Nearly half the Californians born in 2015 will develop cancer at some point in their lives. And it is likely that one in five of them will die of the disease, the second most common cause of death in the United States.

As cancer places an increasing burden on the health care system, painting an accurate picture of cancer care becomes ever more important. While several screening measures are routinely reported, information on the quality of cancer treatment has long been missing from clinical quality measures reported to the public. No single source of data has the full information needed for accurate, meaningful measurement.

This White Paper reports on a project that endeavored to fill that gap by bringing together complementary sources of cancer data to measure quality of cancer treatment and provider performance. Health plan commercial claims and California Cancer Registry (CCR) data were successfully linked, and quality measure results were generated at the statewide, regional and physician organization (PO) levels. The project broke new ground in that claims and registry data had not been combined previously in California. The findings are summarized here—along with a number of suggestions for future endeavors aimed at improving the quality of cancer care and patient outcomes.

**PROJECT SUMMARY AND HIGHLIGHTS**
California consumers have long lacked meaningful information on the quality of cancer treatment delivered in the state. Available data sources alone are insufficient. Health plan data lack critical clinical information, and cancer registry data lack services provided in certain settings and details on subsequent therapy or recurrences.

This project linked these two sources, developing specifications to make the combined data available as standardized measures. Results—which were calculated statewide, by various geographies, and for physician organizations serving commercial HMO enrollees in California—revealed geographic variation in the quality of cancer treatment.

Challenges in the project included lengthy data access processes, low data linkage rates, and insufficient sample size for reliable measurement and analysis at the physician organization level.

The project team offers tangible suggestions for improving the linkages and timeliness of data, expanding the allowable uses of data, and streamlining the data access process.

**BOTH CLAIMS AND REGISTRY DATA NEEDED FOR FULL PICTURE**
Health plan administrative data, composed of health plan claims and encounter data, contain a wealth of information on diagnoses and treatment from all providers in most settings—including inpatient, outpatient and pharmacy. However, those data currently lack clinical information critical for evaluating cancer care—such as confirming the diagnosis of malignancy, cancer stage and details of tumor pathology.

Cancer registry data contain a definitive diagnosis of cancer and the clinical information missing in claims, in addition to initial treatment information, but often do not contain detailed information on subsequent therapy or recurrence. Together, these sources can bring a new level of accuracy and depth to cancer treatment quality measurement.

Exhibit 1 illustrates where different pieces of data are captured throughout a sample episode of colorectal cancer, and together, how they capture a more complete understanding.

**Exhibit 1. Sample Colorectal Cancer Episode**

<table>
<thead>
<tr>
<th>Timeline</th>
<th>Diagnosis and Treatment Path</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Colonoscopy (Diagnostic procedure)</td>
<td>Claims data</td>
</tr>
<tr>
<td>Step 2</td>
<td>Biopsy to confirm malignancy</td>
<td>Registry data (Results and stage)</td>
</tr>
<tr>
<td>Step 3</td>
<td>Colectomy with lymph node resection</td>
<td>Claims data (Surgical procedure)</td>
</tr>
<tr>
<td>Step 4</td>
<td>Adjuvant chemotherapy (For high risk stage II and all stage III)</td>
<td>Claims data</td>
</tr>
<tr>
<td>Step 5</td>
<td>Metastatic disease (Multiple chemotherapy treatment options)</td>
<td>Claims data</td>
</tr>
</tbody>
</table>
Linked claims and cancer registry data have proven useful in other settings for understanding cancer outcomes, as well as variations in cancer care and quality. For example, the Surveillance Epidemiology and End Results (SEER) Medicare database linked cancer registry and Medicare claims data for patients aged 65 and older, and has been used for many studies. These previous efforts informed this project, which was the first effort in California to link registry data to commercial claims data.

**PROJECT OBJECTIVES AND APPROACH**

In conducting this project, the Integrated Healthcare Association (IHA) partnered with Truven Health Analytics, which offers data management and analytic services to health care payers and providers. The focus was on California’s variations in quality at the physician organization (PO) level, generating results based on commercial health plan claims data linked to population-based cancer registry data. Results were also analyzed by 19 geographic regions (those used in Covered California, the state’s health insurance exchange), as well as by five “super regions” in California.

