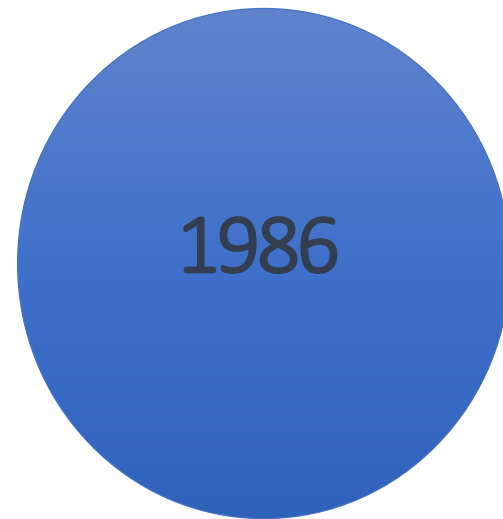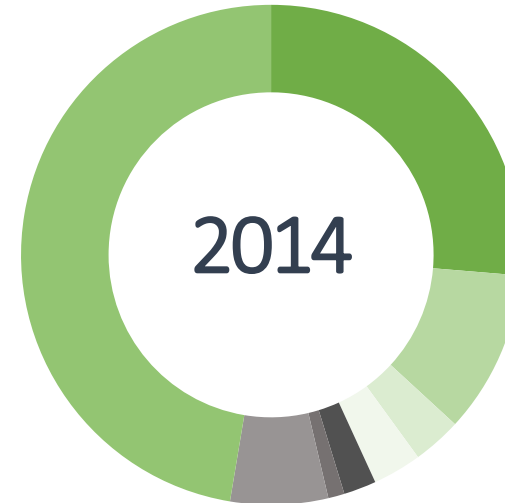# Using "Big Data" in Cancer to Improve Care and Accelerate Learning

