Technical Realization and First Insights of the Multicenter Integrative Breast Cancer Registry INTREST

Thomas Ostermann1, Sebastian Unger1, Michaela Warzecha2, Sebastian Appelbaum1, Daniela Rodrigues Recchia1, Holger Cramer3 and Heidemarie Haller4

1Department for Psychology and Psychotherapy, Witten/Herdecke University, Alfred Herrhausen-Straße 50, 58448 Witten, Germany
2Medical Informatics, University of Applied Sciences Dortmund, Emil-Figge-Str. 42 44227 Dortmund, Germany
3Institute for General Practice and Interprofessional Care, University Hospital Tuebingen and Bosch Health Campus, Stuttgart, Auerbachstraße 110, 70376 Stuttgart, Germany
4Department of Internal and Integrative Medicine, Evang. Kliniken Essen-Mitte, Faculty of Medicine, University of Duisburg-Essen, Am Deimelsberg 34a, 45276 Essen, Germany

Keywords: Clinical Registry, Breast Cancer, Database, Baseline Characteristics.

Abstract: Cancer is one of leading causes of mortality worldwide. According to GLOBOCAN database, 19.3 million new cancer cases and 10 million cancer deaths worldwide were counted in 2020. Thus, there is an absolute necessity for statistical data on cancer incidence and treatments. This is mainly done by cancer registries, which aim at collecting, managing, and analyzing health and demographic data on individuals diagnosed with cancer. As more and more patients make use of integrative oncology to optimize their health and quality of life during and after cancer treatment, it is important to gather clinical registry data of complementary as well as conventional cancer care. The INTREST registry is the first approach that aims to identify predictors of treatment-response in women undergoing individualized, integrative breast cancer treatment. This article reports on the technical realization and representativity of the registry based on 3,341 eligible women and 885 cases included in interim statistical analysis. The analyses show that the INTREST sample of women suffering from breast cancer does not significantly differ from population-based registries and pragmatic trial data of breast cancer patients in Germany with respect to main sociodemographic and clinical cancer data. However, completeness, particularly in tumor classification, currently is a major limitation.

1 INTRODUCTION

Cancer is still one of the leading causes of mortality worldwide. According to the GLOBOCAN database, 19.3 million new cancer cases and 10 million cancer deaths worldwide were counted in 2020 (Sung et al., 2021; Ferley et al., 2021). Thus, there is an absolute necessity for statistical data on cancer incidence and treatments. This is mainly done by cancer registries, which aim at collecting, managing, and analyzing health and demographic data on persons diagnosed with cancer (Jensen et al., 1991). Cancer registries can be classified into three general types:

1. Hospital based registries, which maintain data on all patients diagnosed and/or treated for cancer at their facility and report cancer cases to the central or state cancer registry as required by law.

2. Population-based central registries, which collect data on all cancer patients within certain geographical areas.

3. Special purpose registries, providing data on a particular type of cancer and/or treatment.

The INTREST cancer registry belongs to the third class of registries and collects data of women diagnosed with breast cancer with a special focus on integrative oncological treatment approaches. Integrative Oncology has its origins in the United States and per definition combines conventional cancer care with evidenced-based complementary therapies (CM). The main goal of Integrative Oncology is to reduce side effects of oncological treatments and to improve patient’s quality of life with a first medical guideline being published in 2007.

Common symptoms, accompanying with the diagnosis and treatment of cancer, include fatigue,
sleep disturbances, pain, neuropathy, and affective disorders (Cheng et al., 2013; Patrick et al., 2004; Singer et al., 2021). In order to improve quality of life, women with breast cancer frequently use CM (Molassiotis et al., 2005; Boon et al., 2007). However, patients often do not mention the use of CM to their physicians unless they are explicitly asked about it (Koenig et al., 2015; Samuels et al., 2017). This lack of communication can lead to undesired interactions between conventional and CM therapies that, at worst, negatively impacts quality and quantity of life (Alsana et al., 2014; Ben-Arye., 2015; Bode & Dong, 2015; Zeller et al., 2013).

