Considerations for Developing External Control Arm from Real-World Data

Somnath Sarkar, PhD
Vice President, Head of Quantitative Sciences
Flatiron Health

BBS Spring Seminar
May 10, 2019
Roche Auditorium, Viaduktstrasse, Basel
Regulatory perspective

Flatiron’s real-world evidence generation platform

Real-world control
Regulatory perspective
Change in oncology drug development paradigm
Expanded Phase I or Single Arm Phase II often leads straight to Pivotal

“the desire to provide earlier access to highly effective drugs should encourage further use of seamless expansion cohort”

“greater attention to statistical rationale and analysis plan, more careful selection of drugs to be studied in this fashion”
Number of drug approvals in oncology has increased with availability of novel therapies

- In 2018 over 20 drugs were recommended for approval by EMA\(^1\)
- In the US ~50 new cancer drugs or combinations were approved in 2018, compared with 2 in 2005\(^2\)
  - FDA granted 25 Breakthrough Designations
  - Over last 25 years: 72% of AAs were from single-arm trials\(^3\)

---

3. JCO Editorial, Volume 36, June 20, 2018
5. CHOICE OF CONTROL GROUP AND RELATED ISSUES IN CLINICAL TRIALS E10
Recent work by EMA exploring the use of RWD in regulatory decisions

Creating a sustainable, quality assured, flexible framework delivering rapid access to and analysis of representative, longitudinal RWD throughout a product’s lifecycle

EMA Regulatory Science to 2025
Strategic reflection

E.g. “development of a framework to articulate for what questions and contexts RWE may be acceptable across the product life cycle”

“strongly support exploration of novel analytics approaches”

EMA released a discussion paper on methodological and operational considerations in the use of patient disease registries for regulatory purposes.

Discussion paper:
Use of patient disease registries for regulatory purposes – methodological and operational considerations

The Cross-Committee Task Force on Patient Registries
FDA has legislative mandate to **explore IF and WHEN RWE may support new indications (approved drugs)/post marketing requirements**

**Wide Spectrum of Potential Uses of RWD / RWE in Clinical Studies**

Different challenges and opportunities for each approach

<table>
<thead>
<tr>
<th>Randomized Intervenational</th>
<th>Intervenional non-randomized</th>
<th>Non-randomized / non-interventional</th>
</tr>
</thead>
<tbody>
<tr>
<td>RWD to assess enrollment criteria / trial feasibility</td>
<td>eCRF + selected outcomes identified using EHR / claims data</td>
<td>Mobile technology used to capture supportive endpoints (e.g., to assess ambulation)</td>
</tr>
<tr>
<td>RWD to support site selection</td>
<td>Uniform RCT (use of eCRF) +/- eHR data</td>
<td>Using existing databases</td>
</tr>
</tbody>
</table>

Source: Dr. Jacqueline Corrigan-Curay (Director Office of Medical Policy), FDA, “Framework for FDA's Real-World Evidence Program”, webinar on March 15, 2019
Considerations for Generating RWE Fit for Regulatory Purposes: Duke Margolis White Paper

**Regulatory Context**
What specific decision is FDA considering?
- New indication
- Labeling revision
- Safety revision
- Benefit-risk profile

**Clinical Context**
Can the clinical question be reliably addressed with RWE?
- Prevalence of the disease
- Clinical equipoise
- Expected treatment effect size

**Data Considerations**
Are the real-world data sources of sufficient quality?
- Minimal missing data
- Sufficient data reliability and validity
- Established data quality assurance procedures

**Methods Considerations**
Are the methodological approaches of sufficient rigor?
- Interventional or observational
- Prospective, retrospective, or hybrid
- Appropriate analytic approach
- Established credibility (protocol developed and replication of results achieved/planned)

Considerations for Generating RWE Fit for Regulatory Purposes: Duke Margolis White Paper

Regulatory Context
What specific decision is FDA considering?
- New indication

Example of Real-world Control:
- Rare diseases
- New indication
- Expected large treatment benefit

Data Considerations
Are the real-world data sources of sufficient quality?
- Minimal missing data
- Sufficient data reliability and validity
- Established data quality assurance procedures

Methods Considerations
Are the methodological approaches of sufficient rigor?
- Interventional or observational
- Prospective, retrospective, or hybrid
- Appropriate analytic approach
- Established credibility (protocol developed and replication of results achieved/planned)

Fit-for-purpose RWE
Flatiron’s real-world evidence generation platform
Rapid adoption of EHRs in oncology

President Obama to Sign ARRA’s HITECH provisions Tuesday, February 17, 2009, in Denver, CO

The Senate joined the House on Friday evening, February 13, 2009, in passing the American Recovery and Reinvestment Act, which includes provisions relating to Health Information Technology. Title XIII of Division A and Title IV of Division B together are known as the “Health Information Technology for Economic and Clinical Health Act” or the “HITECH Act.” We will be highlighting attributes of the HITECH Act of February 17, 2009, in summarizing this legislation.

