Content

02  Advance Discovery
03  Foreword
06  Key Findings
17  Authors
Advance Discovery

Learn from the experiences and outcomes of every cancer patient.

Improved clinical outcomes and the future of patient care depend on learning from the experiences of every cancer patient. CancerLinQ Discovery® sits at the forefront of real-world cancer research, providing access to harmonized, codified, and curated sets of aggregated, de-identified real-world cancer patient data representing the diverse pool of 1.5 million patients in the CancerLinQ® network. This powerful database can then be used to uncover unseen patterns in patient characteristics and unlock actionable insights—shaping the future of cancer care.

Comparison

Compare the effectiveness and value of alternative treatment options

Study the use of cancer treatments in populations typically excluded from clinical trials to generate new knowledge to improve patient care

Deliver insights to inform and continuously improve practice guidelines and quality measures

In June 2020, CancerLinQ Discovery will be made more broadly available to the academic, nonprofit and government research communities through the newly launched CancerLinQ Discovery Research Platform. Visit Discovery.CancerLinQ.org to learn more.
Transformational improvement in the outcomes of patients with cancer will only be fully realized by the creation of an oncology rapid learning health system, where routine patient care data seamlessly informs scientific discovery, and, reciprocally, research informs practice. This vision inspired the American Society of Clinical Oncology (ASCO) to create CancerLinQ®, a real-world oncology data platform whose purpose is to improve cancer care quality and accelerate discovery.

CancerLinQ enables learning to occur from every patient with cancer by securely compiling, harmonizing, and analyzing vast amounts of information on patient characteristics (e.g., molecular profiles, comorbidities), treatments, clinical outcomes, and long-term side effects. By using data from more than 1.5 million patients in near real time, CancerLinQ can identify trends and associations between myriad variables, thereby enabling physicians and researchers to generate new hypotheses and rapidly apply those conclusions to improve care in real-world settings.
In the past decade, remarkable advances in the understanding of the molecular basis of cancer have led to the development of multiple new targeted agents and immunotherapy drugs which are revolutionizing cancer treatment. However, these agents are only able to be tested in the 3% of adults with cancer who enroll in clinical trials, and the populations studied are neither clinically nor ethnically representative of patients typically encountered in most practice settings in the U.S., especially community oncology clinics. And while clinical trials are the gold standard for testing the efficacy of new therapies, they are expensive, slow to accrue, may fail to capture long-term outcomes and the patient experience of illness, and in many cases lack external validity especially when rapid changes in practice make the chosen control group therapy obsolete. Therefore, a robust source of actionable real-world data is required to complement the results from traditional trials and fully realize the benefits of these advances.

To that purpose, in 2017, CancerLinQ launched CancerLinQ Discovery®, now one of the largest real-world oncology research data sets in the world, incorporating harmonized, curated, and de-identified data from the 1.5M+ cancer patients whose health care providers are a part of the CancerLinQ network. Today, CancerLinQ Discovery represents patients from over 60 oncology practices in the United States, with data integrated from over 10 different electronic health records (EHRs). Through CancerLinQ Discovery, CancerLinQ is committed to making it possible for researchers to learn from the experiences of patients throughout the CancerLinQ network, dramatically shifting oncology research to include data from a diverse network of rural, community, academic, and health system care settings. This inclusiveness is critical to the scalability of a rapid learning health system and enables research results to be generalizable to large segments of the patient population, particularly because 76% of oncologists practice at small clinics (1-5 oncologists per clinic) in the community.

In June 2020, CancerLinQ will be launching a new CancerLinQ Discovery Research Platform, where researchers from healthcare providers, medical specialty societies, grant organizations and foundations, academic institutions, federal, state and local governments and ASCO members can view and request access to CancerLinQ Discovery data sets, and approved researchers can access their data-sets and work via secure Amazon Web Services (AWS) workspaces. The platform will go live with de-identified data products across five disease areas, including: breast cancer, chronic lymphocytic leukemia/small lymphocytic lymphoma, lung cancer, ovarian cancer, and prostate cancer. Visit discovery.cancerlinq.org to learn more.

CancerLinQ Discovery has been serving as a source of real-world oncology insights since its inception, and at this 2020 ASCO Virtual Scientific Meeting, a number of potentially valuable research findings derived from aggregated CancerLinQ Discovery data will be unveiled.
Key Findings
Machine learning imputation of Eastern Cooperative Oncology Group performance status (ECOG PS) scores from data in CancerLinQ Discovery

Agrawal et. al. | Abstract #e19318

**BACKGROUND:**

ECOG PS is a prognostic indicator of outcomes, and scores of 0-1 (good ECOG PS) are often required for clinical trial enrollment. Patients treated in non-trial settings often lack ECOG PS scores limiting the ability of Real World Data from these patients to be used in external control arms (ECAs) or to provide optimal specificity for clinical effectiveness research. Machine Learning can be used to impute ECOG PS scores from other clinical data at various points during treatment.

**RESULTS:**

AUC-ROC values of up to 0.81 could be obtained for imputing a patient’s final ECOG PS, with lower AUC values when imputing ECOG PS at initial and metastatic diagnosis using large numbers (i.e. thousands) of features. We developed more interpretable models with 110 or 40 features with reduced but still satisfactory AUC, with accuracy of predicting good ECOG PS scores of around 80%.

