BMJ Open

Cohort Profile: Epidemiological Strategy and Medical Economics (ESME) Research Program, the French Metastatic Breast Cancer (MBC) Cohort

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023568
Article Type:	Cohort profile
Date Submitted by the Author:	12-Apr-2018
Complete List of Authors:	Pérol, David; Léon Bérard Cancer Centre, Unit of Biostatistics and Therapeutic Evaluation Robain, Mathieu; R&D UNICANCER Arveux, F; Registre des cancers du sein de Côte d'Or, centre Georges-François-Leclerc Mathoulin-Pélissier , S; INSERM, Chamorey, Emmanuel; Centre Antoine-Lacassagne Asselain, Bernard Berchery, Delphine; Institut Claudius Regaud Gourgou, Sophie; Institut regional du Cancer de Montpellier Breton, Mathias; Centre Eugene Marquis Delaine-Clisant , Stéphanie; Centre Oscar Lambret, Clinical Research Unit Mons, Muriel; Gustave Roussy Diéras, Véronique; Ensemble Hospitalier de l'Institut Curie Carton, Matthieu; Institut Curie, Biométrie; Guizard, Anne-Valérie ; Centre Francois Baclesse Centre de Lutte Contre le Cancer Laborde, Lilian; Institut Paoli-Calmettes Laurent, Carine; Institut de Cancerologie de Lorraine Loeb, Agnès; Centre Henri Becquerel Mouret-Reynier, Marie-Ange ; Centre Jean Perrin Parent, Damien; Institut Jean Godinot Perrocheau, Geneviève; Institut de Cancérologie de l'Ouest Campion, Loïc; Institut de Cancérologie de l'Ouest Velten, Michel; Centre Paul Strauss Cailliot, Christian; R&D UNICANCER Ezzalfani, Monia; Centre Leon Berard Simon, Gaëtane; R&D UNICANCER,
Keywords:	Real-world data, Data Platform, Epidemiology < ONCOLOGY, Patient medical record, Quality control, Metastatic breast cancer

TITLE PAGE

Epidemiological Strategy and Medical Economics (ESME) Research Program, the French Metastatic Breast Cancer (MBC) Cohort: Design and current status

List of authors:

David Pérol (Centre Léon Bérard, Lyon), Mathieu Robain (R&D UNICANCER, Paris), Patrick Arveux (Centre Georges-François Leclerc, Dijon), Simone Mathoulin-Pelissier (Institut Bergonié, Bordeaux), Emmanuel Chamorey (Centre Antoine Lacassagne, Nice), Bernard Asselain (University Paris Sud), Delphine Berchery (Institut Claudius Régaud, Toulouse), Sophie Gourgou (Institut régional du Cancer Montpellier, Montpellier), Mathias Breton (Centre Eugène Marquis, Rennes), Stéphanie Clisant (Centre Oscar Lambret, Lille), Muriel Mons (Gustave Roussy, Villejuif), Véronique Diéras (Ensemble Hospitalier de l'Institut Curie, Paris & Saint-Cloud), Matthieu Carton (Ensemble Hospitalier de l'Institut Curie, Paris & Saint-Cloud), Anne-Valérie Guizard (Centre François Baclesse, Caen), Lilian Laborde (Institut Paoli-Calmettes, Marseille), Carine Laurent (Institut de Cancérologie de Lorraine, Nancy), Agnès Loeb (Centre Henri Becquerel, Rouen), Marie-Ange Mouret-Reynier (Centre Jean Perrin, Clermont-Ferrand), Damien Parent (Institut Jean Godinot, Reims), Geneviève Perrocheau (Institut de Cancérologie de l'Ouest, Nantes & Angers), Loïc Campion (Institut de Cancérologie de l'Ouest, Nantes & Angers), Michel Velten (Centre Paul Strauss, Strasbourg), Christian Cailliot (R&D UNICANCER, Paris), Monia Ezzalfani (Centre Léon Bérard, Lyon), Gaëtane Simon (R&D UNICANCER, Paris)

Corresponding author:

David Pérol (Centre Léon Bérard, Lyon)

E-mail: david.perol@lyon.unicancer.fr

Tel: +33 4.69.85.61.00

ABSTRACT

The global aim of the ESME research program is the centralisation of real-life data on oncology care for epidemiological research purposes. The first covered medical area was the metastatic breast cancer (MBC) using a retrospective data collection called ESME MBC Cohort. The main objective is to describe clinical features, treatment patterns and outcomes over the years. This population-based prospective cohort aims to select all consecutive patients treated for their MBC managed in the 18 French Comprehensive Cancer Centres (FCCCs) from 2008 to 2019. We aim to select 30,000 cases in this cohort. These real-world data will help standardise the management of MBC and improve patient care. Diagnostic, therapeutic and follow up data (demographics, primary tumor, metastatic disease, treatment patterns and vital status) were collected through the course of the disease. To date, over 16,700 patients initially treated for metastatic disease from 2008–2014 with possible follow up period until 2016. A dozen of ancillary research projects have been conducted and few of them are already accepted for publication or published.

Data collection is updated annually. Future aim is to link the data of the cohort to the French national Health Data System (SNDS) for centralizing data on healthcare reimbursement (drugs, medical procedures), inpatient/outpatient stays and visits in primary/secondary care settings. Finally, the ESME research program is expanding to two other areas of oncology: ovarian cancer (OC) and advanced/metastatic lung cancer (AMLC).

KEYWORDS

Real-world data; real-life data; patient medical record; quality control; oncology; metastatic breast cancer

BACKGROUND

Real-world evidence (RWE) studies and observational studies using real world data (RWD) play a growing role in conducting comparative effectiveness research on pharmaceutical products and other healthcare interventions. They aim to bridge the gap between the highly controlled environment of randomised clinical trials (RCTs) and real-life clinical practice [1]. In particular, health authorities are interested in gathering real-world data for long-term benefit-risk assessment and, increasingly, health economic evaluations for reimbursement decisions [2]. With their high internal validity, RCTs are considered the gold standard of evidence for establishing treatment efficacy, although the generalisability of findings to clinical practice may be limited [3]. In fact, cancer survival endpoints in the real-life setting may differ from that measured in traditional RCTs [4].. Furthermore, other limitations such as short follow-up and small sample size have created uncertainty in estimating survival criteria in RCTs, and thus surrogate endpoints are generally used such as progression free survival (PFS) or time to progression (TTP) [5]. On the other hand, large population-based cohorts with longer follow-up periods can be particularly appropriate to assess long-term clinical outcomes , such as overall survival (OS) outside of the RCT setting [6] and to detect changes in medical practice.

RWE and observational studies are non-experimental research where cancer management (treatments and disease evolution assessment) is left to the choice of care providers and patients [7]. Over the past few years, there was a requirement for high-quality data from large cohorts, which is now driving improvements in research methods and practices [8]. The goal of these studies is to generate complementary information to RCTs based on larger samples and provide answers regarding particular populations in real-life clinical practice. These studies are more prone to biases such as baseline differences between patients (selection bias) or bias due to confounding by indication for example. To minimise sources of such biases, statistical approaches including adjusted analyses, propensity score methods, or instrumental variables may be also employed [9-12].

Metastatic breast cancer (MBC) is one of the leading causes of cancer-related mortality among women in Western countries [13]. A relatively high proportion (approximately 30%) of breast cancer patients develop metastatic disease [14], and while significant treatment advances have been made, the overall prognosis is poor with a five-year survival rate of 25% [15]. The national academic network of cancer centres in France (French Comprehensive Cancer Centres; FCCCs), which together handle over one-third of all breast cancer cases nationally, decided to launch in 2014 a program dedicated focussing on RWD in oncology databases in MBC through Epidemiological Strategy and Medical Economics (ESME) research program.

In this paper, we describe the methodological principles that underpin the ESME research program, and illustrated by the first project, the ESME metastatic breast cancer (MBC) cohort. Design, current status and a preliminary description of population are presented.

MAIN OBJECTIVES

The global aim of the ESME research program is the construction of a comprehensive database on oncology care for epidemiological research purposes in order to improve knowledge on medical practice in real-life setting, on public health and healthcare use, and to provide information to health authorities and other associated bodies. For each cancer covered by the program, data platform will then be created to address this epidemiological research.

Regarding the ESME MBC project (ClinicalTrials.gov ; NCT03275311), the primary objective is to describe the medical care of patients treated for MBC according to their disease characteristics. In this context, medical care encompasses all surgical procedures, radiotherapy and drug treatments, focussing particularly on chemotherapy, targeted and endocrine therapies. A secondary objective of the ESME MBC project is to describe the evolution of the metastatic disease and the outcomes.

METHODS

The ESME MBC cohort is a population-based registry in 18 FCCCs (http://www.unicancer.fr/en/rd-unicancer/esme), which collected data on all consecutive patients treated for MBC from 2008. Annual data collection phases are planned to add new diagnosed cases and update patients' follow up data.

STUDY POPULATION

Eligibility criteria

Patients were eligible to the ESME MBC cohort according to the following criteria: male or female patients aged ≥18 years with MBC whose first metastasis was treated (either completely or partially) in an FCCC between January 1st, 2008 and December 31st, 2014. MBC treatment could include radiotherapy, chemotherapy, targeted therapy, immunotherapy, or endocrine therapy. Patient data follow up was considered on patient basis according to last contact information available in the patient medical records (PMR) at the day of the data collection. This data collection has been performed from October 2014 to October 2016 by trained research assistants.

Patient screening process

Patient screening process involved two phases: an automated case screening and then the validation of selection for each screened case. The automated case screening was based on information retrieved from multiple data sources available within each FCCC: administrative records [French National Computerised Medical Information System] (via MBC-specific International Classification of Diseases (ICD) codes associated with inpatient stays), pharmacy records, PMRs (including multidisciplinary team meeting records and search using relevant keywords), MBC-specific registries. The objective of the automated screening phase was to identify almost all cases meeting the selection criteria and generate the patient screening list. The ICD codes used were: C50 (Malignant neoplasm of breast), C77.- [except

C77.3] (Secondary and unspecified malignant neoplasm of lymph nodes), C78.- (Secondary malignant neoplasm of respiratory and digestive organs) and C79.- (Secondary malignant neoplasm of other and unspecified sites).

Once the patient screening list was finalised in each centre, each patient was given an anonymous ESME number. As the automated screening method did not allow the identification of the first MBC-specific treatment, eligibility criteria were cross-checked for all screened cases using data from the patient medical records. 45,329 cases were screened and 16,711 were selected in the cohort.

Baseline and follow up data

The ESME MBC Cohort is composed of 3 types of data (see **Figure 1**):

- *Patient-related data* are obtained from systematic review of patient medical records (non-structured data) and provide information on patient demographics, cancer family history, characteristics of primary tumour, relapses, metastatic recurrences pathological reports (tumor size, grade, histological type), hormone-receptor status and HER-2 status, therapeutic care (focussing on cancer-related treatment) and settings, reasons for treatment termination, and clinical events.
- *Hospitalisation records* are integrated data from a structured and automated database related to inpatient stays, and primarily used to bill the French National Health Insurance Fund (*Assurance Maladie*). It provides information on patient entry and discharge information, (date and destination code at discharge), ICD codes associated to each stay, diagnostic, medical and therapeutic procedures including radiotherapy and surgery.
- *Pharmacy-dispensed treatment records* includes all data related to anti-cancer treatments obtained from each centre's pharmacy database: drugs (International Non-proprietary Name), administration date, protocol name, patient's height and BMI, cycle ID, UCD code (pharmaceutical form related to the drug dosage), line of treatment, administration in a clinical trial (yes/no). It does not include information on products that are prescribed by the primary care practice.

The detailed raw data collected above and derived data are listed in **Table 1**.

Data management

Any data integrated in the ESME research program are subject to Quality Control procedures.

Patient-related data are entered via an electronic case report form (eCRF) in each centre by trained clinical research assistants (CRAs) between December 2014 and October 2016. Medical support was provided for assisting CRAs, data are quality-controlled and recoded before annual database lock. All procedures are handled according to the Data Management Plan. Importantly, all data are exclusively obtained retrospectively; no attempts are made to recover unavailable data from patients' medical records by contacting healthcare providers or patients.

The clinical data management system (CDMS) used was Statistical Analysis System (SAS) software. For both, ESME MBC database and the eCRF tool administration used an Oracle solution and certified personal data hosting system guarantee data security.

On-site quality review

On-site quality review (QR) of the patient screening process was carried out. This consisted of checking eligibility criteria for samples of selected and non-selected cases in each FCCC. For selected cases, key variables were crosschecked versus the source data. For all Quality Controls (QC) procedures, accepted error limits were 10% for non-selected patients and 5% for selected patients. Subsequently, a central audit was performed by the Unicancer Quality Assurance Department, and an audit on data entry and generation of the screening list was also conducted at the local level.

DATA QUALITY ASSURANCE

The ESME MBC database was authorised (authorization no. 1704113) by the French data protection authority (*Commission Nationale de l'Informatique et des Libertés*) and managed by R&D Unicancer in accordance with guidelines for Good Pharmacoepidemiology Practices and Good Epidemiology Practices [16, 17]. Approval was obtained from an independent ethics committee, which waived the requirement for informed consent.

Governance structure

Three boards monitor the ESME Research program : Scientific Committee, deontology Committee and International Advisory Board. The main role of the Scientific Committee is to: i) ensure that the applicable scientific rules are followed, ii) evaluate any ancillary projects in compliance with defined criteria and scientific pertinence, and iii) monitor the all validated ancillary projects. The deontology Committee monitors any potential conflicts of interest related to experts involved in the program, gives recommendations that may improve the prevention of conflicts, provides opinions on individual or particular situations, and potential collaborations with private partners. The ESME International Advisory Board has a consultative role with regard to coherence of the scientific program and reviews key international communications, formulates recommendations for publication rules or methodology, and reinforces international academic cooperation.

DATA ANALYSIS PRINCIPLES

Academic research teams or private organisations could propose ancillary projects.

For each accepted ancillary project, statistical analyses are conducted according to a detailed statistical analysis plan (SAP) that must be reviewed by the Scientific committee. This article does not aim at providing a comprehensive and exhaustive review on appropriate statistical methods to reduce bias related to analysis based on RWD.

A first ancillary project was published aiming to describe the outcomes (OS and PFS) following first-line paclitaxel treatment with or without bevacizumab [18]. Other analyses of sensibility are in progress to better address the bias potentially found in real-life settings. Two other have been accepted regarding the description of OS over the time and the first-line therapy (Endocrine therapy or chemotherapy) in hormone receptor-positive HER2 negative

cancer subgroup respectively [19, 20]. Other ones have been accepted for communications (abstracts/posters) to major congress in 2017.

RESULTS AND CURRENT STATUS

Of 45,329 patients screeneed,16,711 were selected in the ESME MBC Cohort (see **Figure 2**). Regarding the sensitivity and specificity of the three main screening sources used, sensitivity was highest for administrative records (78% versus 43% for pharmacy records and 28% for BC-specific local registries). On the other hand, specificity was highest for BC-specific local registries (87% versus 67% for pharmacy records and 49% for administrative records).

As shown in **Figure 2**, the main reasons for non-selection of screened patients were: presence of non-metastatic breast cancer or other metastatic cancers (n=14,104 patients), initial MBC treatment received before January 1st, 2008 (n=7486), and first metastasis not initially treated in an FCCC (n=4239). Nine additional were excluded prior to the final database lock due to inconsistencies in the date. 16,702 patients were analysed. Median follow-up duration was 48.55 months for the whole cohort [95% CIs: 47.7–49.38]; see **Figure 3** for median follow-up according to patient selection year.

Table 2 summarises the main demographic and disease characteristics at the time of initial metastatic diagnosis. Patients were nearly all women (99.1%) with a median age of 61 years. Over half (56.2%) had at least visceral metastases present, with 30.2% having at least bone and non-visceral metastases, and 13.6% with skin only, or node only, or at least skin plus node. 20.7% of patients had a HR-negative MBC, 73.7% had a HER2-negative MBC, and 13.9% were classified as triple-negative BC (i.e. HER2 and HR status both negative).

DISCUSSION

Retrospective analysis using real world data is likely to become increasingly important to ensure that medications are accepted by policymakers and adopted by patient practitioners. The ESME Research program is a large-scale initiative to provide access to real-world data in oncology. The program's first project, the ESME MBC cohort, centralises data from 16 711 patients.

The ESME research program provides a unique opportunity to study a diverse range of topics related to MBC care and management in real-life settings. Indeed, there are many potential applications, including study of the factors influencing patient care (e.g. cancer and patient characteristics), description of therapeutic strategies (treatment lines and sequences of therapies, etc.), measurement of clinical events (disease progression, death, persistence of treatment effect). Characterisation of patients enrolled in clinical trials is also possible, as is the reconstruction of "virtual trials" using appropriate statistical methodologies. Potentially, these data could be used for health economics evaluation of management strategies for patients (e.g. rehospitalisation and related ambulatory care), as well as reconstruction of health care trajectories through data modelling.

The ESME research program incorporates alternative approaches to create cohorts that use different types of RWD (clinical data, therapeutic treatment data, long term outcomes, health economics data) in the FCCCs *vs* existing registries in France, Europe and the US (e.g. SEER). It involves rigorous procedures for patient screening and data collection, ensuring

both validity and reliability of data. It uses a fully retrospective approach, with no influence on treatment practice or interaction with oncologists. Unlike prospective interventional or observational research studies, data are not influenced by study design and reflect the real-life management of patients treated. While data recorded for the cohort are defined by experts in the field, the vast majority of data are collected by trained clinical research technicians, minimising any potential risk of data misinterpretation. As discussed above, the ESME MBC Cohort offers a unique opportunity to study a wide range of research questions in a large sample. With respect to evaluation of treatment strategies, the database enables estimation of survival criteria such overall survival (OS) and the surrogates endpoints (PFS …). This is particularly important for showing OS improvement in diseases with a long median post-progression survival time, such as MBC [21-23].

The ESME MBC cohort also has several limitations. For example, the database relies on the collection and restructuring of existing data only, i.e. there is no creation of new data. Furthermore, apart from events reported in the patient medical record impacting therapeutic management, adverse effects are not routinely captured. Conceivably, further in-depth analysis of the data could highlight trends such as treatment interruption or discontinuation due to toxicity, which is important from a risk-management perspective. The main potential sources of bias include selection bias, and information bias due to differences in patient monitoring and non-standardised data collection. Selection bias has been taken into account by using rigorous selection procedures across all 18 FCCCs, and the data management plan and quality control program described above have been designed to limit information bias. Nevertheless, due to the retrospective data collection and the fact that it is based on real-life follow-up, clinical and biological events are not evaluated at predefined time points (unlike in RCTs). For example, objective response, historical endpoint in RCTs, could not be assessed retrospectively without a central review of existing imaging as not systematically documented in routine practice. The information collected therefore depends on the frequency of follow-up visits and clinical and radiological exams prescribed by the patient's doctor. As clinical signs are the only means by which disease metastasis can be identified, the number of disease progressions may be underestimated. With respect to the clinical event of death, all deaths are reported in the patient medical records.

