medidata

Roche

Using Synthetic Control Databases to Accelerate Indication-Specific Safety and Efficacy Evidence

Colin Neate, MSc, Oncology Biostatistics, Roche Mississauga.

15th August 2019

Acknowledgements

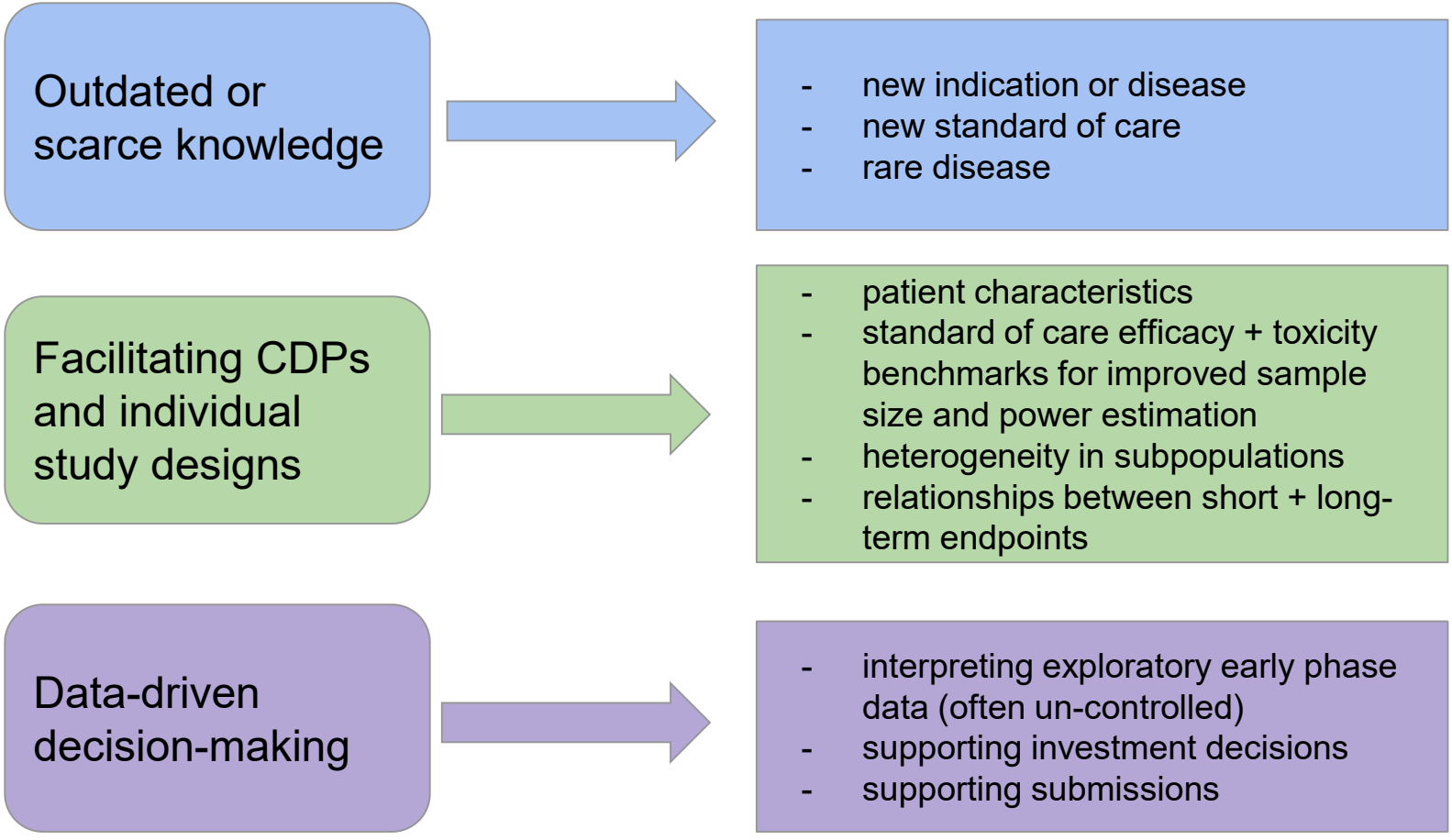
Key collaborator on this project:

