-LUCIA-BB-0648		291.33	61.46	73.29	114.04	11449.12	150.38	01:04.5

Figure 15: representative picture of the output format on the HDP of samples of the patch via the app. In this picture, samples from two volunteers from two different clinical partners were tested.

- **Multiomics (MO)**

In the past reporting period CNAG has worked on the implementation of the process for taking the samples collected in the clinical trial through a standardized workflow.

As a reminder, the low-pass genome skimming and the DNA methylation array analysis were replaced by a single analysis with medium-pass nanopore whole genome sequencing. This allows to capture variants and DNA methylation simultaneously. The output can be used to assign polygenic risk scores and methylation levels.

However, in preparation CNAG needed to ascertain that the DNA prepared from the samples taken from the study participants, needs to be compatible with the nanopore analysis. Clinical samples are coming from the four clinical partners in the Basque Country, Andalusia, Belgium and Latvia. A procedure has been defined for the extraction of DNA, starting with either whole blood or buffycoat. Both substrate turned out to yield DNA of sufficient quality and no differences were detected when this DNA was used for nanopore sequencing (Oxford Nanopore Technologies).

A marked difference in the performance of the DNA coming from the different collection sites was detected. These differences were traced back to the DNA extraction methods applied. As a conclusion, the recommended DNA extraction kit

was from Qiagen and best used in combination with the Qiagen Symphony robotic system.

On the side of the laboratory CANG has worked on streamlining the process to minimize the number of processing steps and has also focused on getting the balancing of the different samples in a pool optimized. In the current strategy a high number of samples (24) are pooled in one flowcell and then run multiple flowcells with rebalancing pools to achieve the same coverage for each sample. CNAG has carried out the first production-like test runs.

Planning, execution and results of Task 4.5 "Contextual-empirical investigations to evaluate the realization of identified values"

- **Methodological overview**

The study was conducted between August and December 2024, following a structured multi-step process aiming to gather insights and address challenges related to value realization in the LUCIA project (see Figure 1 in page 6). It began with a pre-interview survey to collect basic data from clinical partners and assess their perceptions of how well the LUCIA clinical study aligns with the identified values. This was followed by semi-structured interviews with clinical partners to explore challenges related to the study, including the use of LUCIA technologies. The interview results were subjected to thematic analysis, alongside input from other tasks (Task 1.1 and Task 1.4), as well as reviewer comments from the Periodic Report Review and comments of the LUCIA Stakeholder Advisory Board. This process identified four main themes, which were presented to the LUCIA consortium along with the relevant risks and values. To explore potential solutions, a Value Realisation Workshop was held during the LUCIA Consortium meeting. Finally, key recommendations from the workshop were summarized and shared with the consortium.

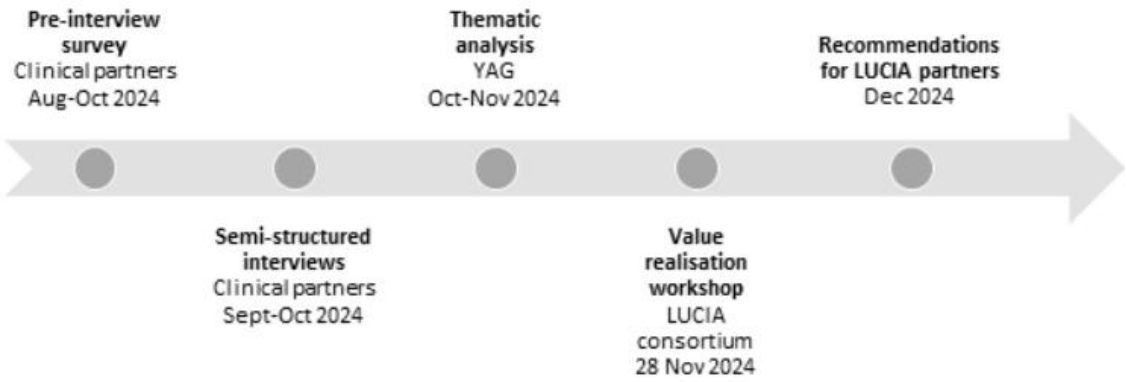

Figure 16: Methodological overview of the study

- **Pre-interview survey**:

The pre-interview survey was sent to all clinical partners in August 2024.

Clinical partners were advised to fill in the survey the latest around 1 week before their interview for the interview facilitator to prepare for the interview based on the survey questions (Table 7). They were asked to send the survey to anyone familiar with the study in their center, but at least one person per partner. For most clinical partners one person filled in the survey.

Table 7: Description of the data collection points

	UL	CHUL	SAS	BB
Date of survey filled in	13/09/2024	19/09/2024	30/09/2024	02/10/2024
Participants in survey	Ilmārs Stonāns	Benoit Ernst	Laura Vangas Gonzalez; Luis Grabriel Luque Romero	Jon Eneko Idoyaga Uribarrena
Recruitment				
Date of first recruitment	17/07/2024	Mid-April/2024	04/07/2024	06/06/2024
No. of people tested (at the time of survey)	7	10	45	411

At the time of the survey, all clinical partners had started the recruitment and testing with LUCIA technologies. The date of the first participant recruited to the clinical study ranged from mid-April to mid-July 2024. Number of participants tested were ranging between 7 and 411 participants (Table 7).

There were 7 statements for legal requirement and 7 statements listed for the ethical and social aspects of AI. These statements were formed based on Deliverable 1.1. Survey participants were asked to rate on the scale of 1 (Strongly disagree) and 5 (Strongly agree) how much they agreed with each statement, in the context of implementing and using the LUCIA technologies at their clinical center. After each section, there was a free text box to leave comments and explain if any choices were below 5 (Strongly agree).

For the statements concerning legal requirements, all participants marked "Agree (4)" or "Strongly agree". CHUL participant marked "Neither agree or disagree" for Storage limitation and left a free-text comment: "In the processing of personal data, there are always risks. One cannot state that adverse effects on the data processing will for sure be avoided as there are always practical and technical constraints that create risks of data breach for example. Regarding data storage limitation, it is planned to protect, limit and erase data access once data will not be needed anymore, but this is yet to come thus I cannot state on this at present." No other comments were made.

For the statements concerning the ethical and social aspects of AI three participants marked "Neither agree or disagree (3)" for all statements, with comments added on 1) lack of information about these aspects within LUCIA, 2) lack of competence and 3) noting the importance but the improper timing of these survey questions suggesting to complete the survey later when participants have more insights on the AI models used in the clinical study. Two participants skipped almost all questions. One marked "Agree" or "Strongly agree" on most, except for the statement on Transparency receiving a "Neither agree or disagree (3)" mark without a note.