The study focused on the two types of cancer that are most prevalent in the state: breast and colorectal. The estimated number of new cases in 2015 in California was 25,270 for female breast cancer and 14,510 for colorectal cancer.

**Measure Identification**
The project team reviewed all National Quality Forum (NQF)-endorsed breast and colorectal cancer treatment quality measures (www.qualityforum.org/QPS) and selected nine (see Exhibit 2) that could be run

### Exhibit 2. Selected Quality Measures for Breast and Colorectal Cancers

<table>
<thead>
<tr>
<th>Measure Number</th>
<th>Breast Cancer Measures</th>
<th>Measure Developer/ Steward</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC 219</td>
<td>Post Breast Conservation Surgery Irradiation</td>
<td>Radiation therapy administered within one year for women under age 70 receiving breast conserving surgery</td>
</tr>
<tr>
<td>BC 220</td>
<td>Adjuvant Hormonal Therapy</td>
<td>Tamoxifen or third generation aromatase inhibitor considered or administered within one year of diagnosis for women with American Joint Committee on Cancer Stage I, II or III progesterone or estrogen receptor positive</td>
</tr>
<tr>
<td>BC 387</td>
<td>Oncology: Hormonal Therapy</td>
<td>Tamoxifen or aromatase inhibitor received within one year of diagnosis by patients with American Joint Committee on Cancer Stage I, II or III progesterone or estrogen receptor positive</td>
</tr>
<tr>
<td>BC 559</td>
<td>Combination Chemotherapy</td>
<td>Combination chemotherapy considered or administered within four months of diagnosis for women under 70 with American Joint Committee on Cancer Stage I, II or III progesterone and estrogen receptor negative</td>
</tr>
<tr>
<td>BC 1857</td>
<td>Trastuzumab and Negative or Undocumented HER2</td>
<td>Trastuzumab not administered when Human Epidermal Growth Factor Receptor2/neu negative or undocumented</td>
</tr>
<tr>
<td>BC 1878</td>
<td>HER2 Testing</td>
<td>Test for HER2/neu overexpression or gene amplification completed within 90 days of diagnosis of American Joint Committee on Cancer Stage I, II, III or IV</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure Number</th>
<th>Colorectal Cancer Measures</th>
<th>Measure Developer/ Steward</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC 223</td>
<td>Adjuvant Chemotherapy</td>
<td>Adjuvant chemotherapy considered or administered within four months of diagnosis for patients under age 80, American Joint Committee on Cancer Stage III (lymph node positive)</td>
</tr>
<tr>
<td>CC 225</td>
<td>Twelve Node Examination</td>
<td>At least 12 regional lymph nodes removed and pathologically examined for resected colon cancer</td>
</tr>
<tr>
<td>CC 1859</td>
<td>KRAS Gene Mutation Testing</td>
<td>KRAS testing for patients with metastatic colorectal cancer completed prior to receiving anti-Epidermal Growth Factor Receptor monoclonal antibody therapy</td>
</tr>
</tbody>
</table>
from the combined CCR and health plan claims and
enrollment data.\textsuperscript{6}

**Measure Specifications**

Obtaining measure specifications proved challenging. Although the measures were endorsed by the National Quality Forum, the NQF-endorsed specifications only indicated general clinical algorithms to determine the quality of care provided.

Registry-based specifications were available, but not programmable specifications with standardized diagnosis and procedure codes for measures generated from claims data alone or claims data linked to cancer registry data. Therefore, the project team embarked on a rigorous process to develop measure specifications programmable against the linked CCR and health plan data.

Claims code sets were created for each measure—and verified with oncologists, radiation oncologists, surgeons, and pharmacists. The verified codes were used in conjunction with CCR data fields to develop algorithms applying the NQF-endorsed clinical algorithm to the linked dataset. The end result was a set of project-specific measure specifications programmable against the combined CCR and health plan data.

