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

flatiron

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"

Phase 1a/1b/single arm Phase 2 → **Phase 3**

Potential for submission

Confirmatory Studies

Learning ⟷ **Confirming**

Regulatory Guidance recognized option of using external historical controls[4, 5]

- Disease is rare and has no satisfactory treatment
- New treatment appears very promising based on preliminary data

flatiron

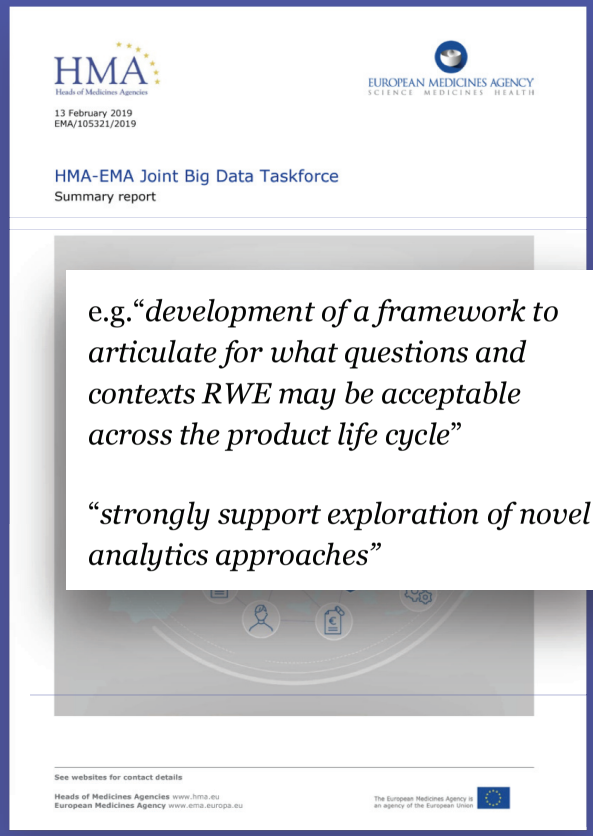
Number of drug approvals in oncology has increased with availability of novel therapies

- 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]

[1] https://www.ema.europa.eu/en/documents/report/human-medicines-highlights-2018_en.pdf
[2] Blumenthal GM, Pazdur R. Approvals in 2018: a histologyagnostic new molecular entity, novel end points and real-time review. Nat Rev Clin Oncol. 2019 Mar;16(3):139-141
[3] JCO Editorial, Volume 36, June 20, 2018
[4] Blood Cancer Journal (2016) 6, e473; doi:10.1038/bcj.2016.84
[5] CHOICE OF CONTROL GROUP AND RELATED ISSUES IN CLINICAL TRIALS E10

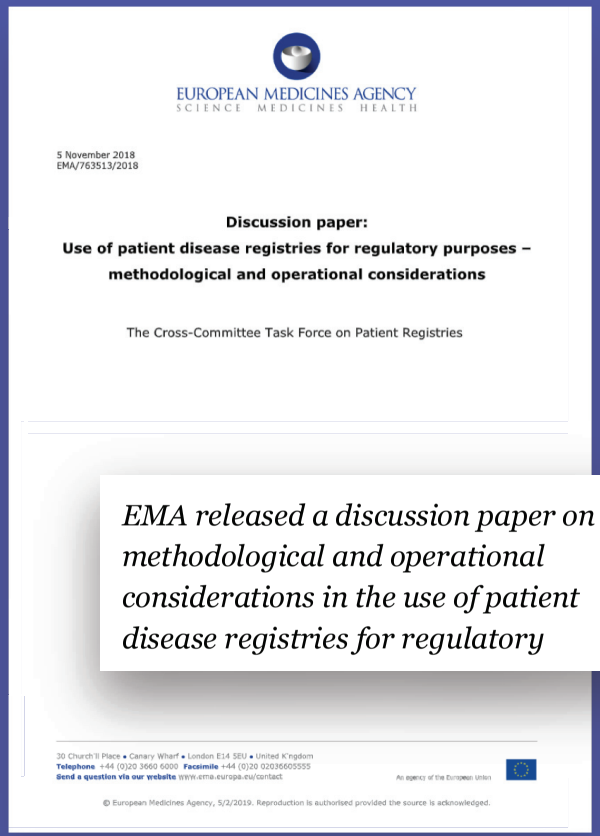
Recent work by EMA exploring the use of RWD in regulatory decisions

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



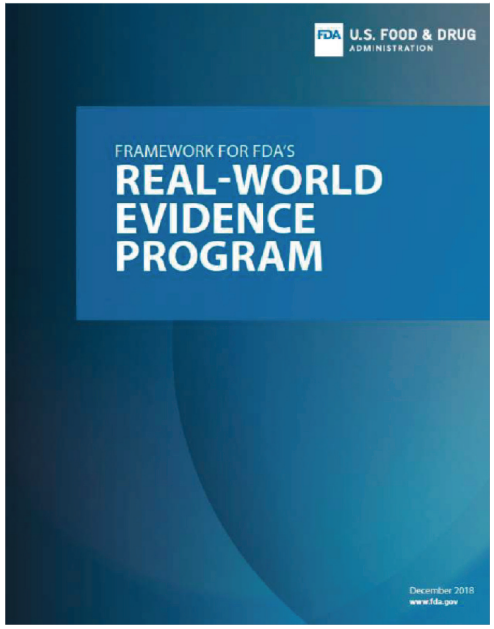
e.g."*development of a framework to articulate for what questions and contexts RWE may be acceptable across the product life cycle*"

"*strongly support exploration of novel analytics approaches*"



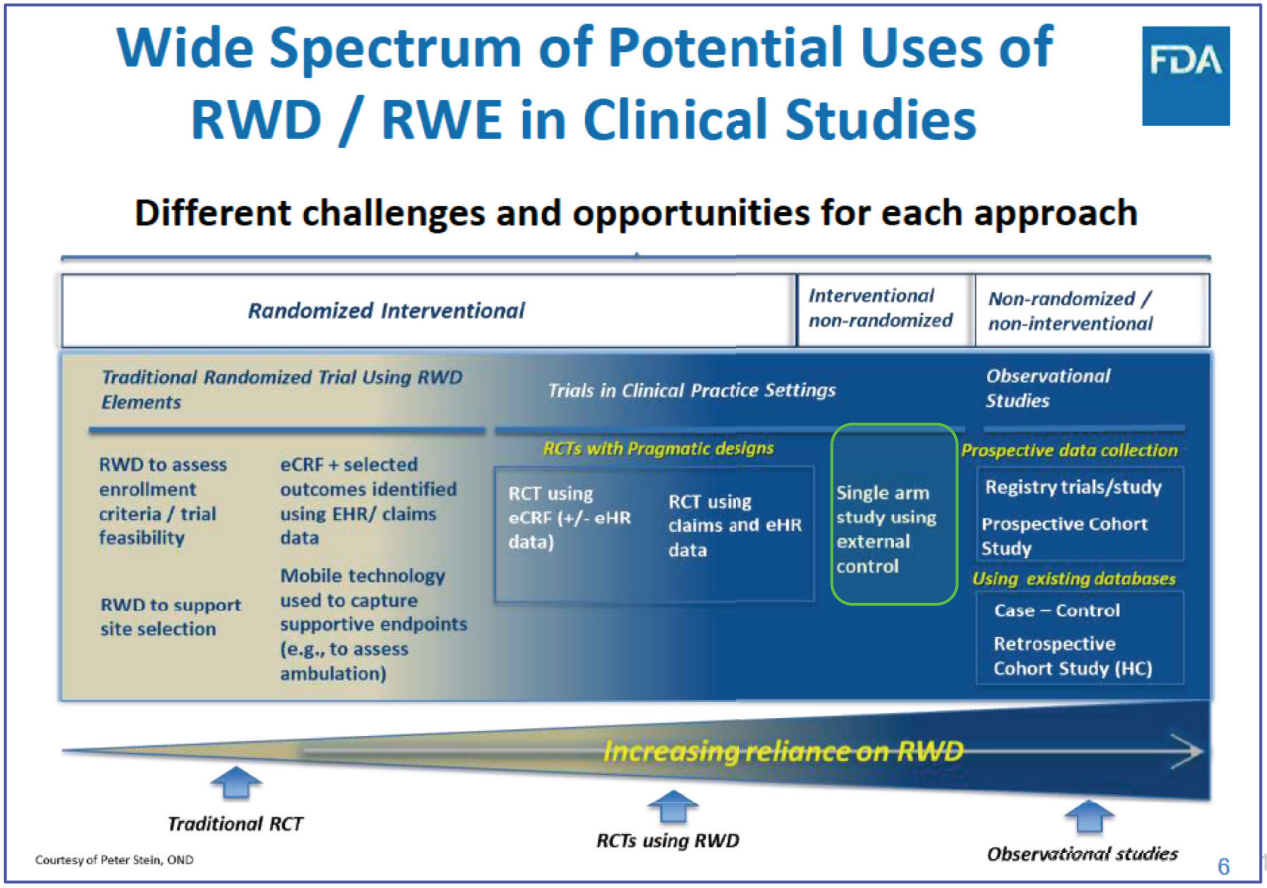
EMA released a discussion paper on methodological and operational considerations in the use of patient disease registries for regulatory

