Plan Overview

A Data Management Plan created using DMPonline

Title: External Validation of Thirteen Survival Prediction Models for Patients with Metastatic Spinal Disease in The Netherlands

Creator: Harmen Kuijten

Principal Investigator: Jorrit-Jan Verlaan

Affiliation: UMC Utrecht

Template: UMC Utrecht DMP

ORCID iD: 0000-0001-8105-6660

Project abstract:

Prognostic models predicting survival in patients with metastatic cancer may aid physicians in clinical practice. To date, ten prediction models exist to predict survival in patients with metastatic spinal disease, but these models have numerous limitations.

The existing models were developed and validated with patient data ranging from 1986 to 2016. Current treatment therapies increase survival rates for metastatic spinal disease, which limits the applicability of some models. Also, the existing models were mainly validated using only surgically-treated patients, whereas non-surgical treatment alternatives such as stereotactic and particle radiation therapy have increased exponentially. Therefore, to determine their current clinical value, we intend to validate the existing prognostic models using a large prospective dataset of patients with spinal metastases treated with radiotherapy, surgery, or both.

Furthermore, all existing models focus on survival in metastatic spinal disease but the patient’s quality of life (QOL) after treatment has become the more important goal. For instance, some models indicate that patients with a short life expectancy should not undergo surgery, but minimally invasive surgery has such reduced patient demand that some patients may now very well profit from surgical intervention in terms of better QOL. Therefore, we intend to develop a model predicting QOL using the same retrospective cohort on which the existing models were validated. QOL questionnaires were prospectively collected for included patients in the cohort. Additionally, we will identify factors associated with QOL after treatment to guide new or to-be-updated prognostic models.

Determining the optimal treatment plan for metastatic spinal disease is based on the best trade-off between survival and QOL. Our future goal is to develop a well-performing open-access tool that can predict survival and QOL for metastatic spinal disease. The new tool may better facilitate shared decision-making for treating patients with spinal metastases.

ID: 104655

Start date: 27-10-2022

End date: 01-06-2023

Last modified: 25-11-2022

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

External Validation of Thirteen Survival Prediction Models for Patients with Metastatic Spinal Disease in The Netherlands

1. General features

1.1. Please fill in the table below. When not applicable (yet), please fill in N/A.

DMP template version	30 (don't change)
ABR number *(only for human-related research)*	
METC number *(only for human-related research)*	TBD
DEC number *(only for animal-related research)*	
Acronym/short study title	EVMSD
Name Research Folder	22-962_EVMSD
Name Division	Division Surgical Specialty, Orthopaedic Surgery
Name Department	Orthopedics
Partner Organization	
Start date study	27-10-2022
Planned end date study	01-06-2023
Name of datamanager consulted*	Dax Steins
Check date by datamanager	19-11-2022

1.2 Select the specifics that are applicable for your research.

- Retrospective study
- Non-WMO
- Monocenter study
- Use of Questionnaires
- Observational study

Quality of life will be assessed by varies questionnaires (i.e. Brief Pain Inventory, EQ-5D-3L, EORTC QLQ-BM22, EORTC QLQ-CP15-PAL, and SOSGOQ). The questionnaires are part of standard data collection and have already been acquired.

2. Data Collection

2.1 Give a short description of the research data.

For this retrospective cohort study, we shall use previously collected QoL questionnaires from patients with symptomatic metastatic spine disease presenting to the University Medical Center Utrecht (UMCU) in the Netherlands.

Clinical data will be reused from two different retrospective studies (METC 17-777/18-841/C), a monocentre registry (PRESENT; METC 13-361/C), and a multicentre AO Spine registry (MTRON 18-314/C), which include adult patients with spinal metastases treated with radiotherapy and/or surgery between 01-01-2016 and 01-11-2022.