Results of the pre-interview survey were discussed in more details throughout the interviews.

- **Value realization interviews**

Interviews were conducted in 90 minutes online throughout September and October of 2024 with each clinical partner. The semi-structured interviews followed a semi-structured interview design, with a detailed interview guide. All interviews were conducted by the same research project manager at Yaghma. The description of the interview details can be found in Table 8.

Table 8: Description of the interview details

	UL	CHUL	SAS	BB
Date of interview	25/09/2024	01/10/2024	08/10/2024	21/10/2024
Participants in interview	Ilmārs Stonāns, Alvils Krams	Benoit Ernst	Laura Vangas Gonzalez; Encarnacion Gil	Eunate Arana Arri; Jon Eneko Idoyaga Uribarrena

Participant titles	Research coordinator¶ Principal investigator, clinician	Assisting the principal investigator	Nurses	Clinician (family and emergency physician). PI Clinical Researcher.

The interviews had three parts:

1. Participant and clinical site information
2. Collecting information on challenges with the implementation of LUCIA technologies (alone and in combination)
3. Discussing pre-interview survey result

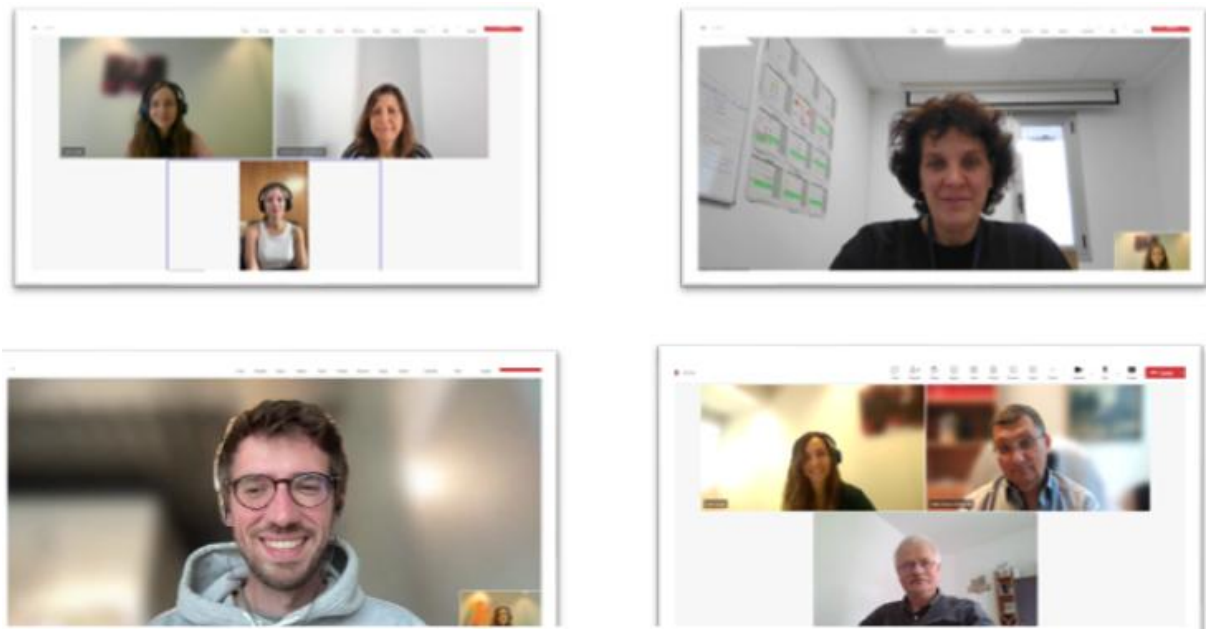

Figure 17: Value realisation interviews with clinical partners

- *Participant and clinical site information*

The status of recruitment was asked both in the pre-interview survey and then in the interview. All clinical study sites had been using the BAN device and collecting blood samples for SPOC and the genetic testing. By the time of the interview none of them had conducted the analysis of blood samples yet. CHUL noted their laboratory team has all the material needed for doing the first

SPOC analysis later that month (planned for mid-October 2024), BB has carried out several tests with samples donated by the centre's staff to fine-tune the device. Various tests have also been required to adapt the extraction and validation material to the SPOC requirements. The analysis of samples with LUCIA volunteers will begin in November 2024.

CHUL noted to be the latest to receive the BAN devices (2 September 2024), explaining the difference in the number of participants giving blood samples and tested with the BAN device. All partners have received the skin patches by the time of the interview, but none of them have started using them on study participants yet because they are waiting for the mHealth app to arrive which will analyse the measurement and because Phase II of the clinical study has not started yet at any of the sites. BB noted this mHealth app will also function as the patient facing interface of the questionnaires planned, now ran either on paper of with the help of online surveying software (Google Forms – BB). None of the partners have started using any risk prediction model at the time of the interview, as there are complications with finding an externally developed lung cancer risk prediction tool that is in alignment with the inclusion criteria of the LUCIA study. This topic will be discussed on the next Consortium meeting (27-28 November 2024, Seville).

Table 9: Status of participant recruitment and use of technologies

	UL	CHUL	SAS	BB/OSA
Recruitment				
Date of first recruitment	17/07/2024	Mid-April/2024	04/07/2024	06/06/2024
No. of people tested (at the time of interview)	9 (1 excluded due to inclusion criteria not fulfilled)	10 enrolled in blood sampling and 4-5 in BAN (115 recruited)	68	482 (1100+ recruited)
Technologies used				
BAN	yes	yes	yes	yes
SPOC	yes*	yes*	yes*	yes*
Genetic tests	yes	yes	yes	yes***
Skin patch	no**	no**	no**	no**
Risk prediction model	no	no	no	no
mHealth app	no	no	no	no

*Samples collected and stored, not analyzed yet

**Skin patches received, not started using them as it comes in Phase 2 and once the relating application will be ready for the readings

*** Samples have been sent to CNAG for analysis of the extraction quality, which was satisfactory. Waiting for the next partners to send samples to evaluate the quality and continue sending samples.

- *Collecting information on challenges with the implementation of LUCIA technologies (alone and in combination)*

  After clarifying which of the LUCIA technologies have been used in the specific clinical site, the next part of the interviews included discussing in details through the following questions:

  - Have you experienced any kind of challenges or problems with regards to the technology as **study coordinators**, when setting up the system and training staff?
  - Have the **healthcare professionals** experienced any challenges or problems with regards to the technology? If yes, please explain what happened and why it may be a problem.
  - Have the **study participants** experienced any challenges or problems with regards to the technology? If yes, please explain what happened and why it may be a problem.
  - Were there cases that study participants, that are eligible for the technology sampling, **were not sampled at all** (refused to give sample) with the technology for any reason?

