Cardiovascular Phenotyping in Breast Cancer Patients Treated With Her2 Targeted Therapies Using Informatics Approaches

Michael A. Harris
Georgetown University Medical Center
Brian Conkright
Georgetown University Medical Center
Robert Johnson
Georgetown University Medical Center
Simina M. Boca
Georgetown University Medical Center
Shahla Riazi
MedStar Georgetown University Hospital
Rebecca Torguson
Medstar Cardiovascular Research Network
Adil Alaoui
Georgetown University Medical Center
Krithika Bhuvaneshwar
Georgetown University Medical Center https://orcid.org/0000-0003-4015-7056
Yuriy Gusev
Georgetown University Medical Center
Federico M. Asch
Medstar Cardiovascular Research Network
Paula R Pohlmann
Georgetown Lombardi Comprehensive Cancer Center
Ana Barac
Georgetown Lombardi Comprehensive Cancer Center
Subha Madhavan (✉ sm696@georgetown.edu)

Research

Keywords: Breast Cancer, Data Wrangling, Visual Analytics, Statistical Modeling

DOI: https://doi.org/10.21203/rs.3.rs-34395/v1

Abstract

Background

Cardiotoxicity is a serious adverse event associated with some of the most effective breast cancer therapies. Currently, it is difficult to predict which patients will develop cardiotoxicity due to the multiplicity of clinical, behavioral, and biological factors involved.

Methods

Here we describe an effort to apply biomedical informatics approaches to patient data from MedStar Health's EHR systems to discover and characterize factors that contribute to cardiotoxicity in a real world breast cancer population.

Results

Data wrangling techniques including merging data from disparate clinical systems, data transformation, and de-identification of personal health information (PHI)were appliedto the raw clinical data to produce a structured integrated dataset for predictive analysis and hypothesis generation. Using this dataset as input, weshowed howpredictive models can be developed to identify patients at high risk for cardiotoxicity.

Conclusions

We demonstrate how suchmodels can be used for hypothesis generation and data exploration with the ultimate goal of developing applications for precision medicine.

Background

Electronic health records (EHR) contain a variety of rich information regarding patient diagnoses, treatments, and health outcomes. Analysis of these data using informatics techniques can uncover important information regarding drug efficacy, outcomes, and adverse events at both the individual patient and population levels[1-5]. Extracting knowledge from raw data mined from clinical systems requires extensive data wrangling in order to shape, clean, validate, and transform the data so that it is in a form which can be readily consumed by downstream analysis processes[6].

We employ EHR data mining and analysis to address the problem of cardiotoxicity in a real-world cohort of HER2-positive breast cancer patients treated with trastuzumab. HER2-positive tumors are defined by overexpression of human epidermal growth factor receptor 2 (HER2/neu),comprisingapproximately 20%-25% of all breast cancers; if untreated, they have the poorest prognosis among breast cancer subtypes. The monoclonal antibody trastuzumab is among the most effectivetreatments for HER2-positive breast cancer.A recognized potential cardiac side effect of trastuzumabis an asymptomatic decline in left ventricular ejection fraction (LVEF) - the fraction of blood that leaves the left ventricle during contraction - known as left ventricular systolic dysfunction (LVSD)[7] and an increased risk of congestive heart failure (CHF)[3].Trastuzumab is often paired with anthracycline chemotherapy, which appears to lead to potentially more permanent LVSD and greater incidence of CHFthan the use of trastuzumab alone [7-9].

Despite widespread use,many questions remain unanswered about which patients will be the most susceptible to side-effects from trastuzumab and the optimal treatment protocols[10].

Here we report on our efforts to extract, integrate, and analyze data from EHR systems atthe MedStar Washington Hospital Center (MWHC), a hospital in the MedStar Health network.This project was evaluated by the Georgetown-Howard Universities Center for Clinical and Translational Science Institutional Review Board and was approved for detailed review (GU IRB #: 2016-1255).

Methods

Patient data including demographics, drug administrations, and cardiology information were extracted from MedStar clinical systems and prepared for analysis. We then performed data visualization, statistical modeling, and other analysis to identify the factors most predictive of cardiotoxic events. Our ultimate aim was to develop a framework for clinical decision support and precision medicine. Figure 1 shows our overall analysis workflow.

Clinical EHR Systems

To address the critical need to use EHR data to better understand trastuzumabrelated cardiotoxicity in breast cancer patients,we identifiedpatients with available diagnosis, lab, demographic and cardiology information. This required a cohort discovery strategy to identify data sources residing in disparate systems across the MedStar Health network.Our initial data source was ARIA, the oncology EHR systemthat contains patient demographics, diagnosis, lab results, drug orders, and clinical notes, among other data elements.The multi-modality image management system Xcelerawas used as a data source for echocardiogram data from MWHC.

Clinical Data Extraction, Filtering, Integration, and Cleaning

We investigated patients diagnosed with breast cancer who were treated with trastuzumab and had valid echocardiogram data from MWHC available for analysis. Figure 2illustrates our data extraction and filtering process.

