BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-069214
Article Type:	Original research
Date Submitted by the Author:	19-Oct-2022
Complete List of Authors:	Seesaghur, Anouchka; Amgen Ltd Uxbridge, Centre for Observational Research Egger, Peter; IQVIA Europe Warden, Joshua; IQVIA Europe Abbasi, A; Amgen Ltd Uxbridge, Centre for Observational Research Levick, Bethany; IQVIA Europe Riaz, Majid; IQVIA Europe McMahon, Peter; IQVIA Europe Thompson, Matthew; IQVIA Europe Cheeseman, Sue; The Leeds Teaching Hospitals NHS Trust, REAL Oncology
Keywords:	Breast tumours < ONCOLOGY, ONCOLOGY, Urological tumours < ONCOLOGY, Respiratory tract tumours < ONCOLOGY, Epidemiology < ONCOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

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

Author Affiliations

Anouchka Seesaghur, MBBS MPH[1], Peter Egger, PhD[2], Joshua Warden, MSc[2], Ali Abbasi, PhD[1], Bethany Levick, PhD[2], Majid Riaz, BSc[2], Peter McMahon, MSc[2], Matthew Thompson, BSc[2], Sue Cheeseman MBChB, FRCP[3].

Name and Address of corresponding author: Peter Egger (peter.egger@iqvia.com)

[1] Centre for Observational Research, Amgen Ltd, 4 Uxbridge Business Park, Sanderson Road, Uxbridge UB8 1DH

[2] IQVIA, The Point, 37 North Wharf Road, Paddington, London, W2 1AF

[3] Leeds NHS Teaching Hospitals Trust, St. James's University Hospital, Beckett Street, Leeds LS9 7TF

Manuscript Word Count: 2978 words (excluding abstract, tables, figures, acknowledgments, references, and online-only material)

Abstract: 300 words

Objective: To assess use of bone-targeting agents (BTA) in patients with confirmed bone metastases (BM) from breast cancer (BC), non-small cell lung cancer (NSCLC) or prostate cancer (PC)

Design: Retrospective cohort study

Setting: Regional hospital-based oncology database of approximately 2 million patients in England

Participants: Patients aged ≥18 years with a diagnosis of BC, NSCLC or PC as well as BM between January 1, 2007 to December 31, 2018, with follow-up to June 30, 2020 or death; BM diagnosis ascertained from recorded medical codes and unstructured data using natural language processing (NLP).

Main Outcomes Measures: Initiation or non-initiation of BTA following BM diagnosis, time from BM diagnosis to BTA initiation, time from first to last BTA, time from last BTA to death

Results: This study included 559 BC, 894 NSCLC and 1013 PC with BM; median age (Q1, Q3) was 65 (52-76), 69 (62-77) and 75 (62-77) years respectively. NLP identified BM diagnosis from unstructured data for 92% BC, 92% NSCLC and 95% PC patients. Among patients with BC, NSCLC and PC with BM, 47%, 87% and 88% did not receive a BTA, and 53%, 13% and 12% received at least one BTA, starting a median 65 (27, 167), 60 (28, 162) and 610 (295, 980) days after BM respectively. Median (Q1, Q3) duration of BTA treatment was 481 (188, 816), 89 (49, 195) and 115 (53, 193) days for patients with BC, NSCLC and PC. For those with a death record, median time from last BTA to death was 54 (26-109) for BC, 38 (17, 98) for NSCLC, and 112 (44, 218) days for PC.

Conclusion: In this study identifying BM diagnosis from both structured and unstructured data, a high proportion of patients did not receive a BTA. Unstructured data provide new insights on the real-world use of BTA.

Strengths and Limitations

- Our study is the first attempt to characterize bone-targeting agents (BTA) use in clinical practice using both structured and unstructured data on a large sample of patients with solid tumors within England
- Our study uses natural language processing techniques to identify patients with bone metastasis from unstructured data within multiple electronic medical records
- In this study, prescribing data originates from multiple data sources, and includes both inpatient and outpatient data
- This study relies on the quality and completeness of data collected from hospital records
- Insights from this study are limited to the routine practice in one regional area in the UK

INTRODUCTION

Bone is a frequent site of metastasis for breast cancer (BC), non-small cell lung cancer (NSCLC) and prostate cancer (PC), occurring in approximately 70% of patients with advanced BC [1, 2], in 30-40% of all patients with NSCLC [3, 4] and in 80% of patients with advanced PC [5, 6]. Bone metastasis (BM) is a major cause of morbidity leading to severe pain, mobility difficulties, and bone complications, also known as skeletal-related events (SRE) [7-9]. Bone-targeting agents (BTAs) reduce skeletal morbidity from metastatic bone disease and are used in patients with BMs across several tumour types. For most patients, whether symptomatic or not, clinical guidelines recommend starting a BTA as soon as bone metastases (BMs) are diagnosed [10-12].

Records of BM depend on imaging practices in routine clinical practice. Imaging at baseline is performed to stage the patient and define the patient's ongoing management. Throughout a patient's disease journey, other imaging assessments may occur but repeat scans are not routinely performed unless clinically indicated. In electronic medical records (EMR), BM diagnoses are often not coded using medical codes [13, 14], and may be captured in unstructured free text. Studies relying solely on BM diagnosis identified via structured data, may therefore, lead to an incomplete picture of the management of patients with cancer and BM.

To address these gaps in evidence on BM ascertainment, we used novel techniques to identify BMs in both structured medical code-based data, and unstructured free text data from the hospital-based EMR database of the largest integrated regional cancer center in the UK. This allowed us to identify a comprehensive BM patient population to better understand the management of BM in cancer patients. The current study aims to evaluate the real-world use and non-use of BTAs in patients with BC, NSCLC or PC with a BM diagnosis.

METHODS

Data source

This hospital-based cohort study used EMR data from the REAL-Oncology database of Leeds NHS teaching hospital trust (LTHT). REAL-Oncology receives patient-level data directly from various clinical information systems, and each data source is linked at the patient-level via the patient's unique identifier. (Figure 1)

A two-phase approach was adopted to assess BTA use in patients with cancer and BM using secondary and tertiary care data. In Phase I, we applied novel techniques to identify patients with confirmed BM across all existing EMRs, whether structured or unstructured. In Phase II, we evaluated the use of BTAs within the identified study cohort.

Phase I: Identification of BM diagnosis

Adult patients (aged ≥18 years at the date of primary cancer diagnosis) with a primary diagnosis of BC, PC and NSCLC (index date) were identified through International Classification of Diseases (ICD)-10 codes (Appendix A) (and additionally ICD-O-3 morphology codes for NSCLC, Appendix B) during the study period from January 1, 2007 to December 31, 2018. Patients with other primary malignancies prior to the index date or enrolled in a randomized controlled trial on BTA were excluded.

We included patients who had a BM either at their first diagnosis of primary cancer or developed BM at any time after initial primary cancer diagnosis. The BM diagnoses were identified via a BTA record, direct coding of BM, and query of unstructured text from imaging, pathology, and clinical summary reports using a Natural Language processing (NLP) approach. Linguamatics I2E, an NLP platform, was used to automate reviewing of unstructured text by looking for inbuilt and predefined keywords and phrases defined by clinical physicians with experience in diagnosing and treating patients with BM. To confirm all cases, all identified BM cases were manually reviewed by a senior physician and a data quality officer.

Phase II: Assessment of BTA use

Study population

From Phase I, all adult patients with BC, NSCLC or PC and a confirmed diagnosis of BM (identified from January 1, 2007 to December 31, 2018) were followed from BM diagnosis date to June 30, 2020 or death.

BTA treatment

BTA treatment was determined through patients EMRS linked to the hospital pharmacy dispensing database JAC covering both in-patient and out-patient prescriptions, and the treatment prescribing database ChemoCare. We reported three phases of medication adherence (initiation, implementation, and persistence) as recommended by the European Society for Patient Adherence, COMpliance, and Persistence (ESPACOMP)[15]. BTAs included two different classes of anti-resorptive agents: bisphosphonates (both IV and oral) and the RANKL inhibitor denosumab.

Statistical analysis

Primary cancers, BM cases, and BTA use, including type of BTA and switches between BTAs, were reported as counts and percentages. Patient characteristics were reported as percentages for categorical variables and medians (Q1, Q3) for continuous variables. The Kaplan Meier method was applied to analyze time-to-event data of BTA records, such as time to first BTA, duration of BTA, and time from last BTA to death. Counts of <6 were marked as such in all results to protect patient privacy. The SAS version 9.4 (SAS, CARY, NC, USA) and R version 3.2[16] was used for all data management and statistical analyses.

Patient and public involvement

Patients were not involved in this study.

RESULTS

Phase I: Identification of BM diagnosis

In Phase I, we identified a total of 6,142 BC, 5,202 NSCLC and 5,382 PC primary cancer patients. Table 1 shows a summary of the different approaches and corresponding results for identifying BM diagnoses. Each of these approaches were reviewed to ascertain confirmation of a BM diagnosis: direct identification by NLP, identification by proxy based on a record of BTA treatment, identification by proxy based on a record of spinal cord compression, and direct identification via diagnosis codes in structured EMR.

Table 1 shows the numbers and percentages of the three different methods of BM identification: NLP of unstructured data, evidence of spinal cord compression and BM in coded EMR fields. For BC, 573 patients were identified, with 527 (92%) via NLP-based querying of unstructured data. For NSCLC the total was 899, with 829 (92%) from unstructured data. For PC the total was 1017 and the results for unstructured data were 963 (95%). Further clinical expert review of all resulting cases detected additional false positives and yielded a final study cohort for BC, NSCLC and PC: 559 (9% of all primary cancer cases), 894 (17%) and 1013 (19%) BM patients, respectively.

Table 1 Attrition table of study patient population in phase I

	Breast Cancer		NSCLC		Prostate Cancer	
Patient size by BM Method	**N**	**%**	**N**	**%**	**N**	**%**
Eligible Cohort at feasibility stage	**6142**	**100**	**5202**	**100**	**5382**	**100**
Patient cohort identified with BM in the medical records	573	9.3	899	17.3	1017	18.9
Patients identified by NLP of reports *	527	92.0	829	92.2	963	94.7
Patients identified as receiving BTA treatment *	309	53.9	118	13.1	129	12.7
Patients identified as patient having SCC *	19	3.3	41	4.6	42	4.1
Patients identified in coded EMR field *	49	8.6	<75	-	55	5.4
Final cohort of patients with BM after further clinical review	559	9.0	894	17.0	1013	18.8

*Number of patients identified as BM by each method of the overall eligible cohort and confirmed after review by two clinical physicians (patient can be identified in multiple methods)

Abbreviations: BM: Bone metastasis; BTA: Bone-targeting agents; EMR: Electronic medical record; NLP: Natural language processing; NSCLC: Non-small cell lung cancer; SCC: Spinal cord compression.

Phase II: Assessment of BTA use

Table 2 shows the patient demographic and clinical characteristics as well as the treatment histories of the final study cohort stratified by tumour type, and by BTA use/non-use. BTA initiation, implementation and persistence are shown in Figure 2, followed by further details of the two most frequent BTAs in Table 3.