With respect to evaluation of treatment strategies, analysis of real-life data poses unique challenges, such as accounting for confounding factors between patient groups, although various statistical approaches can be used to address this, as discussed above [24].

Concerning overall generalisability and applicability (external validity), it should be noted that the cohort centralises data from patients treated in specialised cancer centres only. FCCCs may utilise different clinical practices compared to public hospitals and private institutions, and thus patients from FCCCs may not be truly representative of all French breast cancer patients. Potentially, data extrapolation from all French health care organisations could be developed with the Exhaustive National Health Reimbursement System (SNIIRAM; *Système national d'information inter-régimes de l'Assurance maladie*).

The ESME MBC Cohort aims to collect data for up to 25,000-30,000 patients by 2019. As mentioned, future aims might include linking our database with those from other institutions, such as the SNDS database for data on exhaustive healthcare reimbursement, and the INSEE database to provide vital status updates for patients lost to follow-up. Conceivably, the ESME research program have also expanded to include two areas of oncology (ovarian cancer and

advanced/metastatic lung cancer) It is hoped that real world data from the ESME cohorts will help to provide medical recommendations and ultimately improve patient care.

ACKNOWLEDGEMENTS

We thank the 18 French Comprehensive Cancer Centres for providing the data and each ESME contact for coordinating the project at the local level.

We thank the ESME Scientific Committee members for their ongoing support. We also acknowledge Dr. Sarah Hopwood (Scinopsis, France) for editorial assistance with this publication, paid by R&D UNICANCER.

The 18 French Comprehensive Cancer Centres (FCCC) are: Ensemble hospitalier de l'Institut Curie, Paris/Saint-Cloud, Gustave Roussy, Villejuif, Institut de Cancérologie de l'Ouest, Nantes/Angers, Centre François Baclesse, Caen, Institut du Cancer de Montpellier Montpellier, Centre Léon Bérard, Lyon, Centre Georges-François Leclerc, Dijon, Centre Henri Becquerel, Rouen; Institut Claudius Régaud, Toulouse; Centre Antoine Lacassagne, Nice; Institut de Cancérologie de Lorraine, Nancy; Centre Eugène Marquis, Rennes; Institut Paoli-Calmettes, Marseille; Centre Jean Perrin, Clermont-Ferrand; Institut Bergonié, Bordeaux; Centre Paul Strauss, Strasbourg; Institut Jean Godinot, Reims; Centre Oscar Lambret, Lille.

ESME central coordinating staff: Head of Research and Development: Christian Cailliot. Program director: Mathieu Robain. Data Managers: Irwin Piot and Olivier Payen. Project Managers: Coralie Courtinard, Tahar Guesmia and Gaëtane Simon. Project assistant: Esméralda Pereira. Software designer: Alexandre Vanni. ESME local coordinating staff: Patrick Arveux, Thomas Bachelot, Jean-Pierre Bleuse, Delphine Berchery, Mathias Breton, Stéphanie Clisant, Emmanuel Chamorey, Valérie Dejean, Véronique Diéras, Anne-Valérie Guizard, Anne Jaffré, Lilian Laborde, Agnès Loeb, Muriel Mons, Damien Parent, Geneviève Perrocheau, Marie-Ange Mouret-Reynier, Carine Laurent, Michel Velten.

Funding: This work was supported by R&D UNICANCER and funded by an industrial consortium (Roche, Pierre Fabre, Pfizer, and Astrazeneca). The funders did not have any input into maintenance of the database (project design, data collection, management) or analysis of data. They were not involved in the writing of the manuscript or decision to submit for publication.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest: D Perol has received personal fees (honoraria and travel/accommodation expenses) from Laboratoire Roche, outside the submitted work. B Asselain has received personal fees for board membership and for consultancy from Roche Pharma, outside the submitted work. S Gourgou has received personal fees for board membership from Celgene and for consultancy from Roche, outside the submitted work. All other authors declare that they have no conflict of interest.

Ethical approval: Approval for the ESME MBC database was obtained from an independent ethics committee, which waived the requirement for informed consent.

REFERENCES

1. Chavez-MacGregor M, Giordano SH. Randomized Clinical Trials and Observational Studies: Is There a Battle? J Clin Oncol. 2016;34(8):772-3. doi:10.1200/JCO.2015.64.7487

JCO.2015.64.7487 [pii].

2. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part I. Value Health. 2009;12(8):1044-52. doi:10.1111/j.1524-4733.2009.00600.x

S1098-3015(10)60308-7 [pii].

3. Hahn OM, Schilsky RL. Randomized controlled trials and comparative effectiveness research. J Clin Oncol. 2012;30(34):4194-201. doi:10.1200/JCO.2012.42.2352

JCO.2012.42.2352 [pii].

4. Meyerhardt JA, Li L, Sanoff HK, Carpenter Wt, Schrag D. Effectiveness of bevacizumab with first-line combination chemotherapy for Medicare patients with stage IV colorectal cancer. J Clin Oncol. 2012;30(6):608-15. doi:10.1200/JCO.2011.38.9650

JCO.2011.38.9650 [pii].

5. Prasad V, Kim C, Burotto M, Vandross A. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-level meta-analyses. JAMA Intern Med. 2015;175(8):1389-98. doi:10.1001/jamainternmed.2015.2829

2323416 [pii].

6. Visvanathan K, Levit LA, Raghavan D, Hudis CA, Wong S, Dueck A et al. Untapped Potential of Observational Research to Inform Clinical Decision Making: American Society of Clinical Oncology Research Statement. J Clin Oncol. 2017;35(16):1845-54. doi:10.1200/JCO.2017.72.6414.

7. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. Value Health. 2012;15(2):217-30. doi:10.1016/j.jval.2011.12.010

S1098-3015(12)00007-1 [pii].

8. Morton SC, Costlow MR, Graff JS, Dubois RW. Standards and guidelines for observational studies: quality is in the eye of the beholder. J Clin Epidemiol. 2016;71:3-10. doi:10.1016/j.jclinepi.2015.10.014 S0895-4356(15)00501-6 [pii].

9. Rosenbaum PR, Rubin DB. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. The American Statistician. 1985;39(1):33-8. doi:10.1080/00031305.1985.10479383.

10. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac

management on AMI survival using propensity score and instrumental variable methods. JAMA. 2007;297(3):278-85. doi:297/3/278 [pii] 10.1001/jama.297.3.278.

11. D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998;17(19):2265-81. doi:10.1002/(SICI)1097-0258(19981015)17:19<2265::AID-SIM918>3.0.CO;2-B [pii].

12. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. Annu Rev Public Health. 1998;19:17-34. doi:10.1146/annurev.publhealth.19.1.17.

13. Institut National du Cancer (INCa). Information sur le cancer. http://www.e-cancer.fr/. Accessed 04 Jul 2017.

14. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005;365(9472):1687-717. doi:S0140-6736(05)66544-0 [pii]

10.1016/S0140-6736(05)66544-0.

15. Binder-Foucard F, Belot A, Delafosse P, Remontet L, Woronoff AS, Bossard N. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012. Partie 1 – Tumeurs solides. Saint-Maurice (Fra): Institut de veille sanitaire, 2013.

16. Association des Epidémiologistes de Langue Française (ADELF). Déontologie et bonnes pratiques en épidémiologie, recommendations révisées, 2007.

17. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiol Drug Saf. 2008;17(2):200-8. doi:10.1002/pds.1471.

18. Delaloge S, Perol D, Courtinard C, Brain E, Asselain B, Bachelot T et al. Paclitaxel plus bevacizumab or paclitaxel as first-line treatment for HER2-negative metastatic breast cancer in a multicenter national observational study. Ann Oncol. 2016;27(9):1725-32. doi:10.1093/annonc/mdw260 mdw260 [pii].

19. Gobbini E, Ezzalfani M, Dieras V, Bachelot T, Brain E, et Al.Time trends of overall survival among metastatic breast cancer patients in the real life ESME Cohort . Eur J Cancer. (In press)

20. Lardy-Cléaud A, Pistilli B, Frank S, Cottu P, Delaloge S, et Al. Endocrine therapy or chemotherapy as first-line therapy in hormone receptor-positive HER2 negative metastatic breast cancer patients. Eur J Cancer. (In press)

21. Largillier R, Ferrero JM, Doyen J, Barriere J, Namer M, Mari V et al. Prognostic factors in 1,038 women with metastatic breast cancer. Ann Oncol. 2008;19(12):2012-9. doi:10.1093/annonc/mdn424.

22. Asselain B, De Rycke Y, Savignon A, Mould RF. Parametric modelling to predict survival time to first recurrence for breast cancer. Phys Med Biol. 2003;48(12):L31-3.

23. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. J Natl Cancer Inst. 2009;101(23):1642-9. doi:10.1093/jnci/djp369djp369 [pii].

24. Perol D, Robain M, Delaloge S, Cailliot C. Routinely collected data may usefully supplement randomised controlled data on treatment effects for mortality. BMJ. 2016;355:i6745. doi:10.1136/bmj.i6745.

ESME MBC Cohort

TABLES

Table 1 Main data recorded

Patient-related	Hospitalisation-related	Pharmacy record-related
Patient data:	Age	Height
Demographics	Gender	Body Mass Index
Other cancer and family history	Main diagnosis code (ICD code)	Drug (International Non-proprietary Name)
Menopausal status	Linked diagnosis code (ICD code)	Protocol
Initial tumor:	Related significant diagnoses (ICD code)	Cycle ID
Diagnosis	Diagnosis-related group code (reimbursement coding)	Pharmaceutical form and dosage
Relapses	Hospital stay-related group ID (reimbursement coding)	Administration date
Histological results	Entry date	Line treatment number
Medical care	Discharge date	Administered dose
Metastatic disease:	Destination code at discharge	Cumulated dose
Diagnosis	Medical procedures performed	Inclusion in a clinical trial (Yes/No)
Progression	Radiation therapy	
Histological results		
Therapies		
Invasive procedures related to metastasis		
Last contact:		
Vital status		
Date of last contact/death		

ICD International Classification of Diseases

ESME MBC Cohort

Table 2 Characteristics of the at diagnosis of metastatic disease

Characteristic		ESME MBC Population (N=16 702)
Age (years)	Mean (SD)	60.6 (13.8)
	Median [Q1-Q3 range]	61.0 [51.0-71.0]
Sex	Male	149 (0.9%)
	Female	16553 (99.1%)
Histological grade* at primary tumour diagnosis	1	1277 (10.1%)
	2	6438 (50.7%)
	3	4733 (37.3%)
	Not available	240 (1.9%)
	Missing data	2008
Metastatic status	*De novo* MBC	4763 (28.5%)
	Relapsed MBC	11939 (71.5%)
Year of selection (first metastatic treatment)	2008	2651 (15.9%)
	2009	2675 (16.0%)
	2010	2598 (15.6%)
	2011	2515 (15.1%)
	2012	2371 (14.2%)
	2013	2216 (13.3%)
	2014	1676 (10.0%)
Type of metastases	Visceral	9383 (56.2%)
	Bones and not visceral	5047 (30.2%)
	Nodes only	880 (5.3%)
	Skin only	916 (5.5%)
	Skin + nodes	476 (2.8%)
Global HR status ***	Positive	12748 (76.3%)
	Negative	3451 (20.7%)
	Not determined	503 (3.0%)
Global HER2 status***	Positive	2863 (17.1%)
	Negative	12306 (73.7%)
	Not determined	1533 (9.2%)
Triple negative status	Yes	2321 (13.9%)

Data are n (%) unless indicated otherwise

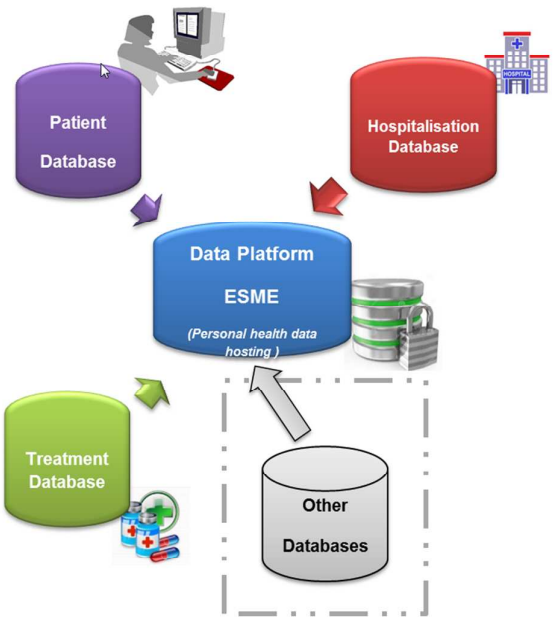
*Histological grade at primary tumour diagnosis: The histological grade at primary tumour diagnosis is defined as the worst histological grade recorded within 1 month (30 days) of the initial diagnosis (primary tumour).

**De novo MBC: Metastatic breast cancer is considered de novo if the diagnosis of metastatic disease occurs within 6 months (180 days) of the initial diagnosis (primary tumour).

***HR and HER2 status:The estrogen receptor (ER) or progesterone receptor status (PR) is considered positive if the pathology report indicates a "positive" result or if the result is not available. The result is considered positive if ≥10% of cells in the sample are positive for ER or PR. The HER2 scoring system and criteria are described in Table 4. If two or more histologic reports are available at the same date the positive status is dominant.The global HR/HER2 status are the status at metastatic diagnosis based on histological results forms related to the primary tumour (if available) or metastatic sites.

FIGURES

Fig. 1 The ESME Data platform



Hospitalisation Database

- Hospitalisations: dates, diagnoses, GHS code
- Medical procedures (inc. Radiotherapy): dates, code

Treatment Database

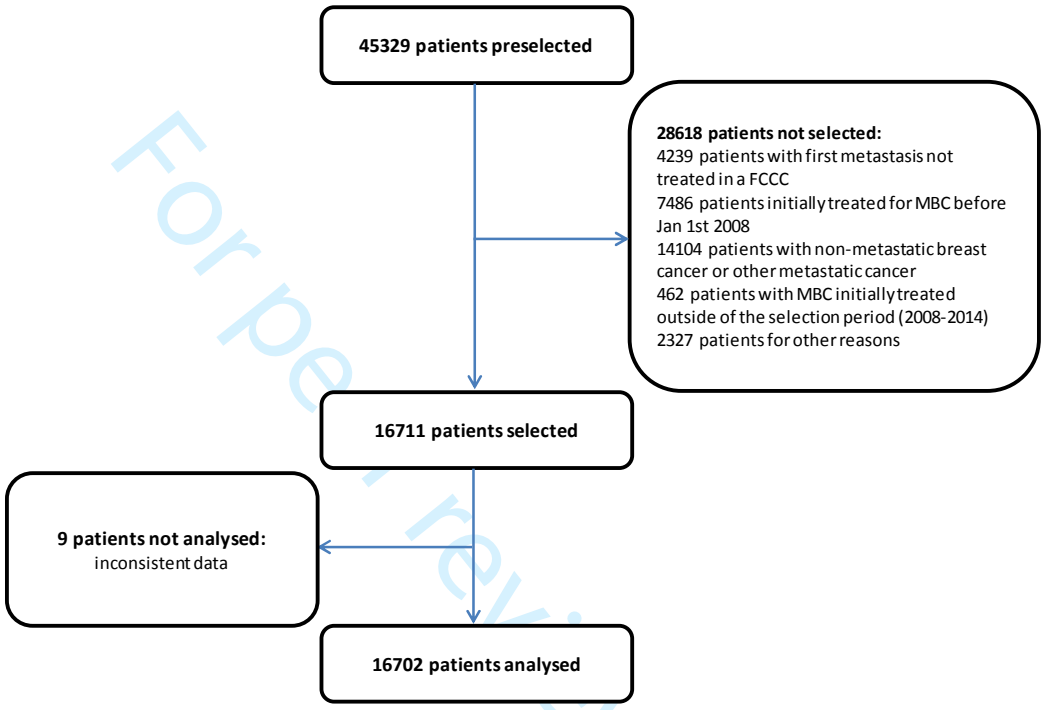
- Pharmacy records: dates, antineoplastic drugs, therapeutic protocol and other concomitant drugs

Patient Database

- Collection based on Electronic medical records
- Patient data : demographics, cancer management, clinical events (progression, relapse), pathological report, metastatic disease, anti-cancer treatment (chemotherapy, targeted therapy, endocrine therapy, immunotherapy), and other therapeutic care (radiotherapy, surgery) or supportive care

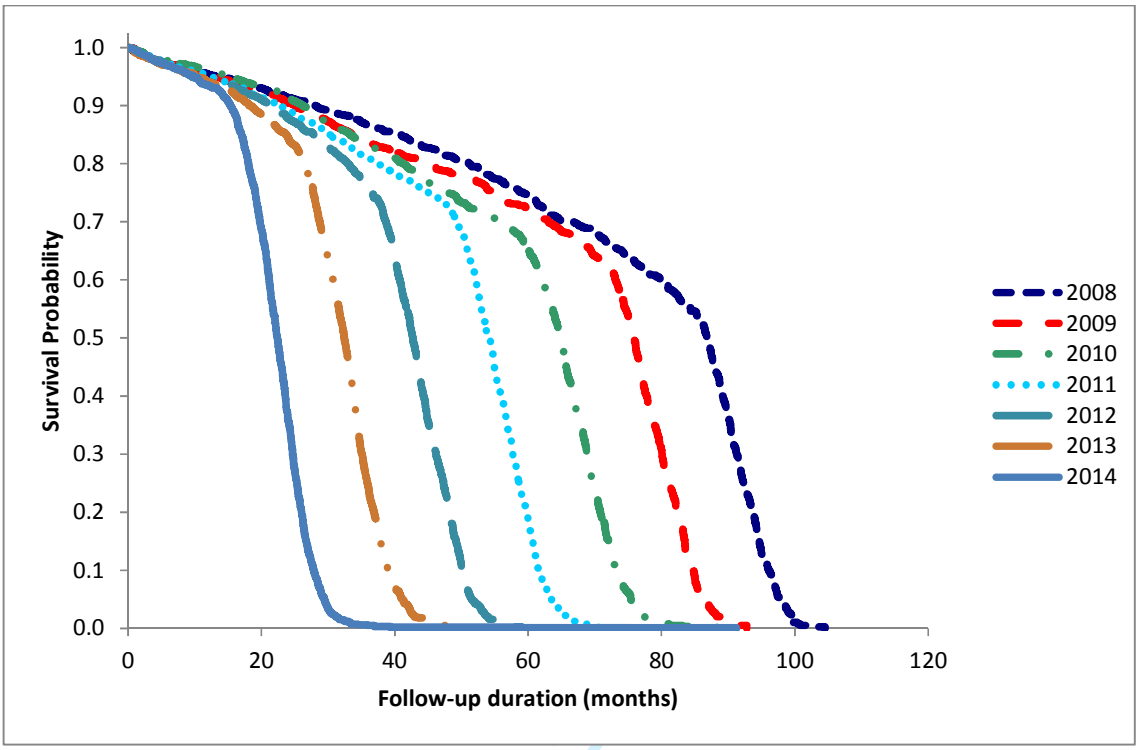
ESME MBC Cohort

Fig. 2 ESME MBC Cohort: Flow chart



ESME MBC Cohort

Fig. 3 Follow-up duration according to year of first-line metastatic treatment



BMJ Open

Cohort Profile: Epidemiological Strategy and Medical Economics (ESME) Research Program, the French Metastatic Breast Cancer (MBC) Cohort