Key features were obtained from lab tests, physical exams, comorbidities, medications, age and metastatic status. The table shows the results of several of these models. Where the models misclassify ECOG PS, the error was rarely greater than 1 grade.

<table>
<thead>
<tr>
<th>TIME POINT</th>
<th>METHOD</th>
<th>FEATURE COUNT</th>
<th>USES PRIOR ECOG PS</th>
<th>USES AJCC STAGE</th>
<th>AUC (POOR ECOG PS)</th>
<th>PRECISION (POOR/GOOD)</th>
<th>RECALL (POOR/GOOD)</th>
<th>F1 SCORE (POOR/GOOD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final</td>
<td>LR</td>
<td>110</td>
<td>N</td>
<td>Y</td>
<td>0.70</td>
<td>0.59/0.70</td>
<td>0.47/0.79</td>
<td>0.53/0.74</td>
</tr>
<tr>
<td>Final</td>
<td>LR</td>
<td>220</td>
<td>Y</td>
<td>Y</td>
<td>0.73</td>
<td>0.65/0.72</td>
<td>0.51/0.82</td>
<td>0.57/0.77</td>
</tr>
<tr>
<td>Final</td>
<td>XGB</td>
<td>110</td>
<td>N</td>
<td>N</td>
<td>0.76</td>
<td>0.66/0.73</td>
<td>0.53/0.82</td>
<td>0.59/0.77</td>
</tr>
<tr>
<td>Final</td>
<td>XGB</td>
<td>110</td>
<td>N</td>
<td>Y</td>
<td>0.77</td>
<td>0.66/0.73</td>
<td>0.53/0.83</td>
<td>0.59/0.77</td>
</tr>
<tr>
<td>Final</td>
<td>XGB</td>
<td>40</td>
<td>N</td>
<td>Y</td>
<td>0.77</td>
<td>0.66/0.72</td>
<td>0.53/0.83</td>
<td>0.59/0.77</td>
</tr>
<tr>
<td>Final</td>
<td>XGB</td>
<td>22000</td>
<td>Y</td>
<td>Y</td>
<td>0.81</td>
<td>0.71/0.76</td>
<td>0.60/0.84</td>
<td>0.65/0.80</td>
</tr>
<tr>
<td>Initial</td>
<td>XGB</td>
<td>200</td>
<td>N</td>
<td>N</td>
<td>0.74</td>
<td>0.50/0.81</td>
<td>0.43/0.85</td>
<td>0.46/0.83</td>
</tr>
<tr>
<td>Initial</td>
<td>XGB</td>
<td>15000</td>
<td>N</td>
<td>Y</td>
<td>0.76</td>
<td>0.51/0.84</td>
<td>0.54/0.82</td>
<td>0.52/0.83</td>
</tr>
<tr>
<td>Metastatic</td>
<td>XGB</td>
<td>16000</td>
<td>N</td>
<td>Y</td>
<td>0.76</td>
<td>0.55/0.81</td>
<td>0.56/0.80</td>
<td>0.56/0.81</td>
</tr>
</tbody>
</table>

**CONCLUSION:**

ECOG PS is subjective, suggesting that ML based cohort assignment will be sufficiently accurate to support their use in research. Further work will be required to assess if the ML predicted cohorts have different outcomes.

**LINK TO STUDY**

Agrawal et. al. | Abstract #e19318

We developed more interpretable models with 110 or 40 features with reduced but still satisfactory AUC, with accuracy of predicting good ECOG PS scores of around 80%.
Identification of transgender people with cancer in electronic health records (EHR): Recommendations based on CancerLinQ observations

Alpert et. al. | Abstract #e19046

**BACKGROUND:**

Data regarding people who are gender minorities are not well-captured in oncology practices or large cohorts. Given this, collection of cancer prevalence and outcomes data, which are necessary to understand disparities in this population, are significantly hampered. Real-world data may be the most readily available source to explore outcomes in transgender populations. A database of EHR data on people with cancer, CancerLinQ, is housed by the American Society of Clinical Oncology and collected from nation-wide practices and multiple EHRs.

**RESULTS:**

Of ~1.3 million records in CancerLinQ at time of case selection, 557 matched inclusion criteria and 242 were abstracted.

<table>
<thead>
<tr>
<th>Dx indication of transgender</th>
<th>NUMBER ABSTRACTED</th>
<th>EVIDENCE OF TRANSGENDER/ NON-BINARY GENDER IDENTITY, N (%)</th>
<th>DX CODING ERROR, N (%)</th>
<th>GENDER CODING ERROR, N (%)</th>
<th>UNKNOWN N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male with female cancer Dx</td>
<td>100</td>
<td>2 (2%)</td>
<td>88 (88%)</td>
<td>8 (8%)</td>
<td>2 (2%)</td>
</tr>
<tr>
<td>Female with male cancer Dx</td>
<td>100</td>
<td>3 (3%)</td>
<td>81 (81%)</td>
<td>14 (14%)</td>
<td>2 (2%)</td>
</tr>
</tbody>
</table>

76% of patients with ICD9/10 gender related diagnosis (Dx) codes had evidence confirming transgender identity. By contrast, only 2% and 3% of the people identified by criteria 2 and 3 had evidence of transgender identity, respectively.