Asking patients and systematically exploring their concurrent CM use is recommended by international clinical practice guidelines (Greenlee et al., 2014; Leitlinienprogramm Onkologie, 2017; Lyman et al., 2018) but implemented only gradually by physicians (Paepke et al., 2020; Grimm et al., 2021) and initial registries (Schad et al., 2013; Dusek et al., 2016).

Standard clinical cancer registries, in contrast, usually do not assess data beyond tumor characteristics, conventional treatment algorithms, and patient survival while other supportive treatments, strengthening the physical and psychosocial resilience of cancer survivors, are not yet included.

This article reports on the technical realization and first results of the data analysis of the INTREST registry, which aims at assessing data on the influence of conventional and CM treatments as well as physical and psychosocial resilience using qualitative and quantitative endpoints.

2 MATERIAL AND METHODS

2.1 Guidelines, Ethics and Partners

The INTREST registry uses an epidemiological, multi-center cohort design according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) and the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) statement (Collins et al., 2015). The INTREST protocol is approved by the respective ethics committees, registered at the World Health Organization (WHO) International Clinical Trials Registry Platform / German Clinical Trials Register (DRKS00014852), and published in 2021 (Haller et al., 2021).

2.2 Technical Realization

In the initial phase, INTREST was developed for local use. The basis was a Windows 10 machine with the XAMPP package installed, an Apache distribution with a MySQL-Database and the scripting language PHP (PHP: Hypertext Preprocessor) using an architecture used in the medical learning context (Ostermann et al., 2018). To make INTREST accessible online, it was migrated to a server that was already fully set up with a similar operating system and software to those of the local machine, where INTREST was previously running. Thus, design and structure, which are briefly presented below, could be retained during the migration without any complications.

2.3 Data Model

All data are stored within a MySQL-Database. The structure of the relational model is provided in Figure 1.

![Figure 1: Relational model of the data.](image)

For reasons of clarity, only one table is shown as representative for all. The name of the table is only a placeholder for the actual names.

The structure of a table mostly follows the same principle: The first column contains the primary key, which consists of an integer value and is automatically incremented when a new entry is created. This is followed by the foreign key column, which is not required only for the account table. Finally, there are two columns for an alphanumeric input in the form of a limited number of characters in the regular tables,
while there are three such columns in the tables for accounts and patients. In general, these columns represent a survey item and the corresponding response. The patient table uses these columns for different types of IDs (identifiers), which are unique in their three-way combination and consist of a sequential center number, an individual five-digit patient code, and a patient identification number of the specific clinic. In the account table, on the other hand, these columns are related to username, password, and name of the study center. Another characteristic of the account table is that two of its columns are unique, since each study center receives only one account with one user.

2.4 System Architecture

INTREST’s architecture can be summarized in three main components (see Figure 2) that follow the principle of a layer concept for smaller application (Richards, 2015). On top, there is a GUI (Graphical User Interface) that allows the communication between client and server through internet. With this study nurses of a center can access INTREST at any time to create new patients or add their individual Case Report Forms (CRFs). Each page of the GUI uses a different script because a page refers to a specific point in time when various items are recorded. After the CRF inputs are transmitted to a script, they are transformed to MySQL queries and redirected to the database at the bottom of INTREST’s architecture. If a page is accessed with data already entered, a message appears stating that the data already exists and can no longer be entered.

2.5 Data Security and Validity

In any application that is connected to the Internet and contain data, especially if it is medical or personal data as with INTREST, certain security precautions are necessary. Hoque et al. (2014) describe many diverse network attacks and that these attacks often target web sites or databases to gather information. Therefore, approaches should be applied to reduce the risk of exposing data in network applications, which might have security issues or process medical data. HTTPS (Hypertext Transfer Protocol Secure) is one of the worldwide used approaches to encrypt the communication between clients and servers. The server, hosting INTREST, supports this protocol in conjunction with an officially authorized certificate, allowing secure data retrieval and transmission.

Since the data is transferred from collected medical records in paper form, special attention is put on this issue. Only medical personnel who have been trained by the respective study center are authorized for this. Their tasks are the formal monitoring for completeness and the input of the paper CRFs into the specially developed GUI. Therefore, they are supported on the software side.

