Leveraging Randomized Designs to Generate RWE

Robert Temple
CDER, Deputy Director Clinical Science
July 11, 2019
Background

There is tremendous interest in making use of the vast amounts of data already collected in healthcare systems (EHRs, claims data, registries) to more efficiently generate evidence of effectiveness. Greatest interest is in use of data in “observational studies,” i.e., NOT randomized trials, but this is not today’s subject and previous publications by FDA personnel [Sherman, et al in 2016] have emphasized the difficulties and lack of experience in doing this. Rather we have urged attention to utilizing RWD in RCTS, and that is the subject of this workshop. But there are many, many issues that will need to be considered and indeed, the program shows this.
Background

The conference is about using data from the healthcare system in randomized trials to generate “RWE.” What people mean by RWE, and the specific study designs to be considered, is not clear and, in fact, the specifics of the data generated by an RCT within a healthcare system could vary tremendously, depending, for example, on whether end points collected are those extracted from the usual records of the healthcare system (claims, EHR, or more detailed records) or are generated as part of the study. I will try to set forth the full range of possibilities, obviously briefly.
Data Quality- In General

Before addressing study design issues, I want to offer one note about the effect on outcome of the quality and precision of the data. Whether, e.g., a person did or did not have an outcome of interest, and the severity of that outcome, is not always obvious, even for “hard endpoints” in CV outcome studies (thus we have adjudication committees), and many endpoints can be even more subjective (if they are recorded at all, e.g., pain, depression.)

Imprecision (noise) can obscure effects, leading to failure in a superiority study or spurious success in a NI study.

All this could depend on the data system. A registry could be better than EHR or claims data.
Easiest Use of RWD: The Low Hanging “Fruit”

Although it is not the major subject of the workshop, which is about USING RWD in the trials, it seems critical to take note of a very obvious use of RWD that could greatly increase study efficiency by facilitating identification of potential study participants, i.e. identifying patients for study using EHR or claims data. This would be very important, as recruitment is a major cause of delay in conducting studies. What could be looked for?

1. The presence of the disease or condition being studied.
2. Critical patient characteristics.
   - Demographics
   - History and potential enrichment characteristics
     – Duration of disease
     – Disease severity
     – History of compliance with treatment
     – Past outcome events such as AMI, stroke, CV surgery
     – Relevant laboratory measures
     – Concomitant illness and concomitant treatments

Obviously, different systems will differ in the amount of available data for screening: claims data are not very detailed; EHR data may be far more complete, and some systems may be especially complete (VA, Mayo.) None of these, however, are as detailed as full medical records, but they could be a screen, to be followed by a more detailed assessment.
Actual Study Could be a Conventional Trial

The actual trial could be simply a conventional trial following the screen for patient identification with:

- Specific study monitoring for effectiveness and safety made outside of the system; i.e., use of typical CRF
- New consent with full information
- More detailed entry criteria: history, lab tests, etc.

Is this still RWE? That depends on the detailed design of the trial, entry criteria, and study endpoints.
As noted, an easy use of RWD is to find patients (a major problem) and it seems possible that patients could be told about the study and consented and possibly queried to obtain more details (i.e., beyond EHR or claims), on past history or lab values, to allow enrichment (e.g., disease severity and duration; prior events such as AMI, stroke, CV surgery, or lab findings to allow enrichment (e.g., disease severity and duration, prior events such as AMI, stroke, and CV surgery, or lab findings such as abnormal renal function, CRP)

You would then do a conventional randomized trial with identified investigators, scheduled monitoring, regular laboratory data collection, complete safety assessments, etc.

But in some cases (probably for marketed drugs where the need for safety data is more limited) clinic visits could be markedly reduced by using novel endpoints to provide some, or even all, of the data. Examples are growing in number, but include clinic visits by telemedicine, use of on-line PROs, and devices such as smart watches and others to detect steps taken, HR, BP, episodes of AF, etc. These endpoints are plainly NOT EHRs or claims data, but are, at least arguably, RWD
Using RWD Alone

Using RWD alone (EHR and claims) to assess the effect of an intervention is obviously possible for some very specific endpoints, notably survival (but with questions about cause-specific mortality), hospitalization (but with concern about cause-specific hospitalization).

It is recognized that the state of very few symptomatic endpoints will be reliably recorded (anxiety, depression, pain, anxiety levels, etc.) in ways that could represent usable endpoints.

What might make a RWD trial more possible?
Possible RWD Trials

If the drug is not marketed you probably need safety monitoring with observation, lab tests and actual visits with the investigator (although, as noted before it may be possible to make use of decentralized features such as local lab tests, telemedicine, etc.) In this case the trial will be a more or less conventional trial, but with use of decentralized data. It will probably still need identified investigators with clear monitoring responsibilities, again with the potential for decentralized interactions.

But suppose the treatment is a one time treatment (TPA of SK in a post-infarction trial) or a maintained, standardized and unchanging treatment with late outcomes, such as an adjuvant chemotherapy trial or a bisphosphonate fracture prevention trial or an anti-platelet drug maintenance trial to prevent late AMI or stroke. In these cases the outcome probably would be detected in EHR or claims and the main concerns would be interim study evaluation and compliance. Although there could surely be safety issues. Could these be done without identified investigators as part of “medical practice,” and would it depend on the health care system (VA seems easier). Could there be standardized queries sent periodically?
These Are Examples of RWD (or almost)

GISSI
In 176 Italian coronary care units patients within 12 hours of an AMI were randomized to SK or placebo. The primary endpoint was 21 day survival (10.7% vs 13%, p < 0.0002) based on registry data (also got other endpoints later from local resources).

TASTE
Is intracoronary aspiration of thrombus prior to PCI randomized to PCI in patients with STEMI beneficial? Used Swedish Coronary Angiography and Angioplasty Registry (SCAAR) to find patients within 24 hr of pain onset and randomized to aspiration + PCI or PCI alone. Primary endpoint was mortality (Nat’I registry) and used other registries (SWEDEHEART) for other endpoints (recurrent MI, stent thrombosis).
Examples

• **Impact AF**
  Randomized trial in Mini-Sentinel Data Partners of the effect on anti-coagulant use in patients with AF (but not receiving anti-coagulants) of a standardized message to physicians. The study will also look at stroke outcomes.

As has been noted before, the Peto/ Collins proposal in 1995 for large simple trials goes a long way toward simplifying conventional randomized trials, bringing them closer to RWD trials by minimizing data collection, exclusions, etc. (e.g., ISIS trials). But they still have many remaining conventional components.
Other Issues

1. Defining investigators and their responsibility

Almost all trials, even “real world trials,” involve investigators going beyond usual care (maybe except when a single dose is used) How does this fit with IND rules and responsibilities

2. Consequences or decreased “rigor” in assessing outcome

“Noise” obscures - much of what is sought in “ pragmatic trials) leads to less precision: some patients may not have the disease or the desired severity of disease; some endpoints are missed or are erroneous; patients are lost to follow up; patients stop the test drug or go onto other similar therapy. All of these problems decrease study power, a bad outcome in a difference-showing trial, and a FATAL (to credibility) outcome in a non-inferiority trial (these are hard enough even without that problem).
"Pragmatic" Trials

• Part, at least, of what people interested in randomized RWE trials are seeking is “pragmatic” trials. The concept goes back a long way, with a proposed framework (PRECIS) set forth in 2009 by Thorpe, Zwarenstein, et al. seeking to define the difference between an “explanatory” trial (typical phase 3 “efficacy” trial to support approval) and a pragmatic trial (“effectiveness” trial).

<table>
<thead>
<tr>
<th></th>
<th>Pragmatic</th>
<th>Explanatory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>All</td>
<td>Selected (high risk, pred enrich)</td>
</tr>
<tr>
<td>Compliance</td>
<td>Not measured</td>
<td>Closely monitored and screened for</td>
</tr>
<tr>
<td>Comparison</td>
<td>Usual practice</td>
<td>Placebo or best alternative (NI or super)</td>
</tr>
<tr>
<td>Follow up</td>
<td>None, registries</td>
<td>Extensive and well-defined</td>
</tr>
</tbody>
</table>

So, not clear to me what the anticipated benefits of the pragmatic trial are and, as noted, no systematic comparison.
Pragmatic

• I have now read a fair amount on the virtues of pragmatic trials and I confess the benefits remain unclear to me. Certainly if patients don’t take the drug the effect will be smaller but I did not need a trial to tell me that. If the patient populations are poorly defined (e.g., if some do not really have the disease) that will reduce effects but why do I want to know that? And if endpoints are missed, that too will reduce effect size but, again, why am I interested in that? And if the effect in people at low risk is harder to detect (it is, that is why we do prognostic enrichment to get enough events), the likely result is a failed study – why is that good? And to detect the effect in the low risk population the trial will need to be a good deal larger, not exactly a cash-saver. So why do it? One reason is that many felt the study populations were excluding too many people (old, black, poor) and that results did therefore not reflect the true user population. That is probably the best argument but recent trials have bee design to overcome that and some maneuvers (seeing older patients in their homes by telemedicine) could help further.”