FDA has legislative mandate to explore IF and WHEN RWE may support new indications (approved drugs)/post marketing requirements



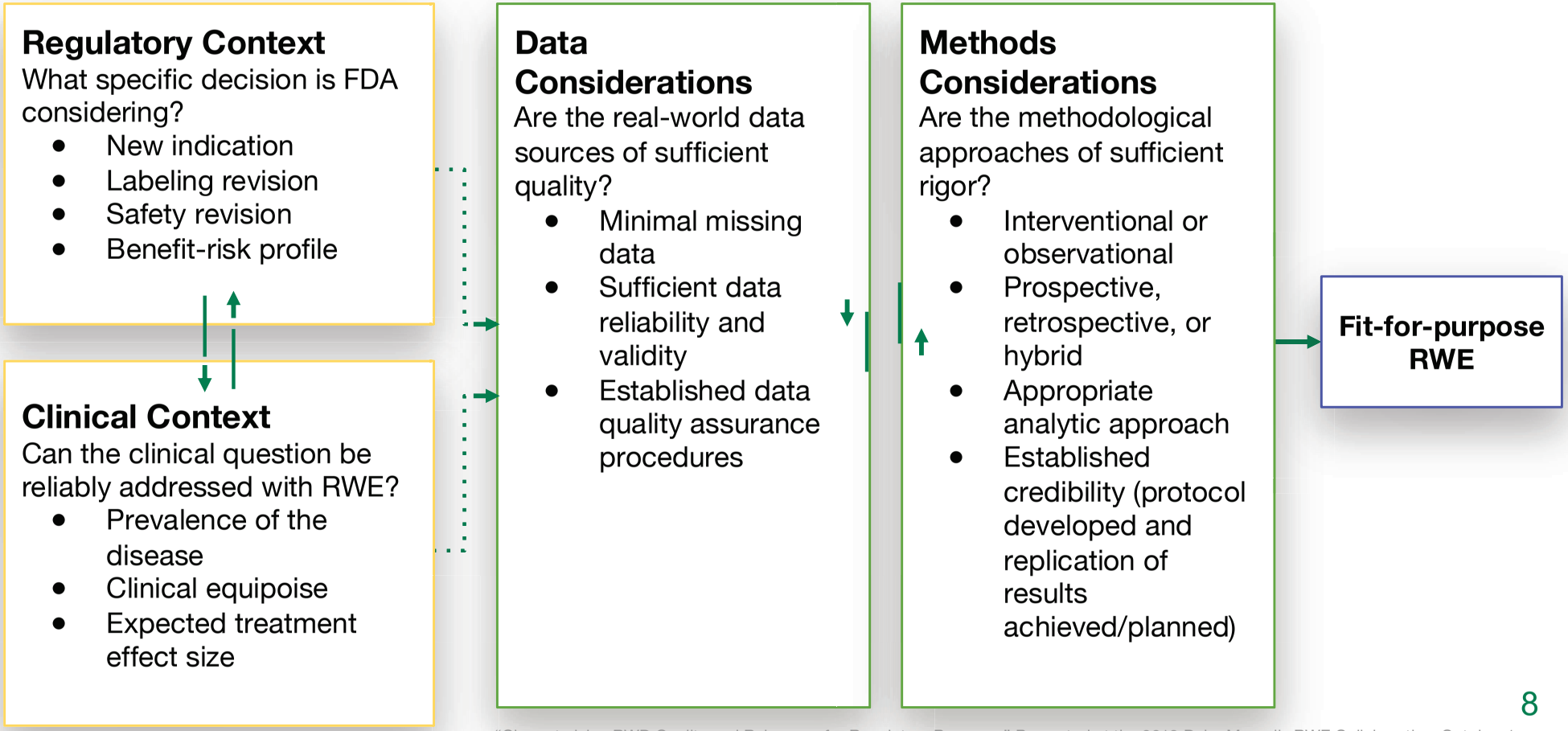
Framework

https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RealWorldEvidence/UCM627769.pdf



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



"Characterizing RWD Quality and Relevancy for Regulatory Purposes" Presented at the 2018 Duke Margolis RWE Collaborative. October 1, 2018. https://healthpolicy.duke.edu/sites/default/files/atoms/files/characterizing_rwd.pdf

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

Regulatory Context
What specific decision is FDA considering?

- New indication
- Prevalence of the disease
- Clinical equipoise
- Expected treatment effect size

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

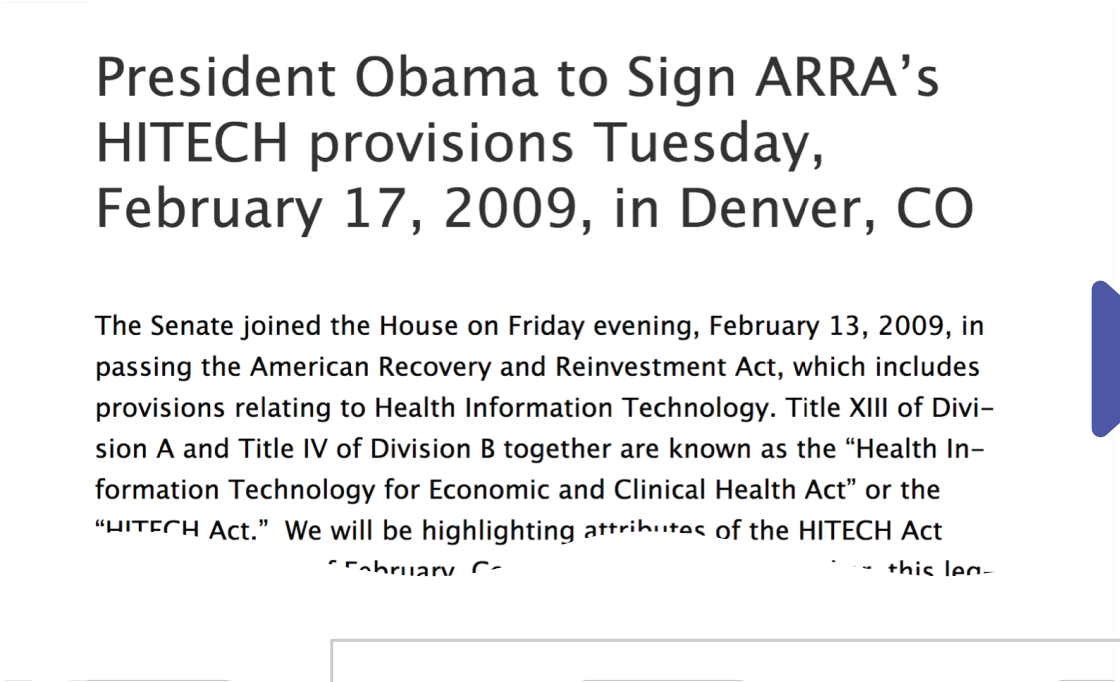
Methods Considerations
Are the methodological approaches of sufficient rigor?

