Genetics in Medicine Open (2025) 3, 101913

Genetics in Medicine OPEN
An Official Journal of the ACMG
www.journals.elsevier.com/genetics-in-medicine-open

ARTICLE

Lessons learned in migrating from one commercial genetics clinical decision-making tool to another: Assessment of data integrity and utilization

Calvin Le[1], Kevin Tatunay[2], Wayne Liu[2], Haibo Lu[3], Nicole-Ann Rodis[2], Thomas Nam[4], Mercy Y. Laurino[2], Marianne E. Dubard-Gault[1,2,5,*]

[1]*University of Washington, Seattle, WA;* [2]*Fred Hutchinson Cancer Center, Seattle, WA;* [3]*CancerIQ, Chicago, IL;* [4]*Ambry Genetics, Laguna Beach, CA;* [5]*Providence Swedish Cancer Institute, Seattle, WA*

ARTICLE INFO

Article history:
Received 22 July 2024
Received in revised form
3 December 2024
Accepted 4 December 2024
Available online 14 December 2024

Keywords:
Data integration
Data management
Family history
Genetics
Health services

ABSTRACT

Purpose: Rapid advancements in information technology have greatly influenced clinicians' engagement with patient data for health maintenance. The electronic health record often contains multiple ways to record risk factors and to identify patients at an elevated genetic risk for cancer. However, these variables exist in multiple forms and disparate locations in each commercial electronic health record solution resulting in significant variations in how family history or genetic data is codified. Furthermore, there is pressure to migrate from one commercial solution to another at times, prompting the need for a process ensuring data integrity during such a transition.

Methods: Between July and December 2023, the genetics team migrated a family history database from one commercial software solution to another. After the data migration of 13,000 patient records, we reviewed 500 randomly selected charts in both support tools to measure the rate of concordance of information transferred.

Results: A total of 461 patient charts were reviewed. Of these, 425 (92.2%) were noted to be concordant. Of the 36 charts that were discordant, 9 had incorrect genetic testing results entered, 19 had missing information, 5 charts contained data recorded on paper before 2017 (legacy data), and 3 had additional information transferred.

Conclusion: There was high data integrity after migration from one commercial software solution to another. Although these results can ease clinicians' concerns after such support tool transitions, our effort also highlights areas for improvement in how family and patient genetic data are recorded and utilized for clinical care and research within an institution.

The APC was waived by the Editor in Chief.

Calvin Le is the first author.

*Correspondence and requests for materials should be addressed to Marianne E. Dubard-Gault, Providence Swedish Cancer Institute, 1221 Madison Street Suite 600, Seattle, WA 98104. *Email address:* marianne.dubard-gault@swedish.org

doi: https://doi.org/10.1016/j.gimo.2024.101913

Introduction

Over the past decade, rapid advancements in health information technology have influenced how clinicians collect, synthesize, and engage with patient health data.[1] In the context of clinical genetics, there exists a wide array of data in patients' electronic health record (EHR) that can be used to identify patients who are at elevated risk for an inherited cancer syndrome, and who may benefit from increased clinical surveillance.[2] These data may exist in several forms, from free-text information that must be abstracted from the EHR to formal pedigrees drawn manually or with the help of commercial software that are not integrated into the EHR.[3] Numerous commercial software solutions are available on the market and similar terminology and conventions are used to record family history. However, there does not exist a universal convention for codifying or combining software-generated, free-text, and patient-reported data in the EHR, leading to wide variations in how these data are available to clinicians.[3,4] Furthermore, there is often a desire or need to migrate from one software solution to another within an institution because different platforms promise improvements in clinical workflow, patient communication, documentation, and tracking.[3,5,6] As a result, ensuring the accuracy of clinical data during these migrations is critical to maintaining the accuracy of subsequent clinical decisions that rely on these data.

The focus of our quality improvement project was to evaluate data integrity in the process of migrating clinical genetics and pedigree information from one commercial software and clinical decision-making tool to another. The results of this project have implications for care delivery and research not only for our institution but for other institutions considering or planning a similar transition.

