

# Use of Electronic Health Records for Development and Feasibility Testing of Clinical Trial Protocols

Manfred Stapff, TriNetX, Cambridge, MA, USA

## Introduction:

Clinical Trial Protocols define the target patient population through various inclusion and exclusion criteria. These criteria may be well justified by safety, scientific, regulatory, or operational reasons, but very often limit the number of available patients by a significant amount.

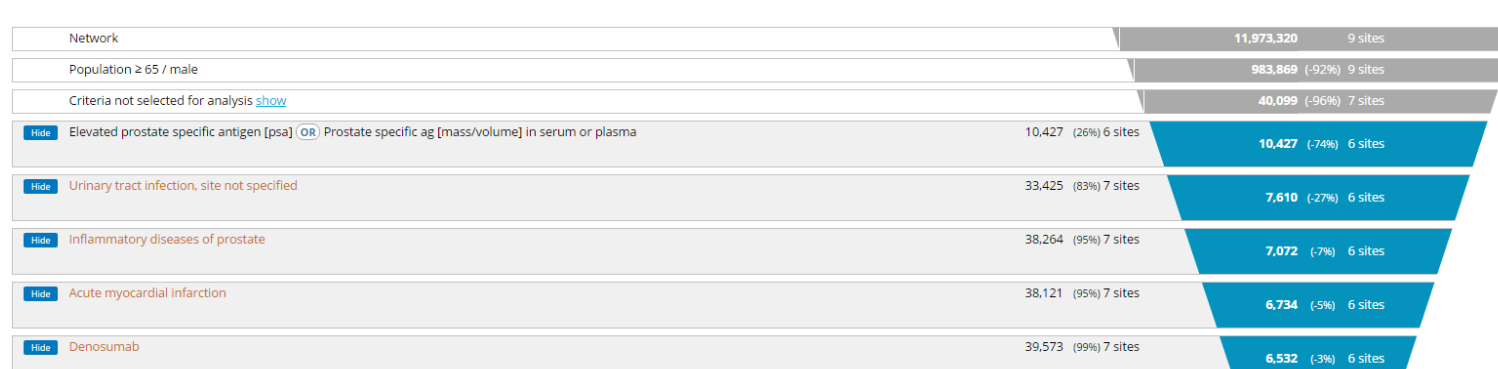
TriNetX is a federated network, providing real world data based on Electronic Health Records (EHR), which can be used to test Clinical Trial Protocols for feasibility, especially evaluating the impact of individual in- or exclusion criteria on patient availability (fig 1). In a second step, the patient cohort as defined by the criteria can then be located at member institutions and can be re-identified there for potential enrollment into the trial.

Although the network is continuously growing and has a diverse portfolio of provider members, it needs to be shown how well the network population represents the general population.

Fig 1: Simplified example of a funnel analysis using EHR

Inclusion criteria: Male patients > 65 years of age with Prostate Cancer and elevated PSA.

Exclusion criteria: Urinary tract infection, prostatitis, Prolia (denosumab) medication



Example: The funnel shows that the requirement for increased PSA and the exclusion of an urinary tract infection have the largest impact on patient availability.

## Objective:

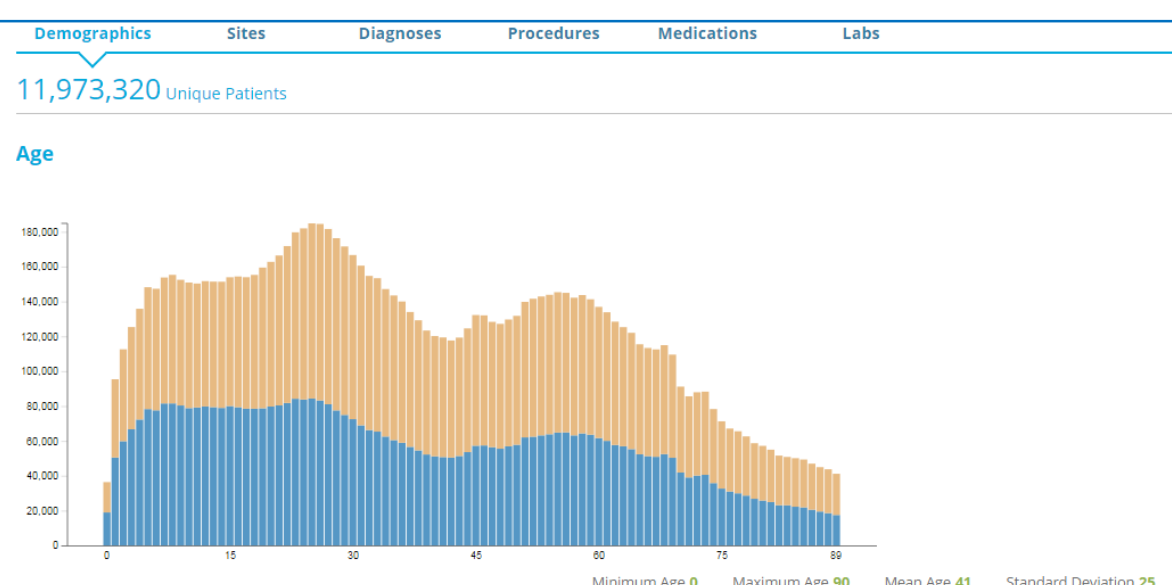
To evaluate whether a federated data network of health care providers supplying aggregate counts on electronic health records (EHR) is correctly representing a real world disease landscape so it can be used for development and feasibility testing of clinical trial protocols in various therapeutic areas.

