

# BMJ Open Long-term prognosis of young breast cancer patients ( $\leq 40$ years) who did not receive adjuvant systemic treatment: protocol for the PARADIGM initiative cohort study

Gwen MHE Dackus,<sup>1,2</sup> Natalie D ter Hoeve,<sup>2</sup> Mark Opdam,<sup>1</sup> Willem Vreuls,<sup>3</sup> Zsuzsanna Varga,<sup>4</sup> Esther Koop,<sup>5</sup> Stefan M Willems,<sup>2</sup> Carolien HM Van Deurzen,<sup>6</sup> Emilie J Groen,<sup>7</sup> Alicia Cordoba,<sup>8</sup> Jos Bart,<sup>9</sup> Antien L Mooyaart,<sup>10</sup> Jan G van den Tweel,<sup>2</sup> Vicky Zolota,<sup>11</sup> Jelle Wesseling,<sup>7</sup> Anna Sapino,<sup>12,13</sup> Ewa Chmielik,<sup>14</sup> Ales Ryska,<sup>15</sup> Frederic Amant,<sup>16</sup> Annegien Broeks,<sup>17</sup> Ron Kerkhoven,<sup>18</sup> Nikolas Stathonikos,<sup>2</sup> Mitko Veta,<sup>19</sup> Adri Voogd,<sup>20,21</sup> Katarzyna Jozwiak,<sup>22</sup> Michael Hauptmann,<sup>22</sup> Marlous Hoogstraat,<sup>23</sup> Marjanka K Schmidt,<sup>1,24</sup> Gabe Sonke,<sup>25</sup> Elsken van der Wall,<sup>26</sup> Sabine Siesling,<sup>21,27</sup> Paul J van Diest,<sup>2</sup> Sabine C Linn<sup>1,2,25</sup>

**To cite:** Dackus GMHE, ter Hoeve ND, Opdam M, *et al*. Long-term prognosis of young breast cancer patients ( $\leq 40$  years) who did not receive adjuvant systemic treatment: protocol for the PARADIGM initiative cohort study. *BMJ Open* 2017;7:e017842. doi:10.1136/bmjopen-2017-017842

► Prepublication history and additional material for this paper are available online. To view, please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2017-017842>).

PJD and SCL contributed equally.

Received 19 May 2017  
Revised 15 August 2017  
Accepted 20 September 2017



For numbered affiliations see end of article.

**Correspondence to**  
Dr Sabine C Linn; [s.linn@nki.nl](mailto:s.linn@nki.nl)

## ABSTRACT

**Introduction** Currently used tools for breast cancer prognostication and prediction may not adequately reflect a young patient's prognosis or likely treatment benefit because they were not adequately validated in young patients. Since breast cancers diagnosed at a young age are considered prognostically unfavourable, many treatment guidelines recommend adjuvant systemic treatment for all young patients. Patients cured by locoregional treatment alone are, therefore, overtreated. Lack of prognosticators for young breast cancer patients represents an unmet medical need and has led to the initiation of the PATients with bReAst cancer DIagnosed preMenopausally (PARADIGM) initiative. Our aim is to reduce overtreatment of women diagnosed with breast cancer aged  $\leq 40$  years.

**Methods and analysis** All young, adjuvant systemic treatment naive breast cancer patients, who had no prior malignancy and were diagnosed between 1989 and 2000, were identified using the population based Netherlands Cancer Registry (n=3525). Archival tumour tissues were retrieved through linkage with the Dutch nationwide pathology registry. Tissue slides will be digitalised and placed on an online image database platform for clinicopathological revision by an international team of breast pathologists. Immunohistochemical subtype will be assessed using tissue microarrays. Tumour RNA will be isolated and subjected to next-generation sequencing. Differences in gene expression found between patients with a favourable and those with a less favourable prognosis will be used to establish a prognostic classifier, using the triple negative patients as proof of principle.

**Ethics and dissemination** Observational data from the Netherlands Cancer Registry and left over archival patient material are used. Therefore, the Dutch law on Research

## Strengths and limitations of this study

- Large cohort of adjuvant systemic treatment-naive patients, ideal for answering prognostic questions
- Clinical data were derived from the Netherlands Cancer Registry, a large prospective registry with reliable data on many clinical variables.
- Cause of death is unknown, as a result, breast cancer-specific survival cannot be addressed.
- Adjuvant treatment allocation in the 1990s was based on nodal status; our cohort of untreated patients consists of mainly node negative patients, results are therefore not generalisable to young patients with node positive disease.