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023568
Article Type:	Cohort profile
Date Submitted by the Author:	12-Apr-2018
Complete List of Authors:	Pérol, David; Léon Bérard Cancer Centre, Unit of Biostatistics and Therapeutic Evaluation Robain, Mathieu; R&D UNICANCER Arveux, F; Registre des cancers du sein de Côte d'Or, centre Georges-François-Leclerc Mathoulin-Pélissier , S; INSERM, Chamorey, Emmanuel; Centre Antoine-Lacassagne Asselain, Bernard Berchery, Delphine; Institut Claudius Regaud Gourgou, Sophie; Institut regional du Cancer de Montpellier Breton, Mathias; Centre Eugene Marquis Delaine-Clisant , Stéphanie; Centre Oscar Lambret, Clinical Research Unit Mons, Muriel; Gustave Roussy Diéras, Véronique; Ensemble Hospitalier de l'Institut Curie Carton, Matthieu; Institut Curie, Biométrie; Guizard, Anne-Valérie ; Centre Francois Baclesse Centre de Lutte Contre le Cancer Laborde, Lilian; Institut Paoli-Calmettes Laurent, Carine; Institut de Cancerologie de Lorraine Loeb, Agnès; Centre Henri Becquerel Mouret-Reynier, Marie-Ange ; Centre Jean Perrin Parent, Damien; Institut Jean Godinot Perrocheau, Geneviève; Institut de Cancérologie de l'Ouest Campion, Loïc; Institut de Cancérologie de l'Ouest Velten, Michel; Centre Paul Strauss Cailliot, Christian; R&D UNICANCER Ezzalfani, Monia; Centre Leon Berard Simon, Gaëtane; R&D UNICANCER,
Keywords:	Real-world data, Data Platform, Epidemiology < ONCOLOGY, Patient medical record, Quality control, Metastatic breast cancer

TITLE PAGE

Epidemiological Strategy and Medical Economics (ESME) Research Program, the French Metastatic Breast Cancer (MBC) Cohort: Design and current status

List of authors:

David Pérol (Centre Léon Bérard, Lyon), Mathieu Robain (R&D UNICANCER, Paris), Patrick Arveux (Centre Georges-François Leclerc, Dijon), Simone Mathoulin-Pelissier (Institut Bergonié, Bordeaux), Emmanuel Chamorey (Centre Antoine Lacassagne, Nice), Bernard Asselain (University Paris Sud), Delphine Berchery (Institut Claudius Régaud, Toulouse), Sophie Gourgou (Institut régional du Cancer Montpellier, Montpellier), Mathias Breton (Centre Eugène Marquis, Rennes), Stéphanie Clisant (Centre Oscar Lambret, Lille), Muriel Mons (Gustave Roussy, Villejuif), Véronique Diéras (Ensemble Hospitalier de l'Institut Curie, Paris & Saint-Cloud), Matthieu Carton (Ensemble Hospitalier de l'Institut Curie, Paris & Saint-Cloud), Anne-Valérie Guizard (Centre François Baclesse, Caen), Lilian Laborde (Institut Paoli-Calmettes, Marseille), Carine Laurent (Institut de Cancérologie de Lorraine, Nancy), Agnès Loeb (Centre Henri Becquerel, Rouen), Marie-Ange Mouret-Reynier (Centre Jean Perrin, Clermont-Ferrand), Damien Parent (Institut Jean Godinot, Reims), Geneviève Perrocheau (Institut de Cancérologie de l'Ouest, Nantes & Angers), Loïc Campion (Institut de Cancérologie de l'Ouest, Nantes & Angers), Michel Velten (Centre Paul Strauss, Strasbourg), Christian Cailliot (R&D UNICANCER, Paris), Monia Ezzalfani (Centre Léon Bérard, Lyon), Gaëtane Simon (R&D UNICANCER, Paris)

Corresponding author:

David Pérol (Centre Léon Bérard, Lyon)

E-mail: david.perol@lyon.unicancer.fr

Tel: +33 4.69.85.61.00

ABSTRACT

The global aim of the ESME research program is the centralisation of real-life data on oncology care for epidemiological research purposes. The first covered medical area was the metastatic breast cancer (MBC) using a retrospective data collection called ESME MBC Cohort. The main objective is to describe clinical features, treatment patterns and outcomes over the years. This population-based prospective cohort aims to select all consecutive patients treated for their MBC managed in the 18 French Comprehensive Cancer Centres (FCCCs) from 2008 to 2019. We aim to select 30,000 cases in this cohort. These real-world data will help standardise the management of MBC and improve patient care. Diagnostic, therapeutic and follow up data (demographics, primary tumor, metastatic disease, treatment patterns and vital status) were collected through the course of the disease. To date, over 16,700 patients initially treated for metastatic disease from 2008–2014 with possible follow up period until 2016. A dozen of ancillary research projects have been conducted and few of them are already accepted for publication or published.

Data collection is updated annually. Future aim is to link the data of the cohort to the French national Health Data System (SNDS) for centralizing data on healthcare reimbursement (drugs, medical procedures), inpatient/outpatient stays and visits in primary/secondary care settings. Finally, the ESME research program is expanding to two other areas of oncology: ovarian cancer (OC) and advanced/metastatic lung cancer (AMLC).

KEYWORDS

Real-world data; real-life data; patient medical record; quality control; oncology; metastatic breast cancer

BACKGROUND

Real-world evidence (RWE) studies and observational studies using real world data (RWD) play a growing role in conducting comparative effectiveness research on pharmaceutical products and other healthcare interventions. They aim to bridge the gap between the highly controlled environment of randomised clinical trials (RCTs) and real-life clinical practice [1]. In particular, health authorities are interested in gathering real-world data for long-term benefit-risk assessment and, increasingly, health economic evaluations for reimbursement decisions [2]. With their high internal validity, RCTs are considered the gold standard of evidence for establishing treatment efficacy, although the generalisability of findings to clinical practice may be limited [3]. In fact, cancer survival endpoints in the real-life setting may differ from that measured in traditional RCTs [4].. Furthermore, other limitations such as short follow-up and small sample size have created uncertainty in estimating survival criteria in RCTs, and thus surrogate endpoints are generally used such as progression free survival (PFS) or time to progression (TTP) [5]. On the other hand, large population-based cohorts with longer follow-up periods can be particularly appropriate to assess long-term clinical outcomes , such as overall survival (OS) outside of the RCT setting [6] and to detect changes in medical practice.

RWE and observational studies are non-experimental research where cancer management (treatments and disease evolution assessment) is left to the choice of care providers and patients [7]. Over the past few years, there was a requirement for high-quality data from large cohorts, which is now driving improvements in research methods and practices [8]. The goal of these studies is to generate complementary information to RCTs based on larger samples and provide answers regarding particular populations in real-life clinical practice. These studies are more prone to biases such as baseline differences between patients (selection bias) or bias due to confounding by indication for example. To minimise sources of such biases, statistical approaches including adjusted analyses, propensity score methods, or instrumental variables may be also employed [9-12].

Metastatic breast cancer (MBC) is one of the leading causes of cancer-related mortality among women in Western countries [13]. A relatively high proportion (approximately 30%) of breast cancer patients develop metastatic disease [14], and while significant treatment advances have been made, the overall prognosis is poor with a five-year survival rate of 25% [15]. The national academic network of cancer centres in France (French Comprehensive Cancer Centres; FCCCs), which together handle over one-third of all breast cancer cases nationally, decided to launch in 2014 a program dedicated focussing on RWD in oncology databases in MBC through Epidemiological Strategy and Medical Economics (ESME) research program.

In this paper, we describe the methodological principles that underpin the ESME research program, and illustrated by the first project, the ESME metastatic breast cancer (MBC) cohort. Design, current status and a preliminary description of population are presented.

MAIN OBJECTIVES

The global aim of the ESME research program is the construction of a comprehensive database on oncology care for epidemiological research purposes in order to improve knowledge on medical practice in real-life setting, on public health and healthcare use, and to provide information to health authorities and other associated bodies. For each cancer covered by the program, data platform will then be created to address this epidemiological research.

Regarding the ESME MBC project (ClinicalTrials.gov ; NCT03275311), the primary objective is to describe the medical care of patients treated for MBC according to their disease characteristics. In this context, medical care encompasses all surgical procedures, radiotherapy and drug treatments, focussing particularly on chemotherapy, targeted and endocrine therapies. A secondary objective of the ESME MBC project is to describe the evolution of the metastatic disease and the outcomes.

METHODS

The ESME MBC cohort is a population-based registry in 18 FCCCs (http://www.unicancer.fr/en/rd-unicancer/esme), which collected data on all consecutive patients treated for MBC from 2008. Annual data collection phases are planned to add new diagnosed cases and update patients' follow up data.

STUDY POPULATION

Eligibility criteria

Patients were eligible to the ESME MBC cohort according to the following criteria: male or female patients aged ≥18 years with MBC whose first metastasis was treated (either completely or partially) in an FCCC between January 1st, 2008 and December 31st, 2014. MBC treatment could include radiotherapy, chemotherapy, targeted therapy, immunotherapy, or endocrine therapy. Patient data follow up was considered on patient basis according to last contact information available in the patient medical records (PMR) at the day of the data collection. This data collection has been performed from October 2014 to October 2016 by trained research assistants.

Patient screening process

Patient screening process involved two phases: an automated case screening and then the validation of selection for each screened case. The automated case screening was based on information retrieved from multiple data sources available within each FCCC: administrative records [French National Computerised Medical Information System] (via MBC-specific International Classification of Diseases (ICD) codes associated with inpatient stays), pharmacy records, PMRs (including multidisciplinary team meeting records and search using relevant keywords), MBC-specific registries. The objective of the automated screening phase was to identify almost all cases meeting the selection criteria and generate the patient screening list. The ICD codes used were: C50 (Malignant neoplasm of breast), C77.- [except

C77.3] (Secondary and unspecified malignant neoplasm of lymph nodes), C78.- (Secondary malignant neoplasm of respiratory and digestive organs) and C79.- (Secondary malignant neoplasm of other and unspecified sites).

Once the patient screening list was finalised in each centre, each patient was given an anonymous ESME number. As the automated screening method did not allow the identification of the first MBC-specific treatment, eligibility criteria were cross-checked for all screened cases using data from the patient medical records. 45,329 cases were screened and 16,711 were selected in the cohort.

Baseline and follow up data

The ESME MBC Cohort is composed of 3 types of data (see **Figure 1**):

- *Patient-related data* are obtained from systematic review of patient medical records (non-structured data) and provide information on patient demographics, cancer family history, characteristics of primary tumour, relapses, metastatic recurrences pathological reports (tumor size, grade, histological type), hormone-receptor status and HER-2 status, therapeutic care (focussing on cancer-related treatment) and settings, reasons for treatment termination, and clinical events.
- *Hospitalisation records* are integrated data from a structured and automated database related to inpatient stays, and primarily used to bill the French National Health Insurance Fund (*Assurance Maladie*). It provides information on patient entry and discharge information, (date and destination code at discharge), ICD codes associated to each stay, diagnostic, medical and therapeutic procedures including radiotherapy and surgery.
- *Pharmacy-dispensed treatment records* includes all data related to anti-cancer treatments obtained from each centre's pharmacy database: drugs (International Non-proprietary Name), administration date, protocol name, patient's height and BMI, cycle ID, UCD code (pharmaceutical form related to the drug dosage), line of treatment, administration in a clinical trial (yes/no). It does not include information on products that are prescribed by the primary care practice.

The detailed raw data collected above and derived data are listed in **Table 1**.

Data management

Any data integrated in the ESME research program are subject to Quality Control procedures.

Patient-related data are entered via an electronic case report form (eCRF) in each centre by trained clinical research assistants (CRAs) between December 2014 and October 2016. Medical support was provided for assisting CRAs, data are quality-controlled and recoded before annual database lock. All procedures are handled according to the Data Management Plan. Importantly, all data are exclusively obtained retrospectively; no attempts are made to recover unavailable data from patients' medical records by contacting healthcare providers or patients.

The clinical data management system (CDMS) used was Statistical Analysis System (SAS) software. For both, ESME MBC database and the eCRF tool administration used an Oracle solution and certified personal data hosting system guarantee data security.

On-site quality review

On-site quality review (QR) of the patient screening process was carried out. This consisted of checking eligibility criteria for samples of selected and non-selected cases in each FCCC. For selected cases, key variables were crosschecked versus the source data. For all Quality Controls (QC) procedures, accepted error limits were 10% for non-selected patients and 5% for selected patients. Subsequently, a central audit was performed by the Unicancer Quality Assurance Department, and an audit on data entry and generation of the screening list was also conducted at the local level.

DATA QUALITY ASSURANCE

The ESME MBC database was authorised (authorization no. 1704113) by the French data protection authority (*Commission Nationale de l'Informatique et des Libertés*) and managed by R&D Unicancer in accordance with guidelines for Good Pharmacoepidemiology Practices and Good Epidemiology Practices [16, 17]. Approval was obtained from an independent ethics committee, which waived the requirement for informed consent.

Governance structure

Three boards monitor the ESME Research program : Scientific Committee, deontology Committee and International Advisory Board. The main role of the Scientific Committee is to: i) ensure that the applicable scientific rules are followed, ii) evaluate any ancillary projects in compliance with defined criteria and scientific pertinence, and iii) monitor the all validated ancillary projects. The deontology Committee monitors any potential conflicts of interest related to experts involved in the program, gives recommendations that may improve the prevention of conflicts, provides opinions on individual or particular situations, and potential collaborations with private partners. The ESME International Advisory Board has a consultative role with regard to coherence of the scientific program and reviews key international communications, formulates recommendations for publication rules or methodology, and reinforces international academic cooperation.

DATA ANALYSIS PRINCIPLES

Academic research teams or private organisations could propose ancillary projects.

For each accepted ancillary project, statistical analyses are conducted according to a detailed statistical analysis plan (SAP) that must be reviewed by the Scientific committee. This article does not aim at providing a comprehensive and exhaustive review on appropriate statistical methods to reduce bias related to analysis based on RWD.

A first ancillary project was published aiming to describe the outcomes (OS and PFS) following first-line paclitaxel treatment with or without bevacizumab [18]. Other analyses of sensibility are in progress to better address the bias potentially found in real-life settings. Two other have been accepted regarding the description of OS over the time and the first-line therapy (Endocrine therapy or chemotherapy) in hormone receptor-positive HER2 negative

cancer subgroup respectively [19, 20]. Other ones have been accepted for communications (abstracts/posters) to major congress in 2017.

RESULTS AND CURRENT STATUS

Of 45,329 patients screeneed,16,711 were selected in the ESME MBC Cohort (see **Figure 2**). Regarding the sensitivity and specificity of the three main screening sources used, sensitivity was highest for administrative records (78% versus 43% for pharmacy records and 28% for BC-specific local registries). On the other hand, specificity was highest for BC-specific local registries (87% versus 67% for pharmacy records and 49% for administrative records).

As shown in **Figure 2**, the main reasons for non-selection of screened patients were: presence of non-metastatic breast cancer or other metastatic cancers (n=14,104 patients), initial MBC treatment received before January 1st, 2008 (n=7486), and first metastasis not initially treated in an FCCC (n=4239). Nine additional were excluded prior to the final database lock due to inconsistencies in the date. 16,702 patients were analysed. Median follow-up duration was 48.55 months for the whole cohort [95% CIs: 47.7–49.38]; see **Figure 3** for median follow-up according to patient selection year.

Table 2 summarises the main demographic and disease characteristics at the time of initial metastatic diagnosis. Patients were nearly all women (99.1%) with a median age of 61 years. Over half (56.2%) had at least visceral metastases present, with 30.2% having at least bone and non-visceral metastases, and 13.6% with skin only, or node only, or at least skin plus node. 20.7% of patients had a HR-negative MBC, 73.7% had a HER2-negative MBC, and 13.9% were classified as triple-negative BC (i.e. HER2 and HR status both negative).

DISCUSSION

Retrospective analysis using real world data is likely to become increasingly important to ensure that medications are accepted by policymakers and adopted by patient practitioners. The ESME Research program is a large-scale initiative to provide access to real-world data in oncology. The program's first project, the ESME MBC cohort, centralises data from 16 711 patients.

The ESME research program provides a unique opportunity to study a diverse range of topics related to MBC care and management in real-life settings. Indeed, there are many potential applications, including study of the factors influencing patient care (e.g. cancer and patient characteristics), description of therapeutic strategies (treatment lines and sequences of therapies, etc.), measurement of clinical events (disease progression, death, persistence of treatment effect). Characterisation of patients enrolled in clinical trials is also possible, as is the reconstruction of "virtual trials" using appropriate statistical methodologies. Potentially, these data could be used for health economics evaluation of management strategies for patients (e.g. rehospitalisation and related ambulatory care), as well as reconstruction of health care trajectories through data modelling.

The ESME research program incorporates alternative approaches to create cohorts that use different types of RWD (clinical data, therapeutic treatment data, long term outcomes, health economics data) in the FCCCs *vs* existing registries in France, Europe and the US (e.g. SEER). It involves rigorous procedures for patient screening and data collection, ensuring

both validity and reliability of data. It uses a fully retrospective approach, with no influence on treatment practice or interaction with oncologists. Unlike prospective interventional or observational research studies, data are not influenced by study design and reflect the real-life management of patients treated. While data recorded for the cohort are defined by experts in the field, the vast majority of data are collected by trained clinical research technicians, minimising any potential risk of data misinterpretation. As discussed above, the ESME MBC Cohort offers a unique opportunity to study a wide range of research questions in a large sample. With respect to evaluation of treatment strategies, the database enables estimation of survival criteria such overall survival (OS) and the surrogates endpoints (PFS …). This is particularly important for showing OS improvement in diseases with a long median post-progression survival time, such as MBC [21-23].

The ESME MBC cohort also has several limitations. For example, the database relies on the collection and restructuring of existing data only, i.e. there is no creation of new data. Furthermore, apart from events reported in the patient medical record impacting therapeutic management, adverse effects are not routinely captured. Conceivably, further in-depth analysis of the data could highlight trends such as treatment interruption or discontinuation due to toxicity, which is important from a risk-management perspective. The main potential sources of bias include selection bias, and information bias due to differences in patient monitoring and non-standardised data collection. Selection bias has been taken into account by using rigorous selection procedures across all 18 FCCCs, and the data management plan and quality control program described above have been designed to limit information bias. Nevertheless, due to the retrospective data collection and the fact that it is based on real-life follow-up, clinical and biological events are not evaluated at predefined time points (unlike in RCTs). For example, objective response, historical endpoint in RCTs, could not be assessed retrospectively without a central review of existing imaging as not systematically documented in routine practice. The information collected therefore depends on the frequency of follow-up visits and clinical and radiological exams prescribed by the patient's doctor. As clinical signs are the only means by which disease metastasis can be identified, the number of disease progressions may be underestimated. With respect to the clinical event of death, all deaths are reported in the patient medical records.