Data Use Agreements

Given the innovative nature of the project combining CCR and commercial health plan claims data, the terms under which each organization would participate and supply the necessary data were carefully structured. IHA worked with management and legal counsel from the participating health plans, data aggregator and CCR to formulate a common framework and develop standardized agreements. The final result was a three-way data use agreement among the data aggregator, CCR and each of the health plans.

**Data Sources**

The data aggregator received enrollment and claims data for 2009-2012 from the four participating health plans for all California commercial health maintenance organization (HMO) and point of service (POS) members, totaling about 2.5 million members per year.

The claims data included the complete range of member services: medical, facility and pharmacy claims. Broad criteria were used to identify health plan members who might have cancer. Every member with a breast or colorectal cancer diagnosis in any encounter or claim at any point in the enrollment period was included, regardless of age, gender or type of visit.

A dataset from the health plan enrollment files of members with one or more claims diagnosis for breast or colorectal cancer was provided to CCR to link to the registry data. Once the files were linked, a subset of CCR variables was attached to the health plan enrollment data file and returned to the data aggregator.

The registry data received were from 2009, 2010 and 2011—the most recent years available—and were linked back to health plan claims files from 2009 through 2012, allowing for one year of follow-up for patients diagnosed in 2011. This created a dataset containing the population of patients with a confirmed diagnosis of breast or colorectal cancer, against which the nine selected quality measures were run. Exhibit 3 summarizes the data linkage and measure generation process.

---

**Opportunities**: Leverage the specifications produced by this study as a foundation for measure specifications for other studies. Optimally, specifications using linked data would be endorsed by NQF and disseminated more widely.

As an alternative, NQF could require a claims version when approving measures.

These standardizations would streamline future efforts and support comparability of results across initiatives.

---

**Data Use Agreements**

Given the innovative nature of the project combining CCR and commercial health plan claims data, the terms under which each organization would participate and supply the necessary data were carefully structured. IHA worked with management and legal counsel from the participating health plans, data aggregator and CCR
RESULTS OF LINKAGE AND QUALITY MEASUREMENTS

There were several instructive challenges during this pioneering project linking health plan claims and CCR data for the first time in California.

Linkage Rates

Out of 36,144 members of four California commercial HMO health plans who had one or more claim with a diagnosis of breast or colorectal cancer, only 8,798 individuals (6,846 for breast cancer and 1,952 for colorectal cancer) were successfully linked to CCR.

There are several reasons why members may not have linked:

• **They did not actually have breast or colorectal cancer.** The claims-based criteria cast a wide net to identify members with cancer. The data may have included members whose claims data diagnoses were “ruled out” after diagnostic testing determined they did not have cancer.

• **They did have cancer, but were not captured in the CCR.** Omissions may be because their treatment did not involve CCR data suppliers or due to incomplete reporting by CCR data suppliers. Data from ambulatory cancer centers are often not submitted to the registry, so cancer cases treated only in an outpatient setting may be completely missing. Cancer centers and hospitals that do submit data rely on registrars at each center to collect and submit the information, which may not be 100% complete.

• **They had cancer and were present in both datasets, but linkage failed.** In certain cases, one or both datasets may have lacked the information required to connect the individuals—for example, no unique identifier such as Social Security Number was available, or there were discrepancies in member names between the two sources.

• **They received treatment as outpatients.** The largest category of non-linking members is individuals with three or more office visits. The question remains whether these members truly had a cancer episode that is not captured in the CCR data because it was treated solely on an outpatient basis, or if they were incorrectly identified by the claims dataset.

**Opportunity:** Develop and validate claims-based algorithms to accurately identify cancer cases to enable more precise linkages between claims and cancer registry data.

Additionally, CCR could collect data from all outpatient cancer treatment centers.

LACKLUSTER LINKING: TAKING A CLOSER LOOK

The project team delved deeper to better understand the causes of the unsuccessful linkages in this study. For the purposes of this analysis, we assumed:

• The greater the number of claims with a cancer diagnosis for a member, the greater the likelihood the member did have cancer.