Involving Human Subjects Act (WMO) is not applicable. The PARADIGM study received a 'non-WMO' declaration from the Medical Ethics Committee of the Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, waiving individual patient consent. All data and material used are stored in a coded way. Study results will be presented at international (breast cancer) conferences and published in peer-reviewed, open-access journals.

## INTRODUCTION

Breast cancer research has increased over the last few decades and led to the development and implementation of multiple novel (neo) adjuvant systemic treatment modalities. The development of these treatment modalities came at a price, with both treatment and developmental costs skyrocketing in recent

years. Biomarkers predicting sensitivity or resistance for these treatments have been developed in parallel, enabling the optimal selection of patients for treatment. Clinical and genetic prognosticators, on the other hand, have been developed to avoid overtreatment and associated side effects among patients who do not require systemic treatment, thereby reducing unnecessary costs. These prognosticators are usually established and validated in relatively small patient populations, often including patients who received adjuvant systemic treatment.<sup>1</sup> Systemic treatment is likely to change the natural course of disease in a significant proportion of patients; prognostic models based on treated patients might therefore produce unreliable risk estimates. Moreover, most prognosticators are tailored towards a specific patient subset, limiting both their applicability and overall prognostic ability.

Young women diagnosed with breast cancer  $\leq 40$  years have not, or only sporadically, been included in studies establishing and validating prognostic models.<sup>1</sup> Breast cancers in young women differ from those in older women, both genetically and clinicopathologically.<sup>2–4</sup> They are more often hereditary, larger in size, high grade, oestrogen receptor (ER) and progesterone receptor (PR) negative, human epidermal growth-factor receptor 2 (HER2) positive and express different genes compared with breast cancers in older patients.<sup>5–8</sup> These unfavourable characteristics usually translate into a shorter time to distant recurrence and worse overall survival (OS).<sup>9</sup> It is therefore not surprising that risk estimates derived from patient series of older women produce unreliable and inconsistent results for young breast cancer patients. Adjuvant!Online, for instance, was reported to overestimate OS in 4%–30% of young patients.<sup>1,10</sup> The realisation that young women do worse compared with their older counterparts has, in the absence of reliable prognostic means, led to the incorporation of age in current treatment guidelines. Early guidelines recommended adjuvant systemic therapy for patients with nodal involvement only.<sup>11</sup> At the fifth international conference of the International Consensus Panel on the Treatment of Primary Breast Cancer, in 1995, it was acknowledged that patients with a young age at diagnosis were considered to have a high risk of recurrence, although an exact age threshold was not defined.<sup>12</sup> It was not until the next meeting in 1998 that age was accepted as a factor that influences treatment decision-making.<sup>13</sup> Currently, Dutch guidelines recommend systemic treatment for all women diagnosed with breast cancer aged  $< 35$  years with tumours any size, except for those with histological grade 1, provided that these tumours are  $\leq 1$  cm in diameter.<sup>14</sup> In the Netherlands, this approach led to a four-fold increase in adjuvant systemic therapy use in young breast cancer patients between 1990 and 2006.<sup>15</sup>

Although the crude incidence ratios from the Netherlands Cancer Registry (NCR) show an increase in breast cancer incidence from 20.08/100 000 women aged 15–39 years in 1989 to 27.18/100 000 in 2015 for the same age groups,<sup>16</sup> the number of breast cancer-specific deaths

in women 15–39 years decreased from 3.79/100 000 in 1989 to 2.12/100 000 in 2015.<sup>16</sup> This decrease in overall mortality might be caused by the increased use of (more effective) adjuvant systemic therapies and preventive measures taken for BRCA germline mutation carriers. In addition, an increased breast cancer awareness among young women, leading to earlier tumour detection in a group of patients that are not eligible for the regular national breast cancer screening programme, inviting only 50–75-year olds, may have contributed to the better outcome of young breast cancer patients as well.

Although their breast cancers are more aggressive in general, a large proportion of young patients have a favourable prognosis and do not require adjuvant systemic therapy following locoregional treatment.<sup>17</sup> The current dilemma is that we cannot tell upfront which patient is ‘low risk’ and which patient is ‘high risk’. Adjuvant systemic treatment of ‘low risk’ young women is considered overtreatment and might induce certain age-related side effects, for instance ovarian suppression and the emergence in time of second primary malignancies, while requiring specific measures such as fertility preservation.<sup>9</sup> To reduce the overtreatment of young breast cancer patients, we need accurate prognosticators which are able to discriminate patients with an extremely favourable prognosis from those with a less favourable prognosis. To do so, a large cohort of systemic treatment naive patients is mandatory. To that end the PATients with bReAst cancer DIAGnosed preMenopausally (PARADIGM) initiative was established.