In order to identify the patients in our cohort, we executed queries against ARIA using the ICD-9 diagnosis codes for female breast cancer (174.0, 174.1, 174.2, 174.3, 174.4, 174.5, 174.6, 174.8, and 174.9) identifying a set of 11,560 patients for further consideration. Next we queried the drug administration tables for these patients and determined that 702 of these patients received trastuzumab at a MedStar facility.

Using medical record numbers (MRNs) from these702 patients,we queried theMWHCXcelera system for LVEF,left ventricular dimensions and mass, and parameters of diastolic function. 307 patients had an MRN associated with MWHC, and we were able to obtain echocardiogram data for 160 of these patients.

Next, we identified a baseline LVEF measurement for each patient from an echocardiogram acquired within a period of two years prior to the first administration of chemotherapy. This required merging the drug administration information with data from echocardiograms and formulating temporal queries to ensure that the LVEFmeasurements occurred within a two year window prior to trastuzumab administration.   Of the 160 patients we were able to identify 95patients with valid baseline LVEF measurementsand additional measurements after trastuzumab administration.

A patient was determined to have a cardiac event if the LVEF dropped below 50 and by more than 10% below baseline or if the LVEF dropped by more than 16% from baseline. This is consistent with clinical guidelines [11]. Using these guidelines, we identified 21 patients with cardiotoxic events.

We then produced a consolidated file containing the study data for downstream analysis. This required de-identification of PHI including patient MRNs,procedure dates, and the calculation of derived variables like age at baseline and time to cardiotoxic event.Table 1 compiles descriptive statistics about our patient cohort.

Table 1
Characteristics of 95 patients with valid baseline and follow-up LVEF measurements

Characteristic	**No event** **(*N* = 74)**	**Event** **(*N* = 21)**
Follow-up time(days)	356 (41, 1738)	417 (102, 2350)
Time to cardiotoxic event (days)	-	217 (51, 1692)
Age at baseline LVEF (years)$^{\pm}$	55.8 (32.5, 79.0)	56.4 (40.5, 78.7)
Deceased at time of study	8	1
Race$^{\pm}$		
Black or African American	48 (65%)	15 (71%)
White	16 (22%)	3 (14%)
Asian	1 (1%)	2 (10%)
Other	1 (1%)	1 (5%)
Declined/Unknown	8 (11%)	0 (0%)
Trastuzumab		
Total dose (mg)	6547 (755, 9942)	4420 (660, 9890)
Number of administrations	8 (2, 74)	8 (3, 26)
Patient measurements$^{\pm}$		
BMI (kg/m^2)	29.5 (18.8, 68.9)	31.0 (20.4, 40.2)
Systolic BP (mm Hg)	131 (100, 184)	126 (103, 152)
Diastolic BP (mm Hg)	75 (46, 129)	72 (58, 87)
Pulse (bpm)	77 (54, 136)	80 (60, 101)
Baseline echocardiogram measurements [number of missing values]$^{\pm}$		
LVEF	60.0 (40.0, 70.0)	55.0 (35.0, 65.0)
Post wall thickness (cm)[4]	0.99 (0.65, 1.41)	0.99 (0.70, 1.23)
Septal thickness (cm)[4]	1.05 (0.70, 2.04)	1.00 (0.80, 1.36)
LVEDD (cm)[4]	4.11 (3.23, 5.34)	4.49 (3.10, 4.88)
LAD (cm)[5]	3.58 (2.60, 4.80)	3.30 (2.45, 4.53)
LV mass (g)[4]	137.00 (68.82, 276.30)	154.80 (62.19, 219.20)

LVM index (g/m^2)[10]	74.56 (42.83, 143.10)	85.07 (53.74, 115.40)
RWT [4]	0.486 (0.319, 0.725)	0.452 (0.375, 0.669)
±Variable used in survival analysis and probabilisticnetwork modeling		
Total counts (percentages) are given for categorical variables, while median (min, max) values are given for continuous variables.		

Results

Through the extensive data wrangling efforts described above we were able to create adataset for the investigation of cardiotoxic events associated with a specific, widespread breast cancer therapy. We further analyzed the dataset using a combination of data visualizationand modeling techniques to extract clinically meaningful information andcreate a prototype clinical decision support framework. Our analysis was able to reproduce findings from other research groups showing that LVEF at baselineis an important factor for predicting cardiotoxicity[9].

Data Visualization

We employed a number of techniques to visualize the data for our patient cohort. The EventFlow software[12, 13]allowed us to visualize, align, and query large amounts of temporal data related to drug dosing and echocardiogram measurements. Figure 3shows patients aligned by diagnosis.Patients 994 and 995 clearly show a series of low LVEF values after trastuzumab administration, indicating a cardiotoxic event. This software greatly enhanced our multidisciplinary team’s collaborative data analysis sessions by allowing for real-time data exploration and hypothesis generation.

Statistical and Probabilistic Network Modeling

A Cox proportional hazards regression framework was considered, with time to cardiotoxic event as the outcome. In addition to the variables indicated in Table 1, we also considered treatment with 12 other drugs (dichotomous yes/no variables). Race was coded as either “Black or African American” or “Other.” We employed the LASSO approach to identify thevariables most associated with this outcomevia the “glmnet” package in R [14, 15], using leave-one-out cross validation (CV) to select the model with the minimum CV error.Due to several missing echocardiogram measurements, we considered 5 missing data imputations using the “mice” R package[16]. For 4 of the 5 imputed datasets, the recommended modelselected only baseline LVEF; for the remaining imputed dataset, it additionally selectedseptal thickness.In a univariate Cox regression,baseline LVEF showed a highly significant association with time to cardiotoxic event (estimated HR per 10 units = 0.45, p=0.0048).

