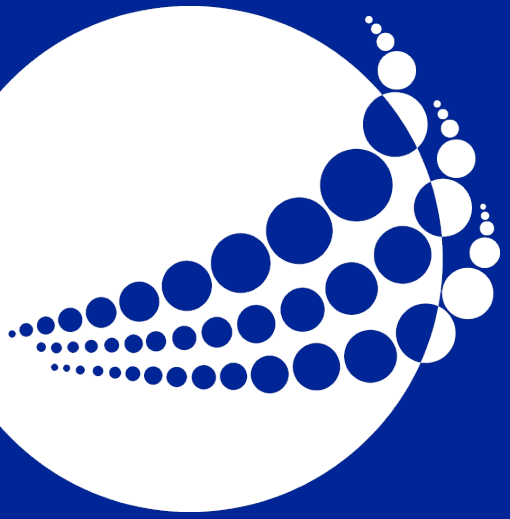
M-CERSI:

Use of Big and Real-World Data by PhRMA: More than Data Warehousing

Aaron Galaznik, MD

Senior Director, Real-World Data and Analytics Pfizer, Inc.

February 11, 2014

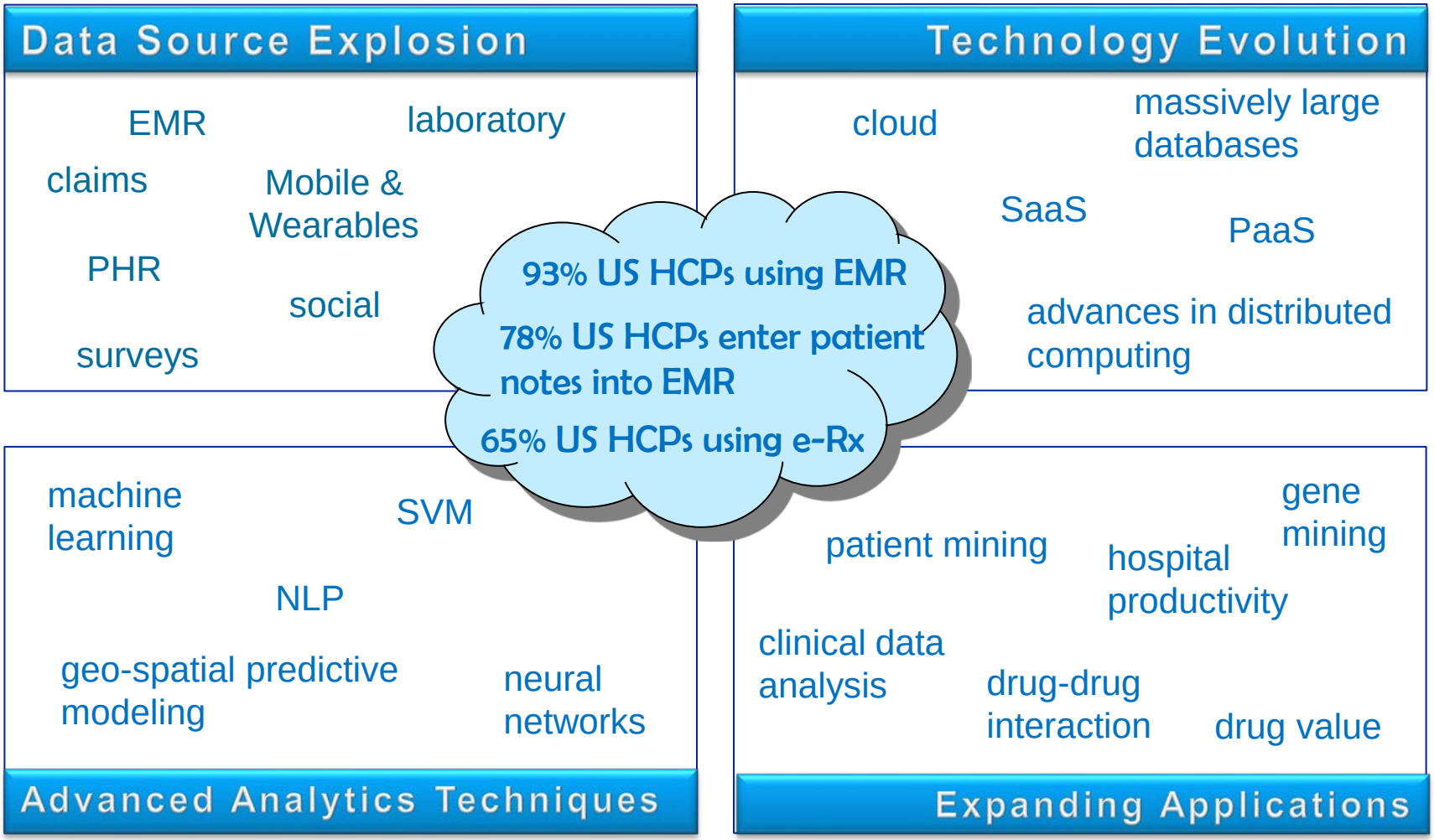


Enhancing Real World Insights Together

Agenda

- Overview of Big Data and Real World Data
- Real World Data vs Clinical Trial Data
- Application of Real World Evidence in the BioPharmaceutical Life Cycle
    - Drug Target Validation
    - Clinical Trial Protocol Development
    - Patient/Disease Pathway Understanding
    - Drug Utilization Understanding
    - Population Cost/Reimbursement Dynamics

Current Landscape



Source: http://newsroom.accenture.com/news/ehr-and-hie-use-increases-among-us-doctors-accenture-annual-survey-finds.htm

BIG DATA (including Real World Data) will transform:

