Maximizing the Utility of Real World Evidence

Integration of Structured EMR Data, Unstructured EMR Data, and Billing Data for Economics and Outcomes Research in Oncology

ISPOR 20th Annual International Meeting
Philadelphia, PA, USA
Tuesday, May 19, 2015
Workshop Session III

Maximizing the Utility of Real World Evidence

Integration of Structured EMR Data, Unstructured EMR Data, and Billing Data for Economics and Outcomes Research in Oncology

Introduction: Real World Evidence in Oncology Outcomes Research
Arliene Ravelo, MPH
Associate Director, *Genentech*, South San Francisco, CA

Selection of Data Sources to Optimize ROI
Kathy Schulman, MS
Principal, *Outcomes Research Solutions*, Waltham, MA

Role of Unstructured EMR Data in Outcomes Research
Mark S. Walker, PhD
Vice President of Scientific Affairs, *Vector Oncology Solutions*, Memphis, TN

EMR and Billing Data in HRU and Cost Analyses
Kim Saverno, PhD
Director of Pharmacoeconomics, *Vector Oncology Solutions*, Memphis, TN*

Conclusions: Real World Evidence in Oncology Outcomes Research
Arliene Ravelo, MPH

*Now with Comprehensive Health Insights, Louisville, KY

Maximizing the Utility of Real World Evidence

Integration of Structured EMR Data, Unstructured EMR Data, and Billing Data for Economics and Outcomes Research in Oncology

Financial Disclosures

Arliene Ravelo, MPH, Associate Director, *Genentech*
- Employment at Genentech
- Stock in Roche

Kathy Schulman, MS, Principal, *Outcomes Research Solutions*
- No financial disclosure

Mark S. Walker, PhD, Vice President of Scientific Affairs, *Vector Oncology*
- Genentech funding to Vector Oncology
- No other disclosures

Kim Saverno, PhD, Director of Pharmacoeconomics, *Vector Oncology*
- Genentech funding to Vector Oncology
- No other disclosures

ISPOR 20th Annual International Meeting, Philadelphia, PA, May 19, 2015

Using Real World Evidence for Oncology Outcomes Research

Arliene Ravelo, MPH
Associate Director
BioOncology Health Economics & Outcomes Research
Genentech

USMA US Medical Affairs

HEOR Health Economics & Outcomes Research

Research Question #1 – PROs and ETDS

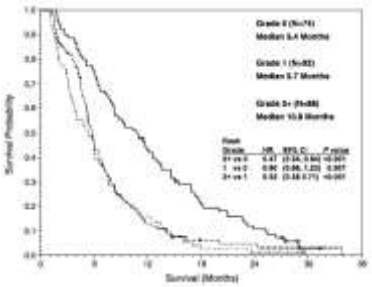
- **What is the relationship between patient-reported symptom burden and early treatment discontinuation or switching (ETDS) in patients with metastatic breast cancer?**

- Goal of treatment in MBC is to prolong progression free survival and to minimize toxicities so that treatment can be delivered at full dose on schedule, while maintaining or improving HRQoL
- Accumulation of symptom burden over time may lead to early treatment discontinuation or switching treatments (ETDS) to reduce symptom burden
- Availability of PROs and symptom data in some real world settings, so it was possible to examine this relationship

Walker MS, Masaquel AS, et al. Breast Cancer Res Treat 2014

Research Question #2 – Rash and Survival

- **Is there an association between rash and overall survival in patients with pancreatic cancer receiving erlotinib?**
- In an exploratory analysis of a phase 3 trial of erlotinib in advanced pancreatic cancer, rash grade ≥2 was correlated with improved survival (Wacker et al., 2007)
- Using data from a community oncology setting, the primary objective was to determine if the association between rash and outcomes observed in clinical trials would be observed using real-world data.