• But do the explanatory trials give different results – success when the pragmatic trial shows no or less effect? In 2016 I asked Sean Tunis what differences between traditional and pragmatic trials had been found. He asked Merrick Zwarenstein who said that these comparisons had rarely (one or two) been done but the pragmatic trials showed smaller effects (not a surprise if people do not take the drug) and he was looking to do more trials. So the answer then was that the idea that the standard trials were in some way misleading was largely unsupported and (my guess) if there is indeed any difference it is related to compliance.

• So I really do not understand why anyone wants to see pragmatic trials where compliance is less than optimal. As noted, we are trying quite hard to have fewer exclusions. But there are very good reasons for continuing to study enriched populations (people who will have outcome events and disease of a severity that allows improvement to be detected)
SO,

- The discussion of theoretical and practical ways to make more use of RWD to gain evidence of effectiveness (?RWE) in randomized trials will cover a wide range of issues, including:
  - Data Quality
  - Clinical relevance
  - Practicality
  - And much, much more

- We are looking forward to the exciting program.
Leveraging Randomized Designs to Generate RWE
Leveraging Randomized Designs to Generate RWE

Experience from the Real World

Lesley H Curtis, PhD
Interim Executive Director, Duke Clinical Research Institute
Chair, Department of Population Health Sciences
Duke University School of Medicine
Strengthen the national capacity to implement cost-effective large-scale research studies that engage healthcare delivery organizations as research partners.

Over 819,000 participants across 869 clinical sites

15 demonstration projects spanning multiple Institutes and Centers
1-year planning phase
Implementation phase
Collaboratory projects span patient populations and interventions

<table>
<thead>
<tr>
<th>Baseline severity of participants</th>
<th>More severe</th>
<th>Therapeutic</th>
<th>Type of Intervention</th>
<th>Less severe</th>
<th>Operational/Educational</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HiLo</td>
<td>TiME</td>
<td>SPOT</td>
<td>GGC4H</td>
<td>PRIM-ER</td>
</tr>
<tr>
<td></td>
<td>ABATE</td>
<td>TSOS</td>
<td>PROVEN</td>
<td>LIRE</td>
<td>PROVEN-Pieces</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PPACT</td>
<td>ACP PEACE</td>
<td></td>
<td>Nudge</td>
</tr>
<tr>
<td></td>
<td></td>
<td>STOP CRC</td>
<td>ICD-Pieces</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

More severe

Less severe
To determine the effect of Albiglutide, when added to standard blood glucose lowering therapies, on major cardiovascular events in patients with type 2 diabetes mellitus

**Objective 1**
Understand how EHR data are used to facilitate trial recruitment and the barriers to that use.

**Objective 2**
Evaluate the fitness of EHR data for use in populating baseline characteristics in the eCRF

**Objective 3**
Evaluate the fitness of EHR data for use in identifying clinical endpoints
Patients with known ASCVD + ≥ 1 “enrichment factor”

- Identified through EHR (computable phenotype) by CDRNs (PPRN patients that are already a part of a CDRN are eligible to participate.)
- Patients contacted with trial information and link to e-consent:
  - Treatment assignment will be provided directly to patient

- Electronic follow-up: Every 3 or 6 months
  - Supplemented with EHR/CDM/claims data

- **Duration:** Enrollment over 39 months; maximum follow-up of 45 months

**Primary endpoint:**
- Composite of all-cause mortality, hospitalization for MI, or hospitalization for stroke

**Primary safety endpoint:**
- Hospitalization for major bleeding

ClinicalTrials.gov: NCT02697916
Experiences from the real world
The only constant is change

- Systems are complex, heterogeneous, and constantly changing.
  - Leadership, provider, and staff turn over
  - Underlying data and IT systems are not static

- Stability of control
  - Competing initiatives and priorities arise
  - Changes in usual care may be unethical to control
Real-world evidence generation is a team sport

- Clinician and health system engagement is critical
  - Partnerships required at all levels and local engagement is a must
  - Large, diverse systems require greater investment in (re)training

- Participant perspectives uniquely inform design, recruitment, and implementation.

- Ethical and regulatory systems engagement is important.
Leveraging existing data and systems may add complexity

- Integrating study-related data elements into the EHR has implications for clinical workflow and compliance.

- Data linkage for complete outcomes is technically straightforward, governance is often not.

- Data may not be available in a timely manner and latency often varies by site.
Implications for our work together

- The dynamic environment has consequences for design, implementation, and monitoring.
- Engagement of clinicians, systems, and participants can mitigate risk and increase the likelihood of success.
- Completeness, timeliness and quality of data in the real world introduces complexity.
Selecting Interventions and Study Designs to Generate RWE
Randomized clinical trials in the “Real World”

Martin Landray
University of Oxford
Robust assessment of health interventions

Most treatments have moderate effects

Distinguishing a moderate effect from no effect requires

RANDOMIZATION (to avoid balance)
+
SCALE (to avoid play-of-chance)
Clinical trials in crisis

• Rising cost & complexity of late-phase trials
  – 2 recent trials of PCSK9 inhibitors cost >$1Bn each
  – 85% commercial trials fail to recruit on time and to target
  – temptation to abandon randomization for lure of observational methods

• Distorted treatment development priorities
  – early decisions to continue treatment development based on limited data
  – away from preventive and long-term treatments for common diseases
  – focus on very expensive drugs for rare conditions
Robust assessment of health interventions

How can we take advantage of technological advances in healthcare, engineering & communications to facilitate randomized assessments of treatment efficacy & safety?
Opportunities for data-driven trials

• **Efficient recruitment** using routine clinical data

• **Effective assessment of safety & efficacy** using routine data (events) & digital technology (symptoms & function)

• **Excellent study quality** through protocol design, software engineering & statistical monitoring

• **Enhanced engagement** for participants & their doctors through integrated communication & consent approaches
Re-imagining the recruitment pathway

- Simple inclusion criteria
- Re-usable algorithms for identifying the right patients
- Innovative methods of pre-screening

1. Research question
2. Protocol design
3. Feasibility
4. Identify patients
5. Invite patients
6. Pre-screening
7. Consent

- Easy access to feasibility tools
- Accessible methods for large-scale invitation
- Digital tools for consent & engagement
National feasibility: Example ORION-4

ELIGIBILITY (Target 12,000 participants)

i) Diagnosis of **HEART ATTACK**
   Or
ii) Diagnosis of **STROKE**
   Or
iii) Surgery for **PERIPHERAL VASCULAR DISEASE**

Hospital Episode Statistics (HES) data held by NHS Digital can be searched using the diagnosis & procedure codes which match the protocol eligibility criteria.

i) Diagnosis of **HEART ATTACK**
   • i.e. ICD9 code: 410*, 412* and/or
   • ICD10 codes: I21*, I22*, I23*, I252 and/or
   • READ codes: G30*
   Or
ii) Diagnosis of **STROKE**
   • i.e. ICD9 codes: 433*, 434* and/or
   • ICD10 codes: I63*, I64* and/or
   • READ codes: G63*, G64*, G66* and/or
   Or
iii) Surgery for **PERIPHERAL VASCULAR DISEASE**
   • i.e. OPCS-4 procedure codes: L16*-28* inclusive, L48*-65* inclusive, L71*
National feasibility: Results for ORION-4

Target: 12,000 pts with prior MI, ischaemic stroke or peripheral revascularization

<table>
<thead>
<tr>
<th>Patients/hospital</th>
<th>Hospitals</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;10,000</td>
<td>96</td>
<td>~1.5M</td>
</tr>
<tr>
<td>&gt;15,000</td>
<td>54</td>
<td>~0.8M</td>
</tr>
<tr>
<td>&gt;20,000</td>
<td>26</td>
<td>~0.5M</td>
</tr>
</tbody>
</table>
Direct-to-patient invitation & recruitment in UK Biobank

- 9 million invitations generated from NHS register
- 500,000 volunteers recruited at 22 assessment centres in 3.5 years
- (89% in England, 7% in Scotland, 4% in Wales)

Similar methods used in multiple previous clinical trials involving >80,000 patients with cardiovascular disease and diabetes (1994-present)
Using routine data to ascertain outcomes in randomized controlled trials

Strengths:

• **Efficient**: enable larger, more “real-world” trials
• **Comprehensive**: minimize loss-to-follow-up
• **Durable**: enable prolonged study of safety & efficacy

Weaknesses:

• **Accessibility**: not all records are easy to access
• **Accuracy**: not all events are well coded
• **Confidence**: not all audiences or regulators are convinced
Assessment of treatment effect on clinical outcomes

Large randomized trials are resistant to small random errors in the data

<table>
<thead>
<tr>
<th></th>
<th>Active (10,000)</th>
<th>Control (10,000)</th>
<th>OR (&amp; 95%CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>True events</td>
<td>800</td>
<td>1000</td>
<td>0.78 (0.71-0.86)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Extra false</td>
<td>+20%</td>
<td></td>
<td>0.81 (0.74-0.89)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Missing real</td>
<td>-20%</td>
<td></td>
<td>0.79 (0.71-0.88)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Data do not need to be perfect
Assessment of treatment effect on clinical outcomes