Research & Development

- Precision Medicine
- Clinical Trial Design

Commercialization

- Market segmentation and targeting
- Adherence / Compliance

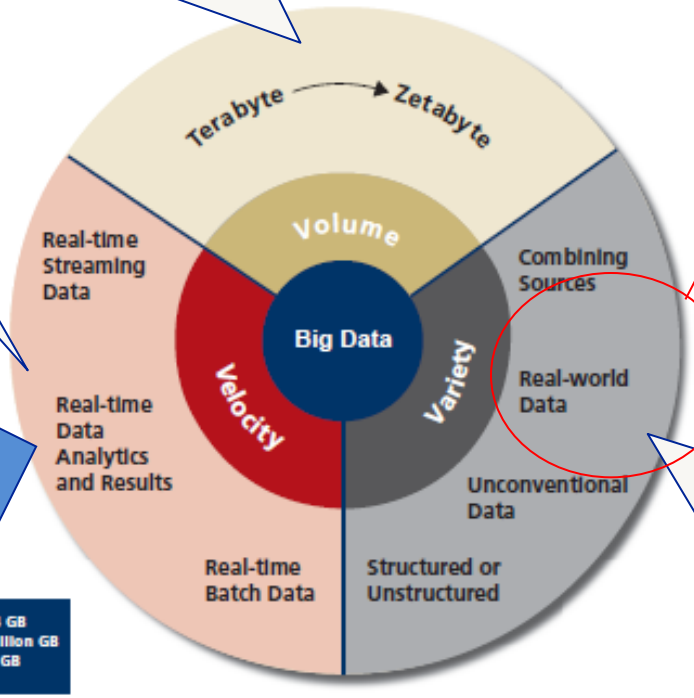
Dynamics of relationships among patients, providers, payers, and developers of new therapeutics

- Patient centered research & development
- Social networking and increasing access to public data sources

What is Big Data in Pharma?

Genomic
Imaging
EMR Unstructured Notes

Sensors and Health Monitoring Devices



Learning Healthcare System

1 Terabyte = 1024 GB
1 Petabyte = 1 million GB
1 Zetabyte = 10^{12} GB

Current focus

Claims, Laboratory
EMR, PHR
Surveys: Health Risk Assessments, Health Status Assessments
Pt Reported Outcomes
Revealed Pt Behaviors and Preferences (ex. Purchasing Habits, Google Trend)

Adapted from: Zikopoulos PC, Eaton C, deRoos D, Deutsch T, Lapis G. *Understanding Big Data. New York: McGraw Hill; 2012*

What is Real World Data?

Real World Data is healthcare data used for decision making that is not collected in conventional randomized controlled trials (RCTs)

Sources of Real World Data:

Focus for Today

Databases	• Cross-sectional and longitudinal databases which essentially provide retrospective data but increasingly offer the opportunity to have prospective add-ins.
Surveys	• Primarily for epidemiological information.
EMRs	• Used to reflect particular insights in patient management.
Cohort studies	• What most people would understand by real life studies.
Pragmatic clinical trials	• Simple experimental trials, where efforts are however made to mimic a real life situation as much as possible.
Registries	• Analyzing all patients treated at a particular center for a particular condition on a continuous basis.

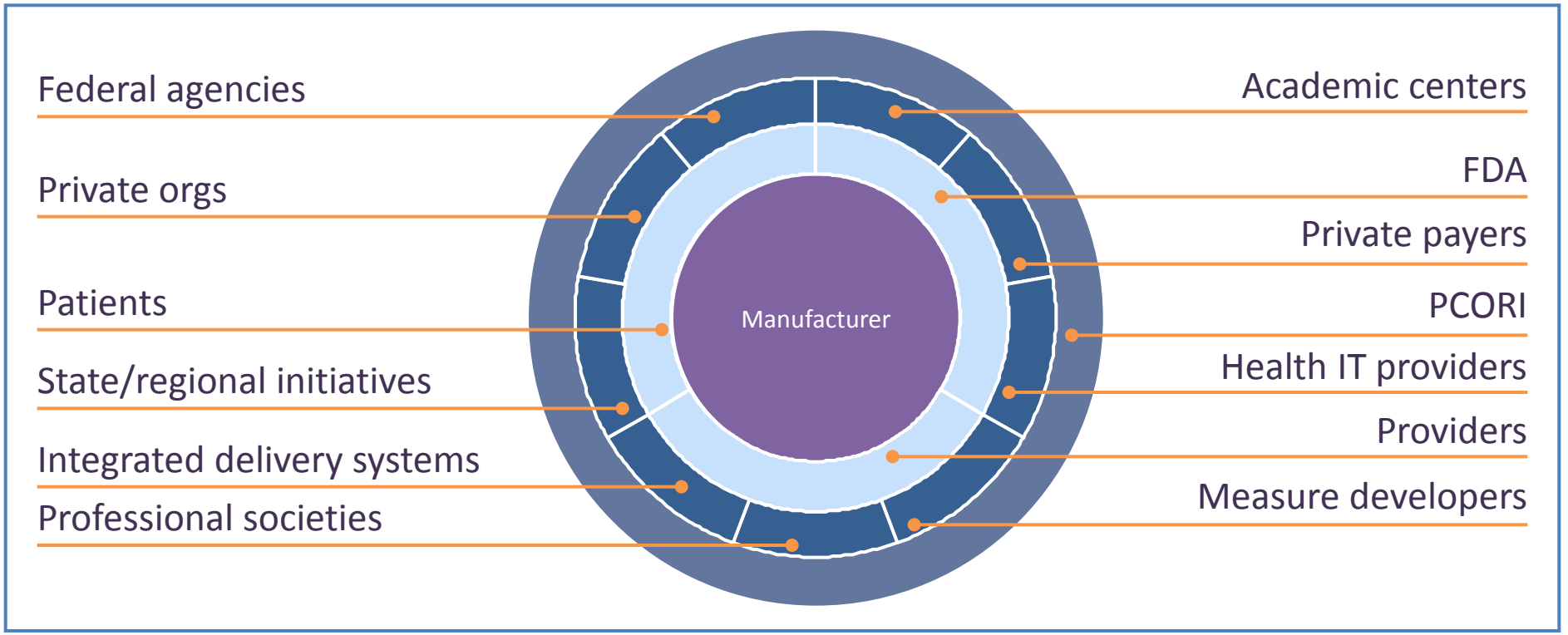
- "Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report," Value in Health, Volume 10, November 5, 2007.
- Annemans, L., Aristides, M., Kubin, M. "Real-Life Data: A Growing Need," ISPOR Connections 2007.

