Patient Journey | August 2020

# OpenPHR: Empowering patients to understand the journey through breast cancer

Authors: Caleb Li, Ronald Marconet, Gregory Falco, Fozoh Saliki

## Abstract

Traditional medicine has relied on scientific studies from 50-100 person clinical trials that can take decades to see results. With the millions of patient posts online that are seeking and providing answers, we believe we can curate the millions of posts into patient journeys and structured patient-reported outcomes ("PRO") database to accelerate personalized medicine to provide targeted treatments to patients. Our mission is to empower patients to be well-informed of potential treatment options by analyzing outcomes from patients around the world.

# Introduction

Currently, the majority of patients today have been researching their diagnosis online and creating community forums to share their journeys. These provide us the opportunity to create powerful insights for the patients and community by aggregating and curating the millions of posts from social media. We have developed algorithms to aggregate over 50,000 patients outcomes just from breast cancer while using data science to search and sort responses, which will help to match the right drug to the right patient and predict other possible drugs. This paper shows how we developed patient journeys for breast cancer by using unstructured social media data to analyze over 50,000 posts from 1,200 unique individuals. Our objective is to leverage our methodology and techniques developed for breast cancer to become a platform to develop patient journeys for other diseases.

## Mission of OpenPHR

Our mission is to empower patients and loved ones with insightful and actionable information on diseases by leveraging big data and machine learning to aggregate, curate, and generate insights. Our objective is to collaborate in an open source environment to eventually become the platform for patient journeys across many disease verticals.

## Problems in the healthcare industry and opportunity for tech

US Healthcare has traditionally been provider-focused and has been an episodic experience for patients that average less than three visits per year for 20 minutes per appointment. Therefore, patients resort to online searches and online communities ("Unstructured Data") to seek information on their disease and to ask questions and share their experiences. However, the information online has generally been unstructured, and patients may misinterpret the information due to their lack of medical background.

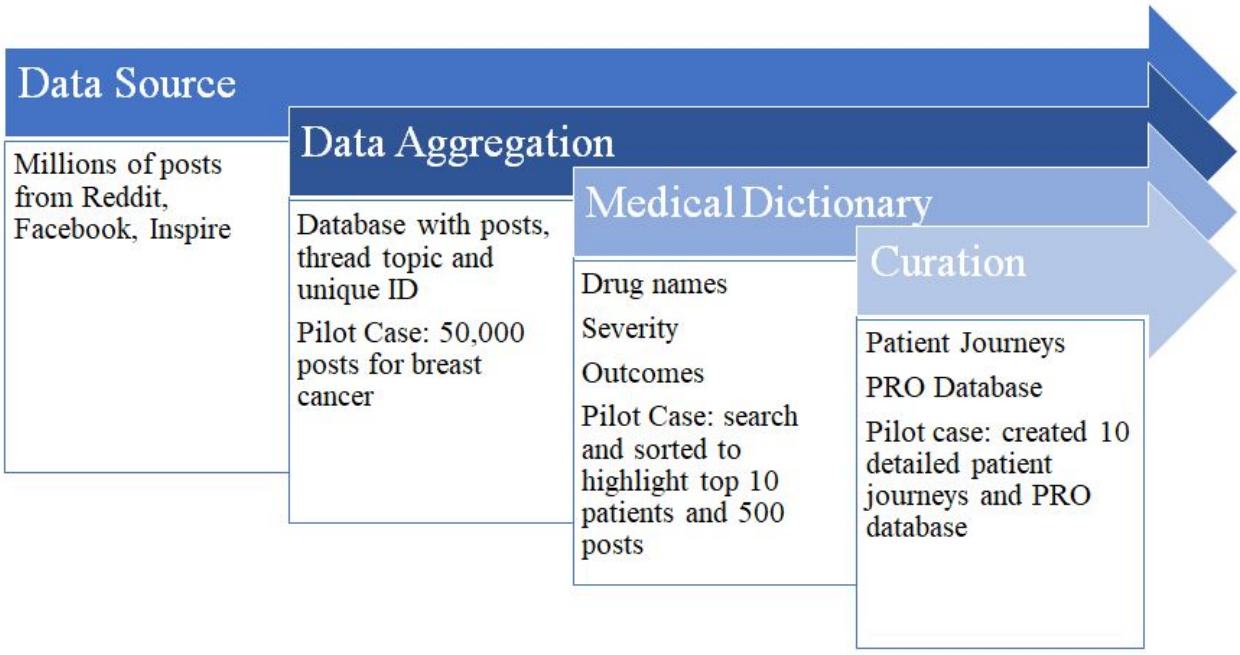
With technologies to aggregate and curate unstructured data, we have taken advantage of the opportunity from the unstructured data to build algorithms that can generate patient journeys. We have worked with professionals with a medical background to review and curate the patient journeys to extract the drugs and outcomes from each line of treatment. We provide patients with a clear and comprehensive view of the patient journey through diseases like breast cancer, giving them an understanding of what others went through. Patients will be able to understand the vast treatment options and outcomes. This information could empower patients to seek second opinions or provide guidance on how to manage symptoms.

