Role of real world data to characterize biomarker testing in routine clinical practice and inform drug and diagnostics development

*40th Annual Midwest Biopharmaceutical Statistics Workshop*

*May 23, 2017*

*Lisa Wang, Sc.D, M.P.H, Genentech, Principal Scientist, Real World Data Science*
Key Message for Today...

Real world clinical and real world biomarker testing data, provides valuable insights across all stages of the drug and diagnostics development lifecycle.
Agenda for Today

1. What is Real World Data (RWD)? Why is it important?

2. Overview of Types of Real World Data Sources
   - Medical claims data, Electronic Medical Records (EMR) and Registries
   - Opportunities and challenges of integrating real world clinical and real world biomarker testing data

3. RWD Examples and Types of research questions that can be addressed

4. Role of RWD in drug and diagnostics development

5. The future of RWD: Challenges and Opportunities
What is Real-World Data (RWD)?

*Healthcare data collected from sources outside of traditional clinical trials* – U.S. FDA

"*Health care data collected under real life practice circumstances; anything that is not an interventional study*"

– International Society for Pharmacoeconomics and Outcomes Research

"*RWD is any data not collected in conventional RCT. It includes data from existing secondary sources and the collection of new data, both retrospectively and prospectively*"

– Agency for Healthcare Research and Quality

**Real World Evidence (RWE):** is RWD that has been organised to inform a conclusion or judgment
Global Challenges for Oncology Drugs

Oncology drug and biologics makers face increasing demands for data that demonstrate the benefit and cost-effectiveness of their products. Not only are treatment cost increases considered unsustainable, but questions about new product effectiveness are also openly raised by physicians and payers alike. Pharmaceutical companies are increasingly expected to demonstrate the overall economic and clinical value (ECV) of new products to justify premium payment.

The End of “Special Status?”

A cancer center says ‘no’ to an $11,000 treatment. Will others follow?

By Sarah Kliff, Updated: October 15, 2012

The United States spends $750 billion annually on health care that does not make us any healthier. The world’s oldest private cancer center, Memorial Sloan-Kettering in New York City, announced Monday a surprising step to bring down that number. It will not offer patients a $11,000 per month cancer drug called Zaltrap.

Simply put, top executives do not believe the drug is worth the price tag. Here’s Peter Bach, Leonard Saltz and Robert Wityes in Monday’s New York Times:

The drug, Zaltrap, has proved to be no better than a similar medicine we already have for advanced colorectal cancer, while its price — at $11,063 on average for a month of treatment — is more than twice as high.
“As we participate in the current data revolution, it is important that FDA consider the possibilities of using so-called “real world” data as an important tool in evaluating not only the safety of medications but also their effectiveness.

To accomplish this will require an understanding of what questions to ask, including how such data can be generated and used appropriately in product evaluation, what the challenges are to appropriate generation and use of these data, and how to address such challenges.”

From Prescription Drug User Fee Act (PDUFA) VI Commitment Letter
Increasing Focus by Regulators to Understand Real World Evidence (RWE)

*Recent Congressional Actions to Explore Use of RWE for Regulatory Decision-Making*

**Impact:** 21st Century Cures Enacted Dec. 13, 2016: Requires FDA to develop framework & guidance for use of RWE

**Impact:** PDUFA VI (Expected by Oct. 1, 2017): Provides user fee funds for RWE activities, establishes pilot studies to inform draft guidance [aimed for 2021]

PDUFA=Prescription Drug User Fee Act
Agenda for Today

1. What is Real World Data (RWD)? Why is it important?

2. Overview of Types of Real World Data Sources
   - Medical claims data, Electronic Medical Records (EMR) and Registries
   - Opportunities and challenges of integrating real world clinical and real world biomarker testing data

3. RWD Examples and Types of research questions that can be addressed

4. Role of RWD in drug and diagnostics development

5. The future of RWD: Challenges and Opportunities
Healthcare Claims Data

• Consists of billable encounters between patients and the healthcare delivery system