  The input of each clinical site representative has been collected per topics according to the following topics for the LUCIA study in general:

  - Equipment and requirements
  - Infrastructure
  - Inclusion/ exclusion criteria
  - Study population bias
  - Recruitment
  - Motivation and follow-up
  - Human resources
  - Usability

- Questionnaires
- Waste management
- Costs
- CT scans
- Next steps

The input of each clinical site representative has been collected per topics according to the following topics for each LUCIA technology:

- Equipment and requirements
- Infrastructure
- Training
- User experience

- *Discussing pre-interview survey results*

In the last part of the interviews pre-interview survey results have been discussed. These discussions were brief, as clinical partners reassured their input to the survey noting that the lung cancer risk prediction model to be used in determining which participants will be followed for Phase II of the clinical study has not been chosen before at the time of the interviews, therefore they cannot comment on the AI's ethical and social aspects. Regarding legal requirements, CHUL has reassured their note on data storage limitation and that data security cannot be fully promised in advance, with the inevitability of practical and technical constraints that create risks of a data breach. SAS noted they have marked concerns with legal requirements by mistake on the pre-interview survey.

- **Thematic analysis**

In the next phase of the research, a thematic analysis was conducted to identify potential risks to the realisation of values. The main input to identifying these **risks** were the interviews with clinical partners, but additional insights were drawn from reviewer comments on the M18 Periodic Report and feedback provided by the LUCIA Stakeholder Advisory Board (Figure 18).

The **values** were derived from previously published LUCIA deliverables (D1.1) on the legal requirements of LUCIA and ethical and social aspects of AI in LUCIA. Additionally, to channel potential values of LUCIA stakeholders', insights from the Social Lab workshops held in May-June 2024 were added to the analysis (Figure 18).

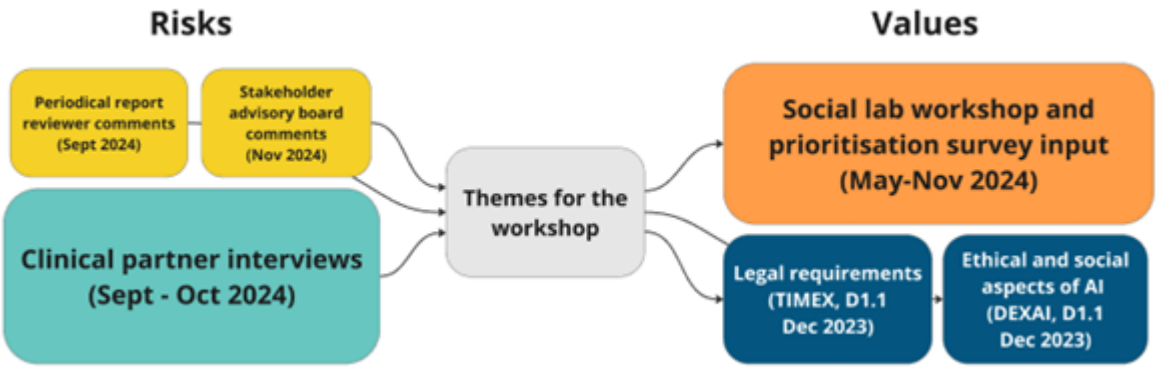

Figure 18: Inputs for risks and values of the thematic analysis

Social Lab is a collaborative, participatory method used to address complex challenges by bringing together diverse stakeholders. It involves collectively identifying barriers, brainstorming solutions, and developing actionable recommendations through structured discussions and activities. The goal is to foster shared understanding and create practical strategies for overcoming challenges. Two Social lab workshops have been organised by the time of the thematic analysis, one in May 2024 connected to the LUCIA Consortium meeting in Mannheim, and one in June 2024 online representatives of stakeholder organisations and other lung cancer research projects. Participants were asked to imagine a hypothetical future, 5 years after the end of the LUCIA project, and to brainstorm on potential barriers that may hinder the effective implementation and use of LUCIA technologies. These barriers have been analysed forming a list of 80 barriers and were prioritised though a prioritisation survey by November 2024.

Barriers have been prioritised on two domains by two groups of respondents 1) LUCIA partners were asked to rank each barrier on a Likert-scale of 1-5 based on the potential influence the LUCIA consortium can have on minimising the impact of the barrier by the end of the project timeline ('LUCIA influence' on Figure 19) and 2) registered participants of the online Social lab workshop in June 2024 representing stakeholder organisations and other (lung) cancer research projects were asked to rank each barrier on a Likert scale of 1-5 based on the potential negative impact the barrier may have on the successful implementation of LUCIA technologies to lung cancer screening in the future. The mean scores of each barrier per the two domains have been added to an 'Impact-influence map' (Figure 19). Those barriers scoring 3 or more than 3 on both domains have been added as input to the thematic analysis.