## METHODS

### Objectives

PARADIGM started in 2012 as part of the Netherlands Breast Cancer Project (NBCP). NBCP aims to become the population-based counterpart of the Early Breast Cancer Trialists’ Collaborative Group, containing population based data from the NCR combined with tissue banking. PARADIGM entirely focuses on the prognostication of young ( $\leq 40$  years at diagnosis) breast cancer patients. Therefore, only women who have not received (neo)adjuvant systemic or hormonal treatment will be included. PARADIGM has set the following initial objectives:

- ▶ *To estimate the true prognosis of breast cancers arising in women  $\leq 40$  years at diagnosis.* The PARADIGM cohort will contain full clinical and long-term follow-up data of all young, Dutch adjuvant systemic treatment naive breast cancer patients diagnosed between 1989 and 2000. Such a cohort is ideal for answering prognostic questions since outcome is known in the absence of systemic treatment effects.
- ▶ *To develop a combined clinicomolecular prognosticator for the triple-negative (TN) patient subset, using clinical characteristics combined with the differences in gene expression between patients with varying outcomes (OS, Recurrence Free Survival).* The TN patients will be used as a proof of principle since they form an ideal group of patients because of their virtually non-existent recurrence

rates after 10 years.<sup>18</sup> When successful, this approach will be expanded to the other immunohistochemical breast cancer subtypes.

- ▶ *To establish an integrated online image database platform for the pathological revision of all cases.* The established database may be used for future teaching and research purposes.

Numerous research questions can be answered with a cohort like PARADIGM. Therefore, after concluding the above-mentioned initial phase of data and tissue acquisition, we will seek funding to:

- ▶ *Develop prognostic models directly derived from the image data using deep learning approaches.*<sup>19 20</sup>
- ▶ *Estimate the prognostic performance of established clinical and molecular prognosticators on breast cancer patients ≤40 years.*
- ▶ *Estimate the prognosis of women with breast cancer during pregnancy in the absence of adjuvant systemic treatment.*
- ▶ *Find drivers and potentially actionable mutations of breast cancers arising in young women for the development of novel targeted therapies.*

### Patient selection and sample collection

Patients for the PARADIGM study have been selected through collaboration with the NCR. The NCR is a population-based prospective cancer registry, with nationwide coverage since 1989. The NCR registers all newly diagnosed, pathologically confirmed malignancies in the Netherlands. Detailed information on patient, tumour and treatment characteristics are collected from hospital records by trained registrars. For PARADIGM, all 3,525 (neo)adjuvant systemic treatment-naive breast cancer patients ≤40 years, diagnosed between 1989 and 2000 with non-metastatic, primary invasive, histologically proven, T<sub>any</sub>N<sub>any</sub>M<sub>0</sub> tumours have been selected. In this cohort, lymph node-negative patients are over-represented since adjuvant treatment allocation in the '90s was mostly based on nodal status.

OS data are available and yearly verified through linkage with the municipality population register; cause of death however is unknown. Recurrence data are not systematically recorded by the NCR. Therefore, NCR registrars will go back to check the individual hospital records of all patients to complete missing information as well as data on disease recurrence (local, regional, distant) and secondary primaries.

Patients identified by the NCR will be coupled anonymously to the nationwide pathology registry in The Netherlands (PALGA)<sup>21</sup> by a trusted third party to obtain tissue specimens. We aim to collect formalin fixed paraffin embedded (FFPE) blocks from normal tissues, tumours, lymph nodes and lymph node metastases whenever available. Based on previous experience, we expect to be able to retrieve tissue samples for >75% of patients.

If we want to assess the prognosis of treatment-naive women with breast cancer during pregnancy, a linkage with either Statistics Netherlands or the Dutch Data Protection Authority is mandatory since information