- 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



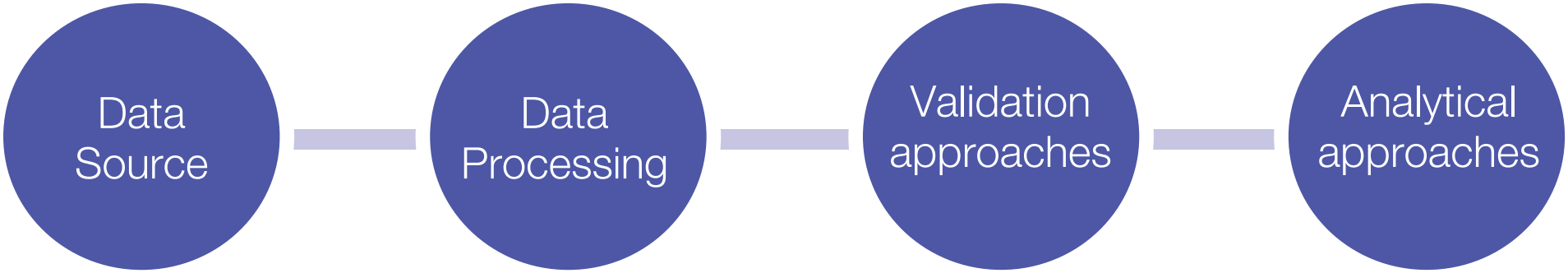
President Obama to Sign ARRA's HITECH provisions Tuesday, February 17, 2009, in Denver, CO

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

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

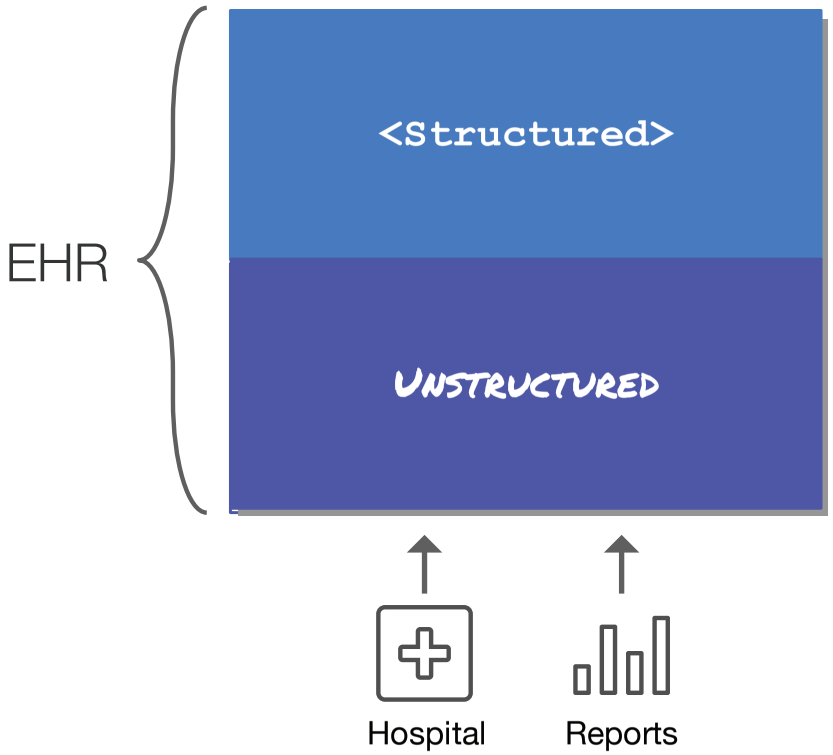
Recent adoption in US parallels UK ambition to shift to a "paperless NHS"

….Journey to Research/Regulatory Grade Real-World Evidence

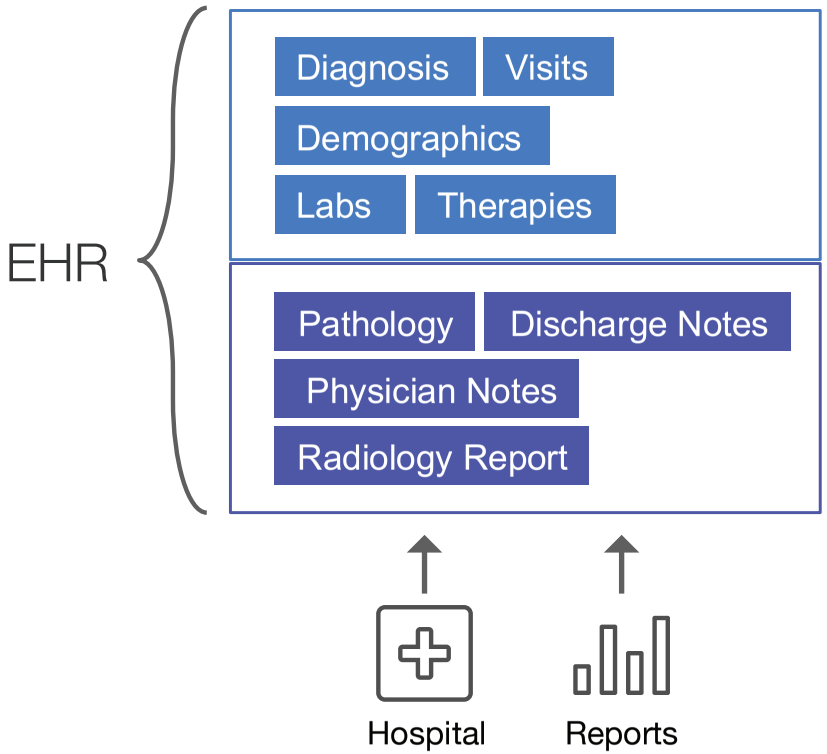


"RWE is derived from RWD through the application of research methods"