Large *randomized* trials are resistant to small random errors in the data

<table>
<thead>
<tr>
<th></th>
<th>Active (10,000)</th>
<th>Control (10,000)</th>
<th>OR (&amp; 95%CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>True events</td>
<td>800</td>
<td>1000</td>
<td>0.78 (0.71-0.86)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Extra false events (evenly distributed)</td>
<td>+ 20%</td>
<td>980</td>
<td>1180</td>
<td>0.81 (0.74-0.89)</td>
</tr>
</tbody>
</table>

Data do not need to be perfect
Assessment of treatment effect on clinical outcomes

Large *randomized* trials are resistant to small random errors in the data

<table>
<thead>
<tr>
<th></th>
<th>Active (10,000)</th>
<th>Control (10,000)</th>
<th>OR (&amp; 95%CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>True events</td>
<td>800</td>
<td>1000</td>
<td>0.78 (0.71-0.86)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Extra false events (evenly distributed)</td>
<td>980</td>
<td>1180</td>
<td>0.81 (0.74-0.89)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Missing real events (unevenly distributed)</td>
<td>640</td>
<td>800</td>
<td>0.79 (0.71-0.88)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Data do not need to be perfect
The impact of outcome adjudication on estimates of treatment effect in randomized trial

- 16,580 medical documents (159,000 pages) retrieved, redacted & scanned
- 16,117 potential events adjudicated by 3 adjudicators & 17 assistants over 3 years

### Before

<table>
<thead>
<tr>
<th>Type of Event</th>
<th>Anacetrapib (N=15225)</th>
<th>Placebo (N=15224)</th>
<th>Rate Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of participants with events (%)</td>
<td>Anacetrapib Better</td>
<td>Placebo Better</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular death</td>
<td>481 (3.2)</td>
<td>510 (3.3)</td>
<td>0.94 (0.83–1.07)</td>
<td>0.009</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>589 (3.9)</td>
<td>694 (4.6)</td>
<td>0.85 (0.76–0.94)</td>
<td>0.032</td>
</tr>
<tr>
<td>Cardiovascular death or myocardial infarction</td>
<td>984 (6.5)</td>
<td>1099 (7.2)</td>
<td>0.89 (0.82–0.97)</td>
<td>0.009</td>
</tr>
<tr>
<td>Stroke</td>
<td>579 (3.8)</td>
<td>605 (4.0)</td>
<td>0.96 (0.85–1.07)</td>
<td></td>
</tr>
<tr>
<td>Cardiovascular death, myocardial infarction or stroke</td>
<td>1480 (9.7)</td>
<td>1594 (10.5)</td>
<td>0.93 (0.86–0.99)</td>
<td></td>
</tr>
</tbody>
</table>
The impact of outcome adjudication on estimates of treatment effect in randomized trial

• 16,580 medical documents (159,000 pages) retrieved, redacted & scanned

• 16,117 potential events adjudicated by 3 adjudicators & 17 assistants over 3 years

<table>
<thead>
<tr>
<th>Type of Event</th>
<th>Anacetrapib (N=15225)</th>
<th>Placebo (N=15224)</th>
<th>Rate Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of participants with events (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular death</td>
<td>520 (3.4)</td>
<td>564 (3.7)</td>
<td>0.92 (0.82–1.04)</td>
<td>0.008</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>669 (4.4)</td>
<td>769 (5.1)</td>
<td>0.87 (0.78–0.96)</td>
<td>0.93</td>
</tr>
<tr>
<td>Cardiovascular death or myocardial infarction</td>
<td>1057 (6.9)</td>
<td>1179 (7.7)</td>
<td>0.89 (0.82–0.97)</td>
<td>0.008</td>
</tr>
<tr>
<td>Stroke</td>
<td>547 (3.6)</td>
<td>559 (3.7)</td>
<td>0.98 (0.87–1.10)</td>
<td>0.045</td>
</tr>
<tr>
<td>Cardiovascular death, myocardial infarction or stroke</td>
<td>1497 (9.8)</td>
<td>1604 (10.5)</td>
<td>0.93 (0.87–1.00)</td>
<td>0.045</td>
</tr>
</tbody>
</table>
Effect of adjudication on estimate of treatment effect

<table>
<thead>
<tr>
<th>Trial</th>
<th>Experimental site adj</th>
<th>Control site adj</th>
<th>Weight</th>
<th>ROR [95%CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onsite assessor blinded</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topol 1993</td>
<td>31</td>
<td>97</td>
<td>0.283</td>
<td>0.78 [0.37, 1.63]</td>
</tr>
<tr>
<td>Shumaker 2004</td>
<td>31</td>
<td>26</td>
<td>0.248</td>
<td>0.84 [0.38, 1.68]</td>
</tr>
<tr>
<td>IMPACT-II 1997</td>
<td>12</td>
<td>134</td>
<td>0.098</td>
<td>0.57 [0.31, 1.07]</td>
</tr>
<tr>
<td>Winston 2012</td>
<td>22</td>
<td>14</td>
<td>0.138</td>
<td>1.08 [0.30, 2.58]</td>
</tr>
<tr>
<td>PURSUIT 2001</td>
<td>223</td>
<td>247</td>
<td>0.021</td>
<td>0.99 [0.74, 1.07]</td>
</tr>
<tr>
<td>Shumaker 2003</td>
<td>230</td>
<td>92</td>
<td>0.216</td>
<td>0.92 [0.84, 1.02]</td>
</tr>
<tr>
<td>Schulman 2004</td>
<td>209</td>
<td>30</td>
<td>0.447</td>
<td>0.93 [0.52, 1.66]</td>
</tr>
<tr>
<td>CHARM 2003</td>
<td>324</td>
<td>333</td>
<td>0.265</td>
<td>1.03 [0.74, 1.42]</td>
</tr>
<tr>
<td>McMurray 2006</td>
<td>796</td>
<td>587</td>
<td>0.676</td>
<td>0.98 [0.40, 2.12]</td>
</tr>
<tr>
<td>Fiesinger 2005</td>
<td>43</td>
<td>124</td>
<td>0.300</td>
<td>0.95 [0.47, 1.92]</td>
</tr>
<tr>
<td>O’Connor 2005</td>
<td>39</td>
<td>135</td>
<td>0.271</td>
<td>0.95 [0.45, 2.01]</td>
</tr>
<tr>
<td>GUSTO-IIb 1998</td>
<td>510</td>
<td>409</td>
<td>0.488</td>
<td>1.03 [0.68, 1.57]</td>
</tr>
<tr>
<td>Kinlaw 2007</td>
<td>1243</td>
<td>804</td>
<td>0.139</td>
<td>1.06 [0.65, 1.72]</td>
</tr>
<tr>
<td>CURE 2001</td>
<td>752</td>
<td>625</td>
<td>0.143</td>
<td>1.08 [0.74, 1.57]</td>
</tr>
<tr>
<td>HOPE 1998</td>
<td>513</td>
<td>278</td>
<td>0.223</td>
<td>1.12 [0.82, 1.54]</td>
</tr>
<tr>
<td>CREATE 2006</td>
<td>543</td>
<td>424</td>
<td>0.223</td>
<td>1.08 [0.82, 1.42]</td>
</tr>
<tr>
<td>OASIS-5 2006</td>
<td>566</td>
<td>570</td>
<td>0.233</td>
<td>0.75 [0.56, 1.00]</td>
</tr>
<tr>
<td>Bresalier 2005</td>
<td>77</td>
<td>45</td>
<td>0.213</td>
<td>1.04 [0.69, 1.61]</td>
</tr>
<tr>
<td>Wiviott 2007</td>
<td>328</td>
<td>251</td>
<td>0.430</td>
<td>1.15 [0.96, 1.36]</td>
</tr>
<tr>
<td>OASIS-2 1999</td>
<td>186</td>
<td>193</td>
<td>0.896</td>
<td>1.00 [0.78, 1.28]</td>
</tr>
<tr>
<td>OASIS-6 2000</td>
<td>569</td>
<td>583</td>
<td>0.488</td>
<td>0.95 [0.68, 1.34]</td>
</tr>
<tr>
<td>Bolland 2013</td>
<td>40</td>
<td>72</td>
<td>0.271</td>
<td>1.06 [0.55, 2.09]</td>
</tr>
<tr>
<td>PARAGON-B 2002</td>
<td>269</td>
<td>256</td>
<td>0.464</td>
<td>0.92 [0.68, 1.25]</td>
</tr>
<tr>
<td>Hata 2013</td>
<td>1018</td>
<td>851</td>
<td>0.345</td>
<td>1.07 [0.76, 1.50]</td>
</tr>
<tr>
<td>Ninomiya 2009</td>
<td>340</td>
<td>301</td>
<td>0.453</td>
<td>1.08 [0.80, 1.46]</td>
</tr>
<tr>
<td>BIP 2000</td>
<td>289</td>
<td>303</td>
<td>0.756</td>
<td>0.91 [0.71, 1.16]</td>
</tr>
<tr>
<td>HOPE 2000</td>
<td>895</td>
<td>654</td>
<td>0.038</td>
<td>1.12 [0.88, 1.43]</td>
</tr>
<tr>
<td>OPPORTUNITY 2078</td>
<td>506</td>
<td>535</td>
<td>0.194</td>
<td>1.10 [0.87, 1.39]</td>
</tr>
<tr>
<td>Wallentin 2009</td>
<td>504</td>
<td>533</td>
<td>0.194</td>
<td>1.10 [0.87, 1.39]</td>
</tr>
<tr>
<td>Naslund 1999</td>
<td>98</td>
<td>83</td>
<td>0.138</td>
<td>1.15 [0.88, 1.52]</td>
</tr>
<tr>
<td>EPIC 1994</td>
<td>64</td>
<td>59</td>
<td>0.184</td>
<td>1.04 [0.70, 1.58]</td>
</tr>
<tr>
<td>Prince 2006</td>
<td>28</td>
<td>18</td>
<td>0.135</td>
<td>1.25 [0.90, 1.74]</td>
</tr>
<tr>
<td>Verhamme 2013</td>
<td>16</td>
<td>20</td>
<td>0.161</td>
<td>1.05 [1.03, 1.09]</td>
</tr>
<tr>
<td>Hawke 2007</td>
<td>366</td>
<td>390</td>
<td>0.056</td>
<td>1.03 [0.78, 1.35]</td>
</tr>
<tr>
<td>RE Model for Subgroup</td>
<td>1.00</td>
<td>1.00</td>
<td>1.000</td>
<td>1.00 [0.96, 1.04]</td>
</tr>
</tbody>
</table>