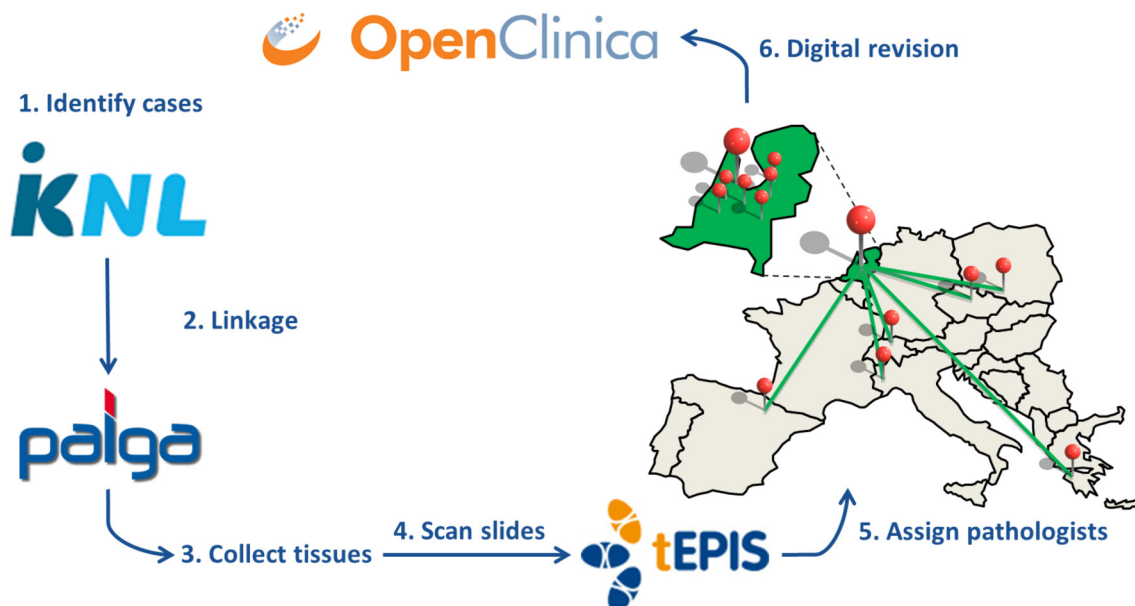
pertaining to pregnancies and their outcome are not registered by the NCR.

### Pathological and molecular characterisation

Fresh slides will be cut and stained with haematoxylin and eosin (H&E), from all FFPE blocks collected, for two reasons. First, old H&E slides often do not represent the remaining tissue in a block. Second, over the years slide quality may deteriorate complicating proper digitalisation.

All slides will be scanned onto the online trait Enhanced Pathology Image Sharing (tEPIS) system using the Philips ultrafast scanner 1.6.1.3 RA (Philips, Amsterdam, The Netherlands) or Nanozoomer XR C12000-21/-22 (Hamamatsu photonics, Hamamatsu, Shizuoka, Japan). The tEPIS system is an online image platform for the sharing of whole slide pathology images among both researchers and pathologists.<sup>22</sup> The images on tEPIS can be integrated on a case-by-case basis in electronic case report forms (eCRF) using a link. These eCRFs will be specifically developed for PARADIGM on the web-based OpenClinica database (OpenClinica, Waltham, Massachusetts, USA).<sup>23</sup> A total of 16 specialised European breast pathologists agreed to revise cases (figure 1) and enter data in OpenClinica. Pathologists will be given data manager rights, in such a way they can be treated as unique study sites, blinding them for the cases and scores of other pathologists. Items scored will be tumour cell percentage, histological subtype, density of tumour infiltrating lymphocytes, histological grade and the presence of lymphovascular invasion, pushing border, fibrotic focus, central necrosis, carcinoma in situ, normal tissue and lymph node (metastases) (see online supplementary table 1).

Tissue microarrays (TMA) will be constructed in duplicate, using the TMA Grandmaster (3DHitech, Budapest, Hungary), whenever sufficient tumour tissue is available according to the H&E slides. When blocks do not fit in the TMA Grandmaster, manual non-duplicated TMAs will be constructed. Three 0.6 mm cores from the tumour centre and periphery will be used per patient, with a maximum of 59 patients per Grandmaster TMA or 55 patients in case of manual TMAs. Immunohistochemistry (IHC) for ER, PR and HER2 will be performed on TMA slides using the Ventana BenchMark ULTRA (Ventana Medical Systems, Basel, Switzerland) to derive clinical immunohistochemical breast cancer subtypes.<sup>24</sup> Whole slides will be cut and stained in case insufficient tumour material is present for TMA construction. Thresholds for ER and PR tumour positivity are set at ≥10% cells staining positive. For HER2, both IHC and silver in situ hybridisation (SISH) will be performed. Tumours will be considered positive on IHC if a strong, complete membranous staining in >10% of tumour cells (IHC-3+) is present. Tumours will be considered positive for HER2 SISH if on average seven or more HER2 copies are present in the nucleus (SISH low and SISH high level amplified). Tumours scoring positive for at least one of the two modalities (IHC or SISH) will be considered HER2 positive.



