Original Study

Improving Collection of Real-World Data: The Experience of the *Joven & Fuerte* Prospective Cohort for Mexican Young Women With Breast Cancer

Cynthia Villarreal-Garza,[1,2,#] Ana S. Ferrigno,[1,2,#] Fernanda Mesa-Chavez,[1,2] Alejandra Platas,[1,3] Melina Miaja,[1,2] Alan Fonseca,[1,3] Marlid Cruz-Ramos,[1,3] Alejandro Mohar,[1,3] Juan E. Bargallo-Rocha,[1,3] Judy-Anne W. Chapman[4]

Abstract

The prospective collection of clinical data can generate detailed information on heterogeneous populations. This article reviews the strengths and limitations of the collection of real-world data and provides insight into the feasibility of routine collection of high-quality evidence even in a resource-constrained setting. The acquisition of high-quality data to assess the clinical and psychosocial needs of young Mexican patients with breast cancer has been enhanced through the use of preplanned, standardized data definitions and instrumentation to provide internally and externally comparable results, optimization of data collection with web-based surveys, engagement of participants to minimize missing data, and routine review for data consistency. A similar approach by other research groups could improve the quality of real-world data and accomplish enhanced inference of information.

Clinical Breast Cancer, Vol. 000, No.xxx, 1–6
Keywords: Clinical research, Global oncology, Point-of-care data, Research design, Standardized instrumentation

Introduction

The practice of evidence-based medicine has led the research community to build a hierarchy of the strength of evidence according to research design.[1] Results from randomized clinical trials (RCTs) are considered of the highest quality, as their strict methodology minimizes confounding effects and permits inference about causal relationships. Furthermore, RCTs can provide insights into the studied population by disclosing information regarding participant characteristics, prior and ongoing treatment strategies, prognosis, patient-reported outcomes, and experienced adverse events. However, narrow eligibility criteria, inability to study complex interactions, and elevated costs are considerable limitations for investigations with this design.[2,3]

Recently, the collection of real-world data (RWD) has garnered substantial interest as a research method that can provide complementary detailed information to findings from RCTs.[3] The strengths of RWD are high external validity, flexibility in possible explorations, and the ability to study complex associations. Although traditional RCTs remain the gold standard for assessing the impact of an intervention, observational data can explore diverse concepts such as disease burden, treatment patterns, and outcomes of heterogeneous populations.[4] Nonetheless, limitations in interpreting observational RWD include the intrinsic embedding of clinical management decisions on broader determinants of patient status, such as prevalent comorbidities,[5] available resources, and regional treatment preferences.

In the last decade, compiling RWD has been enhanced by technological advances that permit the low-cost ubiquitous collection of patient data, which can be used for both research purposes and informing the clinical management of current patients. One of the main strengths of RWD is that the data can be collected from a wide variety of sources that can be broadly categorized into two groups: health-related datasets (including electronic health records, pharmacy dispensing data, insurance claims data, personal digital device data, and electronic instruments measuring patient-reported outcome measures) and non–health-related datasets (such as internet activity data and employment data).[6,7] The purposeful prospec-

[1]Joven & Fuerte: Programa para la Atencion e Investigacion de Mujeres Jovenes con Cancer de Mama, Ciudad de Mexico, Mexico
[2]Centro de Cancer de Mama, Hospital Zambrano Hellion TecSalud, Tecnologico de Monterrey, San Pedro Garza Garcia, Mexico
[3]Departamento de Tumores Mamarios y Departamento de Investigacion, Instituto Nacional de Cancerologia, Ciudad de Mexico, Mexico
[4]Public Health Sciences Department and Canadian Cancer Trials Group, Queen's University, Kingston, Ontario, Canada (retired)

Submitted: Jan 12, 2021; Revised: Mar 20, 2021; Accepted: Apr 11, 2021; Epub: xxx

Address for correspondence: Judy-Anne W. Chapman, 11 Dayman Court, Kitchener, ON N2M 3A1, Canada
E-mail contact: jachapma@aol.com

[#] C.V.-G. and A.S.F. contributed equally to this work as first authors.

1526-8209/$ - see front matter
https://doi.org/10.1016/j.clbc.2021.04.005

Please cite this article as: Cynthia Villarreal-Garza et al, Improving Collection of Real-World Data: The Experience of the Joven & Fuerte Prospective Cohort for Mexican Young Women With Breast Cancer, Clinical Breast Cancer, https://doi.org/10.1016/j.clbc.2021.04.005

tive collection of patient data in real-world settings provides a foundation for designing point-of-care studies that can provide data on diverse clinical scenarios.[2] Thus, the collection of data from representative populations has the potential to achieve a seamless acquisition of disease knowledge that can be extracted from usual clinical practice.