Real World Data and Clinical Trial Data

	Efficacy (Clinical Trial Data)	Effectiveness (Real-World Data)
Objective	Does it work under ideal circumstances	Does it work under usual circumstances
Setting / Design	Controlled clinical trial	Real-world clinical practice
Purpose	Regulatory approval (FDA)	Drug performance in real-world
Intervention or treatment	Fixed regimen	Flexible regimen
Comparator	Placebo	Active comparator/usual care
Subjects	Homogenous/highly selective (stringent inclusion/exclusion criteria)	Heterogeneous / any subjects
Compliance	High	Low to High
Outcomes	Clinical endpoints (e.g. BP, HbA1c, LDL)	Example: Cardiovascular events, hospitalizations
Internal Validity	High	Low
External Validity (generalize to other populations)	Low to medium	Medium to high

Why is Real World Data a critical focus ?

- Increasing use of real-world evidence in decision making
  - Enhanced post-marketing safety surveillance
  - Comparative Effectiveness; Value-based purchasing; Risk-sharing contracts
- Growing universe of developers of real-world evidence
  - Bio-Pharm R&D companies can either develop/analyze real-world evidence as part of development & commercialization; or others will do it for/to them.



Mini-Sentinel Initiative

- Mini-Sentinel is a pilot project sponsored by the U.S. Food and Drug Administration (FDA) to create an active surveillance system - the Sentinel System - to monitor the safety of FDA-regulated medical products.
- Mini-Sentinel uses pre-existing electronic healthcare data from multiple sources.Collaborating Institutions provide access to data as well as scientific and organizational expertise.

- The Mini-Sentinel pilot provides FDA the ability to:
  - Work through the "nuts and bolts" of designing safety assessments using multiple existing electronic healthcare data systems
  - Develop and evaluate scientific methods to increase the precision of active safety surveillance efforts
  - Identify and address barriers and challenges to building a practical, accurate, and timely system for active safety surveillance

Mini-Sentinel

Meta-Data and A Glimpse of the Future

J Med Info Assoc

Brief communication

Web-scale pharmacovigilance: listening to signals from the crowd

Ryen W White,[1] Nicholas P Tatonetti,[2] Nigam H Shah,[3] Russ B Altman,[4] Eric Horvitz[1]

▸ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/amiajnl-2012-001482).

[1]Microsoft Research, Redmond, Washington, USA
[2]Department of Biomedical Informatics, Columbia University, New York, New York, USA
[3]Department of Medicine, Stanford University, Stanford, California, USA
[4]Departments of Bioengineering and Genetics, Stanford University, Stanford, California, USA

Correspondence to
Dr Ryen W White, Microsoft Research, Redmond, WA 98052, USA; ryenw@microsoft.com

Received 9 November 2012
Revised 8 January 2013
Accepted 13 January 2013

ABSTRACT
Adverse drug events cause substantial morbidity and mortality and are often discovered after a drug comes to market. We hypothesized that Internet users may provide early clues about adverse drug events via their online information-seeking. We conducted a large-scale study of Web search log data gathered during 2010. We pay particular attention to the specific drug pairing of paroxetine and pravastatin, whose interaction was reported to cause hyperglycemia after the time period of the online logs used in the analysis. We also examine sets of drug pairs known to be associated with hyperglycemia and those not associated with hyperglycemia. We find that anonymized signals on drug interactions can be mined from search logs. Compared to analyses of other sources such as electronic health records (EHR), logs are inexpensive to collect and mine. The results demonstrate that logs of the search activities of populations of computer users can contribute to drug safety surveillance.

BACKGROUND
The US Food and Drug Administration and other organizations collect reports on drug side effects from physicians, pharmacists, patients, and drug companies.[1–3] These reports provide valuable clues about drug-related adverse events, but are incomplete and biased.[4–6] As a result, adverse event alerts for single drugs are often delayed as evidence accumulates.[7 8] These challenges are compounded in the setting of adverse events resulting from multiple drugs that interact in unexpected ways.

Given that a significant use of the internet is for health searches, we hypothesized that internet users may provide early clues about adverse drug events via their online information-seeking activities.[9] Previous research on tracking seasonal influenza has demonstrated that search logs can form an implicit sensor network for health monitoring.[10 11] In that work, search logs accurately estimated the weekly levels of influenza activity in different regions of the USA, with a reporting delay of approximately 1 day. The authors showed that health-seeking activity captured in queries to online web search services mirrors trends in data gathered by tradi-

case an interaction between paroxetine (an antidepressant) and pravastatin (a cholesterol-lowering drug), which was recently reported to create hyperglycemia.[13 14] This association was extracted from the US Food and Drug Administration adverse event reporting system (AERS) using a data-mining algorithm that aggregates reports to
interactions.[13] The finding was
spective analysis of the electron
three regionally distinct medical
firmed in a mouse model.[14] W
patients taking these two dru
symptoms of hyperglycemia and
internet searches on these sym
related to hyperglycemia before
reported in 2011.