In the long-term, we believe this "big data" has a wealth of information at a scale that could generate evidence that impacts regulation and real-world evidence. In traditional clinical trials, they may enroll only tens to hundreds of patients to closely monitor for novel treatments. By leveraging big data, our database of breast cancer patients already has over 10,000 unique users

and 50,000 posts, more than 100x a clinical trial. By developing technology to curate the unstructured "big data," we can significantly impact US Healthcare.

## Methodology

Social media has provided a platform for patients to provide an abundance of patient-reported outcomes. However, the quality, varieties, and velocity of data become challenging to manage when it becomes abundant. Consequently, the varieties of posts and websites could result in patients making wrong decisions. Hence, we propose a big data-driven approach where data scientists and medical experts can analyze patients' insights from unstructured data.




Our system architecture and principles include three approaches: 1) Big data warehouse, 2) Proprietary medical dictionary, and 3) Curated patient-reported outcomes. The approaches are detailed in the subsequent section below:

Our initial pilot case focused on breast cancer, and we aggregated 50,000 posts from 1,200 unique patients. Despite using particular keywords to collect data across social media, the data

was not structured enough to provide quality patients' journeys. Hence, our big data tools helped us learn, analyze, and model the raw data. Consequently, we present herein a curated response of ten patients to show the quality of responses we seek, which upon training gives us a qualitative patient's journey. Having trained the data, we integrated the codes to obtain a medical dictionary, as shown in the subsequent section.

## 1) Data aggregation to create a big data warehouse

Patients often have access to social media and online forums to share their journey, most of which are unmoderated. Therefore, patients could get confused from these posts, which potentially lead to troublesome outcomes.

With big data algorithms and data science, there is an opportunity to curate unstructured data to become meaningful insights. We developed a python-based web scraper to collect and aggregate the data. Hence, we aggregated unstructured forum posts in a disease category across social media sites, including Facebook, Reddit, and Inspire, and was able to build a database of 50,000 social media posts about breast cancer.

## 2) Proprietary Medical Dictionary

With 50,000 posts, we needed to develop a proprietary medical dictionary to search and sort through posts and prioritize the posts that should be part of our structured database of patient insights. We created a medical dictionary that started with keywords like drug names, “Herceptin”, severity of disease, “Stage 4”, and outcomes, “stable for 2 years”, to help search and sort posts to prioritize.

With over 50,000 posts and 1,200 unique users, we leveraged the medical dictionary to prioritize the users with in-depth posts with many drug names, several lines of treatments, and many patient reported outcomes. In this pilot case, we decided to focus on the top 10 patients to build the patient journey.

## 3) Curated Patient-Reported Outcomes

With the database of social media posts and proprietary medical dictionary, our next step is to create high quality patient journeys and a structured database of patient-reported outcomes. Our goal here was to mimic how a physician may see their patients at the clinics.

For the initial pilot, we extracted our top 10 user posts with the longitudinal history and worked with clinicians to develop a chronological patient journey. After curating 10 patient journeys, we created a dashboard of patient reported outcomes and extracted the data from the patient journeys. Below will provide our summary results and current progress. We hope this initial phase of work can demonstrate the value of OpenPHR for patients and their loved ones.

## Summary Results of Initial Phase

Below is a summary of the results of our initial work. This is a snapshot of a potential database that can be developed from the unstructured social media data.

| Code Name | Joy | Mandy | Joan |
|---|---|---|---|
| Initial Diagnosis | Stage 3 | Stage 4 | Stage 4 |
| Operations/Surgeries | California | Ohio | New York / USA |
| Treatment 1 | Chemotherapy and radiotherapy that are not stated | Tamoxifen + Herceptin | Letrozole and 10 sessions of Radiotherapy on the spine |
| Objective Response | Not stated | Stable disease | Stable disease |
| Duration of Response | 6 months | Tamoxifen - 2 years | 3 years |
| Number of treatment-related adverse events | *2 (Osteopenia and cardiomyopathy)* | *4 tamoxifen (Nausea- Increased uterus thickness - Pulmonary embolism - Decreased Hemoglobin level) 1 Herceptin ( Decreased ejection fraction)* | *NA* |
| Treatment 2 | Letrozole | Radiotherapy + Tamoxifen + Herceptin | 6 cycles of Ibrance with Arimidex |
| Objective Response | Progressive disease | Partial response to Herceptin | Stable disease |
| Duration of Response | 2.5 years | 5 years for Herceptin | 5.5 months |
| Time to Progression | 2.5 years | 5 years | NA |
| Number of treatment-related adverse events | 1 (Hot flushes) | 4 tamoxifen (Nausea- Increased uterus thickness - - Pulmonary embolism - Decreased Hemoglobin level) 1 Herceptin ( Decreased ejection fraction) | 1 Arimidex (Joint pains) 2 Ibrance (Elevated liver enzymes - hair thinning) |

