# TITLES

Predicting dying: metabolomic changes, a model and the dying process in patients with lung cancer

# Authors

Séamus Coyle PhD1,2, Elinor Chapman PhD3,4, David Hughes PhD5, James Baker MRes3, Rachael Slater PhD3 , Andrew S. Davison PhD6,7, Brendan P. Norman PhD7, Ivayla Roberts8, Amara C. Nwosu PhD9,10,11, Mark T. Boyd PhD1,3, Catriona R. Mayland MD12,13, Douglas B Kell8,14, Stephen Mason PhD13, John Ellershaw MA9,13, Chris Probert MD3.

**Corresponding Author**

Dr Séamus Coyle,

Liverpool Head and Neck Cancer Centre,

University of Liverpool,

200 London Road,

Liverpool,

L3 9TA

UK

Tel: +44 (0)151 794 8806

Email: s.coyle@liverpool.ac.uk

**Author Affiliations**

1. Liverpool Head and Neck Cancer Centre, University of Liverpool, UK.

2. Department of Palliative Medicine, Clatterbridge Cancer Centre, Liverpool, UK.

3. Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, UK

4. School of Medical and Health Sciences, Bangor University, Bangor, UK

5. Department of Health Data Science, University of Liverpool, UK.

6. Department of Clinical Biochemistry and Metabolic Medicine, Liverpool Clinical Laboratories, Liverpool University Hospitals Foundation Trust, Liverpool, UK.

7. Institute of Life Course & Medical Sciences, University of Liverpool, UK.

8. Centre for Metabolomics Research (CMR), Department of Biochemistry and Systems Biology, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, UK

9. Academic Palliative & End of Life Care Department, Liverpool University Hospitals NHS Foundation Trust, Liverpool, UK.

10. Lancaster Medical School, Lancaster University, Lancaster, UK.

11. Marie Curie Hospice, Liverpool, UK.

12. Department of Oncology and Metabolism, University of Sheffield, UK.

13. Palliative Care Unit, Institute of Life Course & Medical Sciences, University of Liverpool, Liverpool, UK.

14. Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark, Building 220, Chemitorvet, 2000, Kgs Lyngby, Denmark

**Word count** 3425 / 3500 for manuscript only (not including Abstract and Research in context)

# ABSTRACT

Background:

Accurately recognising that a person may be dying is central for improving their experience of care. Yet recognising dying is difficult and predicting dying frequently inaccurate.

Methods:

Urine samples from patients (n=112 and 49, training and validation cohorts, respectively)) with lung cancer were analysed using high resolution mass spectrometry. Cox lasso regression was engaged to develop a multivariable model predicting the probability of survival within the last 30 days of life. ANOVA and volcano plot analysis demonstrated metabolites that changed in the last weeks of life. Further analysis identified potential biological pathways affected.

Findings:

A model predicting time to death using 7 metabolites had excellent accuracy in the training

cohort (AUC = 0⋅85, 0⋅85, 0⋅88 and 0⋅86 on days 5, 10, 20 and 30) and validation cohort (AUC = 0⋅86, 0⋅83, 0⋅90, 0⋅86 on days 5, 10, 20 and 30). The model enabled classification of patients at low, medium and high risk of dying on a Kaplan-Meier survival curve. 124 metabolites changed. ANOVA analysis identified 93 metabolites and volcano plot analysis 85 metabolites. 53 metabolites changed using both approaches. Pathways altered in the last weeks included those associated with decreased oral intake, muscle loss, decreased RNA and protein synthesis, mitochondrial dysfunction, disrupted β-oxidation and one carbon metabolism. Epinephrine and cortisol increased in the last 2 weeks and week respectively.

Interpretation:

Metabolomic analysis identified metabolites and their associated pathways that change in the last days of life in patients with lung cancer. Prognostic tests, based on the metabolites identified in this study, could aid clinicians in the early recognition of people who may be dying, and have the potential to influence clinical practice and improve the care of dying patients.

Funding

Wellcome Trust Seed Award for Science; North West Cancer Research; University of Liverpool Enterprise Innovation Fund; Novo Nordisk Foundation.

Word count 300/300

# Research in Context

**Evidence before this study**

Accurately recognising that a person may be dying and in the last weeks or days of life is central to improving peoples’ experience of care. However, physicians’ predictions of dying are frequently inaccurate and overoptimistic. A recent comparison of five validated prognostic tools showed the best model was only as accurate as expert multidisciplinary clinician judgement. Therefore, an objective prognostic model that better discriminates when people are entering the last days of life is needed to influence clinical practice and improve patient care.

**Added value of this study**

There is a process to dying. Metabolites generated can be modelled to predict the dying phase in patients with lung cancer.

**Implications of all the available evidence**

Prognostic tests, based on the metabolites identified in this study, could aid in the early recognition of people who may be dying, and have the potential to influence clinical practice and improve the care of dying patients.

# INTRODUCTION

Accurately recognising that a person may be dying and in the last weeks or days of life is central to improving peoples’ experience of care. It enables families, medical teams and health-care providers to plan and provide the best care possible. However, physicians’ predictions of dying are frequently inaccurate and overoptimistic 1. The United Kingdom National Audit of Care at the End of Life of hospital deaths in 2019 found the recognition of the dying was challenging. When dying was recognised, 20% of people died within 8 hours; the median time to death was 36 hours; and importantly 50% of patients lacked the capacity to be directly involved in any decision-making 2.

Globally, there were 19⋅3 million new cancer cases and almost 10 million cancer deaths in 2020; lung cancer had the highest mortality, responsible for 1⋅8 million deaths 3. Predicting when a patient with advanced cancer is likely to die is a challenge and currently no diagnostic test is available. Several validated prognostic tools aim to predict survival of patients with advanced cancer 4. A recent comparison of five validated prognostic tools showed the best model, PiPs-B, was as accurate as expert multidisciplinary clinician judgement 5. However, existing models only consider a binary outcome of death from a particular time point e.g. 30 days. An objective model estimating risk of death over a range of time periods including the last days of life is needed.

We do not know how people die from cancer. In the last two weeks of life, there is evidence for deranged respiratory and renal function variables 6, although few patients have evidence of organ failure. Pulmonary embolus and infection are thought to be the major causes of death based on post-mortem studies 7,8. However, it is unusual for people with cancer to die suddenly as anticipated from a pulmonary embolus. About a third of patients with advanced cancer admitted to specialist palliative care units have a femoral deep vein thrombosis. Therefore, thromboembolism is considered a manifestation of advanced disease, rather than a cause of premature death 9. There is also a difference between the physiological deterioration leading to death in the acutely unwell patient compared to people dying from cancer; there is no evidence of sepsis in those with cancer 6. This suggests cancer patients generally do not die from organ failure and in those that die with an infection or pulmonary embolus; they die with them, rather than from these events.