In addition, a probabilistic network model of cardiotoxicity was developed using the BayesiaLab software [17]. The Markov Blanket learning algorithm[18]was applied to the data using the K-Means method with K=3 bins to discretize the continuous variables and a structural complexity influence coefficient of 0.35. This parameter balances data fitting versus network complexity in learning algorithm. In addition to the variables indicated in Table 1, the TNM staging variables (stage, primary tumor, regional lymph nodes, distant metastasis, and grade) and treatment information with 14 other drugs – including 14 patients treated with doxorubicin, an anthracycline – were included in the analysis. The algorithm identified baseline LVEF and left ventricular end diastolic dimension (LVEDD)as important factors for predicting cardiotoxicity. After performing 5-Fold CV only LVEF was found to be robust as it was included in 4 of the 5 models generated. LVEDD was only found in 2 of the 5 models generated.The model had an in-sample classification accuracy of 82% and 5-Fold CV accuracy of 75%. Probabilistic networks are valuable for hypothesis generation and can be dynamically queried to produce belief measures given information that is known about a patient. For example, our multidisciplinary team was able to play “what if” by setting the model variables to specific values of interest to dynamically compute the probability of a cardiotoxic event.

In summary, we successfully created a valuabledataset for use in cardiotoxicity research and demonstrated the use of this dataset by creating predictive models of cardiotoxicity. We hope that others can benefit from our experiences and that this methodology can be extended to other disease areas.

Discussion

Our study provided useful data for assessing cardiotoxicity in breast cancer patients treated with trastuzumab, as well as an important set of lessons learned. A major challenge in using clinical data in Hospital EHR systems is that EHR systems are largely designed for single patient clinical care and not forresearch on large patient cohorts. Bulk EHR data extraction is a major challenge and requires intimate knowledge of backend data stores as well as technical and administrative access. The majority of the time spent on this effort involved getting access to and wrangling data from numerous clinical systems in order to create a structured dataset which could be used for statistical analysis and modeling. A further challenge we faced was incomplete or missing data. Out of a total of 11,560 female breast cancer patients, only 702 were recorded to have received trastuzumab. However, we would expect 20-25% of the breast cancer patients to have HER2-positive tumorsand most of those patients to be treated with trastuzumab, which means we are only capturing 24%-30% of the expected patient population. Many patients were also excluded from our analysis because we could not locate valid LVEF measurements for a patient or determine a baseline LVEF value. This illustrates the general problem of incomplete data in EHR systems that can be due to patients being seen at outside institutions or being lost to follow-up. In addition, some of the patients with missing LVEF data may have had Multi Gated Acquisition Scans (MUGA) rather than echocardiograms and therefore their LVEF values would not be represented in our data set. In the future MUGA data can be extracted from patient records increasing the number of patients in the analysis.

Data completeness can also be improved by using a proxy diagnoses to supplement missing ICD-9 codes. For example a female treated with Trastuzumab having evidence of breast cancer in her medical record can be regarded as a breast cancer patient even if there is no breast cancer associatedICD-9 code found. This technique will allow more patients to be considered in the analysis.

Another issue is the possible overrepresentation of cases due to medical surveillance bias. Patients having complete data in our data set may have been followed closely by the treating physician because they were determined to be at high risk for a cardiotoxic event. This would result in an overrepresentation of high risk patients in our analysis set.We found 21 potential events in the 95 patients with sufficient data for analysis. This represents a 22% incidence rate, whichis significantly higherthan the incidence rate typically found in clinical practice when using trastuzumab alone or with non-anthracycline chemotherapy agents. For example, a retrospective analysis of patients enrolled in seven phase II and III clinical trials found that 3% - 7% of patients receiving only trastuzumab had cardiac dysfunction, while 27% of patients receiving trastuzumab and anthracycline plus cyclophosphamide had cardiac dysfunction; however, most of the patients receiving only trastuzumab had been on anthracycline therapy previously[19].In our study, only 14 patients were treated concomitantly treated with doxorubicin – a type of anthracycline, 4 of them having cardiotoxic events, which made it difficult to perform inference on this group.

Once created, the structured dataset allowed us to conduct collaborative data exploration and analysis with a multidisciplinary team of clinician researchers and informatics scientists. The common variable found to be associated with cardiotoxicity across our analyses was low baseline LVEF. We note that 9 patients in our study had LVEF below the normal limit of 50. In practice, the decision to consider trastuzumab needs to weigh risks and benefits. We also know that at least 3 of these 9 patients were in the SAFE-HEaRt study, which enrolled patients with baseline LVEF between 40 and 50% who are on applicable heart medications[20].We also cannot exclude the possibility that some of them had a later baseline LVEF measurement – possibly through MUGA – which we missed.