METHODS
We analyzed the search logs of
ing web users who opted to sh
with Microsoft via the install
add-on, spanning a 12-month p
and comprising searches on
Yahoo!. An anonymous iden
instance of the browser add-on
the drugs and symptom queries
formed over time (note that we
tinguish between multiple u
machine). Searches for informa
drugs are common. We found t
people (0.43%) pursued infor
one of the top 100 best-selling
including paroxetine and prav
tions that we focus on here.[15]

By examining words used i
sought evidence that searches fr
pravastatin and paroxetine over t
2010) would have a higher
hyperglycemia-associated words
ing for only one of the d
hyperglycemia-related terminolo
included in the supplementary n
mentary table S1, available onli
ated the list based on a review
The list is broad to ensure that we covered a majority
of related symptoms. Although there are many pos-

The New York Times

March 6, 2013

Unreported Side Effects of Drugs Are Found Using Internet Search Data, Study Finds:

"Using data drawn from queries entered into Google, Microsoft and Yahoo search engines, scientists at Microsoft, Stanford and Columbia University have for the first time been able to detect evidence of unreported prescription drug side effects before they were found by the Food and Drug Administration's warning system."

Adapted from PhRMA

"The Pink Sheet®"

Adverse Events In Social Media: FDA Expects Signal Detection "Revolution"

By Sarah Karlin / Email the Author / View Full Issue

Drug Safety / Word Count: 2555 / Article # 00140127001 / Posted: January 27 2014 12:02 AM

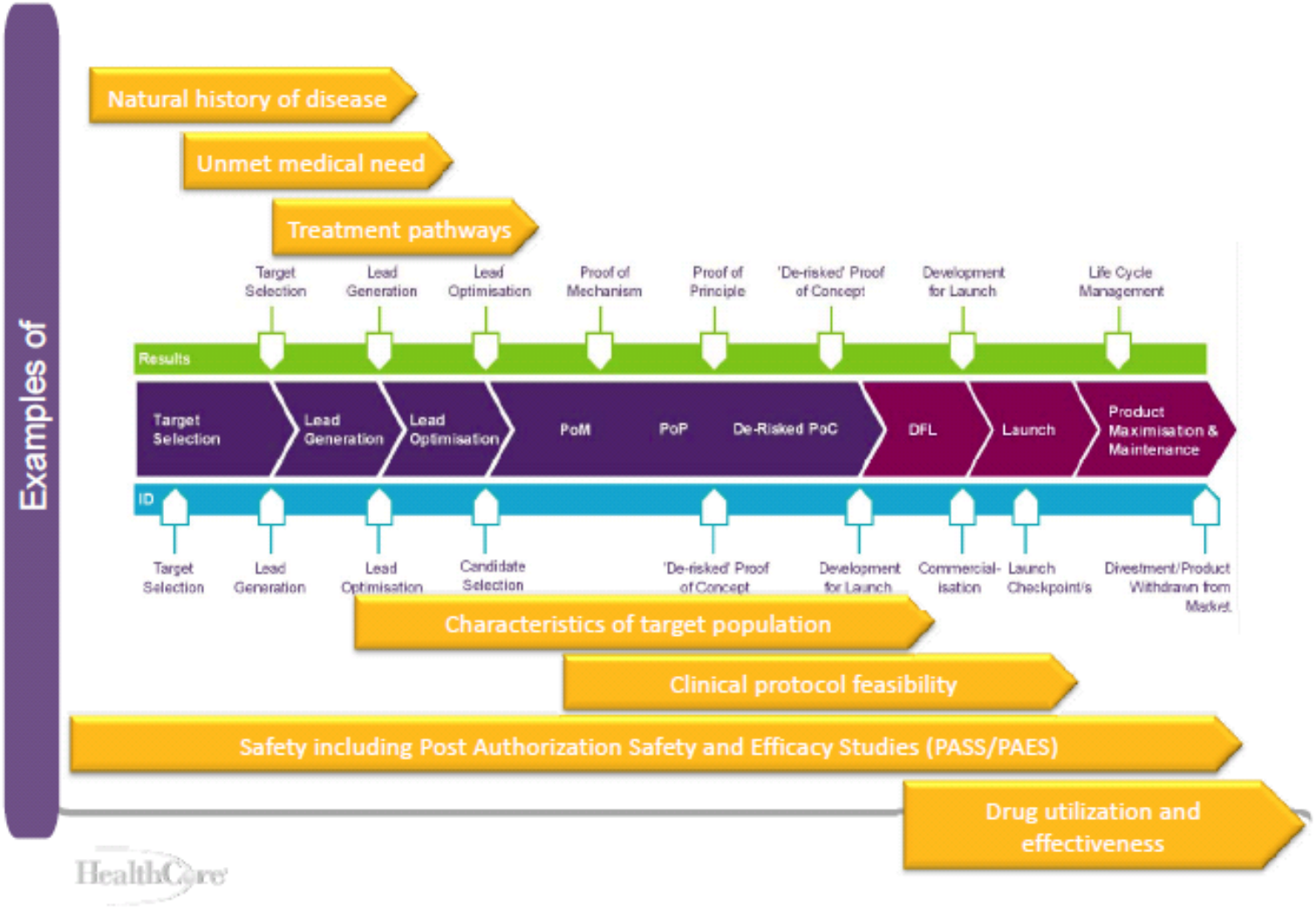
Executive Summary

Technology is ready now, and agency anticipates Twitter and Facebook will become part of the "key instruments" of pharmacovigilance in the next five years.

Advances in information technology are expected to allow FDA to vastly expand how it collects and analyzes adverse event information on social media platforms like Twitter and Facebook.

While FDA now tends to use a single system of adverse event reporting to detect drug safety signals post-market, that is going to change, Henry Francis, director for data mining and informatics evaluation and research in the Center for Drug Evaluation and Research's Office of Translational Science, said.