With respect to evaluation of treatment strategies, analysis of real-life data poses unique challenges, such as accounting for confounding factors between patient groups, although various statistical approaches can be used to address this, as discussed above [24].

Concerning overall generalisability and applicability (external validity), it should be noted that the cohort centralises data from patients treated in specialised cancer centres only. FCCCs may utilise different clinical practices compared to public hospitals and private institutions, and thus patients from FCCCs may not be truly representative of all French breast cancer patients. Potentially, data extrapolation from all French health care organisations could be developed with the Exhaustive National Health Reimbursement System (SNIIRAM; *Système national d'information inter-régimes de l'Assurance maladie*).

The ESME MBC Cohort aims to collect data for up to 25,000-30,000 patients by 2019. As mentioned, future aims might include linking our database with those from other institutions, such as the SNDS database for data on exhaustive healthcare reimbursement, and the INSEE database to provide vital status updates for patients lost to follow-up. Conceivably, the ESME research program have also expanded to include two areas of oncology (ovarian cancer and

advanced/metastatic lung cancer) It is hoped that real world data from the ESME cohorts will help to provide medical recommendations and ultimately improve patient care.

ACKNOWLEDGEMENTS

We thank the 18 French Comprehensive Cancer Centres for providing the data and each ESME contact for coordinating the project at the local level.

We thank the ESME Scientific Committee members for their ongoing support. We also acknowledge Dr. Sarah Hopwood (Scinopsis, France) for editorial assistance with this publication, paid by R&D UNICANCER.

The 18 French Comprehensive Cancer Centres (FCCC) are: Ensemble hospitalier de l'Institut Curie, Paris/Saint-Cloud, Gustave Roussy, Villejuif, Institut de Cancérologie de l'Ouest, Nantes/Angers, Centre François Baclesse, Caen, Institut du Cancer de Montpellier Montpellier, Centre Léon Bérard, Lyon, Centre Georges-François Leclerc, Dijon, Centre Henri Becquerel, Rouen; Institut Claudius Régaud, Toulouse; Centre Antoine Lacassagne, Nice; Institut de Cancérologie de Lorraine, Nancy; Centre Eugène Marquis, Rennes; Institut Paoli-Calmettes, Marseille; Centre Jean Perrin, Clermont-Ferrand; Institut Bergonié, Bordeaux; Centre Paul Strauss, Strasbourg; Institut Jean Godinot, Reims; Centre Oscar Lambret, Lille.

ESME central coordinating staff: Head of Research and Development: Christian Cailliot. Program director: Mathieu Robain. Data Managers: Irwin Piot and Olivier Payen. Project Managers: Coralie Courtinard, Tahar Guesmia and Gaëtane Simon. Project assistant: Esméralda Pereira. Software designer: Alexandre Vanni. ESME local coordinating staff: Patrick Arveux, Thomas Bachelot, Jean-Pierre Bleuse, Delphine Berchery, Mathias Breton, Stéphanie Clisant, Emmanuel Chamorey, Valérie Dejean, Véronique Diéras, Anne-Valérie Guizard, Anne Jaffré, Lilian Laborde, Agnès Loeb, Muriel Mons, Damien Parent, Geneviève Perrocheau, Marie-Ange Mouret-Reynier, Carine Laurent, Michel Velten.

Funding: This work was supported by R&D UNICANCER and funded by an industrial consortium (Roche, Pierre Fabre, Pfizer, and Astrazeneca). The funders did not have any input into maintenance of the database (project design, data collection, management) or analysis of data. They were not involved in the writing of the manuscript or decision to submit for publication.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest: D Perol has received personal fees (honoraria and travel/accommodation expenses) from Laboratoire Roche, outside the submitted work. B Asselain has received personal fees for board membership and for consultancy from Roche Pharma, outside the submitted work. S Gourgou has received personal fees for board membership from Celgene and for consultancy from Roche, outside the submitted work. All other authors declare that they have no conflict of interest.

Ethical approval: Approval for the ESME MBC database was obtained from an independent ethics committee, which waived the requirement for informed consent.

REFERENCES

1. Chavez-MacGregor M, Giordano SH. Randomized Clinical Trials and Observational Studies: Is There a Battle? J Clin Oncol. 2016;34(8):772-3. doi:10.1200/JCO.2015.64.7487

JCO.2015.64.7487 [pii].

2. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part I. Value Health. 2009;12(8):1044-52. doi:10.1111/j.1524-4733.2009.00600.x

S1098-3015(10)60308-7 [pii].

3. Hahn OM, Schilsky RL. Randomized controlled trials and comparative effectiveness research. J Clin Oncol. 2012;30(34):4194-201. doi:10.1200/JCO.2012.42.2352

JCO.2012.42.2352 [pii].

4. Meyerhardt JA, Li L, Sanoff HK, Carpenter Wt, Schrag D. Effectiveness of bevacizumab with first-line combination chemotherapy for Medicare patients with stage IV colorectal cancer. J Clin Oncol. 2012;30(6):608-15. doi:10.1200/JCO.2011.38.9650

JCO.2011.38.9650 [pii].

5. Prasad V, Kim C, Burotto M, Vandross A. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-level meta-analyses. JAMA Intern Med. 2015;175(8):1389-98. doi:10.1001/jamainternmed.2015.2829

2323416 [pii].

6. Visvanathan K, Levit LA, Raghavan D, Hudis CA, Wong S, Dueck A et al. Untapped Potential of Observational Research to Inform Clinical Decision Making: American Society of Clinical Oncology Research Statement. J Clin Oncol. 2017;35(16):1845-54. doi:10.1200/JCO.2017.72.6414.

7. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. Value Health. 2012;15(2):217-30. doi:10.1016/j.jval.2011.12.010

S1098-3015(12)00007-1 [pii].

8. Morton SC, Costlow MR, Graff JS, Dubois RW. Standards and guidelines for observational studies: quality is in the eye of the beholder. J Clin Epidemiol. 2016;71:3-10. doi:10.1016/j.jclinepi.2015.10.014 S0895-4356(15)00501-6 [pii].

9. Rosenbaum PR, Rubin DB. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. The American Statistician. 1985;39(1):33-8. doi:10.1080/00031305.1985.10479383.

10. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac

management on AMI survival using propensity score and instrumental variable methods. JAMA. 2007;297(3):278-85. doi:297/3/278 [pii] 10.1001/jama.297.3.278.

11. D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998;17(19):2265-81. doi:10.1002/(SICI)1097-0258(19981015)17:19<2265::AID-SIM918>3.0.CO;2-B [pii].

12. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. Annu Rev Public Health. 1998;19:17-34. doi:10.1146/annurev.publhealth.19.1.17.

13. Institut National du Cancer (INCa). Information sur le cancer. http://www.e-cancer.fr/. Accessed 04 Jul 2017.

14. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005;365(9472):1687-717. doi:S0140-6736(05)66544-0 [pii]

10.1016/S0140-6736(05)66544-0.

15. Binder-Foucard F, Belot A, Delafosse P, Remontet L, Woronoff AS, Bossard N. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012. Partie 1 – Tumeurs solides. Saint-Maurice (Fra): Institut de veille sanitaire, 2013.

16. Association des Epidémiologistes de Langue Française (ADELF). Déontologie et bonnes pratiques en épidémiologie, recommendations révisées, 2007.

17. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiol Drug Saf. 2008;17(2):200-8. doi:10.1002/pds.1471.

18. Delaloge S, Perol D, Courtinard C, Brain E, Asselain B, Bachelot T et al. Paclitaxel plus bevacizumab or paclitaxel as first-line treatment for HER2-negative metastatic breast cancer in a multicenter national observational study. Ann Oncol. 2016;27(9):1725-32. doi:10.1093/annonc/mdw260 mdw260 [pii].

19. Gobbini E, Ezzalfani M, Dieras V, Bachelot T, Brain E, et Al.Time trends of overall survival among metastatic breast cancer patients in the real life ESME Cohort . Eur J Cancer. (In press)

20. Lardy-Cléaud A, Pistilli B, Frank S, Cottu P, Delaloge S, et Al. Endocrine therapy or chemotherapy as first-line therapy in hormone receptor-positive HER2 negative metastatic breast cancer patients. Eur J Cancer. (In press)

21. Largillier R, Ferrero JM, Doyen J, Barriere J, Namer M, Mari V et al. Prognostic factors in 1,038 women with metastatic breast cancer. Ann Oncol. 2008;19(12):2012-9. doi:10.1093/annonc/mdn424.

22. Asselain B, De Rycke Y, Savignon A, Mould RF. Parametric modelling to predict survival time to first recurrence for breast cancer. Phys Med Biol. 2003;48(12):L31-3.

23. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. J Natl Cancer Inst. 2009;101(23):1642-9. doi:10.1093/jnci/djp369djp369 [pii].

24. Perol D, Robain M, Delaloge S, Cailliot C. Routinely collected data may usefully supplement randomised controlled data on treatment effects for mortality. BMJ. 2016;355:i6745. doi:10.1136/bmj.i6745.

ESME MBC Cohort

TABLES

Table 1 Main data recorded

Patient-related	Hospitalisation-related	Pharmacy record-related
Patient data:	Age	Height
Demographics	Gender	Body Mass Index
Other cancer and family history	Main diagnosis code (ICD code)	Drug (International Non-proprietary Name)
Menopausal status	Linked diagnosis code (ICD code)	Protocol
Initial tumor:	Related significant diagnoses (ICD code)	Cycle ID
Diagnosis	Diagnosis-related group code (reimbursement coding)	Pharmaceutical form and dosage
Relapses	Hospital stay-related group ID (reimbursement coding)	Administration date
Histological results	Entry date	Line treatment number
Medical care	Discharge date	Administered dose
Metastatic disease:	Destination code at discharge	Cumulated dose
Diagnosis	Medical procedures performed	Inclusion in a clinical trial (Yes/No)
Progression	Radiation therapy	
Histological results		
Therapies		
Invasive procedures related to metastasis		
Last contact:		
Vital status		
Date of last contact/death		

ICD International Classification of Diseases

ESME MBC Cohort

Table 2 Characteristics of the at diagnosis of metastatic disease

Characteristic		ESME MBC Population (N=16 702)
Age (years)	Mean (SD)	60.6 (13.8)
	Median [Q1-Q3 range]	61.0 [51.0-71.0]
Sex	Male	149 (0.9%)
	Female	16553 (99.1%)
Histological grade* at primary tumour diagnosis	1	1277 (10.1%)
	2	6438 (50.7%)
	3	4733 (37.3%)
	Not available	240 (1.9%)
	Missing data	2008
Metastatic status	*De novo* MBC	4763 (28.5%)
	Relapsed MBC	11939 (71.5%)
Year of selection (first metastatic treatment)	2008	2651 (15.9%)
	2009	2675 (16.0%)
	2010	2598 (15.6%)
	2011	2515 (15.1%)
	2012	2371 (14.2%)
	2013	2216 (13.3%)
	2014	1676 (10.0%)
Type of metastases	Visceral	9383 (56.2%)
	Bones and not visceral	5047 (30.2%)
	Nodes only	880 (5.3%)
	Skin only	916 (5.5%)
	Skin + nodes	476 (2.8%)
Global HR status ***	Positive	12748 (76.3%)
	Negative	3451 (20.7%)
	Not determined	503 (3.0%)
Global HER2 status***	Positive	2863 (17.1%)
	Negative	12306 (73.7%)
	Not determined	1533 (9.2%)
Triple negative status	Yes	2321 (13.9%)

Data are n (%) unless indicated otherwise

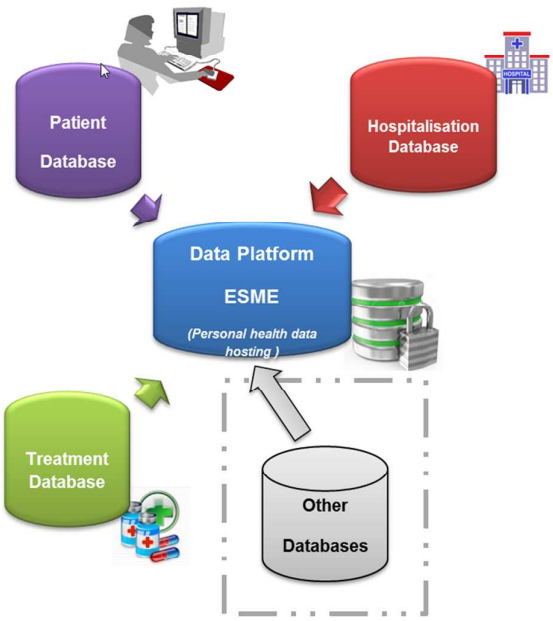
*Histological grade at primary tumour diagnosis: The histological grade at primary tumour diagnosis is defined as the worst histological grade recorded within 1 month (30 days) of the initial diagnosis (primary tumour).

**De novo MBC: Metastatic breast cancer is considered de novo if the diagnosis of metastatic disease occurs within 6 months (180 days) of the initial diagnosis (primary tumour).

***HR and HER2 status:The estrogen receptor (ER) or progesterone receptor status (PR) is considered positive if the pathology report indicates a "positive" result or if the result is not available. The result is considered positive if ≥10% of cells in the sample are positive for ER or PR. The HER2 scoring system and criteria are described in Table 4. If two or more histologic reports are available at the same date the positive status is dominant.The global HR/HER2 status are the status at metastatic diagnosis based on histological results forms related to the primary tumour (if available) or metastatic sites.

FIGURES

Fig. 1 The ESME Data platform



Hospitalisation Database

- Hospitalisations: dates, diagnoses, GHS code
- Medical procedures (inc. Radiotherapy): dates, code

Treatment Database

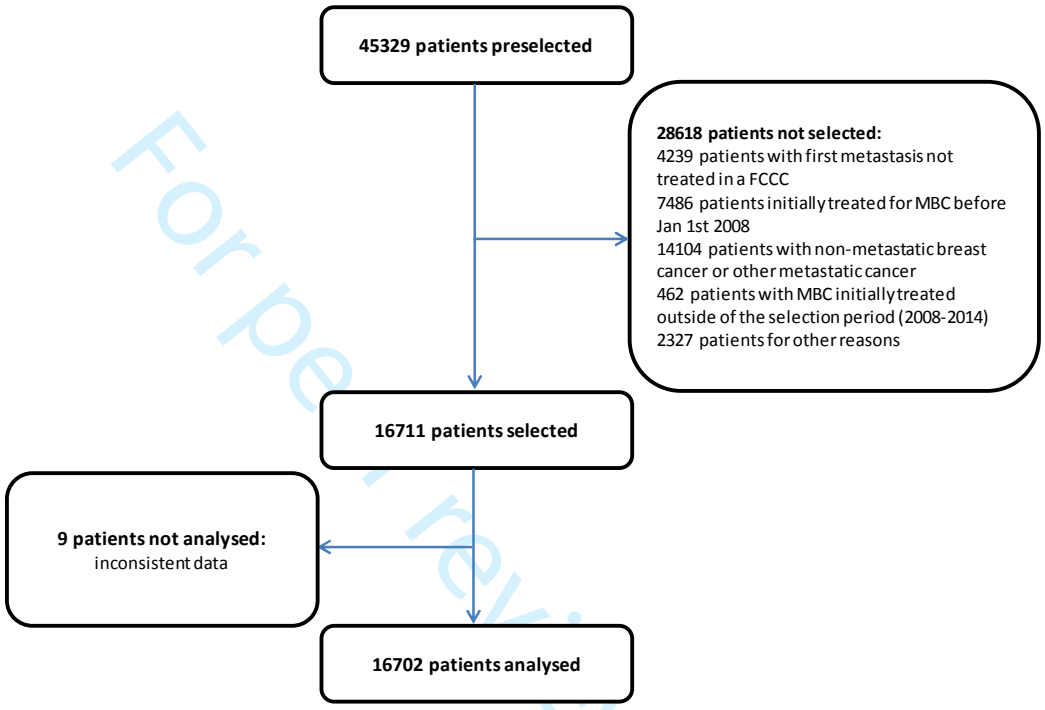
- Pharmacy records: dates, antineoplastic drugs, therapeutic protocol and other concomitant drugs

Patient Database

- Collection based on Electronic medical records
- Patient data : demographics, cancer management, clinical events (progression, relapse), pathological report, metastatic disease, anti-cancer treatment (chemotherapy, targeted therapy, endocrine therapy, immunotherapy), and other therapeutic care (radiotherapy, surgery) or supportive care

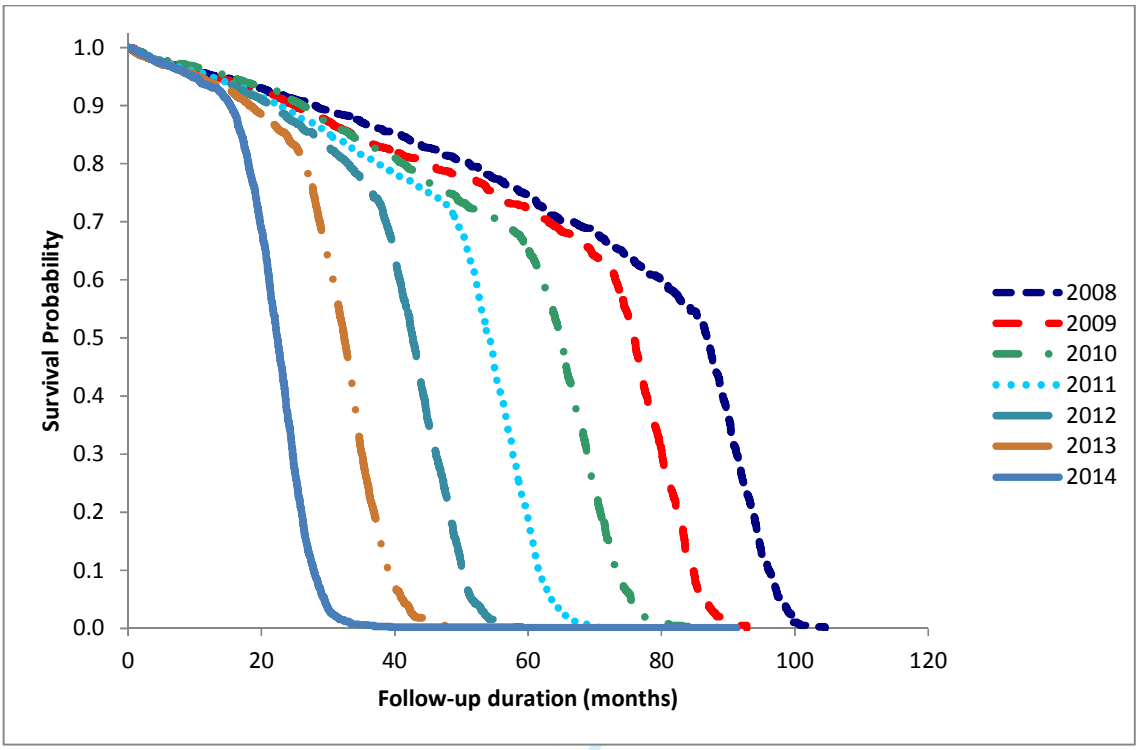
ESME MBC Cohort

Fig. 2 ESME MBC Cohort: Flow chart



ESME MBC Cohort

Fig. 3 Follow-up duration according to year of first-line metastatic treatment



BMJ Open

Cohort profile of the ongoing French Metastatic Breast Cancer (MBC) cohort: the example-based methodology of the Epidemiological Strategy and Medical Economics (ESME)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023568.R1
Article Type:	Cohort profile
Date Submitted by the Author:	13-Jul-2018
Complete List of Authors:	Pérol, David; Léon Bérard Cancer Centre, Unit of Biostatistics and Therapeutic Evaluation Robain, Mathieu; R&D UNICANCER Arveux, F; Registre des cancers du sein de Côte d'Or, centre Georges-François-Leclerc Mathoulin-Pélissier , S; INSERM, Chamorey, Emmanuel; Centre Antoine-Lacassagne Asselain, Bernard Berchery, Delphine; Institut Claudius Regaud Gourgou, Sophie; Institut regional du Cancer de Montpellier Breton, Mathias; Centre Eugene Marquis Delaine-Clisant , Stéphanie; Centre Oscar Lambret, Clinical Research Unit Mons, Muriel; Gustave Roussy Diéras, Véronique; Ensemble Hospitalier de l'Institut Curie Carton, Matthieu; Institut Curie, Biométrie; Guizard, Anne-Valérie ; Centre Francois Baclesse Centre de Lutte Contre le Cancer Laborde, Lilian; Institut Paoli-Calmettes Laurent, Carine; Institut de Cancerologie de Lorraine Loeb, Agnès; Centre Henri Becquerel Mouret-Reynier, Marie-Ange ; Centre Jean Perrin Parent, Damien; Institut Jean Godinot Perrocheau, Geneviève; Institut de Cancérologie de l'Ouest Campion, Loïc; Institut de Cancérologie de l'Ouest Velten, Michel; Centre Paul Strauss Cailliot, Christian; R&D UNICANCER Ezzalfani, Monia; Centre Leon Berard Simon, Gaëtane; R&D UNICANCER,
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Medical management
Keywords:	Real-world data, Data Platform, Epidemiology < ONCOLOGY, Patient medical record, Quality control, Metastatic breast cancer