**CONCLUSION:**

Given the need for data regarding transgender people with cancer and the deficiencies of current data resources, a national concerted effort is needed to broaden terminology in EHRs to include whether people are transgender or not as routine and required data elements, provided by patients at their discretion.
Artificial intelligence model to predict slow progression for advanced non-small cell lung cancer (aNSCLC) patients receiving second-line therapies

Charest et. al. | Abstract #e21596

BACKGROUND:

There are ongoing efforts to understand and predict exceptional response to existing cancer therapies, but few clinical characteristics of these patients are known. We trained a machine learning model using the Concerto HealthAI database of oncology EMR data that includes clinical data from CancerLinQ Discovery to predict slow progression, a proxy for exceptional response, in aNSCLC in the second line setting.

RESULTS:

The final model was able to predict slow progression with an AUCROC of 0.75 (F-score 0.48, precision 0.39, recall 0.6). The performance compares favorably to that of a logistic regression model (0.66 AUCROC). Top features that indicated slow progression included a low number of prior progression events or regimens, absence of metastatic disease, lower stage/t-stage/ECOG, absence of COPD, previous treatment with an EGFR inhibitor, normal Alk-Phos/WBC (versus elevated), absence of tachycardia, and a normal BMI (versus low).

Patients met selection criteria of the study.

2205

and 429 were set aside for model validation.

CONCLUSION:

Machine learning and real world-data provided promising results in predicting slow progression in aNSCLC and may be useful in discovering novel drivers of favorable response.
Low rates of *BRCA1* and *BRCA2* testing for patients with ovarian cancer in ASCO’s CancerLinQ, a real-world database

Dewdney et al. | Abstract #6041

**BACKGROUND:**

Ovarian cancer is the deadliest gynecological cancer and has limited screening options for early stage diagnosis. Genetic mutations in genes such as *BRCA1* and *BRCA2* increase the risk of ovarian cancer, and if identified, patients can undergo risk-reducing surgery. It is recommended and well accepted to test any new ovarian cancer patient for genetic mutations, particularly *BRCA1* and *BRCA2*. If a *BRCA1/2* mutation is found in a patient (somatic or germ line), this information can be used to guide therapy. We sought to analyze the characteristics of genetic testing in a real-world database, ASCO’s CancerLinQ.

**RESULTS:**

2,654

Of the 2,654 patients meeting inclusion criteria

600

600 had been tested for a *BRCA1/2* mutation (22.6%)

63%

Of those tested, 63% were stage III/IV

14%

14% stage I/II

21.8%

and 21.8% an unknown stage.

The majority of the histologies were serous (76%), followed by undifferentiated (21.2%). The majority of patients tested were white (69.9%), with 18.8% unknown, and 9.9% black. The rate of a positive *BRCA1* or *BRCA2* mutation in this population was 17.2%. Of the patients with a *BRCA1/2* mutation, the majority had serous histology (87%), followed by 18.5% undifferentiated, and 3.9% transitional cell. The majority of the patients found to have a *BRCA1/2* mutation were age >50 (57.3%).

**CONCLUSION:**

Since 2008 evidence-based guidelines have recommended that all ovarian cancer patients be tested for *BRCA1* and *BRCA2* mutations, but in this real-world database only 22.6% have a recorded test. Of those tested, we found a *BRCA1* or *BRCA2* mutation rate of 17.2%. Our data is limited by what is recorded in the database and may not represent the true number of patients tested because of data missing from the EHR; however, these percentages appear similar to previous studies. Not only is testing important for cancer prevention for family members of patients, it now impacts the type of treatments for which these patients are eligible. Since genetic testing remains low at only 22.6% in this population, significant opportunities exist to impact cancer prevention and treatment.
Predicting cardiac adverse events in patients receiving immune checkpoint inhibitors: A machine learning approach

Dreyfus et. al. | Abstract #15075

BACKGROUND:

Many oncology treatments have been associated with cardiovascular (CV) adverse events. Cases of CV events, including myocarditis have been reported for PD-1 and PD-L1 therapies. We created a machine learning model to predict potential CV events in PD-(L)1 patients using the CancerLinQ database.

RESULTS:

The model predicted serious cardiac events within 100 days of index with an AUC-ROC of 0.75 in all patients and 0.79 in PD-(L)1 patients. The top predictors of cardiac risk in PD-(L)1 patients included a history of heart disease, weight loss, the % lymphocyte count, and median LDH. The % lymphocyte count and weight loss were noticeably more predictive in PD-(L)1 patients than in non-PD-(L)1 patients. However, in general SHAP summary plots of all and PD-(L)1 patients were nearly identical, suggesting that both cohorts’ cardiac risk is determined in a similar way. PD-(L)1 and autoimmune disease associated factors did not appear in the top 40 most predictive risk factors.

CONCLUSION:

Using traditional cardiac risk factors, our model was able to predict potential cardiac events in PD-(L)1 patients. Our model found that high lymphocyte count may be protective while weight loss and a history of cardiac disease (e.g. heart failure) could indicate a poor prognosis.

A total of 27,172 advanced cancer patients were included in our study.

The model was trained on 21,758 patients and 5,414 patients were set aside for testing.

27,172
A total of 27,172 advanced cancer patients were included in our study.

21,758
The model was trained on 21,758 patients

5,414
and 5,414 patients were set aside for testing.

4,966 received PD-(L)1 therapy.

4,966
4,966 received PD-(L)1 therapy.