First, entered data is validated, using programmed validation checks, e.g., checks for required values, item types, and item ranges. And second, if a response of an item is not recognizable, this item is stored in the database with a discrepancy note. In regular data review meetings, all such discrepancies are discussed and clarified by comparing the entries in the database with the source data.

Another aspect of data security concerns the storage of data. Since it should not be possible to draw conclusions about an individual participant, the data are exclusively pseudonymized during transmission to the registry. This even applies to the statistical analysis, where pseudonymized data is transferred to a CSV (Comma-Separated Values) file. At this stage, the data is only checked for accuracy and completeness by randomly comparing a set of items with the original database.

2.6 Patients and Outcomes

Female patients diagnosed with primary breast cancer stage I-III according to the pTNM (pathological Tumor-Node-Metastasis) classification, who received individualized integrative cancer treatments in one of the participating study centers, were included in the registry. Cancer diagnosis and treatment data as well as those on progression were retrieved from medical records, while women were asked to complete
sociodemographic data and the following Patient Reported Outcomes (PROs);

- Cancer-related quality of life and fatigue, assessed by the Functional Assessment of Cancer Therapy General (FACT-G) (Brucker et al., 2005) and the associated Fatigue Scale (FACIT-F) (Yost & Eton, 2005),
- Distress assessed by the Questionnaire on Distress in Cancer Patients (QSC) (Book et al., 2011),
- Depression assessed by the Center for Epidemiologic Studies Depression Scale (CESD) (Stafford et al., 2014),
- Hopelessness assessed by the Brief Hopelessness measure (BH) (Fraser et al., 2014),
- State anxiety assessed by the Patient-Reported Outcomes Measurement Information System Emotional Distress Anxiety Form (PROMIS-EDA) (Schaelet et al., 2016) and progression anxiety assessed by the Fear of Relapse/Recurrence Scale (FRRS) (Thewes et al., 2012),
- Emotion regulation assessed by the Emotion Regulation Questionnaire (ERQ) (Gross et al., 2003),
- Sleep disturbance assessed by the Patient-Reported Outcomes Measurement Information System Sleep Disturbance Form (PROMIS-SD) (Yu et al., 2011),
- Spiritual well-being assessed by the Functional Assessment of Chronic Illness Spiritual Well-Being Scale (FACIT-SP) (Bredle et al., 2011),
- Social support assessed by the perceived Available Support subscale of the Berlin Social Support Scales (BSSS) (Schulz et al., 2003),
- Physical activity assessed by International Physical Activity Questionnaire (IPAQ) (Craig et al., 2003),
- Healthy diet assessed by the Mediterranean Diet Adherence Screener (MEDAS) (Schroder et al., 2011),
- CM attitudes assessed by the CAM Health Belief Questionnaire (CHBQ) (Lie et al., 2004),
- Interest in CM assessed by a numeric rating scales (NRS),
- Use of CM assessed by an extended version of the International Complementary and Alternative Medicine Questionnaire (I-CAM-Q) (Quant et al., 2009),
- Adverse events assessed by the Memorial Symptom Assessment Scale (MSAS) (Chang et al., 2000) and
- Therapy satisfaction assessed by the Client Satisfaction Questionnaire (CSQ) (Attkisson et al., 1982).

2.7 Statistical Analysis

Statistical analysis included univariate analyses of frequencies using Chi-Square statistics and analyses of mean differences using a t-test with respect to group differences. For all analyses, due to the high sample size, a p-value of .01 was considered to be significant.

3 RESULTS

Originally developed at the Department of Internal and Integrative Medicine, KEM, University of Duisburg-Essen and the KEM Breast Unit (Start in September 2017), three additional German cancer centers have joint into the INTREST-registry: the Department of Gynecology at the Robert-Bosch-Hospital (Stuttgart in January 2018), the Breast Unit of the St. Franziskus-Hospital (Münster in September 2019), and the Breast Unit of Hall (Hall in November 2020).