Stepanski EJ, Reyes C, et al. Pancreas 2013

7

Maximizing The Utility of Real World Evidence: Selection of Data Sources to Optimize ROI

MAY 19, 2015

KATHY L. SCHULMAN, MS

PRINCIPAL, OUTCOMES RESEARCH SOLUTIONS, INC.
WALTHAM MA

Secondary Data Source Options in Oncology

8

- Electronic health records/medical records (EHR / EMR)
  - Structured vs unstructured
  - Linked to patient reported outcomes and/or billing records or claims data
- Tumor registries
  - Linked to billing/claims and/or EMR
- Claims databases with or w/o laboratory data
- Regional delivery networks with or w/o insurance subsidiaries
- Some data sources are dependent on code sets with limited clinical detail which can make cohort ascertainment difficult

- How many audience members have used each of these?

What is the Clinical Context?

9

- The best data source for your study will depend on
  - tumor type (solid, lymphoma, hematologic)
  - ICD 9/10 diagnosis code specificity
  - availability of a validated algorithm for use in claims or structured EMR data which is NOT linked to a registry
  - the importance of biomarkers or other prognostic factors
  - exclusionary conditions that need to be eliminated
  - the need to cull specific treatment groups or histologic subsets
  - comprehensiveness of follow-up data
  - the role of cancer stage (outcome, population subset, covariate)
  - the choice of outcome: overall survival, disease progression, recurrence, patient reported, economic

Case Study: Metastatic Breast Cancer

10

- Early Treatment Discontinuation/Switching in First-Line MBC
- Registry data only captures MBC at diagnosis; must have claims/EMR linkage to identify pts who develop MBC after dx
- Breast cancer diagnosis codes are not stage specific
  - Algorithms may be complex, PPV (66-75%); not very sensitive (51-62%) (Chawla 2014; Whyte 2013, Nordstrom 2012)
  - Optimal thresholds for sensitivity, specificity, PPV must be determined by investigators in consideration of their objectives
- No validated algorithm to identify first line treatment in BC
  - CRC study demonstrated PPV of 72% (Dacosta Byfield, 2013)
  - Ability to identify resection in kidney, colon and rectal cancers is good (PPV >90%, sensitivity>90%) (Miller 2009, Li 2012)

Case Study: Metastatic Breast Cancer

11

- Primary endpoint defined as a treatment change (cessation, or switch) not associated with disease progression defined as
  - EMR statement that patient discontinued early
  - observed duration of therapy < stated planned duration
  - EMR maximum cycle number < planned number of cycles
  - duration of planned therapy ≤ 6 weeks
- Data source in this study had to include unstructured EMR data since billing/claims database, and most structured EMRs, do not include prescribed regimens
- Counting actual chemotherapy cycles completed is difficult in any data source but can be done using unstructured EMR

Not Always That Clear Cut….

12

- The clinical detail available in registry and EMR databases is critical when you are evaluating
  - outcomes in histologic subpopulations that may be prognostic e.g. inflammatory breast cancer, women with triple negative breast cancer or are dependent upon a specific cancer stage
  - safety and/or efficacy outcomes between treatment regimens, especially non-standard multi-agent regimens
  - complex outcomes such as disease progression, recurrence (Chubak 2012, Hassett 2014)
- Increased accuracy in identifying treatment cohorts, exposure windows, disease progression/recurrence and the ability to control for disease stage, cumulative exposures may be more important than larger sample size, more heterogeneous patient and provider populations

Conversely….

13

- Databases constrained w/r/t clinical detail generally have greater sample size, more heterogeneous patient/provider population (geographically, payer, treatment setting, etc.)
- Important for assessing prevalence, characterizing care patterns, developing national estimates of disease burden
- May be adequate for treatment based cohort comparisons in cancers which are less dependent on prognostic clinical metrics or in cancers with heavy use of standard chemotherapy regimens (Lamont 2014)
- Use caution in employing metrics which require capture of cancer stage: good specificity and acceptable PPV may not be enough, poor sensitivity may prevent good predictive ability

Other Considerations: Is Follow-Up Sufficient & Comprehensive?

