concePTION
SAFETY EVIDENCE ECOSYSTEM

Protocol
Version 1.0

DP5. Studying drug exposure when disease is measured through accurate identification of an incident case: application to breast cancer in pregnancy

This study will be conducted within the ConcePTION project of the Innovative Medicines Initiative under grant agreement No 821520

Title	Studying drug exposure when disease is measured through accurate identification of an incident case: application to breast cancer and pregnancy
Protocol version identifier	1.0
Date of last version of protocol	09/30/2021
EU PAS register number	Study will be registered in the EU PAS register.
Active substance	Antineoplastic agents (ATC class L01) and endocrine therapy (ATC class L02).
Medicinal product	

Research question and objectives	The objective is to evaluate which pharmaco-epidemiological methods are best suited to assess treatment modalities, including drug utilisation in pregnancy for malignant disease. The goal is to study drug exposure when disease is measured through accurate identification of an incident case. Therapies for pregnancy associated breast cancer (PABC) are used as motivating examples. We will particularly focus on improving methods for developing measurements of medication exposure in hospital settings / secondary and tertiary care. This drug utilisation study will describe patterns of medication use in PABC and in breast cancer in non-pregnant women (non-PABC). We will also assess whether time at breast cancer diagnosis and timing of medication use in pregnancy impacts maternal survival and pregnancy outcomes.
Countries in the study	Finland, Spain (Valencian Region), UK (Wales), Germany, Scotland, and possibly other countries, all pending on results from the Data characterization (WP7).
Primary Authors	Dr Maarit Leinonen and Prof Mika Gissler
Contributors	Janssen Research and Development Dina Gifkins NIHW (THL) Finnish Institute for Health and Welfare, NIHW Maarit Leinonen, Mika Gissler, Visa Martikainen DAP Valencian Region (Spain): Rare Disease Research Unit, FISABIO (The Foundation for the Promotion of Health and Biomedical Research of Valencian Region). Clara Cavero-Carbonell, Laia Barrachina-Bonet, Laura García-Villodre and Óscar Zurriaga BIPS, Leibniz Institute for Prevention Research and Epidemiology Tania Schink, Miriam Heinig, Katja Oppelt, Ulrike Haug USWAN, University of Swansea Sue Jordan, Daniel Thayer

List of abbreviations

Abbrev.	Term
ATC	anatomical therapeutic chemical
BMI	Body mass index
CDM	Common Data Model
CI	Confidence intervals
DAP	Data access provider
DDD	Defined daily dose
EMA	European Medicines Agency
ETL	Extract, transform, load
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Reusable
FISABIO	The Foundation for the Promotion of Health and Biomedical Research of Valencian Region
GA	gestational age
GePARD	The German Pharmacoepidemiological Research Database
GLST	Generalized Least Squares for trend
GVP	Guideline on good pharmacovigilance practices
ICD-O-3	International Classification of Diseases for Oncology
IUGR	Intrauterine Growth Retardation
HR	Hazard ratio
MEGLM	Multilevel mixed-effects generalized linear model
MI	Multiple imputation
MICE	Multiple imputation by chained equations
PABC	Pregnancy Associated Breast Cancer
PASS	Post-authorisation safety study
SAILS	Secure Anonymised Information Linkage
SERMS	Selective oestrogen receptor modulators
SES	Socioeconomic status
SGA	small-for-gestational age
STROBE	Strengthening the reporting of observational studies in epidemiology
TNM	TNM classification of malignant tumours
WHO	World Health Organization

1. Responsible parties

Responsible parties are:

Name	Role	Working institution/company
Maarit Leinonen **Mika Gissler** **Visa Martikainen**	PI, Protocol development lead, DAP Finland Data analysis	Finnish Institute for Health and Welfare (THL), Finland
Dina Gifkins	Protocol development, co-lead	Janssen
WP7 representative	SAP review, R Coding and analytics of SAP	UMCU
Joan Morris	Statistical expert, meta-analysis	St George's University of London, UK
Sue Jordan, Daniel Thayer	DAP Wales	University of Swansea, Wales
Tania Schink	DAP Germany	Leibniz Institute for Prevention Research and Epidemiology
Clara Cavero-Carbonell **Laia Barrachina-Bonet** **Laura García-Villodre** **Óscar Zurriaga**	DAP Spain (Valencian Region)	Rare Disease Research Unit, FISABIO: The Foundation for the Promotion of Health and Biomedical Research of Valencian Region
Tom MacDonald*	DAP Scotland	University of Dundee Scotland

*protocol review lacking

2. Abstract

Title
Studying drug exposure when disease is measured through accurate identification of an incident case: application to breast cancer and pregnancy

Main authors:
Dr Maarit K Leinonen, Finnish Institute for Health and Welfare, Finland
Prof Mika Gissler, Finnish Institute for Health and Welfare, Finland

Rationale and background
Pregnancy-associated breast cancer (PABC) is usually defined as breast cancer diagnosed during pregnancy or within one year after delivery. It is a rare disease, but incidence is increasing as women are postponing childbearing. Several studies have reported poorer prognosis of PABC compared with women diagnosed with breast cancer prior to pregnancy or 1-2 years after delivery. It is uncertain to what extent this is attributed to less aggressive or delayed therapy secondary to concerns regarding foetal effects, delays in diagnosis and later stage at diagnosis due to difficulties of diagnosing PABC, or underlying immune-suppression in pregnancy, or underlying tumour characteristics. Women with history of breast cancer are recommended to wait at least two years from remission prior to conceiving. Women of childbearing age with breast cancer face unique challenges, such as reassurance that maternal outcome is not adversely affected by pregnancy and possible teratogenicity of cancer therapy exposure in-utero. Management guidelines are based on case reports, case series and small cohorts, and limited by the retrospective nature and heterogenous treatment regimens. Most studies have been hampered by low power and inability to control for tumour characteristics due to missing data.

Research question and objectives

1. What is the incidence of PABC and non-PABC in women of reproductive age in European countries?
2. Is the pattern of cancer treatment (including surgery, radiation, chemotherapy, endocrine therapy and targeted therapy) similar in PABC and non-PABC patients?
3. What medications are used and how does the pattern of treatment change over the course of a pregnancy (e.g. prior to pregnancy, during, after pregnancy), by cancer severity, by country, and time period?
4. What is the pregnancy outcome (termination of pregnancy, live birth, stillbirth, preterm birth, small for gestational age (SGA)/ intrauterine growth retardation (IUGR), as available) and mode of delivery for women with PABC and non-PABC and does that vary by cancer severity, by country, by time period?
5. What is the 5-year relative survival for women with PABC and non-PABC when adjusted for tumour characteristics and is there a difference in survival between PABC patients diagnosed during pregnancy and those diagnosed one year postpartum?

Study design
A multinational cohort study conducted in following countries: Finland, UK (Wales), Spain (Valencian Region), Germany, Scotland and possibly others, all pending on results from the Data characterization (WP7).

Study population
All women free of an identifiable cancer diagnosis prior to age 15 diagnosed with breast cancer during reproductive age. Women diagnosed with breast cancer during pregnancy or within 12 months after childbirth (live births and stillbirths) will be defined as PABC. Women who have no indication within the data source of a pregnancy at diagnosis or diagnosed > 1 year after delivery will be defined as non-pregnant breast cancer (non-PABC).

Study period

Study period will start from the first year cancer incidence and birth outcomes are available from the data source (whichever is the latest) and will end at the most recent date of the data source where maternal death is available. Analyses will be stratified by age groups where appropriate.

Variables

Exposure: The exposure will be defined based on maternal record of one or more prescriptions or procedures (prescribed, dispensed, reimbursed or administered in hospital) of medications or administration of medication for breast cancer in pregnancy. Medications for breast cancer will be classified according to the Anatomical Therapeutic Chemical (ATC) classification system and the primary exposure is antineoplastic agents (ATC class L01) and endocrine therapy (ATC class L02).

Event: PABC

Maternal outcome: 5-year relative survival

Secondary outcomes: mode of delivery and the following pregnancy outcomes (termination of pregnancy, live birth, stillbirth, preterm birth, SGA/ IUGR, congenital anomalies) in women with PABC and non-PABC.

Study size

Assuming 25% of annual breast cancer cases occur in women of childbearing age and the ratio of PABC vs. non-PABC to be 1 to 6, to detect a 40% increased risk for death in PABC vs. non-PABC patients with 80% power and type I error rate of 0.05, we would require 2 965 breast cancer cases and a sample size of around 3.27 million women years in women of childbearing age.

Data analysis Milestones

An important methodological focus will be on exposure misclassification as well as co-exposure effects.

3. Amendments and updates

Number	Date	Section of study protocol	Amendment or update	Reason
1	Date	Text	Text	Text
2	Date	Text	Text	Text
…	Date	Text	Text	Text

4. Milestones

Milestone	Planned date
Registration in the EU PAS register	Date: 1st October 2021
Final report of study results	Date: March 2023

5. Rationale and background

Malignant disease occurring during pregnancy brings particular challenges for diagnostics and treatment. Changes in the breast associated with hormonal changes in pregnancy and lactation make detection and evaluation of breast masses difficult both clinically and radiologically (Ahn 2003, Kakoulidis 2015, Expert Panel on Breast Imaging 2018). More importantly, ionizing radiation used in diagnostics and treatment may lead to death of the embryo and during organogenesis, ionizing radiation exceeding a dose of 50 mGy may cause congenital anomalies, microcephaly, intrauterine growth retardation, and severe cognitive impairment (Navrozoglou 2008, Kakoulidis 2015, MacDonad 2020). Only radiographic studies which will affect management during the course of the pregnancy should be performed and radiotherapy should be postponed after delivery (Navrozoglou 2008, Amant 2012, Cardonick 2014).

Pregnancy-associated breast cancer (PABC) is a very rare type of cancer. Most data on the prognosis of patients with PABC are based on case reports, case series and small cohorts and limited by the retrospective nature and heterogenous treatment regimens (Zagouri 2013, Rojas 2019). Most studies have been hampered by low power and inability to control for tumour characteristics due to missing data. Also, lack of data on medications administered in hospitals is a common limitation encountered in pharmacoepidemiolocial studies.

Another clinical challenge is advising women who have previously been diagnosed and treated for breast cancer and who subsequently desire to become pregnant. These women should not be discouraged from future pregnancy (Rojas 2019). Many clinical guidelines have advised that premenopausal women with breast cancer diagnosis should wait two years after treatment before attempting conception. This can mean over 10 years waiting for women receiving endocrine therapy with a subsequent significant reduction in woman´s ability to conceive (Navrozoglou 2008, Gerstl 2018).