A systematic review of biomarkers associated with dying identified common themes in cancer patients, irrespective of the type of malignancy: systemic inflammation, haematological changes, cachexia, hepatic dysfunction, renal dysfunction, and electrolyte changes. 10. Given the common features shared in patients dying from cancer, a "dying process" has been proposed 10 but has not yet been described.

We hypothesised a dying process is associated with metabolic changes. Here, we developed a model to predict dying based on urinary metabolites: training and validation data are presented. Some of the metabolites are known to be involved in particular pathways and the relevant pathways are described and discussed.

# METHODS.

## Setting and Participants

The study was conducted at eight sites (hospitals and hospices) in the North West of England (UK) from June 2016 to March 2020. Ethical approval was provided by North Wales (West) Research Ethics Committee (REC reference 15/WA/0464). Patients with incurable lung cancer were recruited prospectively and multiple urine samples were collected up to three times a week as previously described 11. Each sample was retrospectively assigned a day before death when this was known. In order not to bias the results, only one sample per patient, the one closest to death, was included for analysis. The validation cohort was recruited after the initial training cohort and analysed over 2 years after the training cohort.

## LC-QTOF-MS Method

Untargeted urine metabolomic studies were performed using liquid chromatography quadrupole time of flight mass spectrometry (LC-QTOF-MS) 12. In brief, analysis was performed on an Agilent 1290 Infinity LC coupled to an Agilent 6550 QTOF-MS equipped with a dual AJS electrospray ionization source (Agilent, UK), using two separate chromatographic methods (for further details see the Supplementary Appendix).

Targeted feature extraction was performed on each dataset based on matching of metabolite chemical features againstan in-house compound library accurate mass and retention time (AMRT) database that included a broad range of metabolites involved in intermediary metabolism. This was previously generated from analysis of the IROA Technology MS compound library of standards by each LC method described above, combined with the same QTOF analytical parameters used in this study 12 (databases publicly available: <https://doi.org/10.6084/m9.figshare.c.4378235.v2>). In addition to accurate mass and retention time MS/MS was also used in the confirmation of metabolite identity (i.e. level 1 identification) as per Sumner et al 13. Compound identification of unknown metabolites in our model was attempted using SIRIUS software (version 5.6.3) tools including SIRIUS molecular formula identification and CSI:FingerID fingerprint prediction.

## Statistical analysis

Data were normalised by probabilistic quotient normalisation (PQN) and glog transformed as previously described 14. Missing values were replaced by one half of the minimum positive value for each variable . Unequal variance was assumed. Principal component analyses and volcano plots were performed using MetaboAnalyst 15. Volcano plots measured differences in compound abundance: fold change (FC>2, FC<-2) and p<0⋅05 (false discovery rate [FDR] adjusted) were considered significant. Analysis of Variance (ANOVA) to look for difference in metabolite abundance in the last weeks of life compared to controls was performed using RStudio version 2021.09.0 16. Although multiple samples were collected from patients, only the final sample was included in the analysis.

## Prediction modelling

PROBAST guidelines for reporting prediction models were followed 17. A Cox proportional hazards model with lasso penalty to derive a prediction model was used for assessing the last days of life in our cohort 18. This approach is similar to the standard Cox model but shrinks parameter estimates towards zero, reducing over-fitting due to the large number of potential metabolites to consider as possible predictors of death. Administrative censoring was applied if the individual was still alive 30 days after their sample was supplied.

A penalty parameter (lambda) was imposed to determine the amount of smoothing chosen when 10-fold cross validation was performed. The value of lambda that gave minimum mean cross-validated error was used for both the prediction model and internal validation.

The model was internally validated using bootstrap resampling methods with 1000 bootstrap samples. The penalty parameter was fixed from the original Cox lasso model to fit to the whole dataset, and then, for each bootstrap sample, a Cox lasso model was fitted and time-dependent area under curve was calculated. Model calibration was assessed with each bootstrap sample by comparing the observed and expected survival probabilities, splitting the predicted risks into 3 groups (denoted low/medium/high survival). The validation cohort was externally validated. Calibration was performed at 10, 20 and 30 days. Kaplan-Meier curves were used to visually inspect the survival probabilities based on 30-day predicted risk. Log-rank tests were used to statistically compare the survival curves. Analysis was performed in R Studio version 1.4.1717 and used the packages “glmnet”, “survival”, and “hdnom” 16.

## Pathway Analysis

The pathway analysis was undertaken using MetaboAnalyst version 5⋅0 15 and combines pathway enrichment analysis with pathway topology analysis to identify the most relevant pathways involved with the conditions under study . Metabolites that showed a significant difference between groups were collated into relevant human KEGG physiological pathways to visualise which were altered towards dying. Metabolites were matched using the publicly available HMDB, PUBCHEM and KEGG databases.

# RESULTS

## Patients

We prospectively collected multiple samples of urine from patients with lung cancer. There were two independent cohorts; 112 patients in the training cohort and 49 patients in the validation cohort (Table 1). Our analysis was in four parts, i) reporting metabolites that change in the last weeks of life, ii) developing a training model to predict death, iii) validating the model in a separate cohort of patients and iv) identifying biochemical pathways affected in the last weeks of life.

### Table 1

**Clinical characteristics of the patients in the training cohort and the validation cohort included in this study. † NSCLC Non-small cell lung cancer; ‡ SCLC Small cell lung cancer; § Based on Multidisciplinary Team discussion**

|  |  |  |
| --- | --- | --- |
| **Patients** | **Training cohort** | **Validation cohort** |
|  | **Absolute number**  **/ 112 (%)** | **Absolute number**  **/ 49 (%)** |
| **Sex**  Female : male | 45 (40) : 67 (60) | 27 (55) : 22 (45) |
| **Diagnosis**  NSCLC† (Adenocarcinoma)  NSCLC† (Squamous)  NSCLC† (Unspecified)  SCLC‡  Mesothelioma  Radiological Diagnosis§ | 35 (31)  27 (24)  -  22 (20)  3 (3)  25 (22) | 2 (4)  1 (2)  18 (37)  11 (22)  2 (4)  15 (31) |
| **Age (years)**  Median (range)  40-49  50-59  60-69  70-79  80-90  >90 | 71 (47-89)  4 (4)  14 (13)  32 (29)  41 (37)  21 (19)  - | 71 (48-94)  1 (2)  4 (8)  14 (29)  19 (39)  10 (20)  1 (2) |
| **Ethnicity**  Mixed - White & Black African  White – British | 1 (1)  111 (99) | -  49 (100) |
| **Current Smoker status**  Ex-smoker  Current  Never  Unknown | 19 (17)  28 (25)  65 (58)  - | 5 (10)  13 (27)  29 (59)  2 (4) |
| **Time of urine sample in relationship to death (weeks/month before death)**  Week 1  Week 2  Week 3  Week 4  Month 2-3  Month >3 | **Samples analysed**  26 (23)  18 (16)  12 (11)  2 (2)  13 (12)  41 (37) | 13 (27)  8 (16)  5 (10)  5 (10)  1 (2)  17 (35) |