TABLE

Potential subjects from which data can be used to perform the study (standard available data) on first three rows. Additional data will be requested through the UPOD for all potentially eligible patients (129 + 332 + 2100) =~2500 patients, fourth row. We estimate that 700-1000 patients are eligible for the study, fifth row.

Subjects	Volume	Data Source	Data Capture Tool	File Type	Format	Storage space
Human	129	eCRF	RedCap (MTRON)	Quantitative	.csv	0-10 GB
Human	332	eCRF	SPSS (Retrospective studies METC 17-777/18-841/C)	Quantitative	.csv	0-10 GB
Human	2100	eCRF	Castor (PRESENT)	Quantitative	.csv	0-10 GB
Human	~2500 (= all subjects from MTRON, PRESENT and the two retrospective studies)	EPD (HiX)	UPOD	Quantitative	.csv	0-10 GB
Human	700-1000 (eligible subjects)	.csv	Python	Quantitative	.csv	0-10 GB

2.2 Do you reuse existing data?

- Yes, please specify
- No, please specify

Yes: We do reuse data examples:

We use data previously collected in the following registries and retrospective studies:

Clinical data will be reused from two different retrospective studies (METC 17-777/18-841/C), data from these studies comes from patients with symptomatic spinal metastases who received surgery in the University Medical Center Utrecht between 2009-2018. The studies included 332 participants.

Metastatic Tumor Research and Outcomes Network (MTRON): registry for patients with symptomatic spinal metastases who received surgery in the University Medical Center Utrecht between 2018-now [18-314]. Currently, the registry has enrolled around 130 participants. The MTRON registry is the successor of the GSTSG registry.

Prospective Evaluation of Interventional Studies on Bone Metastases (PRESENT): registry of patients with symptomatic bone metastases who received radiotherapy in the University Medical Center Utrecht between 2013-now [13-261]. Currently, the registry has enrolled around 2200 participants.

No:

Additionally, variables not part of the standard data collection for each database will be collected with pseudonymized data from the Utrecht Patient Oriented Database (UPOD). The involved data-manager or PI will share the (non-pseudonymized) patient data with UPOD, and UPOD will return pseudonymized data to the investigators. All additional variables are part of standard care.

The variables **not** part of standard data collection include:

Retrospective studies:

Measurements: length, weight, body mass index.
Clinical variables: Charlson comorbidity, visceral metastasis, liver metastasis, lung metastasis, previous systemic therapy, previous spine radiotherapy, pathological fracture, disseminated metastasis.

Laboratory variables: calcium, creatinine, total bilirubin, alkaline phosphatase, LDH, albumin, CRP, hemoglobin, platelet count, white blood cell count, absolute neutrophil count, absolute lymphocyte count, international normalized ratio (or prothrombin time).

MTRON:

Clinical variables: liver metastasis, lung metastasis, previous chemotherapy, disseminated metastasis, amount of extra spinal metastasis
Laboratory variables: calcium, creatinine, total bilirubin, alkaline phosphatase, LDH, albumin, CRP, hemoglobin, platelet count, white blood cell count, absolute neutrophil count, absolute lymphocyte count, international normalized ratio (or prothrombin time).

PRESENT:

Clinical variables: ASA classification
Laboratory variables: calcium, creatinine, total bilirubin, alkaline phosphatase, LDH, albumin, CRP, hemoglobin, platelet count, white blood cell count, absolute neutrophil count, absolute lymphocyte count, international normalized ratio (or prothrombin time).

2.3 Describe who will have access to which data during your study.

Table 1a. regarding the two original registries MTRON and GSTSG

The datamanager is authorized to link different datasets of the selected patient group and thus has access to personal data such as patientID. The key table linking study specific IDs to patient IDs is available to the datamanager and members of the research team with a care relationship to the patient. Other members of the research team receive a pseudonymized dataset and have no access

to direct personal data or the key table.