## Method:

ICD-9 diagnoses provided by a network of 7 health care providers in the US, consisting of 11 million patient lives, were analyzed. TriNetX cohort analyzes were compared to select published epidemiologic data in therapeutic areas of interest for the pharmaceutical industry with special focus on:

- Frequency of diagnoses
- Demographics, age distribution, including smoking status
- Correlation of various risk factors and complications

Fig 2: Network Demographics (as of January 2016)



## Results:

### Most frequent diagnoses:

As of January 2016, seven organizations provide anonymized statistics on 11,973,320 patients in the age of 1 months to 89 years. The information is updated approximately every two to four weeks. The three most frequent individually coded diagnoses were essential hypertension, hyperlipidemia and diabetes mellitus in 979,822 (8.6%), 883,441 (7.8%) and 429,999 (3.8%) of all patients, respectively. These diagnoses were identical to the three most common conditions in the US as published in the literature. Other frequent diagnoses were acute upper respiratory infection 548,073 (4.8%), asthma 401,074 (3.5%), anxiety 345,527 (3.0%), and lumbago 322,890 (2.8%).

### Smoking status:

ICD9 code 305.1 for tobacco use was provided for 179,999 patients, 96,731 (53.7%) male and 83,268 (46.3%), comparable to the gender distribution published by the American Heart Association (55.5% and 44.5%).

### Diabetes and cardiovascular risk:

Of the 266,592 patients in the network with ICD9-249 or -250 for diabetes mellitus, a total of 5747 (2.2%) patients had at least one episode of hypoglycemia, compared to 2.1% in the ADVANCE study. 11146 (4.2%) of the patients with diabetes mellitus had a myocardial infarction which constitutes a 12-fold higher frequency compared to 27251 (0.3%) patients in the TriNetX population without diabetes.

### Diagnosis demographics:

There were 47,588 patients with a diagnosis of prostate cancer in the network. 46,887 (98.5%) were aged 50 years and older, 60% white and 16% black or African American. While the mean age was 73 years in the network, CDC's National Program of Cancer Registries (NPCR) reports a general trend to a decreasing mean age (67.2 years).

### TriNetX Network Population:

1. Hypertension 979,822 (8.6%)
2. Hyperlipidemia 883,441 (7.8%)
3. Diabetes mellitus 429,999 (3.8%)
4. Upper respiratory infection 548,073 (4.8%)
5. Asthma 401,074 (3.5%)
6. Anxiety 345,527 (3.0%)
7. Lumbago 322,890 (2.8%)