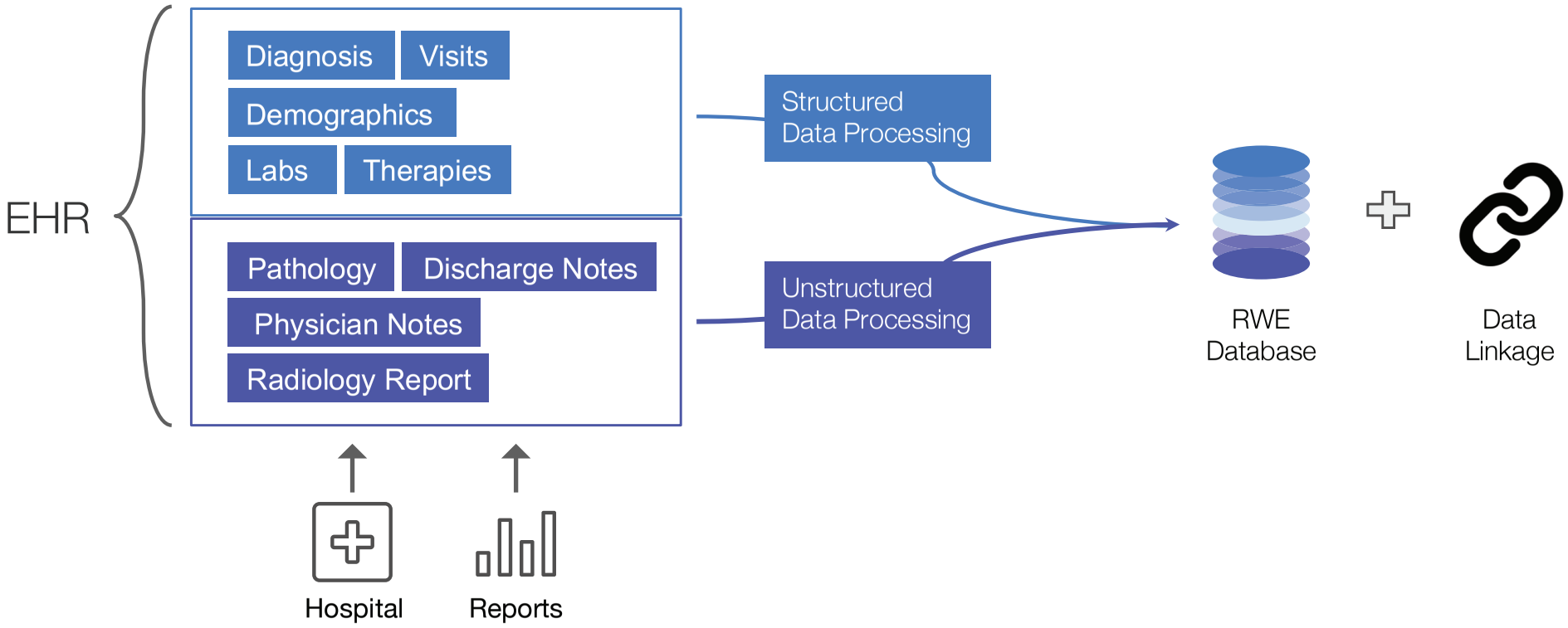
Data source and curation



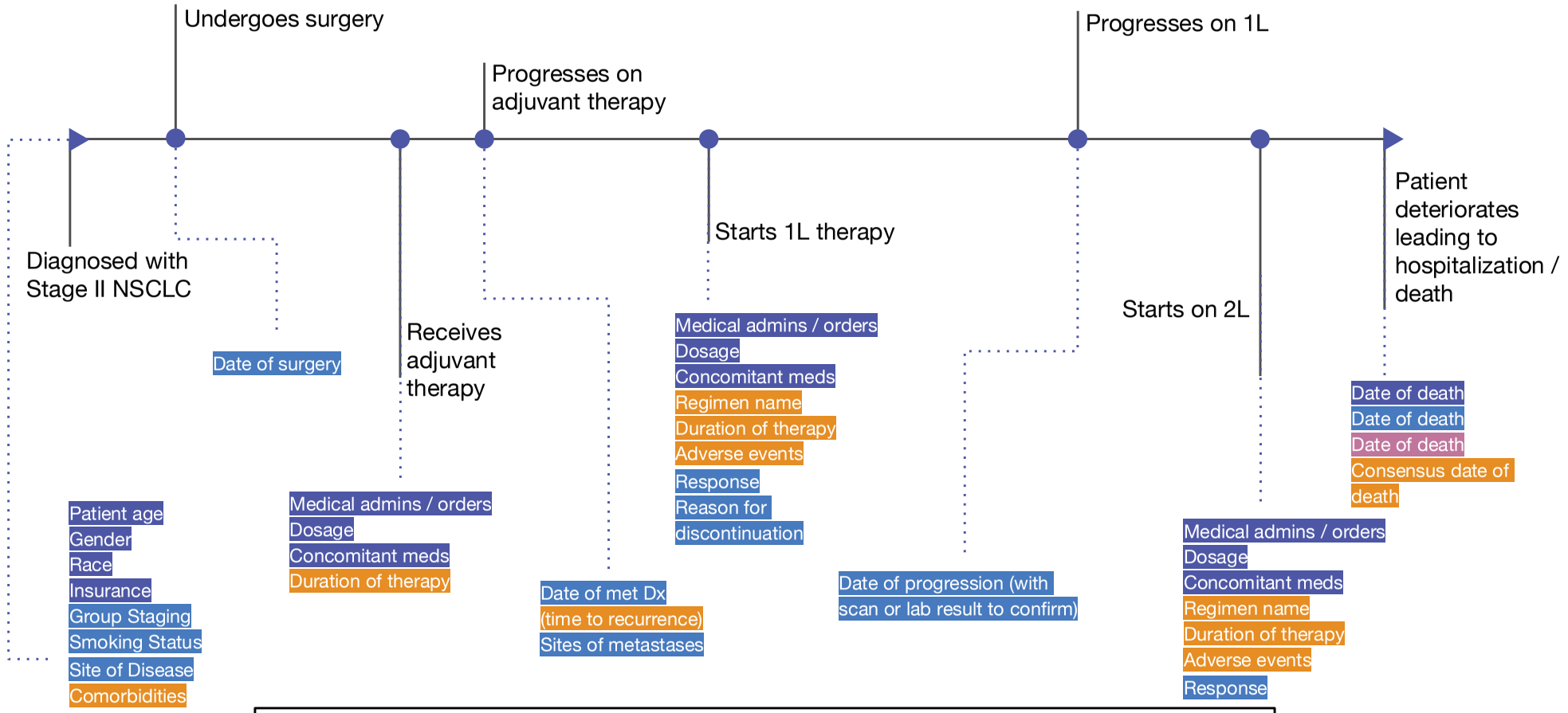
Data source and curation



Data source and curation

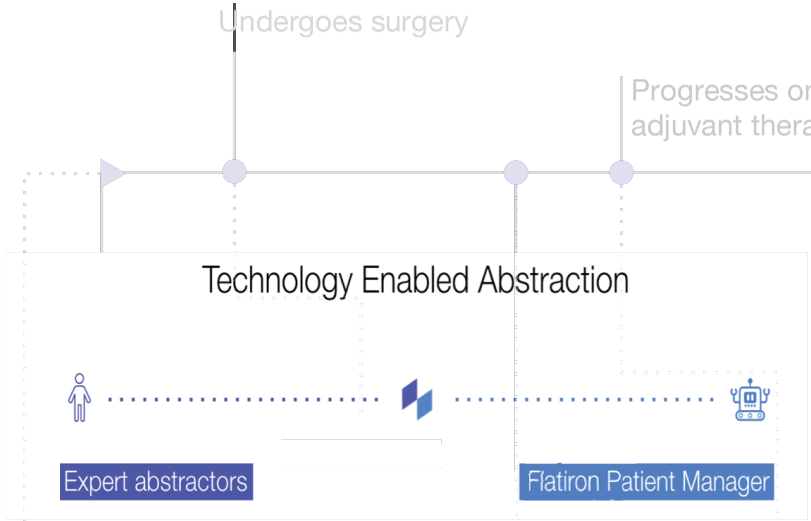


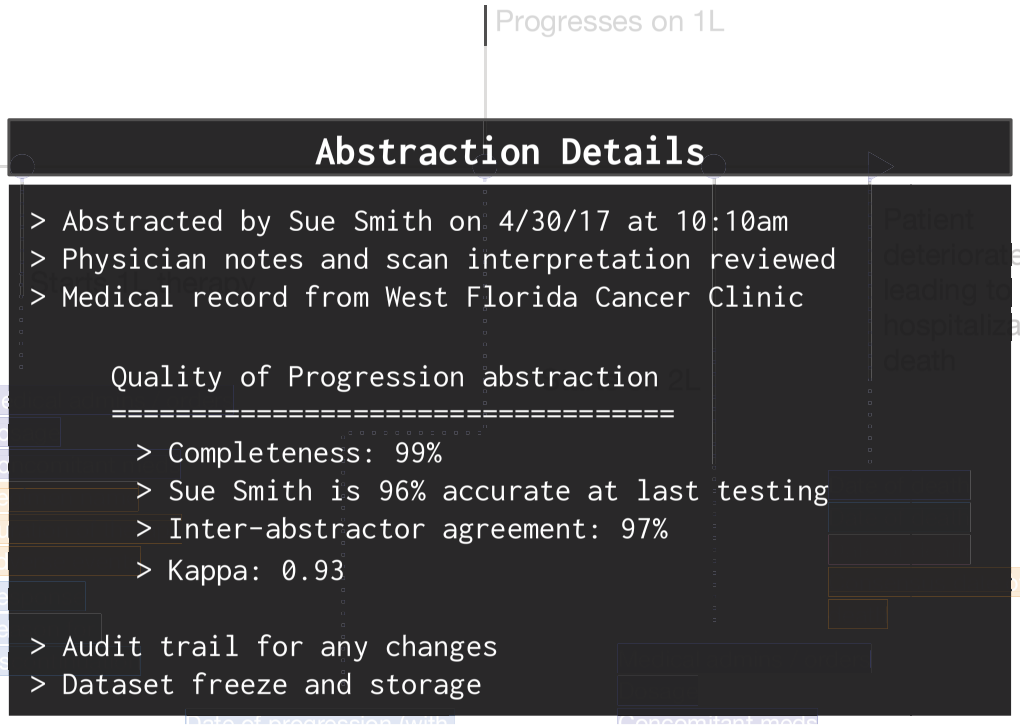
A comprehensive view of the patient journey



*Relative timing not exact

Technology Enabled Abstraction for curating variables





Patient	Stage at Dx	Biomarkers	2L Treatment	**Progression**	Date of Death
Jane Doe	II	EGFR-, ALK-, PD-L1-	nivolumab	**2017-03-08**	2017-04-12

Structured EMR data | Unstructured EMR data | External mortality data | Combined / derived data

**Relative timing not exact*

flatiron

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-abstractor agreement (proxy for accuracy) for derived variables from unstructured data

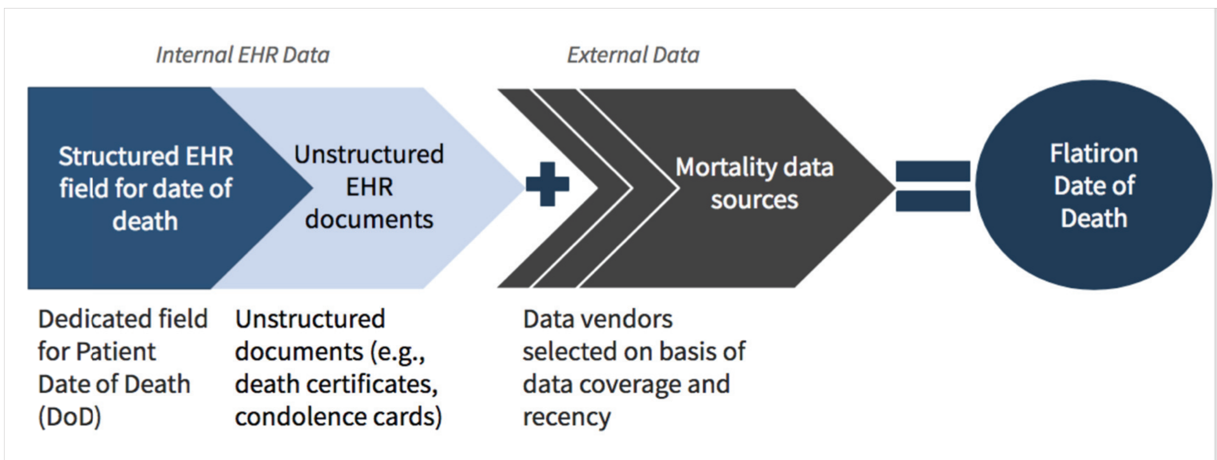
- Data completeness
- Missing data and impact on analysis?