14

- Cancer patients are seen across treatment settings
  - Patients may seek care at tertiary care institutions (resection, chemotherapy, radiation) but get basic care at home e.g., labs done locally
  - Chemotherapies include oral, pharmacy dispensed agents
  - How will you capture at home hospice services
- Cancer patients may be more likely to go on long-term disability or if they do not have disability coverage, to lose employment
- Are there reimbursement issues that may affect claims completeness (e.g. coordination of benefits, dual eligible, capitation)?

Conclusion

- No perfect data source
- Know the biology & natural history of your cancer
- Review published criteria, especially case ascertainment algorithms, use caution in developing new algorithms
- List the strengths and weakness of each data source
- Conduct sensitivity analyses to address limitations in the data source you've chosen

VECTOR ONCOLOGY

Role of Unstructured EMR Data in Outcomes Research

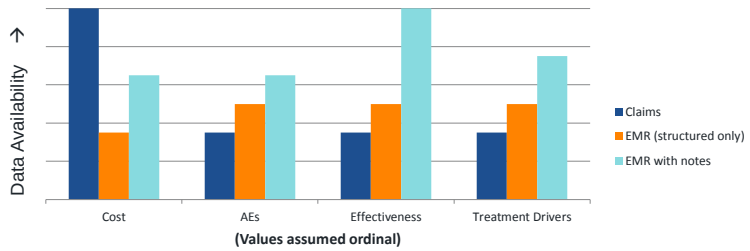
May 19, 2015

Availability of Clinical Data in Claims vs. EMR Sources

Research questions for which approach depends on data availability

Research Question
Cost of care / payments
Toxicity / AEs
Effectiveness (e.g., PFS, Response)
Drivers of treatment selection

Data Source
Claims
EMR (structured only)
EMR w/ unstructured data



(Values assumed ordinal)

VECTOR ONCOLOGY

Completeness of EMR Data

Variable	EMR (structured only)	EMR (w/ notes)
Race/ethnicity	70% - 80%	95% - 100%
Stage of disease	50% - 60% [1]	95% - 100%
Disease progression	Varies / <25%	~100%
Tumor response	N/A	65% - 90%
Performance status indication	<25%	30% - 70% → 100% [2]
Comorbid disease	Limited	Varies by condition: 50% - 100% [3]
Sites of metastasis	< 50%	90%+
Adverse events [4]	Limited, varies [4]	Varies, 50% - 100%
Biomarker testing / status	Limited, varies by marker	90% - 100% for established actionable markers [5]
Patient Reported Symptoms	Depends on Source	N/A

[1] Rate improves with recency
[2] 100% including presence v. absence of informal text record of impairment
[3] E.g., diabetes, hypertension shown at rates close to self-reported population levels for age
[4] In structured data, relies on labs and supportive treatments
[5] E.g., KRAS, HER2. Rates lower for more recent markers, but reflect lower uptake

VECTOR ONCOLOGY

Data Source and Research Questions

- Research questions should influence choice of data source
- Data characteristics should inform the research question

"EMR data are a poor substitute for claims"

"Claims are a poor substitute for EMR data"

- If you are using EMR data, don't think just in terms of claims questions
  - EMR data let you investigate a wider range of research questions
  - EMR data can provide insights into real world oncology care not available in other data sources

VECTOR ONCOLOGY

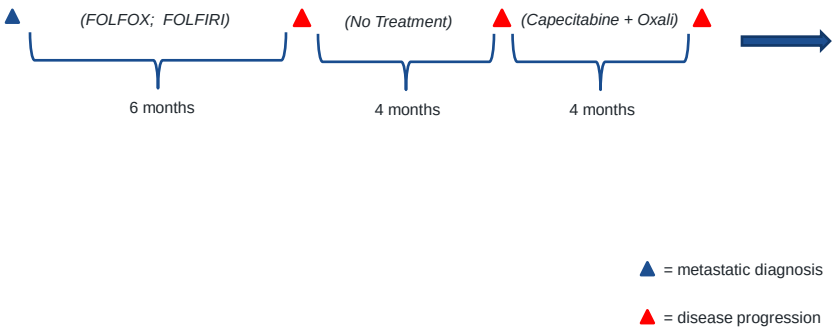
How Data Can Shape the Research?