**Figure 1** PARADIGM workflow. All (neo)adjuvant systemic treatment-naïve breast cancer patients  $\leq 40$  years, diagnosed between 1989 and 2000 with non-metastatic, primary invasive, histologically proven,  $T_{any}N_{any}M_0$  tumours have been selected through the Netherlands Cancer Registry, 3,525 patients in total.<sup>1</sup> These patients will be linked to PALGA (Dutch Surgical pathology Registry),<sup>2</sup> if linkage is possible formalin-fixed paraffin-embedded tissue specimens will be requested from the respective Dutch pathology laboratory.<sup>3</sup> New slides will be cut, H&E stained and digitalised.<sup>4</sup> When all available tissues for a patient are digitalised, the patient's case will be uploaded to the online digital imaging platform tEPIS. The images will be linked on a case-by-case (patient-by-patient) basis to the online PARADIGM database in OpenClinica. A patient case will be appointed to one of 16 specialised breast pathologists who help revise the PARADIGM cases according to the current standard.<sup>5</sup> Pathologists will review the slides in tEPIS and fill out the items asked (see online supplementary table 1) in the PARADIGM OpenClinica electronic case report form.<sup>6</sup> PARADIGM: PATients with bReAst cancer DIAGnosed preMenopausally; tEPIS: trait Enhanced Pathology Image Sharing.

For the triple-negative patients, RNA will be isolated using the AllPrep DNA/RNA isolation kit (QIAGEN, Hilden, Germany), according to manufacturer's protocol. By using the AllPrep method, both RNA and DNA are isolated in the same instance using only 5 to 10  $10\mu\text{m}$  tumour slides, depending on tumour size. Library preparation for RNA sequencing (RNAseq) will be conducted using the TruSeq RNA Access Kit (Illumina, San Diego, California, USA), developed to capture coding RNA sequences of RNA isolated from FFPE. RNAseq will be performed at the Genomics Core Facility of the Netherlands Cancer Institute.<sup>25</sup> Sufficient DNA and RNA will be left for future molecular characterisation and will be stored in the PARADIGM biobank at the Core Facility Molecular Pathology & Biobanking of the Netherlands Cancer Institute.

### Ethical considerations

The PARADIGM initiative will use observational data from the NCR and left over archival patient material. All data and material on the young breast cancer patients involved in this study will be used in a coded way. Neither interventions nor active recruitment of study participants will take place within PARADIGM. As a result, the Dutch law on Research Involving Human Subjects Act (WMO) is not applicable. Therefore, the PARADIGM study received a 'non-WMO' declaration from the Medical Ethics Committee of the Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital (NKI),

waiving individual patient consent, on 31 October 2012 (PTC 12.1489/NBCP project). In addition, approval from the NKI translational research board (TRB) was obtained. The TRB oversees that the Codes of Conduct for data and tissue, jointly published by the Dutch Federation of Biomedical Scientific Societies, the Royal Dutch Medical Association and several patient groups,<sup>26</sup> are adhered to by the respective researchers.

### DISCUSSION

The PARADIGM project is the largest collection of young, adjuvant systemic treatment-naïve breast cancer patients with long-term follow-up. This project provides us with a unique opportunity to study the natural course of breast cancer outcome in the absence of adjuvant systemic treatment. This will form the basis for testing, validating and improving current prognosticators, since most of them were not validated in young breast cancer patients before.

Furthermore, knowing the disease course in the absence of systemic treatment will enable us to discriminate patients with a very favourable prognosis from those with a less favourable prognosis. The clinical and molecular differences between these two patient groups will be used to establish a prognosticator specifically tailored towards young TN breast cancer patients. PARADIGM thus aims to reduce overtreatment of young breast cancer patients through adequate prognostication.

A large part of the project is dedicated to tissue retrieval and pathological characterisation of all the tumours diagnosed. To this end, a digital revision pipeline was established, integrating the tEPIS viewer and an OpenClinica based database. This approach has several advantages over a conventional, hands-on microscopy approach. For instance, it is not necessary to send around slides with the chance of them getting lost or broken. Digitalising all slides enables pathologists to work on the project from every physical location, even from outside of the hospital, as long as there is a computer with decent internet connection available. Sharing slides with pathologists over the internet is easy, secure and it even saves patient material by making better use of already available tissues. Digitalised H&E slides remain stable over time and can be shared with other researchers over and over again without having to cut new slides. The coupling of an online viewer and database limits the amount of errors made because the case-by-case linkage of images to CRFs cannot be interchanged. The OpenClinica-tEPIS set-up may well serve as an example for similar large studies, irrespective of the disease being studied. In addition, the availability of large pathologically annotated digitalised H&E datasets with paired clinical data could be used for teaching purposes and the application of state-of-the-art machine learning techniques to derive prognostic classifiers.