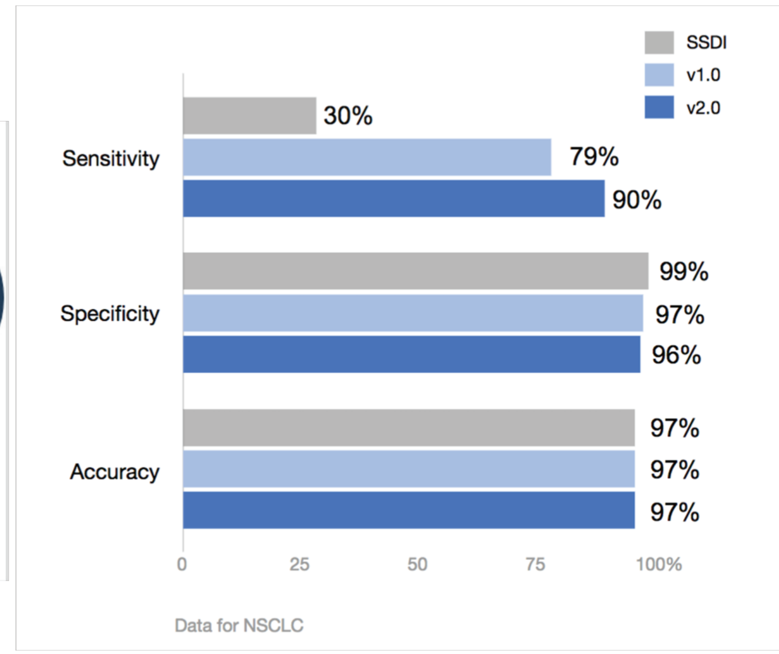
- Provenance
- Variable versioning

- ***Harnessing the power of RWE: A checklist to ensure regulatory grade data quality****, Clinical Pharmacology and Therapeutics, 2017; Rebecca Miksad, and Amy Abernathy*
- ***Evaluation of the impact of Missing Deaths on Overall Survival Analyses Using A Real-World Mortality Endpoint*** *Gillis Carrigan, Samuel Whipple, Michael D. Taylor, Aracelis Z. Torres, Anala Gossai, Brandon Arnieri, Melisa Tucker, Philip P. Hofmeister, Peter Lambert, Sandra D. Griffith, William B. Capra, PDS 2019*

Dataset linkage → Composite endpoint

Evaluate underlying data quality (gold standard = NDI)



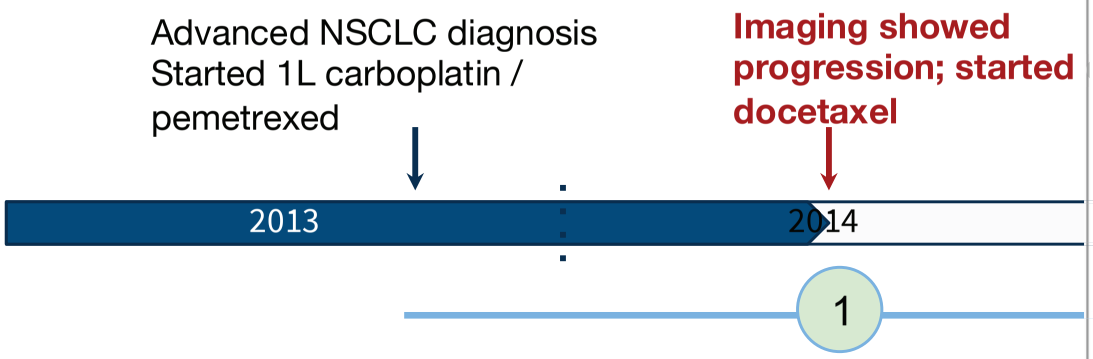


Melissa D. Curtis, Sandra D. Griffith, Melisa Tucker, Michael D. Taylor, William B. Capra, Gillis Carrigan, Ben Holzman, Aracelis Z. Torres, Paul You, Brandon Arnieri, and Amy P. Abernethy, Health Services Research 2018

rwP as a clinician-based endpoint

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



Assessment
Pet Ct evidence of Progressive Non Small cell lung cancer Status post alimta/carboplatin induction therpy followed by 12 cycles of maintenance of Alimta

Histological subtype Adenocarcinoma

History of Tobacco abuse.

Good performance status.

Normocytic Anemia
Disease Status: Progression of disease.

Recommendation/Plan
1. Discussed pet/ct results and the fact that she has evidence of disease progression.Pros and cons of further treatment options were discussed.

2. Incurable nature of disease was emphasized

3. Given her good performance status and the fact that she wants to pursue with further therapy the game plan is to proceed with salvage therapy utilising single agent taxotere at a dose of 60mg/m2 along with neulasta support.

4 Also would continue her on Exjeva which we would give her every 6 weeks.

- Clinically anchored with radiology and pathology reports serving as corroborative evidence
- Most practical and scalable

flatiron

Validation: Patient-level correlation between rwPFS and OS

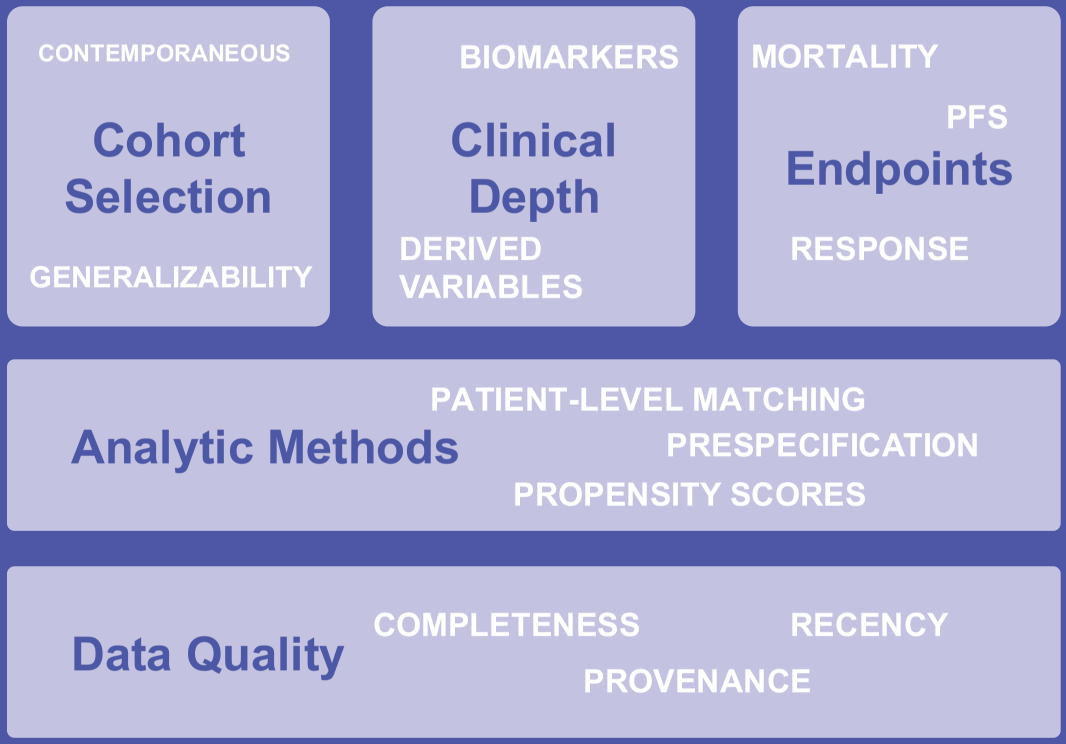
Methods:

- 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

Correlation	N	ρ (95% CI)
rwPFS vs OS	20,020	0.76 (0.75, 0.77)
rwTTP vs OS	11,902	0.69 (0.68, 0.70)
rwTTNT vs OS	9,269	0.61 (0.60, 0.62)

Real-world Control

Foundation for rwCA Development

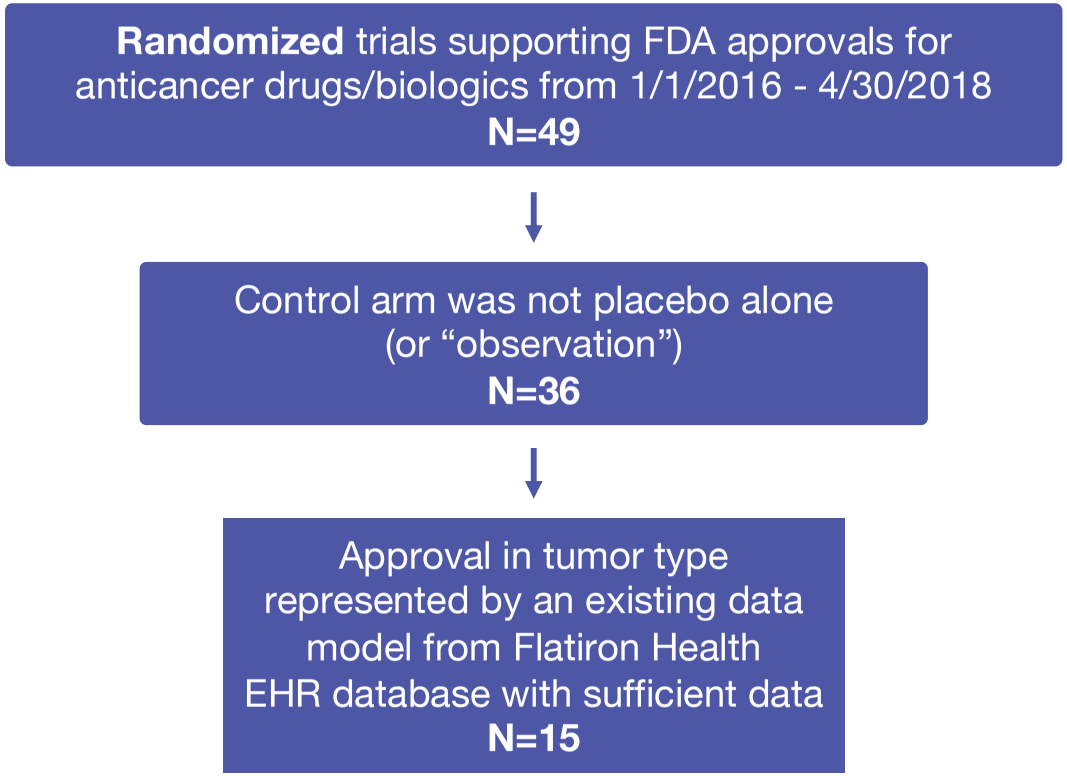


Validation through Replication:

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



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

Approvals included initial, supplemental, accelerated & regular approval following accelerated approval

- **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

flatiron

Summary of randomized clinical trials included in analyses

Tumor type	Name of trial	Front line setting?	Primary endpoint(s)
melanoma	CheckMate-067	yes	PFS, OS
kidney	METEOR	no	PFS
kidney	NCT01136733	no	PFS
NSCLC	OAK	no	OS
NSCLC	POPLAR	no	OS
breast	MONARCH-3	yes	PFS
breast	MONALEESA-2	yes	PFS
myeloma	POLLUX	no	PFS
NSCLC	KEYNOTE-024	yes	PFS
head and neck	CheckMate-141	no	OS
myeloma	CASTOR	no	PFS
NSCLC	AURA3	no	PFS
breast	PALOMA-2	yes	PFS
NSCLC	KEYNOTE-021	yes	ORR
urothelial	KEYNOTE-045	no	PFS, OS
NSCLC	ALEX	yes	PFS
kidney	CABOSUN	yes	PFS
breast	OlympiAD	no	PFS
kidney	CheckMate-214	yes	ORR, PFS, OS
breast	MONARCH-2	no	PFS
breast	PALOMA-3	no	PFS

Weighting real-world patients to published trials

- Derive “inverse odds” weights (w_i = Pr(C_i=0 | x_i) / Pr(C_i=1 | 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

References: (1) Signorovitch JE, et al. Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research. *Value in Health; 15 (2012) 940 –947.* (2) Westreich et al. Transportability of Trial Results Using Inverse Odds of Sampling Weights. American Journal of Epidemiology, Volume 186, Issue 8, 15 October 2017, Pages 1010–1014.

OAK
comparison

Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial

Achim Rittmeyer, Fabrice Barlesi, Daniel Waterkamp, Keunchil Park, Fortunato Ciardiello, Joachim von Pawel, Shirish M Gadgeel, Toyoaki Hida, Dariusz M Kowalski, Manuel Cobo Dols, Diego L Cortinovis, Joseph Leach, Jonathan Polikoff, Carlos Barrios, Fairooz Kabbinavar, Osvaldo Arén Frontera, Filippo De Marinis, Hande Turna, Jong-Seok Lee, Marcus Ballinger, Marcin Kowanetz, Pei He, Daniel S Chen, Alan Sandler, David R Gandara, for the OAK Study Group*

Cohort Selection and data completeness report

Cohort Selection

Table 1. Flatiron cohort attrition

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

Completeness summary

Table 2. Completeness of key data elements in replicating I/E criteria

Data element	% Complete
Neutrophils or granulocytes	91.3%
Creatinine	99.5%
Platelets	98.9%
Hemaglobin	98.9%
Albumin	90.7%
Lymphocytes	96.2%
White blood cells	98.9%
ALT	96.7%
AST	96.7%
Calcium	95.1%
Bilirubin	91.8%
ECOG	36.6%

flatiron

OAK
comparison

Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial

Achim Rittmeyer, Fabrice Barlesi, Daniel Waterkamp, Keunchil Park, Fortunato Ciardiello, Joachim von Pawel, Shirish M Gadgeel, Toyoaki Hida, Dariusz M Kowalski, Manuel Cobo Dols, Diego L Cortinovis, Joseph Leach, Jonathan Polikoff, Carlos Barrios, Fairooz Kabbinavar, Osvaldo Arén Frontera, Filippo De Marinis, Hande Turna, Jong-Seok Lee, Marcus Ballinger, Marcin Kowanetz, Pei He, Daniel S Chen, Alan Sandler, David R Gandara, for the OAK Study Group*

Baseline Comparison

Baseline patient characteristics

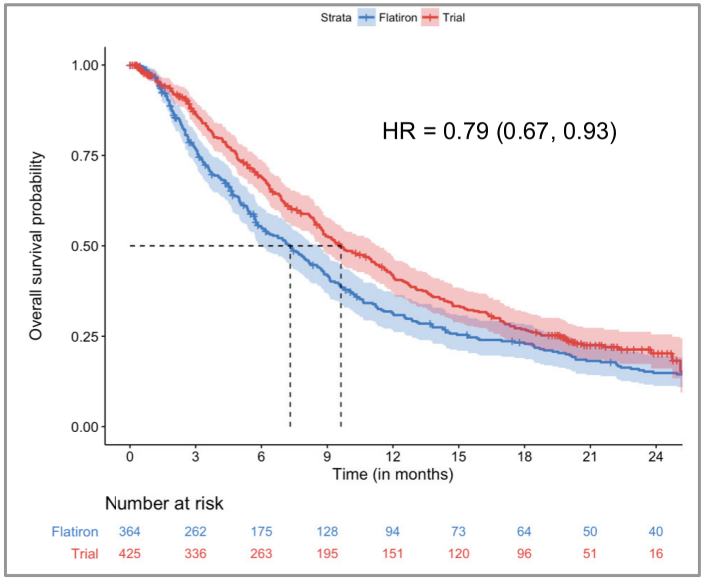
	RWD I/E aligned	Weighted	Trial Control	Trial Treatment
% <Med Age	35	50	51	50
% Male	56	61	61	61
% White	71	71	70	71
% Squamous	12	11	26	26
% EGFR+	4	6	10	10
% KRAS +	12	12	8	6
% 1 Prior Line	52	54	75	75
% Smoking History	92	80	83	80