• If the cancer diagnosis appeared only on lab or radiology claims, there was a greater likelihood the member did not have cancer—that is, the purpose of the test was to “rule out” a cancer diagnosis.

<table>
<thead>
<tr>
<th>Claims/Encounter Experience for Non-Linking Individuals</th>
<th>Likelihood of Actual Cancer Diagnosis</th>
<th>Breast Cancer</th>
<th>Colorectal Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Non-linking Individuals</td>
<td></td>
<td>21,395</td>
<td>5,951</td>
</tr>
<tr>
<td>Three or more facility* claims with a cancer diagnosis</td>
<td>HIGH</td>
<td>1,193 (6%)</td>
<td>437 (7%)</td>
</tr>
<tr>
<td>One or more inpatient hospitalization claims with a cancer diagnosis</td>
<td>HIGH</td>
<td>607 (3%)</td>
<td>440 (7%)</td>
</tr>
<tr>
<td>Identified through lab and radiology claims</td>
<td>LOW</td>
<td>2,139 (10%)</td>
<td>518 (9%)</td>
</tr>
<tr>
<td>One to two office visits with a cancer diagnosis</td>
<td>LOW - MEDIUM?</td>
<td>7,866 (37%)</td>
<td>1,981 (33%)</td>
</tr>
<tr>
<td>Three or more office visits with a cancer diagnosis</td>
<td>?</td>
<td>8,429 (39%)</td>
<td>1,675 (28%)</td>
</tr>
</tbody>
</table>

* Includes inpatient facilities, radiation treatment centers and ambulatory surgery centers.
Quality Measure Results
This project advanced cancer quality measures from common screening measures toward cancer treatment quality measures. The nine quality measures selected were run against the linked dataset comprised of members of four commercial HMO health plans who had one or more diagnosis of breast or colorectal cancer and whose diagnosis was confirmed via the CCR. Statewide, regional and PO results were generated.

Statewide Findings
At the statewide level, performance was strong across eight of nine cancer treatment quality measures, with average statewide rates ranging from 80 to 98% compliance. For example, the result for BC 219 shows that 98% of women under age 70 receiving breast conserving surgery in California received radiation therapy within one year. The ninth measure, which relied only on CCR data, showed about 50% compliance. See Exhibit 4; measure descriptions are provided in Exhibit 2.

Although the statewide rates were quite high, when the linked dataset was used to calculate results at the PO and regional levels, there was a range of performance. This may highlight an opportunity to improve the quality of cancer treatment. However, denominator sizes were generally small, limiting the analyses and conclusions.

Physician Organization Findings
About 160 physician organizations (POs) contract with the four participating health plans for commercial HMO, and the project team produced results of the cancer treatment quality measures at the PO level. Those results were mostly unreliable, with no valid results for four of the nine measures. Study results were considered to be reliable and valid when the denominator was 30 or more people—a number commonly required when measuring for accountability. For the measure with the largest number of POs having valid rates, only 18% had valid results.

Regional Findings
The project team also produced geographic results by the rating regions used in Covered California, the state health insurance exchange. At this level, there were also insufficient denominator sizes; less than half of the regions had reliable results for at least six of the measures. The results were therefore rolled up to five “super regions” used in IHA’s Value Based Pay for Performance program. Exhibit 5 shows the validity of the results by unit.

Exhibit 6 shows performance rates by super region. For breast cancer treatment quality measures, there
were notable differences across super regions, with up to 21 percentage points for Oncology: Hormonal Therapy (BC 387) between the best and worst performing regions. The Bay Area/Sacramento region generally performed the best, and Los Angeles and Inland Empire generally had the lowest performances.

Colorectal cancer treatment quality performance rates were lower overall and showed less variation across super regions and more variation across measures. Performance on Adjuvant Chemotherapy (CC 223) ranged from 81% to 93% compliance, while performance for Twelve Node Examination (CC 225) showed the most variation and the greatest opportunity for improvement, ranging from 41% to 57%. The measure of KRAS Gene Mutation Testing (CC 1859) lacked sufficient sample size across all super regions.