Table 2 Patient characteristics of final BM patient cohort

	BC		NSCLC		PC	
N	559		894		1013	
Follow-up: median days (Q1-Q3)	458 (128-933)		87 (37-205)		682 (357)	
Age at BM diagnosis, median (Q1-Q3)	65 (52-76)		69 (62-77)		75 (62-77)	
Stage IV at primary cancer diagnosis	30%		86%		72%	
History of SRE at BM diagnosis						
In the 56-day pre-BM-diagnosis period	35%		94%		25%	
Less than 16 days before BM diagnosis	1%		3%		1%	
Within 16-32 days before BM diagnc	3%		8%		5%	
	BTA	**no BTA**	**BTA**	**no BTA**	**BTA**	**no BTA**
N	294	265	117	777	121	892
Sex-female	100%	100%	45%	45%	0%	0%
Primary Cancer stage IV	29%	30%	85%	86%	74%	71%
CRPC diagnosis	NA*	NA*	NA*	NA*	82%	56%
At BM diagnosis	NA	NA	NA	NA	NA*	NA*
ECOG present	64%	39%	86%	82%	30%	19%
ECOG 0-2	94%	81%	67%	49%	27%	18%
eGFR** median (Q1-Q3)	81 (67-90)	83 (60-90)	90 (72-90)	86 (66-90)	82 (65-90)	73 (57-90)
eGFR** <60	15%	22%	14%	17%	17%	26%
Hypercalcaemia classification at BM diagnosis						
<2.75 mmol/L	90.51%	88.33%	82.20%	94.27%	75.21%	71.65%
Mild	3.05%	NA	6.78%	2.00%	0.00%	NA
Moderate	NA	NA	6.78%	NA	0.00%	NA
Severe	NA	0.00%	NA	NA	0.00%	0.00%
Missing/Unknown	NA	8.95%	NA	2.40%	24.79%	27.76%
Renal disease						
Yes	NA	3.4%	NA	3.6%	0.8%	3.8%
missing	70.4%	56.6%	23.1%	21.5%	57.0%	47.2%
CCI						
0	21.4%	31.3%	57.3%	51.1%	33.1%	39.4%
1	5.4%	4.2%	14.5%	16.6%	4.1%	5.9%
2	NA	4.9%	NA	6.8%	5.0%	4.9%
3+	NA	3.0%	NA	4.0%	0.8%	2.6%
missing	70.4%	56.6%	23.1%	21.5%	57.0%	47.2%
Estrogen receptor status						
Positive	84.4%	72.1%	NA	NA	NA	NA
Missing	NA	3.0%				
Progesterone receptor status						
Positive	65.0%	55.1%	NA	NA	NA	NA
Missing	6.1%	6.4%				
HR/HER2 status						
HR-/HER2-	10.0%	17.0%	NA	NA	NA	NA
HR-/HER2+	3.4%	6.8%	NA	NA	NA	NA
HR+/HER2-	58.2%	46.4%	NA	NA	NA	NA
HR+/HER2+	5.8%	5.3%	NA	NA	NA	NA
missing	22.8%	24.5%	NA	NA	NA	NA

	BC		NSCLC		PC	
EGFR mutation	NA	NA				
Pathogenic	NA	NA	9.4%	5.2%	NA	NA
Wildtype	NA	NA	45.3%	29.6%	NA	NA
missing	NA	NA	45.3%	65.3%	NA	NA
ALK mutation	NA	NA			NA	NA
Pathogenic	NA	NA	NA	2.2%	NA	NA
Wildtype	NA	NA	41.0%	21.8%	NA	NA
missing	NA	NA	55.6%	76.1%	NA	NA
PDL1 mutation	NA	NA			NA	NA
High (>=50%)	NA	NA	NA	4.1%	NA	NA
Intermediate (1-49%)	NA	NA	10.3%	NA	NA	NA
Low (<1%)	NA	NA	13.7%	5.0%	NA	NA
Not done	NA	NA	25.6%	19.1%	NA	NA
missing	NA	NA	44.4%	68.5%	NA	NA
Histopathological stage	NA	NA	NA	NA	NA	NA
Squamous-cell carcinoma	NA	NA	23.1%	15.6%	NA	NA
Other specified NSCLC	NA	NA	NA	2.7%	NA	NA
NSCLC NOS	NA	NA	12.0%	16.6%	NA	NA
Non-squamous NSCLC	NA	NA	49.6%	39.6%		
missing			11.1%	25.5%		
Therapy before BTA administration* **	NA	NA	NA	NA	NA	NA
None	12.9%	10.6%	59.8%	68.5%	63.6%	72.9%
RT	NA	NA*	NA	2.3%	8.3%	6.1%
RT & Surgery	15.0%	11.7%	NA	1.9%	5.8%	4.2%
SACT	NA	0	7.7%	2.3%	NA	NA*
SACT & RT	NA	0	NA	2.1%	NA	0.8%
SACT & Surgery	11.9%	7.9%	NA	3.5%	NA	NA
SACT, RT & Surgery	32.7%	33.2%	6.8%	3.2%	NA	NA
Surgery	25.9%	34.7%	13.7%	16.2%	16.5%	14.6%

* NA means <6 patients
** eGFR units: ml/min/1.73m2)
*** The time period for these therapies includes the time from primary cancer to BM diagnosis
Abbreviations: ALK: anaplastic lymphoma kinase; BM: Bone metastasis, BTA: Bone-targeting agent; CCI: Charlson Comorbidity Index; CRPC: castration-resistant prostate cancer; eGFR: Estimated glomerular filtration rate; ECOG: Eastern Cooperative Oncology Group; HER: Human Epidermal Growth Factor Receptor; HR: hormone receptor; NA: Not available; NOS: Non-otherwise specified; NSCLC: Non-small cell lung cancer; PD-L1: Programmed cell death receptor ligand-1; RT: radiotherapy; SACT: Systemic anticancer therapy

Breast Cancer

Among 559 patients with BC and BM, 47% (n=265) did not have a BTA prescription, and 53% (n=294) received at least one BTA prescription, starting a median (Q1, Q3) of 65 (27, 167) days from their BM diagnosis date (inclusive) to their first BTA initiation date (excludes 9 patients with a BTA before BM). Median (Q1, Q3) duration of BTA therapy from first to last BTA record was 481 (188, 816) days and median (Q1, Q3) time from last BTA to death was 54 (26, 109) days (Figure 2) . Most patients (86%, n= 254) received only one type of BTA. Table 3 provides details of two specific BTAs of different classes that were administered, the RANKL inhibitor denosumab (n=56, 19.1%) and the bisphosphonate zoledronic acid (n=229, 77.9%), both with the most frequent cycle duration of 28 days. During the follow-up period, a total of 52 switches were observed between BTAs. Of those, switches between denosumab and zoledronic acid

were the most frequent: 30% (17/56) of all denosumab administrations ended in a switch to zoledronic acid, within a median (Q1, Q3) time to switch of 32 (28, 57) days, and 5% (11/229) of all zoledronic acid administrations ended in a switch to denosumab, within a median (Q1, Q3) time of 78 (35, 216) days. Patients with BTAs had a numerically higher percentage of oestrogen receptor status positive, progesterone receptor status positive, human epidermal growth factor receptor (HR+/HER)- status as well as a history of surgery compared to patients without a BTA (Table 2).

NSCLC

Among the 894 patients with NSCLC and BM, 87% (n=777) did not receive a BTA prescription and 13% (n=117) received at least one BTA prescription, starting a median (Q1, Q3) of 60 (28, 162) days from their BM diagnosis date (inclusive) to their BTA initiation date (excludes 8 patients with a BTA before BM). Median (Q1, Q3) duration of BTA therapy from first to last BTA record was 89 (49, 195) days and median (Q1, Q3) time from last BTA to death was 38 (16, 98) days (Figure 2) . A total of 12 patients with NSCLC received denosumab and 93 patients received zoledronic acid (Table 3), both with the most frequent cycle duration of 28 days. The median number of administrations per patient was 2 (Q1,Q3: 1,11) for denosumab, and 1 (Q1,Q3: 1,3) for zoledronic acid. A total of 114 (97%) patients received only one type of BTA and <6 switches occurred between BTAs. Patients with BTAs had a numerically higher percentage of estimated glomerular filtration rate (eGFR), anaplastic lymphoma kinase (ALK) and PD-L1 mutation data missing as well as a higher percentage of a history of systemic anticancer therapy (SACT) compared to patients without a BTA (Table 2).

Prostate Cancer

Among the 1013 patients with PC and BM, 88% (n=892) did not receive a BTA prescription and 12% (n=121) received at least one BTA prescription, starting a median (Q1, Q3) of 611 (295, 980) days from their BM diagnosis date (inclusive) to their BTA initiation date (excludes 1 patient with a BTA before BM). Median (Q1, Q3) duration of BTA therapy from first to last BTA record was 115 (53, 193) days and median (Q1, Q3) time from last BTA to death was 112 (44, 218) days (Figure 2) There were no patients on denosumab while 113 patients received zoledronic acid (Table 3), with the most frequent cycle duration of 28 days. The median number of administrations per patient was 2 (Q1, Q3: 1, 4) for zoledronic acid. PC BTA patients only had a record of one unique BTA with no switching recorded. Patients with BTA prescriptions had a numerically higher percentage of history of RT, SACT or surgery compared to patients without a BTA prescription (Table 2).

Table 3 BTA administration in BM patients across the 3 cancers

		BC	NSCLC	PC
Total unique agents	**N**	295	117	121
	Median	1.0	1.0	1.0
	Min-Max	1.0-3.0	1.0-2.0	1.0-1.0
Denosumab Administrations	**N**	56 (19.1%)	12 (10.3%)	0 (0.0%)
	Only received once	11	<6	-
	Median	6.50	2.0	-
	Min-Max	1.0-61.0	1.0-14.0	-
Zoledronic acid administrations	**N**	229 (77.9%)	93 (79.5%)	113 (93.4%)
	Only received once	32	52	51
	Median	9.0	1.0	2.0
	Min-Max	1.0-50.0	1.0-21.0	1.0-34.0

Abbreviations: BC: breast cancer; NSCLC: non-small cell lung cancer; PC: prostate cancer
The remaining %s that add up to the total 100% in the table include BTAs other than denosumab or zoledronic acid.

DISCUSSION

Use of structured and unstructured data to identify BM patients within LTHT

In this study, over 90% of all BM cases were identified through NLP-based querying of unstructured data HealthCare professionals typically record BM detected during different diagnostic procedures in both structured and unstructured formats. Restricting the analysis to structured medical codes would have significantly underestimated the occurrence of BM in the three cancer cohorts in this hospital-based setting. Hence, use of NLP greatly enhanced the efficiency of the identification of BM cases from multiple unstructured data sources. The need for clinical review to eliminate false positive cases shows that further refinement of NLP models is still required.