flatiron

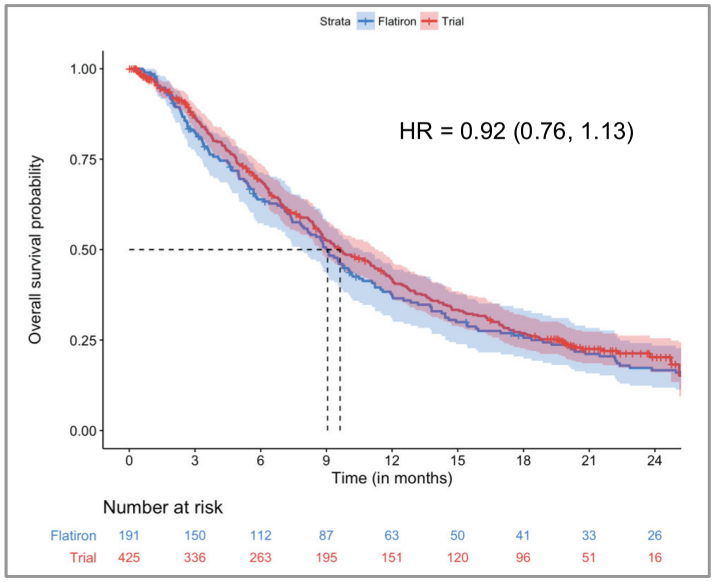
OAK
comparison

Overall survival
(Primary endpoint)

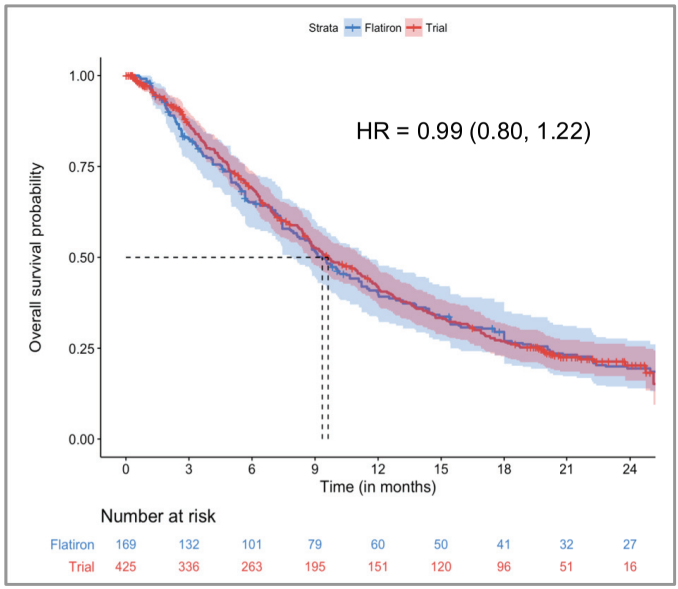
Naive real-world control



Alignment with I/E criteria



Weighted

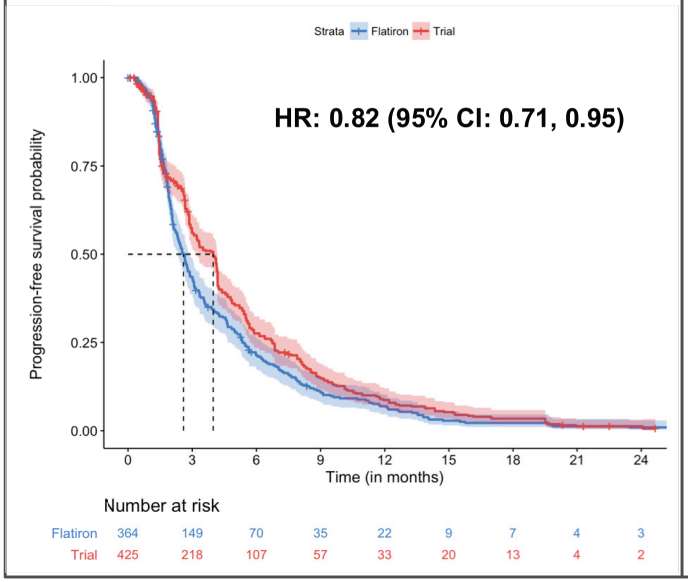


flatiron

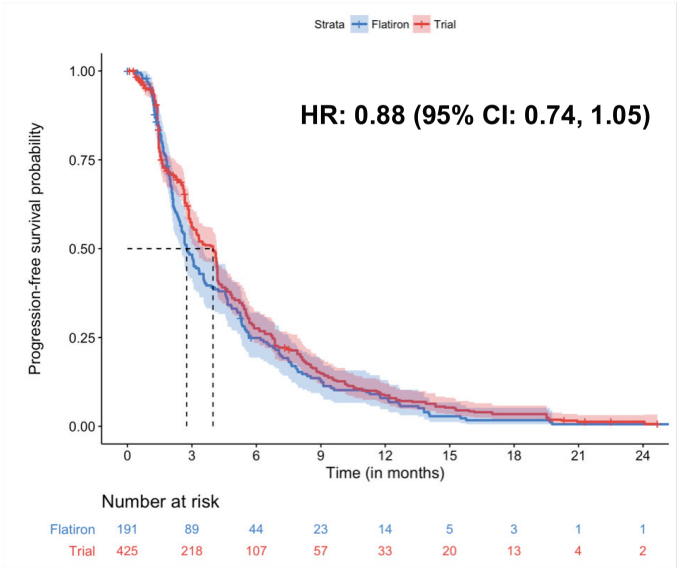
OAK
comparison

Progression-free survival
(secondary endpoint)

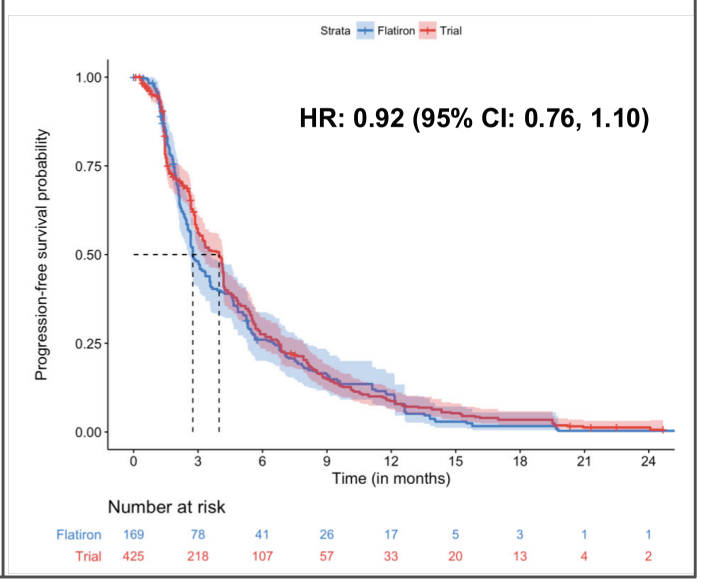
Naive real-world control



Alignment with I/E criteria



Weighted



flatiron

Planned sensitivity analyses

Comparing patients in original trial's control arm versus :