Type of data	Who has access
Direct identifying personal data	Research team with care relationship to patient, Datamanager (surgical specialty)
Key table linking study specific IDs to Patient IDs	PI (with care relationship to patient), Datamanager
Pseudonymized data	Research team, Datamanager

Flow regarding table 1a.

The .csv file with pseudonymized data will be stored in the specific RFS map for this study.

Table 1b. regarding the original registry PRESENT

The datamanager from the imagin department is authorized to link different datasets of the selected patient group and thus has access to personal data such as patientID. The key table linking study specific IDs to patient IDs is available to the datamanager and members of the research team PRESENT. Other members of the research team (not the PRESENT research team) receive a pseudonymized dataset and have no access to direct personal data or the key table.

Type of data	Who has access
Direct identifying personal data	Research team of PRESENT, Datamanager (imaging)
Key table linking study specific IDs to Patient IDs	Research team of PRESENT, Datamanager
Pseudonymized data	Research team of PRESENT, Datamanager

Flow regarding table 1b.

The .csv file with pseudonymized data will be stored in the specific RFS map for this study.

Table 2. regarding the data used by UPOD.

2. Additional data will be collected from the Utrecht Patient Oriented Database (UPOD). Only UPOD datamanagers have direct access to the non-pseudonymized UPOD data. The data will be pseudonymized with a study-specific study ID. The research team gets the data extract, pseudonymized.

Type of data	Who has access
Direct identifying personal data	UPOD Datamanager
Key table linking study specific IDs to Patient IDs	PI (with care relationship to patient) for registries, PI (with care relationship to patient) for PRESENT, Datamanagers (imaging and surgical specialty), UPOD Datamanager
Pseudonymized data	Research team, Datamanagers, UPOD Datamanager

Flow regarding table 2.

The PI from MTRON and GSTSG or datamanger Surgical Specialty will provide the key link to the UPOD Datamanager, and the PI from PRESENT or datamanager Imaging will provide the key link to the UPOD Datamanager.

Important note:

The UPOD datamanger will generate new study specific IDs. However, it is important that the original pseudonymized ID's from the original datasets must be saved in the same key link table as the new study specific IDs. All involved datamanagers (UPOD/Surgical Specialty/Imaging) and PI's have access to this key.

2.4 Describe how you will take care of good data quality.

Research data from included subjects will be collected from varies data sources, namely an SPSS database (IRB number: 18-841, 17-695), RedCap (MTRON 18-314), and Castor EDC (PRESENT 13-261). These datasets will be merged and matched by study subject code, and expended with UPOD data. Data quality will be checked by members of the research team. Data collection will be frozen before analysis in Python.

#	Question	Yes	No	N/A
1.	Do you use a certified Data Capture Tool or Electronic Lab Notebook?		x	
2.	Have you built in skips and validation checks?		x	
3.	Do you perform repeated measurements?		x	
4.	Are your devices calibrated?			x
5.	Are your data (partially) checked by others (4 eyes principle)?	x		
6.	Are your data fully up to date?	x		
7.	Do you lock your raw data (frozen dataset)	x		
8.	Do you keep a logging (audit trail) of all changes?	x		
9.	Do you have a policy for handling missing data?	x		
10.	Do you have a policy for handling outliers?			x

2.5 Specify data management costs and how you plan to cover these costs.

#	Type of costs	Division ("overhead")	Funder	Other (specify)
1.	Time of datamanager	X		
2.	Storage	x		
3.	Archiving	x		

2.6 State how ownership of the data and intellectual property rights (IPR) to the data will be managed, and which agreements will be or are made.

Apart from the standard collected data for the MTRON registry, the UMC Utrecht is and remains the owner of all readily collected and newly collected data for this study.

The AO Foundation is the owner of standard collected data for the MTRON registry.

Apart from published data, the raw data will not be accessible for other researchers or institutions.

3. Personal data (Data Protection Impact Assessment (DPIA) light)

Will you be using personal data (direct or indirect identifying) from the Electronic Patient Dossier (EPD), DNA, body material, images or any other form of personal data?

