JMIR Medical Informatics
Editor-in-Chief: Christian Lovis, MD, MSc, MBA
Published since 2013 by JMIR Publications, ISSN 2291-9694, URL: http://medinform.jmir.org/
Monthly (Online) journal, JMIR Publications, Phone: 416 585 2226, Fax: 416 423 3399, Email: jmedinform@jmir.org

This article is open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.

Preliminary Attainability Assessment of Real-World Data for Assessing Major Clinical Decisions in Breast Cancer Brain Metastasis: Framework Development and Validation Study

Monitoring Editor: Alexander Mazurenko

Reviewed by Ewelina Rogalska, Ewa Rogalska, and Krzysztof Sterkowicz

Yelin Jeong-Kim, MPH[1,2,3]; Hyun Ji Choi, MD[4]; Hyun Jung Park, MD[5]; Chan Woo Wee, MD[6]; Yoon Sim Park, MD[7]; Ji Hyun Lee, MD[8]; Jung Hyun Yoon, MD[9]; Jae Hyun Park, MD[10]

[1]Department of Digital Health, Samsung Medical Center, Seoul, Republic of Korea
[2]Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea
[3]Department of Medical Informatics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[4]Department of Medical Oncology, Yeungnam University College of Medicine, Daegu, Republic of Korea
[5]Center for Clinical Epidemiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[6]Department of Medical Oncology, Dongsan Medical Center, Keimyung University School of Medicine, Daegu, Republic of Korea
[7]Department of Medical Oncology, Dong-A University Hospital, Busan, Republic of Korea
[8]Division of Medical Oncology, Department of Internal Medicine, Gachon University Gil Medical Center, Incheon, Republic of Korea
[9]Department of Medical Oncology, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea
[10]Department of Medical Oncology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

Corresponding author:
Jae Hyun Park, MD, PhD, Department of Medical Oncology, Seoul National University Bundang Hospital, 82 Gumi-Ro 173 Beon-Gil, Bundang-Gu, Seongnam, Gyeonggi-do 13620, Republic of Korea, Phone: 82 31 787 7019, Fax: 82 31 787 4052, jhpark@snubh.org

Co-first author
Yelin Jeong-Kim, MPH, Department of Digital Health, Samsung Medical Center, Seoul, Republic of Korea, yelin.jeong@samsung.com

Co-second author
Hyun Ji Choi, MD, Department of Medical Oncology, Yeungnam University College of Medicine, Daegu, Republic of Korea, choihj@yumc.yuhs.ac

Received 02/02/2022; Revised version received 04/04/2022; 11/04/2022; 12/04/2022; 13/04/2022; Accepted 20/05/2022; Accepted for publication 23/05/2022; Published 22/06/2022

Abstract

Background

In recent decades, real-world evidence (RWE) in oncology has rapidly gained traction for its potential to answer clinical questions of interest to clinical practitioners and regulatory decision-makers. However, real-world data (RWD) from clinical research practice contribute to issues surrounding research designs, including extension, augmentation, enrichment, and pragmatic designs. Furthermore, the attainability of RWD for addressing major clinical decisions in breast cancer brain metastasis (BCBM) is unknown. Systematic harmonizing and analytical frameworks to assess the attainability of RWD for addressing major clinical decisions in BCBM are sufficient to answer the research questions before conducting the research with RWD but not yet been established.

Objective

We examined the PAM (Preliminary Attainability Assessment of Real-World Data) framework and assessed its ability to breast cancer brain metastasis (BCBM). We also set another medical unmet but data-attainability screening at the preliminary stage of observational studies that use RWD.

Methods

The PAM framework was proposed to assess data attainability from a particular data source during the early research process. The framework consists of 3 components: (1) clinical question, (2) data sufficiency, and (3) data attainability. We identified 3 major clinical data marking (structured/unstructured), 12 data screening and extraction, and 10 data attainability diagnostics. We identified 3 clinical questions of interest in BCBM and 1 medical unmet but data-attainability screening at the preliminary stage of observational studies that use RWD. We used the Korean Breast Cancer Registry (KBCR) linked with the Korean National Cancer Registry (March 2021, leveraging the institutional's unsynchronized and demilitarized clinical data warehouse platform). The number of breast cancer patients with BCBM from the KBCR was 1.5% (1,000/66,000) and the median follow-up was 5.1 (IQR 2.1-8.1) years. We found that the major clinical data, patient characteristics, tumor characteristics, treatment, and outcomes, were sufficient to answer the research questions from the PAM framework and a procedural diagram of data attainability for each research question.

Results

Data attainability was tested for study feasibility according to the PAM framework with 3 clinical questions for BCBM. We obtained data on the size and location of the primary tumor, the number of brain metastasis, and the treatment modality for local control of brain metastasis from clinical practice-based research variables. In the beta, only 1 question could be answered using these data elements. The other 2 questions could not be answered because the data elements for conditional data variables, for the third, the question could not be feasible for conducting further analysis.

Conclusions

The adaption of the PAM framework was associated with more efficient preliminary clinical research using RWD from BCBM. Furthermore, the framework could be a pragmatic approach for assessing data attainability, improving the efficiency of research reproducibility and increasing the entry barrier for clinical research.