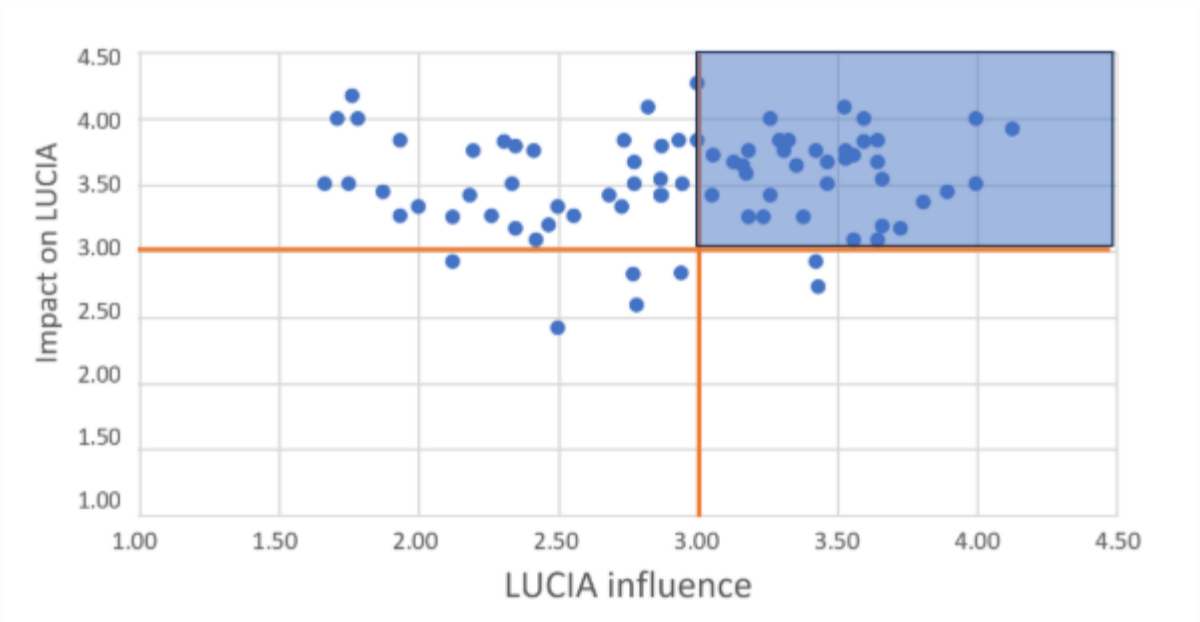

Figure 19: Impact-influence map of barriers collected in the Social lab workshops in May-June 2024 and prioritised through the prioritisation survey in Oct-Nov 2024 :

The thematic analysis consisted of four steps, using an online visualization platform (Miro). The full analysis was performed by one researcher at Yaghma with a background in health policy and health economics research. Validation steps with LUCIA partners have been implemented to minimize this limitation.

Input on potential risks to LUCIA values have been excluded from the thematic analysis for:

- *Information related to the setting up of the system* – these inputs will be valuable for writing recommendations on using the LUCIA technologies, but are not relevant to the LUCIA project progress as all clinical sites have set up their systems and started testing
- *Information related to ease of use of technologies* – these inputs are expected to be handled by technology owners
- *Information related to cost or cost-effectiveness of the technologies* – these inputs will be relevant at later stages of the LUCIA project

First, based on reading all the input draft themes have been set up. Then all input has been categorized into the draft themes using color codes. The phrasing and grouping of themes were changing iteratively, shown on diagrams. After finalizing the diagrams,

they were sent to the whole LUCIA consortium for comments with all raw input listed as pre-reading materials for the Value realization workshop. No comments arrived at this stage. LUCIA partners had the chance to ask clarifying questions or leave comments regarding the identified risks throughout the workshop.

The following four themes have been identified for the value realization of LUCIA:

- **Theme 1**: Study population characteristics (See Figure 20)
- **Theme 2**: Recruiting enough participants (See Figure 21)
- **Theme 3**: Quality of data (See Figure 22)
- **Theme 4**: Informing participants (See Figure 23)

Each of the diagrams show the risks and values potentially connected to the theme. Color coding per input is aligned with that of Figure 18.

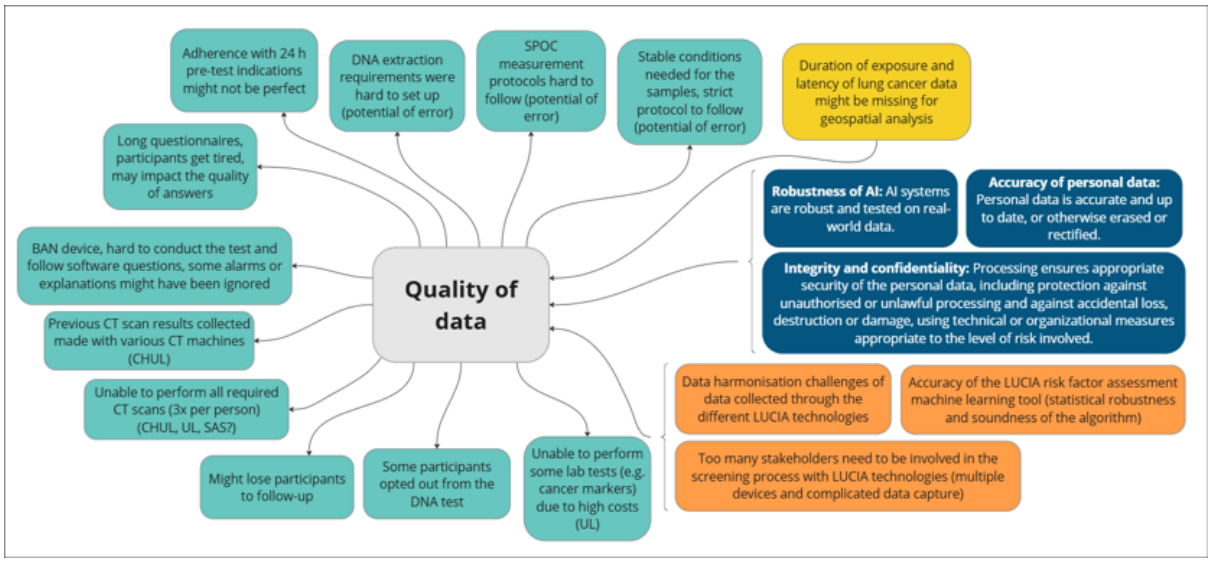

Figure 20: Diagram for Theme 1: Study population characteristics:

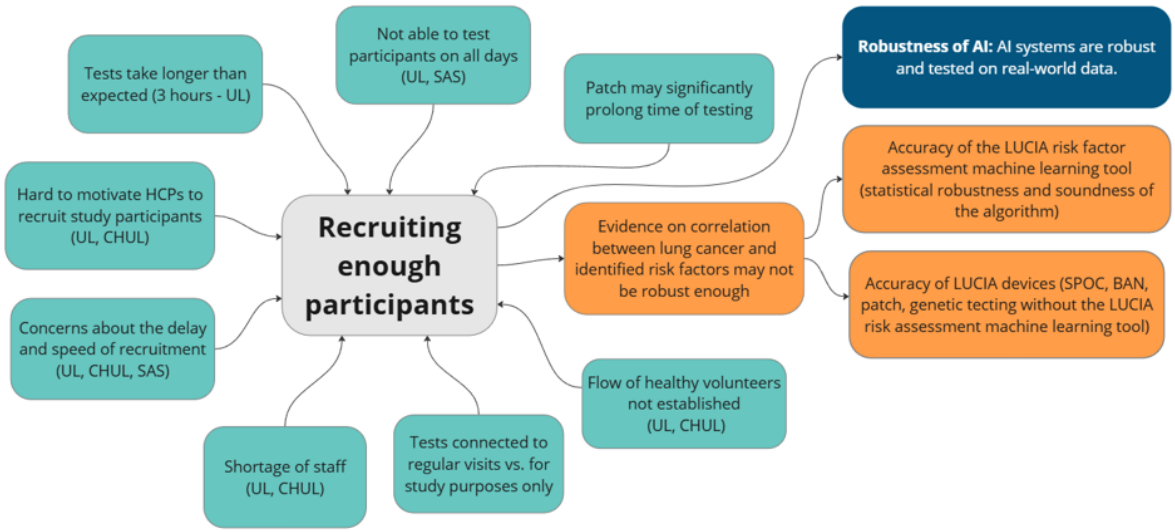

Figure 21: Diagram for Theme 2. Recruiting enough participants

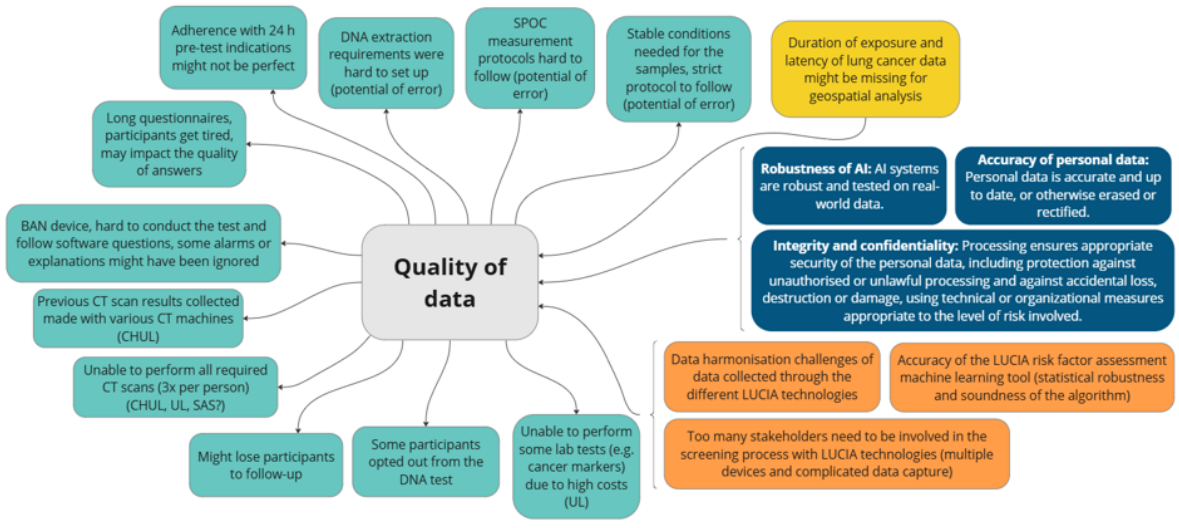

Figure 22: Diagram for Theme 3. Quality of data

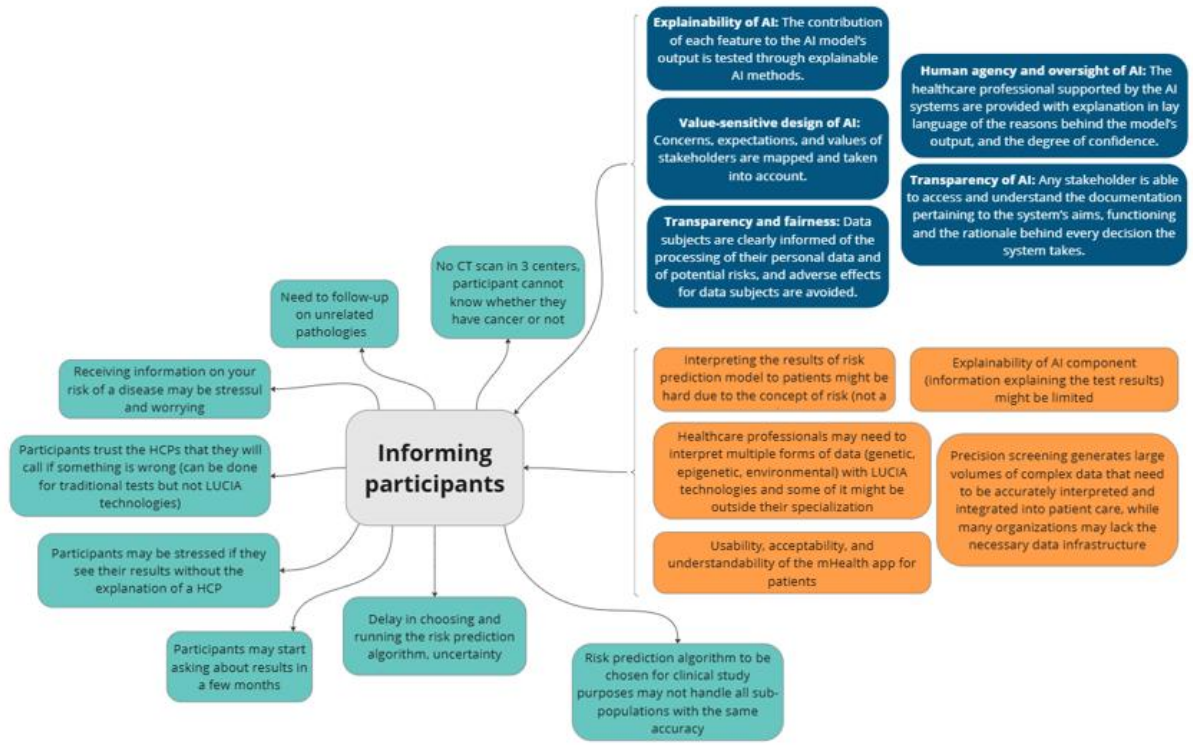

Figure 23: Diagram for Theme 4. Informing participants

These diagrams (Figure 20-23) were presented and discussed in detail on the Value realization workshop.

- **Workshop**

The value realization workshop was held face-to-face in Seville, Spain, as part of the LUCIA Consortium meeting on the 28th of November 2024. Thirty-seven (n=37) members of the LUCIA consortium have actively contributed to the 120-minute workshop. All LUCIA Consortium members have received the four diagrams on the themes (Figure 20-23) as well as the raw input text that was analyzed thematically, with a request to review.

Figure 24: Small group work on the value realization workshop in Seville, Spain on 28 November 2024

The **objective of the workshop** was to brainstorm on potential solutions for the risks listed at each of the 4 themes to be implemented until the end of the project duration. There were four main parts of the workshop. First, data collection and results of the thematic analysis were presented. Individual work brainstorming on solutions followed to give a chance to participants to reflect on their own. Then, participants were asked to form small groups of 3-4 people and continue to brainstorm on solutions together (Figure 24).