Lisa Ensign, PhD, Principal Biostatistician, Medidata Solutions, USA

Outline

- Motivation for and overview of external data sources
- Introduction to SCDs
- Overview of pilot in breast cancer
  - scope
  - example results/learnings
  - challenges
- Further considerations + summary

Why use external data sources in drug development?



What external data sources are there?

Aggregated:

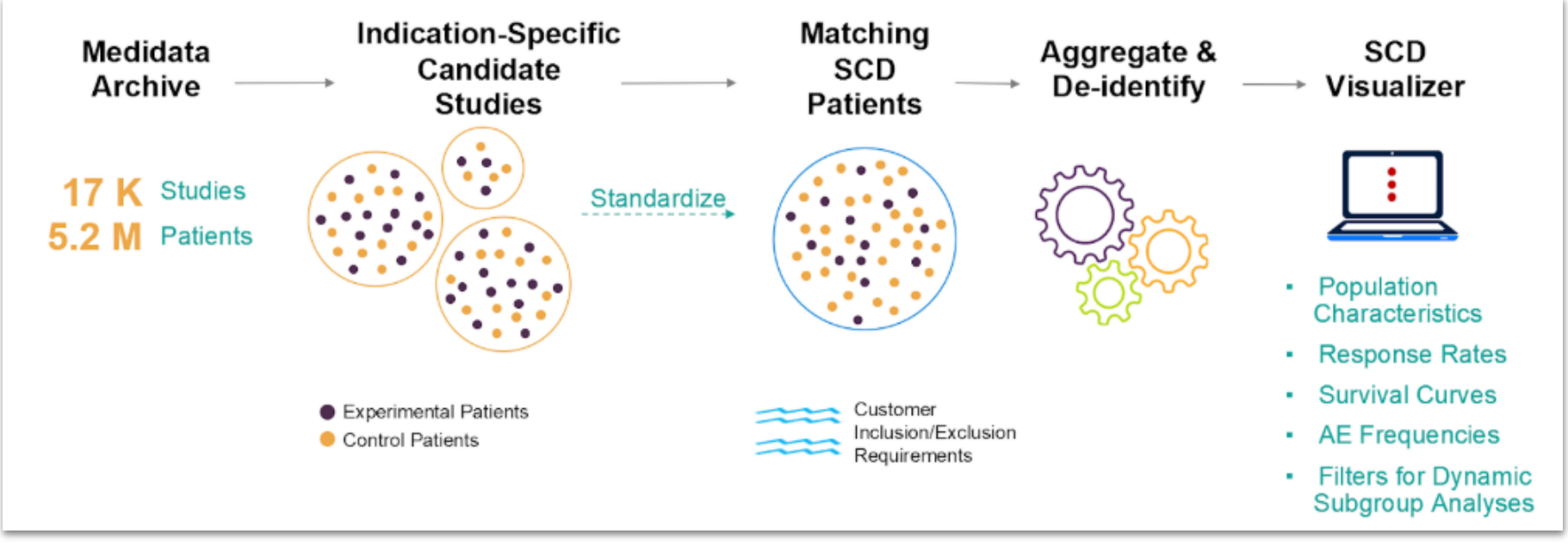
- Journal publications
- ClinicalTrials.gov, EudraCT and other registries
- EMA policy 70 (publication of redacted CSRs);
- FDA and Health Canada pilots
- Aggregated clinical trial data sources, e.g. *Medidata Archive*

Anonymised Individual Patient level data:

