RWD125

Identifying Fit-for-Emulation Data: Adaptation of a Structured Data Feasibility Assessment Process for Real-world Oncology Trial Emulations

Natalie Levy, Paige Sheridan, Ulka Campbell, David Lenis, Inish O'Doherty, Adina Estrin, Nileesa Gautam, Monica Iyer, Sarah McDonald, Lauren Becnel, Drew Belli, Gillis Carrigan, K Arnold Chan, James Chen, Victoria Chia, Neil Dhopeshwarkar, Joy Eckert, Laura Fernandes, Max Goldstein, Joel Greshock, Rachele Hendricks Sturrup, Jenny Huang, XiaoLong Jiao, Sajan Khosla, Orsolya Lunacsek, Lynn McRoy, Yanina Natanzon, Osayi Ovbiosa, Nelson Pace, Simone Pinheiro, Jameson Quinn, Megan Rees, Jennifer Rider, Mothaffar Fahed Rimawi, Travis Robinson, Carla Rodriguez-Watson, Chithra Sangli, Khaled Sarsour, Sebastian Schneeweiss, Mark Shapiro, Mark Stewart, Aliki Taylor, CK Wang, Shirley Wang, Asher Wasserman, Yiduo Zhang, Ann Madsen

BACKGROUND

- Randomized controlled trials (RCTs) are the gold standard for evaluating the efficacy and safety of biomedical products, but they face several practical and ethical barriers (e.g., enrollment difficulties, long study timelines, widespread use of off-label therapies).
- Regulators increasingly recognize the potential of non-interventional studies using real-world data (RWD) to produce effectiveness and safety evidence in oncology.
- RWD studies can complement RCTs by generating new hypotheses, expediting more cost-effective results, representing broader patient populations, reflecting real-world clinical care patterns, and assessing longer-term outcomes.
- Non-interventional RWD studies face threats to internal validity that are not encountered in RCTs (e.g., lack of randomization, substantial missing data).
- The Coalition to Advance Real-World Evidence through Randomized Controlled Trial Emulation (CARE) Initiative seeks to advance understanding of circumstances when RWD can generate valid treatment effectiveness estimates by emulating oncology RCTs using RWD.
- Successful emulation requires *fit-for-emulation* data.

OBJECTIVE

- Describe learnings from a structured feasibility assessment process to evaluate potential datasets for CARE studies, which may inform other RCT emulations.

METHODS

- Candidate RCTs were identified from active comparator trials for common tumor types leading to approvals during 2015-2020.
- Trials with design features that would be difficult to emulate in any RWD source were excluded (e.g., new biomarker indication, very recent approval).
- 6 partner RWD sources (de-identified) were considered.
- Data feasibility assessments proceeded in two phases.
  1. Initial screening confirmed the counts of patients with the treatments of interest and the capture of necessary outcome data. Data sources with available sample size >1.5x the size of the trial proceeded to phase 2.
  2. A more detailed feasibility assessment using a modified version of the Structured Process to Identify Fit-For-Purpose Data (SPIFD2) framework was conducted.
- Key RCT design elements and potential confounders were identified. The ability to operationalize each element was assessed and ranked (1-low, 5-high; Figure 1), based on measure reliability/validation and missingness.

Figure 1. SPIFD2 Ranking Scale

Data requirements not met		Some data requirements are met		Nearly all data requirements met
1	2	3	4	5

RESULTS

- 23 RCTs were selected for initial screening (Figure 2).
- 9 potential RCT-data source (DS) combinations (representing 6 RCTs and 4 potential DSs) progressed to detailed feasibility assessment.
- Table 1 summarizes the detailed data feasibility assessments conducted for emulation of the KEYNOTE-189 trial in 3 DSs.
- 3 emulations across 2 DSs — KEYNOTE-189 in DS 3 and DS 4, and PALOMA-2 (data not shown) in DS 3 — were ultimately selected for protocol development.