Allen Lichter, MD

Former ASCO CEO

Board Member of CancerLinQ

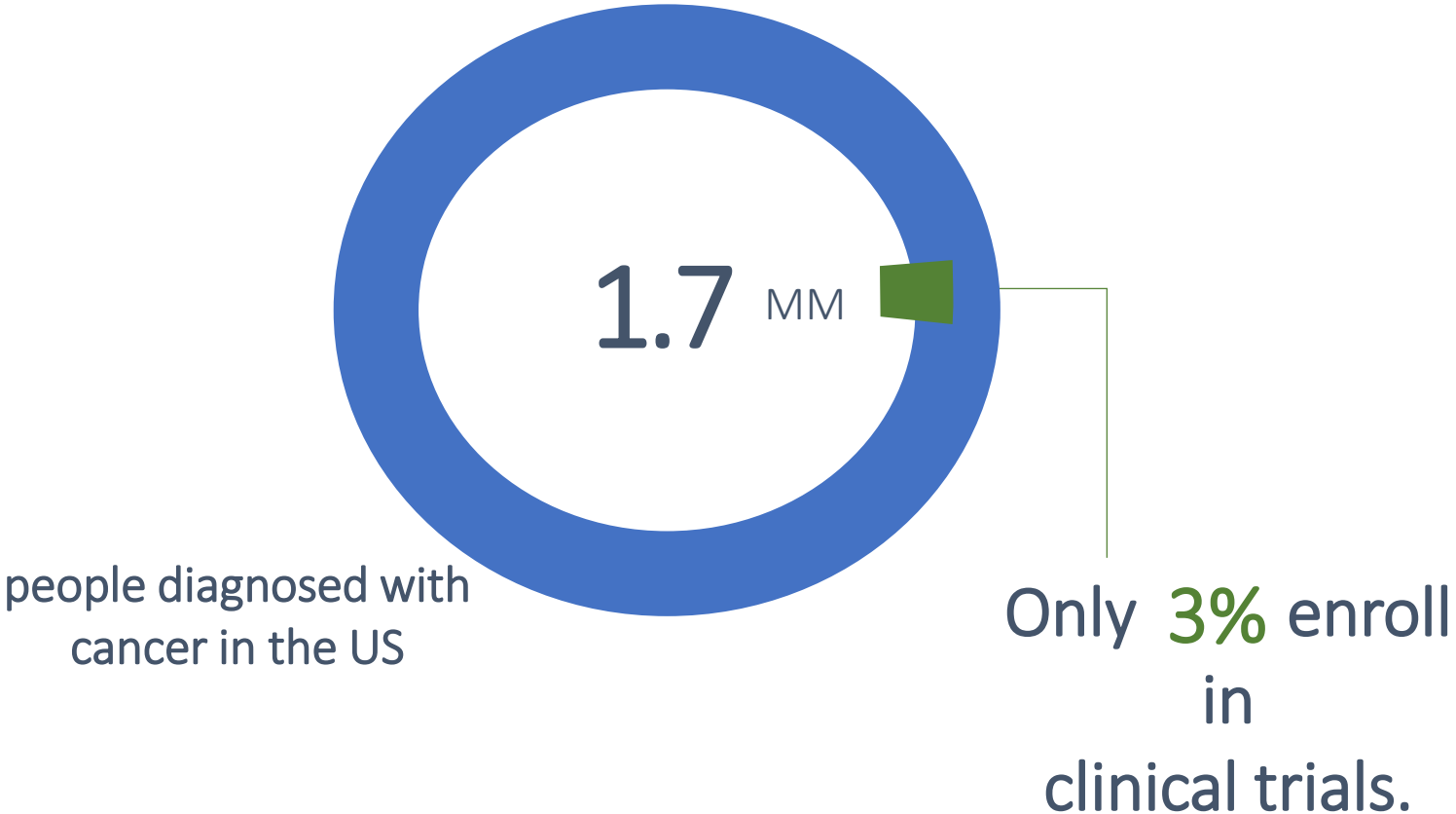# Lung cancer: from one cancer to many

1986
Three diseases

2014
Multiple molecular drivers—and more to be discovered

KRAS
EGFR
BRAF
PIK3CA
AKT1
HER2
EML4-ALK
Unknown

The Oncology Knowledge Base

1.7 MM

people diagnosed with cancer in the US

Only 3% enroll in clinical trials.

Current State of CLQ

- 1,600,000 individual patient records

- 73 participating institutions

- > 2000 Oncologists

- > 100 institutions in the pipeline