Medication safety

Chemotherapy
Chemotherapy plays a key role in improving the survival of patients with early stage breast cancer (Zagouri 2013). During pregnancy, several physiological changes alter the pharmacokinetics of many medicines, including chemotherapeutic agents. For instance, expanded plasma volume, increased renal clearance and increased activity of liver enzymes may affect free-drug levels and raise doubt about effectiveness of chemotherapy during pregnancy (Navrozoglou 2008, Amant 2012, Cardonick 2014). However, since no evidence suggests that standard treatment in PABC is less efficient than in non-PABC, the chemotherapeutic agents should be prescribed for pregnant patients as for non-pregnant breast cancer patients (Amant 2012, Zagouri 2013, Hartman 2016, Rojas 2019, MacDonald 2020).

First trimester chemotherapy is not recommended since there is substantial risk of spontaneous abortions, teratogenesis and foetal anomalies (Navrozoglou 2008, Amant 2012, Basta 2015). Exposure during the second and third trimester (after 14 weeks) has not been associated with teratogenic effects, i.e. the rate of anomalies mirrors the baseline population´s risk of congenital anomalies (Hartman 2016, Rojas 2019). However, growth restriction, prematurity, intrauterine and neonatal death and haemopoietic suppression and sepsis has been reported following cytotoxic treatment in the second and third trimesters (Amant 2012, Zagouri 2012, MacDonald 2020). Chemotherapy must cease approximately three weeks prior to labour affording both the mother and the foetus the necessary period to excrete the drugs and recover from myelosuppression and, thus, avoiding postpartum infection and/or haemorrhage (Navrozoglou 2008, Amant 2012, Zagouri 2013, Cardonick 2014, Rojas 2019, MacDonald 2020). The main risks related to chemotherapeutic agents are summarized in the Table 1.

Table 1. The main risks to the foetus of chemotherapeutic agents

Examples of chemotherapeutic agents	The main risks to the foetus
Plant alkaloids: vincristine, vinblastine, vinorelbine	Preterm delivery, intrauterine growth restriction (IUGR) [67]. Defect in the atrial septum [68].
Anthracycline antibiotics: doxorubicin, daunorubicin, adriamycin, idarubicin, epirubicin, dactinomycine, bleomycin, mitoxantrone	Mid-trimester miscarriage, transient neonatal neutropenia, and sepsis, IUGR [69]. Transient myelosuppression [70].
Alkylating agents: cyclophosphamide, busulfan, ifosphamide, chlorambucil, carmustine, dacarbazine	Absent toes, eye abnormalities, low-set ears, and cleft palate [71]. Oesophageal atresia, abnormal inferior vena cava [72]. Pyloric stenosis, renal agenesis, and liver calcifications [73].
Antimetabolites: Methotrexate, 5-fluorouracil, aminopterin, cytarabine, mercaptopurine.	Spontaneous abortions [68, 74]. Ventriculomegaly, microcephaly, syndactyly, deficient growth and development [75–77].
Cisplatin and carboplatin	Sensorineural hearing loss, respiratory distress syndrome [10, 78].
Trastuzumab	Kidney injury [79] kidney perfusion [80]. Respiratory failure, capillary fragility, and neonatal death [33].

Table 1. The main risks related to chemotherapeutic agents. Figure from Basta 2015.

Endocrine therapy

During pregnancy, hormonal agents such as selective oestrogen receptor modulators, SERMs can disturb the hormonal environment. Of these, tamoxifen (ATC L02BA01) and its metabolite interact with embryonic or fetal tissues, is teratogenic and may lead to severe foetal anomalies. Studies have reported a foetal malformation rate of up to 20%, including craniofacial malformations and ambiguous genitalia. Therefore, it is recommended that endocrine therapy will be delayed until after birth (Navrozoglou 2008, Amant 2012, Zagouri 2013). Oral aromatase inhibitors (ATC L02BG) are not indicated in premenopausal women (Amant 2012, Zagouri 2013). However, as the upper age limit for this investigation is 55, we expect to identify these prescriptions in non-pregnant women.

Targeted therapy

Treatment with trastuzumab (L01XC03) in Her-2-positive breast cancer tumours is contraindicated during pregnancy. HER is strongly expressed in the foetal renal epithelium and exposure to trastuzumab has been associated with renal failure, reduced amniotic fluid, foetal limb anomalies, pulmonary hypoplasia and death (Amant 2012, Zagouri 2013, Rojas 2019, MacDonald 2020). The risk of oligo- and/or an-hydramnios seems to be attributed particularly to exposure after the first trimester (Zagouri 2013). Also the long-term sequelae for the foetus are unknown. HER2-targeted treatment may be discussed in special high risk situations, and if the patient conceives while taking trastuzumab, exposure is not considered an indication for termination of pregnancy (Cardonick 2014, Rojas 2019). No studies exist, yet, demonstrating safety for the use of pertuzumab (L01XC13) in pregnancy (Rojas 2019, MacDonald 2020). There are insufficient data on lapatinib (L01EH01) and bevacizumab (L01XC07) during pregnancy and their use cannot, thus, be recommended (Zagouri 2013).

This demonstration project offers solutions on how to study medication exposure in diseases where accurate diagnosis requires histopathological examination and classification and, due to course of the disease, incidence is a better indicator defining a patient than prevalence. We will focus on improving methods for developing measurements of medication exposure in hospital settings. Cancer registries often lack data on cancer therapies especially those administered at outpatient visits such as chemotherapeutic agents and radiation (Beatty 2011, Caldarella 2012, Gurney 2013, Mallin 2013).

6. Event: Pregnancy-associated breast cancer (PABC)

Pregnancy has a dual effect on breast cancer risk. Full-term pregnancies in early life (below age 30) have consistently been associated with a long-term reduced risk of breast cancer while a transient increased risk immediately after the pregnancy has been observed. A study including 2.3 million Danish women and 1.6 million Norwegian women observed the reduction in breast cancer risk in pregnancies lasting 34 gestational weeks or longer (Husby 2018).

Pregnancy-associated breast cancer (PABC) is generally defined as breast cancer diagnosed during pregnancy or within one year after childbirth (Ahn 2003, Amant 2012, Hartman 2016). Although about 80 % of breast masses that develop during pregnancy and breastfeeding are benign (Ahn 2003, Amant 2012, Expert Panel 2018), up to 3.8% of all breast cancers occur in pregnant and breastfeeding women (Vinatier 2009).

PABC is a very rare type of cancer. However, given PABC occurs in 1/10 000 to 1/3000 pregnancies, it is one of the most commonly diagnosed cancer during pregnancy (Amant 2012, Expert Panel on Breast Imaging 2018). As women delay childbearing into the fourth and fifth decades, PABC is more frequently encountered by oncologists, gynecologists, and obstetricians.

Diagnosis

Ultrasound has the highest, up to 100%, sensitivity for the diagnosis of PABC (Expert Panel 2018, MacDonald 2020). Therefore, breast ultrasound is considered as the first imaging modality for the evaluation of breast lumps during pregnancy and lactation (Ahn 2003, Navrozoglou 2008, Kakoulidis 2015, Expert Panel 2018). Imaging both breasts is important as the incidence of bilateral disease may be as high as 10% (Kakoulidis 2015). If the breast ultrasound is negative or there are suspicious sonographic findings, additional imaging with mammography or digital breast tomosynthesis may be indicated (Expert Panel 2018).

Knowledge of the tumor subtype and grade is crucial for the management and treatment of breast cancer. Biopsies, preferably ultrasound-guided core needle biopsy, can be safely performed for suspicious masses at any gestational age. This allows precise diagnosis and avoids surgical biopsy but pathologists should be alerted to the pregnant or lactational state of breast (Cardonick 2014 Kakoulidis 2015). Core needle biopsy also allows for evaluation of hormone receptor expression, a known predictive biomarker, by immunohistochemistry.

Breast cancer types

As this demonstration project will use cancer registries as the main data source, algorithms will be based on ICD-O-3 (the latest available version), topography C50 and any ICD-O-3 morphology except those for leukemia, lymphoma and Kaposi's sarcoma i.e. morphology <9590 excluding 9140. To identify breast cancer diagnosis and/or treatment episodes from patient registry data, also ICD-10 (C50.xx) and ICD-9 (174.XX) codes can be relevant.

As in non-pregnant women, the most prevalent histological type of PABC is invasive ductal carcinoma (70-90%) followed by invasive lobular carcinoma and inflammatory carcinoma (Ahn 2003, Navrozoglou 2008, Cardonick 2014) and with all breast cancer subtypes represented (MacDonald 2020). To avoid potential effect of breast cancer screening, no precursor lesions but only malignant primary tumors (behavior = 3) of the breast according to the WHO 2012 classification will be considered (Lakhani 2012, Appendix II).

Breast cancer severity

Breast cancer severity and prognosis is determined by tumour biology and stage. TNM staging system, where T stands for tumour, N for node and M for metastasis, is the most common way to evaluate the extent of

disease i.e. stage breast cancers (Sobin 2009, Appendix III). Other prognostic factors include oestrogen receptor (ER) and progesterone receptor (PR) status, human epidermal growth factor receptor 2 (HER2) status and the grade of the cancer. In order of increasing aggressiveness and worse prognosis, the subtypes are Luminal A (ER+ and/or PR+, HER2-), Luminal B (ER+ and/or PR+, HER2+) and triple negative (ER-, PR-, HER2-) breast cancer (Akinyemiju 2015). PABC is more likely to be ER and PR negative and HER2-positive compared to age-matched controls (Johansson 2019, MacDonald 2020).

Changes in breast density during pregnancy and lactation together with absence of screening at young age, contributes to delays in diagnosis, later stage at recognition and consequently poor prognosis in PABC (Ahn 2003, Kakoulidis 2015, Expert Panel on Breast Imaging 2018). Diagnostic delays up to 7 months have been documented for pregnant patients (Navrozoglou 2008, Cardonick 2014, Kakoulidis 2015). One month delay may increase the risk of nodal involvement by 0.9 % to 1.8 % while a 6-month delay increases the risk by 5.1% (Kakoulidis 2015). PABC tumours are larger and there is higher incidence of nodal involvement than in non-PABC (Ahn 2003, Amant 2012, MacDonald 2020). A Swedish study extracting medical records of 273 PABC cases and 273 age- and hospital-matched non-PABC controls found no evidence of delayed diagnosis or treatment in women with PABC following the first health care contact (Johansson 2019). However, there was an indication of longer time between symptoms to first health care contact in woman with PABC (35 days) than in controls (23 days) although the difference was not significant.