LINK TO STUDY
Trends in immunotherapy use in patients with advanced non-small cell lung cancer (aNSCLC) patients: Analysis of real-world data

Kushi et al. | Abstract #e19311

**BACKGROUND:**
Leveraging data from a collaboration with 9 data partners, Friends of Cancer Research convened the Real-world Evidence Pilot 2.0, to examine trends and real world (rw) data endpoints in immunotherapy (IO) use for the front line treatment of aNSCLC.

**RESULTS:**
Seven datasets identified a range of 999 to 4617 patients per dataset for this analysis. Across datasets, 2508, 3446, and 4176 patients initiated treatment in 2015, 2016, and 2017, respectively. No patients received IO or IO + chemo regimens prior to 2015.

<table>
<thead>
<tr>
<th>DATA-SET</th>
<th>IO MEDIAN AGE, YRS</th>
<th>IO STAGE III/IV, %</th>
<th>IO PD-L1 +</th>
<th>IO 2015 1st HALF/ 2nd HALF, %</th>
<th>IO 2016 1st HALF/ 2nd HALF, %</th>
<th>IO 2017 1st HALF/ 2nd HALF, %</th>
<th>IO + CHEMO MEDIAN AGE, YRS</th>
<th>IO + CHEMO STAGE III/IV, %</th>
<th>IO + CHEMO PD-L1 +</th>
<th>IO + CHEMO 2015 1st HALF/ 2nd HALF, %</th>
<th>IO + CHEMO 2016 1st HALF/ 2nd HALF, %</th>
<th>IO + CHEMO 2017 1st HALF/ 2nd HALF, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>70</td>
<td>91</td>
<td>33</td>
<td>50 / &lt; 1</td>
<td>0 / 0</td>
<td>14 / 38</td>
<td>66</td>
<td>45 / 15</td>
<td>51 / 60</td>
<td>0 / 0</td>
<td>6 / 19</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>67</td>
<td>94</td>
<td>89</td>
<td>73 / 74</td>
<td>0 / 12</td>
<td>11 / 20</td>
<td>33 / 25</td>
<td>64 / 97</td>
<td>70 / 70</td>
<td>0 / 0</td>
<td>9 / 19</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>72</td>
<td>-</td>
<td>-</td>
<td>0 / 14</td>
<td>25 / 28</td>
<td>65</td>
<td>-</td>
<td>- / -</td>
<td>- / -</td>
<td>0 / 0</td>
<td>10 / 0</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>70</td>
<td>77</td>
<td>95</td>
<td>62 / 77</td>
<td>0 / 5</td>
<td>9 / 13</td>
<td>33 / 32</td>
<td>66 / 75</td>
<td>68 / 62</td>
<td>0 / 0</td>
<td>8 / 17</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>71</td>
<td>86</td>
<td>90</td>
<td>63 / 50</td>
<td>5 / 11</td>
<td>15 / 19</td>
<td>34 / 32</td>
<td>68 / 91</td>
<td>42 / 80</td>
<td>0 / 0</td>
<td>6 / 37</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>70</td>
<td>97</td>
<td>57</td>
<td>61 / 64</td>
<td>0 / 13</td>
<td>25 / 18</td>
<td>64</td>
<td>96 / 44</td>
<td>0 / 0</td>
<td>0 / 0</td>
<td>0 / 24</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>70</td>
<td>85</td>
<td>87</td>
<td>58 / 27</td>
<td>7 / 10</td>
<td>12 / 20</td>
<td>35 / 40</td>
<td>64 / 94</td>
<td>64 / 50</td>
<td>0 / 0</td>
<td>7 / 6</td>
<td>15 / 32</td>
</tr>
</tbody>
</table>

Initial approvals for IO use in aNSCLC occurred in October 2015 and for first line in metastatic NSCLC in October 2016. When examining survival at 1 year, overall, OS in PD-(L)1 + patients appeared longer than those with a PD-(L)1 - status.

**CONCLUSION:**
RWE analyses may reveal important trends in clinical cancer patient care including patterns of off-label use. The heterogeneity in the timing of IO uptake across datasets ranged from immediately after approval to ~12 months post-approval.

Approval Dates: Mar 2015 (squamous aNSCLC), Oct 2015 (aNSCLC), Oct 2016 (1st line metastatic), May 2017 (no PDL1 req.)

**LINK TO STUDY**
Annual trends in opioid prescribing for patients (Pts) with metastatic non-small cell lung cancer (mNSCLC): CancerLinQ data analysis, 2010 to 2017

Paice et. al. | Abstract #2076

**BACKGROUND:**

Despite opioid misuse and abuse, opioids remain a mainstay for management of cancer pain. Government, payers, and institutions have implemented policies to reduce opioid use. The impact of these restrictions on oncologist prescriptions (Rx) of opioids and management of cancer pain in pts with cancer is not well known.

**RESULTS:**

Overall, 39.8% of pts had opioid Rx in 2010-2017. 18,106 pts with mNSCLC clinical activity between 2010 and 2017 were identified. Overall, 39.8% of pts had opioid Rx in 2010-2017. Annual Rx rates increased from 2010-2015 and fell 2016-2017 (see table). Hydrocodone was the second most frequently prescribed opioid overall (N=4211 pts), but Rx rates began to decline in 2012. Tramadol and acetaminophen + codeine Rx rates gradually increased throughout the time period. DEA initially scheduled Tramadol as schedule IV in 2014.