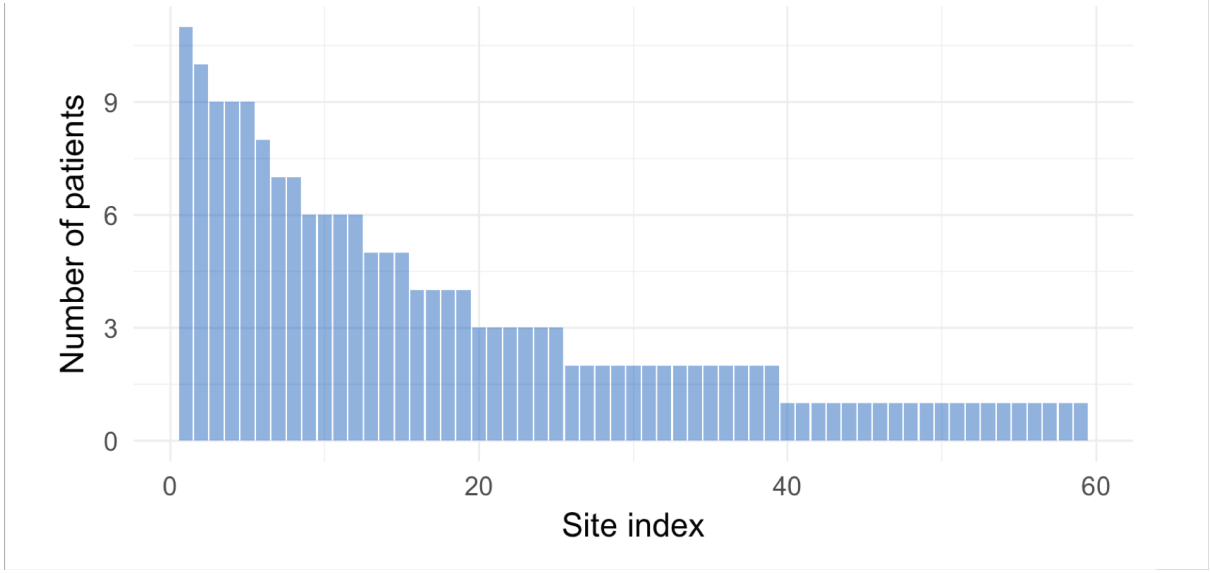
Trial	**Main analyses (weighted)** HR (95% CI)	**Excluding patients with missing ECOG performance status** HR (95% CI)	**Excluding patients with missing laboratory results used to define organ function** HR (95% CI)	**Excluding patients treated before trial enrollment period started** HR (95% CI)
OAK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

OAK
comparison

Diagnostics

Generalizability
(Distribution of sites)

Distribution of Number of patients from each sites in the Flatiron network



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

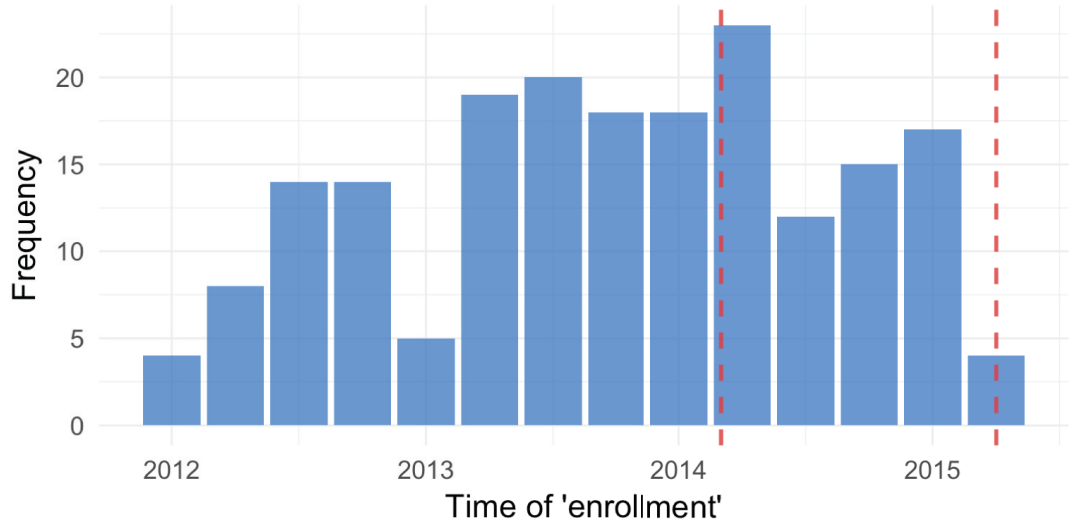
flatiron

OAK
comparison

Diagnostics

Trial recruitment window
(51.5% of Flatiron cohort)

Follow-up Comparisons



- Median follow-up in the Flatiron cohort is 46.2 months.

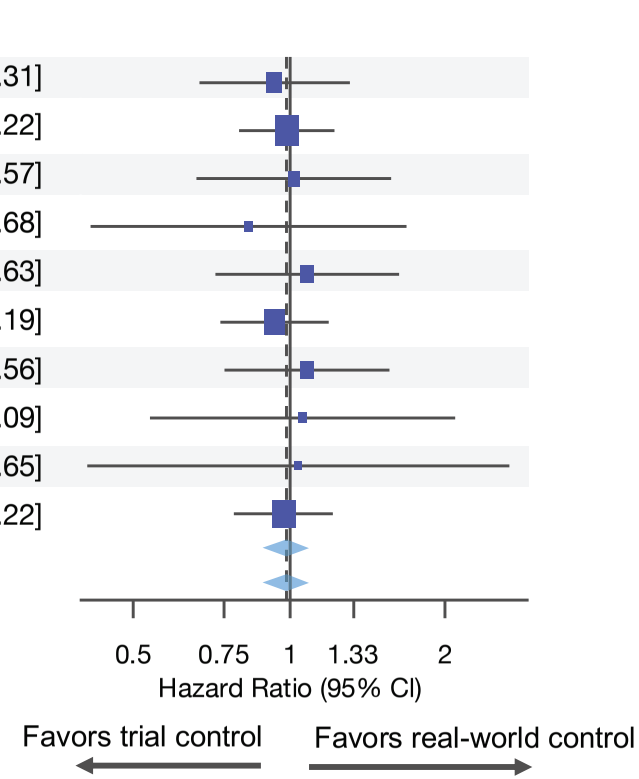
- Concurrent patient cohort
- Possible due to large sample size
- HR=1.01 if cohort is restricted

flatiron

OS Combined HR (95% CI):

0.98 (0.89 - 1.09)

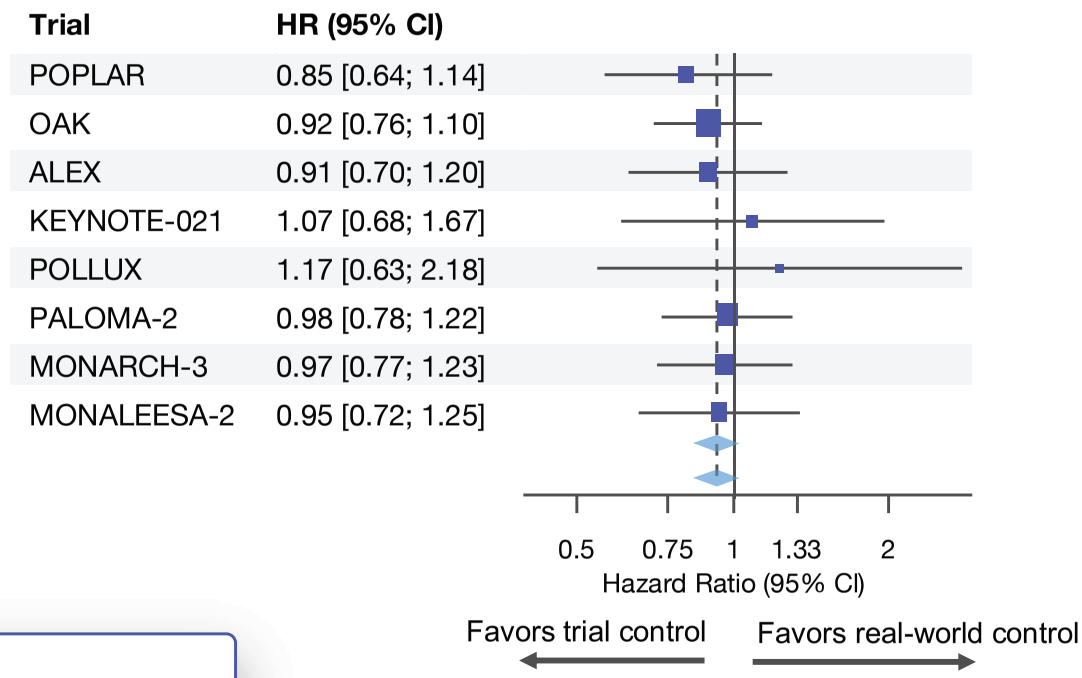
- Overall consistent finding for OS



flatiron

PFS Combined HR (95% CI):

0.94 (0.86 - 1.04)



- Similarly consistent finding for PFS

flatiron

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

flatiron

Cohort selection

Weighting or Matching

Comparing Outcomes

flatiron

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

Cohort selection

Weighting or Matching

Comparing Outcomes

flatiron

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

Cohort selection

Weighting or Matching

Comparing Outcomes

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*

flatiron

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

flatiron

Thank you