Materials and Methods

Institutional setting and data transfer

In 2023, our clinical genetics service migrated their family history survey from commercial pedigree software called Progeny to another commercial pedigree and decision-making tool, CancerIQ. The genetics and genetic counseling team requested a master file from the software Progeny of all patients seen since the first data entry in 2017. A master file was downloaded as a comma-separated variables (CSV) file. In addition, Progeny provided us with a raw data dump of all existing pedigree/information as CSV file(s). Master CSV file was shared with CancerIQ via a permission-based/limited SharePoint external link and data integrity was reviewed during monthly meetings with Prevention, Genetics, and High-Risk team and CancerIQ. Our genetic clinic operates by sending 3 reminders to patients referred for consultation so that they can schedule their appointments. When scheduled, they receive a previsit packet that includes a medical history questionnaire and a link to the software so they can enter family history before the visit. Family history is then reviewed during the visit and utilized to guide counseling and testing decisions, and testing can be organized at the same appointment. The results of testing are uploaded in the chart and reviewed with patients. Family history and genetic results are used subsequently for cancer screening and surveillance decision-making.

Data clean up and upload

After receiving the master CSV file, CancerIQ implemented automated data cleaning scripts interspersed with automated and manual check points for data accuracy. The main challenge encountered during this process was the presence of additional commas embedded in the free text entered at follow-up visits. CancerIQ's solution was to implement a series of anchor columns to help orient the cleaning script and guide where the data should reside. Additional challenges included double-quote/single-quote mismatches, as well as incompatible data typing challenges. These were resolved in an automated fashion for most data rows and manual review was required for a few thousand data rows before the data file was ready to be converted into CancerIQ format. The finalized data sheet was then uploaded via a script into the CancerIQ database where it would be viewable by the genetics team on the front-end interface. For the fields that existed in Progeny that did not have a corresponding CancerIQ field, the decision was made to concatenate these variables to the note section in CancerIQ for that patient or family member. The genetics team also made the decision to remove third-degree relatives from data transfer because it would take the CancerIQ team another 3 to 6 months and data available for these relatives were sparse.

Chart review and data analysis

The study team reviewed the electronic health records of 500 randomly selected patients in both Progeny and CancerIQ. Random selection was performed by generating a list of almost 15,000 numbers and selecting 500 numbers that matched the record ID in the Masterfile CSV. These charts were reviewed to determine (1) the year of the patient's referral or appointment, (2) the clinical indication for referral, (3) whether genetic testing had been ordered as a result of the referral, (4) the result of the genetic testing if ordered, including the gene(s) implicated, (5) the underlying personal or family history, and (6) whether the patient met criteria for genetic testing as defined by the National Comprehensive Cancer Network (NCCN) guidelines. Data abstracted from the above categories were compared between Progeny and CancerIQ. Charts were labeled as concordant if there was agreement in the following categories: (1) whether genetic testing was ordered, (2) the results of the testing including correct identification of the gene(s) involved, and (3) if there was sufficient

documentation to demonstrate that the patient met criteria for testing including review of pedigrees.

Results

Of the 500 patient charts initially reviewed, 461 charts were determined to include enough patient information to include in the final analysis. Of the excluded charts, a manual review of the patients' EHR revealed that 29 of these patients did not come to their scheduled appointment, and the remaining 10 charts did not have any information. A consort diagram outlying the classification of data charts is shown in Figure 1. Of the 461 charts, 425 (92.2%) were determined to be concordant between the 2 software solutions, whereas 36 (7.8%) were determined to be discordant. Of the discordant charts, the most common error was an incorrect transfer of a test result, followed by an incomplete transfer of pedigree or family history data.

Appointment and demographic information are displayed in Table 1. Most patients were referred for a personal and/or family history of breast cancer (194 cases, 42.1% of all cases), followed by colon cancer (55 cases, 11.9% of all cases). When recorded, the most common reason for referral was a personal history of cancer, representing 43.0% of cases.