## Metabolites that change in the last weeks

Our first aim was to demonstrate that metabolites change in the last weeks of life. Using the Training cohort (n=112) we applied two approaches; ANOVA and volcano plot analysis. In total, 124 metabolites changed. ANOVA analysis identified 93 metabolites that varied in abundance. The abundance of 87 metabolites differed significantly in the last week of life (week 1) compared to week 2, week 3, week 4, week 4+ (weeks 4-11) or week 12+ (see summary Table S1 and box-plots in Figure S1 in the Supplementary Appendix). Volcano plot analysis identified 85 metabolites with a greater than 2-fold change for different time intervals; the last 4 weeks of life (0-4 weeks), last 2 weeks (0-2 weeks), last 5 days (0-5 days) and last 3 days (0-3 days) (see Tables S2 and S3). Some metabolite changes were especially large in the last 3 days, for example, creatine 15 fold and sarcosine 17 fold. 53 metabolites were demonstrated to change by both approaches. A table of the molecular mass and retention times (RT) for the Unknown metabolites is in Table S4.

## PREDICTION MODEL – Training cohort

Using a Cox lasso regression approach on our training set (n=112), we derived a multivariable model predicting the probability of survival for each day in the last 30 days. The model utilized seven metabolites (Table 2). Kaplan-Meier survival curves were plotted for patients classified as low, medium and high risk of dying (see Figure 1 A, log rank test p<0⋅001). The low risk group predicted those who are unlikely to be in the last 4 weeks of life; 3% (1/37) died by day 10 and 5% (2/37) by days 20 and 30. The high-risk group predicted those in the last days of life; 59% (22/37) died by 10 days, 89% (33/37) by 20 days and 92% (34/37) by 30 days. The 30-day model had excellent AUC values for every day in the last 30 days; for example, 0⋅85, 0⋅85, 0⋅88 and 0⋅86 on days 5, 10, 20 and 30 (see Figure 1 C). Calibration of the model at days 30, 20 and 10 was good (Figure S2). Compound identification of unknown metabolites in our model was attempted using SIRIUS software. This approach did not yield confident chemical structure identifications based on acquired MS/MS fragmentation spectra but did indicate clear molecular formula predictions for UBRM 4 (C11H14N4O3)and Unknown Metabolite 7 (C11H18N2O6). ANOVA plots of five of the metabolites identified from the prediction model analysis is shown in Figure 2.

### Table 2

**Table of metabolites for the 30-day Cox lasso logistic regression model and the corresponding Hazard Ratio.**

† UBRM 4 Unidentified Bone Derived Metabolite with molecular mass 250⋅1066 and retention time 1⋅54.

‡ Unknown Metabolite 5 with molecular mass 188⋅0472 and retention time 6⋅77

§ Unknown Metabolite 7 with molecular mass 274⋅1148 and retention time 3⋅48

**Metabolite Hazard Ratios**

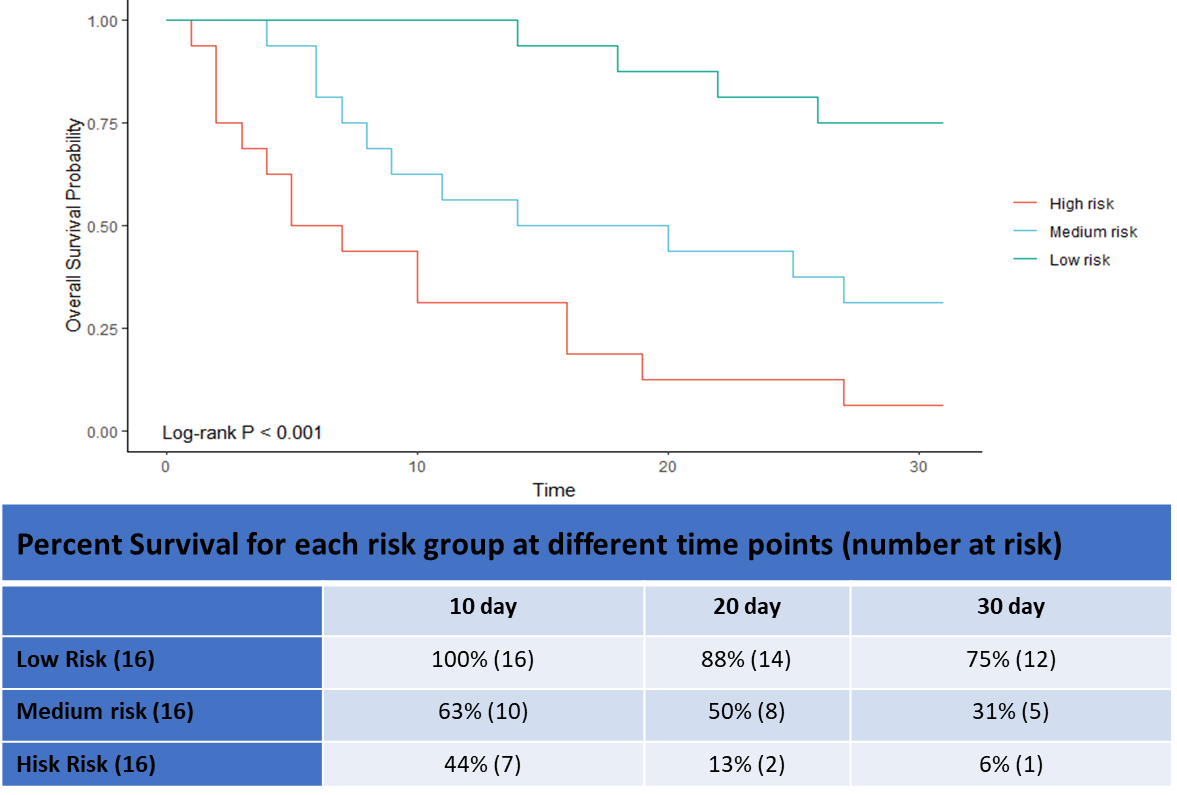
|  |  |  |  |
| --- | --- | --- | --- |
| 1 | Creatine |  | 1⋅66 |
| 2 | Gluconic acid |  | 0⋅75 |
| 3 | Carnitine |  | 1⋅18 |
| 4 | Pyrocatechol |  | 1⋅05 |
| 5 | UBRM 4† |  | 0⋅97 |
| 6 | Unknown Metabolite 5‡ |  | 1⋅59 |
| 7 | Unknown Metabolite 7§ |  | 1⋅05 |