Optimization of RWD

Despite the recent interest in RWD, current investigations with this design are limited by a lack of interoperability among data sources, unstandardized and unstructured raw data that complicate automatic data extraction and hinder comparability, and vulnerability to bias and confounding factors.[6,8] RWD comparability and progress can be enhanced by using common data definitions, established instrumentation, and overall quality improvements in data completeness and internal consistency. Hence, several groups have worked on standardization (or harmonization) of terminology that is independent of study design. Examples of existing standardized definitions are the US National Institutes of Health common data elements (CDEs),[9] cancer Biomedical Informatics Grid Vocabulary and Common Data Element Workspace,[10] standardized laboratory processes, and clinical biomarker cutpoints,[11,12] along with standardized definitions of clinical endpoints.[13] Additionally, online repositories facilitate the contextual comparison of cancer patient attributes; two examples are the European Organisation for Research and Treatment of Cancer (EORTC) established quality of life (QoL) instruments[14] and the National Institutes of Health genomic data-sharing database of Genotypes and Phenotypes (dbGaP).[15] Finally, RWD can be enriched by data aggregation from multiple sources, which can be successfully achieved through patient-centered health-data-sharing platforms in order to provide a more comprehensive picture of the gathered clinical data.[7]

RWD Research Designs

Research involving RWD encompasses a variety of study designs. Particularly, this methodology has been useful for cohort studies, pharmacovigilance trials, and population-based cancer registries.[2,3] Of note, the promotion of large-scale clinical registries with automated data collection and aggregation can generate valuable insights. An example is the US Surveillance, Epidemiology, and End Results (SEER) program, which currently covers population-based registries for approximately 34.6% of the US population.[16] The SEER database includes broad information about patient demographics, primary tumor site, tumor morphology, stage at diagnosis, first-line treatment, and survival.[16] However, no detailed information is provided on disease recurrence, QoL assessment, or biologic tumor investigations. Thus, there is a role for smaller research databases with more in-depth coverage of specific information, especially if these are high-quality databases with data standardization and minimization of missing values. It is important to note that existing databases such as SEER may be improved at specified recorded collection time points, as was illustrated with the SEER April 2020 release; factors in common in different time periods are then assessable using timeline stratification or assessment within and across periods. For example, the new SEER Research Plus version now provides high-level descriptions of first-line radiotherapy and chemotherapy, thus providing more detailed information regarding treatment strategies employed in cancer patients.[17]

RWD in Resource-Limited Settings

In resource-limited settings, RWD is of substantial interest as these regions tend to be under-represented in RCTs because of cultural and infrastructural challenges despite the potential for increased health burden.[18] The obstacles for conducting clinical research in developing countries or other resource-constrained settings such as rural health clinics and non-academic urban health centers include lack of funding, shortage of qualified research and clinical personnel, insufficient motivation to conduct clinical trials, overly complex regulatory systems, inadequate infrastructure, and challenges with data integrity and security, as well as difficult patient recruitment because of cultural and educational barriers.[19-21] Despite these limitations, there has been a recent interest in research utilizing RWD in Latin America,[21] which could be attributed to the flexibility in data sources allowed by this research method and the ability of RWD to be collected in a structured framework by personnel with no research background.

Observational Designs for Under-Serviced Populations

Studies with observational designs are particularly valuable to characterize disease spectra and outcomes in settings where there is limited knowledge about patient characteristics, current management, and psychosocial needs. Such is the case with women $\leq$40 years of age, a unique population affected by breast cancer (BC).[22] Compared with older patients, this group is characterized by an increased prevalence of poor-prognosis biological subtypes, more aggressive clinicopathological characteristics, less favorable survival outcomes, and a higher association with germline pathogenic mutations.[22-24] Furthermore, young patients face age-specific psychosocial issues of vulnerability to emotional distress, decreased adherence to medical recommendations, and susceptibility to treatment-related adverse effects (eg, fertility issues, menopausal symptoms).[22,25-28] The combination of these factors may significantly impact patient management and QoL.[29] However, high-quality data are lacking, as this group is under-represented in RCTs.[22] Hence, RWD could potentially provide important clinical information about this vulnerable population.