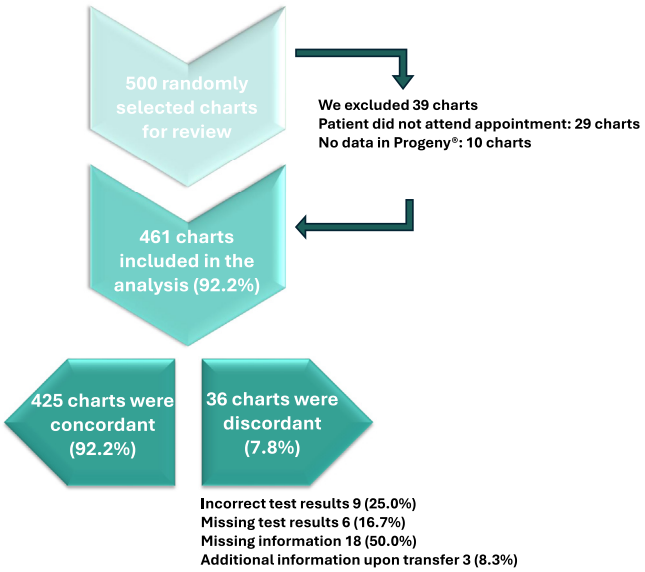

Figure 1 Consort diagram outlining the classification of patient charts included in this study. A total of 500 initial charts were randomly selected, of which 39 were excluded that did not contain any data in the initial Progeny chart. Of these 39 charts, an independent review of the electronic health record confirmed that 29 patients did not attend their appointments. From the 461 remaining charts, 425 were noted to be Concordant, whereas 36 were noted to be Discordant. The source of the error(s) leading to discordance is enumerated, with the incorrect test result being the most common reason. Legacy chart refers to remote records in our system that may not have been fully transferred from paper charts to Progeny in 2016/2017 and are therefore missing pedigree information. "Additional information upon transfer" refers to new diagnoses or tests recorded in CancerIQ that were not present in the original Progeny file. We were not able to identify the source of the additional information.

Table 1 Patient demographic data

Year of Appointment and Referral Indication	Number	Percent (%)
Year of Appointment		
2016	2	0.5
2017	19	4.1
2018	14	3.0
2019	50	10.8
2020	59	12.8
2021	84	18.2
2022	71	15.4
2023	30	6.5
Year of visit not populated	132	28.7
Reason for referral		
Personal history	198	43.0
If personal history, type of cancer		
Breast and ovarian	194	42.1
Colon	55	11.9
Prostate	31	6.7
Other cancer	65	14.1
Not documented	116	25.2
Family history	89	19.3
Test result interpretation	9	2.0
Other reason	165	35.7

Several representative pedigrees encountered during our analysis are included below. Figure 2 depicts a pedigree that was almost identically reproduced upon the transition from Progeny to CancerIQ. In this pedigree, the proband depicted was referred for a personal history of colorectal cancer with intact mismatch repair proteins. Family history was significant for 2 pancreatic cancers and colon and breast cancer on the paternal side. Probands met the criteria for genetic testing on the basis of having had rectal cancer before age 50, and the family met Amsterdam II criteria on maternal side.[7]

From our discordant chart review, we have several examples of incorrect transfer of family or genetic information, including incorrect transfer of parental relationships (Figure 3A and B) and 1 where the personal history of colon cancer and history of maternal aunt with ovarian cancer were not transferred altogether on the pedigree in CancerIQ.

Lastly, in Figure 4, we identified 1 patient who would have missed being offered genetic testing if viewing the family tree from CancerIQ compared with Progeny. The patient met NCCN guidelines on multiple third-degree relatives with breast cancer. Mother of consultand had estrogen-positive breast cancer at 61 and then peritoneal cancer, possibly a mesothelioma.

Discussion

Here, we report on the integrity and quality of a patient-level data transfer involving family history and genetic test results from one commercial software solution to another.