### Figure 1 (A)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |



### Figure 1 (B)

****

### Figure 1 (C)

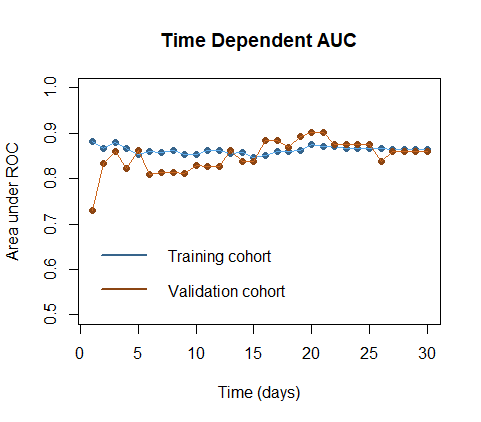
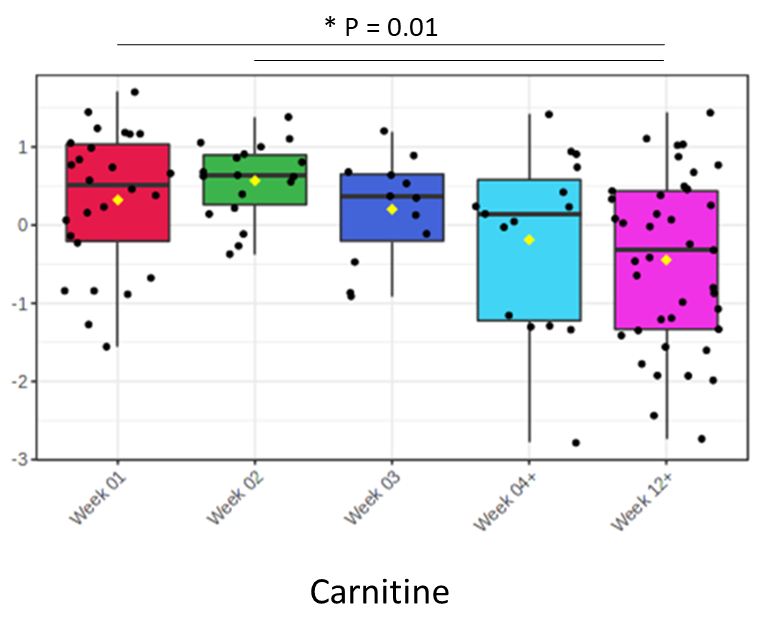