Daikou L, Cochrane Methodology Review Group, 2016
Reliable results from electronic health records

ASCEND trial: Effect of (a) aspirin vs. placebo, and (b) omega-3 fatty acids vs. placebo on Vascular Events*

<table>
<thead>
<tr>
<th></th>
<th>Active</th>
<th>Placebo</th>
<th>Rate ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>812 (10.8%)</td>
<td>903 (12.0%)</td>
<td>0.89 (0.81-0.98)</td>
</tr>
<tr>
<td></td>
<td>692 (9.2%)</td>
<td>761 (10.2%)</td>
<td>0.90 (0.82-1.00)</td>
</tr>
<tr>
<td>Omega-3 fatty acids</td>
<td>855 (11.4%)</td>
<td>860 (11.4%)</td>
<td>1.00 (0.91-1.10)</td>
</tr>
<tr>
<td></td>
<td>723 (9.7%)</td>
<td>730 (9.7%)</td>
<td>1.00 (0.90-1.10)</td>
</tr>
</tbody>
</table>

Vascular Event: MI, ischaemic stroke, TIA, vascular death (exc. intracranial haemorrhage), or arterial revascularization

Electronic health record follow-up only
Routine data provide robust information on clinical efficacy during and for many years after clinical trial.

Long-term reduction in cumulative hospital admissions for coronary heart disease following 7 years of treatment with pravastatin vs. placebo (validation studies show trivial differences between routine collected hospital admission data versus expensively collected and adjudicated trial data).
Greater risk of transplant rejection and serious infection

<table>
<thead>
<tr>
<th></th>
<th>Sirolimus</th>
<th>Tacrolimus</th>
<th>Rate ratio (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunistic</td>
<td>11%</td>
<td>11%</td>
<td>1.00 (0.56-1.81)</td>
<td>0.9</td>
</tr>
<tr>
<td>Non-opportunistic</td>
<td>42%</td>
<td>31%</td>
<td>1.54 (1.11-2.15)</td>
<td>0.01</td>
</tr>
<tr>
<td>Any serious infection</td>
<td>48%</td>
<td>35%</td>
<td>1.51 (1.11-2.06)</td>
<td>0.008</td>
</tr>
</tbody>
</table>

No significant improvement kidney function at 18 months:
eGFR 53.7 mL/min/1.73 m² vs. 54.6 mL/min/1.73 m² (P = 0.50)

Haynes et al. Lancet 2014
Re-inventing randomized trials for the 21st Century

Improved feasibility driven by data & technology

PLUS

Adherence to principles of randomized trials

_SUPPORTED BY_

Proportionate approaches to trials regulations & guidance

FOR THE BENEFIT OF

Patient care and public health
Selecting Interventions and Study Designs to Generate RWE
Randomised controlled trials in the routine care setting: A Sponsor’s perspective

Dr Elaine A Irving PhD,
Head Real World Study Delivery, GSK
Outline

- Importance of ensuring the research question drives study design
- GSK Case Study: INTREPID: randomised controlled trial in the routine care setting.
- The GetReal Initiative: Challenges and opportunities for the conduct of randomised controlled trials in routine care settings
Translating efficacy to effectiveness: Identifying the key research questions and appropriate design

**Drivers of effectiveness**

- Health System Guidelines
- Physician
- Patient

**Effectiveness**

**What additional data is required?**
- Compared to the registration RCTs are there differences in the:
  - Population
  - Intervention
  - Comparator
  - Outcome
- Who are the relevant decision makers and what decision(s) do they need to make with the data?
- What endpoints have most relevance to clinical practice/public health?

**Case Study: INvestigation of TRELEGY**

**Effectiveness: usual Practice Design (ongoing)**

**Patient Population**

**Adherence**

**INTREPID**

**Objective:** To assess the effectiveness and safety of Ellipta FF/UMEC/VI (TRELEGY) compared to non-Ellipta MITT, in COPD patients managed in routine healthcare systems (EU: UK, Germany, Spain, Netherlands, & Sweden)

Randomised, open-label, parallel control arm phase IV relative effectiveness study conducted in the routine care setting

*Study was set up in countries following launch*
Considerations & Challenges: The balancing Act

Robust scientific design with maximum generalisability

Selection of endpoints
- Decision makers: CAT, FeV1, critical errors, exacerbations
- Treatment effect size

Randomisation vs non randomised design
- Choice driven by study objectives and endpoints

Open label vs blinding
- Impact on objectives & operational feasibility

Provision/Payment of drug
- Community Pharmacy/commercial supply

Data quality/consistency
- Use of centralised spirometry/eCRF

Choice of comparator

Safety monitoring

Study monitoring
Key Messages

- Comparative effectiveness research is of value to all key stakeholders
- Designing an effectiveness trial is complicated!
- Sponsor investment for a prospective study is significant with respect to time and cost. Clarity surrounding acceptance and impact of data is needed
- All stakeholders need to work together to enable this research to become embedded in the clinical development path
  - Recognition that this type of research answers different kinds of questions – some uncertainty may always remain and therefore needs a different approach to evaluation
  - A framework that enables routine care physicians to participate in research utilising data generated in the course of routine care
The GetReal Initiative: driving the use of RWE in healthcare decision making

Dr Elaine A Irving, GSK
GetReal Project Leader

Prof Dierdrick Grobbee, UMCU
GetReal Coordinator
The research leading to these results has received support from the EU/EFPIA Innovative Medicines Initiative (IMI) Joint Undertaking GetReal Initiative grant nº 807012.

The GetReal Initiative Partners

IMI Funding: €1.750.000,00
EFPIA Contribution: €1.350.688,00

Total project budget: €3.100.688,00
The GetReal Initiative

To increase the quality of RWE generation and increase adoption of RWE in drug development and regulatory/HTA assessment processes

- Standardisation of approaches
- Increased quality
- Tools to facilitate research
- Education & Training

Increased use and acceptance

The research leading to these results has received support from the EU/EFPIA Innovative Medicines Initiative [2] Joint Undertaking GetReal Initiative grant n° 807012.
Selecting Interventions and Study Designs to Generate RWE
A brief introduction to the PragMagic tool

Iris Goetz, Eli Lilly and Company, UK
Mira Zudigeest, University Medical Center Utrecht, NL
Pragmatic trials: challenges

Generate real world evidence in the early stages of drug development

Limited experience with pragmatic trials

Unexpected operational challenges

Major challenge
- concept
- protocol
- methodology
- feasibility
- stakeholder acceptability
PragMagic:
Decision support & educational tool for design and protocol evaluation of pragmatic trials.

https://www.pragmagic.eu/
What can PragMagic do for you?

- To facilitate the consideration & planning of pragmatic trials:
  - inform as team work where pragmatic choices could be made
  - become aware of possible operational challenges
  - assess expected generalizability of trial protocol

- For educational purposes:
  - Rasing awareness of pragmatic trial design options, operational challenges and solutions
How does it work?
PragMagic outputs

- PragMagic Fingerprint
- Excel sheets
Facilitate the design and planning of more pragmatic trials.

Giving insights into the consequences of design choices and operational challenges.

Maximise the generalisability of findings while ensuring validity and feasibility.

Randomised pragmatic trials with a drug component.

- A decision making tool
- A checklist to ensure compliance
- Quality verdict
THANK YOU!