BTA usage in patients with BC and BM

A European multi-country study (Von Moos et al.[17]) found that 88% of BC patients with BM received BTA treatment, while 53% of BC patients with BM received BTA treatment in our study. There are key differences between the studies. The Von Moos et al. study collected data in a cross-sectional survey of physicians that were treating BM patients who were actively receiving treatment for their cancer. In contrast, 57% of all BM patients in our study did not have a record of SACT, even though some of it may be a result of prescribing recorded outside the available data systems. A prospective study using a German tumour

registry (Schroder et al.[18]) reported a BTA treatment of 89% in BC patients with BM with a median time to treatment from BM diagnosis of 3 weeks. Data collection was prospective and focused on an anticancer treated cohort, including out-patient treatment data. In contrast, our study obtained treatment data retrospectively from potentially incomplete hospital treatment databases and did not include treatment outside the hospital. Furthermore, there may be genuine differences in the use of BTAs in cancer patients with BM between the UK and other European countries. Determinants of BTA prescribing in cancer patients with BM were evaluated in several studies [17, 19, 20]. Findings from Von Moos et al.[17] indicate that some physicians base their BTA treatment decisions not only on clinical guidelines but also consider patients' Eastern Cooperative Oncology Group (ECOG) performance score, disease burden, and the presence of other sites of metastatic disease. For example, a patient with an ECOG performance score of 0-2 is considered fit enough to receive BTA treatment, but in the presence of extensive liver disease and low burden of bone disease, may not routinely receive a BTA. Our study showed a numerical difference in ECOG scores between patients with BTA and without BTA: 94% of patients with a BTA and 81% without a BTA had an ECOG score of 0-2. These findings suggest that BTA treatment is determined on a case-by-case basis within this setting and is not solely reliant on BTA guidelines.

BTA usage in patients with NSCLC and BM

Diel et al.[21] investigated 242 lung cancer patients with a diagnosis of BM and who received at least one BTA treatment in Germany from 2011 to 2015. Of these patients, 15% received denosumab and 63% zoledronic acid, while our study observed 10% of NSCLC BM patients receiving denosumab and 80% zoledronic acid. The probability of patients still on denosumab after 6 months was 87% in Diel et al., compared to 13% in our study. The 2014 European Society for Medical Oncology (ESMO) bone health guidelines, which cover some of the German study time period recommend zoledronic acid or denosumab in patients with a life expectancy of greater than 3 months[10]. The NSCLC patients within our study had a median follow-up time of 87 (Q1, Q3: 37,205) days.

The low proportion of patients receiving BTA within the current study is likely due to the poor prognosis of these patients. Overall, survival data published by LTHT on advanced non-squamous NSCLC patients (stage IIIB-IV) showed patients had a median survival of 4.1 months between 2007-2012 and 5.0 months

between 2013-2017 [22]. The ECOG score further reflects the burden of disease in this population: 67% of BTA patients with a score of 0-2 and 49% in non-BTA patients.

BTA usage in patients with PC and BM

The European multi-country study (von Moos, et al. [17]) included an evaluation of castrate-resistant prostate cancer (CRPC) patients and reported that 77% of CRPC patients received at least one BTA. The von Moos et al. study included patients who were actively receiving anticancer therapy. In our study, over 71% were diagnosed at stage IV and had BM at the time of PC diagnosis and 72% had no record of any other treatment such as SACT or surgery. A US-based study using claims and commercial databases, (Hernandez, et al [7]) identified BTA use in 52% of PC patients with BM in 2012. Median time to first BTA was 35 and 37 days, respectively for the claims and commercial databases. In our study we observed that 12% of BM patients received a BTA, with a median time to first BTA of 610 days. While the National Institute for Health and Care Excellence (NICE) guidelines do not recommend denosumab for PC patients with BM in the UK [12], it is approved for use in PC patients in the US [23].

STRENGTHS AND LIMITATIONS

A key strength of the current study is the use of extensive unstructured data from multiple EMR sources and application of NLP techniques to identify patients with BM. Leveraging unstructured data is especially important because bone metastases are likely identified at different diagnostic investigations and reported in different medical records. Access to multiple data sources and linkage within the LTHT database and the application of NLP methods enabled a more comprehensive account of the patient's medical record data. Our findings show that the vast majority of BM cases would have been missed without evidence from unstructured medical record data, as BMs are typically not recorded through structured medical codes in this particular setting. In addition, the availability of both in-patient and out-patient prescribing data from multiple data sources is a strength of the study.

Nevertheless, the study has some limitations due to the capture and documentation of in-patient BTA prescribing information. The comprehensive hospital drug dispensing data (JAC) is only available for the last 5 years. Although BTA treatment is also included in the oncology treatment database ChemoCare, BTA

treatment is not always recorded within ChemoCare, especially for patients who receive a BTA during an in-patient admission. Hence, medications that were not prescribed using ChemoCare, including hormone therapy, and that were prescribed more than 5 years ago (not included in JAC), are not captured in this study. However, an assessment of BTA prescribing before and after the period of JAC availability showed only a marginal difference in BTA prescribing between the two periods. In addition, insights from this study are limited to the routine practice in the UK and reflect existing restrictions in reimbursement and access to BTA therapy within the country.

CONCLUSION

To our knowledge, this is the first study that retrospectively identified BM patients using both structured and unstructured data within England to characterize BTA use in clinical practice. Applying NLP to unstructured data should be considered as a useful additional strategy to identify BM and ascertain cases which would have been missed if only structured data were used. This study provided a different picture to existing literature on BTA use in Europe and the US, highlighting the underuse of BTA treatment within patients with metastatic bone disease from BC, NSCLC or PC. These findings point to a complex decision-making process to prescribe bone protection therapy to cancer patients. Further work is warranted to better understand individual patient medical need and treatment benefit, including repeating this work in other data sources to assess the benefit of using unstructured data.

ACKNOWLEDGEMENT

The authors acknowledge the contribution of Nina Snelling[3] to the research work.

AUTHOR CONTRIBUTIONS

AS, AA, PM and SC made substantial contributions to the design of the work.

BL, MR, MT, and PE contributed to the analysis of the data.

AS, AA, PE, and SC contributed to the interpretation of data. JW drafted the work; AS and PE, made substantial contributions to substantively revise the manuscript. All authors reviewed the manuscript, provided final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This study was funded by Amgen Ltd. The grant number is N/A

COMPETING INTERESTS

I have read the journal's policy and the authors of this manuscript have the following competing interests. Anouchka Seesaghur was an employee and equity holder in Amgen Inc during the conduct of the study. Peter Egger, Bethany Levick and Matthew Thompson were employed with IQVIA during conduct of the study. Joshua Warden had no conflict of interest to declare. Ali Abbasi reported contract work with Amgen Inc. Majid Riaz was employed with IQVIA and had an honorary contract with Leeds Teaching Hospitals Trust (LTHT) to access the data to produce analysis. Peter McMahon worked for IQVIA during the initial development of the manuscript, and the analysis time period. Sue Cheeseman's part was funded by IQVIA.

ETHICS APPROVAL

Institutional Review Board/Independent Ethics Committee (IRB/IEC): Studies conducted at REAL-Oncology are covered by UK Health Research Authority approvals (HRA); the need for ethics approval for this retrospective real-world analysis is waived.

DATA SHARING STATEMENT

No additional data available

REFERENCES

1. Pulido C VI, Ferreira AR. Bone metastasis risk factors in breast cancer. *Ecancermedicalscience*. 2017;
2. Xiong Z, Deng G, Huang X, et al. Bone metastasis pattern in initial metastatic breast cancer: a population-based study. *Cancer Manag Res*. 2018;10:287-295. doi:10.2147/cmar.S155524
3. Santini D, Barni S, Intagliata S, et al. Natural History of Non-Small-Cell Lung Cancer with Bone Metastases. *Sci Rep*. Dec 22 2015;5:18670. doi:10.1038/srep18670
4. D'Antonio C, Passaro A, Gori B, et al. Bone and brain metastasis in lung cancer: recent advances in therapeutic strategies. *Ther Adv Med Oncol*. May 2014;6(3):101-14. doi:10.1177/1758834014521110
5. Liu D, Kuai Y, Zhu R, et al. Prognosis of prostate cancer and bone metastasis pattern of patients: a SEER-based study and a local hospital based study from China. *Sci Rep*. Jun 4 2020;10(1):9104. doi:10.1038/s41598-020-64073-6
6. Macedo F LK, Pinho F, Saraiva N. Bone Metastases: An Overview. *Oncol Rev*. 2017;
7. Hernandez RK, Adhia A, Wade SW, et al. Prevalence of bone metastases and bone-targeting agent use among solid tumor patients in the United States. *Clin Epidemiol*. 2015;7:335-45. doi:10.2147/clep.S85496
8. Asdahl PH, Sundbøll J, Adelborg K, et al. Cardiovascular events in cancer patients with bone metastases-A Danish population-based cohort study of 23,113 patients. *Cancer Med*. Jul 2021;10(14):4885-4895. doi:10.1002/cam4.4027
9. Hjelholt TJ, Rasmussen TB, Seesaghur A, et al. Risk of infections and mortality in Danish patients with cancer diagnosed with bone metastases: a population-based cohort study. *BMJ Open*. Dec 1 2021;11(12):e049831. doi:10.1136/bmjopen-2021-049831
10. Coleman R, Body JJ, Aapro M, Hadji P, Herrstedt J. Bone health in cancer patients: ESMO Clinical Practice Guidelines. *Ann Oncol*. Sep 2014;25 Suppl 3:iii124-37. doi:10.1093/annonc/mdu103
11. Coleman R, Hadji P, Body JJ, et al. Bone health in cancer: ESMO Clinical Practice Guidelines. *Ann Oncol*. Dec 2020;31(12):1650-1663. doi:10.1016/j.annonc.2020.07.019
12. Overview | Denosumab for the prevention of skeletal-related events in adults with bone metastases from solid tumours | Guidance | (NICE) (2012).
13. Jensen A, Nørgaard M, Yong M, Fryzek JP, Sørensen HT. Validity of the recorded International Classification of Diseases, 10th edition diagnoses codes of bone metastases and skeletal-related events in breast and prostate cancer patients in the Danish National Registry of Patients. *Clin Epidemiol*. Aug 9 2009;1:101-8. doi:10.2147/clep.s5446
14. Liede A, Hernandez RK, Roth M, Calkins G, Larrabee K, Nicacio L. Validation of International Classification of Diseases coding for bone metastases in electronic health records using technology-enabled abstraction. *Clin Epidemiol*. 2015;7:441-8. doi:10.2147/clep.S92209
15. De Geest S, Zullig LL, Dunbar-Jacob J, et al. ESPACOMP Medication Adherence Reporting Guideline (EMERGE). *Ann Intern Med*. Jul 3 2018;169(1):30-35. doi:10.7326/m18-0543
16. Team RC. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. https://www.R-project.org/.
17. von Moos R, Lewis K, Massey L, Marongiu A, Rider A, Seesaghur A. Initiation of bone-targeted agents in patients with bone metastases and breast or castrate-resistant prostate cancer actively treated in routine clinical practice in Europe. *Bone*. Jan 2022;154:116243. doi:10.1016/j.bone.2021.116243
18. Schröder J, Fietz T, Köhler A, et al. Treatment and pattern of bone metastases in 1094 patients with advanced breast cancer - Results from the prospective German Tumour Registry Breast Cancer cohort study. *Eur J Cancer*. Jul 2017;79:139-148. doi:10.1016/j.ejca.2017.03.031

19. Butler AM, Cetin K, Hernandez RK, et al. Treatment dynamics of bone-targeting agents among men with bone metastases from prostate cancer in the United States. *Pharmacoepidemiol Drug Saf*. Feb 2018;27(2):229-238. doi:10.1002/pds.4360

20. McGrath LJ, Overman RA, Reams D, et al. Use of bone-modifying agents among breast cancer patients with bone metastasis: evidence from oncology practices in the US. *Clin Epidemiol*. 2018;10:1349-1358. doi:10.2147/clep.S175063

21. Diel I, Ansorge S, Hohmann D, Giannopoulou C, Niepel D, Intorcia M. Real-world use of denosumab and bisphosphonates in patients with solid tumours and bone metastases in Germany. *Support Care Cancer*. Nov 2020;28(11):5223-5233. doi:10.1007/s00520-020-05357-5

22. Snee M CS, Thompson M, . Temporal Trends in Treatment and Overall Survival Among Patients With Incident NSCLC in the UK: A REAL-Oncology Database Analysis From the I-O Optimise Initiative. presented at: 9th European Lung Cancer Congress (ELCC); 2019; Geneva, Switzerland. https://www.leedsth.nhs.uk/assets/ca7effdffd/28.-Snee_REAL-ONC_44P_ELCC2019_FINAL-Poster_4Apr19.pdf

23. FDA Approves Denosumab for the Treatment of Bone Loss in Patients With Prostate or Breast Cancer. Cancer Network: FDA; 2011.