RWE Support Across the Product Life Cycle



Adapted from J. Singer, HealthCore, Fall 2013 PRISME

Real World Data Examples Throughout the Product Life Cycle:

- Drug Target Validation
- Clinical Trial Protocol Development
- Patient/Disease Pathway Understanding
  - Prevalence of Disease
  - Disease concomitancy
  - Patient Profiling
- Drug Utilization Understanding
  - Treatment persistence
  - Use of treatment by indication
- Population Cost/Reimbursement Dynamics
  - Burden of Illness
  - Impact of Utilization Controls

Use of real world data to understand potential target drug development pathways in Precision Medicine

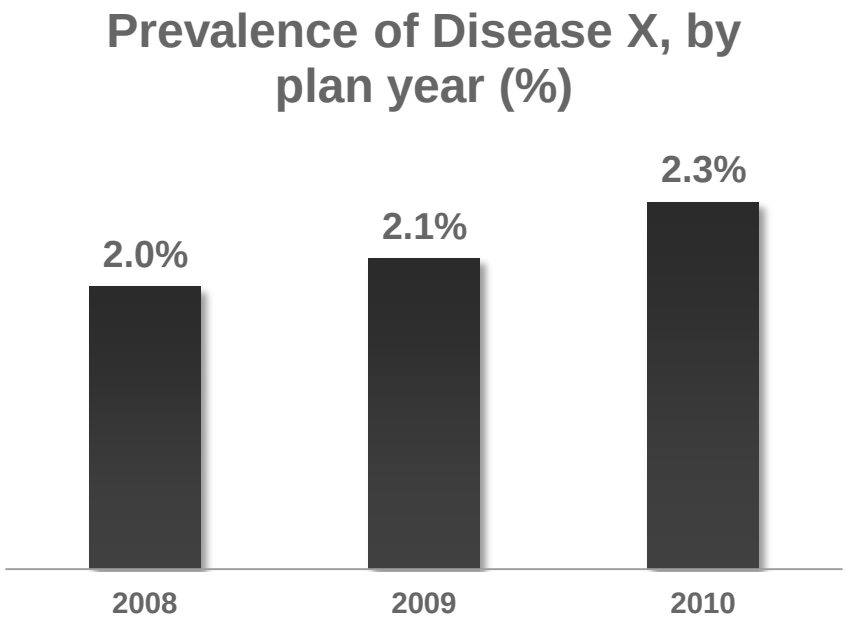
	Medstat	Optum	Humedica
Total Number of Patients	96,561,490	47,848,331	14,513,153
*Disease X Patients (ICD9 – based)	3166	2583	309
T2D patients	6,368,229	2,636,698	998,307
Disease X + T2D	733	735	48
% of total with T2D	6.6%	5.5%	7.3%
% of Disease X with T2D*	23.2%	28.5%	15.5%

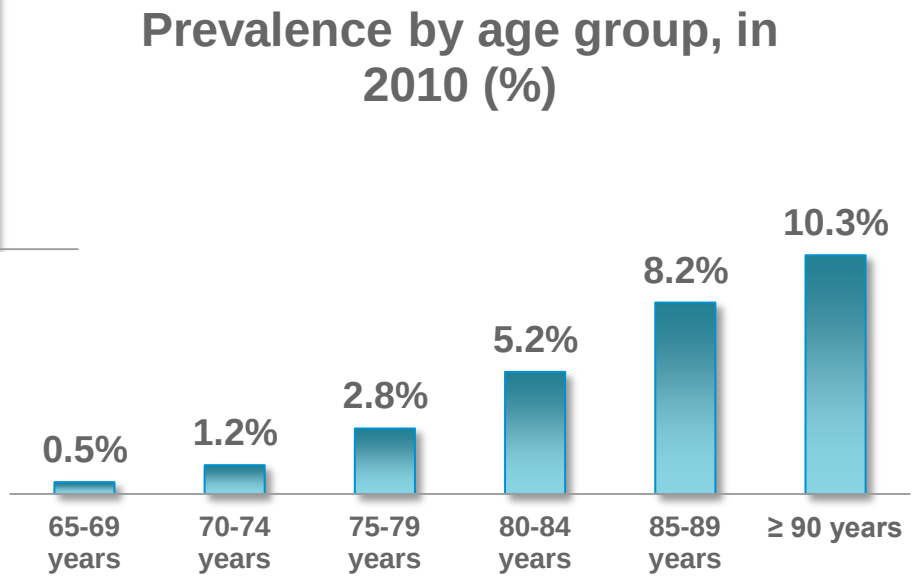
*Rates not yet adjusted for impacts of ascertainment bias