• Submitted as billing codes by physicians, pharmacies, hospitals and other HCPs to payers (eg. commercial or public insurers)
Outpatient claims

- Most frequent type of claims data
- Typically provides the following data elements:
  - Date of outpatient claim
  - Provider specialty, place of service
  - Medical procedure code
  - Multiple diagnosis codes
  - Costs associated with outpatient claim

Inpatient claims

- Organized differently in different databases:
  - Overall hospitalization summary
  - Individual claims related to hospitalization
- Extract information relevant to hospitalization:
  - Total hospitalization cost
  - Length of stay
  - Diagnosis codes relevant to hospitalization
  - Medical procedure code

Pharmacy claims

- Available for patients with prescription drug coverage
- Provides the following data elements:
  - Date of prescription
  - NDC drug code
  - Days supply of drug
  - Quantity of drug dispensed
  - Costs associated with prescription

Patient Demographics

- Enrollment history
- Gender
- Year of birth -- age
- Insurance plan type
- Geographic Region
Medical Claims Data -- US Examples

- **Truven MarketScan**
  - Commercial, Medicare Supplemental, and Medicaid data on more than 180 million patients, ~50 million patients per year

- **IMS PharMetrics Plus (Blue cross/Blue Shield)**
  - Contains data from commercial health plans with medical and pharmacy information for 150 million patients, ~40 million patients per year

- **Optum Clinformatics Datamart (United Health Care)**
  - Contains complete data on ~22 million patients per year
  - With mortality data
  - Access to medical charts
EMR Data

- A digital version of a paper chart that contains all of a patient’s medical history from clinical practice.

- Physicians use EMRs to improve the quality of patient care while reducing medical costs.

- Potential comprehensive and relatively timely clinical information including:
  - physicians’ notes,
  - patient symptoms and history,
  - diagnosis information and
  - planned and actual treatments.

- The number of office-based practices using EMR system is increasing over time.

- Can be leveraged into an analyzable database.
  - However lack of standardization/completeness.

*May also be called electronic health records (EHR)*
EMR Data Example:

Flatiron Oncology Network in the U.S.

- **265** Cancer Clinics
- **2,500** Clinicians
- **1.7M** Active Cancer Patients in Network

FlatIron Provider Network

Darkest color density represents highest patient concentration
The Approach to Data Integration

Flatiron Real World Database

Data Processing

Structured Data
- Demographics
- Diagnosis
- Visits
- Labs
- e-Prescribing

Unstructured Data
- Physician Notes
- Radiology Report
- Pathology Report
- Discharge Notes

Outside Practice
- Hospital
- Lab

Standard EHR Data
No data has everything: Advantage & Disadvantages of Claims Data & EMR Data

<table>
<thead>
<tr>
<th></th>
<th><strong>Claims Data</strong></th>
<th><strong>EMR data</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scope of data</strong></td>
<td>Broad: Captures information from all doctors/providers caring for a patient</td>
<td>Limited: Captures only the portion of care provided by doctors using the EMR</td>
</tr>
<tr>
<td><strong>Scope of patients</strong></td>
<td>Insured patients only</td>
<td>All patients (including uninsured)</td>
</tr>
<tr>
<td><strong>Prescription Data</strong></td>
<td>An accurate record of all prescriptions that were filled including dates of refills</td>
<td>Contains only that a physician prescribed a drug but not whether or not it was filled/refilled</td>
</tr>
<tr>
<td><strong>Data richness</strong></td>
<td>Limited: diagnosis, procedure</td>
<td>Rich: lab results, vital signs, patient surveys, habits (smoking and alcohol use), etc.</td>
</tr>
<tr>
<td><strong>Cost and reimbursement data</strong></td>
<td>Present</td>
<td>Not present</td>
</tr>
</tbody>
</table>
### Summary of the Different Types of RWD Databases