Research and Clinical applications

COVID19 patients with cancer could be at higher risk of adverse cardiac events as a result of cancer treatment[21].NCI has launched the COVID-19 in Cancer Patients Study (NCCAPS)[22], which will help answer questions about COVID-19's impact on cancer patients. The study is now open to adults and will later be expanded to include children. Our work on this paper could be applied to thisusecase to reduce the number of cardiotoxic events experienced by COVID19 patients with cancer.

In 2019, Pishvaian et al [23]presented a new concept called Virtual Molecular Tumor Board (VTMB), which allowedclinicians to combine expert-curated data and data from clinical systems along with data from molecular diagnostics(MolDx) reports to develop consensus on treatments. It usesinterconnected cloud based virtual computing techniques and reduced the time needed for a clinician to assess a patient's tumor profile and suitability for clinical trials from 14 to 4 days. The cleaned data set and the predictive models from this paper could be used in conjunction with the information presented at a VTMB to enable bettermatches for clinical trials and also reduce the number of cardiotoxic events experienced by patients in the clinical trials.

Conclusions And Future Work

We plan to continue our work to develop a rich resource that connects clinical cardiology and cancer data by refining our predictive models and adding new data sources such as data extracted from unstructured clinical notes using natural language processing techniques and data from other institutions. Increasing the number of patients in our analysis will enable us to create more accurate models which will lead to a better understanding of cardiotoxicity in real world breast cancer patients. This will ultimately lead to the development of decision support tools in an oncology setting with the goal of reducing the number of cardiotoxic events experienced by patients.

List Of Abbreviations

PHI - Personal health information

EHR - Electronic health records

HER2/neu - Human epidermal growth factor receptor 2

LVEF - Left ventricular ejection fraction

LVSD - Left ventricular systolic dysfunction

GU IRB - Georgetown-Howard Universities Center for Clinical and Translational Science Institutional Review Board

CHF - Congestive heart failure

MWHC - MedStar Washington Hospital Center

MRN - Medical record numbers

CV - Cross validation

LVEDD - Left ventricular end diastolic dimension

MUGA - Multi Gated Acquisition Scans

VTMB - Virtual Molecular Tumor Board

MolDx - Molecular diagnostics

NCCAPS - COVID-19 in Cancer Patients Study

Declarations

Ethics approval and consent to participate

This project was evaluated by the Georgetown-Howard Universities Center for Clinical and Translational Science Institutional Review Board and was approved for detailed review (GU IRB #: 2016-1255).

Consent for publication

Not applicable

Availability of data and material

This work uses personal identifiable data (PHI) from patients in Electronic Health Records (EHR) and cannot be made public. De-identified data may be made available from the corresponding author upon reasonable request.

Competing interests

None

Funding

This work was funded by the MedStar-Georgetown Partnership Grant, and partly funded by the Lombardi Cancer Center support grant (NCI P30 CA51008) and the Clinical Genome Resource – Expert Curation and EHR Integration grant (U41 HG009650)

Authors' contributions

MH and SM designed the concept, analyzed the data and drafted the manuscript. BC, RJ, SB, AA, SR, RT, FA extracted the data from various electronic systems and helped with data analysis.KB and YG contributed to review and revision of the manuscript. PP and AB provided the clinical context and motivation for the project as well as reviewed and edited the manuscript. All authors read and approved the final manuscript

Acknowledgements

The authors would like to thank Peter B. McGarvey for his assistance with this manuscript.

References

1. Cole TS, Frankovich J, Iyer S, Lependu P, Bauer-Mehren A, Shah NH: **Profiling risk factors for chronic uveitis in juvenile idiopathic arthritis: a new model for EHR-based research**. *Pediatr Rheumatol Online J* 2013, **11**(1):45.