- Clinicalstudydatarequest.com
- TransCelerate's Placebo Standard of Care (PSoC) Initiative
- Disease area specific e.g. Project Datasphere
- RWE data sources (e.g. FlatIron)
- Molecular data sources (e.g. FMI)
- Company-specific

Synthetic Control Databases, SCD™

Aggregated database *of hundreds (or thousands) of patients* ***from recent trials*** *built to match a researcher's inclusion / exclusion criteria for an indication*



Pilot for Metastatic Breast Cancer (mBC)

mBC SCD

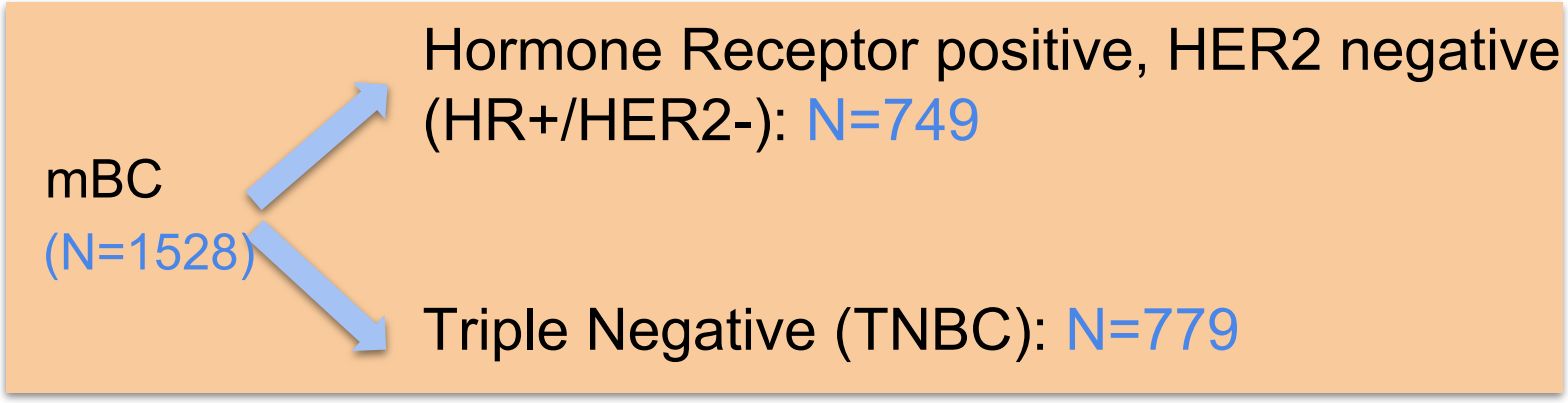
Hormone Receptor positive, HER2 negative (HR+/HER2-) mBC: *to support **design** and **interpretation** of internal programs (heterogeneous disease w/multiple emerging standards of care);*

Triple Negative (TNBC) mBC: *to support **design** and **interpretation** of internal programs (lack of recent internal data)*

Focus on patients pre-treated for mBC "2nd line" (or later)

Specifications developed for patient characteristic, efficacy and safety outcome variables

Current status



SCD data drawn from >10 phase II and III trials that have "completed" their primary analysis time point

Quarterly "updates" to SCD to add new data and functionality

Selected data domains and variables

(*>190* variables included in total)

Adverse Events	Causality, Outcome, Preferred Term (+ all MedDRA levels), SAEs, Severity
Baseline Disease Characteristics	Age at Diagnosis of Disease, Baseline ECOG, Baseline Lactate Dehydrogenase, Baseline Serum Albumin, BC Subtype (TNBC or HER2- / HR+), Diagnosis Preferred Term, Disease Stage (Including TNM), Line of Therapy, Menopausal Status, Relapsed / Refractory Flags, Tumor Grade
Biomarkers	BRCA, BRCA1, BRCA2, CA 15-3, CA 17-19, Estrogen Receptor, HER2, Hormone Receptor, Progesterone Receptor
Concomitant Medications	Ongoing, Preferred Term (+ all ATC levels), Route
Demographics	Age, Baseline Height/Weight, Ethnicity, Gender, Race, Region
Disease Response	Best Overall Response, Clinical Benefit, Death, Progression Free Survival, Overall Survival, Time to First Partial or Complete Response