**TABLE:**

<table>
<thead>
<tr>
<th>YEAR OF ACTIVITY</th>
<th>NUMBER OF PTS WITH DIAGNOSIS N</th>
<th>PTS WITH OPIOID RX N (%)</th>
<th>PTS WITH HYDROCODONE RX AMONG OPIOID RX PTS</th>
<th>PTS WITH TRAMADOL AND/OR ACETAMINOPHEN + CODEINE RX AMONG OPIOID RX PTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>2520</td>
<td>449 (18%)</td>
<td>118 (26%)</td>
<td>15 (3%)</td>
</tr>
<tr>
<td>2011</td>
<td>2647</td>
<td>550 (21%)</td>
<td>212 (39%)</td>
<td>22 (4%)</td>
</tr>
<tr>
<td>2012</td>
<td>4084</td>
<td>878 (21%)</td>
<td>324 (37%)</td>
<td>53 (6%)</td>
</tr>
<tr>
<td>2013</td>
<td>4823</td>
<td>1256 (26%)</td>
<td>440 (35%)</td>
<td>87 (7%)</td>
</tr>
<tr>
<td>2014</td>
<td>4953</td>
<td>1579 (32%)</td>
<td>535 (34%)</td>
<td>136 (9%)</td>
</tr>
<tr>
<td>2015</td>
<td>5336</td>
<td>1807 (34%)</td>
<td>589 (33%)</td>
<td>160 (9%)</td>
</tr>
<tr>
<td>2016</td>
<td>5067</td>
<td>1676 (33%)</td>
<td>541 (32%)</td>
<td>169 (10%)</td>
</tr>
<tr>
<td>2017</td>
<td>4061</td>
<td>1235 (30%)</td>
<td>398 (32%)</td>
<td>145 (12%)</td>
</tr>
</tbody>
</table>

**CONCLUSION:**

Opioids are commonly prescribed by oncologists for patients with mNSCLC. Rx rates have declined since 2015, likely due to increased government, payer, and institutional restrictions on access. Hydrocodone Rx declined since 2012, perhaps exacerbated by reclassification from schedule III to schedule II by the DEA (October 2014). Rxs for schedule IV and III opioids (known to be of lower potency) increased modestly, likely due to comparatively fewer prescribing restrictions. Additional research is needed to understand whether the decline continues and the impact on management of cancer pain, particularly among metastatic patients.

**LINK TO STUDY**

18,106 pts with mNSCLC clinical activity between 2010 and 2017 were identified. Overall, 39.8% of pts had opioid Rx in 2010-2017. Annual Rx rates increased from 2010-2015 and fell 2016-2017 (see table). Hydrocodone was the second most frequently prescribed opioid overall (N=4211 pts), but Rx rates began to decline in 2012. Tramadol and acetaminophen + codeine Rx rates gradually increased throughout the time period. DEA initially scheduled Tramadol as schedule IV in 2014.
Vital status ascertainment in CancerlinQ Discovery (CLQD): Improvement in mortality capture with a supplemental data source

Potter et. al. | Abstract #7064

**BACKGROUND:**

Overall survival (OS) is the gold standard outcome in clinical cancer research but many clinical trials cannot assess long-term OS. Real-world data sources can be used to calculate long-term OS, but only if vital status is accurately captured.

**RESULTS:**

The addition of OBD modestly changes OS estimates (see Table). Among a subset of patients with death dates in both CLQD and OBD, dates were highly correlated for breast ($r = 0.98$), lung ($r = 0.93$), ovarian ($r = 0.99$), and pancreatic ($r = 0.88$) cancers. When date differences existed, they were ≤10 days for > 95% of the patients. These results suggest death dates are reliable in CLQD EMRs. OS curves were as expected, with OS decreasing by stage and age at diagnosis.

**CONCLUSION:**

Incorporating obituarydata.com (OBD) modestly improves OS estimates and shows that when death data is present in CLQD, it is reliable. Future enhancements will focus on improving sensitivity of mortality ascertainment with external data linkages, without compromising specificity.

<table>
<thead>
<tr>
<th></th>
<th>24 MONTH OS</th>
<th>36 MONTH OS</th>
<th>48 MONTH OS</th>
<th>60 MONTH OS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Breast</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMR</td>
<td>0.96 (0.96-0.96)</td>
<td>0.94 (0.94-0.95)</td>
<td>0.93 (0.93-0.93)</td>
<td>0.91 (0.91-0.91)</td>
</tr>
<tr>
<td>EMR+OBD</td>
<td>0.96 (0.95-0.96)</td>
<td>0.93 (0.93-0.93)</td>
<td>0.91 (0.91-0.91)</td>
<td>0.89 (0.89-0.89)</td>
</tr>
<tr>
<td><strong>Lung</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMR</td>
<td>0.59 (0.58, 0.59)</td>
<td>0.51 (0.50, 0.51)</td>
<td>0.46 (0.45, 0.46)</td>
<td>0.42 (0.41, 0.42)</td>
</tr>
<tr>
<td>EMR+OBD</td>
<td>0.54 (0.54, 0.55)</td>
<td>0.46 (0.46, 0.47)</td>
<td>0.40 (0.40, 0.41)</td>
<td>0.36 (0.35, 0.36)</td>
</tr>
<tr>
<td><strong>Ovarian</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMR</td>
<td>0.88 (0.87, 0.89)</td>
<td>0.82 (0.82, 0.83)</td>
<td>0.77 (0.77, 0.78)</td>
<td>0.73 (0.72, 0.74)</td>
</tr>
<tr>
<td>EMR+OBD</td>
<td>0.86 (0.85, 0.86)</td>
<td>0.79 (0.78, 0.80)</td>
<td>0.73 (0.72, 0.74)</td>
<td>0.67 (0.66, 0.68)</td>
</tr>
<tr>
<td><strong>Pancreatic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMR</td>
<td>0.42 (0.41, 0.43)</td>
<td>0.34 (0.33, 0.35)</td>
<td>0.31 (0.30, 0.32)</td>
<td>0.29 (0.27, 0.30)</td>
</tr>
<tr>
<td>EMR+OBD</td>
<td>0.37 (0.36, 0.38)</td>
<td>0.29 (0.28, 0.30)</td>
<td>0.25 (0.24, 0.26)</td>
<td>0.23 (0.22, 0.24)</td>
</tr>
</tbody>
</table>
Overall survival (OS) in advanced non-small cell lung cancer (aNSCLC) patients treated with frontline chemotherapy or immunotherapy by comorbidity: A real-world data (RWD) collaboration  