All breast cancer measures and all colorectal cancer measures were also combined within the super regions to illustrate overall performance. There were consistently strong results for the six breast cancer treatment quality measures combined, with the Bay Area/Sacramento having the highest performance at 95% and Inland Empire (88%) and Los Angeles (89%) having the lowest performance. The same geographic pattern held true for colorectal cancer treatment quality. Performance on the combined colorectal cancer measures ranged from 55% for Inland Empire to 68% for Bay Area/Sacramento.

Sample Size
A key limitation of these results is the small sample size. Given the low prevalence of breast and colorectal cancer in the study population—where about two-thirds of the population are under 45 years old—for most measures, the commercially insured HMO and POS populations did not result in a large enough sample when the data is broken out by physician organization. Insufficient sample size was also a limitation at the regional level, unless results were rolled up to a larger super region geographic area. This constrained analysis of the results.

Opportunity: Expand the study to include other populations, such as additional commercial HMO and PPO plans as well as Medicare and Medicaid. Including the Medicare population would be particularly useful, given that the majority of new cancers (52.8%) occur among Americans age 65 years and older.²
Timeliness of the Data
Data lags were another significant barrier. CCR data are collected from hospitals manually and not released until at least 95% complete. For this project, 2011 data became available in October 2013, meaning this pilot study includes data that are three to four years old. Securing more recent data would have required resubmitting the whole project again for approval—a prohibitively resource-intensive process.

Opportunity: Modify CCR policies for more timely access to registry data, including:
- Securing more timely data from providers, and
- Streamlining the process for adding more current data to an approved study.

Looking Ahead: More Accessible and Meaningful Data
This project demonstrated that linking commercial claims data and a population-based state cancer registry to measure quality performance is both feasible and valuable. Cancer care treatment quality performance results were more accurate than using either data source alone.

Various legal, process and technical challenges made the project burdensome, resource intensive, and impossible to replicate regularly. In addition, the small sample size limited the usefulness of the results.

Going forward, there is a tremendous opportunity to benefit from additional cancer care treatment studies. The following suggestions might help make these studies more feasible and timely.

Data Capture in Cancer Registry and Health Plan Claims
For four measures—two breast cancer and two colorectal cancer—sufficiently detailed service information was not available to generate measures from linked data.

Three of the measures were calculated solely based on claims data and one colorectal cancer measure was derived from registry data only. Because certain data fields were unavailable in one of the sources, it was not possible to use linked data to calculate these measures. Further exploration may yield appropriate alternate fields and allow the specifications to be refined to better assess these areas.

Opportunity: Provide a more detailed data dictionary for CCR data that differentiates between similar fields and provides information about data quality and completeness of each field.

Better Together: Linked Data Most Accurate
The project team also analyzed each of the measures using just registry data; just health plan claims data; and the linked registry-claims data. This allowed a comparative assessment of the two data sources that provided insights regarding the value-add of the combined source. The analysis showed that for all measures generated from the linked dataset, the performance rates relying on the linked dataset were higher and more accurate than either claims or registry data alone.

Specifically, the project found that when the numerators were defined using variables from both CCR and claims data, the results more accurately measured cancer care treatment quality.

Separately, each dataset is missing critical information. For example, the dataset containing only registry data lacks follow up treatment data while claims data is missing important information about stage of cancer. Both types of data are needed to ensure performance rates are consistent with the intent of the measures.

Performance Rates for Linked Data Versus Individual Data Sources
1. **Routinely conduct linkage of CCR and claims data for Commercial, Medicare Advantage, and Medi-Cal populations, and make the linked data readily available for research, measurement, analysis and ongoing quality improvement.**

   The linked CCR and claims dataset joins the clinical and diagnostic richness of the registry data with the completeness of the claims data for treatment in all settings. These benefits are valuable for supporting research, in addition to supporting organizational quality improvement efforts. Specifically, they enable examinations of patterns of cancer care, foster a deeper understanding of the effectiveness of treatment protocols and help identify improvement opportunities.