- Yes, go to next question

I will process personal data. I have consulted the division data manager and I do not have to complete a full DPIA. I therefore fill out this DPIA light and proceed to 3.1.

3.1 Describe which personal data you are collecting and why you need them.

Which personal data?	Why?
Descriptives age, length, weight, body mass index, gender.	These are all the variables needed to use all thirteen prediction models.
Classifications American Society of Anesthesiologists (ASA) classification, Bilsky score, Spinal Instability Neoplastic Score (SINS), functional status, Eastern Cooperative Oncology Group (ECOG), American Spinal Injury Association (ASIA) Impairment Scale, Charlson comorbidity, Karnofsky Performance Scale (KPS), Frankel grading.	These are all the variables needed to use all thirteen prediction models.
Tumor characteristics previously received treatment, primary tumor histology, primary tumor location, visceral metastasis, brain metastasis, liver metastasis, lung metastasis, previous systemic therapy, previous spine radiotherapy, previous chemotherapy, amount of skeletal metastasis, amount of spinal metastasis, amount of extra spinal metastasis, pathological fracture, disseminated metastasis.	These are all the variables needed to use all thirteen prediction models.
Laboratory values calcium, creatinine, total bilirubin, alkaline phosphatase, LDH, albumin, CRP, hemoglobin, platelet count, white blood cell count, absolute neutrophil count, absolute lymphocyte count, international normalized ratio.	These are all the variables needed to use all thirteen prediction models.

3.2 What legal right do you have to process personal data?

- No objection, please explain

We will use both the study-specific informed consent and the no-objection.
Subject who participate in MTRON and PRESENT have signed broad informed consent for the use of their data for future reasearch. The no-objection will be applied to the subjects from MTRON, PRESENT and the two retrospective studies (i.e. passed away and a dataset that contains a large number of patients [>500]).
The datamanager from dHS will perform the no-objection check on 28 or 29 November 2022

3.3 Describe how you manage your data to comply to the rights of study participants.

1. The data of the two retrospective studies, MTRON and PRESENT are pseudonymized and the linking tables to personal data is saved. An authorized person (datamanager) manages the linking table, can re-identify study participants when necessary, and deliver, correct or delete the data.

Participants from MTRON and PRESENT signed broad informed consent for using their data in future studies. Subjects from the two retrospective studies, of which no informed consent is acquired, fall under the general-no-objection. They do not know they participated in the study.

Right	Answers
Right of Access	We have to refuse participant's right of access, because this would make the research impossible to conduct given the large number of participants (n=1000). Subjects from the two retrospective studies, of which no informed consent is acquired, fall under the general-no-objection. They do not know they participated in the study.
Right of Rectification	The authorized person will give the code for which data have to be rectified. Subjects from the two retrospective studies, of which no informed consent is acquired, fall under the general-no-objection. They do not know they participated in the study.
Right of Objection	Will be checked by the datamanager from dHS for all subjects
Right to be Forgotten	In the informed consent we state that the study participant can stop taking part in the research. Removal of collected data from the research database cannot be granted because this would result in a research bias. Subjects from the two retrospective studies, of which no informed consent is acquired, fall under the general-no-objection. They do not know they participated in the study.

3.4 Describe the tools and procedures that you use to ensure that only authorized persons have access to personal data.

We use the secured Research Folder Structure that ensures that only authorized personnel has access to personal data, including the key table that links personal data to the pseudoID.

3.5 Describe how you ensure secure transport of personal data and what contracts are in place for doing that.

We will not transport any personal data outside the UMCU network drives.

4. Data Storage and Backup

4.1 Describe where you will store your data and documentation during the research.

1. All files will be stored in the secured Research Folder Structure of the UMC Utrecht.
2. Also, a separate Castor is used to store the provided additional data by UPOD.

