PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (http://bmjopen.bmj.com/site/about/resources/checklist.pdf) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Assessment of bone-targeting agents use in patients with bone metastasis from breast, lung, or prostate cancer using structured and unstructured electronic health records from a regional UK-based hospital
AUTHORS	Seesaghur, Anouchka; Egger, Peter; Warden, Joshua; Abbasi, A; Levick, Bethany; Riaz, Majid; McMahon, Peter; Thompson, Matthew; Cheeseman, Sue

VERSION 1 – REVIEW

	-
REVIEWER	Fan, Yu
	The Affiliated Hospital of Southwest Medical University
REVIEW RETURNED	13-Dec-2022
GENERAL COMMENTS	This work reported bone-targeting agents use in patients with
	bone metastasis from breast, lung, or prostate cancer using
	structured and unstructured electronic health records. However, it
	does not reach the level of your journal. Because there were many
	research on bone-targeting agents in human cancers. Large
	length of paper is without focus. There are some defects and
	concerns that need to be addressed.
	1.No medical ethical review.
	2. The size of included cases is too small, and it is hard to
	elucidate the problem.
	3. This Natural Language processing (NLP) approach need more
	introduction and cited reference.
	4.Lack of highlights and too many limitations.

REVIEWER	Lu, Jian Peking University Third Hospital, Department of Urology
REVIEW RETURNED	22-Dec-2022

GENERAL COMMENTS	The retrospective cohort study assessed the bone-targeting
	agents (BTA) use in patients with confirmed bone metastases
	(BM) from breast cancer (BC), non-small cell lung cancer (NSCLC)
	or prostate cancer (PC) using structured and unstructured
	electronic health records from a regional UK-based hospital. The
	use of extensive unstructured data from multiple electronic
	medical records (EMR) sources and application of natural
	language processing (NLP) techniques to identify patients with BM
	is key strength of the current study compared with other same type
	research. This is a well-written manuscript, but improvements in
	several parts are still required to make the paper more impactful.
	1. What the basis of hypercalcaemia classification (like Mild,
	Moderate and Severe) at BM diagnosis in Table 2?

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

2. What the definition and content of Systemic anticancer therapy (SACT) in Table 2? As BC, NSCLC and PC all have diverse and complex systemic treatments such as chemotherapy, endocrine therapy and targeted therapy.
3. The authors pointed out the main outcome measures in the part of abstract, it would be better to highlight and explain again in the part of Methods.
4. It was suggested to describe Figure 1 in more detail in the part of Data source, especially the meanings represented by some abbreviations like EPRO, etc.
5. The authors proposed that PC patients with BTA prescriptions had a numerically higher percentage SACT compared to patients without a BTA prescription. However, both BTA and no BTA were represented by NA in Table 2, how could the above finding be obtained? The similar confusion can also be observed in breast cancer. The view about BC patients with BTAs had a numerically
higher percentage of a history of surgery compared to patients without a BTA is not consistent with the results illustrated in Table 2 as the percentage of a history of surgery of BTA (25.9%) is lower than no BTA (34.7%). Please clarified clearly.

REVIEWER	Van Poznak, Cathy University of Michigan
REVIEW RETURNED	24-Dec-2022

GENERAL COMMENTS	The manuscript would benefit from providing additional information
	and rovision
	 Please provide additional information on the methods used for
	record review and maintaining patient confidentiality
	 Please provide data on the validation of the Natural Language
	Processing technology used
	 Please provide additional information on the health care system
	used and the selected study dates
	• Table 2 and Table 3 have data that is outside of the scope of the
	3 Objective Measures, and that of the Methods. Please revise.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Dr. Yu Fan, The Affiliated Hospital of Southwest Medical University

Comments to the Author:

This work reported bone-targeting agents use in patients with bone metastasis from breast, lung, or prostate cancer using structured and unstructured electronic health records. However, it does not reach the level of your journal. Because there were many research on bone-targeting agents in human cancers. Large length of paper is without focus. There are some defects and concerns that need to be addressed.

1.No medical ethical review.