Derivation/Standardisation approach

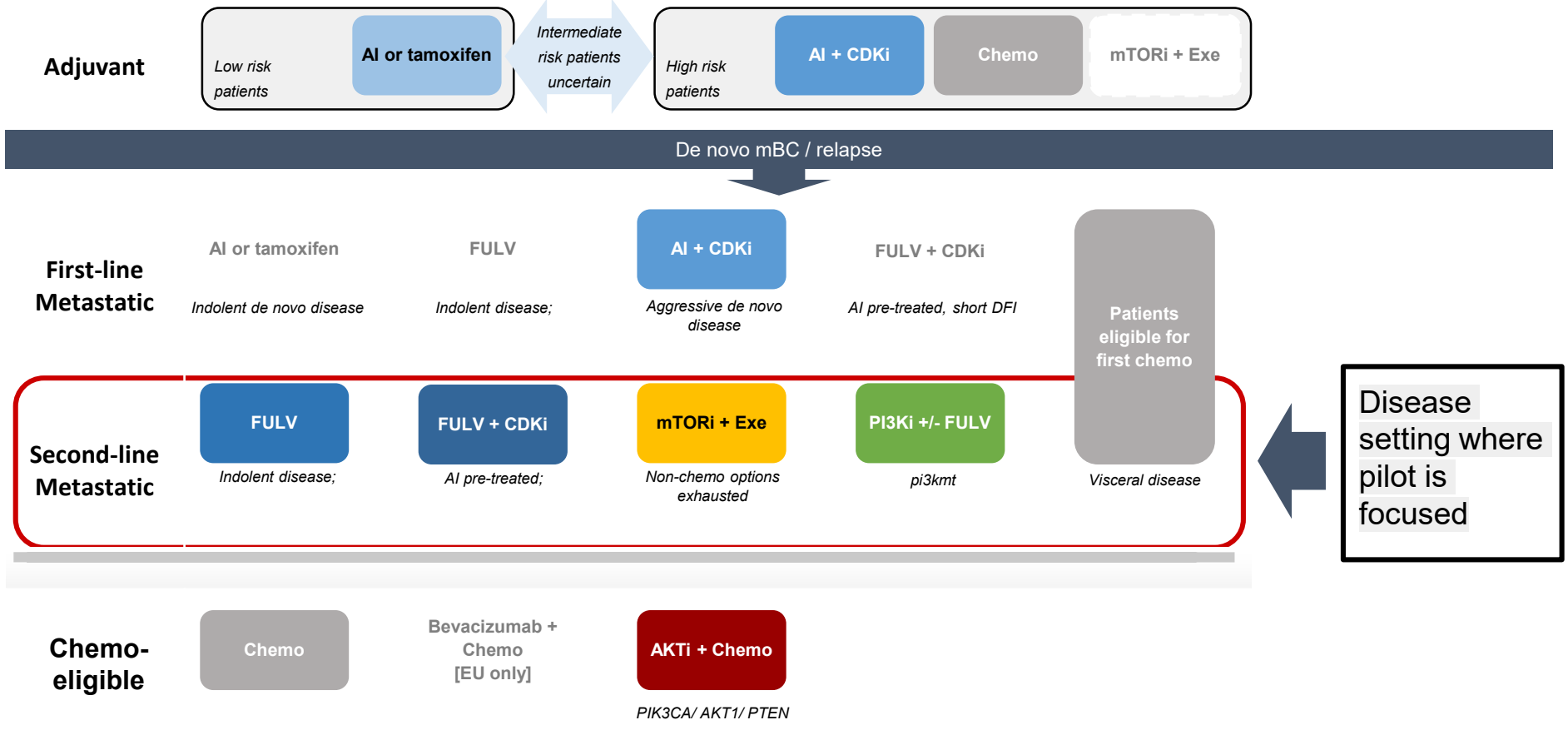
1) Identification of candidate studies/patients ('feasibility' assessment): via protocol title, inclusion criteria, objective information in database
   => does not fully ensure 'just' target studies / patients selected
   => focus on avoiding missing studies/patients ("false negatives")
   => e.g. breast cancer types/lines other than targeted included