4.2 Describe your backup strategy or the automated backup strategy of your storage locations.

1. All (research) data is stored on UMC Utrecht networked drives from which backups are made automatically twice a day by the division IT (dIT).

5. Metadata and Documentation

5.1 Describe the metadata that you will collect and which standards you use.

1. For the data collected in SPSS, RedCap, and Castor, codebooks of the research databases are available in each tool.
2. We do not use metadata standards yet. The data from UPOD will be delivered including a data dictionary. For every variable this data dictionary contains an explanation of the values.

5.2 Describe your version control and file naming standards.

1. We will use GitHub as version control for our code (link to github: XXX).
2. We will keep track of changes using descriptions of changes per datestamp for each file in a separate Word document.
3. We will distinguish versions by indicating the version in the filename of the master copy by adding a code after each edit, for example V1.1 (first number for major versions, last for minor versions). The most recent copy at the master location is always used as the source, and before any editing, this file is saved with the new version code in the filename. The file with the highest code number is the most recent version and older versions are moved to a folder OLD. The major versions will be listed in a version document (projxVersDoc.txt), stating the distinguishing elements per listed version.

** I will update 'XXX' in this answer when available.*

6. Data Analysis

6 Describe how you will make the data analysis procedure insightful for peers.

1. We have written an analysis plan in which we state why we will use which data and which statistical analysis we plan to do in which software. The analysis plan is stored in the project folder, so it is findable for my peers.
2. We will be using Python, version 3.9, for statistical analysis of the data. The scripts will contain comments, such that every step in the analysis is documented and peers can read why I made certain decisions during the analysis phase.
3. I will make an overview of datasets and analysis scripts, such that it is fully clear how the statistical analysis is performed. Peers will be able to repeat the analysis based on my overview.

7. Data Preservation and Archiving

7.1 Describe which data and documents are needed to reproduce your findings.

1. The data package will contain: the raw data, the study protocol describing the methods and materials, the script to process the data, the scripts leading to tables and figures in the publication, a codebook with explanations on the variable names, and a 'read_me.txt' file with an overview of files included and their content and use.

After finishing the project, this documentation will be stored at the UMC Utrecht [path: L:\Onderzoek\Orthopaedie\22-962_EVSM] and is under the responsibility of the Principal Investigator of the research group.

7.2 Describe for how long the data and documents needed for reproducibility will be available.

Data and documentation needed to reproduce findings from this non-WMO study will be stored for at least 15 years.

7.3 Describe which archive or repository (include the link!) you will use for long-term archiving of your data and whether the repository is certified.

After finishing the project, the data package will be stored at the UMC Utrecht Research Folder Structure and is under the responsibility of the Principal Investigator of the research group. When DataverseNL is available, the data package will be published here.

7.4 Give the Persistent Identifier (PID) that you will use as a permanent link to your published dataset.

We will not publish our dataset.

8. Data Sharing Statement

8.1 Describe what reuse of your research data you intend or foresee, and what audience will be interested in your data.

1. My peers will be reusing all research data in the final dataset to generate new research questions
2. The raw data can be of interest for other researchers or for spin-off projects.

8.2 Are there any reasons to make part of the data NOT publicly available or to restrict access to the data once made publicly available?

- Yes (please specify)

As the data is privacy-sensitive, we publish the descriptive metadata in the data repository with a description of how a data request can be made (by sending an email to the corresponding author). In the event that peers like to reuse our data this can only be granted if the research question is in line with the original informed consent signed by the study participants. Every application therefore will be screened upon this requirement. If granted, a data usage agreement is signed by the receiving party.

8.3 Describe which metadata will be available with the data and what methods or software tools are needed to reuse the data.

Along with the publication, scripts of analysis in Python will be available.

8.4 Describe when and for how long the (meta)data will be available for reuse

- (Meta)data will be available as soon as article is published

8.5 Describe where you will make your data findable and available to others.

Data can become available to others upon request as described at 8.2