Keywords

Breast neoplasms (BN) is a major cause of mortality in patients with breast cancer, and it increases the difficulty of treatment. Advancements in treatment and supportive care have improved the survival of patients with BN, and the number of patients with metastatic breast cancer leading to an increased incidence of BN[1,2]. Nonetheless, the opportunity to participate in prospective randomized clinical trials (RCTs) is limited for patients with BN, and the generalizability of the results of clinical trials is limited[3,4]. Real-world evidence (RWE) in oncology has rapidly gained traction for its potential to answer clinical questions of interest to clinical practitioners and patients, and patients, health care providers, and regulatory decision-makers[5]. Consequently, while the incidence and survival duration increase, patients with BN have been underrepresented in RCTs[6].

Meanwhile, real-world evidence (RWE) in oncology has rapidly gained traction in recent decades, with the potential to answer clinical questions of interest to clinical practitioners and patients, and patients, health care providers, and regulatory decision-makers[5]. Real-world data (RWD) from clinical research practice contribute to issues surrounding research designs, including extension, augmentation, enrichment, and pragmatic design[7,8]. To support the use of RWD, pragmatic medical interventions have provided RWD through the construction of clinical research infrastructure and have been implemented in practice[9,10]. However, the attainability of RWD for addressing major clinical decisions in breast cancer brain metastasis (BCBM) remains unknown. Systematic harmonizing and analytical frameworks to assess the attainability of RWD for addressing major clinical questions in the PAM framework, the evaluation of the feasibility of the framework, the working group assessed the preliminary data attainability of the breast clinical questions with a pilot data source, the Korean Breast Cancer Registry (KBCR) linked with the Korean National Cancer Registry.

Data attainability was defined in this study as the availability of data for recommendation and extraction required from a particular data source to answer a clinical question of interest in a research setting.

Data Source and Population

Samsung Medical Center Breast Cancer Registry (SMC BCR) is a hospital-based and time-registry implemented in March 2012 that leverages the institution's unsynchronized and demilitarized CDW platform (the Data Analytics and Research Window for Integrated Environment [DARWIN]). The SMC BCR includes 66,000 patients with BN who underwent treatment at Samsung Medical Center, a tertiary hospital, and collected data from the electronic medical record (EMR) and administrative data (National Health Insurance and National Cancer Registry) under CITI-A (Ethics of the International Classification of Diseases codes, and 2) aged older than 15 years at enrollment[11]. The Korean Breast Cancer Registry (KBCR) is a nationwide, multicenter, prospective, observational cohort of patients with BN, which was established in 2006. The registry has expressed interest in collaborating with the Korean National Cancer Registry (KNCR) to implement a registry provides clinical variables in the forms of structured data, including demographics, diagnostic history, treatment information, and patient-reported outcomes (PROs) and follow-up data. The KNCR is a population-based registry that collects cancer-related data such as pathology, radiology, and treatment reports[12].

The SMC BCR patient identifiers, we used a BCRM data mart in the SMC BCR. The BCRM data mart contains predefined and structured variables, including demographics, tumor characteristics, treatment, and follow-up. We linked the data with the KNCR indicators, defined by clinical experts, as follows: cytotoxicity, trastuzumab (yes/no), endocrine therapy, radiotherapy, treatment, and follow-up. We used the KBCR data set, which was linked with the KNCR data set, to identify the patients with breast cancer with brain metastasis (BCBM). The KBCR data set was evaluated, and the data were extracted, processed, and analyzed. The analyses were reliable to prior data management and data validity was greater than 95% on 10% random sampling.

Clinicians Interested Twice-Daily Review for Key Clinical Questions. Seven breast cancer experts from 6 hospitals in the Republic of Korea participated in a survey to identify the clinical questions of interest in breast cancer brain metastasis (Multimedia Appendix 1).

Working Group for Data Attainability Assessment. A working group was formed for data attainability testing. A computer-aided clinical expert system and 3 experts with 10 years of experience in clinical research across breast oncology, breast cancer, including medical oncologists and epidemiology.

Clinical Question Development. The modified Delphi approach was implemented through 2 rounds of interviews with 7 breast clinical experts in April and June 2021 (Multimedia Appendix 1). From the interviews, we identified clinical questions that reflected the most common practice patterns in BCBM. The working group then selected 3 clinical questions of interest for the preliminary data attainability assessment. Breast cancer brain metastasis (BCBM) was the clinical question. Seven breast cancer experts participated in a survey from September 27 to October 1, 2021. In this survey, clinical experts scored clinical questions, research feasibility, data availability, and clinical relevance (1-9 points). The 3 clinical questions with the highest scores underwent a pilot test for data attainability assessment with a pilot study. All data analysis and calculations were performed using Microsoft Excel 2019 (Microsoft Corporation, Redmond, WA) and were designed using R version 4.0.2 R Foundation for Statistical Computing.

List of clinical questions developed from interviews with experts.

Clinical question A

Is there a difference in survival outcomes between the brain as the primary metastasis site and brain metastasis accompanied by systemic metastasis?

Clinical question B

Does systemic treatment for brain metastases patients affect survival outcomes in patients?

Clinical question C

Does the timing of systemic treatment initiation affect survival outcomes when brain metastasis is accompanied by systemic metastasis?

Clinical question D

Is there a difference in the predilection of brain metastasis in patients who previously received trastuzumab doses for neoadjuvant therapy and patients who received pertuzumab and trastuzumab for neoadjuvant therapy?

Clinical question E

Can any record of neurological symptoms described by patients be translated as a surrogate factor of brain metastasis?

Clinical question F

Does systemic treatment for brain metastases patients affect survival outcomes in patients?

Clinical question C

Does the timing of systemic treatment initiation affect survival outcomes when brain metastasis is accompanied by systemic metastasis?