2) Standardisation of variables across all the candidate studies
   => standardisation of core demographic, efficacy and safety data by Medidata with this SCD and broader re-use in mind; mapping to SDTM-like structure and creation of ADaM datasets

3) Deep-dive to select specific target population for specific SCD
   => need to utilise additional eCRF variables, apply algorithms and medical input, e.g. TNBC can be a combination several pieces of info: "HER2" and "HR" status (defined by PgR and EgR)

Then it can get more complicated still...

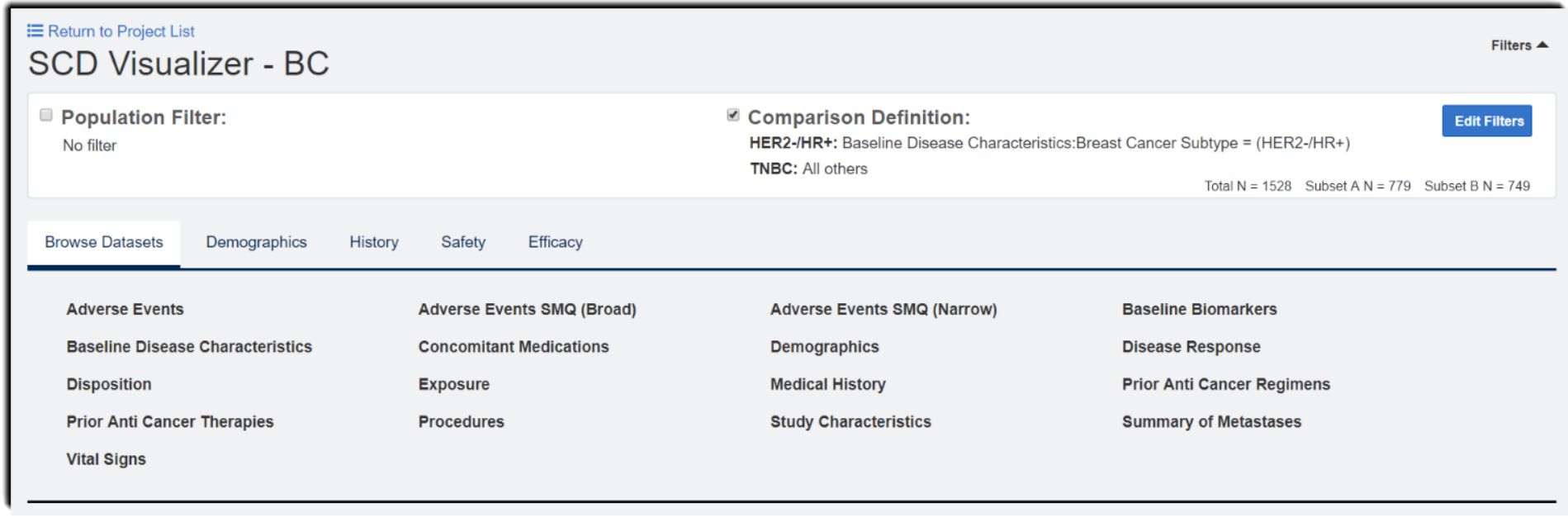
Evolution of SOC in HER2-/HR+ BC



Complex derived variables: extensive collaboration by Subject Matter Experts, Algorithms, Machine Learning