Use of Claims Data for Clinical Trial Enrollment Criteria Development

Flag	Coding	Description	Patient Count
100% Patient Count	Patients with any of either 331.0x or 290.xx ICD-9 code between July 1, 2011 and June 30, 2012, **50+** years of age, with medical and pharmacy benefits		**78,184**
Flag_3310	ICD-9 code: XXXXX		45,355
Flag_2901	ICD-9 code: YYYY		5,655
Flag_290	ICD-9 code: ZZZZZ		49,757
Flag_age50		Age 50+	78,184
Flag_age60		Age 60+	77,070
Flag_age70		Age 70+	72,600
Flag_age80		Age 80+	55,461
Flag_age90		Age 90+	0
Flag_drug_x	Generic Name = drug_x	Patient was treated with drug_x at any point of time during the study period	22,113
Flag_drug_x_range	Generic Name = drug_x	Patient was treated with drug_x at daily dosage at 10mg at any point of time during the study period	16,810
Flag_drug_x_30d	Generic Name = drug_x	daily drug_x therapy at 10 mg for 3 months	7,706
Flag_don_1m	Generic Name = drug_x	Patient must be on drug_x for 1+ months	18,946
Flag_don_2m	Generic Name = drug_x	Patient must be on drug_x for 2+ months	17,556
Flag_don_3m	Generic Name = drug_x	Patient must be on drug_x for 3+ months	16,210
Flag_drug_x_or_y	Generic Name = drug_x and drug_y	Patient must be on both drug_x and drug_y during the study period	9,067
Flag_drug_y	Generic Name = drug_y	Patient was treated with drug_y at any point of time during the study period	15,932
Flag_81401	CPT = 81401	Molecular pathology procedure, Level 2 (eg, 2-10 SNPs, 1 methylated variant, or 1 somatic variant [typically using nonsequencing target variant analysis], or detection of a dynamic mutation disorder/triplet repeat) ABL (c-abl oncogene 1, receptor tyrosine	0
Flag_S3855	CPT = S3855	Genetic testing for detection of mutations in the presenilin - 1 gene	0
Flag_S3852	CPT = S3852	DNA analysis for APOE epsilon 4 allele for susceptibility to Alzheimer's disease	0
Flag_drug_x_10mg_mono	Generic Name = drug_x	Patient must have Alzheimer ICD-9 code, and treated with drug_x at 10mg per day, mono therapy	152

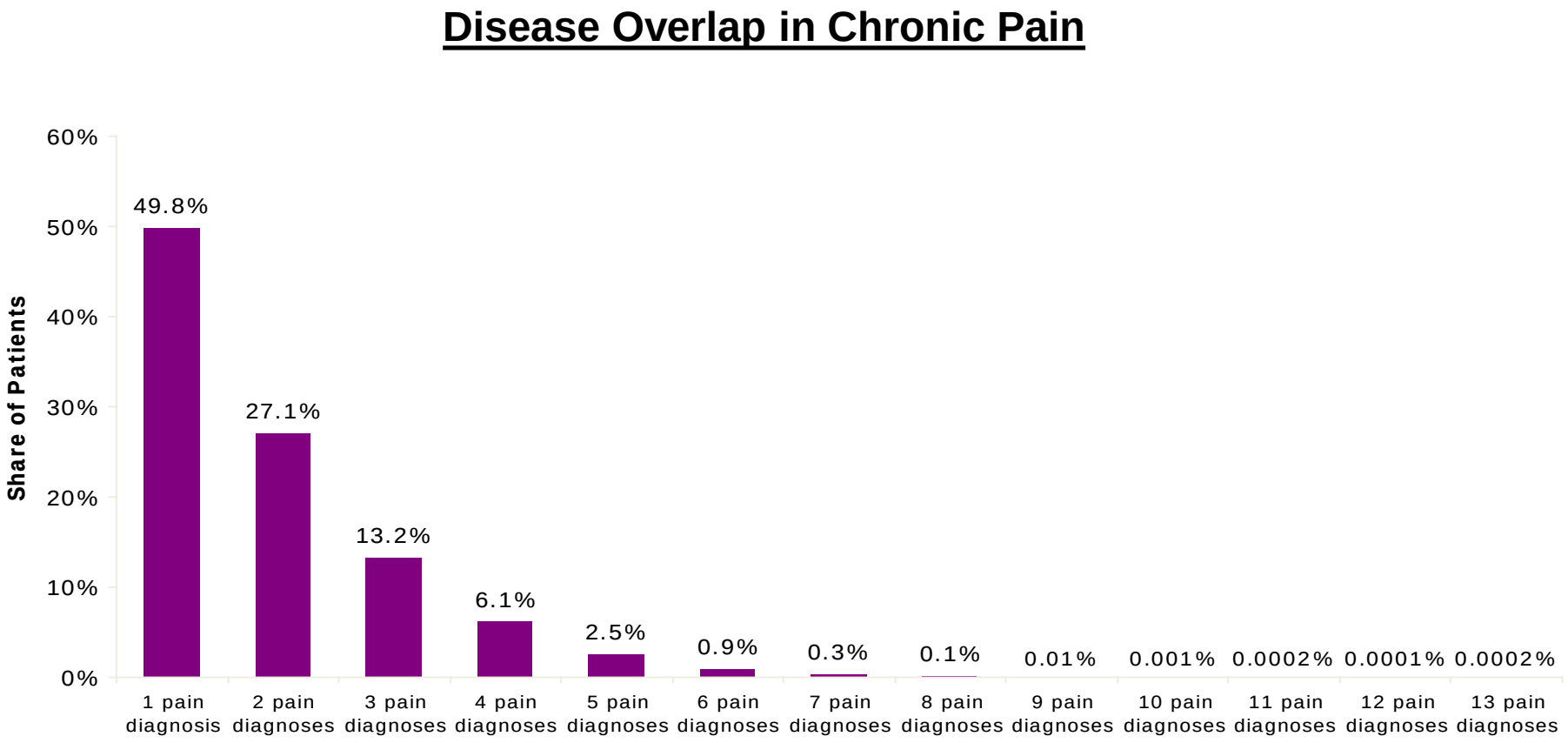
Evaluating Prevalence of Disease





 Annual Prevalence of Disease by ICD-9 Code in a Health Insurance Plan

Estimating Concomitancy of Illness



Sizing and Profiling of Target Patient Populations with EMR Data

		Cohort 1		Cohort 2		Cohort 3	
		N/Mean	%/Std	N/Mean	%/Std	N/Mean	%/Std
Subject count		2,285,277	100.00	474,745	100.00	301,506	100.00
Statin Tx for LDL-lowering		1,148,033	50.24	293,583	61.84	171,854	57.00
Gender	Female	1,173,995	51.37	232,250	48.92	157,731	52.31
Gender	Male	1,106,876	48.44	241,754	50.92	143,437	47.57
Gender	Unknown	4,406	0.19	741	0.16	338	0.11
Average age at 2012 (mean)		60	13.51	66	11.56	64	11.44
Male >45 y.o.		900,405	81.35	223,579	92.48	132,512	92.38
Female >55 y.o.		760,286	64.76	183,128	78.85	112,829	71.53
Current smoker		952,511	41.68	271,475	57.18	165,685	54.95
Average SBP (mean)		127	16.94	128	18.68	127	16.97
Systolic Blood Pressure >150		166,162	7.27	44,205	9.31	22,932	7.61
Diastolic Blood Pressure >90		116,675	5.11	20,927	4.41	12,229	4.06
BP medication		1,246,835	54.56	335,872	70.75	193,469	64.17
HDL <40		341,412	14.94	95,488	20.11	48,547	16.10