Go to:
https://www.pragmagic.eu/

• Open in Google Chrome or Firefox and register a free account to access the tool

• Let us know if you would like more information or a free workshop!
Contact: M.G.P.Zuidegeest@umcutrecht.nl
Selecting Interventions and Study Designs to Generate RWE
mCODE and ICARE

Workshop on Point of Care Evidence
Duke - Margolis Center for Health Policy
11-12 July 2019
Washington, DC

Steven Piantadosi, M.D., Ph.D.
Brigham and Women’s Hospital
HMS Boston MA
• No set of filters will expand a trial population to PoC
• Unnecessary for many questions
• There are other important questions at PoC
Alternative Model

Point of Care Clinical Trials

Data Model

Point of Care

Structured Computable Data

ICARE Questions

Trial Population

Trial Results

mCODE: Minimal Common Oncology Data Elements

Case Report Forms

Access

Content

Eligibility

Exclusions

Data Model
ICARE: Develop and validate mCODE-based outcome measures embedded in the EHR

### Disease Status

**Clinical Assessment**
Based on the data available today (at the time of evaluation), categorize the patient’s disease extent.

**Question Format**

<table>
<thead>
<tr>
<th>Cancer disease status</th>
<th>&lt;lesion evaluated&gt;</th>
<th>&lt;status value&gt;</th>
<th>&lt;reason value&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>primary tumor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>metastatic lesion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>complete response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>partial response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stable disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>progressive disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>not evaluated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>imaging</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pathology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>physical exam</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>markers</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Treatment change

**Clinical Assessment**
Based on your evaluation today, are you making a change in treatment?

**Question Format**

<table>
<thead>
<tr>
<th>Treatment change</th>
<th>&lt;treatment change?&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Yes - disease not responding</td>
<td></td>
</tr>
<tr>
<td>Yes - due to A/E/toxicity</td>
<td></td>
</tr>
<tr>
<td>Yes - pre-planned therapy transition</td>
<td></td>
</tr>
<tr>
<td>Yes - patient request</td>
<td></td>
</tr>
<tr>
<td>Yes - due to other</td>
<td></td>
</tr>
</tbody>
</table>
Selecting Interventions and Study Designs to Generate RWE
Outcome Measurement Based on Real-World Data
Leveraging Real World Data in Clinical Trials: The RELIANCE* Trial Experience

* RofLumilast Versus Azithromycin Therapy to preveNt COPD Exacerbations

Elizabeth A Sugar, PhD
July 11, 2019
RELIENCE Trial

• PCORI-funded multi-center, randomized, parallel, non-inferiority trial
  • Treatments: RofLumilast (FDA approved) vs Azithromycin
  • Patients: 3200 with COPD hospitalized within last 12 months
  • Follow-up: BL, 3m, 6m then every 6m until event/end of trial

• Outcomes
  • Primary: All cause re-hospitalization or death
  • Secondary: medication adherence, cross-over, treatment discontinuation, emergency department and urgent care visits, NIH-PROMIS measurements, PRO outcomes, out of pocket expenses, weight,
  • AEs: e.g. hearing decrement, diarrhea, nausea, suicidal ideation
Data Collection Sources

TRIAL
- Patient Portal
- Investigator Portal
- Call Center Portal

EXTERNAL
- Site EMR
- National Death Index
- CMS Administrative Portal
**FDA Sentinel Collaboration**

- **Goals**
  - Validate and supplement trial data
    - Primary outcome: All cause hospitalizations and death
    - Secondary: medication adherence, cross-over, treatment discontinuation, emergency department and urgent care visits
  - Proof of concept for Distributed Regression Methods

- **Medicare Data**
  - First linkage in 2022 (interim): Annual 2019 & 2020 data
    - Inpatient claims files
    - Enrollment/death files
    - Part D prescription drug dispensing files
  - Second linkage (final): Annual 2021 & Quarterly 2022 (Q1-Q3)
    - Inpatient claims files
    - Enrollment/death files
    - Part D prescription drug dispensing files (only for Annual 2021)
Benefits and Challenges

• **Benefits**
  - Validation of primary outcome
  - Additional adherence data
  - Proof of concept for Distributed Regression Methods

• **Challenges**
  - Annual vs Quarterly files
    - Annual: 13-15 month lag after calendar year, 99.9% complete, Part D available
    - Quarterly: 4.5-6.5 month lag after end of quarter, 93% complete, no Part D
  - Only available for the subset who are on Medicare
  - Limited to subset of outcomes
Acknowledgements
Outcome Measurement Based on Real-World Data
Core Outcomes and Real-World RCTs

Duke Margolis – FDA workshop, Washington, DC. USA
Take Home Messages

• Measuring outcomes that matter most to decision makers (patients, clinicians, regulators, payers, caregivers) has emerged as a primary focus for evidence generation.

• No matter what type of study we are designing, we need to start with a clear understanding of which outcomes matter most.

• Once we know that, we can figure out which RWD elements are valid reflections of those outcomes, or figure out how we are going to add missing but meaningful outcomes data.
PCTs / RW-RCTs: Outcomes

• Peter Tugwell, J Rheumatology, 1993
  o “Clinical trials are only as credible as their outcomes”

• Tunis et al; Practical Clinical Trials, JAMA 2003
  o “measure outcomes of greatest relevance to decision makers”

• PCORI methodology committee standards
  o “identify and include outcomes the population of interest notices and cares about.”

• Califf and Sugarman, Pragmatic Trials, Clinical Trials 2015
  o focus on outcomes that are “directly relevant to participants, funders, communities, and healthcare practitioners”
Defining Value (as in VBHC)

• Health outcomes achieved per dollar spent
  o IOM 2006

• Health outcomes are inherently condition specific and multi-dimensional
  o Michael Porter, NEJM, 2010
## Ratings and Test Results

<table>
<thead>
<tr>
<th>Brand &amp; Model</th>
<th>Overall Score</th>
<th>Speed</th>
<th>Power</th>
<th>Run Time</th>
<th>Charge</th>
<th>Handling</th>
<th>Noise at ear</th>
<th>Weight (lbs)</th>
<th>Volts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hitachi DS19DMR</strong></td>
<td>85</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>Tougher job drill/drivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Makita LXT BDF451</strong></td>
<td>82</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td></td>
<td>⬤</td>
<td>4.9</td>
<td>18</td>
</tr>
<tr>
<td>Tougher job drill/drivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Milwaukee 0824-21</strong></td>
<td>81</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td></td>
<td>⬤</td>
<td>6.2</td>
<td>18</td>
</tr>
<tr>
<td>Tougher job drill/drivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panasonic EY6432GKXW</strong></td>
<td>80</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td></td>
<td>⬤</td>
<td>4.8</td>
<td>15.6</td>
</tr>
<tr>
<td>General use drill/drivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bosch 33613-20</strong></td>
<td>80</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td></td>
<td>⬤</td>
<td>5.6</td>
<td>18</td>
</tr>
<tr>
<td>Tougher job drill/drivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Makita 6347DADE</strong></td>
<td>79</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td></td>
<td>⬤</td>
<td>5.4</td>
<td>18</td>
</tr>
<tr>
<td>Tougher job drill/drivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RYOBI P013</strong></td>
<td>77</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td></td>
<td>⬤</td>
<td>4.5</td>
<td>18</td>
</tr>
<tr>
<td>General use drill/drivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Makita 6900DWADE</strong></td>
<td>75</td>
<td></td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>3.6</td>
<td>12</td>
</tr>
<tr>
<td>Cordless impact drivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RYOBI P230C</strong></td>
<td>74</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>⬤</td>
<td>4.6</td>
<td>10</td>
</tr>
<tr>
<td>Cordless impact drivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brand &amp; Model</td>
<td>Overall Score</td>
<td>Speed</td>
<td>Power</td>
<td>Run Time</td>
<td>Charge</td>
<td>Handling</td>
<td>Noise at ear</td>
<td>Weight (lbs)</td>
<td>Volts</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------</td>
<td>-------</td>
<td>-------</td>
<td>----------</td>
<td>--------</td>
<td>----------</td>
<td>-------------</td>
<td>--------------</td>
<td>-------</td>
</tr>
<tr>
<td>Hitachi DS18DMR</td>
<td>85</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>Makita LXT BDF451</td>
<td>82</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.9</td>
</tr>
<tr>
<td>Milwaukee 0824-21</td>
<td>81</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6.2</td>
</tr>
<tr>
<td>Panasonic EY6432GQK</td>
<td>80</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.8</td>
</tr>
<tr>
<td>Bosch 33813-20</td>
<td>80</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>18</td>
</tr>
<tr>
<td>Makita 6347DADE</td>
<td>79</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ryobi P013</td>
<td>77</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>18</td>
</tr>
<tr>
<td>Makita 6930FDVDE</td>
<td>76</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.6</td>
</tr>
<tr>
<td>Ryobi P230C</td>
<td>74</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.8</td>
</tr>
</tbody>
</table>
Core Outcome Sets

• “An agreed standardised set of outcomes that should be measured and reported, as a minimum, in all clinical research in specific areas of health or health care”
  o Definition from the COMET Initiative
coreHEM