TITLE PAGE

Cohort profile of the ongoing French Metastatic Breast Cancer (MBC) cohort: the example-based methodology of the Epidemiological Strategy and Medical Economics (ESME)

List of authors:

David Pérol (Centre Léon Bérard, Lyon), Mathieu Robain (R&D UNICANCER, Paris), Patrick Arveux (Centre Georges-François Leclerc, Dijon), Simone Mathoulin-Pelissier (Institut Bergonié, Bordeaux), Emmanuel Chamorey (Centre Antoine Lacassagne, Nice), Bernard Asselain (University Paris Sud), Delphine Berchery (Institut Claudius Régaud, Toulouse), Sophie Gourgou (Institut régional du Cancer Montpellier, Montpellier), Mathias Breton (Centre Eugène Marquis, Rennes), Stéphanie Clisant (Centre Oscar Lambret, Lille), Muriel Mons (Gustave Roussy, Villejuif), Véronique Diéras (Ensemble Hospitalier de l'Institut Curie, Paris & Saint-Cloud), Matthieu Carton (Ensemble Hospitalier de l'Institut Curie, Paris & Saint-Cloud), Anne-Valérie Guizard (Centre François Baclesse, Caen), Lilian Laborde (Institut Paoli-Calmettes, Marseille), Carine Laurent (Institut de Cancérologie de Lorraine, Nancy), Agnès Loeb (Centre Henri Becquerel, Rouen), Marie-Ange Mouret-Reynier (Centre Jean Perrin, Clermont-Ferrand), Damien Parent (Institut Jean Godinot, Reims), Geneviève Perrocheau (Institut de Cancérologie de l'Ouest, Nantes & Angers), Loïc Campion (Institut de Cancérologie de l'Ouest, Nantes & Angers), Michel Velten (Centre Paul Strauss, Strasbourg), Christian Cailliot (R&D UNICANCER, Paris), Monia Ezzalfani (Centre Léon Bérard, Lyon), Gaëtane Simon (R&D UNICANCER, Paris)

Corresponding author:

David Pérol (Centre Léon Bérard, Lyon)

E-mail: david.perol@lyon.unicancer.fr

Tel: +33 (0)4.69.85.61.00

ABSTRACT

Purpose

The currently ongoing Epidemiological Strategy and Medical Economics (ESME) research program aims at centralising real-life data on oncology care for epidemiological research purposes. We draw on results from the metastatic breast cancer (MBC) cohort to illustrate the methodology used for data collection in the ESME research program.

Participants

All consecutive ≥18 years patients with MBC treatment initiated between 2008 and 2014 in one of the 18 French Comprehensive Cancer Centres (FCCCs) were selected. Diagnostic, therapeutic, and follow-up data (demographics, primary tumor, metastatic disease, treatment patterns and vital status) were collected through the course of the disease. Data collection is updated annually.

Finding to date

With a recruitment target of 30,000 MBC patients by 2019, we currently screened a total of 45,329 patients, and over 16,700 patients with a metastatic disease treatment initiated after 2008 have been selected. 20.7% of patients had a HR-negative MBC, 73.7% had a HER2-negative MBC, and 13.9% were classified as triple-negative BC (i.e. HER2 and HR status both negative). Median follow-up duration from MBC diagnosis was 48.55 months for the whole cohort.

Future plans

These real-world data will help standardise the management of MBC and improve patient care. A dozen of ancillary research projects have been conducted and some of them are already accepted for publication or ready to be issued. The ESME research program is expanding to ovarian cancer (OC) and advanced/metastatic lung cancer (AMLC). Our ultimate goal is to achieve a continuous link to the data of the cohort to the French national Health Data System (SNDS) for centralising data on healthcare reimbursement (drugs, medical procedures), inpatient/outpatient stays, and visits in primary/secondary care settings.

KEYWORDS

Real-world data; real-life data; patient medical record; quality control; oncology; metastatic breast cancer

Strengths and limitations

- The epidemiological strategy and medical economics (ESME) research program aims at centralising real-life data on oncology care for epidemiological research purposes. The ongoing screening in the metastatic breast cancer (MBC) cohort currently reached over 16,700 patients with a metastatic disease treatment initiated after 2008 have been collected, currently contributing to the development of one of the most important cohort of patients with treated MBC.
- Screening process of patients and data collection (diagnostic, therapeutic, and follow-up data) through the course of the disease provide a solid base of knowledge for real-world survival.
- Significant resources are deployed to achieve a high quality level for the validation of the data, including systematic consultancy of the source folder for data collection, and the implementation of effective quality control, and regular audit.
- Despite the current large-scale recruitment of patients and greater than one third of French MBC patients managed in FCCCs, future studies should integrate the diversity of management options adopted in any health institutions.
- These real-world data will help standardise the management of MBC and will contribute to the improvement of medical expertise through the use, the interpretation and the analysis of the generated real-world data collection.

BACKGROUND

Real-world evidence (RWE) studies and observational studies using real world data (RWD) play a growing role in conducting comparative effectiveness research on pharmaceutical products and other healthcare interventions. They aim to bridge the gap between the highly controlled environment of randomised clinical trials (RCTs) and real-life clinical practice [1]. In particular, health authorities are interested in gathering real-world data for long-term benefit-risk assessment and, increasingly, health economic evaluations for reimbursement decisions [2]. With their high internal validity, RCTs are considered the gold standard of evidence for establishing treatment efficacy, although the generalisability of findings to clinical practice may be limited [3]. In fact, cancer survival endpoints in the real-life setting may differ from that measured in traditional RCTs [4]. Furthermore, other limitations such as short follow-up and small sample size have created uncertainty in estimating survival criteria in RCTs, and thus surrogate endpoints are generally used such as progression free survival (PFS) or time to progression (TTP) [5]. On the other hand, large population-based cohorts with longer follow-up periods can be particularly appropriate to assess long-term clinical outcomes , such as overall survival (OS) outside of the RCT setting [6] and to detect changes in medical practice.

RWE and observational studies are non-experimental research where cancer management (treatments and disease evolution assessment) is left to the choice of care providers and patients [7]. Over the past few years, a broad consensus arose around the requirement for high-quality data from large cohorts to strengthen and drive improvements in research methods and practices [8]. The goal of these studies is to generate complementary information to RCTs based on larger samples and provide answers regarding particular populations in real-life clinical practice. These studies are more prone to biases such as baseline differences between patients (selection bias) or bias due to confounding by

indication for example. To minimise sources of such biases, statistical approaches including adjusted analyses, propensity score methods, or instrumental variables may be also employed [9-12].

Metastatic breast cancer (MBC) is one of the leading causes of cancer-related mortality among women in Western countries [13]. A relatively high proportion (approximately 30%) of breast cancer patients develop metastatic disease [14], and while significant treatment advances have been made, the overall prognosis is poor with a five-year survival rate of 25% [15]. The national academic network of cancer centres in France (French Comprehensive Cancer Centres; FCCCs), which together handle over one-third of all breast cancer cases nationally, decided to launch in 2014 a program dedicated focussing on RWD in oncology databases in MBC through Epidemiological Strategy and Medical Economics (ESME) research program.

The ESME research program aims to build a comprehensive database on oncology care for epidemiological research purposes to improve knowledge on medical practice in real-life setting, on public health and healthcare use, and to provide information to health authorities and other associated bodies.

Several studies based on real-life data collection have been developed in this program. Different cohorts of patients with ovarian cancer, patients with advanced/metastatic lung cancer are currently recruiting.

Some results from the ongoing ESME MBC project (ClinicalTrials.gov NCT03275311) whose aim is to describe the medical care of patients treated for MBC according to their disease characteristics the evolution of their metastatic disease and their outcomes have been recently published. Recent analyses explored overall survival in different subgroups of MBC patients, and first line therapy in HER2 negative MBC patients, and in hormone receptor-positive HER2-negative MBC patients [16-18].

In this paper, we describe the methodological principles that underpin the ESME research program, and illustrated this innovative approach through this first ESME metastatic breast cancer (MBC) cohort. Design, brief description of the selected population and current status are reported.

METHODS

The ESME MBC cohort is a population-based registry in 18 FCCCs (http://www.unicancer.fr/en/rd-unicancer/esme), which collected data on all consecutive patients treated for MBC from 2008. Annual data collection phases are planned to add new diagnosed cases and update patients' follow up data.

STUDY POPULATION

Eligibility criteria

Patients eligible to the ESME MBC cohort were male or female patients aged ≥18 years with MBC whose first metastasis had been either completely or partially treated between January

1st, 2008 and December 31st, 2014 in one of the FCCC. MBC treatment could include radiotherapy, chemotherapy, targeted therapy, immunotherapy, or endocrine therapy.

Patient screening process

Patient screening process involved two steps: an automated case screening followed by the validation of selection for each screened case. The automated case screening was based on information retrieved from multiple data sources available within each FCCC: administrative records [French National Computerised Medical Information System] (via MBC-specific International Classification of Diseases (ICD) codes associated with inpatient stays), pharmacy records, patient medical records (PMRs), including multidisciplinary team meeting records and search using relevant keywords, MBC-specific registries.

The objective of the first step -automated screening step- was to identify all cases with inpatient stays or therapeutic management for MBC in one of the FCCC during the selection period and generate the patient screening list. The ICD codes used were: C50 (Malignant neoplasm of breast), C77.- [except C77.3] (Secondary and unspecified malignant neoplasm of lymph nodes), C78.- (Secondary malignant neoplasm of respiratory and digestive organs) and C79.- (Secondary malignant neoplasm of other and unspecified sites). Once the patient screening list was finalised in each centre, data were subsequently anonymised and each patient had been assigned an ESME number. The first screening step did not allow to precisely identify the date for first MBC-specific treatment, and the second step was performed to cross-check eligibility criteria for all screened cases and specify the dates related to the initiation of the MBC first-line treatment, using data from the patient medical records.

ETHICS AND PATIENT DATA PROTECTION

The ESME MBC database received approval from the French data protection authority (*Commission Nationale de l'Informatique et des Libertés,* authorisation no. 1704113)). No informed consent was required.

The ESME research program was managed by R&D Unicancer in accordance with guidelines for Good Pharmacoepidemiology Practices and Good Epidemiology Practices [19, 20].

Ancillary projects analyses were notified to an independent ethics committee (Lyon Sud-Est II) on December 17th, 2015.

Data collection

The data of the selected patients were planned to be collected by trained research assistants and annually updated. The data collection was performed in two phases from October 2014 to October 2016. A first data collection phase conducted in 2014-2015 collected the data from patients with MBC treatment initiated between January 1st, 2008 and December 31st, 2013 in one of the FCCC. A second phase of data collection performed in 2016 added to the ongoing database the data from patients with MBC treatment initiated between January 1st, and December 31st, 2014, and follow-up data for the global cohort were consequently updated. Hence, the ongoing database provide an overview of all the data from patients with a MBC treatment initiated from 2008; information are updated with last contact information available in the PMR at the date of the data collection.

Baseline and follow-up data

The ESME MBC Cohort is composed of 3 types of data set (**Figure 1**):

- *Patient-related data* are obtained from systematic review of patient medical records (non-structured data) and provide information on patient demographics, cancer family history, characteristics of primary tumour, relapses, metastatic recurrences pathological reports (tumor size, grade, histological type), hormone-receptor status and HER-2 status, therapeutic care (focussing on cancer-related treatment) and settings, reasons for treatment termination, and clinical events.
- *Hospitalisation records* are integrated data from a structured and automated database related to inpatient stays, and primarily used to bill the French National Health Insurance Fund (*Assurance Maladie*). It provides information on patient entry and discharge information (date and destination code at discharge), ICD codes associated to each stay, diagnostic, medical and therapeutic procedures including radiotherapy and surgery.
- *Pharmacy-dispensed treatment records* includes all data related to anti-cancer treatments obtained from each centre's pharmacy database: drugs (International Non-proprietary Name), administration date, protocol name, patient's height and BMI, cycle ID, UCD code (pharmaceutical form related to the drug dosage), line of treatment, administration in a clinical trial (yes/no). It exclusively includes information on products that are prescribed by each FCCC.

The detailed raw data collected above and derived data are listed in **Table 1**.

DATA MANAGEMENT

Any data integrated in the ESME research program are subject to Quality Control procedures.

Patient-related data are registered via an electronic case report form (eCRF) in each centre by trained clinical research assistants (CRAs) between December 2014 and October 2016. Medical support to assist CRAs was provided to ensure quality-controlled data and appropriate recoding was performed before annual database lock when required. All ESME procedures are handled according to the Guidelines for good pharmacoepidemiology practices [20]. Importantly, all data are exclusively obtained retrospectively; no attempts are made to recover unavailable data from patients' medical records by contacting healthcare providers or patients.

The clinical data management system (CDMS) used was Statistical Analysis System (SAS) software. For both, ESME MBC database and the eCRF tool administration used an Oracle solution and certified personal data hosting system guarantee data security.

On-site quality review

On-site quality review (QR) of the patient screening process was carried out. This consisted of checking eligibility criteria on samples of selected and non-selected cases in each FCCC. For selected cases, key variables were crosschecked versus the source data. For all Quality Controls (QC) procedures, accepted error limits were 10% for non-selected patients and 5% for selected patients. A central audit was subsequently performed by the Unicancer Quality

Assurance Department, and an audit on data registration and generated screening list was also conducted at the local level.

DATA QUALITY ASSURANCE

Governance structure

Three boards monitor the ESME Research program : Scientific Committee, deontology Committee and International Advisory Board. The main role of the Scientific Committee is to: i) ensure that the applicable scientific rules are followed, ii) evaluate any ancillary projects in compliance with defined criteria and scientific pertinence, and iii) monitor the all validated ancillary projects. The deontology Committee monitors any potential conflicts of interest related to experts involved in the program, gives recommendations that may improve the prevention of conflicts, provides opinions on individual or particular situations, and potential collaborations with private partners. The ESME International Advisory Board has a consultative role with regard to coherence of the scientific program and reviews key international communications, formulates recommendations for publication rules or methodology, and reinforces international academic cooperation.

DATA ANALYSIS PRINCIPLES

Academic research teams or private organisations could propose ancillary projects.

For each accepted ancillary project, statistical analyses are conducted according to a detailed statistical analysis plan (SAP) that must be reviewed by the scientific committee. This article does not aim at providing a comprehensive and exhaustive review on appropriate statistical methods to reduce bias related to analysis based on RWD.

A first ancillary project reported the outcomes (OS and PFS) following first-line paclitaxel treatment with or without bevacizumab [16]. Other ongoing analyses of sensibility will better address the bias potentially found in real-life settings. Two analyses reported the description of OS in different subgroups of MBC patients over the time, and results for the first-line therapy (Endocrine therapy or chemotherapy) in hormone receptor-positive HER2 negative cancer subgroup respectively [17,18]. Other series accepted for communications (abstracts/posters) to major congress in 2017 reported epidemiological analyses (i.e. impact of age at diagnosis, …), therapeutic management (i.e. impact of loco-regional treatment on OS, …) and specific analyses for drug use in routine practice (i.e. use of vinorelbine, everolimus, etoposide…).

Different other sub-populations are currently being considered such as subgroups of metastatic Triple Negative Breast Cancer, metastatic HER2 positive MBC…

PATIENT AND PUBLIC INVOLVEMENT STATEMENT.

Patients and/or public were not involved.

RESULTS AND CURRENT STATUS

16,711 out of the 45,329 patients screeneed were selected in the ESME MBC Cohort (see **Figure 2**).

The sensitivity and specificity of the three main screening sources used were explored. Sensitivity was highest for administrative records (78% *versus* 43% for pharmacy records, and 28% for BC-specific local registries). On the other hand, specificity was highest for BC-specific local registries (87% *versus* 67% for pharmacy records, and 49% for administrative records).

The main reasons for non-selection of screened patients were: presence of non-metastatic breast cancer or other metastatic cancers (n=14,104 patients), initial MBC treatment received before January 1st, 2008 (n=7486), and first metastasis not initially treated in an FCCC (n=4239). Nine additional were excluded prior to the final database lock due to inconsistencies in the dates. A total of 16,702 patients was analysed **(Figure 2).**

Table 2 summarises the main demographic and disease characteristics at the time of initial metastatic diagnosis. Patients were nearly all women (99.1%) with a median age of 61 years. Over half (56.2%) had at least visceral metastases present, with 30.2% having at least bone and non-visceral metastases, and 13.6% with skin only, or node only, or at least skin plus node. 20.7% of patients had a HR-negative MBC, 73.7% had a HER2-negative MBC, and 13.9% were classified as triple-negative BC (i.e. HER2 and HR status both negative).

Median follow-up duration from MBC initiation treatment was 48.55 months for the whole cohort [95% CIs 47.7–49.38].

DISCUSSION

Retrospective analysis using real world data is likely to become increasingly important to ensure that medications are accepted by policymakers and adopted by patient practitioners. The ESME Research program is a large-scale initiative to provide access to real-world data in oncology. This ongoing ESME MBC cohort currently centralises data from 16 711 patients.