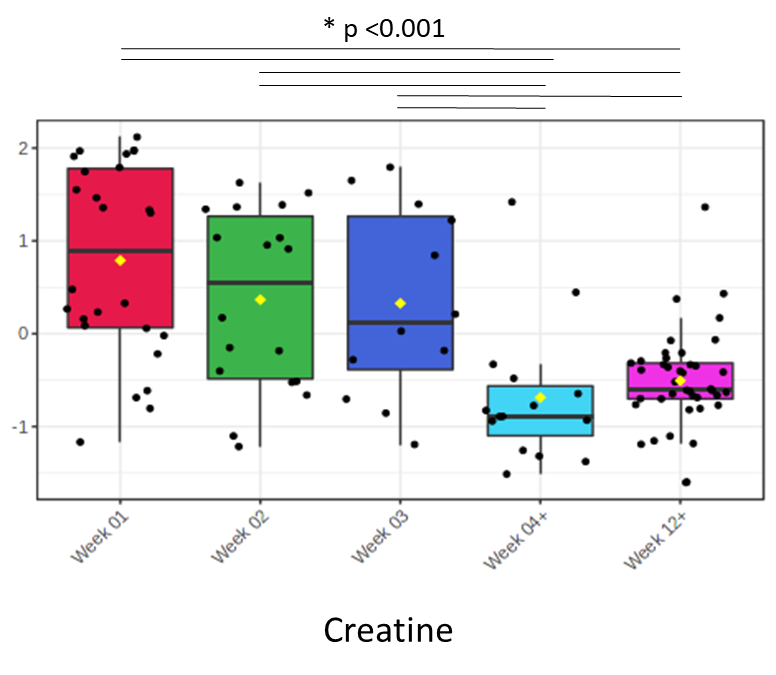
Figure 1

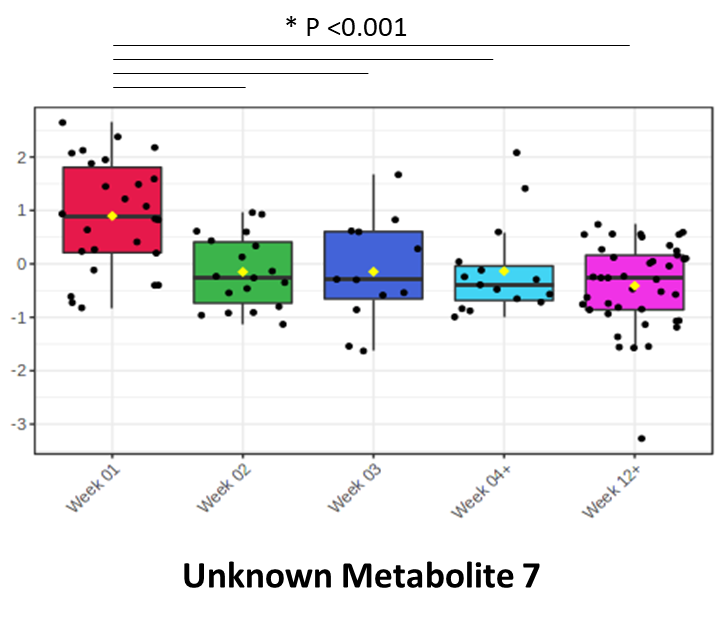
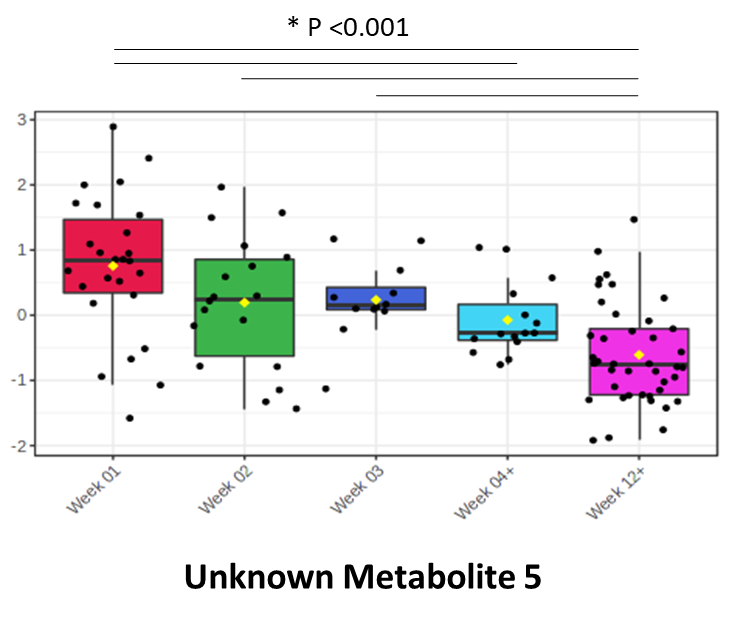
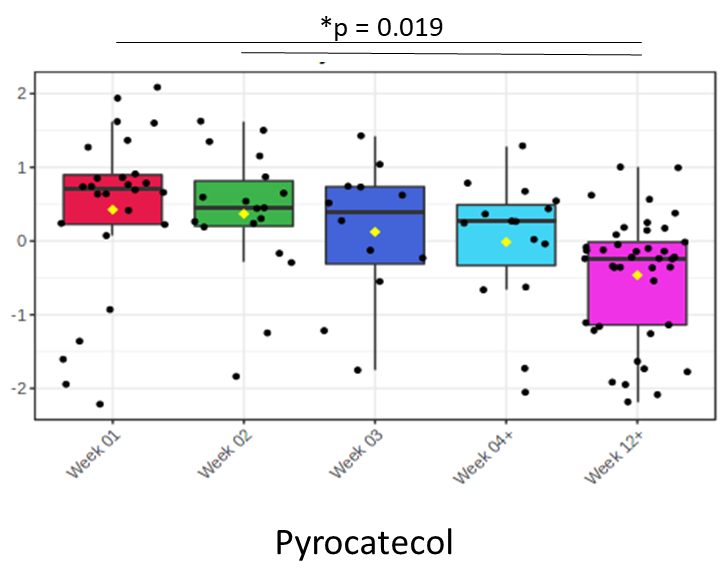
**Kaplan-Meier survival curves a using the 30-day Cox lasso logistic regression model for the Training (A) and Validation (B) cohorts showing High, Medium and Low risk of dying. The Time dependent AUC (C) comparing both cohorts for each day in the last 30 days of life.**

### Figure 2

**ANOVA plots for each metabolite from the prediction model identified as significant (**i.e. **FDR adjusted p-value <0⋅05.).** **Data were normalised by reference feature, g log transformed and auto scaled (centred around the mean and divided by the standard deviation of each variable). Metabolites were identified from the LC-QTOF-MS protocol using a DC18 column in Positive mode.**

****

****

****

## PREDICTION MODEL – Validation cohort

We validated our model in an independent validation cohort (n=49) see Table 1. Thirty one patients were in the last month of life. Kaplan-Meier survival curves were plotted for patients classified as low, medium and high risk of dying (see Figure 1 B, log rank test p<0⋅001). The low risk group predicted those unlikely to be in the last 4 weeks of life; 0% (0/16) died by day 10, 12% (2/16) by days 20 and 25% (4/16) by days 30. The high-risk group predicted those in the last days of life; 56% (9/16) died by 10 days, 87% (14/16) by 20 days and 94% (15/16) by 30 days.The 30-day model had excellent AUC values for every day in the last 30 days; for example, for example, 0⋅86, 0⋅83, 0⋅90 and 0⋅86 on days 5, 10, 20 and 30 (see Figure 1 C). Calibration of the model at days 30, 20 and 10 was good (Figure S2).

## Pathway analysis

Finally, we explored pathways affected during the dying process. The metabolites identified from the training set are compatible with reduced flux though biosynthetic pathways or, in the case of degradative products, increased breakdown of pathway intermediates. The total number of metabolites identified, the number of these identified by KEGG and the number of metabolites used in the pathway analysis for the last 2 weeks and last 3 days from the different LC-QTOF-MS analysis approaches is shown in Table S5. KEGG pathway analysis, using the significant metabolites for the last 2 weeks and last 3 days of life, identified several associated biochemical pathways (Table S6). Pathways involved in the last weeks of life are summarised in Table 3.

### Table 3 Pathways involved in the last weeks of life

For each metabolite, a greater than 2 fold change in last 4 weeks, 2 weeks, 5 days and last 3 days of life identified by volcano plot analysis is presented in parenthesis. \* The change is shown in the ANOVA graphs, Figure S1. ‡Disturbed represents a pathway where some metabolites increased and some metabolites decreased in abundance.