Figures

Figure 1 Leeds NHS Teaching Hospitals Trust (LTHT) data sources and linkages to create study dataset

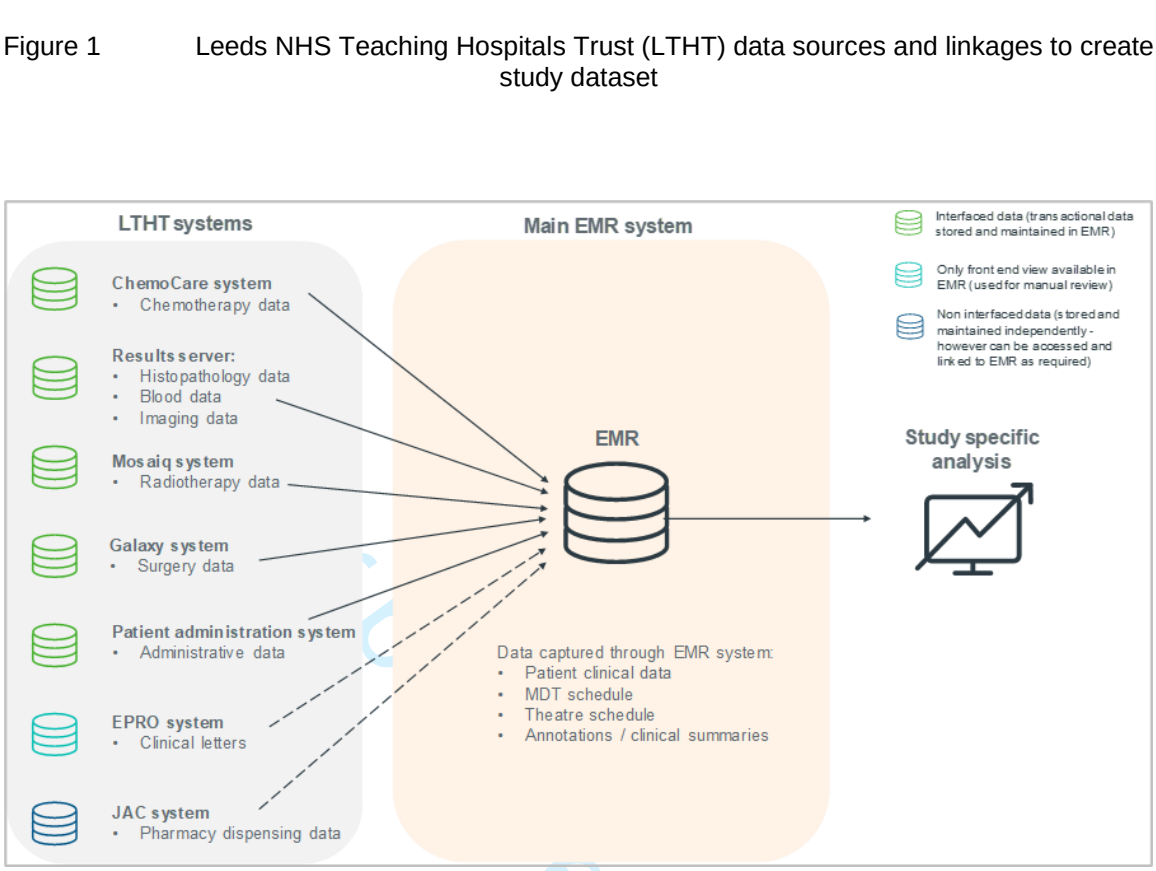
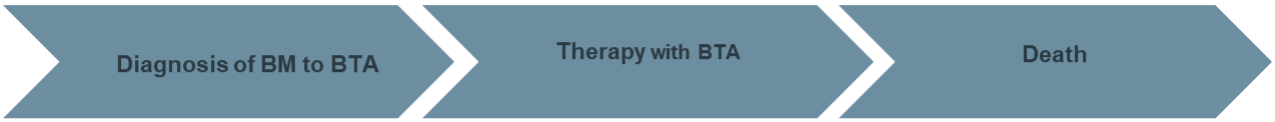


Figure 2 BTA adherence: initiation, implementation, and persistence

BC: 10 weeks
NSCLC: 9 weeks
PC: 87 weeks

BC: 68 weeks
NSCLC: 13 weeks
PC: 16 weeks

BC: 8 weeks
NSCLC: 5 weeks
PC: 16 weeks



Start of therapy		Duration of therapy		End of therapy	
Time from BM to first record of BTA Therapy (days)		Time from start of BTA therapy to end of BTA Therapy (days)		Time from last BTA to death (days)	
Breast Cancer Patients	**N=559**	**Breast Cancer Patients**	**N=294***	**Breast Cancer Patients**	**N=294***
BTA >0 from BM	285 (51.0%)	1> BTA**	250 (84.8%)	Deaths	225 (76.3%)
Median (Q1-Q3)	**65 (27-167)**	**Median (Q1-Q3)**	**481 (188-816)**	**Median (Q1-Q3)**	**54 (26-109)**
NSCLC Patients	**N=894**	**NSCLC Patients**	**N=117**	**NSCLC Patients**	**N=117**
BTA >0 from BM	109 (12.2%)	1> BTA**	54 (46.2%)	Deaths	111 (94.9%)
Median (Q1-Q3)	**60 (28-162)**	**Median (Q1-Q3)**	**89 (49-195)**	**Median (Q1-Q3)**	**38 (17- 98)**
Prostate Cancer Patients	**N= 1013**	**Prostate Cancer Patients**	**N= 121**	**Prostate Cancer Patients**	**N= 121**
BTA >0 from BM	120 (11.9%)	1> BTA**	63 (52.1%)	Deaths	107 (88.4%)
Median (Q1-Q3)	**611 (295-980)**	**Median (Q1-Q3)**	**115 (53-192)**	**Median (Q1-Q3)**	**112 (44-218)**

* includes patients who had the BTA before their BM diagnosis
** number of patients with a duration of at least one day

Title : 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

Appendix A

ICD-10 diagnosis codes for the primary cancers

Condition	ICD10 code	ICD10 Description
Breast Cancer (BC)	C50	Malignant neoplasm of breast
Non-small cell lung cancer (NSCLC)	C34	Malignant neoplasm of lung + morphology codes to identify NSCLC subgroups in Appendix B below
Prostate Cancer	C61	Malignant neoplasm of prostate

Appendix B

ICD-10 Morphology codes for Adenocarcinoma (NON-Squamous NSCLC, Squamous-cell Carcinoma and NSCLC NOS)

	ICD-0-2 Morphology codes
Adenocarcinoma (non-squamous NSCLC)	Adenocarcinoma UNS 81403 Enteric adenocarcinoma 81443 Solid adenocarcinoma with mucin production 82303 MANEC mixed adenoneuroendocrine carcinoma 82443 Adenocarcinoma, bronchiolo-alveolar (BAC), bronchiolar carcinoma, (incl pathologic in situ-variant) 82503 Alveolar adenocarcinoma 82513 Bronchio-alveolar carcinoma 82523 Adenocarcinoma in situ, mucinous 82532 Adenocarcinoma, mucinous bronchiolo-alveolar (BAC) 82533 Bronchio-alveaolar carcinoma, mixed mucinous and non-mucinous 82543 Adenocarcinoma, mixed with other types of carcinoma incl. squamous cell and small-cell carcinoma 82553 Minimally invasive adenocarcinoma, nonmucinous 82563 Minimally invasive adenocarcinoma, mucinous 82573 Papillary adenocarcinoma, NOS 82603 Micropapillary adenocarcinoma 82653 Clear cell adenocarcinoma 83103 Fetal adenocarcinoma 83333 Mucinous cystadenocarcinoma 84703 Mucinous adenocarcinoma 84803 Mucin-producing adenocarcinoma 84813 Signet ring cell carcinoma 84903 Acinar cell carcinoma 85503 Acinar adenocarcinoma 85513
Squamous-cell carcinoma	Papillary squamous cell carcinoma 80523 Keratinizing squamous cell carcinoma 80713 Non-keratinizing squamous cell carcinoma 80723 Squamous cell carcinoma, small cell nonkeratinizing 80733 Squamous cell carcinoma, spindle cell 80743

	Basaloid squamous cell carcinoma 80833 Squamous cell carcinoma, clear cell type 80843
NSCLC NOS	Carcinoma, NOS 80103 Carcinoma, undifferentiated NOS 80203 Carcinoma, anaplastic NOS 80213 Carcinoma, non-small cell unspecified 80463 Large cell carcinoma with rhabdoid phenotype 80143 Sarcomatoid carcinoma, pleomorphic 80223 NUT carcinoma 80233 Spindle cell and giant cell carcinoma 80303 Giant cell carcinoma 80313 Spindle cell carcinoma, NOS 80323 Pseudosarcomatous carcinoma 80333 Basaloid carcinoma 81233 Adenocystic carcinoma 82003 Mucoepidermoid carcinoma 84303 Adenosquamous carcinoma 85603 Epithelial-myoepithelial carcinoma 85623 Blastoma, pulmonary (pneumoblastoma) 89723 Carcinosarkoma, NOS 89803 Myoepithelial carcinoma 89823
Large cell carcinoma (Non-squamous NSCLC)	Large-cell carcinoma, unspecified 80123

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

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5, 6
Bias	9	Describe any efforts to address potential sources of bias	
Study size	10	Explain how the study size was arrived at	
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6
		(*b*) Describe any methods used to examine subgroups and interactions	
		(*c*) Explain how missing data were addressed	

		(*d*) If applicable, explain how loss to follow-up was addressed	
		(*e*) Describe any sensitivity analyses	
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	Report numbers of outcome events or summary measures over time	
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	9 - 12
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	12
Limitations			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	14-15
Generalisability	21	Discuss the generalisability (external validity) of the study results	14-15
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-069214.R1
Article Type:	Original research
Date Submitted by the Author:	20-Feb-2023
Complete List of Authors:	Seesaghur, Anouchka; Amgen Ltd Uxbridge, Centre for Observational Research Egger, Peter; IQVIA Europe Warden, Joshua; IQVIA Europe Abbasi, A; Amgen Ltd Uxbridge, Centre for Observational Research Levick, Bethany; IQVIA Europe Riaz, Majid; IQVIA Europe McMahon, Peter; IQVIA Europe Thompson, Matthew; IQVIA Europe Cheeseman, Sue; The Leeds Teaching Hospitals NHS Trust, REAL Oncology
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Medical management
Keywords:	Breast tumours < ONCOLOGY, ONCOLOGY, Urological tumours < ONCOLOGY, Respiratory tract tumours < ONCOLOGY, Epidemiology < ONCOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

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

Author Affiliations

Anouchka Seesaghur, MBBS MPH[1], Peter Egger, PhD[2], Joshua Warden, MSc[2], Ali Abbasi, PhD[1], Bethany Levick, PhD[2], Majid Riaz, BSc[2], Peter McMahon, MSc[2], Matthew Thompson, BSc[2], Sue Cheeseman MBChB, FRCP[3].