The ESME research program provides a unique opportunity to study a diverse range of topics related to MBC care and management in real-life settings. Indeed, there are many potential applications, including study of the factors influencing patient care (e.g. cancer and patient characteristics), description of therapeutic strategies (treatment lines and sequences of therapies, etc.), measurement of clinical events (disease progression, death, persistence of treatment effect). Characterisation of patients enrolled in clinical trials is also possible, as is the reconstruction of "virtual trials" using appropriate statistical methodologies. Potentially, these data could be used for health economics evaluation of management strategies for patients (e.g. rehospitalisation and related ambulatory care), as well as reconstruction of health care trajectories through data modelling.

ESME MBC Cohort

The ESME research program includes alternative approaches to generate cohorts that use different types of RWD (clinical data, therapeutic treatment data, long term outcomes, health economics data) in the FCCCs *vs* existing registries in France, Europe and the US (e.g. SEER). It involves rigorous procedures for patient screening and data collection, ensuring both validity and reliability of data. It uses a fully retrospective approach, with no influence on treatment practice or interaction with oncologists. Unlike prospective interventional or observational research studies, data are not influenced by study design and reflect the real-life management of patients treated. While data recorded for the cohort are defined by experts in the field, the vast majority of data are collected by trained clinical research technicians, thereby minimising any potential risk of data misinterpretation. As discussed above, the ESME MBC Cohort offers a unique opportunity to study a wide range of research questions in a large sample. With respect to evaluation of treatment strategies, the database enables reliable estimation of survival criteria such as overall survival (OS) and surrogates endpoints (PFS …). OS improvement in diseases with a long median post-progression survival time, such as MBC is a critical endpoint [21-23].

The ESME MBC cohort also has several limitations. For example, the database relies on the collection and restructuring of existing data only, i.e. there is no creation of new data. Furthermore, apart from events reported in the patient medical record impacting therapeutic management, adverse effects are not routinely captured. Conceivably, further in-depth analysis of the data could highlight trends such as treatment interruption or discontinuation due to toxicity, which is important from a risk-management perspective. The main potential sources of bias include selection bias, and information bias due to differences in patient monitoring and non-standardised data collection. Selection bias has been taken into account by using rigorous selection procedures across all 18 FCCCs, and the data management plan and quality control program described above have been designed to limit information bias. Nevertheless, due to the retrospective data collection and the fact that it is based on real-life follow-up, clinical and biological events are not evaluated at predefined time points (unlike in RCTs). For example, objective response, historical endpoint in RCTs, could not be assessed retrospectively without a central review of existing imaging as not systematically documented in routine practice. The information collected therefore depends on the frequency of follow-up visits and clinical and radiological exams prescribed by the patient's doctor. As clinical signs are the only means by which disease metastasis can be identified, the number of disease progressions may be underestimated. With respect to the clinical event of death, all deaths are reported in the patient medical records.

With respect to evaluation of treatment strategies, analysis of real-life data poses unique challenges, such as accounting for confounding factors between patient groups, although various statistical approaches can be used to address this, as discussed above [24].

Concerning overall generalisability and applicability (external validity), it should be noted that the cohort centralises data from patients treated in specialised cancer centres only. FCCCs may utilise different clinical practices compared to public hospitals and private institutions, and thus patients from FCCCs may not be truly representative of all French breast cancer patients. Potentially, data extrapolation from all French health care organisations could be developed with the Exhaustive National Health Reimbursement System (SNIIRAM; *Système national d'information inter-régimes de l'Assurance maladie*).

The ESME MBC Cohort aims to collect data for up to 30,000 patients by 2019. As mentioned, future aims might include to continuously link our database to those from other

institutions, such as the SNDS database for data on exhaustive healthcare reimbursement, and the INSEE database to provide vital status updates for patients lost to follow-up. The ESME research program has further expanded to ovarian cancer and advanced/metastatic lung cancer. Real world data from the ESME cohorts should help to provide medical recommendations and ultimately improve patient care.[25]

ACKNOWLEDGEMENTS

We thank the 18 French Comprehensive Cancer Centres for providing the data and each ESME contact for coordinating the project at the local level.

We thank the ESME Scientific Committee members for their ongoing support. We also acknowledge Dr. Sarah Hopwood (Scinopsis, France) for editorial assistance with this publication, paid by R&D UNICANCER.

The 18 French Comprehensive Cancer Centres (FCCC) are: Ensemble hospitalier de l'Institut Curie, Paris/Saint-Cloud, Gustave Roussy, Villejuif, Institut de Cancérologie de l'Ouest, Nantes/Angers, Centre François Baclesse, Caen, Institut du Cancer de Montpellier Montpellier, Centre Léon Bérard, Lyon, Centre Georges-François Leclerc, Dijon, Centre Henri Becquerel, Rouen; Institut Claudius Régaud, Toulouse; Centre Antoine Lacassagne, Nice; Institut de Cancérologie de Lorraine, Nancy; Centre Eugène Marquis, Rennes; Institut Paoli-Calmettes, Marseille; Centre Jean Perrin, Clermont-Ferrand; Institut Bergonié, Bordeaux; Centre Paul Strauss, Strasbourg; Institut Jean Godinot, Reims; Centre Oscar Lambret, Lille.

ESME central coordinating staff: Head of Research and Development: Christian Cailliot. Program director: Mathieu Robain. Data Managers: Irwin Piot and Olivier Payen. Project Managers: Coralie Courtinard, Tahar Guesmia and Gaëtane Simon. Project assistant: Esméralda Pereira. Software designer: Alexandre Vanni. ESME local coordinating staff: Patrick Arveux, Thomas Bachelot, Jean-Pierre Bleuse, Delphine Berchery, Mathias Breton, Stéphanie Clisant, Emmanuel Chamorey, Valérie Dejean, Véronique Diéras, Anne-Valérie Guizard, Anne Jaffré, Lilian Laborde, Agnès Loeb, Muriel Mons, Damien Parent, Geneviève Perrocheau, Marie-Ange Mouret-Reynier, Carine Laurent, Michel Velten.

Funding: This work was supported by R&D UNICANCER and initially funded by an industrial consortium (Roche, Pierre Fabre, Pfizer, and Astrazeneca), and support was subsequently provided by MSD, Eisai, and Daiichi Sankyo. The funders did not have any input into maintenance of the database (project design, data collection, management) or analysis of data. They were not involved in the writing of the manuscript or decision to submit for publication.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest: D Perol has received personal fees (honoraria and travel/accommodation expenses) from Laboratoire Roche, outside the submitted work. B Asselain has received personal fees for board membership and for consultancy from Roche Pharma, outside the submitted work. S Gourgou has received personal fees for board

membership from Celgene and for consultancy from Roche, outside the submitted work. All other authors declare that they have no conflict of interest.

Ethical approval: Approval for the ESME MBC database was obtained from an independent ethics committee, which waived the requirement for informed consent.

REFERENCES

1. Chavez-MacGregor M, Giordano SH. Randomized Clinical Trials and Observational Studies: Is There a Battle? J Clin Oncol. 2016;34(8):772-3.

2. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part I. Value Health. 2009;12(8):1044-52..

3. Hahn OM, Schilsky RL. Randomized controlled trials and comparative effectiveness research. J Clin Oncol. 2012;30(34):4194-201.

4. Meyerhardt JA, Li L, Sanoff HK, Carpenter Wt, Schrag D. Effectiveness of bevacizumab with first-line combination chemotherapy for Medicare patients with stage IV colorectal cancer. J Clin Oncol. 2012;30(6):608-15.

5. Prasad V, Kim C, Burotto M, Vandross A. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-level meta-analyses. JAMA Intern Med. 2015;175(8):1389-98.

6. Visvanathan K, Levit LA, Raghavan D, Hudis CA, Wong S, Dueck A et al. Untapped Potential of Observational Research to Inform Clinical Decision Making: American Society of Clinical Oncology Research Statement. J Clin Oncol. 2017;35(16):1845-54.

7. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. Value Health. 2012;15(2):217-30.

8. Morton SC, Costlow MR, Graff JS, Dubois RW. Standards and guidelines for observational studies: quality is in the eye of the beholder. J Clin Epidemiol. 2016;71:3-10.

9. Rosenbaum PR, Rubin DB. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. The American Statistician. 1985;39(1):33-8.

10. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. JAMA. 2007;297(3):278-85.

11. D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998;17(19):2265-81.

12. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. Annu Rev Public Health. 1998;19:17-34.

13. Institut National du Cancer (INCa). Information sur le cancer. http://www.e-cancer.fr/. Accessed 04 Jul 2017.

14. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005;365(9472):1687-717.

15. Binder-Foucard F, Belot A, Delafosse P, Remontet L, Woronoff AS, Bossard N. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012. Partie 1 – Tumeurs solides. Saint-Maurice (Fra): Institut de veille sanitaire, 2013.

16. Delaloge S, Perol D, Courtinard C, Brain E, Asselain B, Bachelot T et al. Paclitaxel plus bevacizumab or paclitaxel as first-line treatment for HER2-negative metastatic breast cancer in a multicenter national observational study. Ann Oncol. 2016;27(9):1725-32.

17. Gobbini E, Ezzalfani M, Dieras V, Bachelot T, Brain E, et al. Time trends of overall survival among metastatic breast cancer patients in the real life ESME Cohort . Eur J Cancer. Eur J Cancer. 2018 Jun;96:17-24.

18. Jacquet E, Lardy-Cléaud A, Pistilli B, Frank S, Cottu P, Delaloge S, et al. Endocrine therapy or chemotherapy as first-line therapy in hormone receptor-positive HER2 negative metastatic breast cancer patients. Eur J Cancer. Eur J Cancer. 2018 May;95:93-101.

19. Association des Epidémiologistes de Langue Française (ADELF). Déontologie et bonnes pratiques en épidémiologie, recommendations révisées, 2007.

20. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiol Drug Saf. 2008;17(2):200-8.

21. Largillier R, Ferrero JM, Doyen J, Barriere J, Namer M, Mari V et al. Prognostic factors in 1,038 women with metastatic breast cancer. Ann Oncol. 2008;19(12):2012-9.

22. Asselain B, De Rycke Y, Savignon A, Mould RF. Parametric modelling to predict survival time to first recurrence for breast cancer. Phys Med Biol. 2003;48(12):L31-3.

23. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. J Natl Cancer Inst. 2009;101(23):1642-9.

24. Perol D, Robain M, Delaloge S, Cailliot C. Routinely collected data may usefully supplement randomised controlled data on treatment effects for mortality. BMJ. 2016;355:i6745.

25. Berdaï D, Thomas-Delecourt F, Szwarcensztein K, d'Andon A, Collignon C, Comet D. Requests for post-registration studies (PRS), patients follow-up in actual practice: Changes in the role of databases. Therapie. 2018;73:13-24.

Formatted: Font: 12 pt, French (France)

Formatted: French (France)

ESME MBC Cohort

TABLES

Table 1 Main data recorded

Patient-related	Hospitalisation-related	Pharmacy record-related
Patient data:	Age	Height
Demographics	Gender	Body Mass Index
Other cancer and family history	Main diagnosis code (ICD code)	Drug (International Non-proprietary Name)
Menopausal status	Linked diagnosis code (ICD code)	Protocol
Initial tumor:	Related significant diagnoses (ICD code)	Cycle ID
Diagnosis	Diagnosis-related group code (reimbursement coding)	Pharmaceutical form and dosage
Relapses	Hospital stay-related group ID (reimbursement coding)	Administration date
Histological results	Entry date	Line treatment number
Medical care	Discharge date	Administered dose
Metastatic disease:	Destination code at discharge	Cumulated dose
Diagnosis	Medical procedures performed	Inclusion in a clinical trial (Yes/No)
Progression	Radiation therapy	
Histological results		
Therapies		
Invasive procedures related to metastasis		
Last contact:		
Vital status		
Date of last contact/death		

ICD International Classification of Diseases

ESME MBC Cohort

Table 2 Characteristics of the patients at of the date of metastatic disease diagnosis.

Characteristic		ESME MBC Population (N=16 702)
Age (years)	Mean (SD)	60.6 (13.8)
	Median [Q1-Q3 range]	61.0 [51.0-71.0]
Sex	Male	149 (0.9%)
	Female	16553 (99.1%)
Histological grade* at primary tumour diagnosis	1	1277 (10.1%)
	2	6438 (50.7%)
	3	4733 (37.3%)
	Not available	240 (1.9%)
	Missing data	2008
Metastatic status**	*De novo* MBC	4763 (28.5%)
	Relapsed MBC	11939 (71.5%)
Year of first metastatic treatment	2008	2651 (15.9%)
	2009	2675 (16.0%)
	2010	2598 (15.6%)
	2011	2515 (15.1%)
	2012	2371 (14.2%)
	2013	2216 (13.3%)
	2014	1676 (10.0%)
Type of metastases	Visceral	9383 (56.2%)
	Bones and not visceral	5047 (30.2%)
	Nodes only	880 (5.3%)
	Skin only	916 (5.5%)
	Skin + nodes	476 (2.8%)
Global HR status***	Positive	12748 (76.3%)
	Negative	3451 (20.7%)
	Not determined	503 (3.0%)
Global HER2 status***	Positive	2863 (17.1%)
	Negative	12306 (73.7%)
	Not determined	1533 (9.2%)
Triple negative status	Yes	2321 (13.9%)

Data are n (%) unless indicated otherwise

*Histological grade at primary tumour diagnosis: The histological grade at primary tumour diagnosis is defined as the worst histological grade recorded within one month (30 days) after the initial diagnosis (primary tumour).

***de novo* MBC: Metastatic breast cancer is considered *de novo* if the diagnosis of metastatic disease occurs within 6 months (180 days) after the initial diagnosis (primary tumour).

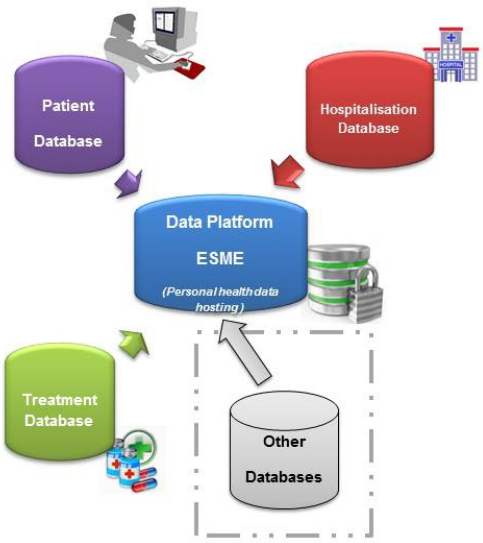
***HR and HER2 status: The estrogen receptor (ER) or progesterone receptor status (PR) is considered positive if the pathology report indicates a "positive" result, or considered as

positive when ≥10% of cells in the sample are positive for ER, or PR respectively. The HER2 status is considered positive if the pathology report indicates for the immunohistochemistry (IHC) result "3+", "2+" or not available, the result will be considered positive if the Fluorescent In Situ Hybridization (FISH) or Chromogenic In Situ Hybridization (CISH) result is positive. If two or more histologic reports are available at the same date, the positive status is preferred. The global HR/HER2 status indicates the status at metastatic diagnosis based on histological results forms related to the primary tumour (if available) or metastatic sites.

FIGURES

Figure 1. The ESME Data platform

Figure 2. ESME MBC Cohort: Flow chart



Hospitalisation Database

- Hospitalisations: dates, diagnoses, GHS code
- Medical procedures (inc. Radiotherapy): dates, code

Treatment Database

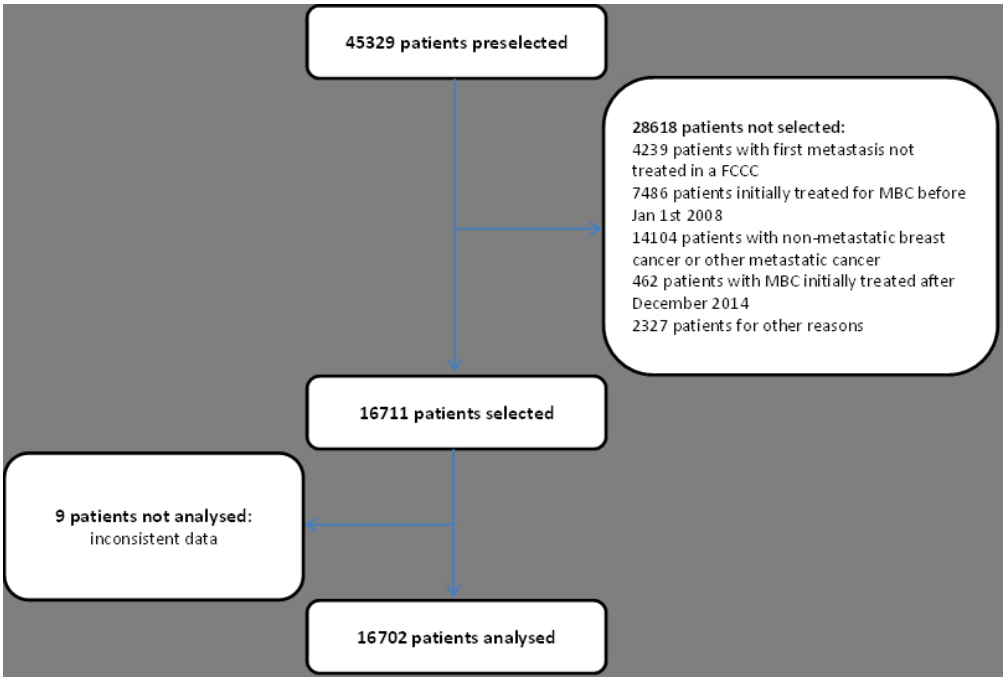
- Pharmacy records: dates, antineoplastic drugs, therapeutic protocol and other concomitant drugs

Patient Database

- Collection based on Electronic medical records
- Patient data : demographics, cancer management, clinical events (progression, relapse), pathological report, metastatic disease, anti-cancer treatment (chemotherapy, targeted therapy, endocrine therapy, immunotherapy), and other therapeutic care (radiotherapy, surgery) or supportive care

The ESME Data platform

254x190mm (96 x 96 DPI)



The ESME MBC Cohort: Flow chart

213x143mm (96 x 96 DPI)

BMJ Open

Cohort profile of the ongoing French Metastatic Breast Cancer (MBC) cohort: the example-based methodology of the Epidemiological Strategy and Medical Economics (ESME)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023568.R2
Article Type:	Cohort profile
Date Submitted by the Author:	19-Nov-2018
Complete List of Authors:	Pérol, David; Léon Bérard Cancer Centre, Unit of Biostatistics and Therapeutic Evaluation Robain, Mathieu; R&D UNICANCER Arveux, F; Registre des cancers du sein de Côte d'Or, centre Georges-François-Leclerc Mathoulin-Pélissier , S; INSERM, Chamorey, Emmanuel; Centre Antoine-Lacassagne Asselain, Bernard Berchery, Delphine; Institut Claudius Regaud Gourgou, Sophie; Institut regional du Cancer de Montpellier Breton, Mathias; Centre Eugene Marquis Delaine-Clisant , Stéphanie; Centre Oscar Lambret, Clinical Research Unit Mons, Muriel; Gustave Roussy Diéras, Véronique; Ensemble Hospitalier de l'Institut Curie Carton, Matthieu; Institut Curie, Biométrie; Guizard, Anne-Valérie ; Centre Francois Baclesse Centre de Lutte Contre le Cancer Laborde, Lilian; Institut Paoli-Calmettes Laurent, Carine; Institut de Cancerologie de Lorraine Loeb, Agnès; Centre Henri Becquerel Mouret-Reynier, Marie-Ange ; Centre Jean Perrin Parent, Damien; Institut Jean Godinot Perrocheau, Geneviève; Institut de Cancérologie de l'Ouest Campion, Loïc; Institut de Cancérologie de l'Ouest Velten, Michel; Centre Paul Strauss Cailliot, Christian; R&D UNICANCER Ezzalfani, Monia; Centre Leon Berard Simon, Gaëtane; R&D UNICANCER,
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Medical management
Keywords:	Real-world data, Data Platform, Epidemiology < ONCOLOGY, Patient medical record, Quality control, Metastatic breast cancer