- Question of interest: Chemo backbone treatment and sequencing (mCRC)
  - FOLFOX → FOLFIRI
  - FOLFOX { --- 2 months --- } FOLFIRI
- With information about disease progression
  - FOLFOX { --- 2 months --- } ▲ FOLFIRI
  - FOLFOX ▲ { --- 2 months --- } FOLFIRI
  - FOLFOX ▲ { --- 2 months --- } ▲ FOLFIRI

▲ = disease progression

VECTOR ONCOLOGY

How Data Can Shape the Research?



VECTOR ONCOLOGY

Accessing Unstructured Medical Record Data

- **Manual review of paper charts**
  - Traditional chart review
  - *Con*: getting access, inefficient
- **Electronically accessed medical record data**
  - Manual review conducted electronically
  - Efficiency depends on data structure
- **Key consideration**
  - Extent of technology support for human review
  - Structure / technology increases efficiency, but disaggregates the data and limits view of the "whole" patient.
    - Abstracting a complete patient record vs. abstracting a datum from a document

VECTOR ONCOLOGY

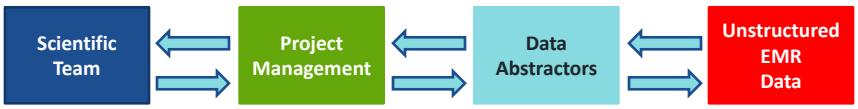
Uses of Unstructured EMR Data

- **Complete case / chart review**
  - Assumes at least some chart review for each case
  - Can be tailored to maximize use of structured EMR data
- **Selective augmentation of structured EMR data**
  - **Some but not all cases are reviewed**
  - Charts are reviewed to augment missing structured data
- **Text mining / Algorithm development & validation**
  - General
    - Create addition structured data based on text
    - Supports case identification, project specific case review
  - Project specific
    - E.g., use of text to identify record of KRAS testing

VECTOR ONCOLOGY

Feedback: Let the Process Inform the Research

Each part of the research process can inform the other



Involvement and proximity to the data collection process improves the quality of research design and the research itself

VECTOR ONCOLOGY

EMR and Billing Data in HRU and Cost Analyses

May 19, 2015

Kim R. Saverno, PhD, Director of Pharmacoeconomics, Vector Oncology Solutions

EMRs as Source of HRU Data

- EMRs have a record of all services provided during visit
  - CPT codes
  - HCPCS codes
- These records serve as a basis of billing information
- Services provided reflect type of provider and facility
- Unstructured text notes may contain additional information about utilization of services
- EMRs are a reliable data source for examining services associated with a particular clinic/facility or condition
- Services unrelated to visit or provided outside of clinic may not be represented well

EMRs as Source of Medication Utilization Data

- **EMRs provide valuable information regarding medications prescribed, including those that are self-administered**
  - Information found within unstructured text notes
  - Medication start date, frequency, dose and end date, frequency, and dose
  - Changes in dose, or discontinuation and stated reason for change
- **Medications administered during clinic visit are well documented in EMR**
  - Information housed within structured data
  - Non-self administered medications
  - Exact doses
  - Dates of administration
- **Medication data not contained in EMR**
  - Medications unrelated to condition or prescribed outside the clinic/facility
  - Records of medications received from pharmacy

VECTOR ONCOLOGY

EMRs as Source of Cost Analyses

- Build upon the EMR HRU data
- What should be used as the cost of services?
- Billing records may also be available, depending on the data source
  - Amount charged
  - Paid amounts
- Even when billing records are available, some services of interest may not have been provided at facility, e.g. oral medications
- Many methods available for overcoming these limitations
- State clearly chosen methodology for determining cost of services, including data source and any assumptions

VECTOR ONCOLOGY

EMRs versus Administrative Claims for HRU and Cost Analyses

EMRs

- Costs and HRU documented in record depend on type of facility and condition
- Billing records may not be available. More complex methodologies may be needed for cost analyses
- Detailed information about medications prescribed, but lacks information on paid pharmacy claims
- Data abstraction often required, which can be time-consuming
- Can be combined with rich clinical information from unstructured text data.