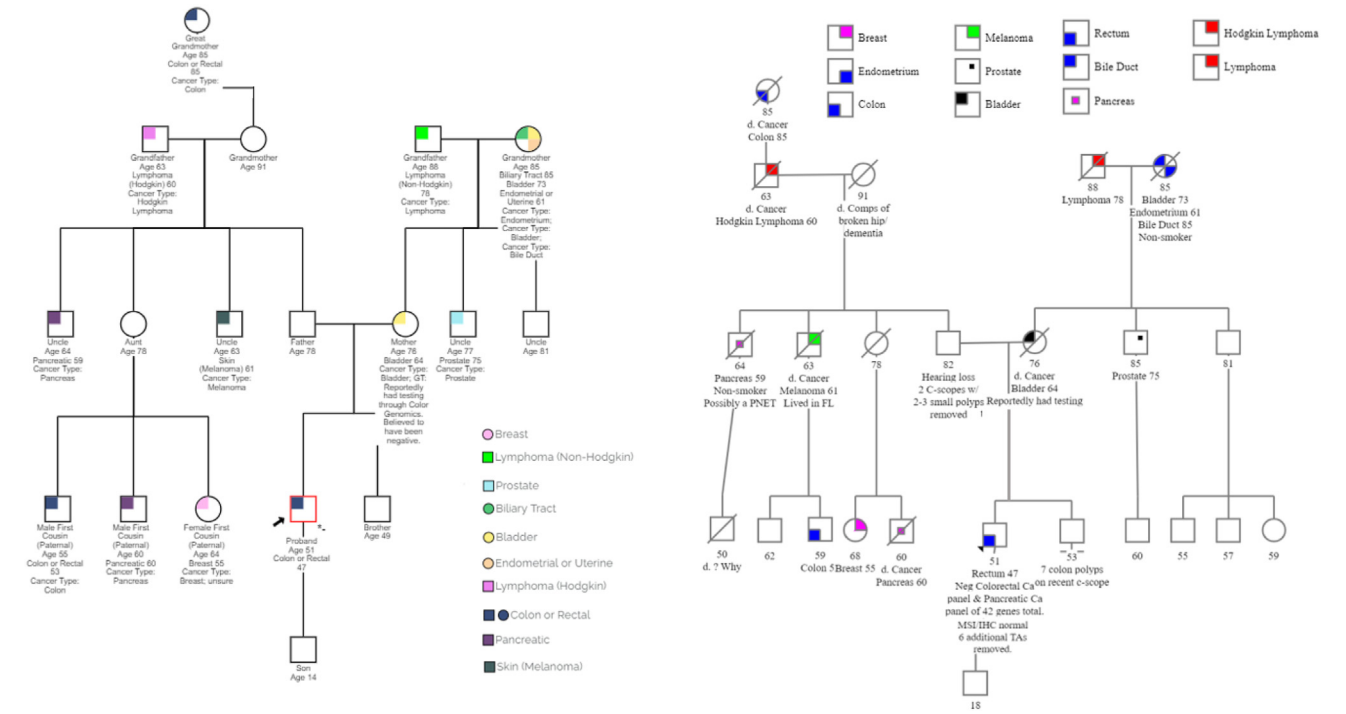

Figure 2 **Pedigree was reproduced almost identically between left panel with CancerIQ and right panel with Progeny**.