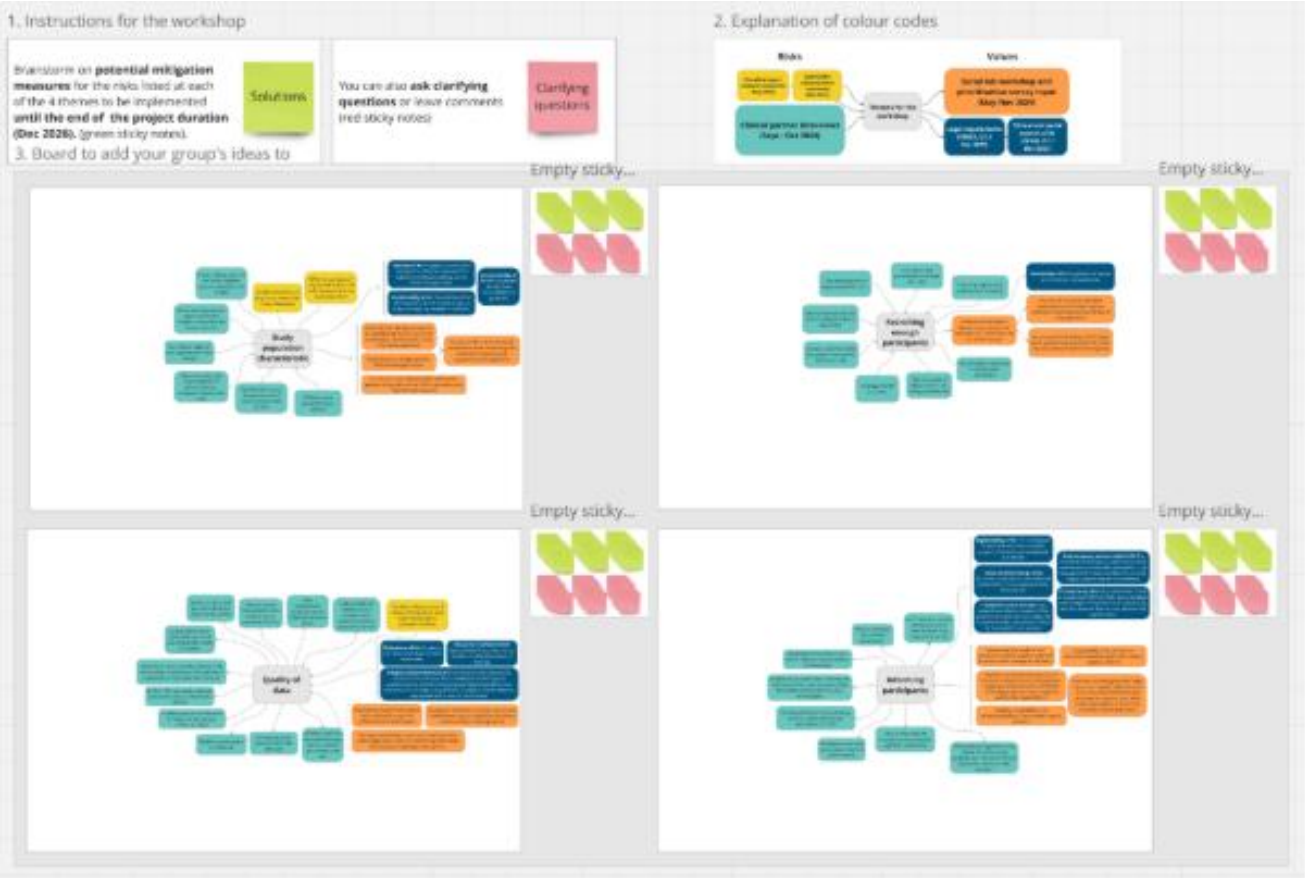

Figure 25: The online collaborative platform board used for the workshop (Miro) before the workshop input

Twenty (n=12) teams were formed. An online collaborative work platform has been used for the workshop (Miro). All groups had access to the same platform online through their personal laptops with the diagrams on themes, and they were able to add input and link it to specific risks. There were two types of input to choose from: 1) potential solutions (green 'sticky notes') or 2) clarifying questions and comments (red 'sticky notes') (Figure 25).

All small groups had to discuss all 4 themes. Each theme contained 8-11 potential risks to discuss (Table 10). After the brainstorming, each theme had 13-24 potential solutions collected, connected to one of multiple risks per theme.

Table 10: Summary of number of potential risks, solutions and resulting action points per themes

Themes	No. of potential risks collected through the thematic analysis before the workshop	No. of potential solutions collected on the workshop	No. of recommendations formed after the workshop
Theme 1: Study population characteristics	8	20	10
Theme 2: Recruiting enough participants	8	16	13
Theme 3: Quality of data	11	24	12
Theme 4: Informing participants	8	13	8
SUMMARY	35	73	43

After brainstorming in groups, each group had to vote on the most pressuring risks to discuss together with the consortium on that day. Each group had 8 votes, and they were allowed to distribute the votes unevenly between themes. Each risk could only be voted once per group.

The top three most voted topics on potential risks were discussed together with all participants of the workshop. Potential solutions and clarifying questions were presented, and relevant partners were asked to comment. The following topics have been discussed:

- Tests take longer than expected (3 hours – UL) (10 votes)
- Need to involve lung cancer patients to validate devices – exclusion criteria in the study (9 votes)
- Concerns about the delay and speed of recruitment (UL, CHUL, SAS) (7 votes)

For tests taking longer than expected in one center, BB/OSA presented their solution with parallel testing of multiple patients, streamlining the testing process and digitalizing the questionnaires making it easier to track if questions were left blank. A discussion on the number of people they have working on the study has unfolded. CHUL noted Belgian rules are specific in who can do the tests, therefore, together with the available human resources parallel testing is not an option for them. The topic on the need to involve lung cancer patients was discussed next and the potential misalignment with the prospective study protocol. BB/OSA clarified Phase III of the clinical study has an addition to involve newly diagnosed lung cancer patients to

ensure enough lung cancer patients are involved in the study. The discussion on the definition of newly diagnosed lung cancer patient needs to be continued and aligned within centers. For the concerns about the delay and speed of recruitment, again, BB/OSA has shared their practices and ensured that they are happy to discuss and share with interested centers. The need to involve GPs actively and promote the study to participants directly (e.g. through posters in GP offices) was highlighted as a working practice for BB/OSA.

After the workshop, input from the online collaborative platform (Figure 26) was analyzed and formed into recommendations. Partners to consider the recommendations were coupled with each recommendation. The draft list has been sent to LUCIA Work package 4 members to comment and amend, including relevant partners.