Clinical question D

Is there a difference in the predilection of brain metastasis in patients who previously received trastuzumab doses for neoadjuvant therapy and patients who received pertuzumab and trastuzumab for neoadjuvant therapy?

Clinical question E

Can any record of neurological symptoms described by patients be translated as a surrogate factor of brain metastasis?

PAM Framework. The PAM framework was proposed to assess the data attainability of a data source from the perspective of a clinical research practitioner. The framework consists of 3 components: (1) clinical question, (2) data sufficiency, and (3) data attainability of the clinical question (Figure 1).

Stage 1. Operational Definitions of Variables. At this stage, we identify variables relevant to the clinical question, describe them in operational definitions, and mark them as structured or unstructured. The operational definitions of variables include dependent or independent variables at the descriptive level of the research hypothesis. The other level is a data variable level, which is a variable that can be measured in a research setting. The data variable level includes the variable's name, the data type, and the data source. The data source includes the data collection method, the data collector, and the data source that contains the research variable.

Stage 2. Data Screening and Extraction. This stage includes mapping the data variables with specific data fields in the data source selected at the research material. Data fields refer to stored data elements, such as the column name of the data table, the cell of a form, or a field in a form. We then extract the data according to the data type, such as a text, a number, a date, or a binary variable. We then extract the data according to the data type, such as a text, a number, a date, or a binary variable.

Stage 3. Data Screening and Extraction. Here, the actual data values are extracted and reviewed from the data source for each variable identified in the previous stage.

Ethical Considerations. The study protocol was approved by the Institutional Review Board of the SMC (2021-06-028), which waived the need for informed consent because the study used de-identified data. Data were extracted from the CDW. This study followed the Standards for Quality Improvement Reporting Excellence (SQUIRE) guidelines [13].

Results

The working group assessed data attainability pertaining to the 3 clinical questions through stages 1 to 3. In stage 1, the research team identified 3 major clinical data marking (structured/unstructured), 12 data screening and extraction, and 10 data attainability diagnostics. We identified 3 clinical questions of interest in BCBM and 1 medical unmet but data-attainability screening at the preliminary stage of observational studies that use RWD (Figure 2). We used the Korean Breast Cancer Registry (KBCR) linked with the Korean National Cancer Registry (March 2021, leveraging the institutional's unsynchronized and demilitarized clinical data warehouse platform). The number of breast cancer patients with BCBM from the KBCR was 1.5% (1,000/66,000) and the median follow-up was 5.1 (IQR 2.1-8.1) years. We found that the major clinical data, patient characteristics, tumor characteristics, treatment, and outcomes, were sufficient to answer the research questions from the PAM framework and a procedural diagram of data attainability for each research question.

Discussion

Data attainability was tested for study feasibility according to the PAM framework with 3 clinical questions for BCBM. We obtained data on the size and location of the primary tumor, the number of brain metastasis, and the treatment modality for local control of brain metastasis from clinical practice-based research variables. In the beta, only 1 question could be answered using these data elements. The other 2 questions could not be answered because the data elements for conditional data variables, for the third, the question could not be feasible for conducting further analysis.

Conclusions

The adaption of the PAM framework was associated with more efficient preliminary clinical research using RWD from BCBM. Furthermore, the framework could be a pragmatic approach for assessing data attainability, improving the efficiency of research reproducibility and increasing the entry barrier for clinical research.

Accelerated and reliable preliminary clinical research assessment, observational study, clinical data warehouse, PAM framework, breast metastasis, breast cancer

Introduction

Breast metastasis (BM) is a major cause of mortality in patients with breast cancer, and it increases the difficulty of treatment. Advancements in treatment and supportive care have improved the survival of patients with BM, and the number of patients with metastatic breast cancer leading to an increased incidence of BM[1,2]. Nonetheless, the opportunity to participate in prospective randomized clinical trials (RCTs) is limited for patients with BM, and the generalizability of the results of clinical trials is limited[3,4]. Real-world evidence (RWE) in oncology has rapidly gained traction for its potential to answer clinical questions of interest to clinical practitioners and patients, health care providers, and regulatory decision-makers[5]. Consequently, while the incidence and survival duration increase, patients with BM have been underrepresented in RCTs[6].

Meanwhile, real-world evidence (RWE) in oncology has rapidly gained traction in recent decades, with the potential to answer clinical questions of interest to clinical practitioners and patients, and patients, health care providers, and regulatory decision-makers[5]. Real-world data (RWD) from clinical research practice contribute to issues surrounding research designs, including extension, augmentation, enrichment, and pragmatic design[7,8]. To support the use of RWD, pragmatic medical interventions have provided RWD through the construction of clinical research infrastructure and have been implemented in practice[9,10]. However, the attainability of RWD for addressing major clinical decisions in breast cancer brain metastasis (BCBM) remains unknown. Systematic harmonizing and analytical frameworks to assess the attainability of RWD for addressing major clinical questions in the PAM framework, the evaluation of the feasibility of the framework, the working group assessed the preliminary data attainability of the breast clinical questions with a pilot data source, the Korean Breast Cancer Registry (KBCR) linked with the Korean National Cancer Registry.

Data attainability was defined in this study as the availability of data for recommendation and extraction required from a particular data source to answer a clinical question of interest in a research setting.