Adoption of EHRs in Oncology clinics go from ~10% → 95%

Recent adoption in US parallels UK ambition to shift to a “paperless NHS”
“RWE is derived from RWD through the application of research methods”
Data source and curation

Structured

UNSTRUCTURED

Hospital

Reports

EHR

© Flatiron Health 2019
Data source and curation

EHR

- Diagnosis
- Visits
- Demographics
- Labs
- Therapies
- Pathology
- Discharge Notes
- Physician Notes
- Radiology Report

Hospital

Reports
Data source and curation

EHR

- Diagnosis
- Visits
- Demographics
- Labs
- Therapies
- Pathology
- Discharge Notes
- Physician Notes
- Radiology Report

Structured Data Processing

Unstructured Data Processing

RWE Database

Hospital
Reports

Data Linkage

© Flatiron Health 2019
Diagnosed with Stage II NSCLC
Undergoes surgery
Receives adjuvant therapy
Progresses on adjuvant therapy
Progresses on 1L
Starts 1L therapy
Starts on 2L
Patient deteriorates leading to hospitalization / death

**Relative timing not exact**
Diagnosed with Stage II NSCLC

Undergoes surgery

Receives adjuvant therapy

Progresses on adjuvant therapy

Progresses on 1L

Patient deteriorates leading to hospitalization / death

Starts 1L therapy

Starts on 2L

Patient: Jane Doe

Stage at Dx: II

Biomarkers: EGFR-, ALK-, PD-L1-

2L Treatment: nivolumab

Progression: 2017-03-08

Date of Death: 2017-04-12

*Relative timing not exact
Data quality across three dimensions

**Accuracy**
- Validity of data elements
- Logical plausibility of results
- Data consistency for a given patient

**Completeness**
- Extent of missingness in data
- Possible root cause and impact of missing data

**Traceability**
- Transparency in data provenance and transformation
- Defined business logic for key variables

**Key metrics:**
- Inter-abstraction agreement (proxy for accuracy) for derived variables from unstructured data
- Data completeness
- Missing data and impact on analysis?
- Provenance
- Variable versioning

*Harnessing the power of RWE: A checklist to ensure regulatory grade data quality*, Clinical Pharmacology and Therapeutics, 2017; Rebecca Miksad, and Amy Abernathy

Dataset linkage → Composite endpoint
Evaluate underlying data quality (gold standard = NDI)

**Definition of rwP:** All distinct episodes in which the treating clinician concludes that there has been overall growth or worsening of the disease of interest.

**NSCLC Patient Example**

- Advanced NSCLC diagnosis
- Started 1L carboplatin / pemetrexed

- Imaging showed progression; started docetaxel

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td></td>
</tr>
</tbody>
</table>

- Clinically anchored with radiology and pathology reports serving as corroborative evidence
- Most practical and scalable
Validation: Patient-level correlation between rwPFS and OS

Methods:
- Real-world time to progression or death was calculated and plotted against time to death for each patient
- Patients without a date of death were excluded from this analysis
- Correlation was calculated using Spearman’s rank correlation coefficient

<table>
<thead>
<tr>
<th>Correlation</th>
<th>N</th>
<th>ρ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rwPFS vs OS</td>
<td>20,020</td>
<td>0.76 (0.75, 0.77)</td>
</tr>
<tr>
<td>rwTTP vs OS</td>
<td>11,902</td>
<td>0.69 (0.68, 0.70)</td>
</tr>
<tr>
<td>rwTTNT vs OS</td>
<td>9,269</td>
<td>0.61 (0.60, 0.62)</td>
</tr>
</tbody>
</table>
Foundation for rwCA Development

- **Contemporaneous Cohort Selection**
  - Generalizability

- **Biomarkers Clinical Depth**
  - Derived Variables

- **Mortality**
  - PFS Endpoints
  - Response

- **Analytic Methods**
  - Patient-Level Matching Prespecification
  - Propensity Scores

- **Data Quality**
  - Completeness
  - Recency
  - Provenance
Validation through Replication:

Can we replicate the outcomes observed in the control arms of recent clinical trials using Flatiron’s real-world data?