TITLE PAGE

Cohort profile of the ongoing French Metastatic Breast Cancer (MBC) cohort: the example-based methodology of the Epidemiological Strategy and Medical Economics (ESME)

List of authors:

David Pérol (Centre Léon Bérard, Lyon), Mathieu Robain (R&D UNICANCER, Paris), Patrick Arveux (Centre Georges-François Leclerc, Dijon), Simone Mathoulin-Pelissier (Institut Bergonié, Bordeaux), Emmanuel Chamorey (Centre Antoine Lacassagne, Nice), Bernard Asselain (University Paris Sud), Delphine Berchery (Institut Claudius Régaud, Toulouse), Sophie Gourgou (Institut régional du Cancer Montpellier, Montpellier), Mathias Breton (Centre Eugène Marquis, Rennes), Stéphanie Clisant (Centre Oscar Lambret, Lille), Muriel Mons (Gustave Roussy, Villejuif), Véronique Diéras (Ensemble Hospitalier de l'Institut Curie, Paris & Saint-Cloud), Matthieu Carton (Ensemble Hospitalier de l'Institut Curie, Paris & Saint-Cloud), Anne-Valérie Guizard (Centre François Baclesse, Caen), Lilian Laborde (Institut Paoli-Calmettes, Marseille), Carine Laurent (Institut de Cancérologie de Lorraine, Nancy), Agnès Loeb (Centre Henri Becquerel, Rouen), Marie-Ange Mouret-Reynier (Centre Jean Perrin, Clermont-Ferrand), Damien Parent (Institut Jean Godinot, Reims), Geneviève Perrocheau (Institut de Cancérologie de l'Ouest, Nantes & Angers), Loïc Campion (Institut de Cancérologie de l'Ouest, Nantes & Angers), Michel Velten (Centre Paul Strauss, Strasbourg), Christian Cailliot (R&D UNICANCER, Paris), Monia Ezzalfani (Centre Léon Bérard, Lyon), Gaëtane Simon (R&D UNICANCER, Paris)

Corresponding author:

David Pérol (Centre Léon Bérard, Lyon)

E-mail: david.perol@lyon.unicancer.fr

Tel: +33 (0)4.69.85.61.00

ABSTRACT

Purpose

The currently ongoing Epidemiological Strategy and Medical Economics (ESME) research program aims at centralising real-life data on oncology care for epidemiological research purposes. We draw on results from the metastatic breast cancer (MBC) cohort to illustrate the methodology used for data collection in the ESME research program.

Participants

All consecutive ≥18 years patients with MBC treatment initiated between 2008 and 2014 in one of the 18 French Comprehensive Cancer Centres (FCCCs) were selected. Diagnostic, therapeutic, and follow-up data (demographics, primary tumor, metastatic disease, treatment patterns and vital status) were collected through the course of the disease. Data collection is updated annually.

Finding to date

With a recruitment target of 30,000 MBC patients by 2019, we currently screened a total of 45,329 patients, and over 16,700 patients with a metastatic disease treatment initiated after 2008 have been selected. 20.7% of patients had a HR-negative MBC, 73.7% had a HER2-negative MBC, and 13.9% were classified as triple-negative BC (i.e. HER2 and HR status both negative). Median follow-up duration from MBC diagnosis was 48.55 months for the whole cohort.

Future plans

These real-world data will help standardise the management of MBC and improve patient care. A dozen of ancillary research projects have been conducted and some of them are already accepted for publication or ready to be issued. The ESME research program is expanding to ovarian cancer (OC) and advanced/metastatic lung cancer (AMLC). Our ultimate goal is to achieve a continuous link to the data of the cohort to the French national Health Data System (SNDS) for centralising data on healthcare reimbursement (drugs, medical procedures), inpatient/outpatient stays, and visits in primary/secondary care settings.

KEYWORDS

Real-world data; real-life data; patient medical record; quality control; oncology; metastatic breast cancer

Strengths and limitations

- The epidemiological strategy and medical economics (ESME) research program aims at centralizing real-life data on oncology care for epidemiological research purposes. The ongoing screening of the metastatic breast cancer (MBC) cohort reached over 16,700 patients with metastatic disease treatment initiated after 2008, currently contributing to the development of one of the most important cohort of patients with treated MBC.
- Screening process of patients and data collection (diagnostic, therapeutic, and follow-up data) through the course of the disease provide a solid base of knowledge for real-world survival.
- The significant resources deployed allowed to achieve a high quality level data validation, including systematic consultancy of the source folder for data collection, and the implementation of effective quality control, and regular audit.
- The main limitations are i) the lack of availability of electronic medical records data required to describe the global MBC management, due to the low level of standardization of current electronic medical records, and ii) the retrospective patient selection - data collection, notwithstanding the prospective compilation of real-life follow-up, clinical and biological events, preventing to assess several endpoints classically defined for randomized clinical trials such as progression disease at predefined time points.
- Despite the current large-scale recruitment of patients and greater than one third of French MBC patients managed in FCCCs, future studies should integrate the diversity of management options adopted in any health institutions.

BACKGROUND

Real-world evidence (RWE) studies and observational studies using real world data (RWD) play a growing role in conducting comparative effectiveness research on pharmaceutical products and other healthcare interventions. They aim to bridge the gap between the highly controlled environment of randomised clinical trials (RCTs) and real-life clinical practice [1]. In particular, health authorities are interested in gathering real-world data for long-term benefit-risk assessment and, increasingly, health economic evaluations for reimbursement decisions [2]. With their high internal validity, RCTs are considered the gold standard of evidence for establishing treatment efficacy, although the generalisability of findings to clinical practice may be limited [3]. In fact, cancer survival endpoints in the real-life setting may differ from that measured in traditional RCTs [4]. Furthermore, other limitations such as short follow-up and small sample size have created uncertainty in estimating survival criteria in RCTs, and thus surrogate endpoints are generally used such as progression free survival (PFS) or time to progression (TTP) [5]. On the other hand, large population-based cohorts with longer follow-up periods can be particularly appropriate to assess long-term clinical outcomes , such as overall survival (OS) outside of the RCT setting [6] and to detect changes in medical practice.

RWE and observational studies are non-experimental research where cancer management (treatments and disease evolution assessment) is left to the choice of care providers and patients [7]. Over the past few years, a broad consensus arose around the requirement for high-quality data from large cohorts to strengthen and drive improvements in research methods and practices [8]. The goal of these studies is to generate complementary information to RCTs

based on larger samples and provide answers regarding particular populations in real-life clinical practice. These studies are more prone to biases such as baseline differences between patients (selection bias) or bias due to confounding by indication for example. To minimise sources of such biases, statistical approaches including adjusted analyses, propensity score methods, or instrumental variables may be also employed [9-12].

Metastatic breast cancer (MBC) is one of the leading causes of cancer-related mortality among women in Western countries [13]. A relatively high proportion (approximately 30%) of breast cancer patients develop metastatic disease [14], and while significant treatment advances have been made, the overall prognosis is poor with a five-year survival rate of 25% [15]. The national academic network of cancer centres in France (French Comprehensive Cancer Centres; FCCCs), which together handle over one-third of all breast cancer cases nationally, decided to launch in 2014 a program dedicated focussing on RWD in oncology databases in MBC through Epidemiological Strategy and Medical Economics (ESME) research program.

The ESME research program aims to build a comprehensive database on oncology care for epidemiological research purposes to improve knowledge on medical practice in real-life setting, on public health and healthcare use, and to provide information to health authorities and other associated bodies.

Several studies based on real-life data collection have been developed in this program. Different cohorts of patients with ovarian cancer, patients with advanced/metastatic lung cancer are currently recruiting.

Some results from the ongoing ESME MBC project (ClinicalTrials.gov NCT03275311) whose aim is to describe the medical care of patients treated for MBC according to their disease characteristics the evolution of their metastatic disease and their outcomes have been recently published. Recent analyses explored overall survival in different subgroups of MBC patients, and first line therapy in HER2 negative MBC patients, and in hormone receptor-positive HER2-negative MBC patients [16-18].

In this paper, we describe the methodological principles that underpin the ESME research program, and illustrated this innovative approach through this first ESME metastatic breast cancer (MBC) cohort. Design, brief description of the selected population and current status are reported.

METHODS

The ESME MBC cohort is a population-based registry in 18 FCCCs (http://www.unicancer.fr/en/rd-unicancer/esme), which collected data on all consecutive patients treated for MBC from 2008. Annual data collection phases are planned to add new diagnosed cases and update patients' follow up data.

STUDY POPULATION

Eligibility criteria

Patients eligible to the ESME MBC cohort were male or female patients aged ≥18 years with MBC whose first metastasis had been either completely or partially treated between January 1st, 2008 and December 31st, 2014 in one of the FCCC. MBC treatment could include radiotherapy, chemotherapy, targeted therapy, immunotherapy, or endocrine therapy.

Patient screening process

Patient screening process involved two steps: an automated case screening followed by the validation of selection for each screened case. The automated case screening was based on information retrieved from multiple data sources available within each FCCC: administrative records [French National Computerised Medical Information System] (via MBC-specific International Classification of Diseases (ICD) codes associated with inpatient stays), pharmacy records, patient medical records (PMRs), including multidisciplinary team meeting records and search using relevant keywords, MBC-specific registries.

The objective of the first step -automated screening step- was to identify all cases with inpatient stays or therapeutic management for MBC in one of the FCCC during the selection period and generate the patient screening list. The ICD codes used were: C50 (Malignant neoplasm of breast), C77.- [except C77.3] (Secondary and unspecified malignant neoplasm of lymph nodes), C78.- (Secondary malignant neoplasm of respiratory and digestive organs) and C79.- (Secondary malignant neoplasm of other and unspecified sites). Once the patient screening list was finalised in each centre, data were subsequently anonymised and each patient had been assigned an ESME number. The first screening step did not allow to precisely identify the date for first MBC-specific treatment, and the second step was performed to cross-check eligibility criteria for all screened cases and specify the dates related to the initiation of the MBC first-line treatment, using data from the patient medical records.

ETHICS AND PATIENT DATA PROTECTION

The ESME MBC database received approval from the French data protection authority (*Commission Nationale de l'Informatique et des Libertés,* authorisation no. 1704113)). No informed consent was required.

The ESME research program was managed by R&D Unicancer in accordance with guidelines for Good Pharmacoepidemiology Practices and Good Epidemiology Practices [19, 20].

Ancillary projects analyses were notified to an independent ethics committee (Lyon Sud-Est II) on December 17th, 2015.

Data collection

The data of the selected patients were planned to be collected by trained research assistants and annually updated. The data collection was performed in two phases from October 2014 to October 2016. A first data collection phase conducted in 2014-2015 collected the data from patients with MBC treatment initiated between January 1st, 2008 and December 31st, 2013 in one of the FCCC. A second phase of data collection performed in 2016 added to the ongoing

database the data from patients with MBC treatment initiated between January 1st, and December 31st, 2014, and follow-up data for the global cohort were consequently updated. Hence, the ongoing database provide an overview of all the data from patients with a MBC treatment initiated from 2008; information are updated with last contact information available in the PMR at the date of the data collection.

Baseline and follow-up data

The ESME MBC Cohort is composed of 3 types of data set (**Figure 1**):

- *Patient-related data* are obtained from systematic review of patient medical records (non-structured data) and provide information on patient demographics, cancer family history, characteristics of primary tumour, relapses, metastatic recurrences pathological reports (tumor size, grade, histological type), hormone-receptor status and HER-2 status, therapeutic care (focussing on cancer-related treatment) and settings, reasons for treatment termination, and clinical events.
- *Hospitalisation records* are integrated data from a structured and automated database related to inpatient stays, and primarily used to bill the French National Health Insurance Fund (*Assurance Maladie*). It provides information on patient entry and discharge information (date and destination code at discharge), ICD codes associated to each stay, diagnostic, medical and therapeutic procedures including radiotherapy and surgery.
- *Pharmacy-dispensed treatment records* includes all data related to anti-cancer treatments obtained from each centre's pharmacy database: drugs (International Non-proprietary Name), administration date, protocol name, patient's height and BMI, cycle ID, UCD code (pharmaceutical form related to the drug dosage), line of treatment, administration in a clinical trial (yes/no). It exclusively includes information on products that are prescribed by each FCCC.

The detailed raw data collected above and derived data are listed in **Table 1**.

DATA MANAGEMENT

Any data integrated in the ESME research program are subject to Quality Control procedures.

Patient-related data are registered via an electronic case report form (eCRF) in each centre by trained clinical research assistants (CRAs) between December 2014 and October 2016. Medical support to assist CRAs was provided to ensure quality-controlled data and appropriate recoding was performed before annual database lock when required. All ESME procedures are handled according to the Guidelines for good pharmacoepidemiology practices [20]. Importantly, all data are exclusively obtained retrospectively; no attempts are made to recover unavailable data from patients' medical records by contacting healthcare providers or patients.

The clinical data management system (CDMS) used was Statistical Analysis System (SAS) software. For both, ESME MBC database and the eCRF tool administration used an Oracle solution and certified personal data hosting system guarantee data security.

On-site quality review

On-site quality review (QR) of the patient screening process was carried out. This consisted of checking eligibility criteria on samples of selected and non-selected cases in each FCCC. For selected cases, key variables were crosschecked versus the source data. For all Quality Controls (QC) procedures, accepted error limits were 10% for non-selected patients and 5% for selected patients. A central audit was subsequently performed by the Unicancer Quality Assurance Department, and an audit on data registration and generated screening list was also conducted at the local level.

DATA QUALITY ASSURANCE

Governance structure

Three boards monitor the ESME Research program : Scientific Committee, deontology Committee and International Advisory Board. The main role of the Scientific Committee is to: i) ensure that the applicable scientific rules are followed, ii) evaluate any ancillary projects in compliance with defined criteria and scientific pertinence, and iii) monitor the all validated ancillary projects. The deontology Committee monitors any potential conflicts of interest related to experts involved in the program, gives recommendations that may improve the prevention of conflicts, provides opinions on individual or particular situations, and potential collaborations with private partners. The ESME International Advisory Board has a consultative role with regard to coherence of the scientific program and reviews key international communications, formulates recommendations for publication rules or methodology, and reinforces international academic cooperation.

DATA ANALYSIS PRINCIPLES

Academic research teams or private organisations could propose ancillary projects.

For each accepted ancillary project, statistical analyses are conducted according to a detailed statistical analysis plan (SAP) that must be reviewed by the scientific committee. This article does not aim at providing a comprehensive and exhaustive review on appropriate statistical methods to reduce bias related to analysis based on RWD.

A first ancillary project reported the outcomes (OS and PFS) following first-line paclitaxel treatment with or without bevacizumab [16]. Other ongoing analyses of sensibility will better address the bias potentially found in real-life settings. Two analyses reported the description of OS in different subgroups of MBC patients over the time, and results for the first-line therapy (Endocrine therapy or chemotherapy) in hormone receptor-positive HER2 negative cancer subgroup respectively [17,18]. Other series accepted for communications (abstracts/posters) to major congress in 2017 reported epidemiological analyses (i.e. impact of age at diagnosis, …), therapeutic management (i.e. impact of loco-regional treatment on OS, …) and specific analyses for drug use in routine practice (i.e. use of vinorelbine, everolimus, etoposide…).

Different other sub-populations are currently being considered such as subgroups of metastatic Triple Negative Breast Cancer, metastatic HER2 positive MBC…

PATIENT AND PUBLIC INVOLVEMENT STATEMENT.

Patients and/or public were not involved.

RESULTS AND CURRENT STATUS

16,711 out of the 45,329 patients screeneed were selected in the ESME MBC Cohort (see **Figure 2**).

The sensitivity and specificity of the three main screening sources used were explored. Sensitivity was highest for administrative records (78% *versus* 43% for pharmacy records, and 28% for BC-specific local registries). On the other hand, specificity was highest for BC-specific local registries (87% *versus* 67% for pharmacy records, and 49% for administrative records).

The main reasons for non-selection of screened patients were: presence of non-metastatic breast cancer or other metastatic cancers (n=14,104 patients), initial MBC treatment received before January 1st, 2008 (n=7486), and first metastasis not initially treated in an FCCC (n=4239). Nine additional were excluded prior to the final database lock due to inconsistencies in the dates. A total of 16,702 patients was analysed **(Figure 2).**

Table 2 summarises the main demographic and disease characteristics at the time of initial metastatic diagnosis. Patients were nearly all women (99.1%) with a median age of 61 years. Over half (56.2%) had at least visceral metastases present, with 30.2% having at least bone and non-visceral metastases, and 13.6% with skin only, or node only, or at least skin plus node. 20.7% of patients had a HR-negative MBC, 73.7% had a HER2-negative MBC, and 13.9% were classified as triple-negative BC (i.e. HER2 and HR status both negative).

Median follow-up duration from MBC initiation treatment was 48.55 months for the whole cohort [95% CIs 47.7–49.38].

DISCUSSION

Retrospective analysis using real world data is likely to become increasingly important to ensure that medications are accepted by policymakers and adopted by patient practitioners. The ESME Research program is a large-scale initiative to provide access to real-world data in oncology. This ongoing ESME MBC cohort currently centralises data from 16 711 patients.

The ESME research program provides a unique opportunity to study a diverse range of topics related to MBC care and management in real-life settings. Indeed, there are many potential applications, including study of the factors influencing patient care (e.g. cancer and patient characteristics), description of therapeutic strategies (treatment lines and sequences of therapies, etc.), measurement of clinical events (disease progression, death, persistence of treatment effect). Characterisation of patients enrolled in clinical trials is also possible, as is the reconstruction of "virtual trials" using appropriate statistical methodologies. Potentially, these data could be used for health economics evaluation of management strategies for patients (e.g. rehospitalisation and related ambulatory care), as well as reconstruction of health care trajectories through data modelling.