Reliable and safe staging is necessary to choose between starting with local or systemic therapy (Rojas 2019). After a definitive diagnosis, mammography (with abdominal shielding) of the unbiopsied breast is recommended to exclude contralateral disease. Because PABC is often diagnosed at an advanced stage, systemic staging is often necessary. Lungs, bone and liver are the most common metastatic sites. To exclude pulmonary metastasis, chest radiography (with abdominal shielding) can be carried out safely during pregnancy. A liver ultrasound is the preferred method to detect liver metastases. In stage I and II disease the incidence of bone metastasis is low and bone scanning or magnetic resonance imaging (MRI) can then be delayed until postpartum (Cardonick 2014, Kakoulidis 2015).

MRI is useful in staging of breast cancer in non-pregnant women and also in assessing the response to neoadvujant therapy (Kakoulidis 2015). MRI requires gadolinium which crosses the placenta and enters fetal circulation and amniotic fluid. Gadolinium exposure is associated with foetal anomalies; therefore, routine use of MRI is not currently recommended for the evaluation and treatment of PABC (Cardonick 2014, Kakoulidis 2015, Expert Panel 2018, Rojas 2019). Furthermore, the prone positioning necessary for breast MRI may lead to prolonged pressure on the gravid uterus, disrupting uterine blood flow (Rojas 2019). In one study including 53 PABC patients, preoperative MRI had large impact on clinical management as it changed surgical management for 28 % patients. Also, MRI showed pathologically proven larger tumor size or multicentric disease and greater extent of disease than mammography or ultrasound (Myers 2017). Decision regarding use should be made on individual basis weighing risks and benefits (MacDonald 2020). Breast MRI performed with iv gadolinium is safe during lactation since less than 0.04% of the administered gadolinium will be excreted into the breast mill (Myers 2017).

Therapeutic strategies are determined by tumour biology, stage, gestational stage and the patient´s wish. Cancer treatment should adhere as much as possible to treatment guidelines for non-PABC and should also be discussed in a multidisciplinary setting including obstetricians, gynaecologists, medical and surgical oncologists, pediatricians and hematologists (Amant 2012, Zagouri 2013, Basta 2015, Rojas 2019). Whether the patient already has children, her desire to continue the present pregnancy and to maintain fertility determines her choices for management of PABC (Navrozoglou 2008). A number of fertility preservation procedures such as ovarian suppression and oocyte and embryo cryopreservation exist and these should be discussed and ideally initiated before the onset of systemic therapy (Gerstl 2018). A proposed algorithm for the diagnosis and treatment of PABC is provided in Appendix IV.

Prognosis

In Europe, the relative 5-year survival rate for breast cancer irrespective of tumour type, stage and age at diagnosis is estimated to be 81-82%, and for those with early stage disease over 90 % (Allemani 2013, Sant 2015, Simoes 2018). European estimates for 1-year and 3-year relative survival are 95% and 87%, respectively (Sant 2015). PABC is a rare disease which limits the possibilities to conduct large powered controlled studies to address the question about survival. Women diagnosed during pregnancy with stages I and IIA have similar survival rates compared to non-pregnant women (Cardonick 2014). However, PABC is two and half times more likely to be diagnosed with advanced disease than non-PABC (Amant 2012).

A meta-analysis of 30 studies found that PABC patients had a significantly higher risk of death compared with those with non-PABC, pooled hazard ratio (HR) being 1.44 (95 % CI 1.27–1.63). The same results remained when adjusted for tumor stage (tumor size, nodal status or both) pooled HR being 1.40 (1.17–1.67) (Azim 2012). The most recent meta-analysis adopted a broader definition of PABC whereby cases diagnosed during pregnancy or up to five years postpartum were included. When measured through overall survival, there was an increased risk of death for PABC patients pooled HR being 1.57 (95 % CI 1.35-1.82). When cases where limited to those diagnosed up to one year postpartum, the pooled HR was 1.97 (95 % CI 1.88-2.06) and it did not change when two years period postpartum was used (Hartman 2016). The same meta-analysis found that women who have had a previous diagnosis of breast cancer and who subsequently become pregnant have reduced risk of death compared to those who do not become pregnant following breast cancer diagnosis. The result remained when accounted for the healthy mother effect i.e. selection bias whereby women who have had earlier stage disease and favourable outcomes are more likely to conceive than those who have relapsed.

There is ongoing controversy as to whether delayed diagnosis and young age at diagnosis account for the poor prognosis of PABC or if tumour biology of PABC is more aggressive than non-PABC when matched for age and stage (Expert Panel 2018). A study including breast cancer patients mainly from Germany and Belgium in 2003-2011 did not find a significant difference in disease free survival (DFS) or overall survival (OS) between 311 PABC and those of 865 non-PABC controls matched for known prognostic factors such as stage, age, hormonal receptors and type of treatment (Amant 2013). A study from Sweden including 778 women with PABC and 1661 breast cancers in nulliparous women found that woman with PABC, and particularly those diagnosed 0-12 months postpartum, had more advanced tumours, higher proportion of ER/PR negative, HER2 positive and triple negative tumours. Compared to nulliparous women, women with PABC had increased hazard ratios for mortality but when adjusted for tumour characteristics, the HRs were attenuated and nonsignificant suggesting that poorer prognosis is attributed to tumour characteristics (Johansson 2018). A retrospective chart review of 99 PABC cases and 186 non-PABC controls matched by age and year of diagnosis conducted in New York showed PABCs to be more often ER and PR negative but no difference in Her2/neu overexpression. No significant difference in disease-free or overall survival between PABC and non-PABC patients was observed. Authors concluded that PABC is not an independent negative prognostic factor when controlling for receptor status and stage (Murphy 2012).

Breast cancer arising in the postpartum period is significantly associated with poor overall survival and risk of relapse or disease progression measured through disease free survival compared to patients diagnosed during pregnancy (Azim 2012, Hartman 2016). In a study from Germany including 25 PABC patients, 5-year survival rate was 76 % and 10-year survival rate 68 %. 10-year-survival was only 50 % for patients diagnosed postpartum (Simoes 2018). Thus, the current literature suggests existence of two distinct subgroups of PABC: those diagnosed during pregnancy and those affected after delivery and it is important to specifically research these two subsets of PABC in greater detail.

There is no evidence that termination of pregnancy would improve maternal outcome i.e. provide survival benefit for PABC patients (Navrozoglou 2008, Cardonick 2010, Amant 2012, Cardonick 2014). The most significant fetal sequalae of PABC are from iatrogenic prematurity. Carrying a pregnancy to near term should be a management goal (Zagouri 2013, MacDonald 2020). The mode of delivery should be determined based on obstetrical indication i.e. vaginal delivery is preferred over caesarean section due to

shorter recovery period (Zagouri 2013, Cardonick 2014, Rojas 2019). However, relatively high rates of caesarean sections have also been reported (Simoes 2018). Induced deliveries indicated prior to the term should be limited to late preterm deliveries between 35-37 weeks for patients who complete chemotherapy by 32-33 in order to complete the treatment (Cardonick 2014).

7. Research question and objectives

ConcePTION is a consortium setting up a platform to generate accurate information about safety of medications in pregnancy and breastfeeding. There are overall five demonstration projects (DP) concerned with specific topics and methodological issues. The goal of this DP is to study drug exposure when disease is measured through accurate identification of an incident case. Therapies for PABC are used as motivating example. The objective of this drug utilisation study will be to describe patterns of chemotherapy, endocrine therapy and targeted therapy use before, during, and after pregnancy, during time periods available within each data source.

The main research questions are:

1. What is the incidence of PABC and non-PABC in women of reproductive age in European countries?
2. Is the pattern of cancer treatment (including surgery, radiation, chemotherapy, endocrine therapy and targeted therapy) similar in PABC and non-PABC patients?
3. What medications are used and how does the pattern of treatment change over the course of a pregnancy (e.g. prior to pregnancy, during, after pregnancy), by cancer severity and by country?
4. What is the pregnancy outcome (terminations of pregnancy, live birth, stillbirth, preterm birth, SGA/ IUGR, as available) and mode of delivery for women with PABC and non-PABC and does that vary by cancer severity and by country?
5. What is the 5-year relative survival for women with PABC and non-PABC when adjusted for tumour characteristics and is there a difference in survival between PABC patients diagnosed during pregnancy and those diagnosed on year postpartum?

8. Research methods

8.1 Study setting

Contributing countries or databases: Finland, UK (Wales), Spain (Valencian Region), Germany and Scotland (see table in Appendix I). Data availability by DAP, pending on results from the Data characterization (WP7):

DAP	Study period	Comments related to data availability
Finland	1996-2019	
Spain (Valencian Region)	2007-2019 (or latest available)	In drugs database, recommended to use since 2010
Germany	2004-2018	
UK (Wales)	1998-2019	
Scotland		protocol review lacking

8.2 Study design

Study design will be a cohort study. Study population is women who were free of cancer prior to age 15 (i.e. childhood cancers excluded) and diagnosed with breast cancer at reproductive. Women diagnosed with breast cancer during pregnancy or within 12 months after childbirth (live births and stillbirths included) will be defined as pregnancy-associated breast cancer (PABC). Women who have no indication within the data source of a pregnancy at diagnosis or diagnosed > 1 year after delivery will be defined as non-pregnant breast cancer (non-PABC).

Study period will start from the first year cancer incidence and birth outcomes are available from the data source (whichever is the latest) and will end to most recent date of the data source where maternal death is available.

8.3 Study material

Exposures
Cancer therapies (surgery, radiation, chemotherapy, endocrine therapy and targeted therapy) will be evaluated as binary variables (treatment yes/no) overall and by trimesters. Specific analyses on medications at a substance level before, during and after pregnancy will be performed whenever possible. Medications will be classified according to the (ATC) classification system (https://www.whocc.no/atc_ddd_index/). Primary exposure is antineoplastic agents (ATC class L01) and endocrine therapy (ATC class L02). Full list of medications is provided in Appendix VI.

Primary outcome – PABC and maternal survival
Maternal 5-year relative survival in woman with PABC vs non-PABC patients.

Secondary outcomes - pregnancy outcomes
Mode of delivery and the following pregnancy outcomes (termination of pregnancy, live birth, stillbirth, preterm birth, small for gestational age SGA/ IUGR) in women with PABC and non-PABC.

Diagnostic codes and quality indicators from the ConcePTION data characterization will be employed. We will use the event definitions and algorithms to identify these pregnancy outcomes as agreed in the ConcePTION Consortium.

Variables

Variables that are important: maternal age at diagnosis, date of cancer diagnosis, grade TNM stage, gestational age (GA) at diagnosis, tumour morphology (histology), oestrogen receptor and progesterone receptor status, HER2/neu receptor status, tumour grade, GA age at first cycle of chemotherapy, GA at delivery, mode of delivery, pregnancy outcome, birth weight, preterm birth, congenital anomalies, and childhood development, if available.