Name and Address of corresponding author: Peter Egger (peter.egger@iqvia.com)

[1] Centre for Observational Research, Amgen Ltd, 4 Uxbridge Business Park, Sanderson Road, Uxbridge UB8 1DH

[2] IQVIA, The Point, 37 North Wharf Road, Paddington, London, W2 1AF

[3] Leeds NHS Teaching Hospitals Trust, St. James's University Hospital, Beckett Street, Leeds LS9 7TF

Manuscript Word Count: 3191 words (excluding abstract, tables, figures, acknowledgments, references, and online-only material)

Abstract: 300 words

Objective: To assess use of bone-targeting agents (BTA) in patients with confirmed bone metastases (BM) from breast cancer (BC), non-small cell lung cancer (NSCLC) or prostate cancer (PC)

Design: Retrospective cohort study

Setting: Regional hospital-based oncology database of approximately 2 million patients in England

Participants: Patients aged ≥18 years with a diagnosis of BC, NSCLC or PC as well as BM between January 1, 2007 to December 31, 2018, with follow-up to June 30, 2020 or death; BM diagnosis ascertained from recorded medical codes and unstructured data using natural language processing (NLP).

Main Outcomes Measures: Initiation or non-initiation of BTA following BM diagnosis, time from BM diagnosis to BTA initiation, time from first to last BTA, time from last BTA to death

Results: This study included 559 BC, 894 NSCLC and 1013 PC with BM; median age (Q1, Q3) was 65 (52-76), 69 (62-77) and 75 (62-77) years respectively. NLP identified BM diagnosis from unstructured data for 92% BC, 92% NSCLC and 95% PC patients. Among patients with BC, NSCLC and PC with BM, 47%, 87% and 88% did not receive a BTA, and 53%, 13% and 12% received at least one BTA, starting a median 65 (27, 167), 60 (28, 162) and 610 (295, 980) days after BM respectively. Median (Q1, Q3) duration of BTA treatment was 481 (188, 816), 89 (49, 195) and 115 (53, 193) days for patients with BC, NSCLC and PC. For those with a death record, median time from last BTA to death was 54 (26-109) for BC, 38 (17, 98) for NSCLC, and 112 (44, 218) days for PC.

Conclusion: In this study identifying BM diagnosis from both structured and unstructured data, a high proportion of patients did not receive a BTA. Unstructured data provide new insights on the real-world use of BTA.

Strengths and Limitations

- Our study uses both structured and unstructured patient medical history data to address the study aims
- The unstructured data is evaluated through Natural Language Processing techniques
- Prescribing data originates from multiple data sources, and includes both inpatient and outpatient data
- This study relies on the quality and completeness of data collected from hospital records
- Insights from this study are limited to the routine practice in one regional area in the UK

INTRODUCTION

Bone is a frequent site of metastasis for breast cancer (BC), non-small cell lung cancer (NSCLC) and prostate cancer (PC), occurring in approximately 70% of patients with advanced BC [1, 2], in 30-40% of all patients with NSCLC [3, 4] and in 80% of patients with advanced PC [5, 6]. Bone metastasis (BM) is a major cause of morbidity leading to severe pain, mobility difficulties, and bone complications, also known as skeletal-related events (SRE) [7-9]. Bone-targeting agents (BTAs) reduce skeletal morbidity from metastatic bone disease and are used in patients with BMs across several tumour types. For most patients, whether symptomatic or not, clinical guidelines recommend starting a BTA as soon as bone metastases (BMs) are diagnosed [10-12].

Records of BM depend on imaging practices in routine clinical practice. Imaging at baseline is performed to stage the patient and define the patient's ongoing management. Throughout a patient's disease journey, other imaging assessments may occur but repeat scans are not routinely performed unless clinically indicated. In electronic medical records (EMR), BM diagnoses are often not coded using medical codes [13, 14], and may be captured in unstructured free text. Studies relying solely on BM diagnosis identified via structured data, may therefore, lead to an incomplete picture of the management of patients with cancer and BM.

To address these gaps in evidence on BM ascertainment, we used novel techniques to identify BMs in both structured medical code-based data, and unstructured free text data from the hospital-based EMR database of the largest integrated regional cancer center in the UK. This allowed us to identify a comprehensive BM patient population to better understand the management of BM in cancer patients. The current study aims to evaluate the real-world use and non-use of BTAs in patients with BC, NSCLC or PC with a BM diagnosis.

METHODS

Outcomes Measures

The main outcome measures were initiation or non-initiation of BTA following BM diagnosis, time from BM diagnosis to BTA initiation, time from first to last BTA and time from last BTA to death. Further details on BTAs used including extent of use were provided. Patient demographic and clinical characteristics as well as the treatment histories by tumour type, and by BTA use/non-use were also described.

Data source

This hospital-based cohort study used EMR data from the REAL-Oncology database of England National Health Service (NHS) Leeds NHS teaching hospital trust (LTHT). REAL-Oncology receives patient-level data directly from various clinical information systems, and each data source is linked at the patient-level via the patient's unique identifier. (Figure 1)

A two-phase approach was adopted to assess BTA use in patients with cancer and BM using secondary and tertiary care data. In Phase I, we applied novel techniques to identify patients with confirmed BM across all existing EMRs, whether structured or unstructured. In Phase II, we evaluated the use of BTAs within the identified study cohort. The study complied with the Hospital Trust's Information Governance requirements. All data was fully anonymized and patients who had opted out of data sharing were removed from the study. Researchers do not work with identifiable data and work within a secure environment on a secure NHS network.

Phase I: Identification of BM diagnosis

Adult patients (aged ≥18 years at the date of primary cancer diagnosis) with a primary diagnosis of BC, PC and NSCLC (index date) were identified through International Classification of Diseases (ICD)-10 codes (Appendix A) (and additionally ICD-O-3 morphology codes for NSCLC, Appendix B) during the study period

from January 1, 2007 to December 31, 2018. Patients with other primary malignancies prior to the index date or enrolled in a randomized controlled trial on BTA were excluded.

We included patients who had a BM either at their first diagnosis of primary cancer or developed BM at any time after initial primary cancer diagnosis. The BM diagnoses were identified via a BTA record, direct coding of BM, and query of unstructured text from imaging, pathology, and clinical summary reports using a Natural Language processing (NLP) approach. The NLP platform Interactive Information Extraction (I2E), that was developed by the company Linguamatics (https://www.linguamatics.com/products/i2e), was used to automate reviewing of unstructured text by 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.

Phase II: Assessment of BTA use

Study population

From Phase I, all adult patients with BC, NSCLC or PC and a confirmed diagnosis of BM (identified from January 1, 2007 to December 31, 2018) were followed from BM diagnosis date to June 30, 2020 or death.

BTA treatment

BTA treatment was determined through patients EMRS linked to the hospital pharmacy dispensing database JAC covering both in-patient and out-patient prescriptions, and the treatment prescribing database ChemoCare. We reported three phases of medication adherence (initiation, implementation, and persistence) as recommended by the European Society for Patient Adherence, COMpliance, and Persistence (ESPACOMP)[15]. BTAs included two different classes of anti-resorptive agents: bisphosphonates (both IV and oral) and the RANKL inhibitor denosumab.

Statistical analysis

Primary cancers, BM cases, and BTA use, including type of BTA and switches between BTAs, were reported as counts and percentages. Patient characteristics were reported as percentages for categorical variables and medians (Q1, Q3) for continuous variables. The Kaplan Meier method was applied to analyze time-to-event data of BTA records, such as time to first BTA, duration of BTA, and time from last BTA to death. Counts of <6 were marked as such in all results to protect patient privacy. The SAS version 9.4 (SAS, CARY, NC, USA) and R version 3.2[16] was used for all data management and statistical analyses.

Patient and public involvement

Patients were not involved in this study.

RESULTS

Phase I: Identification of BM diagnosis

In Phase I, we identified a total of 6,142 BC, 5,202 NSCLC and 5,382 PC primary cancer patients. Table 1 shows a summary of the different approaches and corresponding results for identifying BM diagnoses. Each of these approaches were reviewed to ascertain confirmation of a BM diagnosis: direct identification by NLP, identification by proxy based on a record of BTA treatment, identification by proxy based on a record of spinal cord compression, and direct identification via diagnosis codes in structured EMR.

Table 1 shows the numbers and percentages of the three different methods of BM identification: NLP of unstructured data, evidence of spinal cord compression and BM in coded EMR fields. For BC, 573 patients were identified, with 527 (92%) via NLP-based querying of unstructured data. For NSCLC the total was 899, with 829 (92%) from unstructured data. For PC the total was 1017 and the results for unstructured data were 963 (95%). Further clinical expert review of all resulting cases detected additional false positives and yielded a final study cohort for BC, NSCLC and PC: 559 (9% of all primary cancer cases), 894 (17%) and 1013 (19%) BM patients, respectively.

Table 1 Attrition table of study patient population in phase I

Patient size by BM Method	**N**	**%**	**N**	**%**	**N**	**%**
Eligible Cohort at feasibility stage	**6142**	**100**	**5202**	**100**	**5382**	**100**
Patient cohort identified with BM in the medical records	573	9.3	899	17.3	1017	18.9
Patients identified by NLP of reports *	527	92.0	829	92.2	963	94.7
Patients identified as receiving BTA treatment *	309	53.9	118	13.1	129	12.7
Patients identified as patient having SCC *	19	3.3	41	4.6	42	4.1
Patients identified in coded EMR field *	49	8.6	<75	-	55	5.4
Final cohort of patients with BM after further clinical review	559	9.0	894	17.0	1013	18.8

*Number of patients identified as BM by each method of the overall eligible cohort and confirmed after review by two clinical physicians (patient can be identified in multiple methods)

Abbreviations: BM: Bone metastasis; BTA: Bone-targeting agents; EMR: Electronic medical record; NLP: Natural language processing; NSCLC: Non-small cell lung cancer; SCC: Spinal cord compression.

Phase II: Assessment of BTA use

Table 2 shows the patient demographic and clinical characteristics as well as the treatment histories of the final study cohort stratified by tumour type, and by BTA use/non-use. BTA initiation, implementation and persistence are shown in Figure 2, followed by further details of the two most frequent BTAs in Table 3.