<table>
<thead>
<tr>
<th>Data</th>
<th>Characteristics</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insurance Claims</td>
<td>• Collected for insurance and reimbursement purposes</td>
<td>TRUVEN HEALTH ANALYTICS</td>
</tr>
<tr>
<td></td>
<td>• Often include a number of health plans</td>
<td>IMS LifeLink PharMetrics Plus™</td>
</tr>
<tr>
<td></td>
<td>• Often with &gt;10s of millions currently enrolled pts</td>
<td>OPTUM™</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KAISER PERMANENTE®</td>
</tr>
<tr>
<td>Health Provider</td>
<td>• Smaller population than insurance claims database</td>
<td>GE Healthcare</td>
</tr>
<tr>
<td>Claims</td>
<td>• Higher data integrity, complete knowledge of database</td>
<td>CPRD</td>
</tr>
<tr>
<td></td>
<td>• Possible to link with provider’s EMRs</td>
<td>FLATIRON</td>
</tr>
<tr>
<td>EMR</td>
<td>• Data collected for quality of care, performance measure, utilization, clinical research</td>
<td>SEER</td>
</tr>
<tr>
<td></td>
<td>• Type of patient records included in EMR system may vary (eg. GP only vs GP+ specialist visits +hospital stays)</td>
<td>AMERICAN ACADEMY OF OPHTHALMOLOGY</td>
</tr>
<tr>
<td></td>
<td>• Valuable details in unstructured data (notes)</td>
<td></td>
</tr>
<tr>
<td>Registry</td>
<td>• Can be disease-specific or product-specific</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Essential to study rare conditions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Access to data can vary</td>
<td></td>
</tr>
</tbody>
</table>
Integrating Real World Clinical and Biomarker Data: Opportunities

• Biomarker testing has the promise to make medicine more ‘precise’, ‘individualized’ and better (eg. Precision Medicine Initiative) and is becoming part of routine clinical practice

• Growing numbers of available diagnostics tests and patients having ‘biomarker’ testing (eg. genomic testing)

• More efforts to aggregate different RWD sources (e.g. claims, EMR, biomarker data) or build databases with integrated real world clinical and biomarker testing data
  – Eg. U.S. Precision Medicine Initiative -1 million patient cohort;
  – Genomics England-100,000 Genomes Project
Integrating Real World Clinical and Biomarker Data: *Challenges*

- Heterogeneity and lack of standardization in structure and storage of biomarker data

- Wide variety in:
  - types of assay methods and technologies (eg. sequencing, RNA-based, IHC, FISH);
  - groups conducting the test (eg. in-hospital labs, large and small commercial labs);
  - Regulatory requirements (eg. IVD vs LDTs);
  - Number of biomarkers tested (eg. one gene, large panel of genes, whole genome sequencing);
  - how results are reported (eg. from pathology lab, in-house lab, outsourced lab);

- Patients clinical data (eg. diagnosis, treatment, outcomes) and genomic/molecular testing result data tends to be siloed, not easily combined and not sufficiently comprehensive
Example: Flatiron & Foundation Medicine Partnered to create a Real World Clinico-Genomic Database

**FLATIRON**
Real-world, longitudinal patient-level clinical data from Electronic Health Records (EHRs) from cancer clinics

**FOUNDATION MEDICINE**
Deep NGS profiling across hundreds of cancer-related genes for each patient’s tumor

Real-World Clinico-Genomic Database
First-in-class, continuously refreshed data platform that analyzes and integrates two very large, real-world clinical and genomic datasets

© 2017 Flatiron Health, Inc. Proprietary and confidential
Agenda for Today

1. What is Real World Data (RWD)? Why is it important?

2. Overview of Types of Real World Data Sources
   - Medical claims data, Electronic Medical Records (EMR) and Registries
   - Opportunities and challenges of integrating real world clinical and real world biomarker testing data

3. RWD Examples and Types of research questions that can be addressed

4. Role of RWD in drug and diagnostics development

5. The future of RWD: Challenges and Opportunities
RWD Example: Analysis of Biomarker and Diagnostics Testing using Flatiron EMR Data