Variables that may have an impact on pregnancy outcome, cancer risk and/or treatment choices such as country, calendar year, parity, family history, BMI, socioeconomic status, smoking, alcohol intake, substance misuse, breastfeeding, and relevant co-morbidities as available.

8.4 Data sources and management

This study will utilize data from five countries with geographic spread across Europe using the ConcePTIONcommon data model (CDM). Data will remain in the country of origin, and only aggregated results will be loaded to ConcePTION platform. The following data sources have been selected (See the list of all DAP in Appendix I):

- Finland
- UK: SAIL database (Wales)
- Spain: Rare Disease Research Unit. FISABIO (Valencian Region)
- Germany: GePARD
- Scotland

A description of data sources participating in this project

Germany (GePaRD)
GePaRD is based on claims data from four statutory health insurance providers in Germany and currently includes information on approximately 25 million persons who have been insured with one of the participating providers since 2004 or later. Per data year, there is information on approximately 20% of the general population and all geographical regions of Germany are represented. In addition to demographic data, GePaRD contains information on dispensations of reimbursable prescription drugs as well as outpatient (i.e., from general practitioners and specialists) and inpatient services and diagnoses. **The Leibniz Institute for Prevention Research and Epidemiology – BIPS** will be Data Access Provider for the GePaRD data. GePaRD data have been used for vaccine safety studies and pregnancy studies. GePaRD is listed under the ENCePP resources database.

Finland (linkage of several nationwide registries)
Universal health insurance coverage is accessible for all citizens and permanent residents in the country. Municipalities (currently around 200) are responsible for arranging and funding health care. Health services are divided into primary health care and specialized medical care. The data that THL provides access to is the majority of healthcare registries covering the whole population of Finland (around 5.6 million inhabitants). The core data of the Drugs and Pregnancy project includes data from Medical Birth Register, Register of Congenital Malformations and Register of Induced Abortions from 1996 onwards. Drugs and Pregnancy Database also includes following registries maintained by the Kela: Special refund codes and diagnoses three months before pregnancy to three months following delivery or abortion and drug purchases and reimbursements from three months before pregnancy to three months following delivery or abortion also 1996 onwards. The core data of the Drugs and Pregnancy currently includes all pregancies ending in delivery or induced abortion in 1996-2018 the total amount being aroung 1,5 million pregnancies. Additional data sources maintained or accessed by THL and mapped to ConcePTION CDM are Care Register for Health Care (HILMO), Register of Primary Health Care visits (Avohilmo), Finnish Cancer Registry and Cause of Death Registry for women diagnosed with cancer and for children up to one year of age. Data collection is mandatory by law and does not require informed consent from the recorded subjects. Data is stored on an individual level and can be linked by the personal identification number assigned to all citizens and permanent residents in Finland at birth or upon immigration.

Spain (FISABIO) Rare Disease Research Unit of FISABIO integrates, as described in WP-7: Prescription and dispensations dataset (GAIA), Morbidity through Hospital discharges database (CMBD), Cancer Registry (RTC), Perinatal Mortality Registry (RMPCV), Mortality Registry (RMCV), Birth Registry (MetaB) and Congenital anomaly Registry (RPAC-CV).
A set of multiple, public, population-wide electronic databases for the Valencian Region will be used. Valencian Region is the fourth most populated Spanish region, with ≈5 million inhabitants and an annual birth cohort of 48 000 newborns representing 10.7% of the Spanish population and around 1% of the European population. Together, all the included databases will provide exhaustive longitudinal information including sociodemographic and administrative data (sex, age, nationality, etc.), clinical (diagnoses, procedures, etc.), drug information (prescription, dispensation) and healthcare utilization data from hospital. It also includes a set of associated population databases and registries of significant care areas such as cancer, rare diseases, congenital anomalies, and also public health databases from the population screening programmes. All electronic health systems use the ICD-9-CM and/or the ICD-10 and its derivates. All the information in the databases can be linked at the individual level through a single personal identification code, the health number. The databases were initiated at different moments in time, but it is recommended to use them since 2010 (due to high quality improvements) until the last year available (it could differ between databases).

United Kingdom: Wales (SAIL)
The Secure Anonymised Information Linkage (SAIL) Databank sources, accesses, links and analyses prospectively collected routine health and population data, within a governed infrastructure that is safe and secure. All datasets are anonymised and encrypted by a third party, and returned to SAIL for linkage. Data are held on 5, 400,000 people, since 1998. Data are available within 3 months of events. SAIL holds linkable,

anonymised national datasets, including: Accident and emergency care from 2009, Critical care from 2016, Congenital Anomaly Register and Information Service for Wales (CARIS), In-patient and out-patient PEDW records, Maternity dataset from 2015 for additional data on childbirth, National Community Child Health Database (NCCHD, includes gestation (ultrasound), birth centiles, childbirth, infant feeding, developmental screening and vaccinations), National Pupil Database Wales (education attainment to 16), ONS births and deaths (compulsory registration), Primary care data (including all prescriptions and diagnoses) from ~75% of Welsh GP practices. **Swansea University** will be Data Access Provider for the SAIL data in this project.

Scotland
The National Health Service (NHS) in the UK is a publicly-funded health service that is free at the point of care for everyone resident in the UK. Information Services Division (ISD) is a part of NHS National Services Scotland and holds national level NHS health and health-related data for over 5 million people in Scotland. The Service holds health related data which in some cases cover an individual from before birth, with the mother's antenatal records, through to that individual's death registration record. For the Data Characterisation study ISD will use the most pertinent National Dataset and registries to fill the ConcePTION CDM main columns: NRS Births, Stillbirth and Infant Deaths, Teenage pregnancy, Prescribing Information System for Scotland (PIS), Maternity Inpatient and Day Case, Scottish Birth Record, Scottish Drug Misuse Database, National Record of Scotland Deaths Data, Notification of Abortion Statistics (AAS), Outpatient data, Hospital admission, Mental Health.

Each DAP will perform the following tasks:
1) Obtain required ethical and legal permissions to use the data in this project
2) Extract and transform the individually linked data locally into ConcePTION CDM
2) Check and run script distributed to the DAP by the ConcePTION coordinating centre
3) Run standard scripts to check data quality (quality assessment of data)
4) Run the scripts for this specific study
2) Send aggregated results to the remote ConcePTION secure platform while data will remain with the DAP.

8.5 Study size

Assuming 25% of annual breast cancer cases occur in women of childbearing age (15-54 years) and 3.8% of breast cancers occur in pregnant and breastfeeding woman (i.e. the ratio of PABC vs. non-PABC to be 1 to 6) and 5-year relative survival at 82%, to detect a 40% increased risk for death in PABC vs. non-PABC patients with 80% power and type I error rate of 0.05, we would require 2 965 breast cancer cases. Using the EU-27 breast cancer incidence of 90.7 per 100,000 in women aged 15-54, we will need a sample size of around 3.27 million women years in women of childbearing age. https://sample-size.net/sample-size-survival-analysis/

The prevalence of low birth weight in the general population is estimated to be around 5% and the prevalence of preterm birth 6 to 10 %.

8.6 Data analysis

Statistical analyses will be carried out through the ConcePTION ecosystem with a common data model and common statistical analysis plan. Data will remain local, and only aggregated results or effect estimates will be submitted for pooling. Initial pilot modelling of the statistical analysis plan (SAP) will be carried out using the Finnish data. Then scripts coded in R will be circulated to all DAPs through the ConcePTION task management system.

Descriptive analysis of cancer therapies used to treat breast cancer over the course of a pregnancy (prior to, during, after pregnancy) and during pregnancy (1st, 2nd or 3rd trimester) will be provided.

Relative survival has become the “gold standard” method for estimating cancer survival in population-based data. Relative survival is the ratio of the observed probability of survival (*S*) of cancer patients and the probability of survival that would have been expected (*E*) if patients had had the same survival probability as in the standardized general population

$$R(t) = \frac{S(t)}{E(t)}$$

where *R, S* and *E* are the relative, observed and expected survival probabilities, respectively, at time *t*. The expected survival derives from the general population mortality using life tables stratified by age, sex and calendar period (Nur 2010).

5-year relative survival analyses will be carried out using e.g. Cox proportional hazard regression and flexible parametric models to elucidate the complex associations between time-since-conception, medication exposure, breast cancer incidence and survival and to control for important confounders. Effects will be presented as relative risk estimates with CI describing the precision of the estimate (95% CI). Survival analysis is increasingly used also in perinatal epidemiology to assess time-varying risk factors for pregnancy outcomes (Ahrens 2012). Maternal survival and pregnancy outcomes will be compared according to breast cancer stage to disentangle the impact of the medication from the underlying illness (Wood 2010).

Stage at diagnosis is a very strong predictor for prognosis but data can be incomplete in the data source. Furthermore, information on stage is more often incomplete in patients with advanced tumours and poor survival. Thus, complete case analysis may restrict the dataset substantially, introduce bias and lead to incorrect conclusions (Nur 2010).

Analyses will be stratified by country and time period.

Handling of missing data

Patterns of missingness will be explored and handled as appropriate (Sterne 2009, Nur 2010, Perkins 2018). If supported by ConcePTION tools and necessary variables to predict missing values are included in source data, we will perform multiple imputation by chained equations (MICE) for missing values in covariates to improve measurement of medication exposure and prognostic factors. MI will be done including exposure variables, covariates and outcome variables and estimates from imputed data sets will be combined by Rubin´s rules (Rubin 1987, Graham 2007).

Meta-analyses will be used to combine the aggregate data obtained from each DAP as appropriate. For the meta-analysis, effect estimates will be pooled using the inverse variance method for weighting i.e. weighting the country-specific log hazard ratios by the inverse of the within countries’ variances (Selmer 2016). The meta-analysis on aggregate date will allow for adjustment for country-optimized covariates (See Appendix X: Meta analytic techniques for use in ConcePTION DPs).

Sensitivity analyses to assess the robustness of results

As the strongest known modifier of a woman´s breast cancer risk is her reproductive history (Husby 2018), a sensitivity analysis using only nulliparous women as a control group will be done.

Socioeconomic status (SES) has been associated with both breast cancer incidence and survival (Woods 2006, Sprague 2011, Quaglia 2013, Akinyemiju 2015). Previous studies have been unable to assess any difference in survival for women according to oestrogen receptor status with pregnancy (Hartman 2016). If supported by the available data, a sensitivity analysis adjusting for SES and sensitivity analysis according to receptor status to ascertain whether there are any differences in outcomes for those who become pregnant with endocrine-responsive tumours compared to non-responsive tumours will be done.