Rivera et. al. | Abstract #e19270

BACKGROUND:
Friends of Cancer Research convened 9 data partners to identify data elements and common definitions for real world (rw) endpoints to evaluate populations typically excluded from clinical trials. Here we report on rwOS by frontline treatment and comorbidities.

RESULTS:

A total of 33,649 patients were included (N 972-17,454).

<table>
<thead>
<tr>
<th>BRAIN METASTASES</th>
<th>NO EVIDENCE OF BRAIN METASTASES</th>
<th>ECOG 2+*</th>
<th>ECOG 0/1*</th>
<th>MODERATE/SEVERE KD*</th>
<th>NORMAL/ MILD KD*</th>
<th>MODERATE/SEVERE LD*</th>
<th>NORMAL/ MILD LD*</th>
<th>CKD</th>
<th>NO EVIDENCE OF CKD</th>
<th>CLD</th>
<th>NO EVIDENCE OF CLD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Platinum Doublet (n = 812 to 14240)</strong></td>
<td>Population % Median (Range)</td>
<td>12 (2-28)</td>
<td>88 (73-98)</td>
<td>76 (34-82)</td>
<td>2 (1-2)</td>
<td>99 (98-99)</td>
<td>1 (0-2)</td>
<td>99 (98-100)</td>
<td>4 (1-17)</td>
<td>97 (83-100)</td>
<td>4 (0-20)</td>
</tr>
<tr>
<td>12 mo OS % Median (Range)</td>
<td>44 (25-58)</td>
<td>52 (44-66)</td>
<td>56 (48-64)</td>
<td>33 (30-53)</td>
<td>46 (43-62)</td>
<td>29 (22-31)</td>
<td>46 (43-64)</td>
<td>42 (32-60)</td>
<td>51 (43-64)</td>
<td>58 (34-65)</td>
<td>50 (43-64)</td>
</tr>
<tr>
<td><strong>IO mono or combo (n = 34 to 3214)</strong></td>
<td>Population % Median (Range)</td>
<td>13 (0-40)</td>
<td>87 (60-100)</td>
<td>74 (22-80)</td>
<td>2 (2-3)</td>
<td>97</td>
<td>29 (22-31)</td>
<td>46 (43-64)</td>
<td>42 (32-60)</td>
<td>51 (43-64)</td>
<td>58 (34-65)</td>
</tr>
<tr>
<td>12 mo OS % Median (Range)</td>
<td>60 (47-73)</td>
<td>38 (29-48)</td>
<td>61 (45-68)</td>
<td>54 (39-62)</td>
<td>56 (47-57)</td>
<td>22 (22)</td>
<td>54 (47-58)</td>
<td>63 (48-70)</td>
<td>50 (46-71)</td>
<td>50 (39-71)</td>
<td>59 (47-71)</td>
</tr>
</tbody>
</table>

* Among patients with available data Strata with < 10 patients were excluded from the summary

There was a broad range of comorbidity prevalence across datasets and patients with evidence of comorbidity had comparatively shorter 12-month OS (Table).

CONCLUSION:
RWD analyses can generate expanded evidence on patient outcomes for populations routinely excluded from clinical trials and may help inform decision making where sparse data exist on appropriate treatment approaches. Additional understanding of data missingness, sensitivity of definitions, and covariate adjustment are needed to make direct comparisons across regimens and data sources.
Development of an algorithm using natural language processing to identify metastatic breast cancer patients from clinical notes

Swaminathan et. al. | Abstract #e14056

**BACKGROUND:**

Determination of the metastatic status of a patient is important for outcomes research and candidacy for clinical trials. Structured data in EMR may not always capture the metastatic status, and it is useful to extract it automatically from physician notes. Contextual understanding of the notes is important to resolve issues such as a) local vs distal metastasis b) statements involving family history of metastasis or physician instructing the patient to look for certain signs of metastasis c) text indicating suspicion of metastasis or absence of metastasis d) indirect utterances, e.g. cancer has spread to the bone. e) corrections to previous findings.

**RESULTS:**

At a sentence level, we obtained an accuracy of 0.85 for the distal/local vs suspicious vs irrelevant model; and 0.97 for the distal vs not distal metastasis model.