Table 2 Patient characteristics of final BM patient cohort

Characteristics	BC		NSCLC		PC	
N	559		894		1013	
Follow-up: median days (Q1-Q3)	458 (128-933)		87 (37-205)		682 (357)	
Age at BM diagnosis, median (Q1-Q3)	65 (52-76)		69 (62-77)		75 (62-77)	
Stage IV at primary cancer diagnosis	30%		86%		72%	
History of SRE at BM diagnosis						
In the 56-day pre-BM-diagnosis period	35%		94%		25%	
Less than 16 days before BM diagnosis	1%		3%		1%	
Within 16-32 days before BM diagnc	3%		8%		5%	
	BTA	**no BTA**	**BTA**	**no BTA**	**BTA**	**no BTA**
N	294	265	117	777	121	892
Sex-female	100%	100%	45%	45%	0%	0%
Primary Cancer stage IV	29%	30%	85%	86%	74%	71%
CRPC diagnosis	#*	#*	#*	#*	82%	56%
At BM diagnosis	#	#	#	#	#*	#*
ECOG present	64%	39%	86%	82%	30%	19%
ECOG 0-2	94%	81%	67%	49%	27%	18%
eGFR** median (Q1-Q3)	81 (67-90)	83 (60-90)	90 (72-90)	86 (66-90)	82 (65-90)	73 (57-90)
eGFR** <60	15%	22%	14%	17%	17%	26%
Hypercalcaemia* classification at BM diagnosis**						
<2.75 mmol/L	90.51%	88.33%	82.20%	94.27%	75.21%	71.65%
Mild	3.05%	#	6.78%	2.00%	0.00%	#
Moderate	#	#	6.78%	#	0.00%	#
Severe	#	0.00%	#	#	0.00%	0.00%
Missing/Unknown	#	8.95%	#	2.40%	24.79%	27.76%
Renal disease						
Yes	#	3.4%	#	3.6%	0.8%	3.8%
missing	70.4%	56.6%	23.1%	21.5%	57.0%	47.2%
CCI						
0	21.4%	31.3%	57.3%	51.1%	33.1%	39.4%
1	5.4%	4.2%	14.5%	16.6%	4.1%	5.9%
2	#	4.9%	#	6.8%	5.0%	4.9%
3+	#	3.0%	#	4.0%	0.8%	2.6%
missing	70.4%	56.6%	23.1%	21.5%	57.0%	47.2%
Estrogen receptor status						
Positive	84.4%	72.1%	#	#	#	#
Missing	#	3.0%				
Progesterone receptor status						
Positive	65.0%	55.1%	#	#	#	#
Missing	6.1%	6.4%				
HR/HER2 status						
HR-/HER2-	10.0%	17.0%	#	#	#	#
HR-/HER2+	3.4%	6.8%	#	#	#	#
HR+/HER2-	58.2%	46.4%	#	#	#	#
HR+/HER2+	5.8%	5.3%	#	#	#	#
missing	22.8%	24.5%	#	#	#	#

Characteristics	BC		NSCLC		PC	
EGFR mutation						
Pathogenic	#	#	9.4%	5.2%	#	#
Wildtype	#	#	45.3%	29.6%	#	#
missing	#	#	45.3%	65.3%	#	#
ALK mutation						
Pathogenic	#	#	#	2.2%	#	#
Wildtype	#	#	41.0%	21.8%	#	#
missing	#	#	55.6%	76.1%	#	#
PDL1 mutation						
High (>=50%)	#	#	#	4.1%	#	#
Intermediate (1-49%)	#	#	10.3%	#	#	#
Low (<1%)	#	#	13.7%	5.0%	#	#
Not done	#	#	25.6%	19.1%	#	#
missing	#	#	44.4%	68.5%	#	#
Histopathological stage						
Squamous-cell carcinoma	#	#	23.1%	15.6%	#	#
Other specified NSCLC	#	#	#	2.7%	#	#
NSCLC NOS	#	#	12.0%	16.6%	#	#
Non-squamous NSCLC	#	#	49.6%	39.6%		
missing			11.1%	25.5%		
Therapy before BTA administration* **						
None	12.9%	10.6%	59.8%	68.5%	63.6%	72.9%
RT	#	#*	#	2.3%	8.3%	6.1%
RT & Surgery	15.0%	11.7%	#	1.9%	5.8%	4.2%
SACT*****	#	0	7.7%	2.3%	#	#*
SACT & RT	#	0	#	2.1%	#	0.8%
SACT & Surgery	11.9%	7.9%	#	3.5%	#	#
SACT, RT & Surgery	32.7%	33.2%	6.8%	3.2%	#	#
Surgery	25.9%	34.7%	13.7%	16.2%	16.5%	14.6%

* # means <6 patients
** eGFR units: ml/min/1.73m2)
*** 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+
**** The time period for these therapies includes the time from primary cancer to BM diagnosis
***** SACT was cancer-specific and included chemotherapy, endocrine therapy and targeted therapy.
Abbreviations: ALK: anaplastic lymphoma kinase; BM: Bone metastasis, BTA: Bone-targeting agent; CCI: Charlson Comorbidity Index; CRPC: castration-resistant prostate cancer; eGFR: Estimated glomerular filtration rate; ECOG: Eastern Cooperative Oncology Group; HER: Human Epidermal Growth Factor Receptor; HR: hormone receptor; NA: Not available; NOS: Non-otherwise specified; NSCLC: Non-small cell lung cancer; PD-L1: Programmed cell death receptor ligand-1; RT: radiotherapy; SACT: Systemic anticancer therapy

Breast Cancer

Among 559 patients with BC and BM, 47% (n=265) did not have a BTA prescription, and 53% (n=294) received at least one BTA prescription, starting a median (Q1, Q3) of 65 (27, 167) days from their BM diagnosis date (inclusive) to their first BTA initiation date (excludes 9 patients with a BTA before BM). Median (Q1, Q3) duration of BTA therapy from first to last BTA record was 481 (188, 816) days and median (Q1, Q3) time from last BTA to death was 54 (26, 109) days (Figure 2) . Most patients (86%, n= 254) received only one type of BTA. Table 3 provides details of two specific BTAs of different classes that were administered, the RANKL inhibitor denosumab (n=56, 19.1%) and the bisphosphonate zoledronic acid (n=229, 77.9%), both with the most frequent cycle duration of 28 days. During the follow-up period, a total

of 52 switches were observed between BTAs. Of those, switches between denosumab and zoledronic acid were the most frequent: 30% (17/56) of all denosumab administrations ended in a switch to zoledronic acid, within a median (Q1, Q3) time to switch of 32 (28, 57) days, and 5% (11/229) of all zoledronic acid administrations ended in a switch to denosumab, within a median (Q1, Q3) time of 78 (35, 216) days. Patients with BTAs had a numerically higher percentage of oestrogen receptor status positive, progesterone receptor status positive, human epidermal growth factor receptor (HR+/HER)- status compared to patients without a BTA (Table 2).

NSCLC

Among the 894 patients with NSCLC and BM, 87% (n=777) did not receive a BTA prescription and 13% (n=117) received at least one BTA prescription, starting a median (Q1, Q3) of 60 (28, 162) days from their BM diagnosis date (inclusive) to their BTA initiation date (excludes 8 patients with a BTA before BM). Median (Q1, Q3) duration of BTA therapy from first to last BTA record was 89 (49, 195) days and median (Q1, Q3) time from last BTA to death was 38 (16, 98) days (Figure 2) . A total of 12 patients with NSCLC received denosumab and 93 patients received zoledronic acid (Table 3), both with the most frequent cycle duration of 28 days. The median number of administrations per patient was 2 (Q1,Q3: 1,11) for denosumab, and 1 (Q1,Q3: 1,3) for zoledronic acid. A total of 114 (97%) patients received only one type of BTA and <6 switches occurred between BTAs. Patients with BTAs had a numerically higher percentage of estimated glomerular filtration rate (eGFR), anaplastic lymphoma kinase (ALK) and PD-L1 mutation data missing as well as a higher percentage of a history of overall RT or SACT or surgery compared to patients without a BTA (40.2% (=100%-59.8%) vs 31.5% (=100%-68.5%)) (Table 2).

Prostate Cancer

Among the 1013 patients with PC and BM, 88% (n=892) did not receive a BTA prescription and 12% (n=121) received at least one BTA prescription, starting a median (Q1, Q3) of 611 (295, 980) days from their BM diagnosis date (inclusive) to their BTA initiation date (excludes 1 patient with a BTA before BM). Median (Q1, Q3) duration of BTA therapy from first to last BTA record was 115 (53, 193) days and median (Q1, Q3) time from last BTA to death was 112 (44, 218) days (Figure 2) There were no patients on denosumab while 113 patients received zoledronic acid (Table 3), with the most frequent cycle duration of 28 days. The median number of administrations per patient was 2 (Q1, Q3: 1, 4) for zoledronic acid. PC BTA patients only had a record of one unique BTA with no switching recorded. Patients with BTA prescriptions had a numerically higher percentage of history of overall RT or SACT or surgery compared to patients without a BTA prescription (36.4% (=100%-63.6%) vs 17.1% (=100%-72.9%)) (Table 2).

Table 3 BTA administration in BM patients across the 3 cancers

		BC	NSCLC	PC
Total unique agents	**N**	295	117	121
	Median	1.0	1.0	1.0
	Min-Max	1.0-3.0	1.0-2.0	1.0-1.0
Denosumab Administrations	**N**	56 (19.1%)	12 (10.3%)	0 (0.0%)
	Only received once	11	<6	-
	Median	6.50	2.0	-
	Min-Max	1.0-61.0	1.0-14.0	-
Zoledronic acid administrations	**N**	229 (77.9%)	93 (79.5%)	113 (93.4%)
	Only received once	32	52	51
	Median	9.0	1.0	2.0
	Min-Max	1.0-50.0	1.0-21.0	1.0-34.0

Abbreviations: BC: breast cancer; NSCLC: non-small cell lung cancer; PC: prostate cancer
The remaining %s that add up to the total 100% in the table include BTAs other than denosumab or zoledronic acid.

DISCUSSION

Use of structured and unstructured data to identify BM patients within LTHT

In this study, over 90% of all BM cases were identified through NLP-based querying of unstructured data HealthCare professionals typically record BM detected during different diagnostic procedures in both structured and unstructured formats. Restricting the analysis to structured medical codes would have significantly underestimated the occurrence of BM in the three cancer cohorts in this hospital-based setting. Hence, use of NLP greatly enhanced the efficiency of the identification of BM cases from multiple unstructured data sources. The need for clinical review to eliminate false positive cases shows that further refinement of NLP models is still required.