If it is possible to perform MICE as a primary analysis, then complete case analysis will follow as a sensitivity analysis.

8.7 Quality control

The studies will be conducted in line with the ENCePP Code of Conduct for scientific independence and transparency, and the FAIR (Findable, Accessible, Interoperable, Reusable) principles of the ConcePTION.

Each DAP will be responsible for the extraction, transformation, and loading (ETL) of their original data to the ConcePTION CDM. Standardized scripts in R to run against data in the ConcePTION CDM will be written by the group of statisticians and computer scientists and delivered through the task management system. Result outputs from the scripts will be submitted to a computing platform that can be accessed remotely by DAPs and ConcePTION partners and participating DAPs using authentication. Access to each DAP's results on the platform will be limited to the data access provider, WP1 public partner statisticians, and WP7 public partner statisticians.

The data quality and characterization checks will take place in collaboration with partners. All data will remain local and only summary measures will be reviewed by DAPs in collaboration with WP7. This process will proceed iteratively in collaboration with each DAP until consensus on fitness for purpose has been reached between WP7 and the DAP, the result of this consensus process and some core results will be made available on the catalogue in a private area for inspection by investigators and DAPs. For all indicators and characterization outputs resulting in a cell count less than 5, counts will not be reported and will be replaced with "<5" programmatically (see Protection of Human Subjects, below for details on numbers 1-4, and missing data).

Level 1 data checks review the completeness and content of each data table of the ConcePTION CDM to ensure that mandatory variables contain data and conform to the formats specified by the CDM specifications (e.g., data types, variable lengths, formats, acceptable values, etc.). Level 1 data checks are to verify that ETL procedure to convert from source data to the ConcePTION CDM has been completed as expected. Formats for all values will be assessed and compared to a list of acceptable formats. Frequency tables of variables with finite allowable values will be created to identify unacceptable values. Distributions of days and months of birth to assess any rounding will be constructed.

Level 2 data checks assess the logical relationship and integrity of data values within a variable or between two or more variables within and between tables. Level 2 data checks will assess records occurring outside recorded person time (i.e. before birth, after death, or outside of recorded observation periods) and implausible combinations such as high birth weight and preterm birth.

8.8 Limitations of the research methods

Coverage and validity of information on medication administered in hospitals and how well that is linkable with the cancer registry data is currently unknown. This may create observation gaps for treatment and there may be some patients classified as unexposed who are actually treated. However, the potential misclassification of treatment is assumed to be similar in PABC and non-PABC patients for which it should not affect our comparison.

Considering the optimal method of analysis i.e. MICE, it may not be possible to identify all the important covariates that are needed to predict the missingness and several rounds of data updates and calibration of the model may not be feasible in the ConcePTION setting.

All the potential data sources willing to contribute have been identified for the study. We may still have limited power to adjust for the most important prognostic factors. Also, when the exposure and outcome

are rare, there might be countries without an exposed event or DAP may prohibit disclosing the numbers which may limit the possibilities to analyse the data (see below, 9).

9. Protection of human subjects

This is non-interventional study based on secondary use of data. Therefore, the reporting of suspected adverse reactions as individual case safety reports is not required as per the EMA Guideline on Good Pharmacovigilance Practices. This study is not considered a PASS because it does not aim to "identify, characterize or quantify a safety hazard, confirm the safety profile of the medicinal product, nor measure the effectiveness of risk management measures." When the data characterization is done, the marketing authority holder shall monitor the results and consider possible implications for the risk-benefit balance of the medicinal product concerned.

This study is compliant with the provisions of the ENCePP Code of Conduct, Revision 4. The protocol and governance arrangements must be reviewed and approved by a properly constituted institutional review board/independent ethics committee/research ethics board according to the national guidelines before the study start. A signed and dated statement that the approval has taken place and waiver of informed consent must be given to the principal investigator before study initiation.

DAPs may prohibit the public release of numbers 1-4 in any data category (except 'information missing'). This applies to all documents in the public domain and communications outside secure links (e.g. emails). This not only applies to text and tables, but also to reporting that could lead to the derivation or calculation of a low number in any category, for example:

1. Where an unadjusted OR or RR is reported for a contingency table, and the denominators and numerator in the larger category are available, it is easy to calculate the missing value.
2. Where a proportion is reported in a figure or graph or table, and the total number of cases is reported either in the same report or another report or publication, the number can be calculated.
3. Where numbers in categories across Europe are low, we only have permission to say the 'Wales contributed data'. We would breach our conditions of approval to say 'Wales contributed cases'. *low can only be defined with reference to the number of cases and countries.

Wales' European projects have permission to pass low numbers (1-4) to the centre responsible for analysis, via secure links to authorised colleagues on the above conditions. These numbers are to be aggregated before reporting. https://www.ncbi.nlm.nih.gov/books/NBK350762/

10. Plans for disseminating and communicating study results

The results of this study will be published as ConcePTION report and scientific papers in peer-reviewed journals. Manuscripts will be prepared independently by the investigators and in accordance with the current guidelines of STrengthening the Reporting of OBservational studies in Epidemiology (STROBE), the ENCePP standards and EMA guidelines.

11. References

Ahn BY, Kim HH, Moon WK, Pisano ED, Kim HS, Cha ES, Kim JS, Oh KK, Park SH. Pregnancy- and lactation-associated breast cancer: mammographic and sonographic findings. J Ultrasound Med 2003;22(5):491-7; quiz 498-9.

Akinyemiju TF, Pisu M, Waterbor JW, et al. Socioeconomic status and incidence of breast cancer by hormone receptor subtype. SpringerPlus 2015;4:508.

Allemani CA, Sant M, Weir HK, et al. Breast cancer survival in the US and Europe: a CONCORD high-resolution study. Int J Cancer 2013; 132(5):1170-1181.

Amant F, Loibl S, Neven P, et al. Breast cancer in pregnancy. Lancet 2012; 379: 570-579.

Amant F, von Minckwitz G, Han SN, et al. Prognosis of women with primary breast cancer diagnosed during pregnancy: results from an international collaborative study. J Clin Oncol 2013; 31(20):2532-2539.

Ahrens K, Lash TL, Louik C, et al. Correcting for exposure misclassification using survival analysis with a time-varying exposure. Ann Epidemiol 2012; 22:799-806.

Azim HA Jr, Santoro L, Russell-Edu W, Pentheroudakis G, Pavlidis N, Peccatori FA. Prognosis of pregnancy-associated breast cancer: a meta-analysis of 30 studies. Cancer Treat Rev. 2012 Nov;38(7):834-42.

Basta P, Bak A, Roszkowski K. Cancer treatment in pregnant women. Contemp Oncol (Pozn). 2015;19(5):354-60.

Beatty JD, Adachi M, Bonham C, et al. Utilization of cancer registry data for monitoring quality of care. Am J Surg 2011;201:645–649.

Caldarella A, Amunni G, Angiolini C, et al. Feasibility of evaluating quality cancer care using registry data and electronic health records: a population-based study. Int J Qual Health Care 2012;24:411–418.

Cardonick E. Pregnancy-associated breast cancer: optimal treatment options. Int J Womens Health. 2014 Nov 4; 6:935-43.

Expert Panel on Breast Imaging:, diFlorio-Alexander RM, Slanetz PJ, Moy L, Baron P, Didwania AD, Heller SL, Holbrook AI, Lewin AA, Lourenco AP, Mehta TS, Niell BL, Stuckey AR, Tuscano DS, Vincoff NS, Weinstein SP, Newell MS. ACR Appropriateness Criteria® Breast Imaging of Pregnant and Lactating Women. J Am Coll Radiol. 2018 Nov;15(11S):S263-S275.

Gerstl B, Sullivan E, Ives A, et al. Preganncy outcomes after a breats cancer diagnosis: a systematic review and meta-analysis. Clin Breast Cancer 2018; 18(1):e79-e88.

Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci. 2007;8:206–213.

Gurney J, Sarfati D, Dennett E, et al. The completeness of cancer treatment data on the National Health Collections. N Z Med J 2013;126:69–74.

Hartman EK, Eslick GD. The prognosis of women diagnosed with breast cancer before, during and after pregnancy: a meta-analysis. Breast Cancer Res Treat. 2016 Nov;160(2):347-360.

Husby A, Wohlfahrt J, Øyen N, Melbye M. Pregnancy duration and breast cancer risk. Nat Commun 2018 Oct 23;9(1):4255.

Johansson ALV, Andersson TM, Hsieh CC, Jirström K, Cnattingius S, Fredriksson I, Dickman PW, Lambe M. Tumor characteristics and prognosis in women with pregnancy-associated breast cancer. Int J Cancer 2018 Apr 1;142(7):1343-1354.

Johansson ALV, Weibull CE, Fredriksson I, Lambe M. Diagnostic pathways and management in women with pregnancy-associated breast cancer (PABC): no evidence of treatment delays following a first healthcare contact. Breast Cancer Res Treat. 2019 Apr;174(2):489-503.

Kakoulidis I, Skagias L, Politi E. Pregnancy associated breast cancer (PABC): aspects in diagnosis. Breast Dis. 2015;35(3):157-66.

Lakhani SR, Ellis IO, Schnitt SJ, et al. eds. World Health Organization classification of tumours of the breast. 4th ed. Lyon: IARC Press, 2012.

Macdonald HR. Pregnancy associated breast cancer. Breast J. 2020 Jan;26(1):81-85.

Mallin K, Palis BE, Watroba N, et al. Completeness of American cancer registry treatment data: implications for quality of care research. J Am Coll Surg 2013;216:428–437.

Murphy CG, Mallam D, Stein S, et al. Current or recent pregnancy is associated with adverse pathologic features but not impaired survival in early breast cancer. Cancer 2012; 118(13): 3254-3259.

Myers KS, Green LA, Lebron L, Morris EA. Imaging Appearance and Clinical Impact of Preoperative Breast MRI in Pregnancy-Associated Breast Cancer. AJR Am J Roentgenol. 2017 Sep;209(3):W177-W183.

Navrozoglou I, Vrekoussis T, Kontostolis E, Dousias V, Zervoudis S, Stathopoulos EN, Zoras O, Paraskevaidis E. Breast cancer during pregnancy: a mini-review. Eur J Surg Oncol 2008;34(8):837-843.

Nur U, Shack LG, Rachet B, et al. Modelling relative survival in the presence of incomplete data: a tutorial. Int J Epidemiol 2010; 39(1):118-128.

Perkins NJ, et al. Principled Approaches to Missing Data in Epidemiologic Studies. Am J Epidemiol 2018; 187: 568–575.