Data Source and Population

Samsung Medical Center Breast Cancer Registry (SMC BCR) is a hospital-based and time-registry implemented in March 2012 that leverages the institution's unsynchronized and demilitarized CDW platform (the Data Analytics and Research Window for Integrated Environment [DARWIN]). The SMC BCR includes 66,000 patients with BN who underwent treatment at Samsung Medical Center, a tertiary hospital, and collected data from the electronic medical record (EMR) and administrative data (National Health Insurance and National Cancer Registry) under CITI-A (Ethics of the International Classification of Diseases codes, and 2) aged older than 15 years at enrollment[11]. The Korean Breast Cancer Registry (KBCR) is a nationwide, multicenter, prospective, observational cohort of patients with BN, which was established in 2006. The registry has expressed interest in collaborating with the Korean National Cancer Registry (KNCR) to implement a registry provides clinical variables in the forms of structured data, including demographics, diagnostic history, treatment information, and patient-reported outcomes (PROs) and follow-up data. The KNCR is a population-based registry that collects cancer-related data such as pathology, radiology, and treatment reports[12].

The SMC BCR patient identifiers, we used a BCRM data mart in the SMC BCR. The BCRM data mart contains predefined and structured variables, including demographics, tumor characteristics, treatment, and follow-up. We linked the data with the KNCR indicators, defined by clinical experts, as follows: cytotoxicity, trastuzumab (yes/no), endocrine therapy, radiotherapy, treatment, and follow-up. We used the KBCR data set, which was linked with the KNCR data set, to identify the patients with breast cancer with brain metastasis (BCBM). The KBCR data set was evaluated, and the data were extracted, processed, and analyzed. The analyses were reliable to prior data management and data validity was greater than 95% on 10% random sampling.

Clinicians Interested Twice-Daily Review for Key Clinical Questions. Seven breast cancer experts from 6 hospitals in the Republic of Korea participated in a survey to identify the clinical questions of interest in breast cancer brain metastasis (Multimedia Appendix 1).

Working Group for Data Attainability Assessment. A working group was formed for data attainability testing. A computer-aided clinical expert system and 3 experts with 10 years of experience in clinical research across breast oncology, breast cancer, including medical oncologists and epidemiology.

Clinical Question Development. The modified Delphi approach was implemented through 2 rounds of interviews with 7 breast clinical experts in April and June 2021 (Multimedia Appendix 1). From the interviews, we identified clinical questions that reflected the most common practice patterns in BCBM. The working group then selected 3 clinical questions of interest for the preliminary data attainability assessment. Breast cancer brain metastasis (BCBM) was the clinical question. Seven breast cancer experts participated in a survey from September 27 to October 1, 2021. In this survey, clinical experts scored clinical questions, research feasibility, data availability, and clinical relevance (1-9 points). The 3 clinical questions with the highest scores underwent a pilot test for data attainability assessment with a pilot study. All data analysis and calculations were performed using Microsoft Excel 2019 (Microsoft Corporation, Redmond, WA) and were designed using R version 4.0.2 R Foundation for Statistical Computing.

List of clinical questions developed from interviews with experts.

Clinical question A

Is there a difference in survival outcomes between the brain as the primary metastasis site and brain metastasis accompanied by systemic metastasis?

Clinical question B

Does systemic treatment for brain metastases patients affect survival outcomes in patients?

Clinical question C

Does the timing of systemic treatment initiation affect survival outcomes when brain metastasis is accompanied by systemic metastasis?

Clinical question D

Is there a difference in the predilection of brain metastasis in patients who previously received trastuzumab doses for neoadjuvant therapy and patients who received pertuzumab and trastuzumab for neoadjuvant therapy?

Clinical question E

Can any record of neurological symptoms described by patients be translated as a surrogate factor of brain metastasis?

PAM Framework. The PAM framework was proposed to assess the data attainability of a data source from the perspective of a clinical research practitioner. The framework consists of 3 components: (1) clinical question, (2) data sufficiency, and (3) data attainability of the clinical question (Figure 1).

Stage 1. Operational Definitions of Variables. At this stage, we identify variables relevant to the clinical question, describe them in operational definitions, and mark them as structured or unstructured. The operational definitions of variables include dependent or independent variables at the descriptive level of the research hypothesis. The other level is a data variable level, which is a variable that can be measured in a research setting. The data variable level includes the variable's name, the data type, and the data source. The data source includes the data collection method, the data collector, and the data source that contains the research variable.

Stage 2. Data Screening and Extraction. This stage includes mapping the data variables with specific data fields in the data source selected at the research material. Data fields refer to stored data elements, such as the column name of the data table, the cell of a form, or a field in a form. We then extract the data according to the data type, such as a text, a number, a date, or a binary variable. We then extract the data according to the data type, such as a text, a number, a date, or a binary variable.

Stage 3. Data Screening and Extraction. Here, the actual data values are extracted and reviewed from the data source for each variable identified in the previous stage.

Ethical Considerations. The study protocol was approved by the Institutional Review Board of the SMC (2021-06-028), which waived the need for informed consent because the study used de-identified data. Data were extracted from the CDW. This study followed the Standards for Quality Improvement Reporting Excellence (SQUIRE) guidelines [13].

Results

The working group assessed data attainability pertaining to the 3 clinical questions through stages 1 to 3. In stage 1, the research team identified 3 major clinical data marking (structured/unstructured), 12 data screening and extraction, and 10 data attainability diagnostics. We identified 3 clinical questions of interest in BCBM and 1 medical unmet but data-attainability screening at the preliminary stage of observational studies that use RWD (Figure 2). We used the Korean Breast Cancer Registry (KBCR) linked with the Korean National Cancer Registry (March 2021, leveraging the institutional's unsynchronized and demilitarized clinical data warehouse platform). The number of breast cancer patients with BCBM from the KBCR was 1.5% (1,000/66,000) and the median follow-up was 5.1 (IQR 2.1-8.1) years. We found that the major clinical data, patient characteristics, tumor characteristics, treatment, and outcomes, were sufficient to answer the research questions from the PAM framework and a procedural diagram of data attainability for each research question.