BTA usage in patients with BC and BM

A European multi-country study (Von Moos et al.[17]) found that 88% of BC patients with BM received BTA treatment, while 53% of BC patients with BM received BTA treatment in our study. There are key differences between the studies. The Von Moos et al. study collected data in a cross-sectional survey of physicians that were treating BM patients who were actively receiving treatment for their cancer. In contrast, 57% of all BM patients in our study did not have a record of SACT, even though some of it may be a result of prescribing recorded outside the available data systems. A prospective study using a German tumour

registry (Schroder et al.[18]) reported a BTA treatment of 89% in BC patients with BM with a median time to treatment from BM diagnosis of 3 weeks. Data collection was prospective and focused on an anticancer treated cohort, including out-patient treatment data. In contrast, our study obtained treatment data retrospectively from potentially incomplete hospital treatment databases and did not include treatment outside the hospital. Furthermore, there may be genuine differences in the use of BTAs in cancer patients with BM between the UK and other European countries. Determinants of BTA prescribing in cancer patients with BM were evaluated in several studies [17, 19, 20]. Findings from Von Moos et al.[17] indicate that some physicians base their BTA treatment decisions not only on clinical guidelines but also consider patients' Eastern Cooperative Oncology Group (ECOG) performance score, disease burden, and the presence of other sites of metastatic disease. For example, a patient with an ECOG performance score of 0-2 is considered fit enough to receive BTA treatment, but in the presence of extensive liver disease and low burden of bone disease, may not routinely receive a BTA. Our study showed a numerical difference in ECOG scores between patients with BTA and without BTA: 94% of patients with a BTA and 81% without a BTA had an ECOG score of 0-2. These findings suggest that BTA treatment is determined on a case-by-case basis within this setting and is not solely reliant on BTA guidelines.

BTA usage in patients with NSCLC and BM

Diel et al.[21] investigated 242 lung cancer patients with a diagnosis of BM and who received at least one BTA treatment in Germany from 2011 to 2015. Of these patients, 15% received denosumab and 63% zoledronic acid, while our study observed 10% of NSCLC BM patients receiving denosumab and 80% zoledronic acid. The probability of patients still on denosumab after 6 months was 87% in Diel et al., compared to 13% in our study. The 2014 European Society for Medical Oncology (ESMO) bone health guidelines, which cover some of the German study time period recommend zoledronic acid or denosumab in patients with a life expectancy of greater than 3 months[10]. The NSCLC patients within our study had a median follow-up time of 87 (Q1, Q3: 37,205) days.

The low proportion of patients receiving BTA within the current study is likely due to the poor prognosis of these patients. Overall, survival data published by LTHT on advanced non-squamous NSCLC patients (stage IIIB-IV) showed patients had a median survival of 4.1 months between 2007-2012 and 5.0 months

between 2013-2017 [22]. The ECOG score further reflects the burden of disease in this population: 67% of BTA patients with a score of 0-2 and 49% in non-BTA patients.

BTA usage in patients with PC and BM

The European multi-country study (von Moos, et al. [17]) included an evaluation of castrate-resistant prostate cancer (CRPC) patients and reported that 77% of CRPC patients received at least one BTA. The von Moos et al. study included patients who were actively receiving anticancer therapy. In our study, over 71% were diagnosed at stage IV and had BM at the time of PC diagnosis and 72% had no record of any other treatment such as SACT or surgery. A US-based study using claims and commercial databases, (Hernandez, et al [7]) identified BTA use in 52% of PC patients with BM in 2012. Median time to first BTA was 35 and 37 days, respectively for the claims and commercial databases. In our study we observed that 12% of BM patients received a BTA, with a median time to first BTA of 610 days. While the National Institute for Health and Care Excellence (NICE) guidelines do not recommend denosumab for PC patients with BM in the UK [12], it is approved for use in PC patients in the US [23].

STRENGTHS AND LIMITATIONS

A key strength of the current study is the use of extensive unstructured data from multiple EMR sources and application of NLP techniques to identify patients with BM. Leveraging unstructured data is especially important because bone metastases are likely identified at different diagnostic investigations and reported in different medical records. Access to multiple data sources and linkage within the LTHT database and the application of NLP methods enabled a more comprehensive account of the patient's medical record data. Our findings show that the vast majority of BM cases would have been missed without evidence from unstructured medical record data, as BMs are typically not recorded through structured medical codes in this particular setting. In addition, the availability of both in-patient and out-patient prescribing data from multiple data sources is a strength of the study.

Nevertheless, the study has some limitations due to the capture and documentation of in-patient BTA prescribing information. The comprehensive hospital drug dispensing data (JAC) is only available for the last 5 years. Although BTA treatment is also included in the oncology treatment database ChemoCare, BTA

treatment is not always recorded within ChemoCare, especially for patients who receive a BTA during an in-patient admission. Hence, medications that were not prescribed using ChemoCare, including hormone therapy, and that were prescribed more than 5 years ago (not included in JAC), are not captured in this study. However, an assessment of BTA prescribing before and after the period of JAC availability showed only a marginal difference in BTA prescribing between the two periods. In addition, insights from this study are limited to the routine practice in the UK and reflect existing restrictions in reimbursement and access to BTA therapy within the country.

CONCLUSION

To our knowledge, this is the first study that retrospectively identified BM patients using both structured and unstructured data within England to characterize BTA use in clinical practice. Applying NLP to unstructured data should be considered as a useful additional strategy to identify BM and ascertain cases which would have been missed if only structured data were used. This study provided a different picture to existing literature on BTA use in Europe and the US, highlighting the underuse of BTA treatment within patients with metastatic bone disease from BC, NSCLC or PC. These findings point to a complex decision-making process to prescribe bone protection therapy to cancer patients. Further work is warranted to better understand individual patient medical need and treatment benefit, including repeating this work in other data sources to assess the benefit of using unstructured data.

ACKNOWLEDGEMENT

The authors acknowledge the contribution of Nina Snelling[3] to the research work.

AUTHOR CONTRIBUTIONS

AS, AA, PM and SC made substantial contributions to the design of the work.

BL, MR, MT, and PE contributed to the analysis of the data.

AS, AA, PE, and SC contributed to the interpretation of data. JW drafted the work; AS and PE, made substantial contributions to substantively revise the manuscript. All authors reviewed the manuscript, provided final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This study was funded by Amgen Ltd. The grant number is N/A

COMPETING INTERESTS

I have read the journal's policy and the authors of this manuscript have the following competing interests. Anouchka Seesaghur was an employee and equity holder in Amgen Inc during the conduct of the study. Peter Egger, Bethany Levick and Matthew Thompson were employed with IQVIA during conduct of the study. Joshua Warden had no conflict of interest to declare. Ali Abbasi reported contract work with Amgen Inc. Majid Riaz was employed with IQVIA and had an honorary contract with Leeds Teaching Hospitals Trust (LTHT) to access the data to produce analysis. Peter McMahon worked for IQVIA during the initial development of the manuscript, and the analysis time period. Sue Cheeseman's part was funded by IQVIA.

ETHICS APPROVAL

In this retrospective study, all data were fully anonymised, and no participant consent was required. 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).

DATA SHARING STATEMENT

No additional data available

REFERENCES

1. Pulido C VI, Ferreira AR. Bone metastasis risk factors in breast cancer. *Ecancermedicalscience*. 2017;
2. Xiong Z, Deng G, Huang X, et al. Bone metastasis pattern in initial metastatic breast cancer: a population-based study. *Cancer Manag Res*. 2018;10:287-295. doi:10.2147/cmar.S155524
3. Santini D, Barni S, Intagliata S, et al. Natural History of Non-Small-Cell Lung Cancer with Bone Metastases. *Sci Rep*. Dec 22 2015;5:18670. doi:10.1038/srep18670
4. D'Antonio C, Passaro A, Gori B, et al. Bone and brain metastasis in lung cancer: recent advances in therapeutic strategies. *Ther Adv Med Oncol*. May 2014;6(3):101-14. doi:10.1177/1758834014521110
5. Liu D, Kuai Y, Zhu R, et al. Prognosis of prostate cancer and bone metastasis pattern of patients: a SEER-based study and a local hospital based study from China. *Sci Rep*. Jun 4 2020;10(1):9104. doi:10.1038/s41598-020-64073-6
6. Macedo F LK, Pinho F, Saraiva N. Bone Metastases: An Overview. *Oncol Rev*. 2017;
7. Hernandez RK, Adhia A, Wade SW, et al. Prevalence of bone metastases and bone-targeting agent use among solid tumor patients in the United States. *Clin Epidemiol*. 2015;7:335-45. doi:10.2147/clep.S85496
8. Asdahl PH, Sundbøll J, Adelborg K, et al. Cardiovascular events in cancer patients with bone metastases-A Danish population-based cohort study of 23,113 patients. *Cancer Med*. Jul 2021;10(14):4885-4895. doi:10.1002/cam4.4027
9. Hjelholt TJ, Rasmussen TB, Seesaghur A, et al. Risk of infections and mortality in Danish patients with cancer diagnosed with bone metastases: a population-based cohort study. *BMJ Open*. Dec 1 2021;11(12):e049831. doi:10.1136/bmjopen-2021-049831
10. Coleman R, Body JJ, Aapro M, Hadji P, Herrstedt J. Bone health in cancer patients: ESMO Clinical Practice Guidelines. *Ann Oncol*. Sep 2014;25 Suppl 3:iii124-37. doi:10.1093/annonc/mdu103
11. Coleman R, Hadji P, Body JJ, et al. Bone health in cancer: ESMO Clinical Practice Guidelines. *Ann Oncol*. Dec 2020;31(12):1650-1663. doi:10.1016/j.annonc.2020.07.019
12. Overview | Denosumab for the prevention of skeletal-related events in adults with bone metastases from solid tumours | Guidance | (NICE) (2012).
13. Jensen A, Nørgaard M, Yong M, Fryzek JP, Sørensen HT. Validity of the recorded International Classification of Diseases, 10th edition diagnoses codes of bone metastases and skeletal-related events in breast and prostate cancer patients in the Danish National Registry of Patients. *Clin Epidemiol*. Aug 9 2009;1:101-8. doi:10.2147/clep.s5446
14. Liede A, Hernandez RK, Roth M, Calkins G, Larrabee K, Nicacio L. Validation of International Classification of Diseases coding for bone metastases in electronic health records using technology-enabled abstraction. *Clin Epidemiol*. 2015;7:441-8. doi:10.2147/clep.S92209
15. De Geest S, Zullig LL, Dunbar-Jacob J, et al. ESPACOMP Medication Adherence Reporting Guideline (EMERGE). *Ann Intern Med*. Jul 3 2018;169(1):30-35. doi:10.7326/m18-0543
16. Team RC. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. https://www.R-project.org/.
17. von Moos R, Lewis K, Massey L, Marongiu A, Rider A, Seesaghur A. Initiation of bone-targeted agents in patients with bone metastases and breast or castrate-resistant prostate cancer actively treated in routine clinical practice in Europe. *Bone*. Jan 2022;154:116243. doi:10.1016/j.bone.2021.116243
18. Schröder J, Fietz T, Köhler A, et al. Treatment and pattern of bone metastases in 1094 patients with advanced breast cancer - Results from the prospective German Tumour Registry Breast Cancer cohort study. *Eur J Cancer*. Jul 2017;79:139-148. doi:10.1016/j.ejca.2017.03.031