The recruitment in the four study centers of the INTREST project amounts to $N = 1373$ patients with TNM I-III breast cancer of which $N = 885$ were eligible for the present interim analysis at baseline.

For the individual study centers, patient recruitment results are presented in Figure 3.

3.1 Sociodemographic Data

The mean age at baseline is $57.0 \pm 11.6$ years, with the vast majority born in Germany (92 %). 62 % of the patients are married and live with their spouse. The average weight and height are $72.2 \pm 16.2$ kg (kilogram) and $167.3 \pm 6.3$ cm (centimeters), corresponding to a Body Mass Index (BMI) of $25.8 \pm 5.5$. More than half of the sample (61.2 %) is employed. In addition, almost half of the sample had a high school (18.1 %) or university degree (30.0 %).

3.2 Cancer Parameters

Table 1 provides cancer related baseline values compared to similar cohort studies and representative population data from a German/Saarland cancer registry.

With respect to the age at first cancer diagnosis, the INTREST data are significantly lower compared to population data of the Saarland cancer registry ($p$
which, however, includes not only breast cancer but mixed cancer diagnoses. Tumor type shows similar percentages compared with the Saarland registry \((p = 0.017)\), while the distribution of tumor receptor subtypes is not comparable to the TMK cohort \((p < 0.001)\) (Marschner et al., 2019). However, it has to be noted that the TMK cohort is significantly younger than the INTERST sample \((p < 0.001)\), as only women with early breast cancer were included.

Tumor staging is comparable between INREST and the CM trial \((p = 0.11)\;\text{Witt et al., 2015}\), while INTERST shows significantly different percentages compared to the Saarland registry and the TMK cohort \((p < 0.001)\), respectively. Tumor grading significantly differ between the samples \((p < 0.001)\), respectively except the amount of G2 grading.

Status of menopause in the INTERST registry does not significantly differ from the TMK cohort \((p = 0.14)\) and the pragmatic CM trial \((p = 0.15)\).

### 3.3 Quality of Life

Quality of life measured with the FACT-G and FACIT-F showed comparable values with respect to other studies.

Figure 4 displays the FACT-G total score distribution together with the median and interquartile range (IQR) of the female cancer norm (Brucker et al., 2005).

With a mean value of \(78.8 \pm 15.7\) and a median of \(80.7\) (IQR: [68.2; 91.0]) the FACT-G shows an expected distribution. This value is underpinned when comparing it to other, e.g., with the mean value of 76.2 of the trial of Witt et al. (2015) or with the mean of 75.7 \(\pm 15.7\) of the TMK cohort (Marschner et al., 2019), both presented in Table 1.

Figure 5 displays the FACIT-F total score distribution. In contrast to figure 4, there are no quartile norm values for breast cancer. Thus, comparative values were taken from a sample of non-fatigued and fatigued breast cancer patients at baseline from (Courtier et al., 2013).

With a mean value of \(37.8 \pm 10.5\) and a median of \(41.0\) (IQR: [30.0; 46.0]) the FACIT-F shows a distribution between non-fatigued and fatigued patients (M...
± SD: 36.4 ± 11.1; see Table 1) similar to the trial of Witt et al. (2015).

Finally, the interest in integrative cancer treatment was remarkably high. On an NRS from 0 = no interest to 10 = high interest, patients rated 7.7 ± 3.0.

However, their decision to be treated in an integrative hospital was not driven by their interest: 73.4 % reported that integrative medicine was not relevant for choosing the respective clinical center. 12.5 % reported a slight moderating effect and only 14.0 % based their decision for the hospital on the offer of integrative therapies.

This is somehow in accordance with the fact that only half of the patients (51.1 %) previously did not use integrative therapies.

4 DISCUSSION

This paper presents the technical realization and first results on representativity of the INTREST data, a cancer registry for breast cancer patients treated with integrative oncology.