Figure 2. Trial Selection Flowchart

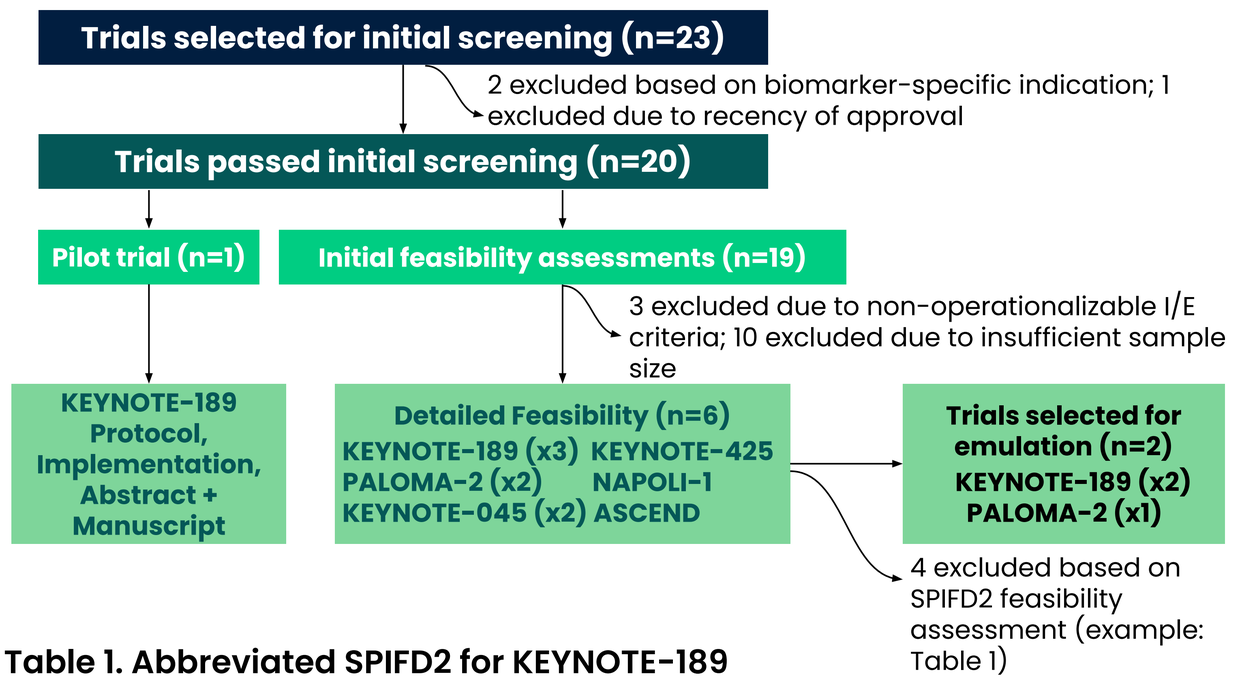

Table 1. Abbreviated SPIFD2 for KEYNOTE-189

DESIGN ELEMENTS	DESCRIBE ORIGINAL CLINICAL TRIAL	DESCRIBE RWD STUDY EMULATION		RWD SOURCE FEASIBILITY ASSESSMENT		
		Minimum criteria	Criteria ranking	DS 2	DS 3	DS 4
OVERALL RATING				3	5	4
GENERAL						
Sample size	**Trial sample size**	**1.5x trial**	Must Have	**Sample size rating**		
Exposed	410	615		5	5	4
Comparator	206	309		5	5	5
VARIABLE-RELATED						
Variable	**Original RCT definition**	**Minimum criteria**	**Criteria ranking**	**Rating**		
Treatment	Exposed and comparator treatments	Treatment data	Must Have	5	5	5
Inclusion 1	>18 years of age	Year of birth	Must Have	5	5	5
Inclusion 2	Metastatic nonsquamous NSCLC	Pathology data	Must Have	5	5	5
Inclusion 3	EGFR-/ALK-	Biomarker data	Must Have	3	5	5
Inclusion 4	ECOG score of 0/1	ECOG data	Must Have	4	5	5
Inclusion 5	>1 measurable lesion per RECIST	RECIST not used in a RW setting	Not Applicable			
Exclusion 1	Symptomatic CNS metastases	Metastatic site data	Nice to Have	5	5	5
Exclusion 2	History of certain other conditions and treatment	Comorbidity and treatment data	Nice to Have	3	4	2
Exclusion 3	Received lung radiation	Radiation data	Nice to Have	2	4	5
Outcome 1	OS	Death data	Must Have	2	4	5
Outcome 2	PFS	Death and progression data	Must Have	2	4	5
Confounding 1	Not applicable in a randomized setting	Sex	Must Have	5	5	5
Confounding 2		Race/ethnicity	Must Have	5	5	5
Confounding 3		Smoking status	Must Have	5	5	5
Confounding 4		PD-L1 status	Must Have	3	5	5

CONCLUSIONS

- Oncology RCT emulations require specific eligibility criteria and outcomes that make identifying *fit-for-emulation* RWD particularly challenging.
- Routinely-captured, non-cancer diagnoses/treatments are absent from high-quality, oncology-specific datasets, which may hinder the operationalization of certain RCT inclusion/exclusion criteria.
- Structured, rigorous, and transparent data feasibility assessments are critical for identifying *fit-for-emulation* RWD, contextualizing results, and identifying gaps in existing datasets.

AETION