Bennette C et al. Use of a curated electronic health records database to create external control arms for cancer clinical trials. In submission.
Constructing real-world control arms

Randomized trials supporting FDA approvals for anticancer drugs/biologics from 1/1/2016 - 4/30/2018
\[N=49\]

Control arm was not placebo alone (or “observation”)
\[N=36\]

Approval in tumor type represented by an existing data model from Flatiron Health EHR database with sufficient data
\[N=15\]

Approvals included initial, supplemental, accelerated & regular approval following accelerated approval
Constructing real-world control arms

**Randomized** trials supporting FDA approvals for anticancer drugs/biologics from 1/1/2016 - 4/30/2018

N=49

- Control arm was not placebo alone (or “observation”)  
  N=36

- Approval in tumor type represented by an existing data model from Flatiron Health EHR database with sufficient data  
  N=15

**Identify** real-world patients with treatment and molecular features consistent with trial’s control arm

**Align** real-world patient population with trial’s inclusion and exclusion criteria

If imbalances wrt key baseline characteristics, **weight** eligible real-world patients to match published baseline characteristics

*Approvals included initial, supplemental, accelerated & regular approval following accelerated approval*
<table>
<thead>
<tr>
<th>Tumor type</th>
<th>Name of trial</th>
<th>Front line setting?</th>
<th>Primary endpoint(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>melanoma</td>
<td>CheckMate-067</td>
<td>yes</td>
<td>PFS, OS</td>
</tr>
<tr>
<td>kidney</td>
<td>METEOR</td>
<td>no</td>
<td>PFS</td>
</tr>
<tr>
<td>kidney</td>
<td>NCT01136733</td>
<td>no</td>
<td>PFS</td>
</tr>
<tr>
<td>NSCLC</td>
<td>OAK</td>
<td>no</td>
<td>OS</td>
</tr>
<tr>
<td>NSCLC</td>
<td>POPLAR</td>
<td>no</td>
<td>OS</td>
</tr>
<tr>
<td>breast</td>
<td>MONARCH-3</td>
<td>yes</td>
<td>PFS</td>
</tr>
<tr>
<td>breast</td>
<td>MONALEESA-2</td>
<td>yes</td>
<td>PFS</td>
</tr>
<tr>
<td>myeloma</td>
<td>POLLUX</td>
<td>no</td>
<td>PFS</td>
</tr>
<tr>
<td>NSCLC</td>
<td>KEYNOTE-024</td>
<td>yes</td>
<td>PFS</td>
</tr>
<tr>
<td>head and neck</td>
<td>CheckMate-141</td>
<td>no</td>
<td>OS</td>
</tr>
<tr>
<td>myeloma</td>
<td>CASTOR</td>
<td>no</td>
<td>PFS</td>
</tr>
<tr>
<td>NSCLC</td>
<td>AURA3</td>
<td>no</td>
<td>PFS</td>
</tr>
<tr>
<td>breast</td>
<td>PALOMA-2</td>
<td>yes</td>
<td>PFS</td>
</tr>
<tr>
<td>NSCLC</td>
<td>KEYNOTE-021</td>
<td>yes</td>
<td>ORR</td>
</tr>
<tr>
<td>urothelial</td>
<td>KEYNOTE-045</td>
<td>no</td>
<td>PFS, OS</td>
</tr>
<tr>
<td>NSCLC</td>
<td>ALEX</td>
<td>yes</td>
<td>PFS</td>
</tr>
<tr>
<td>kidney</td>
<td>CABOSUN</td>
<td>yes</td>
<td>PFS</td>
</tr>
<tr>
<td>breast</td>
<td>OlympiAD</td>
<td>no</td>
<td>PFS</td>
</tr>
<tr>
<td>kidney</td>
<td>CheckMate-214</td>
<td>yes</td>
<td>ORR, PFS, OS</td>
</tr>
<tr>
<td>breast</td>
<td>MONARCH-2</td>
<td>no</td>
<td>PFS</td>
</tr>
<tr>
<td>breast</td>
<td>PALOMA-3</td>
<td>no</td>
<td>PFS</td>
</tr>
</tbody>
</table>
Weighting real-world patients to published trials