There were two clarifying comments on risks identified in the thematic analysis, one highlighting that having participants with a history of lung cancer in the family can be an opportunity from the genetic perspective rather than a risk, and another noting that recruiting patients with a lung disease other than cancer recruited (CHUL) may not be a problem, as there are lung diseases in general population as well.

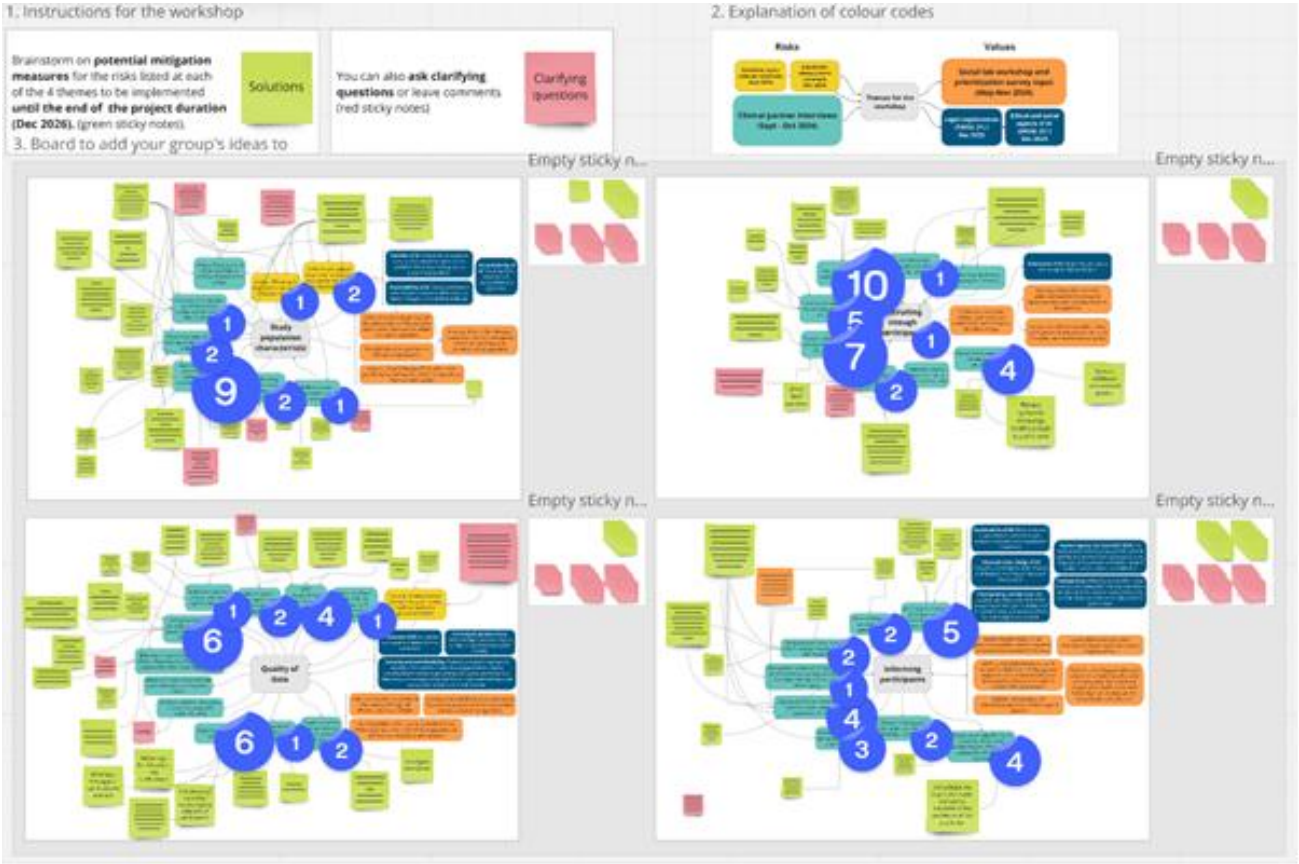

Figure 26:The online collaborative platform board used for the workshop (Miro) after the workshop input (green and red 'sticky notes') and voting (results marked with blue circles

The following recommendations have been formed per themes, ranging 8-13 (Table 11) per each theme (table 12 to Table 15)

Table 11:Recommendations formed to mitigate identified risks and partners to consider for Theme 1: Study population characteristics

Theme 1: Study population characteristics		
Recommendation to mitigate identified risks	**Partner to consider**	**Comment**
- Assess what the original power analysis says about representativeness and sample size	BB	
- Develop a stratified sample target % male/female, age, region, smoker/non-smoker, healthy/non-healthy	BB	

- On recruitment of clinical patients, joint environmental datasets then monthly revisions of recruited samples against targets - Intermediate report to assess distributions of patients and sample the most needed groups from the volunteer queue - Create a dashboard with the characteristics of the population to correct the over representation - Interim recruitment analysis to balance the sample	BB (UL, CHUL, SAS)	
- Provide flexibility in the time slots (e.g. appointments in the evening) and/or give proof of attendance to make sure more people who work can participate in the study	UL	
- Clarify numbers and identify needs of hard-to-reach patients included (e.g. those who work)	BB, UL, CHUL, SAS	
- Diversify the recruitment places so that younger participants can join the study	UL	
- Broaden the ways that the recruitment advertisement is done to reach non-ex smokers (social media, posters, Whatsapp, newspaper, email lists...) / Targeted recruitment by public advertisements or pre-screening by mobile app	BB, UL, CHUL, SAS	
- Analyse differences of subtypes of LC stages	Tech partners	
- Recruit LC Patients from specialized consultancies / need for a parallel recruitment but not inside LUCIA cohort	BB, SAS	
- Agree on a clear definition of newly diagnosed lung cancer patients to be included in Phase 3	BB, UL, CHUL, SAS	

Table 12:Recommendations formed to mitigate identified risks and partners to consider for Theme 2: Recruiting enough participants

Theme 2: Recruiting enough participants		
Recommendation to mitigate identified risks	**Partner to consider**	**Comment**
- For the concerns of the patch significantly prolonging the time of testing from Phase 2 of the study, it is advised to reassess with TECH on how to streamline the process - collaboration between tech providers and clinicians	TECH, BB, UL, CHUL, SAS	