2. Baron KB, Brown JR, Heiss BL, Marshall J, Tait N, Tkaczuk KH, Gottlieb SS: **Trastuzumab-induced cardiomyopathy: incidence and associated risk factors in an inner-city population**. *J Card Fail* 2014, **20**(8):555-559.
3. Collins FS, Hudson KL, Briggs JP, Lauer MS: **PCORnet: turning a dream into reality**. *J Am Med Inform Assoc* 2014, **21**(4):576-577.
4. Di Cosimo S: **Heart to heart with trastuzumab: a review on cardiac toxicity**. *Target Oncol* 2011, **6**(4):189-195.
5. Sawaya H, Sebag IA, Plana JC, Januzzi JL, Ky B, Tan TC, Cohen V, Banchs J, Carver JR, Wiegers SE *et al*: **Assessment of echocardiography and biomarkers for the extended prediction of cardiotoxicity in patients treated with anthracyclines, taxanes, and trastuzumab**. *Circ Cardiovasc Imaging* 2012, **5**(5):596-603.
6. Iyer SV, Harpaz R, LePendu P, Bauer-Mehren A, Shah NH: **Mining clinical text for signals of adverse drug-drug interactions**. *J Am Med Inform Assoc* 2014, **21**(2):353-362.
7. de Azambuja E, Bedard PL, Suter T, Piccart-Gebhart M: **Cardiac toxicity with anti-HER-2 therapies: what have we learned so far?***Targeted oncology* 2009, **4**(2):77-88.
8. Schlitt A, Jordan K, Vordermark D, Schwamborn J, Langer T, Thomssen C: **Cardiotoxicity and oncological treatments**. *Dtsch Arztebl Int* 2014, **111**(10):161-168.
9. Bowles EJ, Wellman R, Feigelson HS, Onitilo AA, Freedman AN, Delate T, Allen LA, Nekhlyudov L, Goddard KA, Davis RL *et al*: **Risk of heart failure in breast cancer patients after anthracycline and trastuzumab treatment: a retrospective cohort study**. *J Natl Cancer Inst* 2012, **104**(17):1293-1305.
10. Plana JC, Galderisi M, Barac A, Ewer MS, Ky B, Scherrer-Crosbie M, Ganame J, Sebag IA, Agler DA, Badano LP *et al*: **Expert consensus for multimodality imaging evaluation of adult patients during and after cancer therapy: a report from the American Society of Echocardiography and the European Association of Cardiovascular Imaging**. *J Am Soc Echocardiogr* 2014, **27**(9):911-939.
11. Plana JC, Galderisi M, Barac A, Ewer MS, Ky B, Scherrer-Crosbie M, Ganame J, Sebag IA, Agler DA, Badano LP: **Expert consensus for multimodality imaging evaluation of adult patients during and after cancer therapy: a report from the American Society of Echocardiography and the European Association of Cardiovascular Imaging**. *Journal of the American Society of Echocardiography* 2014, **27**(9):911-939.
12. **EventFlow: Visual Analysis of Temporal Event Sequences and Advanced Strategies for Healthcare Discovery** [https://hcil.umd.edu/eventflow/] Last Accessed June 9 2020
13. Megan Monroe RL, Catherine Plaisant, Ben Shneiderman: **Temporal Event Sequence Simplification**. In: *TVCG: IEEE Transactions on Visualization and Computer Graphic: 2013*; 2013.
14. Friedman J, Hastie T, Tibshirani R: **Regularization paths for generalized linear models via coordinate descent**. *Journal of statistical software* 2010, **33**(1):1.
15. Simon N, Friedman J, Hastie T, Tibshirani R: **Regularization paths for Cox's proportional hazards model via coordinate descent**. *Journal of statistical software* 2011, **39**(5):1-13.

16. Buuren Sv, Groothuis-Oudshoorn K: **mice: Multivariate imputation by chained equations in R**. *Journal of statistical software* 2011, **45**(3).
17. **BayesiaLab** [https://www.bayesialab.com/] Last Accessed June 9 2020
18. Harris M, Bhuvaneshwar K, Natarajan T, Sheahan L, Wang D, Tadesse MG, Shoulson I, Filice R, Steadman K, Pishvaian MJ *et al*: **Pharmacogenomic characterization of gemcitabine response–a framework for data integration to enable personalized medicine**. *Pharmacogenet Genomics* 2014, **24**(2):81-93.
19. Seidman A, Hudis C, Pierri MK, Shak S, Paton V, Ashby M, Murphy M, Stewart SJ, Keefe D: **Cardiac dysfunction in the trastuzumab clinical trials experience**. *Journal of Clinical Oncology* 2002, **20**(5):1215-1221.
20. **Cardiac Safety Study in Patients With HER2 + Breast Cancer (SAFE-HEaRt)** [https://clinicaltrials.gov/ct2/show/NCT01904903] Last Accessed June 9 2020
21. Ganatra S, Hammond SP, Nohria A: **The Novel Coronavirus Disease (COVID-19) Threat for Patients with Cardiovascular Disease and Cancer**. *JACC CardioOncol* 2020.
22. **NCI COVID-19 in Cancer Patients Study (NCCAPS)** [https://www.cancer.gov/research/key-initiatives/covid-19/coronavirus-research-initiatives/nccaps] Last Accessed June 8 2020
23. Pishvaian MJ, Blais EM, Bender RJ, Rao S, Boca SM, Chung V, Hendifar AE, Mikhail S, Sohal DPS, Pohlmann PR *et al*: **A virtual molecular tumor board to improve efficiency and scalability of delivering precision oncology to physicians and their patients**. *JAMIA Open* 2019, **2**(4):505-515.