Quaqlia A, Roberto L, Carlo M, et al. Socio-economic inequalities: A review of methodological issues and the relationships with cancer survival. Crit Rev Oncol Hematol 2013; 85(3):266-277.

Rojas KE, Bilbro N, Manasseh DM, Borgen PI. A Review of Pregnancy-Associated Breast Cancer: Diagnosis, Local and Systemic Treatment, and Prognosis. J Womens Health (Larchmt). 2019 Jun;28(6):778-784.

Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.

Sant M, Lopez MDC, Agrest R, et al. Survival of women with cancers of breast and genital organs in Europe 1999-2007: Results of the EUROCARE-5 study. Eur J Cancer 2015; 51:2191-2205.

Selmer R, Haglund B, Furu K, et al. Individual-based versus aggregate meta-analysis in multi-database studies of pregnancy outcomes: the Nordic example of selective serotonin reuptake inhibitors and venlafaxine in pregnancy. Pharmacoepidemiol Drug Saf 2016; 25: 1160–1169.

Simoes E, Graf J, Sokolov AN, Grischke EM, Hartkopf AD, Hahn M, Weiss M, Abele H, Seeger H, Brucker SY. Pregnancy-associated breast cancer: maternal breast cancer survival over 10 years and obstetrical outcome at a university centre of women's health. Arch Gynecol Obstet 2018 Aug;298(2):363-372.

Sobin LH, Gospodarowicz MK, Wittekind C, eds. International Union against Cancer (UICC) TNM classification of malignant tumours, 7th edn. New York: Wiley 2009.

Sprague BL, Trentham-Dietz A, Gangnon Re, et al. Socioeconomic status and survival after an invasive breast cancer diagnosis. Cancer 2011; 117(7):1542-1551.

Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393.

Vinatier E, Merlot B, Poncelet E, et al. Breast cancer during pregnancy. Eur J Obstet Gynecol Reprod Biol 2009; 147(1):9-14.

Wood ME, Lapane KL, van Gelder MMHJ, et al. Making fair comparisons in pregnancy medication safety studies: An overview of advanced methods for confounding control. Pharmacoepidemiol Drug DSaf 2018; 27:140-147.

Woods LM, Rachet B, Coleman MP. Origins of socio-economic inequalities in cancer survival: a review. Ann Oncol 2006; 17(1):5-19.

Zagouri F, Psaltopoulou T, Dimitrakakis C, Bartsch R, Dimopoulos MA. Challenges in managing breast cancer during pregnancy. J Thorac Dis. 2013 Jun;5 Suppl 1(Suppl 1):S62-7.

12. Annex 1. List of stand-alone documents

Number	Document reference number	Date	Title
1	Appendix I		Databases and third party DAPs
2	Appendix II		Main categories of malignant primary tumours of the female breast
3	Appendix III		TNM staging system for breast cancer
4	Appendix IV		Proposed algorithm for the diagnosis and treatment of PABC
5	Appendix V		Drugs used to treat event
6	Appendix VI		Procedures used to treat event
7	Appendix VII		DAPs experience on breast cancer
8	Appendix VIII		Covariate items across DAPs

13. Annex 2. ENCePP checklist for study protocols

A copy of the ENCePP Checklist for Study protocols available at http://www.encepp.eu/standards_and_guidances/checkListProtocols.shtml completed and signed by the main author of the study protocol should be included in Annex 2.

Doc.Ref. EMA/540136/2009

ENCePP Checklist for Study Protocols (Revision 4)

Adopted by the ENCePP Steering Group on 15/10/2018

The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) welcomes innovative designs and new methods of research. This Checklist has been developed by ENCePP to stimulate consideration of important principles when designing and writing a pharmacoepidemiological or pharmacovigilance study protocol. The Checklist is intended to promote the quality of such studies, not their uniformity. The user is also referred to the ENCePP Guide on Methodological Standards in Pharmacoepidemiology, which reviews and gives direct electronic access to guidance for research in pharmacoepidemiology and pharmacovigilance.

For each question of the Checklist, the investigator should indicate whether or not it has been addressed in the study protocol. If the answer is "Yes", the section number of the protocol where this issue has been discussed should be specified. It is possible that some questions do not apply to a particular study (for example, in the case of an innovative study design). In this case, the answer 'N/A' (Not Applicable) can be checked and the "Comments" field included for each section should be used to explain why. The "Comments" field can also be used to elaborate on a "No" answer.

This Checklist should be included as an Annex by marketing authorisation holders when submitting the protocol of a non-interventional post-authorisation safety study (PASS) to a regulatory authority (see the Guidance on the format and content of the protocol of non-interventional post-authorisation safety studies). The Checklist is a supporting document and does not replace the format of the protocol for PASS presented in the Guidance and Module VIII of the Good pharmacovigilance practices (GVP).

Study title: Studying drug exposure when disease is measured through accurate identification of an incident case: application to breast cancer and pregnancy

EU PAS Register® number:
Study reference number (if applicable):

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection[1]	☒	☐	☐	8.1, 8.2
1.1.2 End of data collection[2]	☒	☐	☐	8.1, 8.2
1.1.3 Progress report(s)		☒	☐	
1.1.4 Interim report(s)	☐	☒	☐	
1.1.5 Registration in the EU PAS Register®	☒	☐	☐	4
1.1.6 Final report of study results.	☒	☐	☐	4

[1] Date from which information on the first study is first recorded in the study dataset or, in the case of secondary use of data, the date from which data extraction starts.
[2] Date from which the analytical dataset is completely available.

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	7
2.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	5
2.1.2 The objective(s) of the study?	☒	☐	☐	7
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	7
2.1.4 Which hypothesis(-es) is (are) to be tested?	☒	☐	☐	7
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.1 Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☒	☐	☐	8.2
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	9
3.3 Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	6, 8.6
3.4 Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☒	☐	☐	8.6
3.5 Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☐	☐	☒	

Comments:

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.1 Is the source population described?	☒	☐	☐	8.2
4.2 Is the planned study population defined in terms of:				
4.2.1 Study time period	☒	☐	☐	8.1
4.2.2 Age and sex	☒	☐	☐	8.2
4.2.3 Country of origin	☒	☐	☐	8.1
4.2.4 Disease/indication	☒	☐	☐	8.2
4.2.5 Duration of follow-up	☒	☐	☐	8.1, 8.2
4.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	8.2

Comments:

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.1 Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☒	☐	☐	6, 8.4, Appendix VIII
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☒	☐	☐	6, 8.4, Appendix VIII
5.3 Is exposure categorised according to time windows?	☒	☐	☐	7
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	☒	☐	☐	7
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	☒	
5.6 Is (are) (an) appropriate comparator(s) identified?	☒	☐	☐	8.2

Comments:

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	8.3
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	8.3
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☒	☐	☐	8.7
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	☐	☐	☒	
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☒	☐	☐	8.2
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☒	☐	☐	8.8

Comments:

Section 8: Effect measure modification	**Yes**	**No**	**N/A**	**Section Number**
8.1 Does the protocol address effect modifiers? (e.g. collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	☒	☐	☐	7, 8.3, 8.6

Comments:

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☒	☐	☐	6, 8.4, Appendix VIII
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	6, 8.4, Appendix VIII
9.1.3 Covariates and other characteristics?	☒	☐	☐	8.4, Appendix VIII
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☒	☐	☐	8.4, Appendix VIII
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	6, 8.4, Appendix VIII
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	6, 8.4, Appendix VIII
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	6, 8.3, Appendix V-VI
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	CDM documents
9.3.3 Covariates and other characteristics?	☒	☐	☐	CDM documents

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☒	☐	☐	8.4, CDM documents

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	8.6
10.2 Is study size and/or statistical precision estimated?	☒	☐	☐	8.5
10.3 Are descriptive analyses included?	☒	☐	☐	8.6
10.4 Are stratified analyses included?	☒	☐	☐	8.6
10.5 Does the plan describe methods for analytic control of confounding?	☒	☐	☐	8.6
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☒	☐	☐	8.6
10.7 Does the plan describe methods for handling missing data?	☒	☐	☐	8.6
10.8 Are relevant sensitivity analyses described?	☒	☐	☐	8.6

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.1 Does the protocol provide information on data storage? (e.g. software and IT environment, database maintenance and anti-fraud protection, archiving)	☒	☐	☐	8.7
11.2 Are methods of quality assurance described?	☒	☐	☐	8.7
11.3 Is there a system in place for independent review of study results?	☒	☐	☐	8.7

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1 Does the protocol discuss the impact on the study results of:				
12.1.1 Selection bias?	☒	☐	☐	8.8
12.1.2 Information bias?	☒	☐	☐	8.8
12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☒	☐	☐	8.8

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	8.5

Comments:

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	☒	☐	☐	9
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☐	☒	
13.3 Have data protection requirements been described?	☒	☐	☐	9

Comments:

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Section Number**
14.1 Does the protocol include a section to document amendments and deviations?	☒	☐	☐	3

Comments:

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Section Number**
15.1 Are plans described for communicating study results (e.g. to regulatory authorities)?	☐	☒	☐	
15.2 Are plans described for disseminating study results externally, including publication?	☒	☐	☐	10

Comments:

Name of the main author of the protocol: Maarit Leinonen

Date: 30/9/2021

Signature: 

Appendix I: Databases and third-party DAPs

Country	Region	Data sources	DAP Name Name and Email of DAP contact person
Finland	Nationwide	Linkage of several registries	NIHW (THL) Finnish Institute for Health and Welfare, NIHW Maarit Leinonen (Maarit.leinonen@thl.fi)
Spain	Valencian Region	Linkage of several registries	FISABIO The Foundation for the Promotion of Health and Biomedical Research of Valencian Region. (Rare Disease Research Unit) Clara Cavero-Carbonell (cavero_cla@gva.es) Laia Barrachina-Bonet (barrachina_lai@gva.es) Laura García- Villodre Óscar Zurriaga
Germany	17% of population	GePaRD (Claims data)	BIPS Leibniz Institute for Prevention Research and Epidemiology BIPS Tania Schink (schink@leibniz-bips.de) Ulrike Haug (haug@leibniz-bips.de) Miriam Heinig (heinig@leibniz-bips.de) Katja Oppelt (oppelt@leibniz-bips.de)
UK Wales	Nationwide	SAIL Databank	USWAN University of Swansea Sue Jordan (s.e.jordan@swansea.ac.uk) Daniel Thayer (d.e.thayer@swansea.ac.uk)
UK Scotland	Nationwide		University of Dundee Scotland Tom MacDonald (t.m.macdonald@dundee.ac.uk)

Appendix II: Main categories of malignant primary tumours of the female breast

The World Health Organization **classification of tumors of the breast** is the most widely used pathologic classification system for tumours. Due to data availability, coding in cancer registries most likely follows the 4th edition of the WHO series published in 2012.