Transitioning commercial software solutions requires multilayered organizational change within a health care institution. However, few research groups have evaluated the process of health care data migration and its downstream effects on data integrity.[8] To our knowledge, the only comparable study involved transition of immunization records at the Children's Hospital of Philadelphia in which the authors manually reviewed 50 patient charts to evaluate the accuracy of their

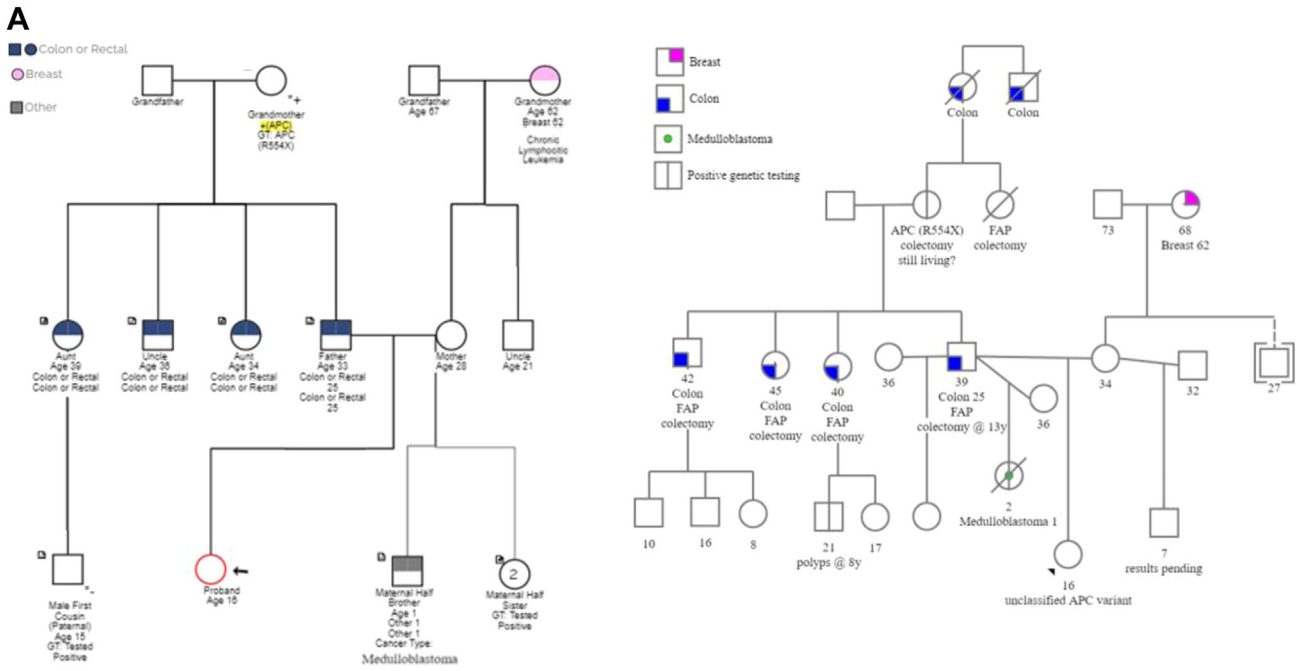

Figure 3 **Pedigrees reproduced with incorrect parental relationships in CancerIQ compared to the one recorded in Progeny.** A. Example of a discordant chart showing incorrect parental relationship in the panel on the left from CancerIQ compared with the panel on the right from Progeny. Other errors include the inclusion of chronic lymphocytic leukemia in the proband's maternal grandmother in CancerIQ. B. Example of a discordant chart showing incorrect parental relationship in the panel on the left from CancerIQ compared with the panel on the right from Progeny.