Although revisions go faster because sharing digital slides is easier and more (remote) pathologists can be reached, processing blocks and slides remains a time-consuming endeavour. Depending on the scanner used, it takes between 45 and 60 seconds to digitalise a tissue piece measuring 15 mm × 15 mm. After the scanned images are JPEG compressed, the file size (one slide) averages 1 gigabyte (GB) (spread 50 megabyte–8 GB). Scanning slides and doing the necessary quality checks requires many person hours, specialised equipment and information technology (IT) infrastructure to host, store and access all slides. The in-depth genetic characterisation of all TN cases through sequencing will pose further data storage and privacy challenges.

Pathologists are also faced with new challenges using digitalised H&E slides rather than glass slides. Digital revisions may take a little more time, depending on network speed and familiarity with the system. In addition, mitoses counting on (digital) whole slide images depends on image quality and is considered more challenging compared with glass slides.<sup>27</sup> Finding expert pathologists motivated to participate free of charge and keeping them engaged is a well-known problem and might be challenging within PARADIGM as well.<sup>28</sup>

The PARADIGM patient cohort could be different from young breast cancer patients diagnosed nowadays due to changes in breast cancer awareness, environmental factors and lifestyle, like a higher body mass index, later age of first child birth, fewer children and less/shorter periods of breast feeding. In addition, the proportion of lymph node-positive patients in PARADIGM is probably going to be small, since they were considered prognostically unfavourable in the '90s and usually received adjuvant systemic treatment.

Lastly, finding adequate funding for research projects is always challenging. The historic nature of this study and the delicate subject of prognosis-driven treatment reduction for young breast cancer patients have made funding particularly difficult. Since correctly assessing a patient's prognosis might be cheaper and hence cost-effective over a one-size-fits-all treatment, practice-changing studies like the PARADIGM initiative should be of great interest to governments and policy-makers, for its research, innovation and teaching purposes.

## CONCLUSION

The PARADIGM initiative aims to reduce the overtreatment of young ( $\leq 40$  years) breast cancer patients by the improvement of established and development of novel prognosticators in a large group of adjuvant systemic treatment-naïve patients. The clinical and molecular characterisation of young women diagnosed with breast cancer will fill in current gaps in prognostication and deepen our understanding of breast cancer in women of this age group. The development of a prognostic tool specifically tailored to TN patients will be the first step towards personalised prognostication of young breast cancers. This approach will be expanded to the other intrinsic subtypes as well in the future. The tEPIS/OpenClinica application that was developed can be used for extensive research collaborations and may well serve as an example for similar large studies in other tumour types and non-oncological diseases.

## Author affiliations

<sup>1</sup>Division of Molecular Pathology, Netherlands Cancer Institute, Amsterdam, Noord-Holland, Netherlands

<sup>2</sup>Department of Pathology, University Medical Center Utrecht, Utrecht, Netherlands

<sup>3</sup>Department of Pathology, Canisius Wilhelmina Ziekenhuis, Nijmegen, Gelderland, Netherlands

<sup>4</sup>Institute of Surgical Pathology, University Hospital Zurich, Zurich, Switzerland

<sup>5</sup>Department of Pathology, Gelre Ziekenhuizen, Apeldoorn, Gelderland, Netherlands

<sup>6</sup>Department of Pathology, ErasmusMC Cancer Institute, Rotterdam, Zuid-Holland, Netherlands

<sup>7</sup>Department of Pathology, Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, Amsterdam, Noord-Holland, Netherlands

<sup>8</sup>Department of Pathology, Complejo Hospitalario de Navarra, Pamplona, Spain

<sup>9</sup>Department of Pathology, IsalaKlinieken Zwolle, Zwolle, Overijssel, Netherlands

<sup>10</sup>Department of Pathology, Leids Universitair Medisch Centrum, Leiden, Zuid-Holland, Netherlands

<sup>11</sup>Department of Pathology, Rion University Hospital, University of Patras, Medical School, Patras, Greece

<sup>12</sup>Candiolo Cancer Institute—FPO, IRCCS, Candiolo, Piemonte, Italy

<sup>13</sup>Department of Medical Sciences, University of Torino, Torino, Italy

<sup>14</sup>Department of Tumor Pathology, Maria Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice, Poland

<sup>15</sup>Faculty of Medicine and University Hospital, Charles University in Prague, Hradec Kralove, Czech Republic