A Core Outcome Set for Gene Therapy in Hemophilia
coreHEM Stakeholders

- clinicians
- government reps
- industry sponsor reps
- US payers
- international payers/HTA
- patients/patient advocates
- academic gene therapy research reps
- methods and epidemiology experts
**coreHEM FINAL CORE SET**

<table>
<thead>
<tr>
<th>Consensus to Select</th>
<th>≥70% of all voters rated the outcome with a score of 7-9 (critical importance).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus to Select (Patient-Important)</td>
<td>&lt;70% of all voters rated 7-9, but the stakeholders in the patient group gave the outcome an average rating of ≥7.</td>
</tr>
</tbody>
</table>

**3 SELECTED**

**8 ADVERSE EVENTS SELECTED**

**3 SELECTED**
HIGH STAKEHOLDER CONSENSUS

Frequency of Bleeds

Rating of Importance


Stakeholder Category
OUTCOME RETAINED DUE TO PATIENT IMPORTANCE
Take Home Messages

• No matter what type of study we are designing, we need to start with a clear understanding of which outcomes matter most

• Suggest caution in moving too far from outcome relevance in service of feasibility
Outcome Measurement Based on Real-World Data
Outcomes in Real-World Data

David Madigan
Columbia University
Terms

Actual outcome

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>0</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Misclassification rate: \( \frac{b + c}{a+b+c+d} \)

Sensitivity: \( \frac{a}{a+c} \) (aka recall)

Specificity: \( \frac{d}{b+d} \)

Positive predictive value: \( \frac{a}{a+b} \) (aka precision)
### OMOP: ‘Ground truth’ for Monitoring Health Outcomes of Interest

#### Drug Classes
- **ACE Inhibitors**
- **Amphotericin B**
- **Antibiotics:**
  - erythromycins, sulfonamides, tetracyclines
- **Antiepileptics:**
  - carbamazepine, phenytoin
- **Benzodiazepines**
- **Beta blockers**
- **Bisphosphonates:**
  - alendronate
- **Tricyclic antidepressants**
- **Typical antipsychotics**
- **Warfarin**

#### Outcome Categories
- **Angioedema**
- **Aplastic Anemia**
- **Acute Liver Injury**
- **Bleeding**
- **Hip Fracture**
- **Hospitalization**
- **Myocardial Infarction**
- **Mortality after MI**
- **Renal Failure**
- **GI Ulcer Hospitalization**

#### Legend
- **True positive’ benefit**
- **True positive’ risk**
- **Negative control’**

#### Total
- **Legend**
- **Total**
  - True positive’ benefit: 2
  - True positive’ risk: 9
  - Negative control’: 44
Acute liver injury definitions

1. Occurrence of at least one broad diagnosis code

2. Occurrence of at least one narrow diagnosis code

3. Occurrence of at least one narrow diagnosis code
   AND (diagnostic procedure <=30d before
   OR treatment procedure >=60d after)

4. Occurrence of at least one narrow diagnosis code
   AND (diagnostic procedure <=30d before
   OR treatment procedure >=60d after)
   AND laboratory results indicative of Hy's law:
   ALT >= 3xULN AND AST >= 3xULN AND Bilirubin >= 2xULN
   within 7 days

5. Laboratory results indicative of Hy's law:
   ALT >= 3xULN AND AST >= 3xULN AND Bilirubin >= 2xULN
   within 7 days

6. Laboratory results strongly indicative of Hy's law:
   ALT >= 10xULN AND AST >= 10xULN AND Bilirubin >= 2xULN
   within 7 days
Estimate sensitivity of RR to HOI definition: Acute Liver Injury

More specific definitions have fewer cases, which increases variability. Estimates are generally stable. Differences observed between HOI definitions vary by method.
Welcome to OHDSI!

The Observational Health Data Sciences and Informatics (OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. All our solutions are open-source.

OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University.

Read more about us, about our goals, and how you can help support the OHDSI community.

Join the Journey

Events
- 2019 OHDSI Symposium
- 2019 slides
- North Simmons

OHDSI on YouTube

Latest News
- Weekly OHDSI Digest - July 16, 2019
- OHDSI Provides Critical Update, Links Rare Skin Study That Proves Improved Results
- OHDSI Center for Surgical
-ワーク流アナライザー is the KIO, a leading organization for improving health care delivery
- Weekly OHDSI Digest - June 26th, 2019
Electronic Health Records ~ 700M
Shiny App Viewer

data.ohdsi.org/PhenotypeLibraryViewer/

Author Submission Template (Example)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotype Title</td>
<td>Rheumatoid Arthritis</td>
</tr>
<tr>
<td>Author(s) and Affiliations</td>
<td>Jane Doe, Example University, John Doe, Example University</td>
</tr>
<tr>
<td>Date of Submission</td>
<td>March 21, 2019</td>
</tr>
<tr>
<td>Modality</td>
<td>Computable</td>
</tr>
</tbody>
</table>

Source Data

<table>
<thead>
<tr>
<th>Link Type</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotype GitHub Page</td>
<td><a href="https://www.github.com">https://www.github.com</a></td>
</tr>
<tr>
<td>Implementation File</td>
<td><a href="https://www.github.com">https://www.github.com</a></td>
</tr>
<tr>
<td>Hash of Implementation File</td>
<td>4a5c0ec985b2b5bd50905a42a612f505a20</td>
</tr>
<tr>
<td>Configuration File</td>
<td><a href="https://www.github.com">https://www.github.com</a></td>
</tr>
</tbody>
</table>

Development

Purpose and Intended Use

This definition is intended to capture patients with a first-observed diagnosis of chronic rheumatoid arthritis (RA), taking care to rule out patients with short-term joint pain or fibromyalgia. Please note this definition is intended to be used with US-only data.

Development Methodology
Combining OHDSI Toolsets

Aphrodite (Juan Banda)

https://github.com/OHDSI/Aphrodite

- **Can create phenotypes** probabilistically by learning good phenotypes from a set of noisy labels

- Built to interface with the OMOP CDM to automatically create and utilize features using all data in your CDM (or a subset, if you choose)

- Machine learning takes into account more features than what could be considered by hand, and labeling heuristic is less time consuming

- Performs internal validation and is easy to share (config file tracks how it was built; binary object output tracks the definition itself)
Combining OHDSI Toolsets

PheValuator (Joel Swerdel)
https://github.com/OHDSI/PheValuator

- **Can evaluate phenotypes** to see how well they perform, offering an alternative to low-powered and time-consuming clinical review

- Uses a diagnostic predictive model to assign a large sample of people a predicted probability of having the condition

- Assess “Truth” based on an extremely specific cohort (xSpec) or extremely sensitive cohort (xSens)

- Produces *all* metrics (not just PPV) for a complete understanding of phenotype definition performance

- Like Aphrodite, will automatically output documentation needed for being a Gold Standard Process.
Measurement Error is Here to Stay!

• Estimating sensitivity/specificity/PPV is a good thing because it allows for you to correct for the error that exists in your data

• Estimating PPV alone is not sufficient, because you can't correct for measurement error if you only have a PPV

• Ignoring measurement error is bad practice. So is reporting about error but not correcting for it
Outcome Measurement Based on Real-World Data

Join the conversation with #RWEregulatory
RCTs in Health Care Settings: Some Issues and Suggested Approaches

William Crown, PhD
Chief Scientific Officer, OptumLabs
High Level Concepts

• Conducting RCTs within health care environments addresses many design limitations associated with observational studies
• But still need to be concerned about non-random attrition, medication adherence, etc.
• Offers ability to combine prospectively collected data with EMR, claims, and other data
• Data quality remains a key issue
Many RWE Guidance Documents


Some Important Data Quality Issues

- Leakage in EMRs
- Unstructured data in EMRs
- Mortality data
Variability in missing data and leakage by provider: Administration of aspirin during MI

<table>
<thead>
<tr>
<th>Source</th>
<th># of AMI hospitalizations</th>
<th>Any meds admin (%)</th>
<th>Meds admin and aspirin use (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S0002</td>
<td>483</td>
<td>100.0</td>
<td>97.3</td>
</tr>
<tr>
<td>S0008</td>
<td>6,708</td>
<td>82.9</td>
<td>85.5</td>
</tr>
<tr>
<td>S0009</td>
<td>7,328</td>
<td>95.1</td>
<td>90.0</td>
</tr>
<tr>
<td>S0019</td>
<td>4,150</td>
<td>92.9</td>
<td>92.5</td>
</tr>
<tr>
<td>S0020</td>
<td>1</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>S0034</td>
<td>10,351</td>
<td>91.5</td>
<td>94.7</td>
</tr>
<tr>
<td>S0039</td>
<td>10,920</td>
<td>99.5</td>
<td>92.1</td>
</tr>
<tr>
<td>S0041</td>
<td>15,227</td>
<td>84.3</td>
<td>40.2</td>
</tr>
<tr>
<td>S0042</td>
<td>535</td>
<td>99.8</td>
<td>91.9</td>
</tr>
<tr>
<td>S0048</td>
<td>3,657</td>
<td>28.4</td>
<td>92.4</td>
</tr>
<tr>
<td>S0057</td>
<td>2,584</td>
<td>73.7</td>
<td>69.3</td>
</tr>
<tr>
<td>S0058</td>
<td>1,055</td>
<td>89.8</td>
<td>30.7</td>
</tr>
<tr>
<td>S0062</td>
<td>2,187</td>
<td>54.1</td>
<td>95.3</td>
</tr>
<tr>
<td>S0068</td>
<td>2,217</td>
<td>81.8</td>
<td>94.3</td>
</tr>
<tr>
<td>S0069</td>
<td>1,748</td>
<td>90.7</td>
<td>94.5</td>
</tr>
<tr>
<td>S0070</td>
<td>2,375</td>
<td>96.4</td>
<td>90.6</td>
</tr>
<tr>
<td>Overall</td>
<td>101,388</td>
<td>88.0</td>
<td>83.0</td>
</tr>
</tbody>
</table>