Epithelial tumors
- microinvasive carcinoma

Invasive breast carcinoma
- invasive breast carcinoma of no special type (NST)
- invasive lobular carcinoma
- tubular carcinoma
- cribriform carcinoma
- mucinous carcinoma
- carcinoma with medullary features
- carcinoma with apocrine differentiation
- carcinoma with signet-ring-cell differentiation
- invasive micropapillary carcinoma
- metaplastic carcinoma
- rare types

Epithelial-myoepithelial tumors
- adenoid cystic carcinoma

Papillary lesions
- intraductal papillary carcinoma
- encapsulated papillary carcinoma
- solid papillary carcinoma

Mesenchymal tumors
- liposarcoma
- angiosarcoma
- rhabdomyosarcoma
- osteosarcoma
- leiomyosarcoma

Fibroepithelial tumors
- phyllodes tumor (malignant and pediductal stromal tumor, low grade)

Tumors of the nipple
- Paget disease of the nipple

Clinical patterns
- inflammatory carcinoma
- bilateral breast carcinoma

Appendix III: TNM staging system for breast cancer

Staging is based on pathological staging (pTNM) when patient undergo surgery and/or clinical staging (cTNM). Tumour (T) describes the size of the tumour, node (N) describes whether the cancer has spread to lymph nodes and metastasis (M) describes whether the cancer has more distant metastases. Overview combining the pTNM and cTNM is provided below.

TX tumour cannot be assessed for size

Tis Carcinoma in situ (ductal carcinoma in situ, DCIS or Paget disease)

T1 tumour is ≤ 2 cm further divided as

- **T1mi** tumour is ≤ 0.1cm
- **T1a** tumour is more than 0.1 cm but ≤0.5 cm
- **T1b** tumour is more than 0.5 cm but not more than 1 cm
- **T1c** tumour is more than 1 cm but not more than 2 cm

T2 tumour is more than 2 cm but ≤ 5 cm

T3 tumour is bigger than 5 cm

T4 tumour is divided into following groups:

- **T4a** tumour has spread into the chest wall
- **T4b** tumour has spread into the skin and the breast might be swollen
- **T4c** tumour has spread to both the skin and the chest wall
- **T4d** tumour is inflammatory carcinoma

NX lymph nodes cannot be assessed for spreading

N0 no signs of cancer cells in lymph nodes or only isolated tumour cells (ITCs)

N1 lymph node spreading is divided into following groups:

- **N1mi** micrometastases in one or more lymph nodes
- **N1a** cancer cells have spread into 1 to 3 lymph nodes and at least one is larger than 2mm
- **N1b** cancer cells in the lymph nodes behind the breastbone found with a sentinel node biopsy
- **N1c** cancer cells in 1 to 3 lymph nodes in the armpit and in the lymph nodes behind the breastbone

N2

- **N2a** cancer cells in 4 to 9 the lymph nodes in the armpit, and at least one is larger than 2 mm
- **N2b** cancer cells in the lymph nodes behind the breast bone and no sign of cancer in lymph nodes in the armpit

N3 lymph node spreading is divided into following groups:

- **N3a** cancer cells in ≥ 10 lymph nodes in the armpit of which at least one is larger than 2mm, or cancer cells in the nodes below the collarbone
- **N3b** cancer cells in lymph nodes in the armpit and lymph nodes behind the breastbone
- **N3c** cancer cells in lymph nodes above the collarbone

M0 no sign of cancer spread

Mo(i+) no sign of the cancer on clinical examination or imaging but cancer cells present in blood, bone marrow or distant lymph nodes.

M1 cancer has spread to another part of the body

Appendix IV: Proposed algorithm for the diagnosis and treatment of PABC

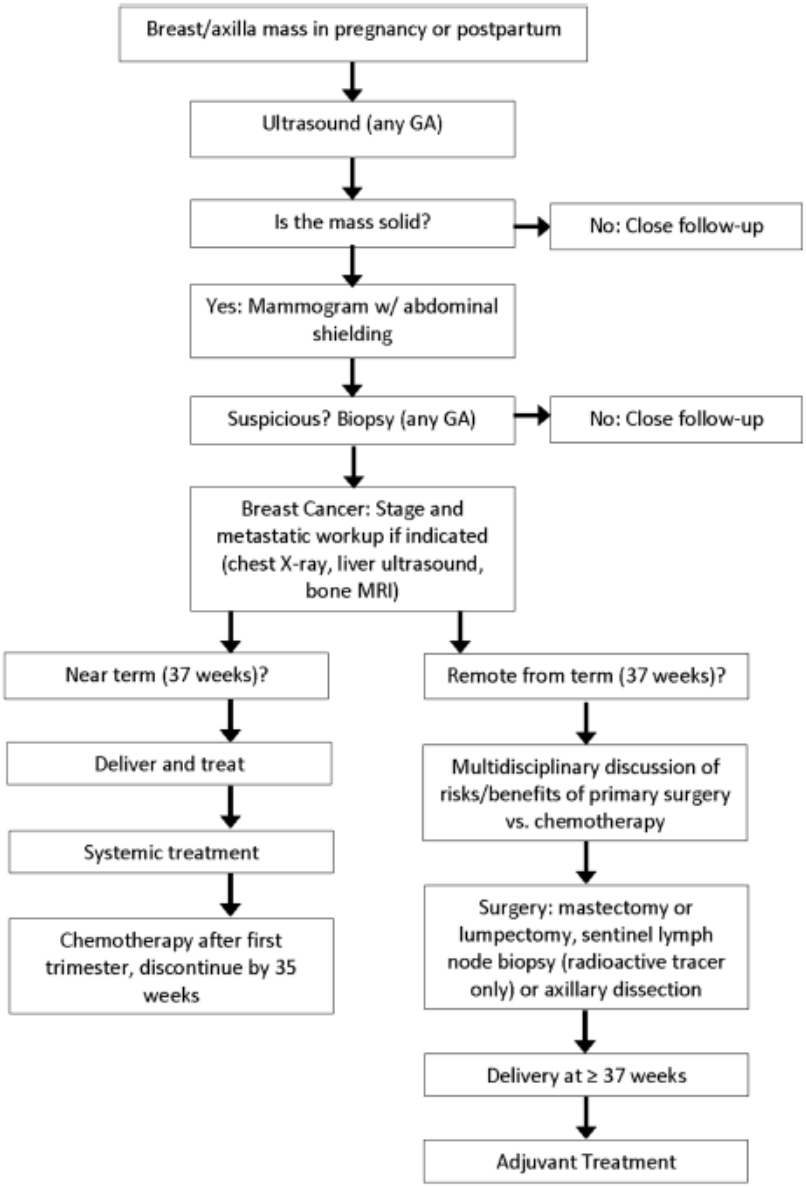

FIG. 1. Algorithm for the diagnosis and treatment of pregnancy-associated breast cancer.

Figure 1. Proposed algorithm for the diagnosis and treatment of PABC. Figure from Rojas 2019.

Appendix V Drugs used to treat event

CHEMOTHERAPY	**ATC5**
Cyclophosphamide	L01AA01
Methotrexate	L01BA01
Pyrimidine analogues	**L01BC**
Cytarabine	L01BC01
Fluorouracil	L01BC02
Gemcitabine	L01BC05
Capecitabine	L01BC06
Vinca alkaloids and analogues	**L01CA**
Vinblastine	L01CA01
Vincristine	L01CA02
Vinorelbine	L01CA04
Taxanes	**L01CD**
Paclitaxel	L01CD01
Docetaxel	L01CD02
Paclitaxel poliglumex	L01CD03
Cabazitaxel	L01CD04
Anthracyclines	**L01DB**
Doxorubicin	L01DB01
Daunorubicin	L01DB02
Epirubicin	L01DB03
Aclarubicin	L01DB04
Idarubicin	L01DB06
Mitoxantrone	L01DB07
Pirarubicin	L01DB08
Other cytotoxic antibiotics	**L01DC**
Mitomycin	L01DC03
Ixabepilone	L01DC04
Platinum agents	**L01XA**
Cisplatin	L01XA01
Carboplatin	L01XA02
TARGETED THERAPY	
Monoclonal antibodies	**L01XC**
Trastuzumab	L01XC03
Bevacizumab	L01XC07
Pertuzumab	L01XC13
Ado-trastuzumab emtansine	L01XC14
Atezolizumab	L01XC32
Cyclin-dependent kinase (CDK) inhibitors	**L01EF**
Palbociclib	L01EF01
Ribociclib	L01EF02
Abemaciclib	L01EF03
mTOR inhibitors	**L01EG**
Everolimus	L01EG02
Human epidermal growth factor receptor 2 (HER2) tyrosine kinase inhibitors	**L01EH**
Lapatinib	L01EH01

Neratinib	L01EH02
Phosphatidylinositol-3-kinase (Pi3K) inhibitors	**L01EM**
Alpelisib	L01EM03
Poly (ADP-ribose) polymerase (PARP) inhibitors	**L01XK**
Olaparib	L01XK01
Talazoparib	L01XK04
Other antineoplastic agents	**L01XX**
Eribulin	L01XX41
ENDOCRINE THERAPY	
Anti-estrogens	**L02BA**
Tamoxifen	L02BA01
Toremifen	L02BA02
Fulvestrant	L02BA03
Aromatase inhibitors	**L02BG**
Anastrozole	L02BG03
Letrozole	L02BG04
Exemestane	L02BG06
Selective estrogen receptor modulators (SERMs)	**G03XC**
Raloxifene	G03XC01
Bazedoxifene	G03XC02
Lasofoxifene	G03XC03

Appendix VI Procedures used to treat event

Radiation therapy	**ICD9**
Therapeutic Radiology and Nuclear Medicine	92.20 – 92.29
Stereotactic Radiosurgery	92.30 – 92.39
Intra-Operative Radiation Procedures	92.40 - 92.41
	ICD10
Encounter for antineoplastic radiation therapy	Z51.0
	CPT
Radiation treatment delivery	77401, 77402, 77407, 77412
Radiation treatment delivery (G codes)	G6003-G6014
IMRT treatment delivery	77385-77386
IMRT treatment delivery (G codes)	G6015-G6016
Port images	77417
IGRT	77387
IGRT (G codes)	G6001, G6002, G6017
CT Guidance	77014
Proton treatment delivery	77520-77525
Neutron beam treatment delivery	77422-77423
SRS treatment delivery	77371-77372
SBRT treatment delivery	77373
Hyperthermia	77600-77620
LDR Brachytherapy	77778
HDR Brachytherapy	77770-77772
Electronic Brachytherapy	0394T-0395T
IORT	77424-77425
Surface application of radiation source	77789
Infusion or installation of radioelement solution	77750
Intracavitary radiation	77761-77763
Supervision and handling	77790
	OPCS
Radiotherapy delivery	X65
Introduction of removable radioactive material into organ	Y35
Introduction of non-removable material into organ NOC	Y36
External beam radiotherapy	Y91
	NCSP
Systemic radiotherapy	WA010, WA019, WA029, WA039, WA099