- Derive “inverse odds” weights \( w_i = \frac{\Pr(C_i=0 \mid x_i)}{\Pr(C_i=1 \mid x_i)} \) that represent odds patient was in the trial \( (C_i=1) \) vs the real-world cohort \( (C_i=0) \) given baseline characteristics \( (x_i) \)

- Approach is analogous to common method of calculating propensity score weights, except we
  - Use inverse odds rather than inverse probability so that we standardize to patients in the trial (and resulting treatment effect can be interpreted in much the same way it would from a randomized trial)
  - Use generalized method of moments rather than maximum likelihood to estimate logistic regression model because we have only summary data for trial

Cohort Selection and data completeness report

Cohort Selection

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>41144</td>
<td>Step 1a: Locally advanced or metastatic NSCLC who received platinum-based therapy in 1st or 2nd line</td>
</tr>
<tr>
<td>22036</td>
<td>Step 1b: Received platinum-based therapy in 1st or 2nd line</td>
</tr>
<tr>
<td>1136</td>
<td>Step 2a: Docetaxel after platinum-based therapy</td>
</tr>
<tr>
<td>377</td>
<td>Step 2b: Docetaxel received before trial enrollment ended</td>
</tr>
<tr>
<td>364</td>
<td>Step 3: Disease progression during or following prior platinum-based therapy</td>
</tr>
<tr>
<td>357</td>
<td>Step 4: No prior docetaxel, anti-CTLA-4, or PD-L1/PD-1 inhibitor</td>
</tr>
<tr>
<td>322</td>
<td>Step 5: Exclude patients with ECOG PS 2+</td>
</tr>
<tr>
<td>191</td>
<td>Step 6: Exclude patients with inadequate organ function (per protocol)</td>
</tr>
</tbody>
</table>

Completeness summary

<table>
<thead>
<tr>
<th>Data element</th>
<th>% Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutrophils or granulocytes</td>
<td>91.3%</td>
</tr>
<tr>
<td>Creatinine</td>
<td>99.5%</td>
</tr>
<tr>
<td>Platelets</td>
<td>98.3%</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>98.3%</td>
</tr>
<tr>
<td>Albumin</td>
<td>90.7%</td>
</tr>
<tr>
<td>Lymphocytes</td>
<td>96.2%</td>
</tr>
<tr>
<td>White blood cells</td>
<td>98.9%</td>
</tr>
<tr>
<td>ALT</td>
<td>96.7%</td>
</tr>
<tr>
<td>AST</td>
<td>96.7%</td>
</tr>
<tr>
<td>Calcium</td>
<td>95.1%</td>
</tr>
<tr>
<td>Bilirubin</td>
<td>91.8%</td>
</tr>
<tr>
<td>ECOG</td>
<td>36.6%</td>
</tr>
</tbody>
</table>
### Baseline patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>RWD I/E aligned</th>
<th>Weighted</th>
<th>Trial Control</th>
<th>Trial Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>% &lt;Med Age</td>
<td>35</td>
<td>50</td>
<td>51</td>
<td>50</td>
</tr>
<tr>
<td>% Male</td>
<td>56</td>
<td>61</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>% White</td>
<td>71</td>
<td>71</td>
<td>70</td>
<td>71</td>
</tr>
<tr>
<td>% Squamous</td>
<td>12</td>
<td>11</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>% EGFR+</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>% KRAS +</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>% 1 Prior Line</td>
<td>52</td>
<td>54</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>% Smoking History</td>
<td>92</td>
<td>80</td>
<td>83</td>
<td>80</td>
</tr>
</tbody>
</table>
Overall survival (Primary endpoint)

OAK comparison

Naive real-world control

Alignment with I/E criteria

Weighted

HR = 0.79 (0.67, 0.93)

HR = 0.92 (0.76, 1.13)

HR = 0.99 (0.80, 1.22)
## Progression-free survival

(secondary endpoint)