- Used the Flatiron EMR database of advanced NSCLC patients in the U.S. to conduct a study looking at the use of Foundation Medicine (FMI) testing and other biomarker testing types in routine clinical practice.
- Did NOT use the linked Flatiron-FMI Clinico-genomic database
- Published as ASCO 2017 abstract titled “Characteristics of advanced NSCLC patients receiving molecular diagnostic testing in US Routine Clinical Practice”

- 30,489 advanced NSCLC pts
- Diagnosed from Jan 1, 2011 to Sept 30, 2016
- Pt follow-up through Sept 30, 2016

Pts classified into 4 groups* based on presence of biomarker testing for EGFR, ALK, KRAS, ROS1, PD-L1

- FMI tested pts (n=1,019)
- NGS tested (non-FMI) pts (n=1,237)
- non-NGS tested pts (n=15,205)
- Pts with no test (n=12,938)

*The 4 groups are mutually exclusive.

Hierarchy such that pts with any FMI test, even if they also had other NGS or non-NGS tests were categorized as an FMI tested pt. Pts with any NGS test (non-FMI), even if they had a non-NGS test would be categorized as NGS tested pts.
### Analysis of FMI and biomarker testing in Flatiron EMR

**- Reported results and types of questions that can be addressed by RWD**

<table>
<thead>
<tr>
<th>Type of Question</th>
<th>Specific Research Question of Interest</th>
<th>Results presented in ASCO abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trends over time (eg. diagnostic testing, treatments, population size)</td>
<td>• Has NGS testing or FMI testing changed in more recent years?</td>
<td><em>the number of pts with FMI or NGS testing has increased in recent years.</em></td>
</tr>
<tr>
<td>Comparison between patient subgroups</td>
<td>• What are the characteristics of patients who received an FMI vs NGS vs non-NGS test? &lt;br&gt; • Do the patient characteristics differ between testing groups?</td>
<td><em>Pts in the FMI group tended to be younger (66 vs. 68-69 years), non-smokers (25% vs. 17-19%) and have squamous cell histology (13% vs. 8-10%) compared to other testing groups.</em></td>
</tr>
<tr>
<td>Timing and sequence of testing and treatment</td>
<td>• When are patients tested in relation to treatment?</td>
<td><em>30% of FMI pts received testing prior to initiating first and 53% of non-NGS pts.</em></td>
</tr>
</tbody>
</table>
Other published examples using RWD

ISPOR 2016 poster: Flatiron EMR data

Real world anaplastic lymphoma kinase (ALK) rearrangement testing patterns, treatment sequences and survival of ALK-inhibitor treated patients
Jessica Davies,1 Michael Martinec,2 Mathieu Coudert,3 Paul Delmar,3 Ursula Becker,3 Gracy Crane1
1Roche Products Ltd, Welwyn, UK; 2F. Hoffmann-La Roche, Boulogne-Billancourt Cedex, France; 3F. Hoffmann-La Roche AG, Basel, Switzerland

2016 manuscript: California Cancer Registry data

Occurrence and outcome of de novo metastatic breast cancer by subtype in a large, diverse population
Li Tao1 · Laura Chu2 · Lisa I. Wang2 · Lisa Moy1 · Melissa Brammer2 · Chunyan Song2 · Marjorie Green2 · Allison W. Kurian3,4 · Scarlett L. Gomez1,3,4 · Christina A. Clarke1,3,4

ASCO GI 2017: English National data and Univ of Leeds hospital cohort

RWD can also be used to look describe and compare clinical outcomes (eg. survival, time to tumor progression) in patients receiving different treatments or tests; or who differ by biomarker status.
Example: assess if a biomarker is prognostic in a broad real-world population
Agenda for Today

1. What is Real World Data (RWD)? Why is it important?

2. Overview of Types of Real World Data Sources
   - Medical claims data, Electronic Medical Records (EMR) and Registries
   - Opportunities and challenges of integrating real world clinical and real world biomarker testing data