### Top diagnoses in the US(1):

1. Hypertension
2. Hyperlipidemia
3. Diabetes
4. Back pain
5. Anxiety
6. Obesity
7. Allergic rhinitis

### Smoking Status in TriNetX Network:

smokers	179999	%
smoker, male	96731	53.7%
smoker, female	83268	46.3%

### Smoking Gender Distribution, AHA (2):

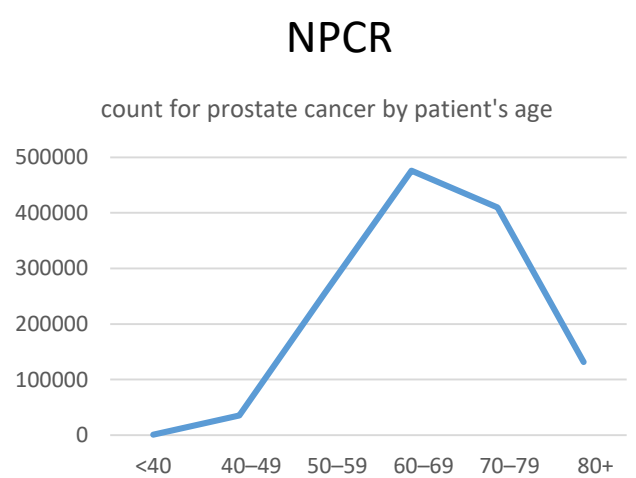
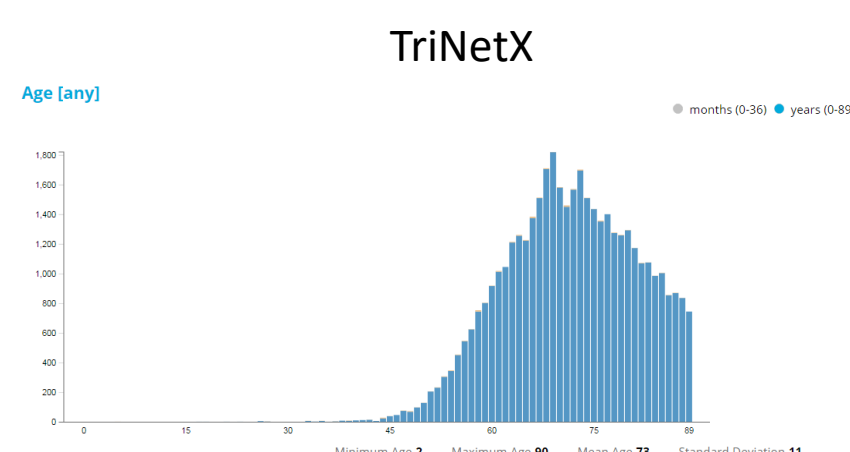
smokers	43415000	%
smoker, male	24080000	55.5%
smoker, female	19298000	44.5%

### Diabetes Complications in TriNetX Network:

Diabetics	266592	
w/hypoglycemic episodes	5747	2.2%
w/myocardial infarction	11146	4.2%
without diabetes	7804537	
w/myocardial infarction	27251	0.3%
Increased risk by		12.0-fold

### Diabetes Complications in the Literature (3, 4):

Diabetics (ADVANCE, 3)	11140	
Hypoglycemic episodes	231	2.1%
Diabetics (NIH, NHDS, 4)	20676427	
w/myocardial infarction	135743	0.7%
without diabetes	226144631	
w/myocardial infarction	135743	0.1%
Increased risk by		10.9-fold



## Conclusion:

The network provides statistics on diagnoses, procedures, demographics, laboratory values, and medications. Since ICD coding is primarily done for billing, its representativity for other purposes was to be validated.

We compared network statistics with select epidemiologic publications (using literature as "gold standard") and found vast consistency:

- The most common diagnoses in the network were identical to the most common conditions diagnosed by physicians in the US, i.e. cardiovascular disease, obesity and diabetes.
- Gender distribution of smokers was similar to the data published by the American Heart Association.
- The percentage of patients with diabetes mellitus (DM) with an episode of hypoglycemia was close to the one observed in the ADVANCE study.
- 6.3% of the patients with DM had a diagnosis of MI, a 12-fold risk compared to the base population. Published data about diabetes-related complications in the United States show a similar risk, i.e. 10.9-fold.
- The age distribution of patients with prostate cancer matched NPCR data. However, NPCR reports a decreasing mean age, lower than in our network. Compared to NPCR the race distribution in our network is slightly shifted towards African Americans (12.3% vs 16%).

Thus, very close correlations between network data and published literature could be seen in diagnoses, demographics, and risk factors. Demographics, race, or ethnicity of an individual diagnosis may differ slightly from other sources, which may be an effect of locations of network providers at the time of analysis. It is expected that these differences will even out upon further growth of the network.

In summary, the data in the analyzed network of health care providers and sponsors/CROs (TriNetX) represent very well the overall disease landscape of the US.

TriNetX' access to EHR of > 11 million patients can therefore be used for design and feasibility testing of clinical trial protocols.

## References:

- (1) Practice Fusion's Research Division data: SAN FRANCISCO – February 9, 2011 The top 20 most common conditions diagnosed by primary care and specialist physicians <http://www.practicefusion.com/pages/pr/health-research-reveals-top-20-chronic-diseases.html> , accessed Jan/20/2016
- (2) Heart Disease and Stroke Statistics—2015 Update; Circulation. 2015;131. Table 3-1
- (3) Heller S.R.; A Summary of the ADVANCE Trial. Diabetes Care. 2009 Nov; 32 (Suppl 2): S357–S361.
- (4) Gregg E.W.; Changes in Diabetes-Related Complications in the United States, 1990–2010. N Engl J Med. 2014; 370: 1514–1523
- (5) Li J.; Recent Trends in Prostate Cancer Incidence by Age, Cancer Stage, and Grade, the United States, 2001–2007. Prostate Cancer, Volume 2012 (2012), Article ID 691380, <http://dx.doi.org/10.1155/2012/691380>

DIA 28<sup>th</sup> Annual  
EuroMeeting

6-8 April 2016 | ID#16101  
CCH Congress Center  
Hamburg, Germany

INnovation • Do You Win by Being IN?