Several prospective cohort studies have been established to identify the specific needs of young women with BC in Europe and North America.[30-32] Notable examples include the PYNK and "Young and Strong" programs to improve the care of young women with BC in the United States[33,34]; the POSH cohort in the United Kingdom, which has provided valuable data regarding the effect of germline deleterious mutations in *BRCA* on early-onset BC[35]; and the PREFER study in Italy, which focuses on fertility and pregnancy issues in this population.[36] Nonetheless, young women account for a variable proportion of BC cases worldwide, with a higher incidence in limited-resource regions such as Latin America and Sub-Saharan Africa.[37] In addition, different demographic, cultural, economic, and health system characteristics make it necessary to study young women with BC in low- and middle-income countries.[38] In this work, we demonstrate the feasibility of collecting high-quality RWD regarding BC patients in a resource-constrained setting using CDEs,

Please cite this article as: Cynthia Villarreal-Garza et al, Improving Collection of Real-World Data: The Experience of the Joven & Fuerte Prospective Cohort for Mexican Young Women With Breast Cancer, Clinical Breast Cancer, https://doi.org/10.1016/j.clbc.2021.04.005

established QoL instrumentation, internal technological controls to improve the completeness of data fields, and planned periodic clinical and biostatistical reviews.

Discussion

High-Quality RWD Collection by Joven & Fuerte

To our knowledge, *Joven & Fuerte* (J&F) is the first program dedicated to the care of young BC patients in Latin America.[37] It was created in 2014 by *Médicos e Investigadores en la lucha contra el Cáncer de Mama* (*MILC*), a non-governmental organization dedicated to promoting quality care of BC patients in Mexico. The program itself has three objectives: (1) provide multidisciplinary care to young BC patients, (2) promote patient and healthcare providers' education about young women's special needs, and (3) study a prospective cohort of young women with BC. This last component is active in three sites: Instituto Nacional de Cancerologia in Mexico City and Hospital San Jose and Hospital Zambrano Hellion in Monterrey, Nuevo Leon. Patients who participate in the research component of the program are followed for a period of 5 years, during which relevant demographic characteristics, clinical information, and patient-reported outcomes are collected. The embedding of the prospective cohort in a program financed by a non-governmental organization has allowed the project to be sustainable with the support of grants from organizations such as Susan G. Komen, Avon Foundation, Cimab Foundation, and pharmaceutical companies, as well as donations from the general public.

Recently, an interim analysis of the J&F pilot phase cohort yielded important information about the characteristics of young Mexican women with BC and the performance of the study protocol.[39] Based on the viability of our approach, we propose several actions that may optimize the general collection of RWD. Specifically, our recommendations focus on (1) preplanned uniform data definitions, (2) optimization of data collection, (3) minimization of selection bias, (4) sustained participants' engagement, (5) promotion of data availability, (6) routine central review for data consistency and validity, and (7) reduction of research costs (Figure 1).

Preplanned Uniform Data Definitions. Preplanned uniform data definitions ensure that the study adequately assesses established endpoints and that the generated data achieve internally and externally comparable results. Demographic and clinical characteristics collected by J&F were selected and culturally adapted by a multidisciplinary team, with inspiration from the PYNK[34] and Young and Strong[33] programs for young women with BC. In addition, previously validated tools for relevant patient-reported issues were incorporated. These instruments include the EORTC QoL Questionnaire Core 30 (QLQ-C30) and Breast Cancer–Specific QoL Questionnaire (QLQ-BR23),[14] the Hospital Anxiety and Depression Scale (HADS),[40] the Female Sexual Functioning Index,[41] and the Sexual Satisfaction Inventory.[42] Moreover, researchers uniformly extract relevant clinicopathological information by filling an electronic survey based on the National Cancer Institute Breast Oncology Local Disease CDEs,[43] and the Standardized Definitions for Efficacy End Points system[13]. The preplanned definitions have allowed the aggregation and comparability of data from patients managed by different physicians in diverse centers. Additionally, the previously validated instruments enable the direct comparability of the results with previous studies.