Reported Term	Standardized Data					
	Preferred Term	Treatment Intent	Best Overall Response	Therapy Study Start Day	Therapy Study End Day	Drug Class
PACLITAXEL	paclitaxel	Curative		-2352	-2268	Taxanes
GOSERELIN	goserelin	Curative	CR	-2170	-1450	Endocrine therapy not otherwise specifed
TAMOXIFEN	tamoxifen	Curative		-2170	-1294	SERM
CYCLOPHOSPHAMIDE	cyclophosphamide	Adjuvant		-1268	-1207	Chemotherapy not otherwise specified
DOCETAXEL	docetaxel	Adjuvant		-1268	-1207	Taxanes
LETROZOLE	letrozole		SD	-1099	-456	AI
DENOSUMAB	denosumab			-988		Supportive Care
EXEMETHSANE	exemestane		PD	-456	-343	AI
EVEROLIMUS	everolimus		PD	-442	-343	Protein Kinase Inhibitors (mTOR inhibitor)
FULVESTRANT	fulvestrant		SD	-343	-175	SERD
NANOPARTICLE ALBUMIN-BOUND PACLITAXEL (ABRAXANE)	paclitaxel		PD	-133	-63	Taxanes

Access to SCD via Visualiser



- Tables and Plots (according to data type)
- Dynamic filtering of sub-populations
- User-specified comparisons

- Researcher limited to generating results in line with preservation of patient and study-level anonymity (at least 5 patients and 2 Sponsors)

Using pilot SCD to generate disease-specific insights - breast cancer characteristics

		TNBC (N = 749)	HER2- / HR+ (N = 779)	Total (N = 1528)
BRCA1 or BRCA2	Positive	379 (50.6%)	431 (55.3%)	810 (53.0%)
Derived Line of Therapy for the Current Treatment	1L	0 (0.0%)	114 (14.6%)	114 (7.5%)
	2L	328 (43.8%)	230 (29.5%)	558 (36.5%)
	2L+	265 (35.4%)	163 (20.9%)	428 (28.0%)
	3L+	156 (20.8%)	272 (34.9%)	428 (28.0%)
Assigned Treatment Arm Drug Class	Antimetabolites	146 (19.5%)	63 (8.1%)	209 (13.7%)
	Antimetabolites + SERD	0 (0.0%)	46 (5.9%)	46 (3.0%)
	Antimetabolites + Targeted-Other	37 (4.9%)	79 (10.1%)	116 (7.6%)
	Chemotherapy not otherwise specified	41 (5.5%)	52 (6.7%)	93 (6.1%)
	PARP inhibitors	270 (36.0%)	300 (38.5%)	570 (37.3%)
	SERD	0 (0.0%)	51 (6.5%)	51 (3.3%)
	SERM + Endocrine therapy not otherwise specified	0 (0.0%)	157 (20.2%)	157 (10.3%)
	Targeted-Other	213 (28.4%)	0 (0.0%)	213 (13.9%)
	Missing	42 (5.6%)	31 (4.0%)	73 (4.8%)
Sites of Metastases at Enrollment	Bone	138 (31.2%)	544 (70.0%)	682 (56.0%)
	Brain	95 (21.5%)	35 (4.5%)	130 (10.7%)
	Liver	95 (21.5%)	270 (34.7%)	365 (29.9%)
	Lung	212 (48.0%)	258 (33.2%)	470 (38.6%)
	Other	250 (56.6%)	471 (60.6%)	721 (59.1%)
Visceral Metastasis	Yes	382 (86.4%)	638 (82.1%)	1020 (83.7%)