|  |  |  |
| --- | --- | --- |
| **Process** | **Metabolites identified**  **(Trend ↑/↓)**  **Fold change ( 4 week, 2 week, 5 day, 3 day)** | **KEGG biochemical pathway**  (P < 0⋅05) |
| **Energy metabolism** | Globally:  NAD↑ (-, -, 2⋅8, 3⋅1)  In muscle:  Creatine↑ (8⋅2, 9⋅2, 13⋅4, 15⋅3)  Creatinine↓\* | Increased  Nicotinate and nicotinamide metabolism  Disturbed‡  Glycolysis / Gluconeogenesis  Pyruvate metabolism |
| **Disrupted mitochondrial fatty acid β-oxidation** | Increased dicarboxylic acids:  Adipic acid ↑ (2⋅3, 2⋅7, -, -)  Ethylmalonic acid ↑ \*  Oxadipic acid ↑ (-, -, -, 2⋅1)  Suberic acid ↑ (-, 2⋅2, -, -)  ↑Dodecanoylcarnitine (-, 2⋅2, 2⋅6, 3⋅0)  ↑12-Hydroxy dodecanoic acid (-, -, 2⋅0, -) | Increased  Synthesis and degradation of ketone bodies |
| **Mitochondrial dysfunction** | Carnitine↑ (-, 2⋅2, -, -)  Hydroxy-3-methyl-glutaric acid↑ (-, -, 2⋅0, 2⋅0) |  |
| **Decreased RNA synthesis** | Uridine monophosphate (UMP) ↑ (-, -, 2⋅9, 3⋅3) |  |
| **Decreased protein synthesis** | Increased essential amino acids:  Histidine↑\*  Isoleucine ↑\*  Phenylalanine↑\*  Decreased RNA synthesis to generate ribosomes:  Uridine monophosphate (UMP)↑ (-, -, 2⋅9, 3⋅3) | Disturbed‡  Aminoacyl-tRNA biosynthesis |
| **Altered 1-carbon metabolism** | Cystathione↑ (5⋅1, 5⋅8, 6⋅5, 7⋅1)  Dihydrofolic acid (DHF)↑ \*  Sarcosine ↑ (8⋅4, 9⋅5, 14⋅3, 17⋅2)  (dTMP not detectable by analysis) | Increased  One carbon pool by folate |
| **Altered nucleoside (purine, pyrimidine) synthesis** | Adenine ↑\*  5-Deoxy, 5-methyl-thio-adenosine ↑\*  Dihydrofolic acid ↑\*  Guanosine ↑ (-, -, 2⋅6, 2⋅7)  Galactose monophosphate ↑\*  Hypoxanthine ↑  \*  Uridine monophosphate (UMP) ↑ (-, -, 2⋅9, 3⋅3)  Xanthine ↑\*  3-Methyl adenine ↓\* | Increased  Purine metabolism  Pyrimidine metabolism  Decreased  Pyrimidine metabolism  Disturbed‡  Amino sugar and nucleotide sugar metabolism |
| **Muscle loss/damage** | 3-Amino-isobutanoic-acid↑\*  3-Carboxypropyl-trimethylammonium↑ (7⋅1, 4⋅4, 6⋅5, -)  Carnitine↑ (-, 2⋅2, -, -)  Carnosine↑\*  Creatine↑ (8⋅2, 9⋅2, 13⋅4, 15⋅3)  Nitrotyrosine↑\*  Sarcosine↑ (8⋅4, 9⋅5, 14⋅3, 17⋅2)  Creatinine↓\*  In addition:  Metabolites of cell membrane breakdown |  |
| **Oxidative stress** | Nitrotyrosine↑\* |  |
| **Cell membrane breakdown / turnover** | Phosphatidylcholine↑\* |  |
| **Altered hormone production:** | Increased  Allotetrahydrocortisol ↑\*  Cortisol ↑\*  Cortexolone ↑\*  Epinephrine ↑ (3⋅3, 3⋅7, 4⋅4, 4⋅8)  Histidine ↑\*  Hydroxytryptophan ↑ (-, -, -, 2⋅2)  Pyrocatechol ↑ (3⋅0, 3⋅3, -, 4⋅6)  Vanillylmandelic acid (VMA) ↑ (-, -, 2⋅3, 2⋅4)  Decreased  Adrenochrome ↓ (-, 2⋅4, -, 3⋅0)  Cortisol-21 acetate ↓ (-, -, 2⋅1, 2⋅2) Dihydroxyphenylacetic acid ↓ (-, -, -, 2⋅2)  L-Dopa ↓ (-, -, 2⋅6, 3⋅2)  Serotonin ↓ (-, -, -, 2⋅3) | Increased  Steroid hormone biosynthesis |
| **Altered amino acid metabolism** | Increased  Histidine ↑\*  Hydroxytryptophan ↑ (-, -, -, 2⋅2)  Isoleucine ↑\*  Kynurenic acid ↑ (-, -, 3⋅0, 3⋅3)  Nitrotyrosine ↑\*  Phenylalanine ↑\*  Phenyl acetate ↑ (-, 2⋅9, 3⋅9, 4⋅4)  L-Tyrosine-methyl-ester-4-sulfate ↑ (-, 2⋅2, -, -)    Decreased  4-Hydroxyphenyllactic acid ↓ (2⋅5, 3⋅5, 4⋅3, 4⋅3)  Tyrosine methylene ketone ↓\* | Increased  β-Alanine metabolism  Alanine, aspartate and glutamate metabolism  Arginine biosynthesis  Arginine and proline metabolism  Cysteine and methionine metabolism  Glycine, serine and threonine metabolism  Histidine metabolism  Lysine degradation  Phenylalanine, tyrosine and tryptophan biosynthesis  Taurine and hypotaurine metabolism  Tryptophan metabolism  Decreased  Tyrosine metabolism  Valine, leucine and isoleucine degradation  Disturbed‡  D-Glutamine and D-glutamate metabolism  Tyrosine metabolism |
| **Kynurenine pathway activation** | Kynurenic acid ↑ (-, -, 3⋅0, 3⋅3)  3-Hydroxyanthranilic acid ↓ (-, 2⋅1, 2⋅8, 2⋅8) |  |
| **Decreased oral intake** | Caffeic Acid ↓ (-, 2⋅5, 4⋅4, 5⋅3)  Caffeine ↓ (2⋅4, 2⋅5, 3⋅4, 3⋅9)  Ferulic acid ↓ (-, 2⋅6, 2⋅7, 4⋅5)  Gluconic acid ↓\*  Paraxanthine ↓ (2⋅5, 2⋅8, 5⋅2, 5⋅4)  Quinic acid ↓ (2⋅5, 2⋅4, 3⋅6, 4⋅6)  Rosmarinic acid ↓ (2⋅1, 2⋅4, 2⋅4, 2⋅3)  Tartaric acid ↓ (-, -, 6⋅2, 7⋅2)  Theobromine ↓ (-, 2⋅2, 3⋅3, 4⋅8)  Theophylline ↓ (2⋅1, 2⋅3, 3⋅2, 3⋅8)  Trigonelline ↓ (2⋅2, 2⋅3, 3⋅2, 3⋅8) |  |

# DISCUSSION

Our work describes an objective model predicting dying based on urinary metabolites and represents the first attempt using a metabolomic approach to describe the dying process. The model was validated on an independent cohort of patients and found to be reproducible. Using a risk score, we were able to categorize patients at risk of dying. A test predicting the last days of life is important; by decreasing clinical uncertainty, it will support clinical practice and improve the care of dying patients 19, enabling families, medical teams and health-care providers to plan and provide the best care possible.

The Cox lasso derived prediction model demonstrates it is possible to use urine metabolites to predict the dying process for each day within the last 30 days of life with good accuracy in both the training and validation cohorts. The high risk score predicted the majority of patients imminently dying. The low risk score identified those not in the last two weeks of life. The model calibration was stable at 30, 20 and 10 days; slightly overestimating those likely to die with a high risk score and underestimating dying with a low risk score. External validation in an independent cohort had excellent AUC values. To our knowledge, it is the only model predicting dying across a range of time points including the last 2 weeks, with an AUC of 0⋅86 and 0⋅83 on days 5 and 10 in the validation cohort. Further work with increased numbers of patients and samples will calibrate our model closer to dying and potentially greater than 30 days.