The purpose of the output above is to provide guidance on the lines of treatments, disease progression, and adverse events. Patients and loved ones including physicians can leverage this

type of database to also learn standard of care across the globe and what outcomes resulted. A full table is in the appendix.

## Conclusion

Unstructured data on social media and forums may contain in-depth insights on patient reported outcomes. Hence, we leverage data science and big data tools to curate unstructured data into patient journeys and a PRO database. Our methodology was the following: aggregate data into a data warehouse, develop proprietary medical dictionary, and curate unstructured data into patient journeys and PRO database. We propose that our solution will give patients and loved ones hope while improving other personalized medicine cases.

## Next Steps

Our pilot case is for breast cancer patients, and our next step is to scale the program to other diseases to help other patients. Our use of Python and R produces a patient's journey in codes, giving it the flexibility to be integrated into other personalized medicine applications for better treatment and diagnosis.

## References


1. Chase JG, Preiser J-C, Dickson JL, Pironet A, Chiew YS,Pretty CG, Shaw GM, Benyo B, Moeller K, Safaei S et al.: Next-generation, personalised, model-based critical care medicine:a state-of-the art review of in silico virtual patient models,methods, and cohorts, and how to validation them. BiomedEng Online 2018, 17:24.
2. Davide Cirillo Alfonso Valencia, Big data analytics for personalized medicine, Current Opinion in Biotechnology 2019, 58:161–167, pg 164.
3. Teli N. Big Data: A Catalyst for Personalized Medicine. 2014 [Online]. Available: http://healthcare-executive-insight.advanceweb.com/Features/Articles/Big-Data-A-Catalyst-for-Personalized-Medicine.aspx.

# Appendix

### Exhibit 1: Sarah's Full Patient Journey

In **June 2014**, when Sarah had just turned **45**, she felt a lump in her right breast. She thought it was just a benign one, similar to that of her mother and aunt. However, when she consulted her doctor, she was advised to go for a **mammogram**. However, both her **mammogram and ultrasounds** came out to be normal. Sarah still felt that there was something wrong with her. However, it eventually proved to be quite a difficult task to convince her doctor to conduct a **biopsy**.

According to Sarah, "I persisted and was eventually diagnosed with a 2.8cm IDC with micropapillary features, lymphovascular and perineural invasion."

Finally, Sarah was diagnosed with bony metastasis and visceral metastasis to lung and liver, and she was already at **Stage 4**.

At first, Sarah had to undergo a **bilateral mastectomy**, which was later accompanied with all sorts of complications including an MRSA infection. Sarah had 4 cycles of ***Taxotere*** and ***Cytoxan***.

Sarah described her reaction to chemotherapy as, "The previous chemotherapy (***Taxotere*** and ***Cytoxan***) nearly killed me (for real)."

Sarah suffered from skin rashes and high fever, after her chemotherapy. Since her liver enzymes were at an elevated level, her chemotherapy got delayed. Moreover, her **alk phos** was **801** and **ALT/AST** was **5-6 times** normal.

Unfortunately, cancer did not respond to chemotherapy and it failed to shrink or regress. Then Sarah was started on **hormone therapy** with ***Letrozole***, but it resulted in her cholesterol levels increasing dramatically (e.g., **LDL 220** and then **290**).

“I had miserable hot flashes, night sweats and all the signs of a second ‘menopause’”.

In **July 2016**, the radiologist identified enlarging intrathoracic lymph nodes. Technically, the lymph nodes were "borderline" enlarged. Her doctor recommended **bronchoscopy** with biopsy of a lymph node, and the pathology report came back as **metastatic breast cancer**, still 100% ER+, 20% PR+, **Her2 negative**. The treatment with ***Letrozole*** was continued for almost **2 years**.