<table>
<thead>
<tr>
<th>Tumor type</th>
<th>INTREST registry (N = 858)</th>
<th>Saarland registry (N = 93,721)</th>
<th>TMK cohort (N = 729)</th>
<th>CM trial (N = 275)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invasive ductal carcinoma</td>
<td>78.9 %</td>
<td>74.0 %</td>
<td>N/A</td>
<td>75.6 %</td>
</tr>
<tr>
<td>Invasive lobular carcinoma</td>
<td>15.8 %</td>
<td>12.8 %</td>
<td>N/A</td>
<td>15.6 %</td>
</tr>
<tr>
<td>Inflammatory breast cancer</td>
<td>0 %</td>
<td>0 %</td>
<td>N/A</td>
<td>0 %</td>
</tr>
<tr>
<td>Other BCs</td>
<td>5.3 %</td>
<td>3.3 %</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Tumor stage (pTNM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I</td>
<td>31.3 %</td>
<td>39.3 %</td>
<td>26.6 %</td>
<td>39.3 %</td>
</tr>
<tr>
<td>Stage II</td>
<td>24.3 %</td>
<td>39.4 %</td>
<td>46.2 %</td>
<td>38.5 %</td>
</tr>
<tr>
<td>Stage III</td>
<td>4.3 %</td>
<td>13.7 %</td>
<td>15.9 %</td>
<td>9.1 %</td>
</tr>
<tr>
<td>Tumor grading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GX</td>
<td>0.1 %</td>
<td>4.7 %</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>G1</td>
<td>13.4 %</td>
<td>3.9 %</td>
<td>N/A</td>
<td>10.9 %</td>
</tr>
<tr>
<td>G2</td>
<td>55.2 %</td>
<td>53.3 %</td>
<td>N/A</td>
<td>45.1 %</td>
</tr>
<tr>
<td>G3</td>
<td>31.2 %</td>
<td>28.2 %</td>
<td>N/A</td>
<td>44.4 %</td>
</tr>
<tr>
<td>Tumor receptor subtype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luminal A</td>
<td>69.2 %</td>
<td>N/A</td>
<td>59.9 %</td>
<td>N/A</td>
</tr>
<tr>
<td>Luminal B</td>
<td>12.5 %</td>
<td>N/A</td>
<td>15.5 %</td>
<td>N/A</td>
</tr>
<tr>
<td>HER2-positive</td>
<td>3.5 %</td>
<td>N/A</td>
<td>6.7 %</td>
<td>N/A</td>
</tr>
<tr>
<td>Triple-negative</td>
<td>14.7 %</td>
<td>N/A</td>
<td>16.2 %</td>
<td>N/A</td>
</tr>
<tr>
<td>Menopause</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-/perimenopausal</td>
<td>38.1 %</td>
<td>N/A</td>
<td>34.4 %</td>
<td>40.4 %</td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>61.9 %</td>
<td>N/A</td>
<td>65.6 %</td>
<td>52.7 %</td>
</tr>
<tr>
<td>FACT-G at baseline</td>
<td>78.8 (15.7)</td>
<td>N/A</td>
<td>75.7 (15.7)</td>
<td>76.2 (N/A)</td>
</tr>
<tr>
<td>FACIT-F subscale at baseline</td>
<td>37.8 (10.5)</td>
<td>N/A</td>
<td>N/A</td>
<td>36.4 (11.1)</td>
</tr>
</tbody>
</table>
Our analyses show that our sample of women suffering on breast cancer does not significantly differ from other registry and pragmatic trial data of breast cancer patients in Germany with respect to main sociodemographic and clinical cancer data. However, completeness particularly in tumor classification currently is a major limitation, which has also been reported in other registries (Ording et al., 2012; Ramos et al., 2015). Whether technical solutions in the sense of machine learning algorithms, e.g., to predict missing TNM-staging (Appelbaum et al., 2023), might be a helpful tool will be discussed when analyzing the missing data more deeply.

In the next step of the analysis, which is planned when the data of the respective follow-up assessment points have been entered into the database and missing data have been imputed according to the strategies described in Haller et al. (2021), logistic regression analyses and other predictive models will be run to identify potential responders and non-responders.

ACKNOWLEDGEMENTS

We would like to thank all the scientific assistants who actively supported us in entering the patient data.

REFERENCES