- Supporting community managers was highlighted to recruit enough participants	BB	
- Parallel testing of multiple participants can help with time management of testing	UL, SAS	
- Disseminate the questionnaires to the subjects before they arrive to the centre can help to streamline completion and shorten time needed for testing	UL	
- Try to merge similar questions to shorten the questionnaires to shorten the time needed for testing	UL	
- Provide incentives to healthcare professionals, especially in primary care to motivate them to join the study (e.g. possibility of co-authorship on publications and other activities that help them to feel part of the project). This opportunity would be prestigious for them and would contribute significantly to their re-certification points.	BB, UL, CHUL, SAS	
- Provide fruit and water to participants to motivate them to join the study	BB, UL, CHUL, SAS	
- Reward system to encourage healthy people to participate	BB, UL, CHUL, SAS	
- Open to additional recruitment centers in primary care	UL, CHUL	
- Encourage primary care professionals to invite healthy people	UL, CHUL	
- Share best practices within clinical partners to speed up recruitment and avoid further delays	BB, UL, CHUL, SAS	
- Divert resources to those centers who can recruit more participants in a timely manner with minimal realistic targets per recruitment center	BB, UL, CHUL, SAS	
- Need to investigate why some participants cancel last minute	BB, UL, CHUL, SAS	

Table 13:Recommendations formed to mitigate identified risks and partners to consider for Theme 3: Quality of data

Theme 3: Quality of data		
Recommendation to mitigate identified risks	**Partner to consider**	**Comment**
- Acknowledge the limitation that duration of exposure and latency of lung cancer data will be missing for geospatial analysis / Given the nature of the data being collected, this point cannot be properly addressed (however, there is a question on previous address in the clinical questionnaire)	ULSTER, Tech partners	
- Hospital staff to visit sequencing center for quality assessment training	CNAG, BB, UL, CHUL, SAS	
- Investigate whether one of the four DNA extraction kits be selected for use by all partners to mitigate risk to the variability of results	CNAG, BB, UL, CHUL, SAS	
- Improve standard operating protocols for SPOC and BAN	NAN, TECH, CNAG, BB, UL, CHUL, SAS	
- More emphasis on preparation procedure (reminder via Whatsapp some hours before) to ensure participants adhere with 24 h pre-test indications / Try to reforce the information before coming to the consultance and send a remind the day before	BB, UL, CHUL, SAS	
- Use google forms to register part of the information of adherence to pre-test indications	BB, UL, CHUL, SAS	
- Provide some kind of standardised explanation in the questionnaires for frequently asked questions to help participants in filling out the questionnaired easier	BB, UL, CHUL, SAS	
- Split questionnaires into several smaller ones to prevent participants getting tires from long questionnaires	BB, UL, CHUL, SAS	
- Provide more support depending on the profile of the participant to fill in the questionnaires	BB, UL, CHUL, SAS	
- Streamline the questionnaires moving forward	BB, UL, CHUL, SAS	
- Whatsapp, message, call or use mobile app to remind participants for the next visit	BB, UL, CHUL, SAS	

- Send information about the project (communications, social media posts, etc.) progress and keep in touch to make sure participants will not be lost for follow-up	BB, UL, CHUL, SAS	
- Use analytical tools that handle missing data in case of loss of participants for follow-up	Tech partners	
- Offer better information besides the materials for the informed consent on the DNA test	CNAG, BB, UL, CHUL, SAS	
- Enhance transparency of handling genetic data	CNAG, BB, UL, CHUL, SAS	
- Do the lab tests missing for high risk participants from phase 1	UL	
- Investigate alternatives to the lab tests not performed in UL	UL	

Table 14: Recommendations formed to mitigate identified risks and partners to consider for Theme 4: Informing participants

Theme 4: Informing participants		
Recommendation to mitigate identified risks	**Partner to consider**	**Comment**
- Clearly inform participants that they are enrolling in a study and provide detailed information about the study's characteristics. This will help avoid any confusion about whether they have been screened for lung cancer in Phase 1 or not (low-dose CT performed or not). Ensure transparency regarding the study's objectives and the specific phase they are involved in.	BB, UL, CHUL, SAS	
- Need to collaborate closely with primary HCP's to hand over this issue for follow up or treatment of unrelated pathologies	BB, UL, CHUL, SAS	
- Involve the primary care doctor of the patient to help with communication about test results and risk for lung cancer	BB, UL, CHUL, SAS	
- Contact participants in a short time to explain the results	BB, UL, CHUL, SAS	
- Program a visit with the health care practitioner for the results of tests as soon as possible after the results are ready	BB, UL, CHUL, SAS	
- Create information sheets for patients handed at first or next consultation to include information like timeframes and expectations	BB, UL, CHUL, SAS	
- Need for a simple explanation for the patient concerning the limitations of the prediction model	BB, UL, CHUL, SAS	

- Acknowledge the bias in the risk prediction model and add an indication of the confidence of the prediction for different sub-populations	BB, UL, CHUL, SAS	

- **Next steps**

In March 2025 YAG requested **feedback from the LUCIA consortium** in writing format on whether the recommendations have been followed/ planned to follow and whether they seem sufficient to mitigate the identified risks.

Those risks, that does not seem handles will be channeled to the next Social lab workshops in May/June 2025 **(Task 1.4)** to find solutions for, together with the barriers not handled through the Value realization workshop (ranking high on impact but low on the influence of the LUCIA consortium within the project timeframe on them). Solutions that proved to be impactful within the LUCIA project concerning the use of LUCIA technologies will be included in the final rrecommendations list of Task 1.4 for future implementation outside of the project scope.

Recommendations marking tech partners as responsible partners will be channeled to **Task 1.3** AI Impact Assessment and followed up on with their potential impact on biasness, explainability and transparency.

Task 4.5 will be continued mid-2026 with another round of interviews with clinical partners and surveying on identified values, to track the realization of values.

Conclusions:

Clinical partners have recruited a total number of **2,855 participants** at month 24 (April 2025). All of those have already performed baseline visit.

There is some information missing in the eCRF about some of the participants due to the great amount of participants recruited. Clinical sites are already finalizing including all data in the eCRF to have the complete picture of the baseline visit.

As some of the devices' test take longer time to be performed and risk calculation is pending in some clinical sites, the performance of the patch test is being done as high risk participants (phase 2) are being identified.

Regarding the SPOC, samples have been frozen to be able to perform the test later in the labs of the sites to ensure the technique is carried out correctly and with utmost quality.

The analysis of the baseline visits of the volunteers recruited in phase 1 is being finalized. Also, after risk stratification and the performance of the LDCT, the final distribution numbers by phases (2 and 3) will be established, and the necessary number of volunteers with LC and IPN to be recruited in the coming months from the clinical services will be estimated.

In the following months, clinical sites will carry out the follow up visits of the participants till the end of study while recruiting Lung Cancer and/or IPN patients to ensure the objectives of the project are met.