Altan, Lin. EHR Data Quality: Use in Regulatory Decisions, Population Health, and Outcomes Research
NASEM Workshop on Understanding Health Care Quality. Washington, DC, October 15, 2018
## Using NLP to Identify Patient-reported measures in an EHR database

<table>
<thead>
<tr>
<th>Functional assessments</th>
<th>Anxiety</th>
<th>Depression</th>
<th>Cognition assessments</th>
<th>Walk/gait/ balance/fall risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADL</td>
<td>GAD</td>
<td>PHQ-9</td>
<td>MMSE</td>
<td>Balance</td>
</tr>
<tr>
<td>ECOG</td>
<td>GAD-7</td>
<td>GDS</td>
<td>Clock drawing</td>
<td>Coordination</td>
</tr>
<tr>
<td>IADL</td>
<td>PHQ-2</td>
<td>Mini-COG</td>
<td></td>
<td>Gait</td>
</tr>
<tr>
<td>FIM</td>
<td>PHQ</td>
<td>BNT</td>
<td></td>
<td>Standing</td>
</tr>
<tr>
<td>Functional activity</td>
<td>Depression screen</td>
<td>Boston naming test</td>
<td>Walks</td>
<td></td>
</tr>
<tr>
<td>Karnofsky</td>
<td>Depression</td>
<td></td>
<td>Balance and gait</td>
<td></td>
</tr>
<tr>
<td>BADL</td>
<td></td>
<td></td>
<td></td>
<td>Berg balance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tinetti balance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gait speed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6-minute walk test</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hendrich II</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dynamic gait index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tinetti gait</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tinetti balance and gait</td>
</tr>
</tbody>
</table>

Altan, Lin. EHR Data Quality: Use in Regulatory Decisions, Population Health, and Outcomes Research
NASEM Workshop on Understanding Health Care Quality. Washington, DC, October 15, 2018
Table 4. Population with valid MMSE results

<table>
<thead>
<tr>
<th></th>
<th>Unique patient count</th>
<th>Cumulative percent</th>
<th>Percent of previous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique baseline patients</td>
<td>46,101,387</td>
<td>100</td>
<td>NA</td>
</tr>
<tr>
<td>With MMSE test any time</td>
<td>179,904</td>
<td>0.3902</td>
<td>0.3902</td>
</tr>
<tr>
<td>With MMSE during 2012–15</td>
<td>126,244</td>
<td>0.2738</td>
<td>70.17</td>
</tr>
<tr>
<td>… and with valid value</td>
<td>107,346</td>
<td>0.2328</td>
<td>85.03</td>
</tr>
</tbody>
</table>

Table 5. Sample view of data after cleaning

<table>
<thead>
<tr>
<th>optum_lab_id</th>
<th>MEAS..._TYPE</th>
<th>MEAS..._DETAIL</th>
<th>MEAS..._DATE</th>
<th>MEAS..._VALUE</th>
<th>idn_indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000aaa</td>
<td>MMSE</td>
<td>30</td>
<td>2014-02-17</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>10008bbb</td>
<td>MMSE</td>
<td>30</td>
<td>2012-03-28</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>10010ccc</td>
<td>MMSE</td>
<td>30</td>
<td>2012-05-09</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>10019ddd</td>
<td>MMSE</td>
<td>NA</td>
<td>2013-12-02</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>100304eee</td>
<td>MMSE</td>
<td>30</td>
<td>2012-04-19</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>100309fff</td>
<td>MMSE</td>
<td>30</td>
<td>2013-02-11</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>100345ggg</td>
<td>MMSE</td>
<td>30</td>
<td>2012-09-19</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>100345ggg</td>
<td>MMSE</td>
<td>30</td>
<td>2013-10-11</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>100345ggg</td>
<td>MMSE</td>
<td>30</td>
<td>2013-07-02</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>100345ggg</td>
<td>MMSE</td>
<td>30</td>
<td>2013-10-11</td>
<td>26</td>
<td>0</td>
</tr>
</tbody>
</table>
Using Deep Learning for Predictive Modeling

Scalable and accurate deep learning with electronic health records


Predictive modeling with electronic health record (EHR) data is anticipated to drive personalized medicine and improve healthcare quality. Constructing predictive statistical models typically requires extraction and transformation of data from raw EHR data, which can be a labor-intensive process that discards the vast majority of the information in each patient’s record. We propose a representation of patients’ entire raw EHR records based on the Fast Healthcare Interoperability Resources (FHIR) format. We present a machine learning approach for accurately predicting multiple medical events from simpler, more patient-specific data harmonization. We validated our approach using de-identified EHR data from two U.S. academic medical centers with 316,221 adult patients hospitalized for at least 24 h. In the sequential format, our approach, which reduces the volume of EHR data to a total of 46,864,534 data points, including clinical notes. Deep learning models achieved high accuracy for tasks such as predicting in-hospital mortality (area under the receiver operating curve (AUROC) across sites 0.81–0.94), 30-day unplanned readmissions (AUROC 0.75–0.76), prolonged length of stay (AUROC 0.85–0.96), and all of a patient’s final discharge diagnoses (frequency-weighted AUROC 0.90). These models outperformed traditional, clinically-used predictive models in all cases. We believe that this approach can be used to create accurate and scalable predictions for a variety of clinical scenarios. In a case study of a particular prediction, we demonstrate that neural networks can be used to identify relevant information from the patients’ chart.
Policy Commentary

Sources for Healthcare Quality and Research
Expansion of Records From the Death Master File

Roza, JD, MS; Giovanni Filardo, PhD, MPH; David Nicewander, MS

Since 2002, this is a "demise of a vital resource" that will
hinder healthcare outcomes research, as well as the abilities
of healthcare organizations to monitor performance on patient
outcome measures. The latter could become critical to hos-
pital financing with the advent of the Centers for Medicare
and Medicaid Services Hospital Readmissions Reduction
Programs and the addition of patient outcome measures to
the performance standards applied in its Hospital Value-Based
Purchasing Program. In addition, given that two vital ele-
ments of the current round of health reform—value-based
purchasing and comparative effectiveness research to identify
the most effective and efficient treatment or disease man-
agement strategies—rely heavily on the ability to accurately
measure or track mortality of study subjects or patient cohorts,
the ramifications of this change to the DMF deserve attention.

What Is the DMF?
The SSA DMF is a public-use file that the SSA created to meet
Freedom of Information Act requests from members of the
public seeking vital status data.7 The DMF contains the Social
Security number, first, middle, and last names; dates of birth
and death; and last known ZIP code for individuals who

OPTUM Labs
Underestimation of Survival Rates When Linking Claims Data to Social Security Administration’s Date of Death

Jessica Chang, MD; William H. Crown, PhD; Henry J. Allen, PhD
OptumLabs, Cambridge, MA; OptumLabs, Eden Prairie, MN

Background:
- Administrative claims and electronic health record databases usually contain comprehensive information about health care encounter and prescription drug use, but lack vital status information on death.
- The Social Security Administration (SSA) is 2 primary sources of vital information. SSA data from 1985 onward includes detailed death records for deceased beneficiaries, including cause of death.

Methods:
- We created a data linkage of claims data from the Optum, the largest commercial US healthcare services company, with Social Security Administration death records for the same time period.
- We used a separate dataset of patients with long cancer and heart disease (CHD) to validate the accuracy of the linkage process.

Results:
- We identified 29,143 decedents with long-term cancer patients and 16,252 decedent with CHD.
- We found that the SSA death data was more accurate than the claims data.

Conclusions:
- Our analysis shows that linking claims data to Social Security death records is a valuable tool for researching the impact of cancer and heart disease on mortality rates.

References:
- [Provide references here]

Impact of the SSA DIA policy change:
- Results of death ascertainment in administrative claims data and the SSA DIA, which is used to determine Social Security Disability Insurance benefits, have been shown to be accurate but have limitations.

Changes in the SSA DIA policy change have been shown to impact the number of deaths identified. The SSA DIA policy change in 2011 has had a significant impact on identifying deaths with less accuracy.