ESME MBC Cohort

The ESME research program includes alternative approaches to generate cohorts that use different types of RWD (clinical data, therapeutic treatment data, long term outcomes, health economics data) in the FCCCs *vs* existing registries in France, Europe and the US (e.g. SEER). It involves rigorous procedures for patient screening and data collection, ensuring both validity and reliability of data. It uses a fully retrospective approach, with no influence on treatment practice or interaction with oncologists. Unlike prospective interventional or observational research studies, data are not influenced by study design and reflect the real-life management of patients treated. While data recorded for the cohort are defined by experts in the field, the vast majority of data are collected by trained clinical research technicians, thereby minimising any potential risk of data misinterpretation. As discussed above, the ESME MBC Cohort offers a unique opportunity to study a wide range of research questions in a large sample. With respect to evaluation of treatment strategies, the database enables reliable estimation of survival criteria such as overall survival (OS) and surrogates endpoints (PFS …). OS improvement in diseases with a long median post-progression survival time, such as MBC is a critical endpoint [21-23].

The ESME MBC cohort also has several limitations. For example, the database relies on the collection and restructuring of existing data only, i.e. there is no creation of new data. Furthermore, apart from events reported in the patient medical record impacting therapeutic management, adverse effects are not routinely captured. Conceivably, further in-depth analysis of the data could highlight trends such as treatment interruption or discontinuation due to toxicity, which is important from a risk-management perspective. The main potential sources of bias include selection bias, and information bias due to differences in patient monitoring and non-standardised data collection. Selection bias has been taken into account by using rigorous selection procedures across all 18 FCCCs, and the data management plan and quality control program described above have been designed to limit information bias. Nevertheless, due to the retrospective data collection and the fact that it is based on real-life follow-up, clinical and biological events are not evaluated at predefined time points (unlike in RCTs). For example, objective response, historical endpoint in RCTs, could not be assessed retrospectively without a central review of existing imaging as not systematically documented in routine practice. The information collected therefore depends on the frequency of follow-up visits and clinical and radiological exams prescribed by the patient's doctor. As clinical signs are the only means by which disease metastasis can be identified, the number of disease progressions may be underestimated. With respect to the clinical event of death, all deaths are reported in the patient medical records.

With respect to evaluation of treatment strategies, analysis of real-life data poses unique challenges, such as accounting for confounding factors between patient groups, although various statistical approaches can be used to address this, as discussed above [24].

Concerning overall generalisability and applicability (external validity), it should be noted that the cohort centralises data from patients treated in specialised cancer centres only. FCCCs may utilise different clinical practices compared to public hospitals and private institutions, and thus patients from FCCCs may not be truly representative of all French breast cancer patients. Potentially, data extrapolation from all French health care organisations could be developed with the Exhaustive National Health Reimbursement System (SNIIRAM; *Système national d'information inter-régimes de l'Assurance maladie*).

The ESME MBC Cohort aims to collect data for up to 30,000 patients by 2019. As mentioned, future aims might include to continuously link our database to those from other institutions, such as the SNDS database for data on exhaustive healthcare reimbursement, and the INSEE

database to provide vital status updates for patients lost to follow-up. The ESME research program has further expanded to ovarian cancer and advanced/metastatic lung cancer. Real world data from the ESME cohorts should help to provide medical recommendations and ultimately improve patient care.[25]

ACKNOWLEDGEMENTS

We thank the 18 French Comprehensive Cancer Centres for providing the data and each ESME contact for coordinating the project at the local level.

We thank the ESME Scientific Committee members for their ongoing support. We also acknowledge Dr. Sarah Hopwood (Scinopsis, France) for editorial assistance with this publication, paid by R&D UNICANCER.

The 18 French Comprehensive Cancer Centres (FCCC) are: Ensemble hospitalier de l'Institut Curie, Paris/Saint-Cloud, Gustave Roussy, Villejuif, Institut de Cancérologie de l'Ouest, Nantes/Angers, Centre François Baclesse, Caen, Institut du Cancer de Montpellier Montpellier, Centre Léon Bérard, Lyon, Centre Georges-François Leclerc, Dijon, Centre Henri Becquerel, Rouen; Institut Claudius Régaud, Toulouse; Centre Antoine Lacassagne, Nice; Institut de Cancérologie de Lorraine, Nancy; Centre Eugène Marquis, Rennes; Institut Paoli-Calmettes, Marseille; Centre Jean Perrin, Clermont-Ferrand; Institut Bergonié, Bordeaux; Centre Paul Strauss, Strasbourg; Institut Jean Godinot, Reims; Centre Oscar Lambret, Lille.

ESME central coordinating staff: Head of Research and Development: Christian Cailliot. Program director: Mathieu Robain. Data Managers: Irwin Piot and Olivier Payen. Project Managers: Coralie Courtinard, Tahar Guesmia and Gaëtane Simon. Project assistant: Esméralda Pereira. Software designer: Alexandre Vanni. ESME local coordinating staff: Patrick Arveux, Thomas Bachelot, Jean-Pierre Bleuse, Delphine Berchery, Mathias Breton, Stéphanie Clisant, Emmanuel Chamorey, Valérie Dejean, Véronique Diéras, Anne-Valérie Guizard, Anne Jaffré, Lilian Laborde, Agnès Loeb, Muriel Mons, Damien Parent, Geneviève Perrocheau, Marie-Ange Mouret-Reynier, Carine Laurent, Michel Velten.

Funding: This work was supported by R&D UNICANCER and initially funded by an industrial consortium (Roche, Pierre Fabre, Pfizer, and Astrazeneca), and support was subsequently provided by MSD, Eisai, and Daiichi Sankyo. The funders did not have any input into maintenance of the database (project design, data collection, management) or analysis of data. They were not involved in the writing of the manuscript or decision to submit for publication.

Contributors: DP Conception and design,; MR Conception and design, writing the first draft; PA Conception and design, data acquisition SM-P Conception and design, interpretation; EC Data acquisition, statistical analysis; BA Statistical analysis; DB Conception and design, data acquisition; SG Statistical analysis; MB, SC, MM, VD Data acquisition; MC Statistical analysis; A-VG, M-AM-R Conception and design, data acquisition; LL, CL, AL, DPa, GP, Data acquisition; LC data acquisition, statistical analysis; MV Conception and design, data acquisition, statistical analysis; CC Conception and design; ME Statistical analysis, writing the first draft; GS Conception and design, statistical analysis, writing the first draft. All authors reviewed and approved the final version of the manuscript.

Data Sharing:

We reported the methodology developed to collect and control the data of the large ESME program and illustrated the methodology with data from the cohort of patients with metastatic breast cancer patients.

Data collected are listed in the Table 1. The database of the ESME program or the database of the MBC cohorts are currently not accessible.

For any specific demand, please contact the corresponding author. Each demand will be examined on a case-by-case basis by the scientific committee.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest: D Perol has received personal fees (honoraria and travel/accommodation expenses) from Laboratoire Roche, outside the submitted work. B Asselain has received personal fees for board membership and for consultancy from Roche Pharma, outside the submitted work. S Gourgou has received personal fees for board membership from Celgene and for consultancy from Roche, outside the submitted work. All other authors declare that they have no conflict of interest.

Ethical approval: Approval for the ESME MBC database was obtained from an independent ethics committee, which waived the requirement for informed consent.

REFERENCES

1. Chavez-MacGregor M, Giordano SH. Randomized Clinical Trials and Observational Studies: Is There a Battle? J Clin Oncol. 2016;34(8):772-3.

2. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part I. Value Health. 2009;12(8):1044-52..

3. Hahn OM, Schilsky RL. Randomized controlled trials and comparative effectiveness research. J Clin Oncol. 2012;30(34):4194-201.

4. Meyerhardt JA, Li L, Sanoff HK, Carpenter Wt, Schrag D. Effectiveness of bevacizumab with first-line combination chemotherapy for Medicare patients with stage IV colorectal cancer. J Clin Oncol. 2012;30(6):608-15.

5. Prasad V, Kim C, Burotto M, Vandross A. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-level meta-analyses. JAMA Intern Med. 2015;175(8):1389-98.

6. Visvanathan K, Levit LA, Raghavan D, Hudis CA, Wong S, Dueck A et al. Untapped Potential of Observational Research to Inform Clinical Decision Making: American Society of Clinical Oncology Research Statement. J Clin Oncol. 2017;35(16):1845-54.

ESME MBC Cohort

7. Berger ML, Dreyer N, Anderson F, Towse A, Sedrakyan A, Normand SL. Prospective observational studies to assess comparative effectiveness: the ISPOR good research practices task force report. Value Health. 2012;15(2):217-30.

8. Morton SC, Costlow MR, Graff JS, Dubois RW. Standards and guidelines for observational studies: quality is in the eye of the beholder. J Clin Epidemiol. 2016;71:3-10.

9. Rosenbaum PR, Rubin DB. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. The American Statistician. 1985;39(1):33-8.

10. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. JAMA. 2007;297(3):278-85.

11. D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998;17(19):2265-81.

12. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. Annu Rev Public Health. 1998;19:17-34.

13. Institut National du Cancer (INCa). Information sur le cancer. http://www.e-cancer.fr/. Accessed 04 Jul 2017.

14. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005;365(9472):1687-717.

15. Binder-Foucard F, Belot A, Delafosse P, Remontet L, Woronoff AS, Bossard N. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012. Partie 1 – Tumeurs solides. Saint-Maurice (Fra): Institut de veille sanitaire, 2013.

16. Delaloge S, Perol D, Courtinard C, Brain E, Asselain B, Bachelot T et al. Paclitaxel plus bevacizumab or paclitaxel as first-line treatment for HER2-negative metastatic breast cancer in a multicenter national observational study. Ann Oncol. 2016;27(9):1725-32.

17. Gobbini E, Ezzalfani M, Dieras V, Bachelot T, Brain E, et al. Time trends of overall survival among metastatic breast cancer patients in the real life ESME Cohort . Eur J Cancer. Eur J Cancer. 2018 Jun;96:17-24.

18. Jacquet E, Lardy-Cléaud A, Pistilli B, Frank S, Cottu P, Delaloge S, et al. Endocrine therapy or chemotherapy as first-line therapy in hormone receptor-positive HER2 negative metastatic breast cancer patients. Eur J Cancer. Eur J Cancer. 2018 May;95:93-101.

19. Association des Epidémiologistes de Langue Française (ADELF). Déontologie et bonnes pratiques en épidémiologie, recommendations révisées, 2007.

20. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiol Drug Saf. 2008;17(2):200-8.

21. Largillier R, Ferrero JM, Doyen J, Barriere J, Namer M, Mari V et al. Prognostic factors in 1,038 women with metastatic breast cancer. Ann Oncol. 2008;19(12):2012-9.

22. Asselain B, De Rycke Y, Savignon A, Mould RF. Parametric modelling to predict survival time to first recurrence for breast cancer. Phys Med Biol. 2003;48(12):L31-3.

23. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. J Natl Cancer Inst. 2009;101(23):1642-9.

24. Perol D, Robain M, Delaloge S, Cailliot C. Routinely collected data may usefully supplement randomised controlled data on treatment effects for mortality. BMJ. 2016;355:i6745.

25. Berdaï D, Thomas-Delecourt F, Szwarcensztein K, d'Andon A, Collignon C, Comet D. Requests for post-registration studies (PRS), patients follow-up in actual practice: Changes in the role of databases. Therapie. 2018;73:13-24.

ESME MBC Cohort

TABLES

Table 1 Main data recorded

Patient-related	Hospitalisation-related	Pharmacy record-related
Patient data:	Age	Height
Demographics	Gender	Body Mass Index
Other cancer and family history	Main diagnosis code (ICD code)	Drug (International Non-proprietary Name)
Menopausal status	Linked diagnosis code (ICD code)	Protocol
Initial tumor:	Related significant diagnoses (ICD code)	Cycle ID
Diagnosis	Diagnosis-related group code (reimbursement coding)	Pharmaceutical form and dosage
Relapses	Hospital stay-related group ID (reimbursement coding)	Administration date
Histological results	Entry date	Line treatment number
Medical care	Discharge date	Administered dose
Metastatic disease:	Destination code at discharge	Cumulated dose
Diagnosis	Medical procedures performed	Inclusion in a clinical trial (Yes/No)
Progression	Radiation therapy	
Histological results		
Therapies		
Invasive procedures related to metastasis		
Last contact:		
Vital status		
Date of last contact/death		

ICD International Classification of Diseases

ESME MBC Cohort

Table 2 Characteristics of the patients at of the date of metastatic disease diagnosis.

Characteristic		ESME MBC Population (N=16 702)
Age (years)	Mean (SD)	60.6 (13.8)
	Median [Q1-Q3 range]	61.0 [51.0-71.0]
Sex	Male	149 (0.9%)
	Female	16553 (99.1%)
Histological grade* at primary tumour diagnosis	1	1277 (10.1%)
	2	6438 (50.7%)
	3	4733 (37.3%)
	Not available	240 (1.9%)
	Missing data	2008
Metastatic status**	*De novo* MBC	4763 (28.5%)
	Relapsed MBC	11939 (71.5%)
Year of first metastatic treatment	2008	2651 (15.9%)
	2009	2675 (16.0%)
	2010	2598 (15.6%)
	2011	2515 (15.1%)
	2012	2371 (14.2%)
	2013	2216 (13.3%)
	2014	1676 (10.0%)
Type of metastases	Visceral	9383 (56.2%)
	Bones and not visceral	5047 (30.2%)
	Nodes only	880 (5.3%)
	Skin only	916 (5.5%)
	Skin + nodes	476 (2.8%)
Global HR status***	Positive	12748 (76.3%)
	Negative	3451 (20.7%)
	Not determined	503 (3.0%)
Global HER2 status***	Positive	2863 (17.1%)
	Negative	12306 (73.7%)
	Not determined	1533 (9.2%)
Triple negative status	Yes	2321 (13.9%)

Data are n (%) unless indicated otherwise

*Histological grade at primary tumour diagnosis: The histological grade at primary tumour diagnosis is defined as the worst histological grade recorded within one month (30 days) after the initial diagnosis (primary tumour).

***de novo* MBC: Metastatic breast cancer is considered *de novo* if the diagnosis of metastatic disease occurs within 6 months (180 days) after the initial diagnosis (primary tumour).

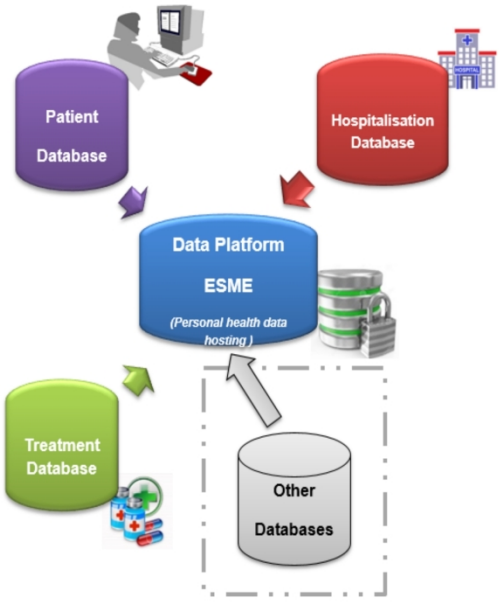
***HR and HER2 status: The estrogen receptor (ER) or progesterone receptor status (PR) is considered positive if the pathology report indicates a "positive" result, or considered as

positive when ≥10% of cells in the sample are positive for ER, or PR respectively. The HER2 status is considered positive if the pathology report indicates for the immunohistochemistry (IHC) result "3+", "2+" or not available, the result will be considered positive if the Fluorescent In Situ Hybridization (FISH) or Chromogenic In Situ Hybridization (CISH) result is positive. If two or more histologic reports are available at the same date, the positive status is preferred. The global HR/HER2 status indicates the status at metastatic diagnosis based on histological results forms related to the primary tumour (if available) or metastatic sites.

FIGURES

Figure 1. The ESME Data platform

Figure 2. ESME MBC Cohort: Flow chart



Hospitalisation Database

- Hospitalisations: dates, diagnoses, GHS code
- Medical procedures (Inc. Radiotherapy): dates, code

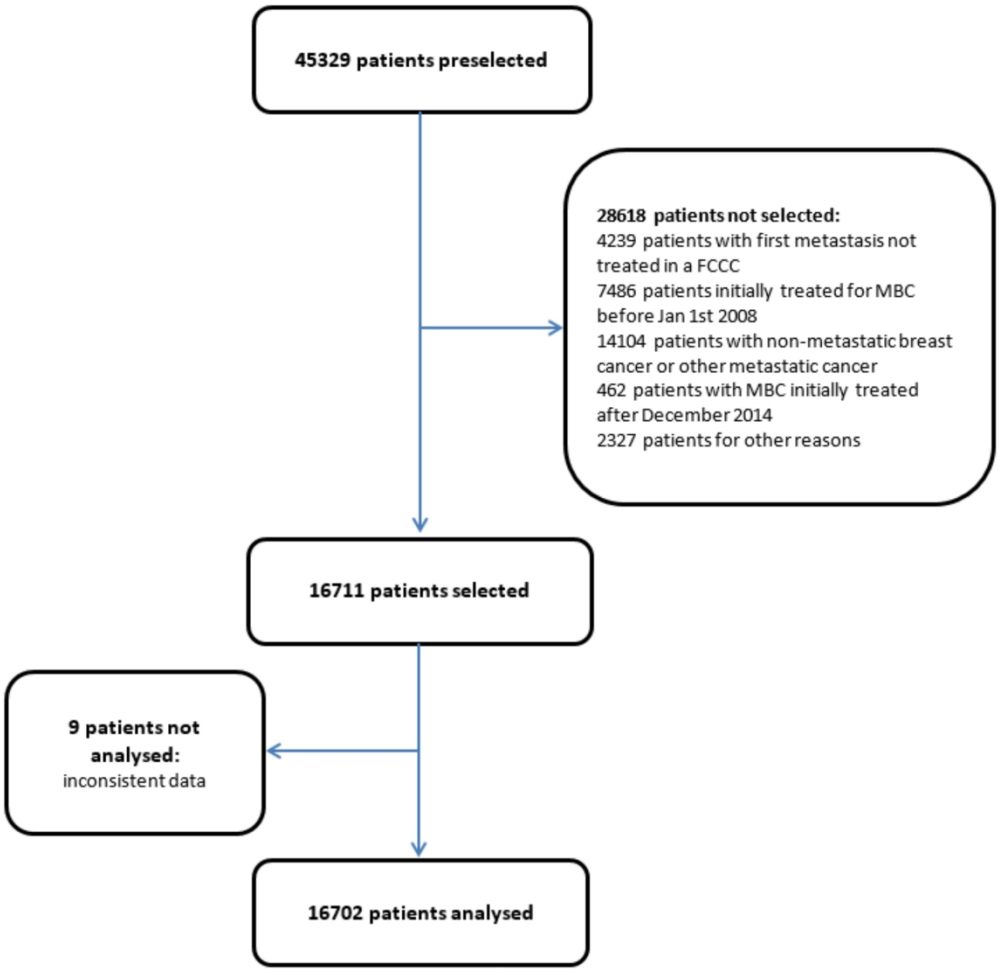
Treatment Database

- Pharmacy records: dates, antineoplastic drugs, therapeutic protocol and other concomitant drugs

Patient Database

- Collection based on Electronic medical records
- Patient data : demographics, cancer management, clinical events (progression, relapse), pathological report, metastatic disease, anti-cancer treatment (chemotherapy, targeted therapy, endocrine therapy, immunotherapy), and other therapeutic care (radiotherapy, surgery) or supportive care

The ESME Data platform



ESME MBC Cohort: Flow chart