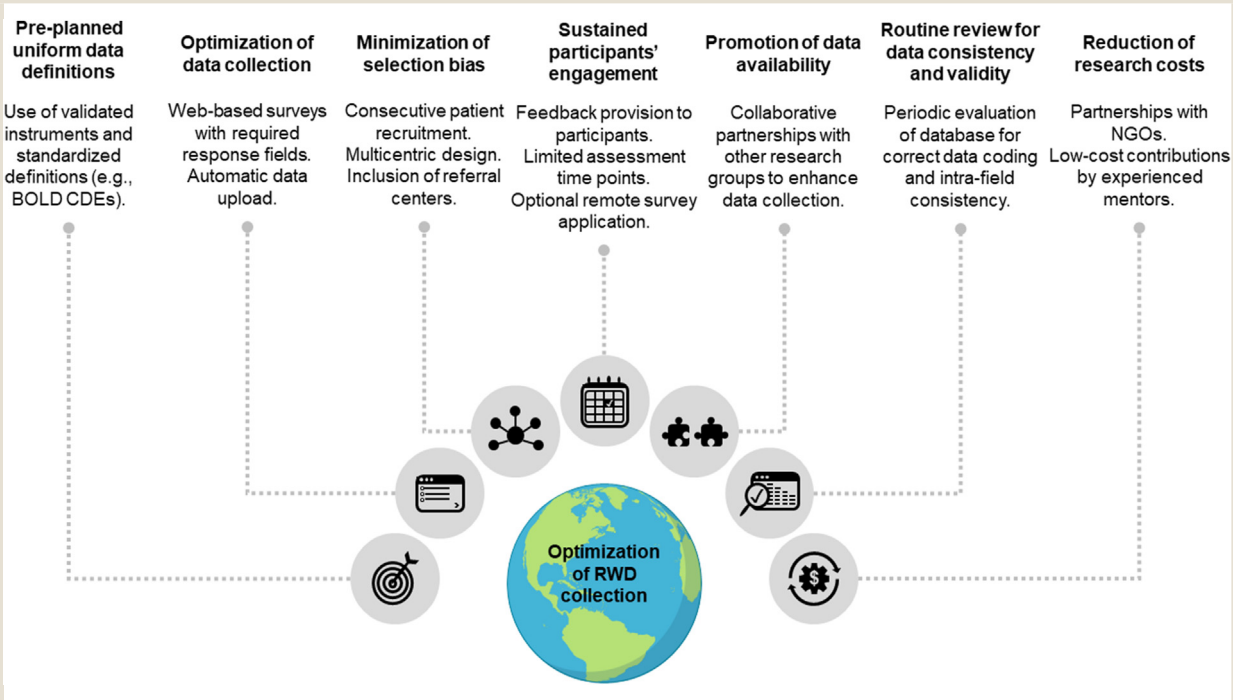
Optimization of Data Collection. In J&F, optimization of data collection is achieved through web-based surveys that are programmed with formatted entry fields that have items that accept responses of only predefined data types (eg, only unannotated numeric or prespecified date formats are accepted for certain items), as well as items with prespecified possible entries that are offered as a drop-down choice or multiple-choice selection. Information is collected by program-supplied tablets at baseline, 6 months, 1 year, and biennially for up to 5 years of follow-up. For both patient-reported outcomes and clinical information surveys, a response is required for each item before proceeding to the next, effectively minimizing missing data. Upon completion, the results of each questionnaire are automatically uploaded into an anonymized central electronic database. Thus, researchers are spared the transcription of responses, and the need for data curation before statistical analysis is reduced.

Minimization of Selection Bias. The minimization of selection bias is promoted in J&F by consecutive patient recruitment at three institutions that are BC referral centers in Northern and Central Mexico with different healthcare coverage schemes (ie, private or public health insurance). For context, the Mexican healthcare system is fragmented into three main components that operate in parallel: (1) social insurance schemes, which cover all salaried employees in the formal sector of the economy (~40% of the population are private sector or federal government employees); (2) public assistance schemes, which traditionally provided basic medical services to the uninsured population (~44% of Mexicans) but are now available to any individual since inauguration of the Institute of Health for Wellbeing (INSABI) in 2020; and (3) private sector, which includes private insurance schemes and private service providers.[44] Although BC management is covered by both social insurance and public assistance schemes (as well as by some private insurance schemes), disparities in access to services and perceived quality of patient-centered care are conspicuous.[45] Thus, the participation of referral centers with different healthcare coverage not only facilitates a continuous accrual of new patients to the J&F cohort but also ensures that a heterogeneous and representative sample of patients is achieved, allowing adequate external validity. However, a limitation is that not all eligible patients have agreed to participate in the research component. For example, 19 of 135 eligible patients (14%) refused accrual to the pilot phase and five of 116 enrolled patients (4%) revoked consent to participate during the follow-up period.[39] Unfortunately, the reasons why the patients refused to participate in the study were not registered. Of 111 patients who maintained enrollment, 21 patients were excluded from the pilot study because of benign breast disease ($n = 7$), distant metastasis at diagnosis ($n = 7$), or death before any survey completion ($n = 7$).[39]