Administrative Claims

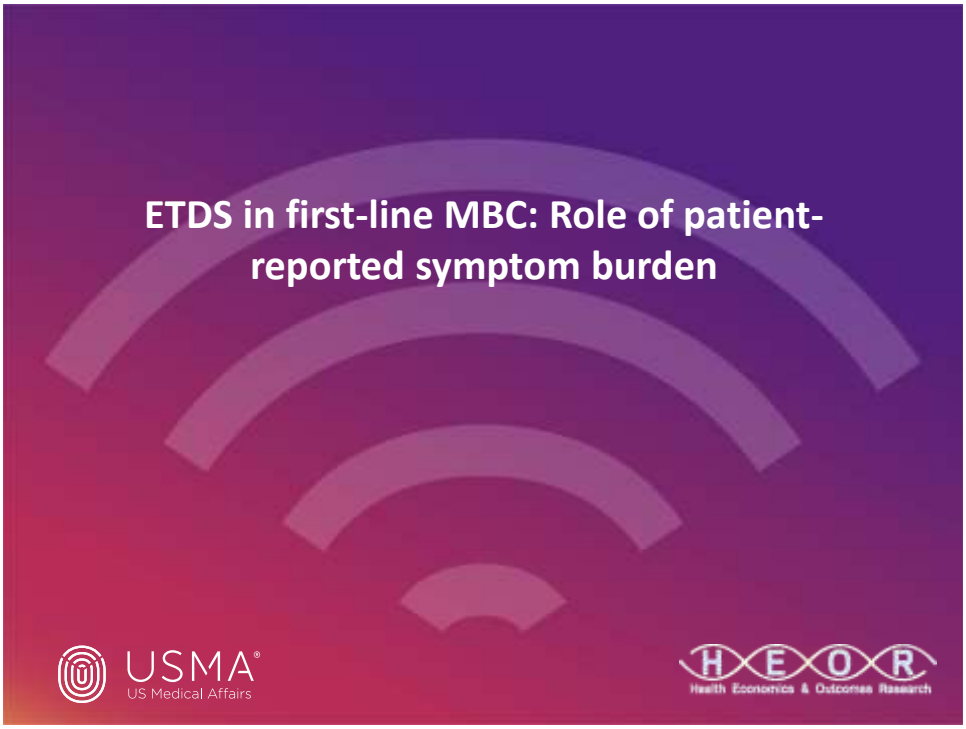
- Total costs and HRU, not site or condition specific
- Claims records for paid/approved services
- Pharmacy records (refills, cost of outpatient medications, compliance), but other assumptions about medication utilization necessary
- Data in structured form. No extra time needed for data abstraction.

VECTOR ONCOLOGY

Using RWE in Oncology Outcomes Research

USMA® US Medical Affairs

HEOR Health Economics & Outcomes Research



Study Objective and Results

Objective: To examine relationship between patient-reported symptom burden and ETDS

Results:

1.Overall rate of ETDS was 24.7%. The ETDS event rate was nominally lower in the hormone therapy group (11.1%) than in the other groups (chemo: 27.6%, targeted: 26.1%)

2.PCM composite score* was a significant predictor of ETDS (HR=1.132, p<0.0001). For each one-point increase in the composite score over time, a patient's risk of ETDS increased by 13.2%

3.Symptoms that were moderate (PCM scores 10-14) and severe (PCM scores 15-22) were associated with increased risk of ETDS (HR = 4.135, p<0.0001; and HR=5.287, p<0;0001, respectively)