As a result, ***Letrozole*** was discontinued and she was put on ***Faslodex*** and ***Ibrance***. Since Sarah was already suffering from osteoporosis, therefore, the doctor started ***Xgeva***. Sarah was given ***Xgeva***, as an injection under the skin, after every **four weeks** but it made her feel miserable for several days. The most side effects that Sarah experienced after having ***Xgeva*** shots **were jaw pain, shortness of breath and unusual muscle or bone pain.**

Within **3 months**, only cancer evident on CT scans was an 8 mm lesion in her thoracic spine. Sarah was able to stay on the highest dose of ***Ibrance*** for about **11 months**. Scans remained stable and her liver function tests also returned to normal. Even **transaminases** normalized after **3 months**. Her **alk phos** levels also dropped to 128 after two months and they continued to show a considerable decrease over the next **two months**. ***Faslodex***, ***Ibrance*** and ***Xgeva*** were continued for more than a year.

The doctor told Sarah that the hot flashes and "menopausal" symptoms were due to hormone therapy (***Letrozole*** followed by ***Faslodex***). Although Sarah had shown great progress with ***Ibrance***, she experienced some side effects as well. These side effects included some **hair thinning, fever and mild fatigue**.

Moreover, it was observed that **both her platelet counts and white blood cell counts (especially neutrophils) had lowered**. ***Ibrance*** also elevated liver enzymes. Sarah managed to stay on ***Ibrance*** for **11 months** and avoided bone marrow toxicity. She had a reduced dose of ***Ibrance*** 100 mg.

Her isoenzymes were elevated for liver, bone, and intestine in **2014**, but after treatment with ***Letrozole, Ibrance, Faslodex and Xgeva***, there was a tremendous decrease in her isoenzymes

levels in **2017**. Sarah is really happy now to have a normal alk phos and she is even hopeful that the normal value means her bone disease is stable and dormant.

Apart from the aggressive cancer treatment, Sarah was also prescribed to have a healthy and balanced diet, with **fruits and vegetables and yogurt**. The diet itself made Sarah feel a whole lot better.

The entire experience of going through the treatment was really difficult not only for her but also for her entire family. Sometimes, she would be really depressed and on some days the pain would be unbearable. But she had her family by her side to support her emotionally.

**Drugs: ibrance, faslodex, letrozole, xgeva, taxotere**

**Stage of cancer:** stage 4

**Side effects:** Jaw pain, shortness of breath and unusual muscle or bone pain, hair thinning, **fever** and mild fatigue

# Exhibit 2: PRO Table

| Code Name | Joy | Mandy | Joan |
|---|---|---|---|
| Initial Diagnosis | Stage 3 | Stage 4 | Stage 4 |
| Operations/Surgeries | California | Ohio | New York / USA |
| Treatment 1 | Chemotherapy and radiotherapy that are not stated | Tamoxifen + Herceptin | Letrozole and 10 sessions of Radiotherapy on the spine |
| Objective Response | Not stated | Stable disease | Stable disease |
| Duration of Response | 6 months | Tamoxifen - 2 years | 3 years |
| Number of treatment-related adverse events | *2 (Osteopenia and cardiomyopathy)* | *4 tamoxifen (Nausea- Increased uterus thickness - Pulmonary embolism - Decreased Hemoglobin level) 1 Herceptin ( Decreased ejection fraction)* | *NA* |
| Treatment 2 | Letrozole | Radiotherapy + Tamoxifen + Herceptin | 6 cycles of Ibrance with Arimidex |
| Objective Response | Progressive disease | Partial response to Herceptin | Stable disease |
| Duration of Response | 2.5 years | 5 years for Herceptin | 5.5 months |
| Time to Progression | 2.5 years | 5 years | NA |
| Number of treatment-related adverse events | *1 (Hot flushes)* | *4 tamoxifen (Nausea- Increased uterus thickness - - Pulmonary embolism - Decreased Hemoglobin level) 1 Herceptin ( Decreased ejection fraction)* | *1 Arimidex (Joint pains) 2 Ibrance (Elevated liver enzymes - hair thinning)* |
| Treatment 3 | Faslodex | Arimidex + Herceptin | Ibrance and Faslodex |
| Objective Response | Complete response | Stable disease | Stable disease |
| Duration of Response | Stated that was NED shortly after she started this treatment and has been taking it for 4 years now | 2 years | one month |
| Number of treatment-related adverse events | 1 (Hot flushes) | 1 (Hot flushes) | 2 (Elevated liver enzymes - Hair thinning) |
| Treatment 4 | Faslodex and Ibrance | Ibrance + Radiotherapy + Herceptin | Kisqali and Faslodex |
| Objective Response | Complete response | Partial response | Stable disease |
| Duration of Response | Stated that was NED shortly after she started this treatment and has been on this treatment for 3.5 years now | 3 years | 1.5 months |
| Number of treatment-related adverse events | 3 (Hair thinning - Hot flushes - Decreased ANC levels)) | 2 (Nosebleeds - Anemia) | 2 (Hair thinning - Nausea) |