<sup>16</sup>Department of Obstetrics and Gynaecology at the Catholic, University of Leuven, Leuven, Belgium

<sup>17</sup>Core Facility Molecular Pathology and Biobanking, Division of Molecular Pathology, Netherlands Cancer Institute, Amsterdam, Noord-Holland, Netherlands

<sup>18</sup>Genomics Core Facility, Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, Amsterdam, Noord-Holland, Netherlands

<sup>19</sup>Medical Image Analysis Group (IMAG/e), Technische Universiteit Eindhoven, Eindhoven, Noord-Brabant, Netherlands

<sup>20</sup>Department of Epidemiology, Maastricht University, Maastricht, Limburg, Netherlands

<sup>21</sup>Department of Research, Netherlands Comprehensive Cancer Organization (IKNL), Utrecht, Utrecht, UK

<sup>22</sup>Department of Epidemiology and Biostatistics, Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, Amsterdam, Noord-Holland, Netherlands

<sup>23</sup>Department of Computational Cancer Biology, Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, Amsterdam, Noord-Holland, Netherlands

<sup>24</sup>Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, Amsterdam, Noord-Holland, Netherlands

<sup>25</sup>Department of Medical Oncology, Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, Amsterdam, Noord-Holland, Netherlands

<sup>26</sup>Division of Internal Medicine and Dermatology, University Medical Center Utrecht, Utrecht, Netherlands

<sup>27</sup>Department of Health Technology and Services Research, MIRA Institute for Biomedical Technology and Technical Medicine, Enschede, Overijssel, Netherlands

**Acknowledgements** The authors would like to thank the registration team of the Netherlands Comprehensive Cancer Organization (IKNL) for the collection of data for the Netherlands Cancer Registry as well as IKNL staff for scientific advice. In addition, the authors would like to thank CTMM-TraIT for providing the IT infrastructure for capturing of clinical research data (TraIT OpenClinica environment), the Dutch Pathology Registry PALGA for providing histopathological data, the PALGA group for FFPE tissue specimens and the NKI-AVL Core Facility Molecular Pathology & Biobanking (CFMPB) for supplying NKI-AVL Biobank material and lab support.

**Contributors** SS, PJD and SCL had the original idea for the study. The study idea was further developed by GMHED, NDH, MO, FA, ABB, RK, NS, MV, ACV and MKS. GMHED wrote the manuscript. All authors listed, critically reviewed the manuscript and approved the manuscript for submission.

**Funding** The authors would like to thank The Netherlands Organisation for Health Research and Development (ZonMW) Project number 836021019, A Sisters Hope, De Vrienden van UMC Utrecht and Mr M Spanbroek for their financial support. Funders had no influence on study design, data collection or project management.

**Competing interests** All authors have completed the Unified Competing Interest form (). Forms are available on request from the corresponding author. Authors declare the following: SCL reports grants from The Netherlands Organization for Health Research and Development (ZonMW) and A Sister's Hope during the conduct of the study. SCL also received non-financial support from AstraZeneca, grants from AstraZeneca, Roche and Genentech and other from Novartis, Cergentis, PhilipsHealth BV, Roche and Astra Zeneca outside the submitted work. In addition, SCL has two patents pending (WO/2015/080585, PCT/NL2014/050813). AR reports grants, personal fees and non-financial support from Pfizer, grants and personal fees from Roche, Astra Zeneca, MSD, BMS, Merck and grants from Boehringer Ingelheim and Novartis outside the submitted work. NS reports grants from University Medical Centre Utrecht during the conduct of the study. SMW reports grants and personal fees from Roche, Pfizer, AstraZeneca and personal fees from MSD outside the submitted work.