Figures

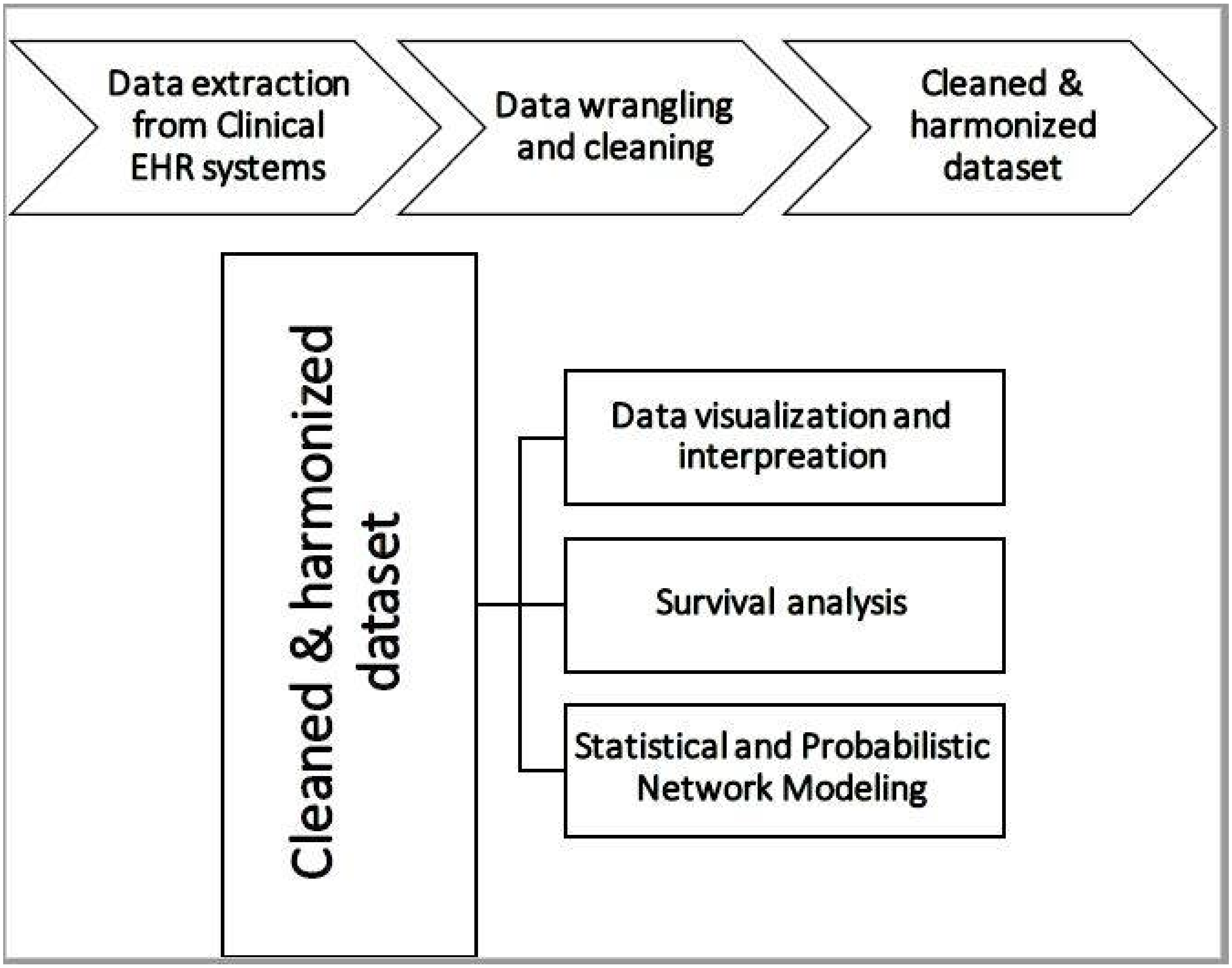
Data extraction from Clinical EHR systems
Data wrangling and cleaning
Cleaned & harmonized dataset
Cleaned & harmonized dataset
Data visualization and interpreation
Survival analysis
Statistical and Probabilistic Network Modeling