- **Onboarding one health system per week**

- Collaborations with major health systems
  - UVA Health System
  - Intermountain Healthcare
  - Vanderbilt University Medical Center
  - Case Western
  - Rush University Medical Center
  - Catholic Health Initiatives
 - Sanford Health

Current State of CLQ

Interfaces with a wide variety of EHR systems:

1. MosaiQ
2. Varian
3. Epic via Data Warehouse
4. Epic via CDA
5. I Know Med via CDA
6. IntrinsiQ
7. GE Centricity
8. All Scripts – Sunrise Clinical Manager (SCM)
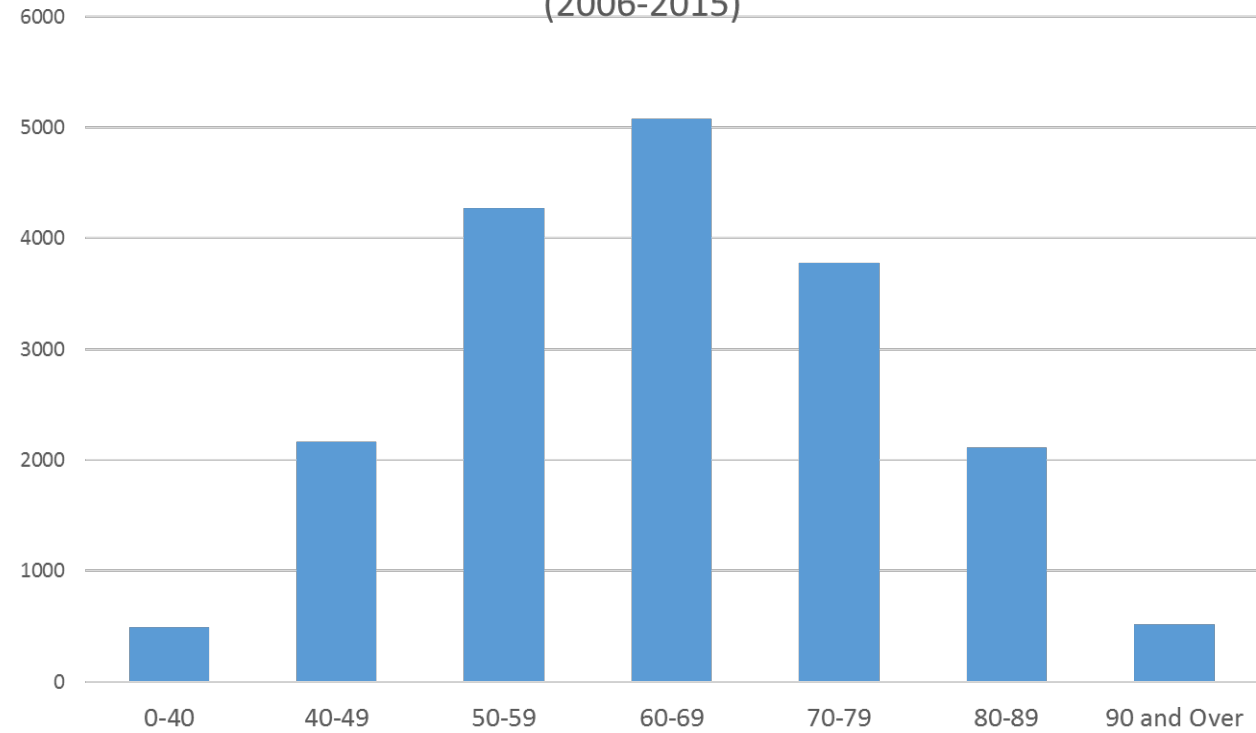9. OncoEMR