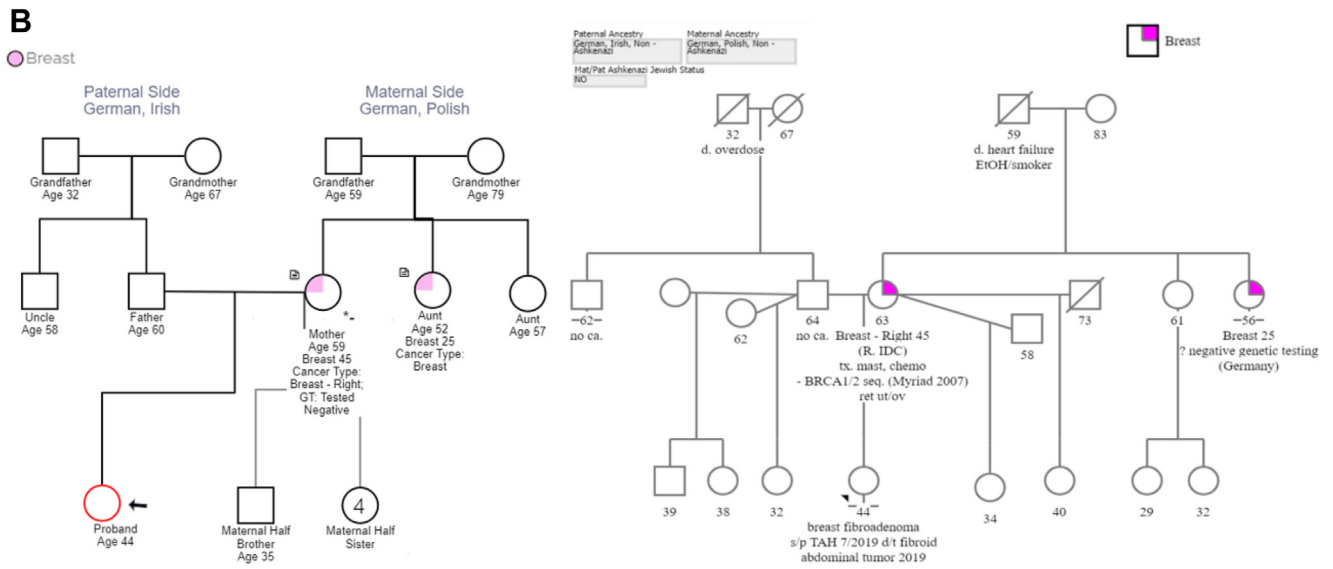

Figure 3 Continued.

data migration.[9] Our quality improvement study shows that there is excellent data agreement (greater than 92%) after data download and migration compared with the 15% error rate found at a large oncology program.[10] These data comprised several modalities, including text and images often portraying complex pedigrees. Successful transfer relied on buy in and efficient team communication between our genetics team and the vendor teams.

Of the recorded errors on data migration, the most common were missing genetic testing results. One potential source of this error is the fragmentation of data within the software. A patient's test result may be recorded in the Notes or Other section of their chart in Progeny, which may not be accurately transferred over to the comments section in CancerIQ. This highlights the importance of establishing conventions for recording patient information and aligning data fields as closely as possible. It is well established that fragmentation of patient data within the EHR has been associated with medical errors across health care settings.[11] Furthermore, even when the data are appropriately entered, idiosyncrasies may exist between 2 software products to the point at which the precise syntax of the entry becomes important. With regard to our experience during this data transfer, one major challenge was free-text fields that included commas within the text. When combined with null values for other fields and the said comma-separated limiters, there were issues at coding the correct values to the correct downstream fields. As an example, a free-text field that says "The patient got tested for *BRCA1*, *BRCA2*, *ATM*, and *CHEK2*" would get naively parsed into 4 columns because of the commas, although it should be only 1 field. To remedy this issue, CancerIQ implemented anchor