**Ethics approval** Medical Ethics Committee (METC) of the NKI.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

## REFERENCES

- Engelhardt EG, Garvelink MM, de Haes JH, *et al.* Predicting and communicating the risk of recurrence and death in women with early-stage breast cancer: a systematic review of risk prediction models. *J Clin Oncol* 2014;32:238–50.
- Partridge AH, Hughes ME, Warner ET, *et al.* Subtype-dependent relationship between young age at diagnosis and breast cancer survival. *J Clin Oncol* 2016;34:3308–14.
- Azim HA, Michiels S, Bedard PL, *et al.* Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling. *Clin Cancer Res* 2012;18:1341–51.
- Anders CK, Hsu DS, Broadwater G, *et al.* Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *J Clin Oncol* 2008;26:3324–30.
- Bharat A, Aft RL, Gao F, *et al.* Patient and tumor characteristics associated with increased mortality in young women (< or =40 years) with breast cancer. *J Surg Oncol* 2009;100:248–51.
- El Saghir NS, Seoud M, Khalil MK, *et al.* Effects of young age at presentation on survival in breast cancer. *BMC Cancer* 2006;6:194.
- Cancellato G, Maisonneuve P, Rotmensz N, *et al.* Prognosis and adjuvant treatment effects in selected breast cancer subtypes of very young women (<35 years) with operable breast cancer. *Ann Oncol* 2010;21:1974–81.
- Colak D, Nofal A, Albakheet A, *et al.* Age-specific gene expression signatures for breast tumors and cross-species conserved potential cancer progression markers in young women. *PLoS One* 2013;8:e63204.
- Anders CK, Johnson R, Litton J, *et al.* Breast cancer before age 40 years. *Semin Oncol* 2009;36:237–49.
- Mook S, Schmidt MK, Rutgers EJ, *et al.* Calibration and discriminatory accuracy of prognosis calculation for breast cancer with the online Adjuvant! program: a hospital-based retrospective cohort study. *Lancet Oncol* 2009;10:1070–6.
- Consensus conference. Adjuvant chemotherapy for breast cancer. *JAMA* 1985;254:3461–3.
- Goldhirsch A, Wood WC, Senn HJ, *et al.* Fifth International Conference on Adjuvant Therapy of Breast Cancer, St Gallen, March 1995. International Consensus Panel on the Treatment of Primary Breast Cancer. *Eur J Cancer* 1995;31A:1754–9.
- Zujewski J, Liu ET. The 1998 St. Gallen's Consensus Conference: an assessment. *J Natl Cancer Inst* 1998;90:1587–9.
- Oncoline-Dutch breast cancer guideline. Hormonal therapy (breast cancer). [https://richtlijndatabase.nl/en/richtlijn/breast\\_cancer/adjuvant\\_systemic\\_therapy/hormonal\\_therapy.html](https://richtlijndatabase.nl/en/richtlijn/breast_cancer/adjuvant_systemic_therapy/hormonal_therapy.html) (accessed 20 Nov 2016).
- van der Sangen MJ, Voogd AC, van de Poll-Franse LV, *et al.* [Breast cancer in young women: epidemiology and treatment dilemmas]. *Ned Tijdschr Geneesk* 2008;152:2495–500.
- Netherlands Cancer Registry - Breast cancer incidence. [http://www.cijfersoverkanker.nl/selecties/dataset\\_2/img5866634cb90cf](http://www.cijfersoverkanker.nl/selecties/dataset_2/img5866634cb90cf) (accessed 30 Dec 2016).
- van der Hage JA, Mieog JS, van de Velde CJ, *et al.* Impact of established prognostic factors and molecular subtype in very young breast cancer patients: pooled analysis of four EORTC randomized controlled trials. *Breast Cancer Res* 2011;13:R68.
- Dent R, Trudeau M, Pritchard KI, *et al.* Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res* 2007;13:4429–34.
- LeCun Y, Bengio Y, Hinton G, *et al.* Deep learning. *Nature* 2015;521:436–44.
- Litjens G, Sánchez CI, Timofeeva N, *et al.* Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep* 2016;6:26286.
- PALGA. The nationwide network and registry of histo- and cytopathology in the Netherlands. <http://www.palga.nl/en/> (accessed 22 Nov 2016).
- tEPIS. Pathology imagemanagement and sharing. <http://www.ctmm-trait.nl/traIt-tools/tepis> (accessed 9 Aug 2016).
- TraIT OpenClinica - Open Source for Clinical Research. <http://www.ctmm-trait.nl/traIt-tools/OpenClinica> (accessed 9 Aug 2016).
- Hugh J, Hanson J, Cheang MC, *et al.* Breast cancer subtypes and response to docetaxel in node-positive breast cancer: use of an immunohistochemical definition in the BCIRG 001 trial. *J Clin Oncol* 2009;27:1168–76.
- The Genomics Core Facility experiment database. <http://gcfdb.nki.nl/> (accessed 22 Nov 2016).
- Federa - Codes of Conduct. <https://www.federa.org/codes-conduct> (accessed 9 Aug 2016).
- Al-Janabi S, van Slooten HJ, Visser M, *et al.* Evaluation of mitotic activity index in breast cancer using whole slide digital images. *PLoS One* 2013;8:e82576.
- Shaw EC, Hanby AM, Wheeler K, *et al.* Observer agreement comparing the use of virtual slides with glass slides in the pathology review component of the POSH breast cancer cohort study. *J Clin Pathol* 2012;65:403–8.