Interfaces in development:

Next Gen

MosaiQ Hosted

Varian Hosted

Cure MD

Proportion of Breast Cancer Patients by Gender (2006-2015)

Male, 1.2%

Female, 98.8%
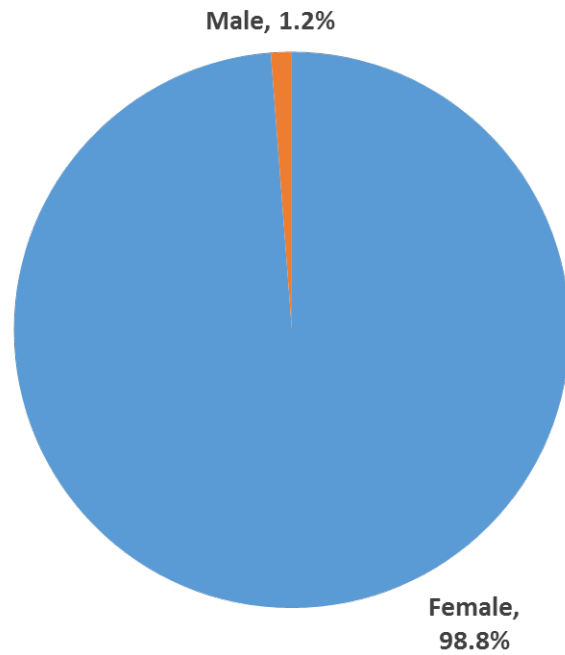
Distribution of Breast Cancer Patients by Age Group (2006-2015)
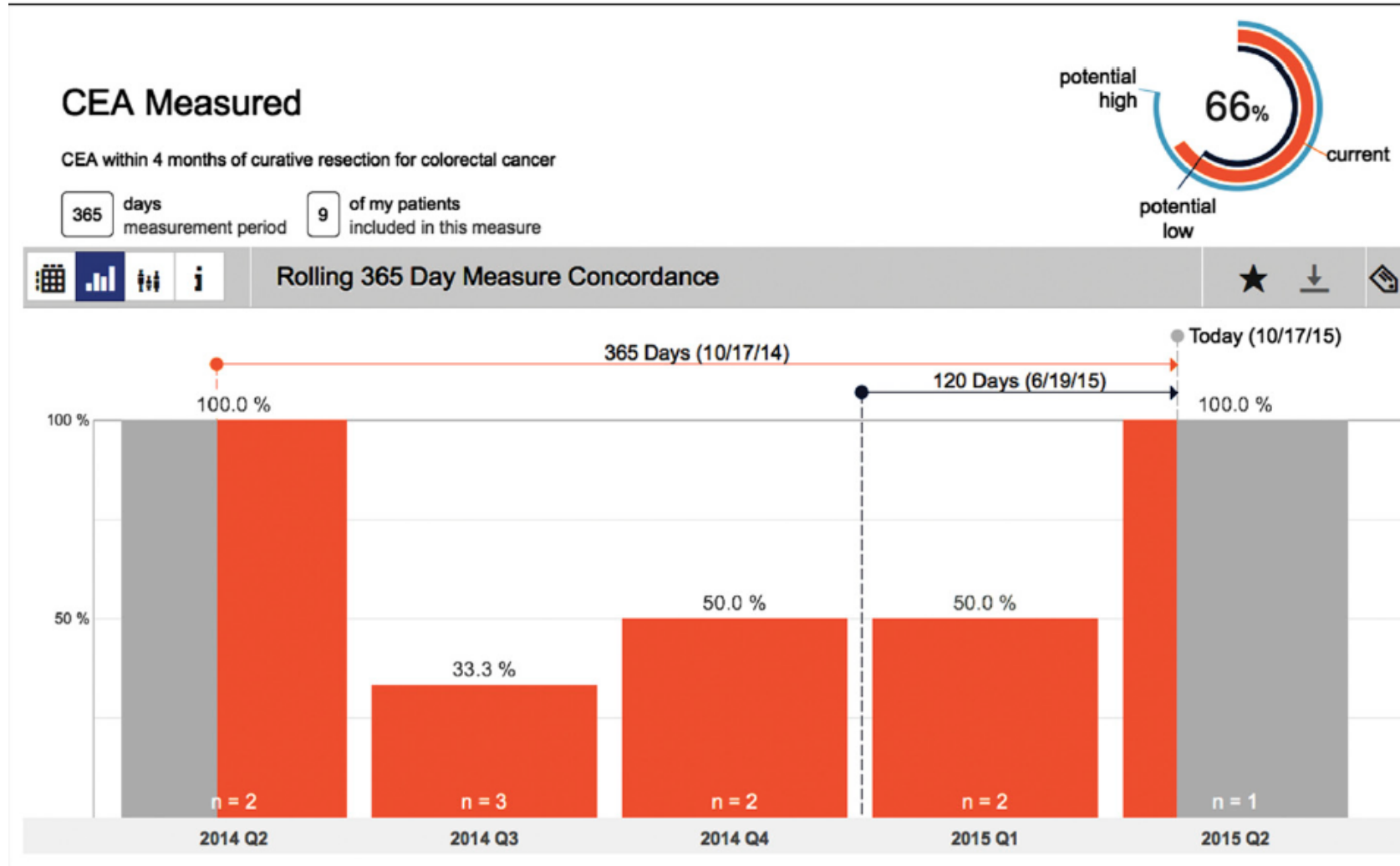
N = 18420 Patients

Data Quality

- Just over 1M reported patients have >800M reported clinical events (observations, lab, imaging, medication orders, medication administration, etc.)

- Yet there are numerous missing data elements
  - Example: Estrogen receptor status in breast cancer is reported ~100% when we use manual chart abstraction and barely 50% when we look for it as a structured data element

- There are key pieces of data sitting in unstructured data (notes, reports, PDFs, etc)