<table>
<thead>
<tr>
<th>Naive real-world control</th>
<th>Alignment with I/E criteria</th>
<th>Weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HR: 0.82 (95% CI: 0.71, 0.95)</strong></td>
<td><strong>HR: 0.88 (95% CI: 0.74, 1.05)</strong></td>
<td><strong>HR: 0.92 (95% CI: 0.76, 1.10)</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number at risk</th>
<th>364</th>
<th>191</th>
<th>169</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flatiron</td>
<td>149</td>
<td>89</td>
<td>78</td>
</tr>
<tr>
<td>Trial</td>
<td>218</td>
<td>44</td>
<td>41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number at risk</th>
<th>35</th>
<th>23</th>
<th>26</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flatiron</td>
<td>70</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>Trial</td>
<td>30</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number at risk</th>
<th>22</th>
<th>14</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flatiron</td>
<td>9</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Trial</td>
<td>7</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number at risk</th>
<th>7</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flatiron</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Trial</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
## Planned sensitivity analyses

Comparing patients in original trial’s control arm versus:

<table>
<thead>
<tr>
<th>Trial</th>
<th>Main analyses (weighted) HR (95% CI)</th>
<th>Excluding patients with missing ECOG performance status HR (95% CI)</th>
<th>Excluding patients with missing laboratory results used to define organ function HR (95% CI)</th>
<th>Excluding patients treated before trial enrollment period started HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAK (OS)</td>
<td>0.99 (0.80, 1.22)</td>
<td>1.21 (0.86, 1.70)</td>
<td>1.04 (0.82, 1.33)</td>
<td>1.05 (0.77, 1.44)</td>
</tr>
<tr>
<td>OAK (PFS)</td>
<td>0.92 (0.76, 1.10)</td>
<td>0.99 (0.75, 1.31)</td>
<td>0.94 (0.77, 1.16)</td>
<td>0.98 (0.75, 1.28)</td>
</tr>
</tbody>
</table>
Diagnostics

Generalizability
(Distribution of sites)

- Total 59 sites
- Both academic and community are well represented
Diagnostics

Trial recruitment window
(51.5% of Flatiron cohort)

Follow-up Comparisons

- Concurrent patient cohort
- Possible due to large sample size
- HR=1.01 if cohort is restricted

Median follow-up in the Flatiron cohort is 46.2 months.
OS Combined HR (95% CI):
0.98 (0.89 - 1.09)

- Overall consistent finding for OS
PFS Combined HR (95% CI): 0.94 (0.86 - 1.04)

- Similarly consistent finding for PFS
Three Key Steps to construct rwCA from curated EHR databases

- **Cohort selection**
  - Aligning with a trial’s eligibility criteria and timing of enrollment
    - Requires transparent application of trial eligibility criteria to real-world patients

- **Weighting or Matching**
  - Matching or weighting real-world patients to trial patients
    - Predefined approach to improve balance between non-randomized groups

- **Comparing Outcomes**
  - Comparing real-world and clinical endpoints
    - Validation of real world endpoints
    - Diagnostics and sensitivity analyses
Comparability of enrollment timeframe and follow-up

- Ideally RW patients treated during the same timeframe of a trial enrollment
- Broader windows may be chosen in absence of no substantial change in SoC and rarer patient population

Documenting which eligibility criteria were implemented

- Identify patients who are similar in prognosis, noting features driving prognosis are different even within a disease setting
- Many of I/E criteria are more difficult to implement (some are feasible with addition abstraction and/or proxies)
- Clinical importance of infeasible eligibility criteria depends on the context
Selection of covariates
- Systematic literature review to identify key prognostic factors that are measured in both datasets

Handling potential missingness in RWD
- e.g. for non-routinely performed lab tests it may be reasonable to assume: absence of a test as absence of the underlying condition (e.g. viral hepatitis tests)
- Attrition diagram and sensitivity analyses

Are the comparison groups balanced on known baseline characteristics?
- Planned sensitivity analysis showing consistency
Validity of real-world endpoints and analysis

- Systematic differences in how the index date was defined may result in biased results
- Ensure that outcome assessments are occurring at reasonable intervals and can be captured reliably
- Evaluate the timing of follow up assessments & censoring patterns to compare to the clinical trial endpoint

“Threshold crossing” framework,

- Anticipated benefit is robust and efficacy threshold is specified a priori
Conclusion

- Oncology drug development/regulatory paradigm continues to change
  - Rapidly changing standard of care
  - Blurring of retrospective & prospective research (recency of RWD)

- EHR data have great potential to provide research/regulatory-grade evidence
  - Important to demonstrate quality, validity and analytical considerations of RWE
Thank you