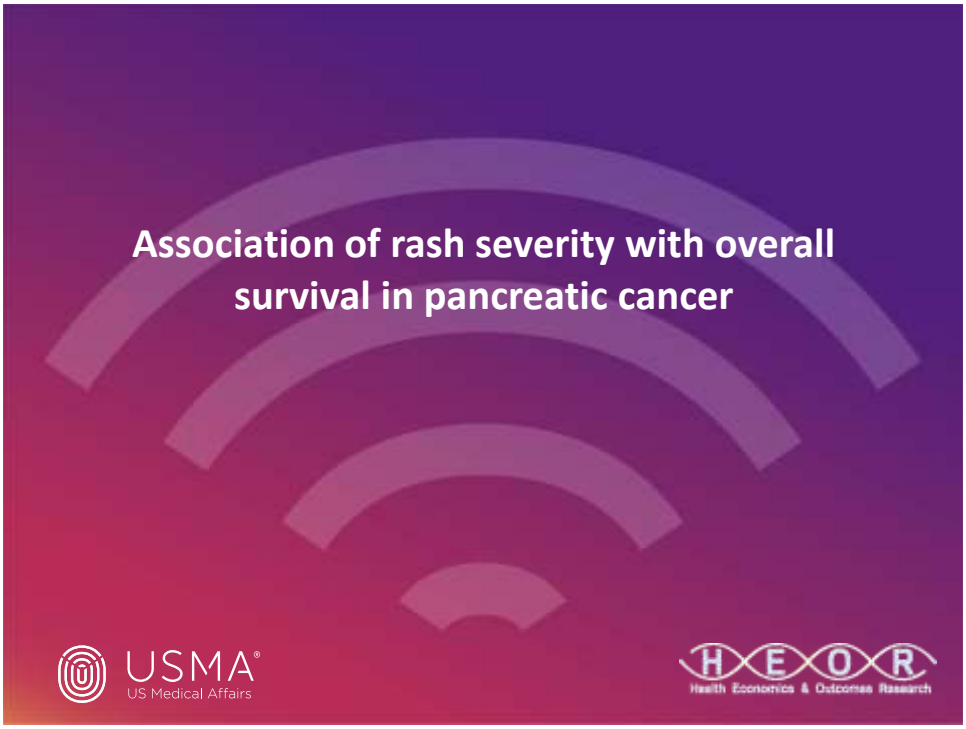
*Patient Care Monitor is an 86-item self-report measure that assesses physical and psychological symptoms, and physical functioning, on an 11-point Likert scale. Higher scores = more severe symptoms

Learnings and Limitations of Real World Data

- Manual chart abstraction was necessary
  - Each record was electronically reviewed by experienced Clinical Research Nurses to verify eligibility and abstract key clinical characteristics
- Inclusion criteria to confirm patient eligibility greatly reduced anticipated patient sample
  - Initial count of 2522 patients → 802 patients meeting inclusion criteria
- Different patients had different numbers of surveys available during first line therapy
  - Required control for this variable in statistical models
- Variables that are potentially relevant to persistence on first-line therapy would require chart review
  - Performance status, stage at initial diagnosis, etc.

HEOR Health Economics & Outcomes Research

USMA® US Medical Affairs33



Study Objective and Results

Objective: To determine if the association between rash and outcomes observed in clinical trials would be observed in real-world community oncology settings

Results:

1. Patients in the High Severity group (n=34) had a longer median overall survival than those in the Low Severity group (n=134). 7.6 vs 5.0 months, p=0.0339
2. Cox regression analysis confirmed a reduced risk of death with High Rash Severity vs Low Severity (HR=0.67, p=0.0389)

Learnings and Limitations of Real World Data

- Manual chart abstraction was necessary
  - Diagnosis of pancreatic cancer was confirmed by pathology reports or based on definitive statements by the treatment physician regarding diagnosis
- Occurrence of symptoms, such as rash, was not systematically documented within the medical records
  - Sample of patients were smaller than anticipated in the High Rash Severity group
- Formal grading of rash following Common Terminology Criteria for Adverse Events (CTCAE v 3.0), as in clinical trials, was generally not performed in RWD
  - Severity of rash was informed from documentation about the presence, absence, and severity of rash in the medical record
  - "moderate" or "severe" were assumed to be comparable to grade 2 or higher rash


Thank You!
ISPOR 20th Annual International Meeting, Philadelphia, PA, May 19, 2015