The classes used for sites of metastasis are Brain, Bone, Lung, Liver, Distant Lymph nodes & Unknown sites. Subset accuracy (mean fraction of labels which match) of 0.93 was obtained on the hold out test set at patient level.

<table>
<thead>
<tr>
<th>CLASS</th>
<th>PRECISION</th>
<th>RECALL</th>
<th>F1-SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distal Metastasis</td>
<td>0.89</td>
<td>0.77</td>
<td>0.82</td>
</tr>
<tr>
<td>No Distal Metastasis</td>
<td>0.92</td>
<td>0.97</td>
<td>0.94</td>
</tr>
</tbody>
</table>

**CONCLUSION:**

Metastatic status & site of metastasis can be reliably extracted automatically from clinical notes using deep learning techniques. This information will be valuable for clinical trial matching, outcomes research and other applications.
Development of an artificial intelligence model to dynamically predict metastatic recurrence of early-stage breast cancer patients

Vaidya et. al. | Abstract #e13078

BACKGROUND:
Models that can dynamically predict risk of metastatic breast cancer (MBC) recurrence based on cumulative historical clinical data could help guide patient care & surveillance decisions. The objectives of this study were to predict risk of MBC recurrence dynamically from any point after 1 year of initial diagnosis in a BC patient’s journey. We show representative results for predicting 4 year risk post 1 year of date of diagnosis. There are established models to predict risk of distant recurrence at the time of diagnosis but we have not found much work on dynamic risk scores.

RESULTS:
The performance of various machine learning (ML) algorithms for predicting metastatic BC recurrence within 4 years post 1 year from date of diagnosis is provided in the table with sensitivity held constant at 0.7. Key variables influencing the results in each model are also indicated. The Extremely Random Forest model for predicting risk of metastatic recurrence within 1 year from 1 year post diagnosis yielded an AUC of 0.814 & a balanced accuracy of 0.719.

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>BALANCED ACCURACY</th>
<th>PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>0.783</td>
<td>0.72</td>
<td>0.35</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.787</td>
<td>0.71</td>
<td>0.34</td>
</tr>
<tr>
<td>XGBoost</td>
<td>0.825</td>
<td>0.74</td>
<td>0.40</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.836</td>
<td>0.75</td>
<td>0.40</td>
</tr>
<tr>
<td>Extremely Random Forest</td>
<td>0.848</td>
<td>0.76</td>
<td>0.44</td>
</tr>
</tbody>
</table>

CONCLUSION:
An AI model to predict risk of metastatic recurrence in breast cancer patients built using a real world dataset yielded promising results. Furthermore, analysis of input variables provided insights not only into the key features driving metastatic recurrence risk such as previous surgery, tumor subtype, stage & age at diagnosis etc. Such a model could be a useful for assessing patient risk & treatment options at various points in a breast cancer patients journey as well as stratify patients for different levels of surveillance.
Authors

01 Abstract #19318. Smita Agrawal, Babu Narayanan, Prajval Chandrashekharaiya, Sai Vinod M, Sourav Nandi, Vivek Vaidya, Ping Sun, Claudia Cabrera, David Svensson, Saajan Khosla, Edward Stepanski, George Anthony Komatsoulis; Concerto HealthAI, Bengaluru, India; Concerto HealthAI, Bangalore, India; Concerto Health AI, Bangalore, India; Concerto HealthAI, Bangalore, India; Concerto HealthAI, Bengaluru, India; AstraZeneca, Gaithersburg, MD; AstraZeneca Pharmaceuticals, Gotthenburg, Sweden; AstraZeneca, Cambridge, United Kingdom; Concerto HealthAI, Boston, MA, American Society of Clinical Oncology’s (ASCO) CancerLinQ, Alexandria, VA

02 Abstract #19046. Ash Alpert, Suanna S. Bruinooge, Don S. Dixon, Becky Kononowski, Elizabeth Garrett-Mayer, Stephen C. Meersman, Robert S. Miller, Danielle Potter, George Anthony Komatsoulis; Wilmot Cancer Center, University of Rochester Medical Center, Rochester, NY; American Society of Clinical Oncology, Alexandria, VA; Massachusetts General Hospital, Boston, MA; Concerto HealthAI, Boston, MA, American Society of Clinical Oncology’s (ASCO) CancerLinQ, Alexandria, VA

03 Abstract #21596. Francois Charest, Afsaneh Ghanbedi, Babu Narayanan; Concerto HealthAI, New York, NY; Concerto HealthAI, New York, NY; Concerto HealthAI, Alexandria, VA; Concerto HealthAI, New York, NY; Concerto HealthAI, Alexandria; Castle Biosciences, Inc., Friendswood, TX; American Society of Clinical Oncology, Alexandria, VA

04 Abstract #6041. Summer Dewdney, Danielle Potter, Joy Larsen Haidle, Peter J Hulick, Mark Rifflon, Federico A. Monzon, Sameer R. Keole, Robert S. Miller, Rush University Medical Center, Chicago, IL; American Society of Clinical Oncology, Alexandria, VA; Humphrey Cancer Center, Robbinsdale, MN; NorthShore University Health System, Evanston, IL; CancerLinQ, Alexandria; Castle Biosciences, Inc., Friendswood, TX; Univ of Florida, Gainesville, FL; American Society of Clinical Oncology’s (ASCO) CancerLinQ, Alexandria, VA