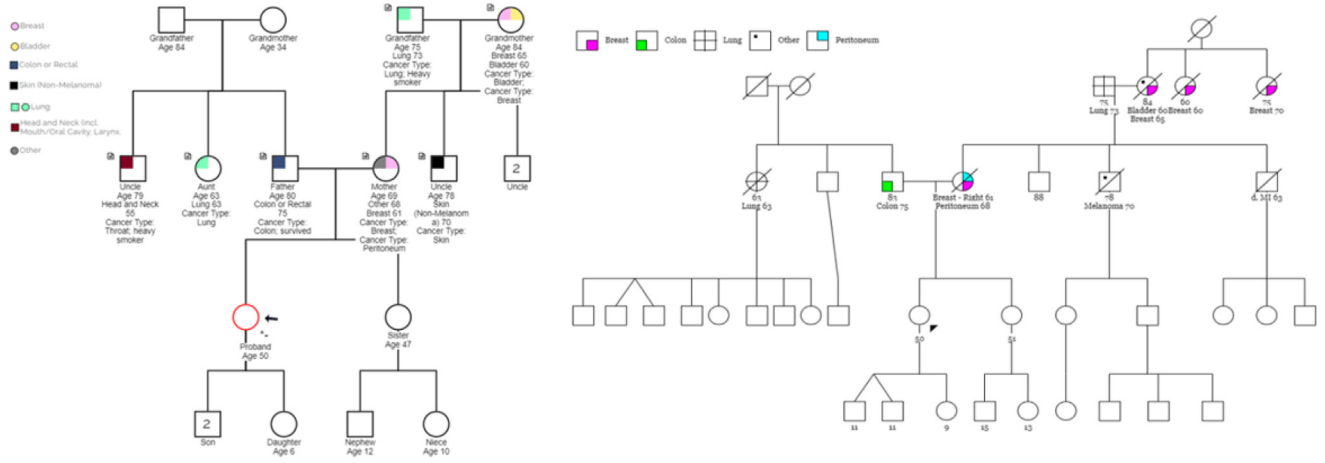

Figure 4 Pedigrees with missing third-degree relative data in right panel from Progeny compared with left panel from CancerIQ. This is an example in which third-degree relative data included on the right would have helped proband meet national guidelines for genetic testing for breast cancer. The pedigree on the left only captured 2 such relatives.

columns to help orient the cleaning script to which columns certain variables should reside. Even with automated data cleanup, the CancerIQ team was presented with several thousand rows of data that were tagged by the script for manual review, before transfer to the final data sheet.

A second common source of data discordance was the incorrect transfer of pedigree information. To our knowledge, no other studies have evaluated the data integrity of pedigree information after such a large data transition. A potential source of this error was a decision made by the genetics clinical team to only transfer first and second-degree relatives of a proband when many charts in Progeny contained limited information on third-degree relatives. This was to facilitate the efficient and timely transfer of data (6 months versus 9-12 months) with our study teams' assumptions that third-degree relative information would not be necessary in determining a patient's eligibility for genetic testing in most cases. However, we recorded one patient chart in which the patient would only meet NCCN criteria for genetic screening for breast and ovarian cancer syndromes if their pedigree included numerous third-degree relatives with a history of breast cancer on the same side of the family. This 1 patient would have been deemed ineligible for genetic testing when, in fact, they were eligible.

Limitations of this study include a short-term assessment of data integrity after transfer without measuring the completeness of data entry within each software. This would have allowed our team to assess whether further in-service training on data entry would be necessary for the new software. Another limitation is having missing data for 2016 and before, given that the study team did not use Progeny until 2017. Most of the pedigrees were obtained manually then and scanned into the Media tab of the EHR without being transferred into Progeny as nondiscrete historical data.

Conclusion

In this quality improvement study, we found that there was excellent data integrity upon transitioning from one commercial pedigree and clinical genetics decision tool to another. This suggests that such data transitions are feasible for institutions looking to switch electronic health software vendors. Future directions of this work are to evaluate the downstream effects of this transition on the workflow of the genetic counseling staff, impact on workload, and availability of discrete genetic results for research studies and clinical trials within our health care institution. We are also planning to participate in efforts to analyze data from multiple institutions using this new software and measure adherence to cancer screening and surveillance.

Data Availability

Please email the corresponding author regarding all data requests.

Acknowledgments

The authors thank the Fred Hutchinson Cancer Center for their strong institutional support in making this transition possible. We also thank the CancerIQ and the Progeny teams for helping ensure compliance with the Health Insurance Portability and Accountability Act of 1996 (HIPAA) during data transfer. An abstract with preliminary results from this project was presented at the 2024 Annual Meeting of the American Society of Clinical Oncology in Chicago, IL.

Ethics Declaration

This project was reviewed by the University of Washington Institutional Review Board and was exempted from needed institutional review board review/approval because the project was deemed a quality improvement project. An exemption letter can be provided upon request.