Discussion

Data attainability was tested for study feasibility according to the PAM framework with 3 clinical questions for BCBM. We obtained data on the size and location of the primary tumor, the number of brain metastasis, and the treatment modality for local control of brain metastasis from clinical practice-based research variables. In the beta, only 1 question could be answered using these data elements. The other 2 questions could not be answered because the data elements for conditional data variables, for the third, the question could not be feasible for conducting further analysis.

Conclusions

The adaption of the PAM framework was associated with more efficient preliminary clinical research using RWD from BCBM. Furthermore, the framework could be a pragmatic approach for assessing data attainability, improving the efficiency of research reproducibility and increasing the entry barrier for clinical research.

Introduction

Breast metastasis (BM) is a major cause of mortality in patients with breast cancer, and it increases the difficulty of treatment. Advancements in treatment and supportive care have improved the survival of patients with BM, and the number of patients with metastatic breast cancer leading to an increased incidence of BM[1,2]. Nonetheless, the opportunity to participate in prospective randomized clinical trials (RCTs) is limited for patients with BM, and the generalizability of the results of clinical trials is limited[3,4]. Real-world evidence (RWE) in oncology has rapidly gained traction for its potential to answer clinical questions of interest to clinical practitioners and patients, health care providers, and regulatory decision-makers[5]. Consequently, while the incidence and survival duration increase, patients with BM have been underrepresented in RCTs[6].

Meanwhile, real-world evidence (RWE) in oncology has rapidly gained traction in recent decades, with the potential to answer clinical questions of interest to clinical practitioners and patients, and patients, health care providers, and regulatory decision-makers[5]. Real-world data (RWD) from clinical research practice contribute to issues surrounding research designs, including extension, augmentation, enrichment, and pragmatic design[7,8]. To support the use of RWD, pragmatic medical interventions have provided RWD through the construction of clinical research infrastructure and have been implemented in practice[9,10]. However, the attainability of RWD for addressing major clinical decisions in breast cancer brain metastasis (BCBM) remains unknown. Systematic harmonizing and analytical frameworks to assess the attainability of RWD for addressing major clinical questions in the PAM framework, the evaluation of the feasibility of the framework, the working group assessed the preliminary data attainability of the breast clinical questions with a pilot data source, the Korean Breast Cancer Registry (KBCR) linked with the Korean National Cancer Registry.

Data attainability was defined in this study as the availability of data for recommendation and extraction required from a particular data source to answer a clinical question of interest in a research setting.

Data Source and Population

Samsung Medical Center Breast Cancer Registry (SMC BCR) is a hospital-based and time-registry implemented in March 2012 that leverages the institution's unsynchronized and demilitarized CDW platform (the Data Analytics and Research Window for Integrated Environment [DARWIN]). The SMC BCR includes 66,000 patients with BN who underwent treatment at Samsung Medical Center, a tertiary hospital, and collected data from the electronic medical record (EMR) and administrative data (National Health Insurance and National Cancer Registry) under CITI-A (Ethics of the International Classification of Diseases codes, and 2) aged older than 15 years at enrollment[11]. The Korean Breast Cancer Registry (KBCR) is a nationwide, multicenter, prospective, observational cohort of patients with BN, which was established in 2006. The registry has expressed interest in collaborating with the Korean National Cancer Registry (KNCR) to implement a registry provides clinical variables in the forms of structured data, including demographics, diagnostic history, treatment information, and patient-reported outcomes (PROs) and follow-up data. The KNCR is a population-based registry that collects cancer-related data such as pathology, radiology, and treatment reports[12].

The SMC BCR patient identifiers, we used a BCRM data mart in the SMC BCR. The BCRM data mart contains predefined and structured variables, including demographics, tumor characteristics, treatment, and follow-up. We linked the data with the KNCR indicators, defined by clinical experts, as follows: cytotoxicity, trastuzumab (yes/no), endocrine therapy, radiotherapy, treatment, and follow-up. We used the KBCR data set, which was linked with the KNCR data set, to identify the patients with breast cancer with brain metastasis (BCBM). The KBCR data set was evaluated, and the data were extracted, processed, and analyzed. The analyses were reliable to prior data management and data validity was greater than 95% on 10% random sampling.

Clinicians Interested Twice-Daily Review for Key Clinical Questions. Seven breast cancer experts from 6 hospitals in the Republic of Korea participated in a survey to identify the clinical questions of interest in breast cancer brain metastasis (Multimedia Appendix 1).

Working Group for Data Attainability Assessment. A working group was formed for data attainability testing. A computer-aided clinical expert system and 3 experts with 10 years of experience in clinical research across breast oncology, breast cancer, including medical oncologists and epidemiology.

Clinical Question Development. The modified Delphi approach was implemented through 2 rounds of interviews with 7 breast clinical experts in April and June 2021 (Multimedia Appendix 1). From the interviews, we identified clinical questions that reflected the most common practice patterns in BCBM. The working group then selected 3 clinical questions of interest for the preliminary data attainability assessment. Breast cancer brain metastasis (BCBM) was the clinical question. Seven breast cancer experts participated in a survey from September 27 to October 1, 2021. In this survey, clinical experts scored clinical questions, research feasibility, data availability, and clinical relevance (1-9 points). The 3 clinical questions with the highest scores underwent a pilot test for data attainability assessment with a pilot study. All data analysis and calculations were performed using Microsoft Excel 2019 (Microsoft Corporation, Redmond, WA) and were designed using R version 4.0.2 R Foundation for Statistical Computing.