Sustained Participant Engagement. Another potential source of bias is non-random patient attrition and non-completion of data at the prespecified time points. This is reduced in J&F by maintaining the engagement of participants. A particular limitation in our

Please cite this article as: Cynthia Villarreal-Garza et al, Improving Collection of Real-World Data: The Experience of the Joven & Fuerte Prospective Cohort for Mexican Young Women With Breast Cancer, Clinical Breast Cancer, https://doi.org/10.1016/j.clbc.2021.04.005

Figure 1 Recommendations for the Collection of High-Quality RWD
Abbreviations: BOLD CDEs = National Cancer Institute Breast Oncology Local Disease Common Data Elements; NGO = non-governmental organization; RWD = real-world data.



setting is that patients are unlikely to attend an appointment solely to complete research surveys due to financial, geographical, and time constraints. Hence, they have been encouraged to continue their participation with feedback about the clinical usefulness of the collected data (eg, if depression is detected in the HADS questionnaire, the medical team would take appropriate measures to address this situation) and by explaining that the information they provided is important to advance the scientific understanding about their disease, with interim results presented at patient educational events. Our experience indicates that such measures have favorably impacted participation. Furthermore, survey completion rates have been improved by (1) approaching patients personally during scheduled medical appointments, (2) a reduction of time points in which patient reported outcomes (PROs) are collected (eg, PRO questionnaires were originally intended to be collected annually up to 5 years after BC but are now assessed at baseline, 6 months, 1 year, 2-3 years, and 4-5 years after diagnosis); (3) flexibility in allowing questionnaire completion during a 3-month window for each time point; and (4) contacting offsite patients via telephone or email to offer remote completion of surveys. These measures have permitted adequate survey completion rates. In total, 73 of 90 patients were expected to answer pilot-phase follow-up surveys regarding PROs. Of these, 69 to 72 (95%-99%) completed the baseline questionnaires, and 40 to 44 of the 61 non-recurrent patients (68%-76%) completed the surveys at 2 years. Therefore, the 2-year instrument completion rates were comparable with those from other research groups (76%-80%).[46]

Promotion of Data Availability. Research collaborations can be encouraged to increase data availability. Due to resource constraints in our setting, few patients have access to specialized services such as genetic testing or genomic studies. Thus, comprehensive biologic characterization is not uniformly available due to socioeconomic and healthcare system constraints. To promote data availability, J&F encourages collaborative partnerships with other research groups in order to improve the characterization of tumor biology in young women with BC. For example, a collaboration with the City of Hope Clinical Cancer Genomics Community Research Network has allowed most of the J&F cohort to receive free genetic cancer risk assessment and genetic testing.

Routine Central Reviews for Data Consistency and Validity. To ensure the collection of high-quality RWD, clinicians, and biostatisticians in J&F periodically perform routine central reviews for data consistency and reliability (Table 1). In this study, data integrity is ensured through the evaluation of data coding compliance and intrafield consistency every 6 months and prior to any interim analyses.

Reduction of Research Costs. Although RWD collection is generally considered less cost-intensive than conducting RCTs, cost-efficient measures in J&F have been needed to engage collaborators from different disciplines (oncologists, psychologists, nurse navigators, and a biostatistician) and to preserve the integrated use of technology (web-based surveys, tablets, and electronic databases).

Please cite this article as: Cynthia Villarreal-Garza et al, Improving Collection of Real-World Data: The Experience of the Joven & Fuerte Prospective Cohort for Mexican Young Women With Breast Cancer, Clinical Breast Cancer, https://doi.org/10.1016/j.clbc.2021.04.005

Table 1 Data Quality Assurance Steps Performed Every 6 Months

1.	Verification of complete data entry for each participant
2.	Corroboration/correction of data entry types (eg, unannotated predefined numeric; date format mm/dd/yy vs. dd/mm/yy)
3.	Evaluation of internal consistency between fields in past assessments or related fields at a single time point
4.	Resolution of all inconsistencies in clinical data against primary medical records to ensure correct inference

Ongoing support of research grants and donations has operationally sustained the protocol, along with low- or no-cost contributions by experienced mentors who have kept expenses manageable.