- Digging these data elements out is going to be a key challenge for us

CancerLinQ Collects PHI

- Resolve the identity of individual patients to create a longitudinal patient record.
- Develop clinically relevant quality analytics capable of determining which individual patients are "actionable".
- Collect unstructured data to capture insights from clinical notes, pathology reports, imaging studies, etc.
- Display individual patient records in a manner not currently available in current EHR systems.
- The PHI is then de-identified and can be used to accelerate our learning about what works and what doesn't.
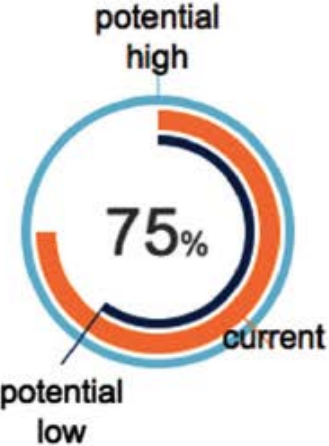
CLQ is First and Foremost a Quality Program
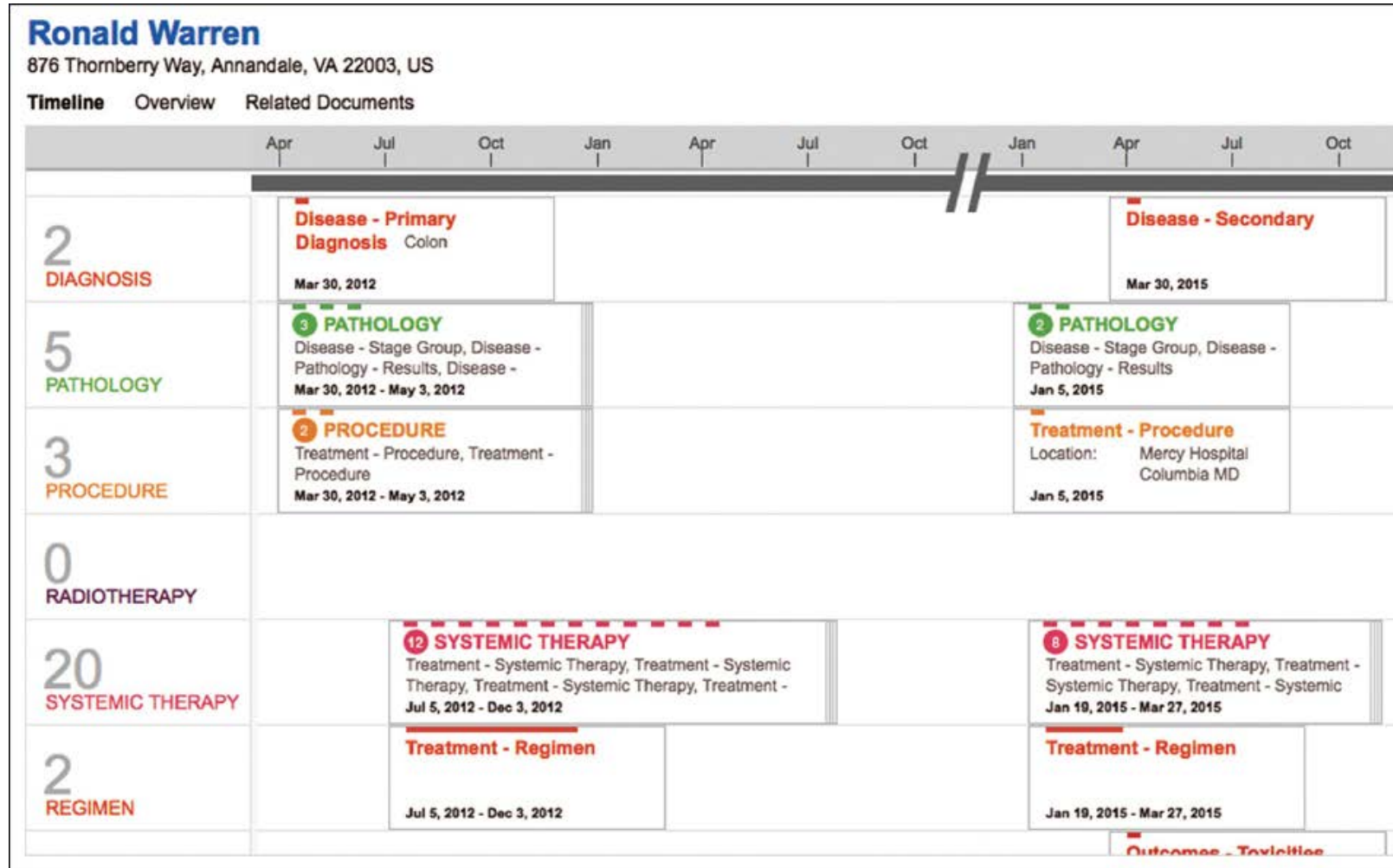
More Than a "Gotcha" Quality Program

Improving Care Through Data Presentation

Summary

- CancerLinQ is launched and running.
- Well over 1M cases have been accessioned into our system to date.
- The response of the physician community as well as the patient community has been very gratifying.
- The multiplicity of EHR systems and their lack of interoperability costs us lots of time and many dollars to overcome
- Data quality is an issue that we will wrestle with for a long time.
- We are excited to continue this journey.