3. RWD Examples and Types of research questions that can be addressed

4. Role of RWD in drug and diagnostics development

5. The future of RWD: Challenges and Opportunities
**Role of RWD in drug and diagnostics development**

-RWD Complements RCT

<table>
<thead>
<tr>
<th>Randomized Clinical Trials</th>
<th>Real World Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controlled setting</td>
<td>Real world, reflect actual practice</td>
</tr>
<tr>
<td>Randomized/ assigned treatment</td>
<td>Observed treatment</td>
</tr>
<tr>
<td>Narrower inclusion criteria</td>
<td>Broad inclusion/ disease based</td>
</tr>
<tr>
<td>Academic/ research institutes</td>
<td>Many treatment settings, e.g. community, public, academic</td>
</tr>
<tr>
<td>Smaller sample size</td>
<td>Typically larger sample size</td>
</tr>
<tr>
<td>Typically shorter follow-up</td>
<td>Typically longer follow-up</td>
</tr>
<tr>
<td>Clinical and safety</td>
<td>Also used to evaluate real world Healthcare Resource</td>
</tr>
<tr>
<td></td>
<td>Utilization and cost</td>
</tr>
</tbody>
</table>
1. **Epidemiology:** What is the estimated number of patients in a specific indication or eligible for a trial? Patient populations may be defined based on biomarker prevalence, treatments received or other patient characteristics.

2. **Patient Characteristics:** What type and frequency of comorbidities and co-medication use are anticipated in our trial patients and in the broader real world population?

3. **Treatment and testing patterns:** What is the frequency and sequence of treatments used? What is real world standard of care treatment? to inform choice of comparator arm in clinical trial design

4. **Unmet need:** How well does the current standard of care treatment work? Can we identify biomarker-defined pts populations where no targeted therapy currently exists?

5. **Burden of Disease:** What is it currently costing payers and patients to treat this disease (financially and in quality of life)?

6. **Drug’s Safety and Effectiveness:** Is our drug safe and effective in the broader, real world population? Does it work in populations not studied in our trials? (eg. older aged, pregnant women, pts with extensive comorbidities or organ dysfunction)
RWD Uses Across Diagnostics Development

Data-driven insights steer product development

Examples for RWD use

1. Identify unmet medical needs and development opportunities in the community setting
2. Quantify testing demand, disease burden and early economic impact
3. Verify real world treatment patterns and endpoints in targeted patient cohorts
4. Project clinical benefits to improve health-decision making and patient outcomes
5. Support market access with real-world evidence driven value dossier
6. Reduce barriers to product adoption and drive payer adoption with real-world and cost-effectiveness evidence, budget-impact considerations
Agenda for Today

1. What is Real World Data (RWD)? Why is it important?

2. Overview of Types of Real World Data Sources
   - Medical claims data, Electronic Medical Records (EMR) and Registries
   - Opportunities and challenges of integrating real world clinical and real world biomarker testing data

3. RWD Examples and Types of research questions that can be addressed

4. Role of RWD in drug and diagnostics development

5. The future of RWD: Challenges and Opportunities
RWD Landscape Is Continually Evolving

More Opportunities For RWD/E To Impact Healthcare and Challenges to Overcome

Growing **new data sources** e.g. mHealth, social media, sensors, etc. Increasing **data storage and integration** capabilities.

Increasing efforts on **standardization**, e.g. terminology. Many initiatives to guide RWD collection and use e.g. IMI

**Emerging science** fragmenting treatment populations, early disease interception, etc. drives innovative evidence generation strategies

**Data sharing regulations** and **privacy concerns** may become barriers for access; yet patients are more willing to be engaged directly.

**Rapidly advancing analytics tools** e.g. machine-learning. **User-friendly interfaces** and **rapid-cycle analytics** promote adoption.

Public/ private initiatives to drive **collaborations** and infrastructure development e.g. mini-Sentinel, EFPIA, etc.