For more information, refer to the full manuscript.
Things that can Help

• Missing data due to leakage
  – Assessment of problem through claims/EMR linkage
  – Focus on sites with more complete data
  – Distinction between signal detection and causal inference
    • ML methods focused on signal detection may work pretty well with incomplete data

• Unstructured data
  – Linkage to registries
  – Primary data collection
  – NLP
  – Deep learning

• Mortality data
  – Linkage to registries
  – Primary data collection
  – Legislation to allow research use of Social Security state DMF
  – Construction of composite indicators
  – Missing mortality data may be less of an issue for relative risk
Key Considerations for Blinding in Randomized Real-World Studies
Key Considerations for Blinding in Randomized Real-World Studies

Simon Skibsted, MD, MPH, PhD
Director, Clinical Development and Outcomes Research
Novo Nordisk Inc
Agenda

1. Background for conducting a trial with a randomization component in a real-world healthcare setting
2. The SEPRA trial
3. Blinding setup for the SEPRA trial
4. Considerations for blinding in the SEPRA trial
Background for conducting a trial with a randomization component in a real-world healthcare setting

- Demonstrate the effectiveness of semaglutide s.c. (once weekly GLP-1 RA) in an externally valid setting of a US healthcare system while maintaining the internal validity of a randomized clinical trial

- Assess novel endpoints related to claims data such as healthcare resource utilization and pharmacy prescription activity

GLP-1 RA, glucagon-like peptide-1 receptor agonist
The SEPRRA Trial - Key Characteristics

- Randomization
- Increased external validity (population and setting)
- Comparator: Standard of care
- Minimal intervention during trial
  - Linkage of claims data
  - Few trial visits and assessments
  - Data collection & monitoring reflects usual care
- Not for regulatory purposes
The SEPRA Trial rationale & objectives

**Trial rationale**
- To inform clinical practice on the comparative effectiveness of semaglutide s.c. versus standard of care in a real-world setting in adult patients with type 2 diabetes

**Trial objectives**
- To investigate long-term comparative effectiveness of semaglutide s.c. versus standard of care according to local clinical practice on:
  - Glycaemic control
  - Body weight
  - Healthcare resource utilisation
The SEPRA trial

Inclusion criteria
- Age ≥18 years
- T2DM
- Inadequately controlled* on up to two OADs

*Inadequately controlled as defined by HCP

Randomization

Semaglutide s.c. once weekly

Standard of care

Duration: 24 months
Few pre-planned visits and assessments in addition to local clinical practice

1 year visit
2 year visit

Trial information
- US only
- Phase 4
- Open label
- Payer partnership
- Multicenter within payer network

- Treatments can be adjusted according to local clinical practice. Switch to, add-on, or discontinuation of anti-diabetic treatment is permitted
- Only switch from standard of care to semaglutide s.c. is not permitted

T2DM, type 2 diabetes mellitus; US, United States; HCP, Healthcare Professional; OAD, Oral Antidiabetic Drug; S.C., subcutaneous
The SEPRA Trial - Selected Endpoints

**Primary endpoint**
- HbA1c <7.0% at year 1

**Other selected effectiveness endpoints**
- Individualised HbA1c target
- Healthcare resource utilisation
- Work productivity
- Medication persistence and adherence
- Number of hypoglycaemic episodes leading to inpatient hospitalisation or emergency room encounter
Blinding setup for the SEPRa trial

Open label approach for patients and physicians

- All drugs investigated in this trial are approved for the treatment of type 2 diabetes

- Once a patient is randomized, the patient will receive a prescription for the drug and pick up the drug at their pharmacy of choice

- In order to minimize the impact of any differential out-of-pocket costs between the treatment groups, Novo Nordisk covers any drug cost above a fixed out-of-pocket amount.
Blinding setup for the SEPRA trial

Novo Nordisk personnel is blinded until DBL

- Randomization & blinding plan
- Statistical analysis plan
Blinding considerations for the SEPRRA trial

**Scientific**
- Level of pragmatism compromised by blinding
- Blinding will lead to deviation from usual clinical practice
- Objective endpoints
- Effectiveness – Hawthorne effect

**Operational**
- Comparator is standard of care
- A maximum out of pocket cost is set
- Using USPI as reference safety information
- Patients will use their own pharmacy – no trial product labelling
Key Considerations for Blinding in Randomized Real-World Studies
Key Considerations for Blinding in Randomized Real-World Studies

Rita F. Redberg, MD, MSc, FACC, FAHA
University of California, San Francisco
Disclosures & Relevant Relationships

Disclosures
- No financial conflicts

Relevant Relationships
- Editor, JAMA Internal Medicine
Is it ethical not to do placebo control?

- Power of placebo is strong
  - Especially for procedures and devices
- FDA approval should assure safety and effectiveness
- Coverage follows
  - No incentive then to randomize
- If devices are later found to be ineffective or unsafe, they cannot be easily removed

Sham controls in Medical Device Trials Redberg RF. *NEJM* 2014
"When former FDA commissioner Robert Califf was asked whether sham controls should be required for device approval, he thought that it was more of a decision for the clinical community: 'Do you want to get the truth or not?'"
Putting the Cart Before the Horse

- ORBITA was the first placebo-controlled study of PCI (cardiac stents) in the 40-year history of the procedure.
- It found no benefit of PCI in addition to medical therapy in relief of angina, exercise time or quality of life.
- Multiple trials have already shown no benefit of PCI compared to medical therapy on MI or mortality.
- It teaches us:
  - that we have to broaden our paradigm of ischemia and angina beyond the “clogged” artery – medical therapy is first line.
  - Placebo controls should be mandatory in device trials.
  - There is multifactorial resistance to accepting changes to the existing paradigm: financial, intellectual, mistrust of science, etc.
Regulatory oversight has not kept pace with increasing complexity of medical device technology

21st Century Cures Act (Dec 2016)
- Allows companies to submit anecdotal instead of scientific evidence
- Shifts burden of evidence to postmarketing
- Subjects millions of Americans to unsafe or untested medical devices
Postapproval Studies – Small and Slow

- Postapproval studies (PAS)
  - Small sample size
  - Mostly prospective cohort studies
  - Only 26% completed
  - Avg 3 years to complete


Opinion

80,000 Deaths. 2 Million Injuries. It’s Time for a Reckoning on Medical Devices.

Patients suffer as the F.D.A. fails to adequately screen or monitor products.

By The Editorial Board
The editorial board represents the opinions of the board, its editor and the publisher. It is separate from the newsroom and the Op-Ed section.

May 4, 2019
THANK YOU!
When Science Fails You, Try Faith

I have no doubt that [insert favorite parable] actually happened. I just can’t prove it.
Key Considerations for Blinding in Randomized Real-World Studies
### Need for Masking (Blinding) Treatment Depends on Magnitude of Potential for Bias

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Potential for Bias</th>
<th>Masking Recommended</th>
<th>Other Remediations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional and/or health state reported by patient</td>
<td>High, especially when patients perceive one treatment as better</td>
<td>Yes, if primary endpoint</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Depending on perception of treatment benefit, clinicians also may be masked if interacting with patients frequently during study</td>
</tr>
<tr>
<td>Event, sign or outcome reported by clinician</td>
<td>Low</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Central adjudication panel or clinical assessor (not treating doc) could be blinded or RW proxies could be used</td>
</tr>
<tr>
<td>Event diaries from patients (counts of observable signs)</td>
<td>Low, since events assessed are typically less influenced by treatment knowledge</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>---</td>
</tr>
</tbody>
</table>

## Need for Masking (Blinding) Treatment Depends on Magnitude of Potential for Bias

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Potential for Bias</th>
<th>Masking Recommended</th>
<th>Other Remediations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry, Hematology</td>
<td>Moderately low, particularly if measures obtained from clinical chart and not patient-reported</td>
<td>No</td>
<td>---</td>
</tr>
<tr>
<td>Imaging +/- (survival, clinical signs)</td>
<td>Moderately low; bias may occur if reader is aware of treatment and expects some benefit</td>
<td>No</td>
<td>Centralized readers can be masked to treatment assignment</td>
</tr>
<tr>
<td>Microbiology</td>
<td>Low</td>
<td>No</td>
<td>---</td>
</tr>
<tr>
<td>Physiologic/functional measures</td>
<td>Low</td>
<td>No</td>
<td>Use wearables as feasible</td>
</tr>
</tbody>
</table>

Regulators around the world are becoming more sophisticated in their understanding of when blinding is essential

*New draft guidance from China on using RWE for Regulatory Decision Support*

Chinese FDA on Pragmatic Randomized Trials (PCT)

• “Since PCTs are conducted in a setting close to real clinical practice, the evidence obtained by PCTs is considered as the most reasonable and practical real-world evidence compared to other research types.”

• “PCTs do not adopt blinding in most cases; therefore sufficient attention should be paid in estimating and adjusting the resulting detection bias.”

Source: Key considerations in using real-world evidence to support drug development, Draft for public review. Dated May, 2019 from the Center for Drug Evaluation, NMPA, China
Nancy A. Dreyer, PhD MPH FISPE

Chief of Scientific Affairs & SVP
Head, Center for Advanced Evidence Generation

IQVIA Real-World & Analytic Solutions
Office: + 1 617 715 6810
nancy.dreyer@IQVIA.com