Treatment Response

Sample Table Best Overall Response

Restore Default | Save | Save as Copy

Export

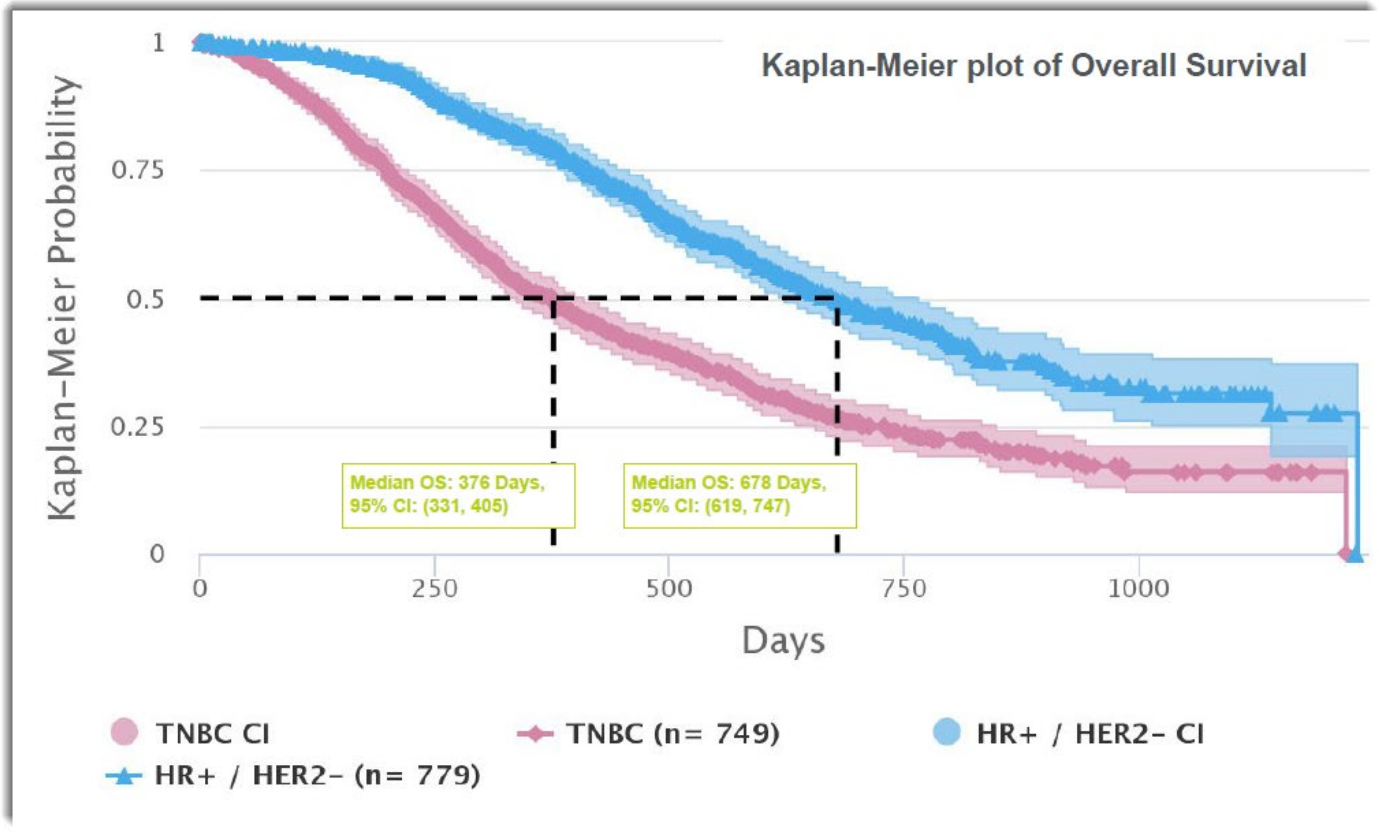
Select Fields Apply

- **Baseline Biomarkers**
- **Baseline Disease Characteristics**
- **Demographics**
- **Disease Response**
  - [x] Best Overall Response
  - [ ] CR
  - [ ] Clinical Benefit Flag
  - [ ] Death Flag
  - [ ] Duration of Response (First Complete to Progression)
  - [ ] Duration of Response (First Partial or Complete to Progression)
  - [ ] Endpoint Assessment Schedule (weeks)
  - [ ] Overall Survival
  - [ ] PR
  - [x] Partial or Complete Response