**New insights** from RWD/E e.g. personalized health services, real-time treatment decisions (ASCO CancerLinQ).

Increasing transparency, scientific rigor and quality promotes **trust**. **Strategic partnership** overcomes barriers for data access.
Agenda for Today

1. What is Real World Data (RWD)? Why is it important?

2. Overview of Types of Real World Data Sources
   - Medical claims data, Electronic Medical Records (EMR) and Registries
   - Opportunities and challenges of integrating real world clinical and real world biomarker testing data

3. RWD Examples and Types of research questions that can be addressed

4. Role of RWD in drug and diagnostics development

5. The future of RWD: Challenges and Opportunities
In conclusion...

*Real world clinical and real world biomarker testing data, provides valuable insights across all stages of the drug and diagnostics development lifecycle.*
Doing now what patients need next
Roche RWDS leadership in public policy forums on use of RWD for Regulatory Decisions

**Workshop on Real World Evidence** (Jan 2016)

**Workshop on Enhancing the Application of Real World Evidence in Regulatory Decision Making** (March 2016)

**Forum: A Blue-Print for Breakthrough: Exploring Utility of Real World Evidence** (June 2016, Nov 2016)

**Multi-stakeholder Workshop on the use of RWD** (July 2016)

**Canadian Science Policy Conference: Unleashing Innovations in Personalized Healthcare** (Nov 2016)
Real-World Endpoints

Traditional surrogate endpoints need to be rethought in the context of real-world data. We need to deliver endpoints for patients in the real world that are:

- Based on existing data captured routinely from the chart in the real world
  - Take advantage of data presented in electronic health records (EHRs) whenever possible
- Tied to source evidence (e.g., radiographic, laboratory, pathologic, clinical assessment)
- Shown to be a meaningful endpoint based on a predefined experimental validation framework
- Accepted by oncologists, researchers, regulatory bodies, and industry, with guidance around suitable applications

Need to develop a solution that works across many different patients, clinicians, documentation habits, EHRs, health systems, and diseases settings
Example of Data Elements for Oncology Cohorts

- **Patient & clinical characteristics**
  - Demographics (e.g., age, gender, race/ethnicity)
  - Comorbidities, Performance status (e.g., ECOG)
  - Medical history, Family history
  - Smoking or alcohol use
  - Visit types and dates (e.g., office, ER)
  - Geographic location of healthcare provider/clinic

- **Tumor characteristics**
  - Cancer diagnosis and date (e.g., for initial diagnosis and for advanced disease)
  - Tumor stage, histology
  - Sites of metastases

- **Treatments and medications**
  - Drug name, drug class
  - Route of administration (IV, oral)
  - Start and end dates of therapy;
  - Date prescription filled, dose
  - Reason for discontinuation

- **Biomarkers**
  - Biomarker measured
  - Test/assay type (e.g., IHC, FISH, PCR)
  - Date of sample collection, date of test result
  - Test result (e.g., +/-; specific mutations detected)
  - Name of Lab performing the test
  - IVD or LDT
  - Test brand name

- **Outcome measures**
  - Vital status, date of death
  - Tumor progression and date
  - Specific adverse events of interest
## Analysis of FMI and biomarker testing in Flatiron EMR

*Reported results and types of questions that can be addressed by RWD*

<table>
<thead>
<tr>
<th>Type of Question</th>
<th>Question</th>
<th>Results presented in ASCO abstract</th>
</tr>
</thead>
</table>
| Testing patterns                  | • Did a patient get more than one test?  
• What other biomarkers were tested?  
• What were the other testing assays used?                                                                                          | *In the FMI group, 528 pts (52%) received at least one other MD test.*  
*EGFR (89%), ALK (83%), ROS1 (39%), KRAS (31%), PD-L1 (18%) and;*  
*test types were 67% FISH, 55% PCR, 17% other NGS and 15% IHC.*                                                                                           |
| Treatments received following testing | • What type of treatment did patients receive after getting an FMI test?                                                                                                                                     | *For 565 pts (of 1,019) with available data on first treatment after FMI testing, 24% (136 pts) initiated a NCCN recommended targeted treatment for LC.*                                                                                                                                                                                                                                                                                                                  |
Retrospective Cohort of ALK+ NSCLC patients using Flatiron EMR data: ISPOR 2016