Figure 1

Overall Analysis workflow

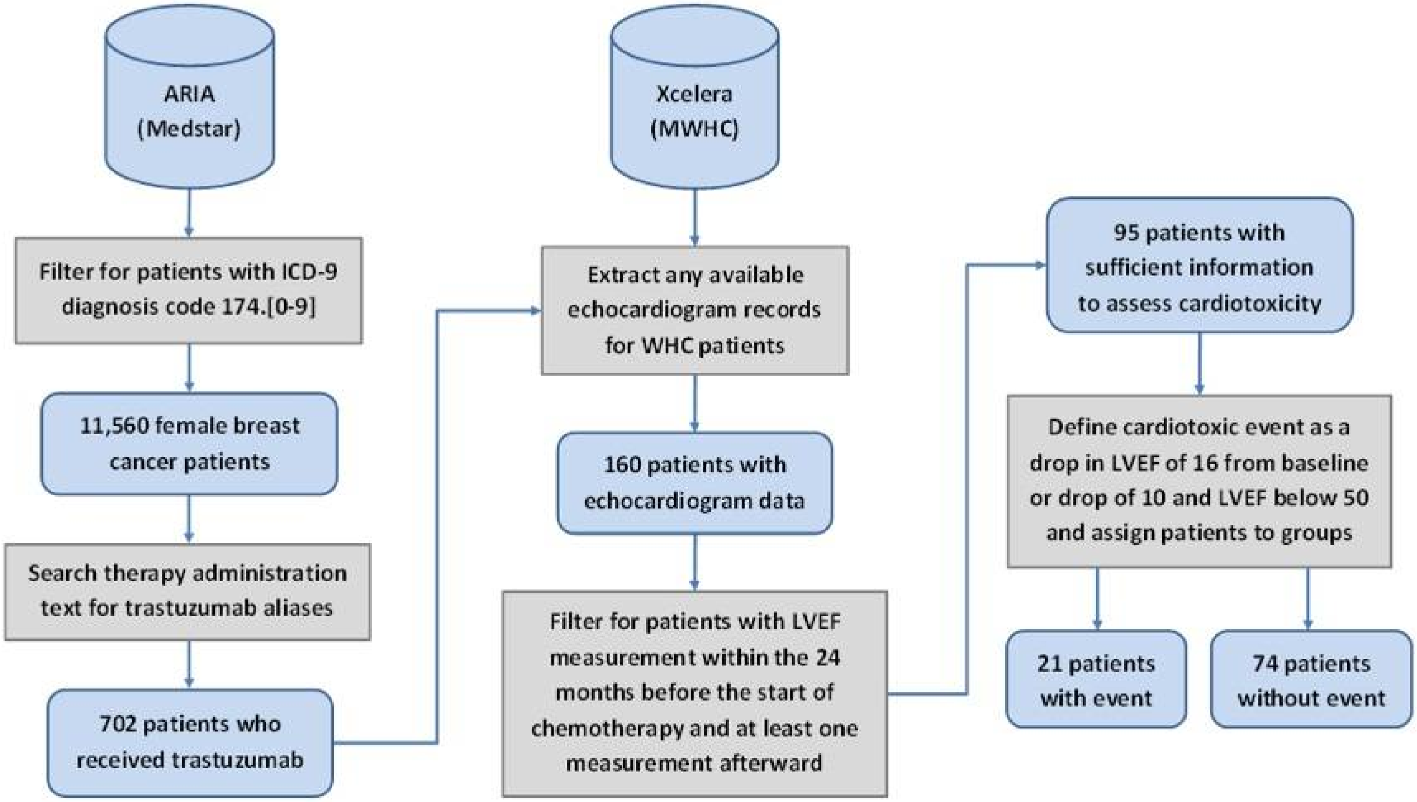
ARIA (Medstar)
Filter for patients with ICD-9 diagnosis code 174.[0-9]
11,560 female breast cancer patients
Search therapy administration text for trastuzumab aliases
702 patients who received trastuzumab
Xcelera (MWHC)
Extract any available echocardiogram records for WHC patients
160 patients with echocardiogram data
Filter for patients with LVEF measurement within the 24 months before the start of chemotherapy and at least one measurement afterward
95 patients with sufficient information to assess cardiotoxicity
Define cardiotoxic event as a drop in LVEF of 16 from baseline or drop of 10 and LVEF below 50 and assign patients to groups
21 patients with event
74 patients without event