Pilot-Phase Clinical Results

The pilot-phase clinical results will be reported when all 90 patients have completed the 5-year follow-up. Longitudinal experience of the ongoing cohort at three recruiting sites and extensions to other sites will permit examinations of temporal and regional validity. Meanwhile, available indicators suggest that the pilot data are representative of the general population of young Mexican patients with BC.[39]

Conclusion

Carefully designed data collection, in conjunction with the widespread availability of electronic capture technology, provides an opportunity to improve the quality of RWD, with increased uniformity and data completeness. Particularly, existing real-world databases could be strengthened by increasing data standardization, optimizing system interoperability, and incorporating efforts to reduce missing data. Database optimization by research groups could improve the quality of RWD available and generate real-world evidence with enhanced inference.

In our cohort of young women with breast cancer, a reasonably complete collection of detailed RWD has been achieved in Mexico through a series of key factors including preplanned data definitions and outcomes, utilization of web-based surveys programmed to minimize missing data, flexibility on how participants are contacted, assurance data reliability, and partnerships with other research groups. Our experience could serve as a precedent for other researchers who aim to collect and analyze RWD, particularly in resource-constrained settings such as ours. Such purposeful point-of-care collection of high-quality data would permit the generation of improved RWD to expedite research advances in representative clinical populations.

CRediT authorship contribution statement

Cynthia Villarreal-Garza: Conceptualization, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration. **Ana S. Ferrigno:** Investigation, Writing – original draft, Writing – review & editing. **Fernanda Mesa-Chavez:** Investigation, Writing – original draft, Writing – review & editing. **Alejandra Platas:** Investigation, Writing – review & editing. **Melina Miaja:** Investigation, Writing – original draft, Writing – review & editing. **Alan Fonseca:** Investigation, Writing – original draft, Writing – review & editing. **Marlid Cruz-Ramos:** Investigation, Writing – original draft, Writing – review & editing. **Alejandro Mohar:** Investigation, Writing – original draft, Writing – review & editing. **Juan E. Bargallo-Rocha:** Investigation, Writing – original draft, Writing – review & editing. **Judy-Anne W. Chapman:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Supervision.

Acknowledgments

We thank Abelardo Meneses, Alexandra Bukowski, Andrea Castro-Sanchez, Ann Partridge (Young and Strong Program for Young Women With Breast Cancer), Bertha Alejandra Martinez-Cannon, Carmen L. Galvez-Hernandez, Cesar J. Marquez-Perez, Christian Aguila, Claudia Pineda, Dione Aguilar, Edna Lopez-Martinez, Ellen Warner (PYNK: Breast Cancer Program for Young Women), Janeth Castro, Jose F. Muñoz-Lozano, Marisol Garcia-Garcia, Mauricio Canavati, Paul Goss (Global Cancer Institute), Pier Ramos-Elias, Regina Barragan-Carrillo, Servando Cardona, Yoatzin Vega, the breast cancer teams of Departamento de Tumores Mamarios–Instituto Nacional de Cancerologia and Centro de Cancer de Mama–Tecnologico de Monterrey, and Richard Zellars of the US National Institutes of Health Breast Oncology Local Disease Task Force.

Disclosure

C. Villarreal-Garza has had a consulting or advisory role with Roche, Novartis, Pfizer, and Eli Lilly; is on the Speakers' Bureau of Roche, Myriad Genetics, and Novartis; has received research funding from AstraZeneca and Roche; and has received travel expense reimbursement from Roche, MSD Oncology, and Pfizer. The remaining authors have stated that they have no conflicts of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.clbc.2021.04.005.

References

1. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. 2011;128:305–310.
2. Khozin S, Blumenthal GM, Pazdur R. Real-world data for clinical evidence generation in oncology. *J Natl Cancer Inst*. 2017;109:djx187.
3. Katkade VB, Sanders KN, Zou KH. Real world data: an opportunity to supplement existing evidence for the use of long-established medicines in health care decision making. *J Multidiscip Healthc*. 2018;11:295–304.
4. Booth CM, Karim S, Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. *Nat Rev Clin Oncol*. 2019;16:312–325.
5. Chapman J-AW, Meng D, Shepherd L, et al. Competing causes of death from a randomized trial of extended adjuvant endocrine therapy for breast cancer. *J Natl Cancer Inst*. 2008;10:252–260.
6. Lipworth W. Real-world data to generate evidence about healthcare interventions: the application of an ethics framework for Big Data in healthcare research. *Asian Bioeth Rev*. 2019;11:289–298.