Understanding Persistency to Treatment

Average Days Supply Received Over 12 Months

	Disease1	Disease 2	Disease 3	Disease 4	Disease 5	Disease 6
Drug 1	141	160	190	132	135	145
Drug 2	58	105	146	70	61	84
Drug 3	58	77	95	63	63	72
Drug 4	136	123	191	167	173	166
Drug 5	25	40	56	51	47	65
Drug 6	26	50	68	50	46	67
Drug 7	48	70	96	62	55	80
Drug 8	128			120		128

Source: Payor Claims Data. Disease defined by ICD-9 Code

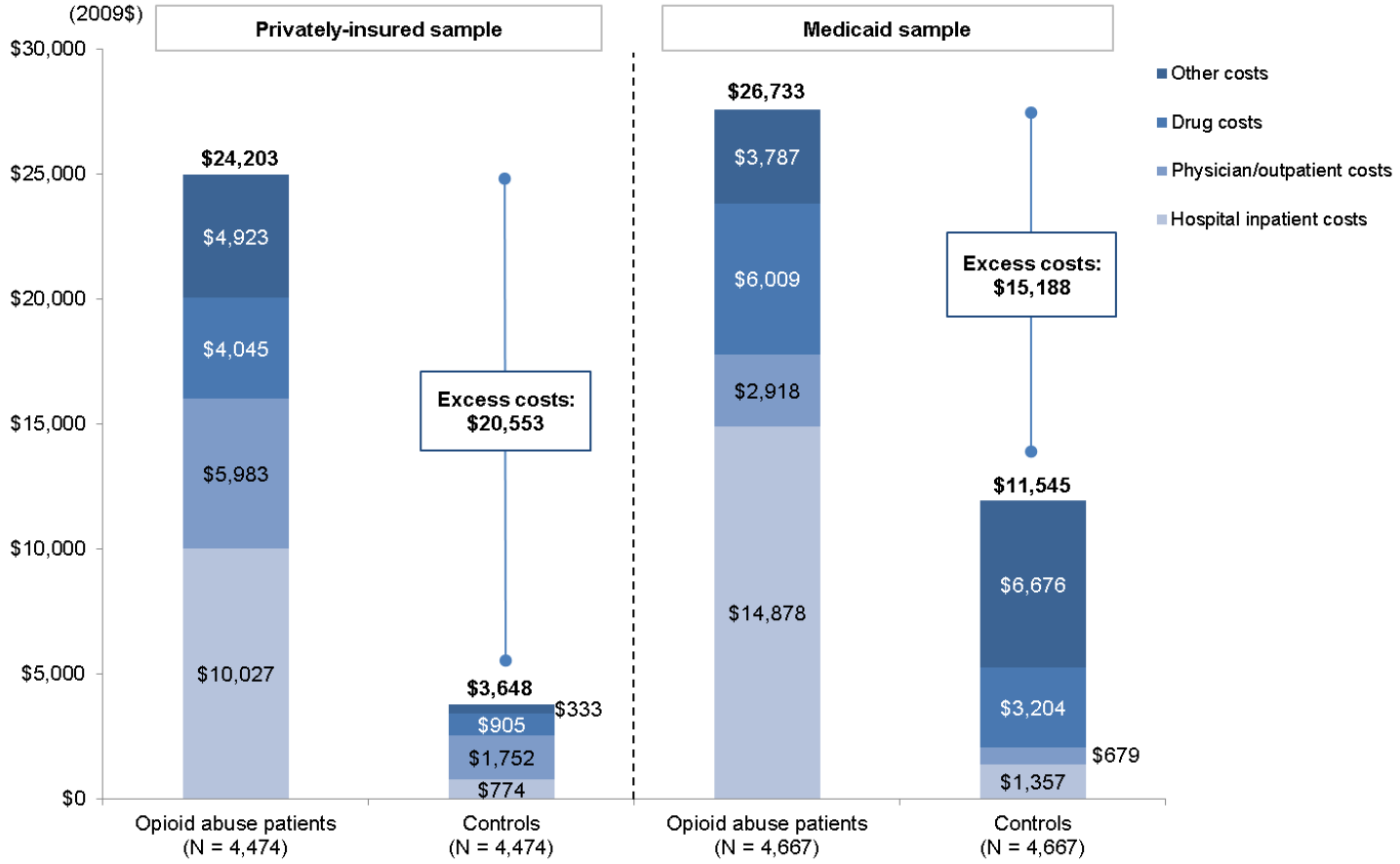
Understanding Inpatient Drug/Procedure Use by Diagnosis