Thank you for your comment. We have clarified the Ethics section (see also above Editor's comment 2):

"In this retrospective study, all data were fully anonymised, and no participant consent wasrequired. Ethics for this study was provided by 3 active Health Research Authority (HRA) Wales approvals for retrospective data- based studies for breast cancer (HRA ref no. 249275), prostate cancer (HRA ref no. 260189) and lung cancer (HRA ref no. 251650)."

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

2. The size of included cases is too small, and it is hard to elucidate the problem.

Thank you for your comment. As shown in table 1, the eligible cohort at feasibility stage ranges from 5202 to 6142. Following the thorough identification process, the included cohorts were smaller but the quality of the data and confirmation of bone metastasis for the included patients were key to a comprehensive assessment of the use of bone-targeting agents. This descriptive study uses one of the largest cancer data sources with bone metastases patients, and included patients identified from unstructured medical records. The latter is a population often not included in larger size studies focusing only on structured data.

3. This Natural Language processing (NLP) approach need more introduction and cited reference. Thank you for your comment. We have added more information in the Methods section:

"The NLP platform Interactive Information Extraction (I2E), that was developed by the company Linguamatics (http: www.linguamatics.com/products/i2e), was used to automate reviewing of unstructured text. That involved looking for inbuilt and predefined keywords and phrases defined by clinical physicians with experience in diagnosing and treating patients with BM. A large percentage of the NLP-identified BM cases were manually checked by the data review team consisting of a senior physician and a data quality officer, and the information from this was used to improve the NLP query in a continuous feedback loop of checking and adjusting. Finally, all identified BM cases were manually reviewed by the data review team to provide final confirmation."

4.Lack of highlights and too many limitations.

Thank you for your comment. We have one paragraph each for the strengths and limitations of our study, and also presented these in 5 bullet points after the abstract (3 main strengths and 2 main limitations). The limitations focus on the data quality (detailed in the corresponding paragraph), and the geographic limitation. The highlights/strengths are numerous, with our study using extensive unstructured data, both inpatient and outpatient data and multiple linked data sources, identifying bone metastasis cases that would otherwise have been missed, and capturing important information on those underrepresented patients.

Reviewer: 2 Prof. Jian Lu, Peking University Third Hospital

Comments to the Author:

The retrospective cohort study assessed the bone-targeting agents (BTA) use in patients with confirmed bone metastases (BM) from breast cancer (BC), non-small cell lung cancer (NSCLC) or prostate cancer (PC) using structured and unstructured electronic health records from a regional UK-based hospital. The use of extensive unstructured data from multiple electronic medical records (EMR) sources and application of natural language processing (NLP) techniques to identify patients with BM is key strength of the current study compared with other same type research. This is a well-written manuscript, but improvements in several parts are still required to make the paper more impactful.

1. What the basis of hypercalcaemia classification (like Mild, Moderate and Severe) at BM diagnosis in Table 2?

Thank you for your comment. We have added the range used for the classification as a footnote to table 2 : "The level of hypercalcemia was based on the following serum calcium levels (mm/L) : mild 2.75-3.00, moderate 3.00-3.40, severe 3.40+"

2. What the definition and content of Systemic anticancer therapy (SACT) in Table 2? As BC, NSCLC and PC all have diverse and complex systemic treatments such as chemotherapy, endocrine therapy and targeted therapy.

Thank you for your comment. We have added a footnote to table 2 : "SACT was cancer-specific and included chemotherapy, endocrine therapy and targeted therapy".

3. The authors pointed out the main outcome measures in the part of abstract, it would be better to

highlight and explain again in the part of Methods.

Thank you for your comment. We have added a short paragraph in the Methods section accordingly.

4. It was suggested to describe Figure 1 in more detail in the part of Data source, especially the meanings represented by some abbreviations like EPRO, etc.

Thank you for your suggestion. Some of the databases are specific to LTHT and some of them are commercially available. The bullet points below each database name in Figure 1 describe the data content. We have added a brief footnote to Figure 1.