Please cite this article as: Cynthia Villarreal-Garza et al, Improving Collection of Real-World Data: The Experience of the Joven & Fuerte Prospective Cohort for Mexican Young Women With Breast Cancer, Clinical Breast Cancer, https://doi.org/10.1016/j.clbc.2021.04.005

7. Dhruva SS, Ross JS, Akar JG, et al. Aggregating multiple real-world data sources using a patient-centered health-data-sharing platform. *NPJ Digit Med*. 2020;3:60.
8. Guinn D, Wilhelm EE, Lieberman G, Khozin S. Assessing function of electronic health records for real-world data generation. *BMJ Evid Based Med*. 2019;24:95–98.
9. National Library of Medicine. Common Data Elements (CDE) Repository. Available at: https://cde.nlm.nih.gov/home. Accessed: June 25, 2020.
10. National Cancer Institute. caBIG-NCI Data Standards. Available at: https://wiki.nci.nih.gov/display/caDSR/caBIG-NCI+Data+Standards. Accessed: June 25, 2020.
11. Allison KH, Hammond MEH, Dowsett M, et al. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. *J Clin Oncol*. 2020;38:1346–1366.
12. Wolff AC, Hammond MEH, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast. *J Clin Oncol*. 2013;31:3997–4013.
13. Hudis CA, Barlow WE, Costantino JP, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol*. 2007;25:2127–2132.
14. Cerezo O, Oñate-Ocaña LF, Arrieta-Joffe P, et al. Validation of the Mexican-Spanish version of the EORTC QLQ-C30 and BR23 questionnaires to assess health-related quality of life in Mexican women with breast cancer. *Eur J Cancer Care*. 2012;21:684–691.
15. National Institute of Allergy and Infectious Diseases. Genomic Data Submission and Release Guidelines. Available at: https://www.niaid.nih.gov/research/genomic-data-submission-and-release-guidelines. Accessed: June 25, 2020.
16. National Cancer Institute. SEER Incidence Data, 1975-2017. Available at: https://seer.cancer.gov/data/. Accessed June 25, 2020.
17. National Cancer Institute. SEER Data Change History. Available at: https://seer.cancer.gov/data/2020-data-changes.html. Accessed: March 10, 2021.
18. Mbuagbaw L, Thabane L, Ongolo-Zogo P, Lang T. The challenges and opportunities of conducting a clinical trial in a low resource setting: the case of the Cameroon mobile phone SMS (CAMPS) trial, an investigator initiated trial. *Trials*. 2011;12:145.
19. Alemayehu C, Mitchell G, Nikles J. Barriers for conducting clinical trials in developing countries-a systematic review. *Int J Equity Health*. 2018;17:37.
20. Rolfo C, Caglevic C, Bretel D, et al. Cancer clinical research in Latin America: current situation and opportunities. Expert opinion from the first ESMO workshop on clinical trials, Lima, 2015. *ESMO Open*. 2016;1.
21. Justo N, Espinoza MA, Ratto B, et al. Real-world evidence in healthcare decision making: global trends and case studies from Latin America. *Value Health*. 2019;22:739–749.
22. Paluch-Shimon S, Pagani O, Partridge AH, et al. ESO-ESMO 3rd international consensus guidelines for breast cancer in young women (BCY3). *Breast*. 2017;35:203–217.
23. Azim HA, Partridge AH. Biology of breast cancer in young women. *Breast Cancer Res*. 2014;16:427.
24. Fredholm H, Eaker S, Frisell J, Holmberg L, Fredriksson I, Lindman H. Breast cancer in young women: poor survival despite intensive treatment. *PLoS One*. 2009;4:e7695.
25. Wassermann J, Gelber SI, Rosenberg SM, et al. Nonadherent behaviors among young women on adjuvant endocrine therapy for breast cancer. *Cancer*. 2019;125:3266–3274.
26. Hubbeling HG, Rosenberg SM, González-Robledo MC, et al. Psychosocial needs of young breast cancer survivors in Mexico City, Mexico. *PLoS One*. 2018;13.
27. Suter MB, Pagani O. Should age impact breast cancer management in young women? Fine tuning of treatment guidelines. *Ther Adv Med Oncol*. 2018;10.