ICD-9 Major Groups Codes for Discharges with Drug X

ICD-9 Major Group Code	Patient Count	Discharge Count
1. INFECTIOUS AND PARASITIC DISEASES (001-139)	891	1,106
2. NEOPLASMS (140-239)	1,012	2,521
3. ENDOCRINE, NUTRITIONAL AND METABOLIC DISEASES, AND IMMUNITY DISORDERS (240-279)	2,805	8,601
4. DISEASES OF THE BLOOD AND BLOOD-FORMING ORGANS (280-289)	1,712	2,656
5. MENTAL DISORDERS (290-319)	888	1,131
6. DISEASES OF THE NERVOUS SYSTEM AND SENSE ORGANS (320-389)	1,741	5,601
7. DISEASES OF THE CIRCULATORY SYSTEM (390-459)	1,633	2,375
8. DISEASES OF THE RESPIRATORY SYSTEM (460-519)	1,272	1,991
9. DISEASES OF THE DIGESTIVE SYSTEM (520-579)	1,027	1,307
10. DISEASES OF THE GENITOURINARY SYSTEM (580-629)	1,027	1,208
11. COMPLICATIONS OF PREGNANCY, CHILDBIRTH, AND THE PUERPERIUM (630-679)	550	573
12. DISEASES OF THE SKIN AND SUBCUTANEOUS TISSUE (680-709)	371	584
13. DISEASES OF THE MUSCULOSKELETAL SYSTEM AND CONNECTIVE TISSUE (710-739)	1,058	2,015
14. CONGENITAL ANOMALIES (740-759)	51	73
16. SYMPTOMS, SIGNS, AND ILL-DEFINED CONDITIONS (780-799)	1,760	2,473
17. INJURY AND POISONING (800-999)	677	931
18. FACTORS INFLUENCING HEALTH STATUS AND CONTACT WITH HEALTH SERVICES (V01-V89)	2,440	4,474
19. EXTERNAL CAUSES OF INJURY AND POISONING (E800-E999)	598	685

Estimating Burden of Illness

FIGURE 1—Average Annual Per Patient Direct Health Care Costs for Opioid Abuse Patients: Privately-Insured and Florida Medicaid, 2003-2007

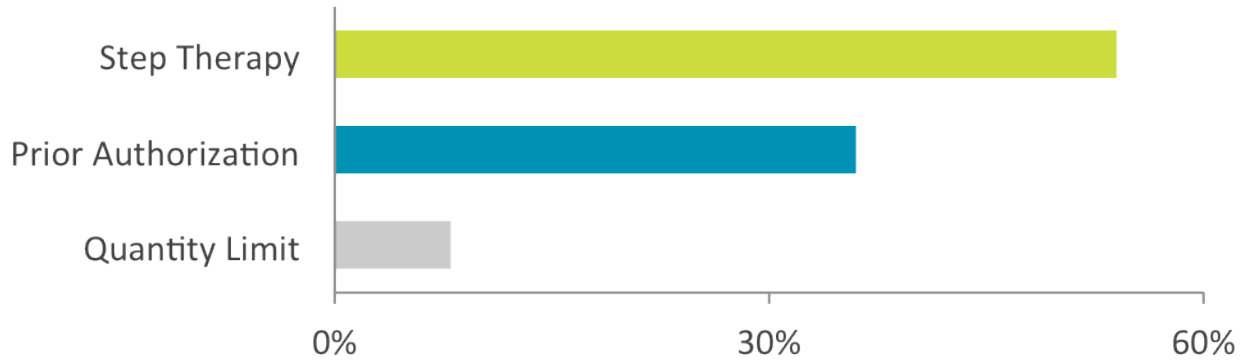


Note.
[a]Costs were measured over a 12-month period (the 6 months before and after the index date).
[b]Other costs include Emergency Department visits, lab services, and treatment at other places of services.

White, AG, et al. *Am J Pharm Benefits* 2011

Impact of Utilization Management Controls on Utilization

Percent of Patients Who Did Not Pursue Any Treatment After a Rejected Drug X Claim



46.2% of patients whose Drug X claim was rejected did not fill any covered smoking cessation medication within 6 months of the rejected claim.

- **Analysis does not capture out-of-pocket purchase of OTC medications subsequent to claim rejection**

Retrospective cohort analysis of pharmacy claims data to determine the impact of utilization management on fills *after rejection*. **Primary outcome** was whether the patient used any medication including Drug XX, Y, Z (prescribed and OTC) in the 6 months following a rejected claim for Drug X. **Inclusion criteria** were: at least 18 years of age, naïve to Drug X, with a rejected Drug X claim in plans that covered the drug between January 2007 and April 2008 due to one of the utilization management methods shown above, continuously enrolled in the same health plan 1 year before and 183 days after the initial rejected Drug X claim, and patient coverage of OTC SC medication.

Total N=15,597; DNC=5,440; SE=1,003; PA=4,745; QL=4,409.

DNC = drug not covered. SE = step edit. PA = prior authorization. QL = quantity limit.

QUESTIONS?

Pfizer's RWDnA DataMart Licensed & Public Datasets

Comprehensive resource of Real World Data (de-identified patient longitudinal or cross-sectional survey information) accessible by any Pfizer team for clinical research or commercial analysis

EMR

- HUMEDICA
  - US: Normalized database of EMR data from 150 US providers
- cegedim Strategic Data
  - UK: THIN database of NHS primary care centers

CLAIMS, ADMIN.

- OPTUMInsight™
  - US: UnitedHealth database of administrative, claims data
- TRUVEN HEALTH ANALYTICS™
  - US: Formerly Thomson MarketScan administrative, claims data
- PREMIER
  - US: Administrative, claims data from over 500 hospitals
- SDI
  - US: IMS subsidiary provider-level written Rx

Linked

- HUMEDICA + OPTUMInsight™
  - US: Linked claims and EMR data

PUBLIC

- BRFSS, NHIS, NHANES, NIS, SEER, DAWN, MEPS, CHIS

Additional Data on Datasources is located at: http://ecfd.pfizer.com/sites/rwdna/SiteContent/pages/Datamart.aspx

Pfizer RWDnA Partnerships

- Humana
- Humedica
- OptumLabs
- IMEDS (OMOP)
- CancerLinQ