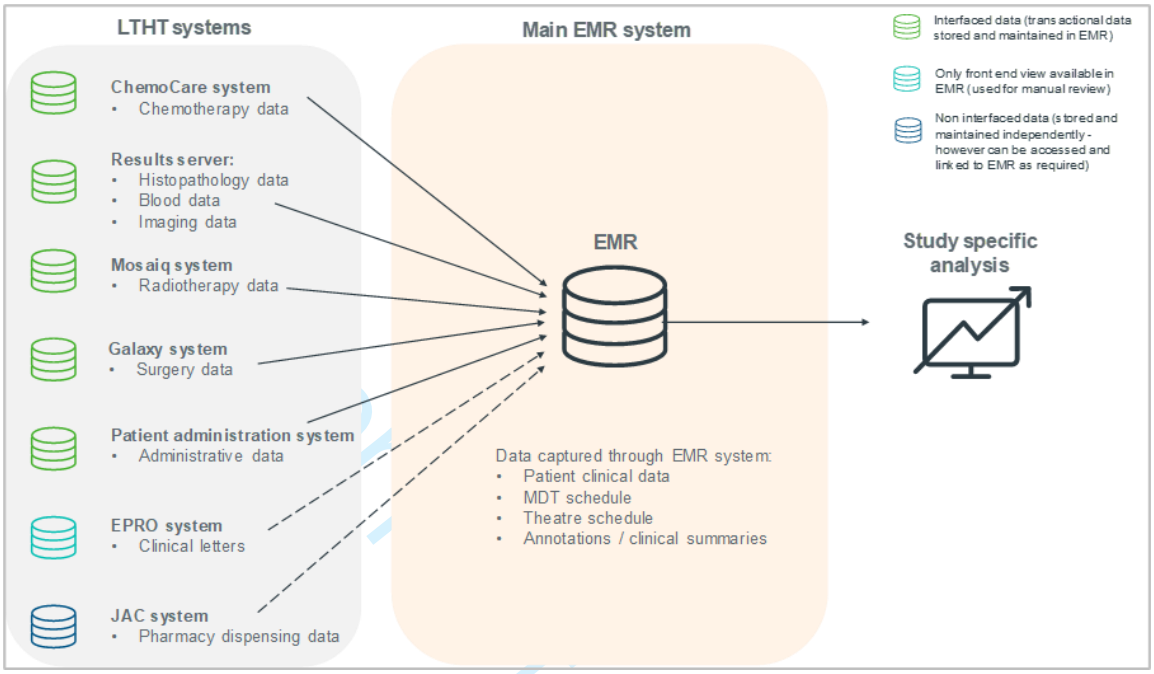
19. Butler AM, Cetin K, Hernandez RK, et al. Treatment dynamics of bone-targeting agents among men with bone metastases from prostate cancer in the United States. *Pharmacoepidemiol Drug Saf*. Feb 2018;27(2):229-238. doi:10.1002/pds.4360
20. McGrath LJ, Overman RA, Reams D, et al. Use of bone-modifying agents among breast cancer patients with bone metastasis: evidence from oncology practices in the US. *Clin Epidemiol*. 2018;10:1349-1358. doi:10.2147/clep.S175063
21. Diel I, Ansorge S, Hohmann D, Giannopoulou C, Niepel D, Intorcia M. Real-world use of denosumab and bisphosphonates in patients with solid tumours and bone metastases in Germany. *Support Care Cancer*. Nov 2020;28(11):5223-5233. doi:10.1007/s00520-020-05357-5
22. Snee M CS, Thompson M, . Temporal Trends in Treatment and Overall Survival Among Patients With Incident NSCLC in the UK: A REAL-Oncology Database Analysis From the I-O Optimise Initiative. presented at: 9th European Lung Cancer Congress (ELCC); 2019; Geneva, Switzerland. https://www.leedsth.nhs.uk/assets/ca7effdffd/28.-Snee_REAL-ONC_44P_ELCC2019_FINAL-Poster_4Apr19.pdf
23. FDA Approves Denosumab for the Treatment of Bone Loss in Patients With Prostate or Breast Cancer. Cancer Network: FDA; 2011.

Figures

Figure 1 Leeds NHS Teaching Hospitals Trust (LTHT) data sources and linkages to create study dataset

Figure 2 BTA adherence: initiation, implementation, and persistence

Figure 1 Leeds NHS Teaching Hospitals Trust (LTHT) data sources and linkages to create study dataset



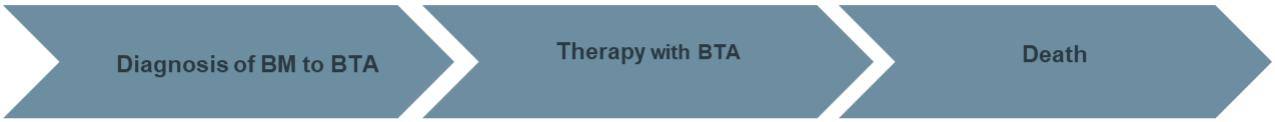
Some of the LTHT systems are specific to LTHT and some of them are commercially available. Further references: Mosaiq® (https://www.elekta.com/products/oncology-informatics/mosaiq-plaza/) and EPRO (https://epro.com/).

Figure 2 BTA adherence: initiation, implementation, and persistence

BC: 10 weeks
NSCLC: 9 weeks
PC: 87 weeks

BC: 68 weeks
NSCLC: 13 weeks
PC: 16 weeks

BC: 8 weeks
NSCLC: 5 weeks
PC: 16 weeks



Start of therapy		Duration of therapy		End of therapy	
Time from BM to first record of BTA Therapy (days)		Time from start of BTA therapy to end of BTA Therapy (days)		Time from last BTA to death (days)	
Breast Cancer Patients	**N=559**	**Breast Cancer Patients**	**N=294***	**Breast Cancer Patients**	**N=294***
BTA >0 from BM	285 (51.0%)	1> BTA**	250 (84.8%)	Deaths	225 (76.3%)
Median (Q1-Q3)	**65 (27-167)**	**Median (Q1-Q3)**	**481 (188-816)**	**Median (Q1-Q3)**	**54 (26-109)**
NSCLC Patients	**N=894**	**NSCLC Patients**	**N=117**	**NSCLC Patients**	**N=117**
BTA >0 from BM	109 (12.2%)	1> BTA**	54 (46.2%)	Deaths	111 (94.9%)
Median (Q1-Q3)	**60 (28-162)**	**Median (Q1-Q3)**	**89 (49-195)**	**Median (Q1-Q3)**	**38 (17- 98)**
Prostate Cancer Patients	**N= 1013**	**Prostate Cancer Patients**	**N= 121**	**Prostate Cancer Patients**	**N= 121**
BTA >0 from BM	120 (11.9%)	1> BTA**	63 (52.1%)	Deaths	107 (88.4%)
Median (Q1-Q3)	**611 (295-980)**	**Median (Q1-Q3)**	**115 (53-192)**	**Median (Q1-Q3)**	**112 (44-218)**

* includes patients who had the BTA before their BM diagnosis
** number of patients with a duration of at least one day

Title : 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

Appendix A

ICD-10 diagnosis codes for the primary cancers

Condition	ICD10 code	ICD10 Description
Breast Cancer (BC)	C50	Malignant neoplasm of breast
Non-small cell lung cancer (NSCLC)	C34	Malignant neoplasm of lung + morphology codes to identify NSCLC subgroups in Appendix B below
Prostate Cancer	C61	Malignant neoplasm of prostate

Appendix B

ICD-10 Morphology codes for Adenocarcinoma (NON-Squamous NSCLC, Squamous-cell Carcinoma and NSCLC NOS)

	ICD-0-2 Morphology codes
Adenocarcinoma (non-squamous NSCLC)	Adenocarcinoma UNS 81403 Enteric adenocarcinoma 81443 Solid adenocarcinoma with mucin production 82303 MANEC mixed adenoneuroendocrine carcinoma 82443 Adenocarcinoma, bronchiolo-alveolar (BAC), bronchiolar carcinoma, (incl pathologic in situ-variant) 82503 Alveolar adenocarcinoma 82513 Bronchio-alveolar carcinoma 82523 Adenocarcinoma in situ, mucinous 82532 Adenocarcinoma, mucinous bronchiolo-alveolar (BAC) 82533 Bronchio-alveaolar carcinoma, mixed mucinous and non-mucinous 82543 Adenocarcinoma, mixed with other types of carcinoma incl. squamous cell and small-cell carcinoma 82553 Minimally invasive adenocarcinoma, nonmucinous 82563 Minimally invasive adenocarcinoma, mucinous 82573 Papillary adenocarcinoma, NOS 82603 Micropapillary adenocarcinoma 82653 Clear cell adenocarcinoma 83103 Fetal adenocarcinoma 83333 Mucinous cystadenocarcinoma 84703 Mucinous adenocarcinoma 84803 Mucin-producing adenocarcinoma 84813 Signet ring cell carcinoma 84903 Acinar cell carcinoma 85503 Acinar adenocarcinoma 85513
Squamous-cell carcinoma	Papillary squamous cell carcinoma 80523 Keratinizing squamous cell carcinoma 80713 Non-keratinizing squamous cell carcinoma 80723 Squamous cell carcinoma, small cell nonkeratinizing 80733 Squamous cell carcinoma, spindle cell 80743

	Basaloid squamous cell carcinoma 80833 Squamous cell carcinoma, clear cell type 80843
NSCLC NOS	Carcinoma, NOS 80103 Carcinoma, undifferentiated NOS 80203 Carcinoma, anaplastic NOS 80213 Carcinoma, non-small cell unspecified 80463 Large cell carcinoma with rhabdoid phenotype 80143 Sarcomatoid carcinoma, pleomorphic 80223 NUT carcinoma 80233 Spindle cell and giant cell carcinoma 80303 Giant cell carcinoma 80313 Spindle cell carcinoma, NOS 80323 Pseudosarcomatous carcinoma 80333 Basaloid carcinoma 81233 Adenocystic carcinoma 82003 Mucoepidermoid carcinoma 84303 Adenosquamous carcinoma 85603 Epithelial-myoepithelial carcinoma 85623 Blastoma, pulmonary (pneumoblastoma) 89723 Carcinosarkoma, NOS 89803 Myoepithelial carcinoma 89823
Large cell carcinoma (Non-squamous NSCLC)	Large-cell carcinoma, unspecified 80123

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

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cohort studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	5, 6
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	Not applicable
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5, 6, 7
Bias	9	Describe any efforts to address potential sources of bias	2, 3
Study size	10	Explain how the study size was arrived at	Not applicable
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7
		(*b*) Describe any methods used to examine subgroups and interactions	Not applicable
		(*c*) Explain how missing data were addressed	Not applicable

		(*d*) If applicable, explain how loss to follow-up was addressed	Not applicable
		(*e*) Describe any sensitivity analyses	Not applicable
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	Not applicable
		(c) Consider use of a flow diagram	8
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9
		(b) Indicate number of participants with missing data for each variable of interest	9, 10
		(c) Summarise follow-up time (eg, average and total amount)	9, 10
Outcome data	15*	Report numbers of outcome events or summary measures over time	9
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	9 - 13
		(*b*) Report category boundaries when continuous variables were categorized	9, 10
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Not applicable
Discussion			
Key results	18	Summarise key results with reference to study objectives	13-15
Limitations			
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	15-16
Generalisability	21	Discuss the generalisability (external validity) of the study results	15-16
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	18

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.