Study population:
- Pts diagnosed with stage IIIB or IV NSCLC from Jan 1, 2011 to Dec 31, 2014
- With documented ALK rearrangement or translocation
- Pts enrolled in clinical trials were excluded
- Pts were followed until death, loss to follow-up or study end date (Feb 29, 2016)

95% of patients received ALK test results before or during first-line treatment.
Gaining biomarker insights from existing stage II colorectal cancer (CRC) patient cohorts in the U.K.: using real-world data to guide treatment decisions.

Sarah Jane Fleming, Eva Morris, Mike Shires, Gemma Hemmings, Lisa Wang, Andrea Muranyi, Shalini Singh, Margaret Elizabeth McCusker, Kandavel Shanmugam, Philip Quirke; University of Leeds, Leeds, United Kingdom; Genentech, San Francisco, CA; Ventana Medical Systems, Inc., Tucson, AZ; Roche Diagnostics International, Pleasanton, CA

Abstract Text:

Background: Trial data, while valuable, does not reflect the importance of a biomarker or treatment in the general population as trials generally exclude patients (pts) due to age or comorbidities. We used merged large population datasets to validate a linked anonymized stage II CRC pt cohort as representative of the population to identify the prognostic and predictive value of deficient mismatch repair (MMR) and a new biomarker Ga-interacting vesicle-associated protein (GIV). Methods: English national data on stage II CRC pts surgically treated from 2001-2015, n = 92,147, were analyzed for survival and clinicopathological parameters. Anonymized data were then linked to pathology and a subset of 405 unselected pts surgically treated from 1990-2003 at the Leeds Teaching Hospitals NHS Trust. These were investigated for 5 antibodies. 4 identified deficient MMR status and one was a new marker; GIV. Results: Population data vs the cohort of 405 pts showed a median age of 73 vs 74 yrs, M:F 55%:45% vs. 53%:47%. These are older than seen in most clinical trials and more reflective of the general stage II CRC population. The median survivals of 108 vs 92 months are as expected based on relative time periods covered. 374 patients yielded valid MMR status on immunohistochemistry (92%); 17% were dMMR and 83% were proficient. dMMR vs pMMR pts did not differ in age (median age 75 vs 74.5) or distribution of T3 (69% vs 66%), but were more likely to be female (71% vs 42%), sited in the right colon (76% vs 30%) and poorly differentiated (42% vs 8%). 11% of dMMR and 8% of pMMR received adjuvant chemotherapy in this cohort. GIV scoring status was also evaluated (405): 30% (121) were negative, 65% (264) positive and 5% (20) were not scored. Survival and risk scores by MMR and GIV status will be presented. Conclusions: Population based data has been created to validate the representativeness of a stage II biomarker cohort. This approach using large, population based data-sets provides opportunities to understand the generalizability of biomarker cohorts. The creation and expansion of such cohorts will more effectively validate existing and new biomarkers and treatments in real-world populations.
Characterizing biomarker defined breast cancer subgroups in the California Cancer Registry — Represents a real world population

Table 1 Demographic and clinical characteristics of female patients diagnosed with invasive breast cancer, by stage at diagnosis, California Cancer Registry, 2005–2011

<table>
<thead>
<tr>
<th></th>
<th>AJCC Stage I–III</th>
<th>De novo metastatic (AJCC Stage IV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of patients</td>
<td>118,817</td>
<td>6,268</td>
</tr>
</tbody>
</table>

**Conclusions** De novo metastatic breast cancer was more likely to be HER2+. Among metastatic tumors, those that were HER2+ had better survival than other subtypes.