List of clinical questions developed from interviews with experts.

Clinical question A

Is there a difference in survival outcomes between the brain as the primary metastasis site and brain metastasis accompanied by systemic metastasis?

Clinical question B

Does systemic treatment for brain metastases patients affect survival outcomes in patients?

Clinical question C

Does the timing of systemic treatment initiation affect survival outcomes when brain metastasis is accompanied by systemic metastasis?

Clinical question D

Is there a difference in the predilection of brain metastasis in patients who previously received trastuzumab doses for neoadjuvant therapy and patients who received pertuzumab and trastuzumab for neoadjuvant therapy?

Clinical question E

Can any record of neurological symptoms described by patients be translated as a surrogate factor of brain metastasis?

PAM Framework. The PAM framework was proposed to assess the data attainability of a data source from the perspective of a clinical research practitioner. The framework consists of 3 components: (1) clinical question, (2) data sufficiency, and (3) data attainability of the clinical question (Figure 1).

Stage 1. Operational Definitions of Variables. At this stage, we identify variables relevant to the clinical question, describe them in operational definitions, and mark them as structured or unstructured. The operational definitions of variables include dependent or independent variables at the descriptive level of the research hypothesis. The other level is a data variable level, which is a variable that can be measured in a research setting. The data variable level includes the variable's name, the data type, and the data source. The data source includes the data collection method, the data collector, and the data source that contains the research variable.

Stage 2. Data Screening and Extraction. This stage includes mapping the data variables with specific data fields in the data source selected at the research material. Data fields refer to stored data elements, such as the column name of the data table, the cell of a form, or a field in a form. We then extract the data according to the data type, such as a text, a number, a date, or a binary variable. We then extract the data according to the data type, such as a text, a number, a date, or a binary variable.

Stage 3. Data Screening and Extraction. Here, the actual data values are extracted and reviewed from the data source for each variable identified in the previous stage.

Ethical Considerations. The study protocol was approved by the Institutional Review Board of the SMC (2021-06-028), which waived the need for informed consent because the study used de-identified data. Data were extracted from the CDW. This study followed the Standards for Quality Improvement Reporting Excellence (SQUIRE) guidelines [13].

Results

The working group assessed data attainability pertaining to the 3 clinical questions through stages 1 to 3. In stage 1, the research team identified 3 major clinical data marking (structured/unstructured), 12 data screening and extraction, and 10 data attainability diagnostics. We identified 3 clinical questions of interest in BCBM and 1 medical unmet but data-attainability screening at the preliminary stage of observational studies that use RWD (Figure 2). We used the Korean Breast Cancer Registry (KBCR) linked with the Korean National Cancer Registry (March 2021, leveraging the institutional's unsynchronized and demilitarized clinical data warehouse platform). The number of breast cancer patients with BCBM from the KBCR was 1.5% (1,000/66,000) and the median follow-up was 5.1 (IQR 2.1-8.1) years. We found that the major clinical data, patient characteristics, tumor characteristics, treatment, and outcomes, were sufficient to answer the research questions from the PAM framework and a procedural diagram of data attainability for each research question.

Discussion

Data attainability was tested for study feasibility according to the PAM framework with 3 clinical questions for BCBM. We obtained data on the size and location of the primary tumor, the number of brain metastasis, and the treatment modality for local control of brain metastasis from clinical practice-based research variables. In the beta, only 1 question could be answered using these data elements. The other 2 questions could not be answered because the data elements for conditional data variables, for the third, the question could not be feasible for conducting further analysis.

Conclusions

The adaption of the PAM framework was associated with more efficient preliminary clinical research using RWD from BCBM. Furthermore, the framework could be a pragmatic approach for assessing data attainability, improving the efficiency of research reproducibility and increasing the entry barrier for clinical research.

Acknowledgments

This study was supported by Patient Medicine 2020 Project of Samsung Medical Center (SMC0191041) and the Korea Institute of Medical Microrobotics (KIMM) under the National Research Council of Science & Technology (NRC) funded by the Ministry of Education (MOE) [EO18-0116-02(0000290)]. We thank the clinical experts who contributed to this study: Ji-Hye Park from Korea University Guro Hospital, Ji-Hyun Lee from Gachon University Gil Medical Center, Jung Hyun Yoon from Gangnam Severance Hospital, Yoon Sim Park from Dong-A University Hospital, Chan Woo Wee from Dongsan Medical Center, Hyun Jung Park from Severance Hospital, Hyun Ji Choi from Yeungnam University Medical Center. The clinical questions were developed and the preliminary data attainability assessment was performed using the Korean Breast Cancer Registry (KBCR) linked with the Korean National Cancer Registry (KNCR). The clinical question development and data attainability assessment were performed using the Data Analytics and Research Window for Integrated Environment (DARWIN) Clinical Data Warehouse. The clinical question development and data attainability assessment were presented among surrogate researchers (Hyun Ji Choi, Chan Woo Wee, Yoon Sim Park, Ji-Hyun Lee, Jung Hyun Yoon, Jae Hyun Park) at a poster at the Jun Internet Breast Cancer Conference Surgery 2022 (December 6-7, 2022, Jeju, Korea).