Funding

No additional external grant or sponsorship was obtained or received for this project.

Author Contributions

Conceptualization: C.L., K.T., M.D.G.; Data Curation: C.L., K.T., H.L., W.L., M.D.G.; Formal Analysis: C.L., K.T., M.D.G.; Funding Acquisition: institutionally funded by Fred Hutch Cancer Center; Investigation: all authors; Methodology: all authors; Project Administration: C.L., K.T., M.D.G.; Resources: H.L., W.L., M.D.G.; Software: H.L., W.L.; Supervision: M.D.G.; Validation: all authors; Visualization: C.L., K.T., M.D.G.; Writing-original draft: all authors; Writing-review and editing: all authors.

Conflict of Interest

The authors acknowledge that Haibo Lu is an employee of CancerIQ and Thomas Nam is an employee of Progeny (now part of Ambry Genetics). All other authors have no conflicts of interest to disclose.

References

1. Lau-Min KS, McKenna D, Asher SB, et al. Impact of integrating genomic data into the electronic health record on genetics care delivery. *Genet Med*. 2022;24(11):2338-2350. http://doi.org/10.1016/j.gim.2022.08.009
2. Payne TH, Zhao LP, Le C, et al. Electronic health records contain dispersed risk factor information that could be used to prevent breast

and ovarian cancer. *J Am Med Inform Assoc*. 2020;27(9):1443-1449. http://doi.org/10.1093/jamia/ocaa152

3. Ritchie JB, Allen CG, Morrison H, et al. Utilization of health information technology among cancer genetic counselors. *Mol Genet Genomic Med*. 2020;8(8):e1315. http://doi.org/10.1002/mgg3.1315
4. Bednar EM, Nitecki R, Krause KJ, Rauh-Hain JA. Interventions to improve delivery of cancer genetics services in the United States: a scoping review. *Genet Med*. 2022;24(6):1176-1186. http://doi.org/10.1016/j.gim.2022.03.002
5. Huang C, Koppel R, McGreevey JD III, Craven CK, Schreiber R. Transitions from one electronic health record to another: challenges, pitfalls, and recommendations. *Appl Clin Inform*. 2020;11(5):742-754. http://doi.org/10.1055/s-0040-1718535
6. Lipton LR, Johnson V, Cummings C, et al. Refining the Amsterdam Criteria and Bethesda Guidelines: testing algorithms for the prediction of mismatch repair mutation status in the familial cancer clinic [published correction appears in *J Clin Oncol*. 2005;23(15):3652]. *J Clin Oncol*. 2004;22(24):4934-4943. http://doi.org/10.1200/JCO.2004.11.084
7. Reynolds TL, Clay B, Rudkin SE, et al. Migrating from one comprehensive commercial EHR to another: perceptions of front-line clinicians and staff. *AMIA Annu Symp Proc*. 2019;2019:765-773.
8. Miake-Lye IM, Cogan AM, Mak S, et al. Transitioning from one electronic health record to another: a systematic review. *J Gen Intern Med*. 2023;38(suppl 4):956-964. http://doi.org/10.1007/s11606-023-08276-3
9. Michel J, Hsiao A, Fenick A. Using a scripted data entry process to transfer legacy immunization data while transitioning between electronic medical record systems. *Appl Clin Inform*. 2014;5(1):284-298. http://doi.org/10.4338/ACI-2013-11-RA-0096
10. Khela H, Khalil J, Daxon N, et al. Real world challenges in maintaining data integrity in electronic health records in a cancer program. *Tech Innov Patient Support Radiat Oncol*. 2024;29:100233. http://doi.org/10.1016/j.tipsro.2023.100233
11. Graber ML, Siegal D, Riah H, Johnston D, Kenyon K. Electronic health record-related events in medical malpractice claims. *J Patient Saf*. 2019;15(2):77-85. http://doi.org/10.1097/PTS.0000000000000240