Our model is only based on 7 metabolites and suggests the involvement of different processes. Creatine and carnitine are both muscle metabolites suggesting increasing muscle damage or cachexia. Carnitine may also indicate disrupted mitochondrial fatty acid β-oxidation and or mitochondrial dysfunction. Gluconic acid, abundant in plants, decreases and suggests decreased oral intake. UBRM 4, a metabolite identified from the *in vitro* growth of osteoclasts on dentine is a breakdown product of bone collagen, suggests bone resorption pathways are affected. Two of the other metabolites are unknown and importantly not identified from a library of 619 intermediary metabolism metabolites or from their accurate mass and retention time on MS/MS despite checking up to date international databases.

A recent comparison of five validated end of life prognostic tools showed the best model, PiPS-B, was as accurate as expert multidisciplinary clinician judgement; the overall accuracy was 61% 20. The PiPS-A model obtained a C-statistic of 0.825 (0.803-0.848) and the PiPS-B model obtained 0.837 (0.810 – 0.863). Whilst not the same as the time-dependent AUC we calculated (which accounts for censoring of observations) it is very similar. Our model obtained slightly higher AUCs in both the training and validation cohorts and gave accurate prediction across a range of time points (between 1 and 30 days).

Towards the end of life there are changes in energy availability; within muscle, increased creatine and decreased creatinine levels suggest a shortage of available phosphocreatine for muscular energy, likely resulting in a shortage of available high energy phosphate (ATP); increased NAD+ further supports the notion of reduced energy availability; as does evidence of disrupted mitochondrial fatty acid β-oxidation indicated by increased urinary dicarboxylic acids; and possible mitochondrial dysfunction, suggested by increased urinary carnitine (essential for fatty acid metabolism), methylglutaric acid and hydroxyl-3-methyl-glutaric acid. We previously demonstrated an increase in acetone, produced during mitochondrial fatty acid β-oxidation, towards the end of life 14.

Numerous changes suggest altered nucleic acid metabolism. Several critical building blocks accumulate, in particular UMP essential for RNA synthesis, as well as adenine and guanosine. In addition, purine degradation products xanthine, and hypoxanthine increase. Collectively these suggest depleting pools of substrates required for nucleic acid anabolism. Crucially this would suppress ribosomal biogenesis, impacting upon cellular capacity for protein synthesis, and importantly cellular stress monitoring and cell viability 21. In addition, ribosome synthesis makes high demands on cellular energy resources which appear to be low given the increase in NAD+. Other changes also likely impact nucleic acid metabolism; alterations in one carbon metabolism, indicated by accumulating dihydrofolic acid, sarcosine, and cystathionine, suggest a reduced ability to generate nucleotides for DNA synthesis. In addition, there will be reduced capacity to deal with reactive oxygen species through decreased production of sulfhydryl-containing reducing agents. Altered one carbon metabolism is highly associated with aging 22, however the consequences of changes in these processes are difficult to predict 23.

There was evidence of muscle damage or breakdown and an increase in amino acids starting approximately three weeks before death. Interestingly, increased muscle protein breakdown and efflux of amino acids is a fundamental response seen in critical illness 24. Numerous hormones, including their essential intermediates were altered; cortisol, epinephrine, histamine and hydroxytryptophan increase; dopamine and serotonin decrease. Approximately 5% of tryptophan is converted to serotonin, therefore the low serotonin levels and increased kynurenine levels imply there is a diversion 25. Cortisol was previously shown to be increased in dying patients including a cohort of patients with lung cancer 26. In critically ill patients, increased cortisol was shown to be related to decreased breakdown in the liver 27.

Our work was developed from a homogenous cancer population i.e. lung cancer. Although any group of lung cancer patients is a population of mixed cancers; adenocarcinoma, squamous and small cell carcinomas predominantly. Our validation cohort was recruited and analysed after the training cohort. Our objective model is only based on seven urinary metabolites and it prognosticates for every day within the 30 days of life with excellent AUC values. The model validated very well. To our knowledge, it is the only model that predicts dying across a range of time points (between 1 and 30 days).

There are a number of limitations to our study. Our training and validation cohorts were based on small sample sizes and the limited number of samples in the last month of life excluded longitudinal analysis. While lung cancer is a heterogeneous group of cancers our work may not represent the dying process of people with other cancers. Our population was homogenous and therefore not a broad cultural cross-section of patients. Two of the metabolites in the prediction model were unknown despite a library with a broad range of metabolites involved in intermediary metabolism and determining an accurate mass and retention time MS/MS for confirmation. We gave an overview of candidate metabolic markers of dying and discussed relevant pathways involving urinary metabolites. While most metabolites are filtered passively from blood to urine, the ideal however is to use blood to discuss relevant biochemical pathways. The hazard ratios for the model predicting dying were derived from semi-quantitative data 28. The training and validation datasets were raw peak area data only i.e. not exact metabolite concentrations. Therefore, the metabolite abundances are only meaningful in the context of each dataset as a whole i.e. the relative abundances between dying and non-dying patients. An important part of translating this model will be to establish concentration reference ranges for these metabolites in dying versus not dying patients. Empirically there are different time frames of dying, which is an added complication to predicting the last days of life. When patients with cancer are recognised to be dying (‘actively dying’) in a hospice setting, they die over two to three days. However, some die within 24 hours and some take longer than a week. By recruiting greater numbers of patients with lung and other solid tumors in the last weeks of life and identifying those cohorts that die in these different ways, we may improve the accuracy of our model. Further work is needed to develop a widely applicable robust clinical tool.

Our work describes an objective validated model predicting dying based on urinary metabolites and represents the first attempt using a metabolomic approach to describe the dying process. Accurate prognostic information at the end of life is essential to co-ordinate and manage care in response to need, whilst avoiding burdensome and unnecessary interventions. The early recognition a person may be dying is central to all the priorities for improving peoples’ experience of care in the last days and hours of life. Prognostic tests, based on the metabolites identified in this study, could aid in the early recognition of people who may be dying, and therefore have the potential to influence clinical practice and improve the care of dying patients.

# Acknowledgements

We would like to acknowledge funding from the Wellcome Trust Seed Award for Science, North West Cancer Research, University of Liverpool Enterprise Innovation fund and Novo Nordisk Foundation. The Liverpool AKU Research Group (Liverpool University Hospitals Foundation Trust) & Prof. Ranganath for spectrometer access. The research nurses who recruited patients and collected samples (F Westwell, C Harrop, S Stanley, P Walker, S Barnes, G Hull, S Dealing, N Collins, M Leach), the Staff of the Liverpool University Hospitals Foundation Trust (including Heather Rogers); The Clatterbridge Cancer Centre; Marie Curie Hospice Liverpool; Willowbrook Hospice, Prescot; St Helens and Knowsley Hospitals Trust (including Jeanette Anders and Paula Scott) ; St Catherines Hospice, Preston; Trinity Hospice, Blackpool; the principal investigators for the study (Dr J Bellieu, Dr A Thompson, Dr A Pope, Dr L Chapman, Dr A Fletcher, Dr A Gadoud) and to E Wright and B Hughes our lay members for their invaluable input.