Abbreviations

BCBM	breast cancer brain metastasis
BCM	breast cancer registry
BN	brain metastases
CDW	clinical data warehouse
DARWIN	Data Analytics and Research Window for Integrated Knowledge
EHR	electronic health records
HER2	human epidermal growth factor receptor 2
HR	hormone receptor
PAM	Preliminary Attainability Assessment of Real-world Data
RCT	randomized clinical trial
RWD	real-world data
RWE	real-world evidence
SMC BCR	Samsung Medical Center Breast Cancer Registry
TNBC	triple-negative breast cancer

Multimedia Appendix 1

Composition of expert assessments for each clinical question was conducted by surveying 7 experts' opinions from September 27th to October 1st, 2021.

Multimedia Appendix 2

The statistical analysis of the survey results utilized correlation analysis between the mean value of each score and the SD.

Notes

Data Availability

The data used for this study were extracted from a real-time breast cancer registry and breast cancer metastasis data warehouse platform, the Data Analytics and Research Window for Integrated Environment (DARWIN) Clinical Data Warehouse. The data are not publicly available.

References

1. Bertozzi S, Jeanton Vallero, Mo P, Cordero S, Seibaldi V, Rubi Rabbia, Mengaldo F, Stesi M. Management of brain and leptomeningeal metastases from breast cancer: a multidisciplinary review. Cancer Treat Rev. 2020 Jun;86:102018. doi: 10.1016/j.ctrv.2020.102018. [PubMed: 32330849]. [Crossref: 32330849]
2. Ali R, Pharaon M, Murtaza G. Preoperative chemotherapy versus primary surgery for breast cancer brain metastasis: a meta-analysis. Ann Surg Oncol. 2012 Jun;19(6):1835-42. doi: 10.1245/s10434-011-1615-7. [PubMed: 21755365]. [Crossref: 21755365]
3. Wilhelm I, Molnár J, Fazakas J, Haskó G, Krizbai I. Transcytosis in brain metastasis. J Neurosci Res. 2017 Jan;95(1-2):339-351. doi: 10.1002/jnr.23849. [PubMed: 27870396]. [Crossref: 27870396]
4. Ali DM, Glass EM, Sperduto PW. Breast cancer brain metastases: the evolving landscape. J Neurosurg. 2014 Jun;121(6):1299-305. doi: 10.3171/2014.7.JNS140256. [PubMed: 25192475]. [Crossref: 25192475]
5. Wolf DM, Yaffe MB, Isakoff SJ, Emdin SO, Fleisher AB, Yagihashi S, Shibata T, Gazdar AF, Minna JD, Kurokawa Y, Tanaka K, Doi O. Breast cancer to the brain without metastasis at diagnosis: a new phenomenon. JAMA. 1995 Jul 26;274(4):299-301. doi: 10.1001/jama.1995.03530040049031. [PubMed: 7609253]. [Crossref: 7609253]
6. Lin NU, Bellon JR, Winer EP. CNS metastases in breast cancer. J Clin Oncol. 2004 Apr 15;22(8):3608-17. doi: 10.1200/JCO.2004.08.075. [PubMed: 15337811]. [Crossref: 15337811]
7. Pestalozzi BC, Peterslund NA, Price KN, Castiglione-Gertsch M, Castiglione M, Cavalli F, Gelber RD, Goldhirsch A, Lindtner J, Nesvold IR, Rey P, Rudenstam CM, Thomas BA, Thurlimann B, Crivellari D, Gusterson BA, Lindtner R, Nesvold IR, Rey P, Rudenstam CM, Thomas BA, Thurlimann B, Crivellari D, Gusterson BA. Prognostic factors in 6,165 women with early breast cancer and synchronous contralateral metastasis: a retrospective study. J Clin Oncol. 2008 Jan 20;26(3):384-90. doi: 10.1200/JCO.2007.12.7.7. [PubMed: 18202417]. [Crossref: 18202417]
8. Pestalozzi E, Gelber RD, Goldhirsch A, Gusterson BA, Mauri D. An overview of real-world data sources for oncology and considerations for research. Lancet Digit Health. 2019 Dec;1(9):e486-e492. doi: 10.1016/s2589-7500(19)30160-6. [PubMed: 33323271]. [Crossref: 33323271]
9. Ali DM, Drilon A, Ahmadi O, Greenblatt J, Zimmerman M, Cole LE, Lasen A, Herling M, Hassanein AM, Gewirtz AN, Gomez D, Brogi E, Morris PG. Central nervous system metastasis in breast cancer: prognostic factors and outcomes. Clin Breast Cancer. 2015 Dec;15(6):467-73. doi: 10.1016/j.clbc.2015.06.003. [PubMed: 26190420]. [Crossref: 26190420]
10. Berrocal A, Berrocal M, Gonzalez-Larriba JL, Lopez-Tarruella C, Martinez-Navarro A, Martinez-Jimenez I, Lara-Lara P, Ramos-Vazquez M, Ramos-Medina R, Ramos-Vazquez M, Martinez-Jimenez I, Lara-Lara P. Central nervous system metastasis in breast cancer patients treated with trastuzumab: incidence, pattern, risk factors and outcomes. Clin Transl Oncol. 2014 Jan;16(1):3-9. doi: 10.1007/s12094-013-1035-1. [PubMed: 23436000]. [Crossref: 23436000]
11. Yoshida T, Ichikawa W, Hamaguchi T, Shimada Y, Hara H, Ishioka C, Katai H, Doi T, Boku N, Sakata Y. Clinical impact of HER2 amplification on survival in patients with metastatic gastric cancer. Int J Cancer. 2015 Jan