5. The authors proposed that PC patients with BTA prescriptions had a numerically higher percentage SACT compared to patients without a BTA prescription. However, both BTA and no BTA were represented by NA in Table 2, how could the above finding be obtained? The similar confusion can also be observed in breast cancer. The view about BC patients with BTAs had a numerically higher percentage of a history of surgery compared to patients without a BTA is not consistent with the results illustrated in Table 2 as the percentage of a history of Surgery of BTA (25.9%) is lower than no BTA (34.7%). Please clarified clearly.

Thank you for your comment. To clarify, the sentence refers to RT or SACT or Surgery, not SACT alone. It is the complement of the category 'none'. We have updated the sentence in the manuscript to further clarify that. Also, please see a more detailed explanation below.

• For Prostate Cancer:

The text on history of RT, SACT or surgery for BTA and no BTA cohorts refers to the % of all three treatments overall, ie. RT OR SACT OR surgery and uses the following results in Table 2: BTA No BTA

None - ie. no RT, SACT or surgery

63.6% 72.9%

Hence, the %s for any of the three treatments overall, are

BTA No BTA

RT OR SACT OR surgery

36.4% (100%-63.6%) 17.1% (100%-72.9%)

Thus, patients with BTA prescriptions had a numerically higher percentage of history of RT OR SACT OR surgery compared to patients without a BTA prescription.

• For Breast Cancer:

The results in Table 2 show higher percentages for two categories of surgery in BTA vs no BTA cohorts and lower percentages in the other two categories:

BTA No BTA

RT & Surgery 15.0% 11.7%

SACT & Surgery 11.9% 7.9%

SACT, RT & Surgery 32.7% 33.2%

Surgery 25.9% 34.7%

Since the %s in the BTA group are higher for two treatment categories and lower in the two other treatment categories compared to the No BTA group, the results are inconclusive, and the corresponding text in the manuscript was deleted.

• For NSCLC:

Similar to PC, it is best to highlights the results for the % of overall RT OR SACT OR surgery: BTA No BTA

None – ie. no RT, SACT or surgery

59.8% 68.5%

Hence the % a for any of t

Hence, the %s for any of the three treatments overall, are

BTA No BTA

RT OR SACT OR surgery

40.2% (100%-59.8%) 31.5% (100%-68.5%)

Thus, patients with BTA prescriptions had a numerically higher percentage of history of RT OR SACT

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

OR surgery compared to patients without a BTA prescription. The text in the manuscript was updated accordingly.

Thank you for your comment on NA. We have changed that to # in Table 2.

Reviewer: 3 Prof. Cathy Van Poznak, University of Michigan

Comments to the Author:

The manuscript would benefit from providing additional information and revision.

• Please provide additional information on the methods used for record review and maintaining patient confidentiality

Thank you for your comment. We added further details in the Data Sources section to address this comment.

• Please provide data on the validation of the Natural Language Processing technology used Thank you for your comment. We have added more information in the Methods section:

"The NLP platform Interactive Information Extraction (I2E), that was developed by the company Linguamatics (http: www.linguamatics.com/products/i2e), was used to automate reviewing of unstructured text. That involved looking for inbuilt and predefined keywords and phrases defined by clinical physicians with experience in diagnosing and treating patients with BM. A large percentage of the NLP-identified BM cases were manually checked by the data review team consisting of a senior physician and a data quality officer, and the information from this was used to improve the NLP query in a continuous feedback loop of checking and adjusting. Finally, all identified BM cases were manually reviewed by the data review team to provide final confirmation."

• Please provide additional information on the health care system used and the selected study dates Thank you for your comment. We have provided additional information in the section 'Data source'. Selected study dates are January 1 2007 to December 31 2018, as detailed in the first paragraph in the section 'Identification of BM diagnosis'.

• Table 2 and Table 3 have data that is outside of the scope of the 3 Objective Measures, and that of the Methods. Please revise.

Thank you for your comment. We have added text in the Methods section to detail the variables presented in tables 2 and 3.

Reviewer: 1 Competing interests of Reviewer: No competing interests

Reviewer: 2 Competing interests of Reviewer: None

Reviewer: 3 Competing interests of Reviewer: None