Radiotherapy for primary tumour	WF001, WF002, WF003, WF004
Surgery	**ICD9**
Operations on the breast	85.0-85.99
Breast reconstruction	S2066-S2068
	ICD10
Acquired absence of breast and nipple	Z90.1x
	CPT
Mastectomy	19160, 19162, 19180, 19182, 19200, 19220, 19240, 19301 - 19307
Breast reconstruction	19340, 19342, 19357, 19361, 19364, 19366, 19367, 19368, 19369
	OPCS
Partial excision of breast	B28.2
Total excision of breast	B27.X
Breast reconstruction	B29.1, B29.3, B29.8, B30.1, B30.8, B30.9, T85.2, T86.2, T87.3, T91.1, B39.1, B39.2, B39.8, B39.9, B39.3, S48.2
	NCSP
Partial excision of mammary gland	HAB00, HAB10, HAB20, HAB30, HAB40, HAB99
Mastectomy	HAC10, HAC15, HAC20, HAC25, HAC30, HAC99
Reconstruction of breast	HAE00, HAE05, HAE10, HAE20, HAE99
Operations for local recurrence of breast cancer	HAF00, HAF10, HAF20, HAF99
Chemotherapy Administration	**ICD9**
Encounter for antineoplastic chemotherapy and immunotherapy	V58.1

Convalescence following chemotherapy	V66.2
Follow up exam following chemotherapy	V67.2
Injection of infusion of cancer chemotherapeutic substance	9925
	ICD10
Encounter for antineoplastic chemotherapy	Z51.1
	CPT
Chemotherapy administration	96400 - 96549
Drug code injections	J8000 – J9999
Chemotherapy administration other than infusion	Q0083-Q0085
Intralesional injection	11900, 11901
	OPCS
Procurement of chemotherapy	X70, X71
Chemotherapy delivery	X72, X73
Intravenous chemotherapy	X35.2
Continuous intravenous infusion of therapeutic substance NEC	X29.2
High cost drugs	X81-X98
	NCSP
Chemotherapy for local primary tumour	WP101, WB103, WB111, WB113, WB121, WB123, WB131, WB133, WB201, WB203, WB211, WB213, WB221, WB223, WB301, WB303, WB311, WB313, WB321, WB323, WB401, WB402, WB501, WB502, WB600, WB610
Chemotherapy for metastized tumour	WD105, WD115, WD125, WD135, WD205, WD215, WD225, WD305, WD315, WD325, WD405, WD415, WD505, WD515

Appendix VII Experience of participating data sources to extract breast cancer

Datasource	**Do you have any experience or expertise to share with respect to extracting this event from the data sources you have access to?**	**Do you have publications where this event was defined? if so can you indicate the PMID? If it is grey literature, can you indicate a link?**	**Can you briefly describe the algorithm(s) you have used to define this event? Please feel free to refer to the data dictionary you have shared earlier with WP7.**	**Can you share any additional comments with respect to this event? In particular: lessons learnt, strengths, weaknesses of the data sources you have access to identifying this event in the corresponding population**	**Have you conducted validation studies, or do you have any quantitative information on validity of this event or it occurrence in the population underlying he databases you have access to?**
05_University_of_Dundee	Yes	22797844; 17855094; 15767381; 11297648; 11290637	22797844: Clinical Practice Research Datalink (patients' demographics, medical diagnoses, referrals to consultants and hospitals, and primary care prescriptions); 17855094: linkage between five breast cancer trials databases and the Scottish Cancer Registry (SCR) 15767381: linked database of acute	Cancer registration is reasonably robust in Scotland so we can be confident on the reliability of these data. CPRD diagnoses depend on whether or not they have been linked to hospitalization data.	We have not personally validated cancer registration data, but we believe that there is a 1% audit of this at ISD.

			hospital discharge (SMR01) records, cancer registrations, maternity (SMR02) database and death records in Scotland 11297648: Scottish Cancer Registry		
19_SIDIAP	Yes	PMID: 31819655	We use the ICD-10 C50 to identify breast cancer	Breast cancer diagnosis has been validated in SIDIAP against data from population-based cancer registries in Catalonia.	Please check the publication mentioned. Sensitivity for breast cancer was found ot be of 89%.
33_THL	Yes	31199509, 22815141, 18226204, 28795403	ICD-10 codes C50 ICD-O-3 topography C50 and ICD-O-3 morphology <9590 excluding 9140 and behaviour code 3.	Registration and coding rules of multiple primaries for pairwise organs must be considered when international research projects use data from the cancer registries.	28350996, 29882462
34_USWAN	Yes	Jordan S. Knight J., Jones J. 2005 Prescription drugs: uses and effects: cytotoxics, disease control. Nursing Standard; 19(27); S1-2*	Read codes		These diagnoses are in the SAIL databases.

Appendix VIII: Covariate items across DAPs

1=THL, 2=UK Sail, 3=FISABIO, 4=GePARD, 5=Scotland
DP5 information items considered important (x) or nice to have (red box)

Information item	1	2	3	4	5	DP 5
Pregnancy timing						
Pregnancy timing	☒	☒	☒	☒	☐	☒
Medication exposure						
Source of medication information						
Primary care/General practitioner	☒	☒	☒	☐	☐	☒
Inpatient	if captured from cancer/ patient registry	☐	☐	☐	☐	☒
Outpatient specialist	if captured from cancer/ patient registry	☐	☒	☐	☐	☒
Prescription records (prescribed or dispensed)	☒	☒	☒	☒	☐	☒
Private prescriptions – private healthcare	☐	☐	☐	☐	☐	☐
Maternal self-report	☐	☐	☐	☐	☐	☐
Details of medication						
Name/ATC code of medication of interest	☒	☒	☒	☒	☐	☒
Date of issued/dispensed prescription, administration or used	☒	☒	☒	☒	☐	☒
Strength	☐	☒	☐	☒	☐	☐
Dosage – amount taken per day	☐	☐	☐	☐	☐	☐
Frequency – per day	☐	☐	☐	☐	☐	☐
Formulation (oral, injection, cream etc).	☐	☒	☐	☒	☐	☐
DDD dispensed	☐	☐	☐	☒	☐	☐
Quantity prescribed or dispensed (tablets)	☒	☐	☐	☒	☐	☐
Prescriber specialty	☒	☐	☐	☒	☐	☐
Co-medications	☒	☒	☐	☒	☐	☐
Maternal disease/medication indication						
Diagnosis						
Diagnosis in healthcare database e.g. ICD10	☒	☒	☒	☒	☐	☒
Diagnosis in disease registry	☒	☒	☒	☐	☐	☒
Type of ward where the diagnosis was given	☐	☐	☒	☐	☐	☐
Intervention in healthcare database as surrogate for disease	☒	☒	☒	☐	☐	☒
Healthcare admission as surrogate for disease/disease severity	☒	☒	☒	☒	☐	☐
Severity of disease						
Health care visit pattern	☒	☒	☐	☐	☐	☐

Co-morbid diagnosis/diagnoses	☒	☒	☒	☒	☐	☒
Co-morbidity – Infection	☒	☒	☒	☒	☐	☐
Outcomes						
Maternal pregnancy outcomes						
Spontaneous abortions	☐	☐	☒	☒	☐	☐
Termination of pregnancy – elective	☒	☐	☐	☒	☐	☒
Termination of pregnancy - for fetal anomaly	☒	☒	☐	☒	☐	☒
Pregnancy related conditions e.g. GD, preeclampsia, hypertension	☒	☒	☒	☒	☐	☒
Mode of delivery	☒	☒	☒	☒	☐	☒
Maternal death	☒	☒	☒	☒	☐	☒
Maternal diagnoses postpartum (e.g. stroke, infection, psychosis, death)	☒	☒	☒	☒	☐	☒
Perinatal outcomes						
Live birth: normal	☒	☒	☒	☒	☐	☒
Stillbirth	☒	☒	☒	☒	☐	☒
Neonatal death	☒	☒	☒	☒	☐	☒
Major congenital anomalies	☒	☒	☒	☒	☐	☒
Gestational age at delivery/preterm birth	☒	☒	☒	☒	☐	☒
Small for gestational age/ IUGR	☒	☒	☐	☒	☐	☐
Birth weight	☒	☒	☒	☒	☐	☒
Head circumference	☒	☐	☐	☐	☐	☐
Length at birth	☒	☐	☐	☐	☐	☐
Apgar score (5, 10 minutes)	☒	☒	☐	☐	☐	☐
Admission to Neonatal Intensive Care Unit	☒	☐	☒	☐	☐	☐
Childhood outcomes						
Death - infant or childhood	☒	☒	☒	☒	☐	☒
Health visitor/public health nurse records	☒ 2012->	☒	☐	☐	☐	☐
Growth in childhood	☐	☐	☐	☐	☐	☐
Diagnosis in a specialist disease registry	☐	☐	☐	☐	☐	☐
Healthcare diagnosis records – ADHD, ASD	☒	☒	☐	☒	☐	☐
Referrals to specialists	☒	☐	☐	☐	☐	☐
Hospital admissions during childhood	☒	☒	☒	☒	☐	☐
Childhood prescriptions	☐	☒	☒	☒	☐	☐
Registered disability in child	☐	☐	☐	☐	☐	☐
Academic results and school performance	☐	☒	☐	☐	☐	☐
Special educational needs/educational support	☐	☒	☐	☐	☐	☐
Psychometric measurements	☐	☐	☐	☐	☐	☐
Confounders/covariates						
Folic acid - pre-conception, first trimester, none	☐	☐	☒	☒	☐	☐
Assisted conception	☒	☐	☒	☒	☐	☐
Maternal age at delivery	☒	☒	☒	☒	☐	☒
Maternal socioeconomic status –or	☒	☒	☒	☒	☐	☐

occupation, employment, income, education etc.						
Smoking status – prior to/ during pregnancy	☒	☒	☒	☒	☐	☐
Alcohol consumption – during pregnancy	☐	☒	☒	☒	☐	☐
Substance misuse services used - during pregnancy	☐	☒	☒	☒	☐	☐
Body mass index	☒ 2005->	☒	☐	☐	☐	☐
Parity	☒	☒	☒	☐	☐	☒