28. Lambertini M, Peccatori FA, Demeestere I, et al. Fertility preservation and post-treatment pregnancies in post-pubertal cancer patients: ESMO clinical practice guidelines. *Ann Oncol*. 2020;31:1664–1678.
29. Rosenberg SM, Partridge AH. Management of breast cancer in very young women. *Breast*. 2015;24:S154–S158.
30. Ruggeri M, Pagan E, Bagnardi V, et al. Fertility concerns, preservation strategies and quality of life in young women with breast cancer: baseline results from an ongoing prospective cohort study in selected European Centers. *Breast*. 2019;47:85–92.
31. Rosenberg SM, Ruddy KJ, Tamimi RM, et al. *BRCA1* and *BRCA2* mutation testing in young women with breast cancer. *JAMA Oncol*. 2016;2:730–736.
32. Maishman T, Cutress RI, Hernandez A, et al. Local recurrence and breast oncological surgery in young women with breast cancer: the POSH observational cohort study. *Ann Surg*. 2017;266:165–172.
33. Partridge AH, Ruddy KJ, Kennedy J, Winer EP. Model program to improve care for a unique cancer population: young women with breast cancer. *J Oncol Pract*. 2012;8:e105–e110.
34. Ali A, Warner E. PYNK: breast cancer program for young women. *Curr Oncol*. 2013;20:e34.
35. Copson ER, Maishman TC, Tapper WJ, et al. Germline BRCA mutation and outcome in young-onset breast cancer (POSH): a prospective cohort study. *Lancet Oncol*. 2018;19:169–180.
36. Lambertini M, Fontana V, Massarotti C, et al. Prospective study to optimize care and improve knowledge on ovarian function and/or fertility preservation in young breast cancer patients: results of the pilot phase of the PREgnancy and FERtility (PREFER) study. *Breast*. 2018;41:51–56.
37. Villarreal-Garza C, Lopez-Martinez EA, Muñoz-Lozano JF, Unger-Saldaña K. Locally advanced breast cancer in young women in Latin America. *Ecancermedicalscience*. 2019;13:894.
38. Gálvez-Hernández CL, González-Robledo MC, Barragán-Carrillo R, Villarreal–Garza C. Special needs of young women with breast cancer in limited resource settings. *Rev Invest Clin*. 2017;69:210–222.
39. Villarreal-Garza C, Platas A, Miaja M, et al. Young women with breast cancer in Mexico: results of the pilot phase of the Joven & Fuerte prospective cohort. *JCO Glob Oncol*. 2020;6:395–406.
40. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67:361–370.
41. Rosen R, Brown C, Heiman J, et al. The female sexual function index (FSFI): a multidimensional self-report instrument for the assessment of female sexual function. *J Sex Marital Ther*. 2000;26:191–205.
42. Álvarez-Gayou Jurgenson JL, Honold Espinosa JA, Millán Álvarez P. ¿Qué hace buena una relación sexual?: percepción de un grupo de mujeres y hombres mexicanos y diseño de una escala. *Archivos Hispanoamericanos de sexología*. 2005;11:91–110.
43. National Cancer Institute. Breast Cancer Steering Committee. Available at: https://www.cancer.gov/about-nci/organization/ccct/steering-committees/nctn/breast-cancer. Accessed: April 6, 2020.
44. González Block MÁ, Reyes Morales H, Hurtado LC, Balandrán A, Méndez E. Mexico: health system review. *Health Syst Transit*. 2020;22:1–222.
45. Doubova SV, Martinez-Vega IP, Infante-Castañeda C, Aranda-Flores CE, Knaul FM, Pérez-Cuevas R. Social inequalities in supportive care needs and quality of patient-centered care of cancer patients in Mexico. *Support Care Cancer*. 2021;29:1355–1367.
46. Velikova G, Williams LJ, Willis S, et al. Quality of life after postmastectomy radiotherapy in patients with intermediate-risk breast cancer (SUPREMO): 2-year follow-up results of a randomised controlled trial. *Lancet Oncol*. 2018;19:1516–1529.

Please cite this article as: Cynthia Villarreal-Garza et al, Improving Collection of Real-World Data: The Experience of the Joven & Fuerte Prospective Cohort for Mexican Young Women With Breast Cancer, Clinical Breast Cancer, https://doi.org/10.1016/j.clbc.2021.04.005