05 Abstract #15075. Brian Dreyfus, Samuel P Heilbroner, Reed Few, Christine Kratt, Andres Gomez, Bristol-Myers Squibb, Lawrence Township, NJ; Concerto Health AI, New York, NY; Concerto Health AI, Boston, N

06 Abstract #19311. Lawrence H. Kushi, Laura Lasiter, Andrew J. Belli, Marley Boyd, Suanna S. Bruinooge, Jennifer Christian, Elizabeth Garrett-Mayer, Eric Hansen, Rebecca Honnold, Ruth A. Benito, Yanina Natanzon, Lori Sokoda, Donna R. R Rivera, Whitney Rhodes, Nicholas J. Robert, Elad Sharon, Connor Sweetnam, Joseph Wagner, Mark S. Walker, Jeff Allen, Kaiser Permanente, Oakland, CA; Friends of Cancer Research, Washington, DC; COTA, Inc., New York, NY; McKesson Life Sciences, The Woodlands, TX; American Society of Clinical Oncology, Alexandria, VA; IQVIA, Research Triangle Park, NC; Tempus Labs, Chicago, Tempus, Chicago, IL; Syapse, San Francisco, CA; Kaiser Permanente Northern California, Division of Research, Oakland, CA; National Cancer Institute, Rockville, MD; Concerto HealthAI, Memphis, TN; 10101 Woodloch Forest, The Woodlands, TX; National Cancer Institute, Bethesda, MD; Syapse Inc., San Francisco, CA; IQVIA, Plymouth Meeting, PA; ACRON Research LLC, Memphis, TN

07 Abstract #2076. Judith A. Pace, Li Chen, Elizabeth Garrett-Mayer, Karen S Haqerty, Kristina Lynne Maletz Novick, Danielle Potter, Mark Rifflon, Whitney Rhodes, Liya Wang, Suanna S. Bruinooge, Northwestern University, Chicago, IL; Concerto HealthAI, Boston, MA, American Society of Clinical Oncology, Alexandria, VA; Univ of Rochester, Rochester, NY; CancerLinQ, Alexandria; Concerto HealthAI, Memphis, TN

08 Abstract #7064. Danielle Potter, Melinda Kaltenbaugh, Shaun Kabadi, Aliki Taylor, Esther Pascal, Jacob Koskimaki, Nancy Zanni, Philip Stoeber, Jose Mena, Elizabeth Garrett-Mayer; American Society of Clinical Oncology, Alexandria, VA; AstraZeneca LP, Gaithersburg, MD; AstraZeneca, Cambridge, United Kingdom; CancerLinQ, Alexandria, VA; CancerLinQ, LLC, Alexandria, VA

09 Abstract #19270. Donna R. R Rivera, Laura Lasiter, Jennifer Christian, Lindsey Ennewold, Janet L. Espirito, Eric Hansen, Henry J. Henk, Lawrence H. Kushi, Daniel Lane, Yanina Natanzon, Ruth A. Benito, Erik Rasmussen, Nicholas J. Robert, Mark Stewart, Connor Sweetnam, Olga Tymejczyk, Emily Valice, Joseph Wagner, Alia Zander, Jeff Allen, National Cancer Institute, Rockville, MD; Friends of Cancer Research, Washington, DC; IQVIA, Research Triangle Park, NC; National Cancer Institute, Bethesda, MD; McKesson Life Sciences, The Woodlands, TX; COTA, Inc., New York, NY; OptumLabs, Shapooke, MN; Kaiser Permanente, Oakland, CA; COTA Healthcare, Boston, MA; Syapse, San Francisco, CA; Tempus, Chicago, IL; Flatiron Health, New York, NY; 10101 Woodloch Forest, The Woodlands, TX; Syapse Inc., San Francisco, CA; Kaiser Permanente Division of Research, Oakland, CA; IQVIA, Plymouth Meeting, PA

10 Abstract #14056. Krishna Kumar Swaminathan, Emma Vendronca, Pranay Mukherjee, Karpagavalli Thirumalai, Rachel Newsome, Babu Narayanan; Concerto HealthAI, Bangalore, India; Concerto HealthAI, Bengaluru, India; Concerto HealthAI, Boston, MA

11 Abstract #13078. Vivek Prabhaikar Vaidya, Smita Agrawal, Sai Vinod M, Sandeep Nagdevani, Prajwal Chandrashekharaiya, Tapasoy Bhardwaj, Babu Narayanan, SymphonyAI, Bengaluru, India; Strad Center for Genomics and Personalized Medicine, Bangalore, India; Concerto Health AI, Bangalore, India; Concerto HealthAI, Bangalore, India; ConcertoAI, WA, India

All abstracts featured herein are © 2020 American Society of Clinical Oncology, Inc.—used with permission. These materials are being provided for informational purposes only. None of the information contained herein is intended as medical, legal, or business advice, or for use in the diagnosis or treatment of individual conditions. The mention of any product, service, organization, activity, or therapy should not be construed as ASCO’s or CancerLinQ’s endorsement. Neither ASCO nor CancerLinQ is responsible for any injury or damage to persons or property arising out of or related to any use or misuse of the foregoing materials, and readers are encouraged to read the full abstracts in their entirety and consult their healthcare practitioner with any questions concerning individual care.
ASCO CancerLinQ™
Shaping The Future Of Cancer Care