   A research dataset similar to the SEER-Medicare database—but using Commercial, Medicare Advantage and Medicaid populations—could be established in California. The dataset should be accompanied by standardized measures, code sets and data dictionaries, such as the ones developed in this project.

   Including the full Commercial, Medicare Advantage and Medicaid populations would allow for robust measurement that could be used for rich analyses and more reliable examination of variation. In addition, standardized tools would streamline future measurement efforts and support comparability of results across initiatives.

   However, such data and tools are only valuable and meaningful if the allowable uses are expanded and more timely data are available.

2. **Expand uses of registry data to include quality measurement and public reporting of provider performance.**

   The enabling legislation currently limits the use of CCR data to public health surveillance and approved research purposes. The legislation should also allow quality measurement and public reporting of provider performance.

   Breast and colorectal cancers are “shoppable” in that people diagnosed with these conditions may seek the most effective providers if they have the resources to do so. However, there is currently little information reliably evaluating the quality of cancer treatment, leaving cancer patients with little guidance when selecting providers. CCR currently prohibits the use of their data for publicly reporting provider results. However, measuring provider performance on cancer treatment quality and making the results available publicly would allow consumers to make informed decisions about where to seek their cancer care, making the data more relevant in a climate where consumers are disposed to comparison-shop.

   The wealth of information contained in linked registry and claims data, including robust benchmarks and regional data for comparison, could also open the door to significant quality improvement opportunities for organizations and providers—and ultimately improve patient outcomes.

3. **Create a more efficient, timely and comprehensive process for collecting cancer registry data.**

   Linked claims and cancer registry data must be more current to support quality improvement and consumer health care decisions. The present lag of two years renders the data outdated and greatly diminishes its value.

   The current CCR data collection system is resource-intensive—relying on registrars at each hospital to collect and submit the information, then holding it hostage until 95% of hospitals have submitted data. Data from ambulatory cancer centers are not always included in the registry, so cancer cases treated only in an outpatient setting may be completely missing.

   Further, the redundant, costly and time-consuming institutional review board (IRB) requirements for obtaining registry data could be eliminated. Currently, in addition to project review and approval by the Committee for the Protection of Human Subjects, an independent IRB review and approval is required. These two review processes have a significant amount of overlap and ask for similar information in different formats. There is also often a fee for independent IRBs, which must be paid annually.

   Data must also be destroyed upon completion of the study. Allowing retention of linked data, as long as data security can be assured, would also enable the use of the data for additional research.

   Registries are currently mandated in every state. California, looked to as a thought leader in many areas
of health care, could also lead by mapping out a path to make its registry data more relevant in today’s health care world and more useful in producing quality measurements—and ultimately, in improving patient care and outcomes.

Notes
2. Ibid.
6. All of these measures are also part of the Quality Oncology Practice Initiative (QOPI), an oncologist-led, practice-based quality assessment program of the American Society of Clinical Oncology.
7. For more on the Value Based Pay for Performance program, see http://www.iha.org/our-work/accountability/value-based-p4p.
9. For draft legislation facilitating this, see “Fighting Cancer With Data: Enabling the California Cancer Registry to Measure and Improve Care,” supra, note 3.

Acknowledgments
This study was generously funded by a grant from the California Health Care Foundation (CHCF). The views expressed here are those of the author and do not necessarily reflect those of CHCF.
This project could not have been completed without the participation of the health plan partners who contributed data. A team from Truven Health Analytics, led by Mahil Senathirajah, conducted the health plan data collection, aggregation, and quality measurement. Mr. Senathirajah, along with Marilyn Novich, MD, also of Truven Health, and Ann Woo, PharmD, provided overall guidance and support throughout the project. Jodi Simon, PhD helped draft this paper and Barbara Kate Repa, JD provided invaluable editing services. The participating plans, Truven Health, and Stephanie Teleki of CHCF reviewed a draft of the white paper and provided important feedback.