**Abbreviations**

ANOVA Analysis of variance

AUC Area under the receiver operating characteristic curve

FDR False discovery rate

LASSO Least absolute shrinkage and selection operator

PCA Principal component analysis

**Authors Contributions**

All authors were involved in critical review of the manuscript and have seen and approved the final version. Specific contributions as follows; Study conception and design: SC, SM, CRM, ACN, MB, JEE, CP; Sample acquisition: SC, CRM, SM, JEE; Sample analyses EC, RS, BPN, ASD, IR, DBK; Data analysis: EC, JB, RS, BPN, ASD, IR, DBK, SC, DH; Drafting the manuscript: SC, EC and CP, Revision of manuscript: SC, EC, RS, BPN, ASD, IR, JB, DH, ACN, CRM, SM, MB, DBK, JEE, CP.

**Funding Statement**

This research received a Wellcome Trust Seed award for Science (202022/Z/16/Z), North West Cancer Research award (SI2018.11), University of Liverpool Enterprise Investment Fund award and Novo Nordisk Foundation (grant NNF20CC0035580).

**Competing Interests**

None

**Data sharing statement**

The raw data and .csv files are uploaded on Metabolomics Workbench, Study ID ST002082 and available at <http://dx.doi.org/10.21228/M8TT44> 29.

.

# References

1. White N, Reid F, Harris A, Harries P, Stone P. A Systematic Review of Predictions of Survival in Palliative Care: How Accurate Are Clinicians and Who Are the Experts? *PloS one* 2016; **11**(8): e0161407.

2. National Audit of Care at the End of Life: First round of audit report. 2019.

3. Sung H, Ferlay J, Siegel R, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. . *CA Cancer J Clin* 2021; **71**: 209-49.

4. Simmons CPL, McMillan DC, McWilliams K, et al. Prognostic Tools in Patients With Advanced Cancer: A Systematic Review. *J Pain Symptom Manage* 2017; **53**(5): 962-70 e10.

5. Stone P, Kalpakidou A, Todd C, et al. Prognostic models of survival in patients with advanced incurable cancer: the PiPS2 observational study. *Health Technol Assess* 2021; **25**(28): 1-118.

6. Taylor P, Crouch S, Howell DA, Dowding DW, Johnson MJ. Change in Physiological Variables in the Last Two Weeks of Life: An Observational Study of Hospitalized Adults With Heart Failure. *J Pain Symptom Manage* 2018; **55**(5): 1335-40.

7. Abdel-Karim IA, Sammel RB, Prange MA. Causes of death at autopsy in an inpatient hospice program. *Journal of palliative medicine* 2007; **10**(4): 894-8.

8. Pautex S, Vayne-Bossert P, Jamme S, et al. Anatomopathological causes of death in patients with advanced cancer: association with the use of anticoagulation and antibiotics at the end of life. *Journal of palliative medicine* 2013; **16**(6): 669-74.

9. White C, Noble SIR, Watson M, et al. Prevalence, symptom burden, and natural history of deep vein thrombosis in people with advanced cancer in specialist palliative care units (HIDDen): a prospective longitudinal observational study. *Lancet Haematol* 2019; **6**(2): e79-e88.

10. Reid VL, McDonald R, Nwosu AC, et al. A systematically structured review of biomarkers of dying in cancer patients in the last months of life; An exploration of the biology of dying. *PloS one* 2017; **12**(4): e0175123.

11. Coyle S, Scott A, Nwosu AC, et al. Collecting biological material from palliative care patients in the last weeks of life: a feasibility study. *BMJ Open* 2016; **6**(11): e011763.

12. Norman BP, Davison AS, Ross GA, et al. A Comprehensive LC-QTOF-MS Metabolic Phenotyping Strategy: Application to Alkaptonuria. *Clin Chem* 2019; **65**(4): 530-9.

13. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007; **3**(3): 211-21.

14. Chapman E, Baker J, Aggarwal P, et al. GC-MS techniques investigating potential biomarkers of dying in the last weeks with lung cancer. *International Journal of Molecular Sciences*; 2023. p. 1591.

15. Xia J, Psychogios N, Young N, Wishart DS. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res* 2009; **37**(Web Server issue): W652-60.

16. RStudio Team. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA 2020.

17. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019; **170**(1): 51-8.

18. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw* 2011; **39**(5): 1-13.

19. Coyle S, Chapman E. Predicting the last days of life will change clinical practice. *BMJ supportive & palliative care* 2020.

20. Stone PC, Kalpakidou A, Todd C, et al. The Prognosis in Palliative care Study II (PiPS2): A prospective observational validation study of a prognostic tool with an embedded qualitative evaluation. *PloS one* 2021; **16**(4): e0249297.

21. Vlatković N BM, Rubbi CP. . Nucleolar control of p53: a cellular Achilles' heel and a target for cancer therapy. *Cell Mol Life Sci* 2014; **Mar;71**(5): 771-91.

22. Suh E C, S-W Friso S,. One-Carbon Metabolism: An Unsung Hero for Healthy Aging. In: Marco Malavolta EM, ed. Molecular Basis of Nutrition and Aging,: Academic Press,; 2016: 513-22,.

23. Maguire M, Nield PC, Devling T, et al. MDM2 Regulates Dihydrofolate Reductase Activity through Monoubiquitination. *Cancer Res* 2008; **May 1;68**(9): 3232-42.

24. Wolfe RR. Regulation of skeletal muscle protein metabolism in catabolic states. *Curr Opin Clin Nutr Metab Care* 2005; **8**(1): 61-5.

25. Cervenka I, Agudelo LZ, Ruas JL. Kynurenines: Tryptophan's metabolites in exercise, inflammation, and mental health. *Science* 2017; **357**(6349).

26. Sandberg AA, Eik-Nes K, Migeon CJ, Samuels LT. Metabolism of adrenal steroids in dying patients. *J Clin Endocrinol Metab* 1956; **16**(8): 1001-16.

27. Boonen E, Vervenne H, Meersseman P, et al. Reduced cortisol metabolism during critical illness. *The New England journal of medicine* 2013; **368**(16): 1477-88.

28. Liu X, Locasale JW. Metabolomics: A Primer. *Trends Biochem Sci* 2017; **42**(4): 274-84.

29. Sud M, Fahy E, Cotter D, et al. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res*2016; **4**(44): D463-70.