Data for SelectedFields		TNBC (N = 749)	HER2-/HR+ (N = 779)	Total (N = 1528)
Best Overall Response	CR	28 (3.74%)	17 (2.18%)	45 (2.95%)
	Missing	142 (18.96%)	128 (16.43%)	270 (17.67%)
	NE	2 (0.27%)	3 (0.39%)	5 (0.33%)
	PD	162 (21.63%)	120 (15.40%)	282 (18.46%)
	PR	178 (23.77%)	200 (25.67%)	378 (24.74%)
	SD	237 (31.64%)	311 (39.92%)	548 (35.86%)
	MISSING	0 (0.00%)	0 (0.00%)	0 (0.00%)
Partial or Complete Response	N	543 (72.50%)	562 (72.14%)	1105 (72.32%)
	Y	206 (27.50%)	217 (27.86%)	423 (27.68%)
	MISSING	0 (0.00%)	0 (0.00%)	0 (0.00%)
Stable Disease > 24 Weeks	N	718 (95.86%)	690 (88.58%)	1408 (92.15%)
	Y	31 (4.14%)	89 (11.42%)	120 (7.85%)
	MISSING	0 (0.00%)	0 (0.00%)	0 (0.00%)

- Abbreviations: N = number of subjects in the population; Q1=25% Percentile; Q3= 75% Percentile; SD=Standard Deviation.
- Number of subjects with non-missing data, used as the denominator

Overall Survival: similar response rates (28%) do not lead to similar OS between breast cancer subtypes



Some Limitations / Challenges

- Blinded studies - cannot link patient characteristics/outcomes to treatment in RAVE
- Variable data standards and information across studies and complexity limits robustness =>*e.g. frequency of clinical response (and other) assessments can vary between studies as can definitions such as of line of therapy*
- Limited to clinical data captured in database - excludes data maintained externally (e.g., molecular)
- Possible publication bias regarding companies allowing data to be shared (although when they do, all studies are typically provided)
- Small sample size in subpopulations
- Lack of access to individual patient level data

With these restrictions, pilot's current focus is on using SCD, alongside RWE to support development programmes

Future considerations

A future step could be to use SCD to facilitate questions regulatory settings. Potentially:

- Building a Synthetic Control Arm (SCA) to serve as an external control for an experimental treatment in an uncontrolled (or partially randomised) trial
  - *e.g. covariate matched, propensity scoring, prognostic scoring => pilot SCA previously developed in AML*
- Supplementary evidence (in combination with other sources) during submissions - e.g. interpreting safety data and quantifying benefit/risk in support of regulatory submissions

Summary

- Pilot has shown feasibility for creating SCDs to support development programmes (2nd pilot in HCC is ongoing)
- Comes with substantial time investment (both at Medidata and Roche) and need to manage stakeholder expectations for what can be achieved
- Key limitations for this pilot have been access to molecular information and to specific treatment subgroups
- ‘Visualizer’ very intuitive (including for non-statisticians)
- SCDs and SCAs are a valuable addition to external data options for planning/executing development programmes

References

Posters available on request from PSI 2019 conference and 2019 DIAglobal meeting

Donald A. Berry, et al., Journal of Clinical Oncology 2017 35:15_suppl, 7021-7021
Creating a synthetic control arm from previous clinical trials: Application to establishing early end points as indicators of overall survival in acute myeloid leukemia (AML) *Poster presented at ASCO 2017*

Questions


thank you
danke
gracias
merci
obrigado
dziękuję
sukriya
go raibh maith agat
teşekkür ederim
ngiyabonga
mochchakkeram
kop khun krap
terima kasih
grazie
arigatô
dakujem
bedankt
спасибо
мерси
dank je
tapadh leat
謝謝
감사합니다
ευχαριστώ