Figure 2

Data extraction and filtering workflow

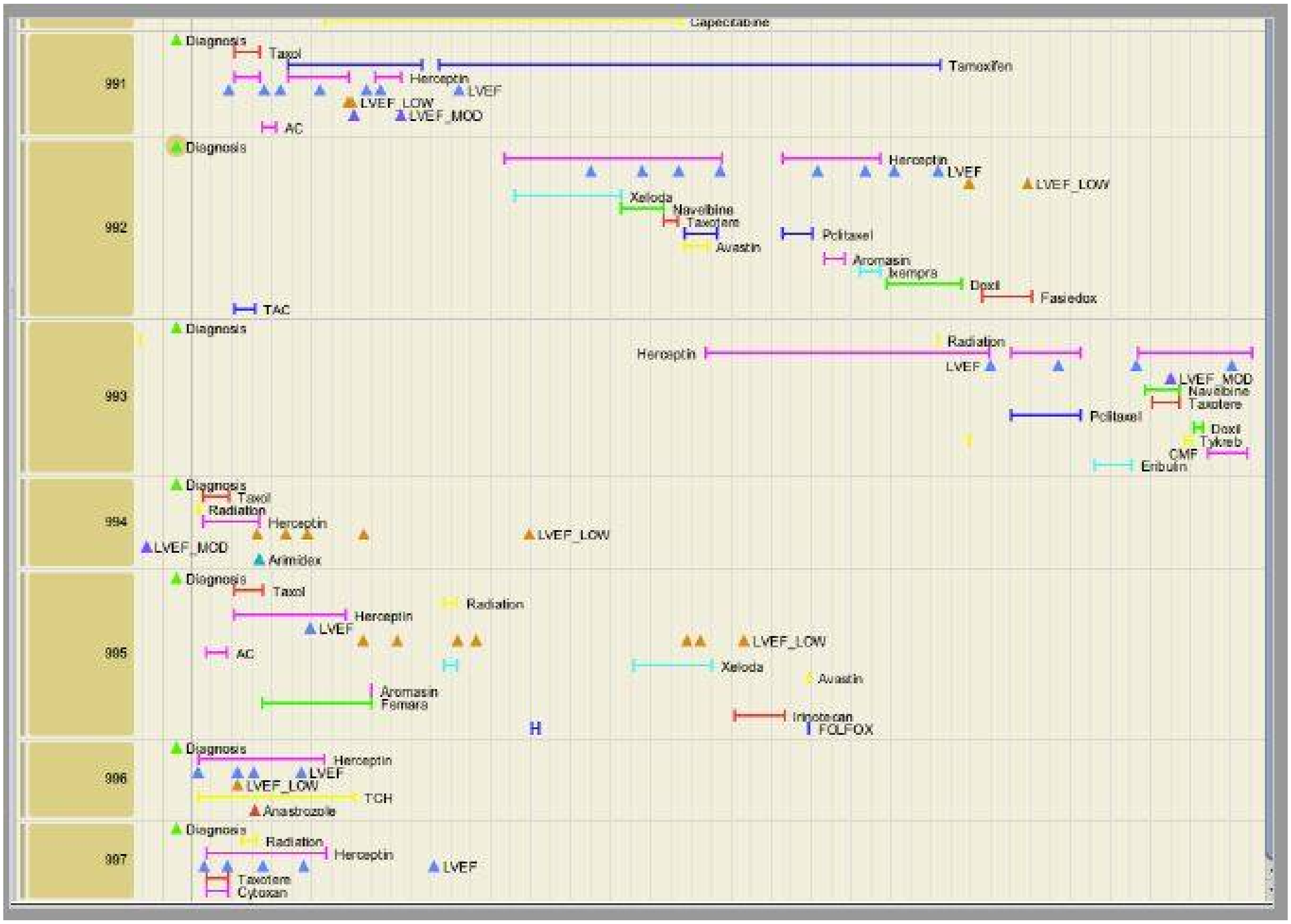

Figure 3

EventFlow visualization of patient data

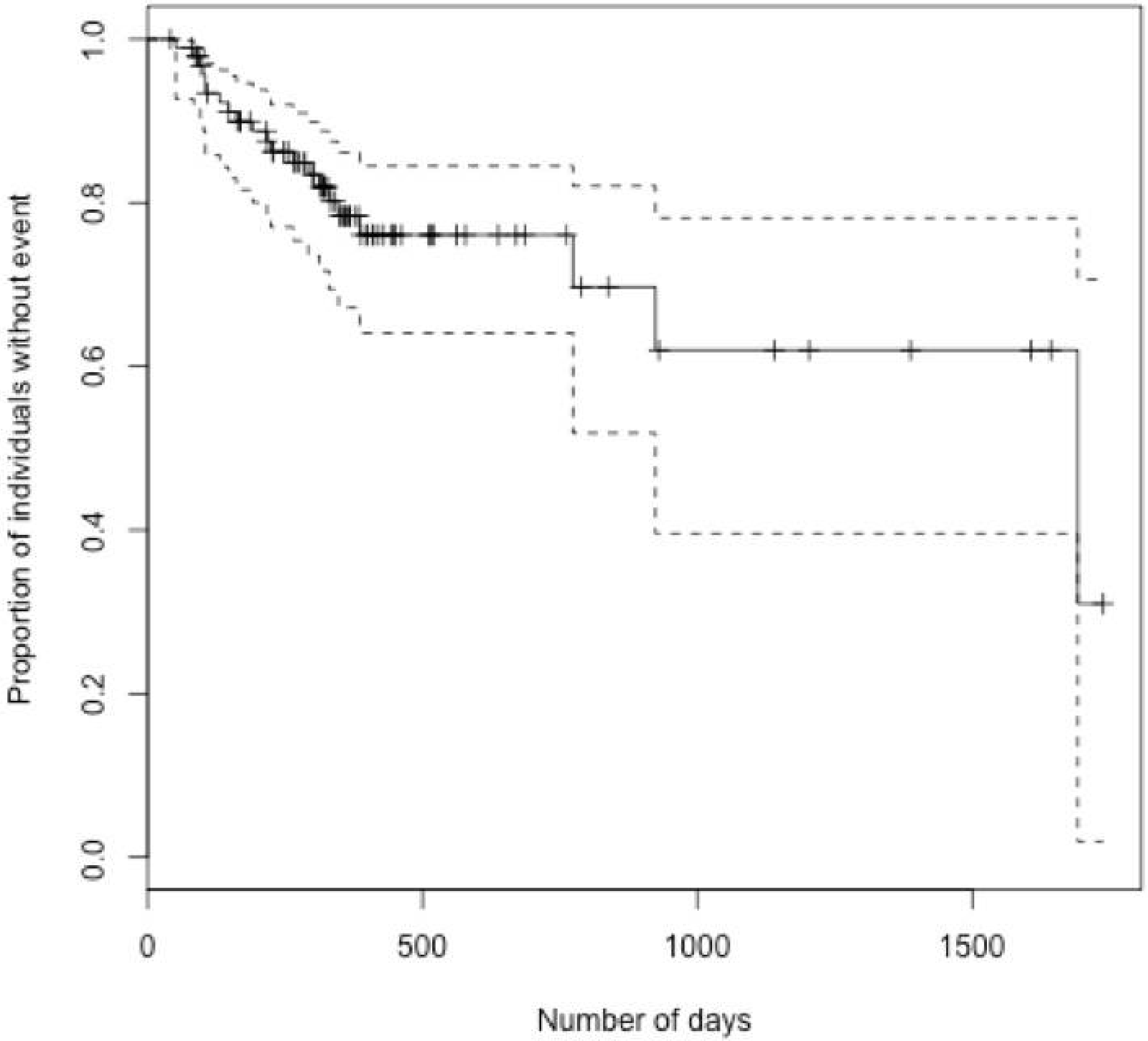

Figure 4

Kaplan-Meier Plot showing time to cardiotoxic event. The dashed lines indicate the 95% confidence interval. The vertical lines show censored individuals