R, Kiely B, Kim S, Lin N, Mertz S, Neciosup S, Offersen B, Ohno S, Pagani O, Prat A, Penault-Llorca F, Rugo H, Sledge G, Thomssen C, Vorobiof D, Wiseman T, Xu B, Norton L, Costa A, Winer E. 5th ESO-ESMO international consensus guidelines for advanced breast cancer (ABC 5) *Ann Oncol.* 2020 Dec;31(12):1623–1649. doi: 10.1016/j.annonc.2020.09.010. https://air.unimi.it/handle/2434/824878 .S0923-7534(20)42460-3 [PMCID: PMC7510449] [PubMed: 32979513] [CrossRef: 10.1016/j.annonc.2020.09.010]

26. Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio I, Zackrisson S, Senkus E. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2019 Aug 01;30(8):1194–1220. doi: 10.1093/annonc/mdz173. https://linkinghub.elsevier.com/retrieve/pii/S0923-7534(19)31287-6 .S0923-7534(19)31287-6 [PubMed: 31161190] [CrossRef: 10.1093/annonc/mdz173]

27. Makady A, de Boer A, Hillege H, Klungel O, Goettsch W, (on behalf of GetReal Work Package 1) What Is real-world data? a review of definitions based on literature and stakeholder interviews. *Value Health.* 2017;20(7):858–865. doi: 10.1016/j.jval.2017.03.008. https://linkinghub.elsevier.com/retrieve/pii/S1098-3015(17)30171-7 .S1098-3015(17)30171-7 [PubMed: 28712614] [CrossRef: 10.1016/j.jval.2017.03.008]

28. Meystre SM, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann CU. Clinical data reuse or secondary use: current status and potential future progress. *Yearb Med Inform.* 2017 Sep 11;26(01):38–52. doi: 10.15265/iy-2017-007. [PMCID: PMC6239225] [PubMed: 28480475] [CrossRef: 10.15265/iy-2017-007]

29. Campion JT, Craven C, Dorr D, Knosp B. Understanding enterprise data warehouses to support clinical and translational research. *J Am Med Inform Assoc.* 2020 Jul 01;27(9):1352–1358. doi: 10.1093/jamia/ocaa089. https://europepmc.org/abstract/MED/32679585 .5873310 [PMCID: PMC7647350] [PubMed: 32679585] [CrossRef: 10.1093/jamia/ocaa089]

30. Hersh W, Weiner M, Embi P, Logan J, Payne P, Bernstam E, Lehmann Harold P, Hripcsak George, Hartzog Timothy H, Cimino James J, Saltz Joel H. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care.* 2013 Aug;51(8 Suppl 3):S30–7. doi: 10.1097/MLR.0b013e31829b1dbd. https://europepmc.org/abstract/MED/23774517 . [PMCID: PMC3748381] [PubMed: 23774517] [CrossRef: 10.1097/MLR.0b013e31829b1dbd]

31. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, DeFalco FJ, Londhe A, Zhu V, Ryan PB. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc.* 2015 May;22(3):553–64. doi: 10.1093/jamia/ocu023. https://europepmc.org/abstract/MED/25670757 .ocu023 [PMCID: PMC4457111] [PubMed: 25670757] [CrossRef: 10.1093/jamia/ocu023]

32. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2) *J Am Med Inform Assoc.* 2010;17(2):124–30. doi: 10.1136/jamia.2009.000893. https://europepmc.org/abstract/MED/20190053 .17/2/124 [PMCID: PMC3000779] [PubMed: 20190053] [CrossRef: 10.1136/jamia.2009.000893]

33. Evans RS, Lloyd JF, Pierce LA. Clinical use of an enterprise data warehouse. *AMIA Annu Symp Proc.* 2012;2012:189–98. https://europepmc.org/abstract/MED/23304288 . [PMCID: PMC3540441] [PubMed: 23304288]

34. Shin S, Kim WS, Lee J. Characteristics desired in clinical data warehouse for biomedical research. *Healthc Inform Res.* 2014 Apr;20(2):109–16. doi: 10.4258/hir.2014.20.2.109. https://europepmc.org/abstract/MED/24872909 . [PMCID: PMC4030054] [PubMed: 24872909] [CrossRef: 10.4258/hir.2014.20.2.109]

Figures and Tables

Figure 1

Illustration of the Preliminary Attainability Assessment of Real-World Data (PAR) framework and study design. CQ: clinical question.

Figure 2

Results at stages 1 and 2 with the breast cancer brain metastasis clinical questions. CQ: clinical question.

Figure 3

Data attainability for the clinical questions. Clinical questions were as follows. (A) Is there a difference in survival outcomes between the brain as the primary metastasis site and brain metastasis accompanied by systemic metastasis? (B) Does systemic treatment for brain metastasis patients affect survival outcomes according to subtype? (C) Does the timing of systemic treatment initiation affect survival outcomes when brain metastasis is accompanied by systemic metastasis? and (D) Is there a difference in the prediction of brain metastasis in patients who previously received trastuzumab alone for neoadjuvant therapy and in patients who received pertuzumab and trastuzumab for neoadjuvant therapy? BCBM: breast cancer brain metastasis; HER2: human epidermal growth factor receptor 2; HR: hormone receptor; TNBC: triple-negative breast cancer.

Figure 4

Preliminary Attainability Assessment of Real-World Data framework evaluation in stages 1 to 4. CQ: clinical question; CDW: clinical data warehouse.

Journal Article Journal Article Journal Article Journal Article