COVID-19 Evidence Accelerator – Summary

CONTENTS

SUMMARY: Given the current circumstance, there have been intense efforts to rapidly deploy data collection efforts to help answer key questions related to the management of COVID-19 patients. While each of these efforts will likely develop into valuable data sources over time, there is an urgency to accelerate and maximize the utility of information in the near term. To do this effectively, the development of a core set of common data elements is needed. This would allow any willing data collection effort to embed the common data elements into their on-going work in a uniform way to allow for rapid aggregation and analysis.

Further, a set of common high priority COVID-19 research questions would be very valuable, especially questions that can be answered by many data and research stakeholders via a variety of approaches. Answering similar questions using several approaches will improve the reliability and credibility of results, especially at a time when traditional formal clinical trials and drug development is less feasible.

GOALS:

1) Establish a common set of core data elements to embed into existing data collection efforts
2) Identify an initial set of immediate questions regarding the therapeutic interventions, treatment settings, and associated outcomes that could be the initial focus of COVID-19 data aggregation efforts.

PROCESS: A broad group of subject matter experts, public and private data and research organizations, and government leaders were requested to review an initial proposed set of common core data elements. This will serve as a feasibility check and seek to develop a degree of consensus to this approach. As part of this effort, we will solicit written submissions of key questions that this initiative can focus toward using the agreed upon core data elements from multiple sources. This would be a rapid process to pull together maximum input in a condensed period – 1 week.

Interested organizations were given 1) DRAFT - core data elements 2) DRAFT - key questions; as well as a template "worksheet" to provide additional feedback and further characterize the Key Questions.

*****41 organizations were provided the documents for review and 25 responses were received and used to develop the summary of feedback.**

KEY QUESTION LIST		<2 weeks	Existing Efforts (n)
General epidemiology of COVID-19		8	8
Predictors of patients at risk for development of severe COVID-19 disease		6	1
Patterns of general outcomes for people with COVID-19 (e.g., death, time to disease resolution)		5	2
Patterns of COVID-19 diagnostic testing and results		8	2
Patterns of development of COVID-19 immunity across the US population		1	1
Can real world data support the evaluation of the performance characteristics of COVID-19 diagnostics?		2	
Are there data that could help identify an evolving COVID-19 hot spot before molecular testing results become available?		3	1
What medications are doctors prescribing for COVID-19 in the real world?		3	1
	What treatments are being prescribed?	7	1
	Which patients are most likely to get which treatments?	6	1
	Which treatments are being prescribed in the context of clinical trials?	1	Registered trials [clinicaltrials.gov]
	Treatment patterns for specific subpopulations (e.g., pregnant women, underlying COPD)	7	1
Patterns of enrollment in COVID-19 clinical trials		1	Registered trials [clinicaltrials.gov]
Can real world data provide initial understanding of safety and effectiveness of therapies used for COVID-19?		4	2
	In particular: safety of hydroxychloroquine and chloroquine, with or without azithromycin	4	1
	Predictors of treatment safety and effectiveness	4	1
	Safety for specific subpopulations (e.g., pregnancy)	3	1
Are there data that can inform risk and/or management of drug shortages (e.g., surge in demand, available drug supply)?		2	
Can data sources be used to help identify patients who can donate convalescent plasma?		1	

Table 1. List of original Key Questions. Provides the total number of respondents that believe the question could be answered in <2 weeks. Catalogues existing efforts that respondents identified that may be able to provide evidence toward each question.

RANKED BY TIMEFRAME (score/respondents)		Responders	Avg	(<2 weeks)
Patterns of COVID-19 diagnostic testing and results		11	**1.36**	8
General epidemiology of COVID-19		13	**1.38**	8
What medications are doctors prescribing for COVID-19 in the real world?		5	**1.40**	3
	What treatments are being prescribed?	12	1.42	7
	Which patients are most likely to get which treatments?	11	1.45	6
	Treatment patterns for specific subpopulations (e.g., pregnant women, underlying COPD)	11	1.45	7
	Which treatments are being prescribed in the context of clinical trials?	6	2.17	1
Are there data that could help identify an evolving COVID-19 hot spot before molecular testing results become available?		7	**1.43**	3
Are there data that can inform risk and/or management of drug shortages (e.g., surge in demand, available drug supply)?		4	**1.50**	2
Predictors of patients at risk for development of severe COVID-19 disease		13	**1.54**	6
Can real world data provide initial understanding of safety and effectiveness of therapies used for COVID-19?		8	**1.75**	4
	In particular: safety of hydroxychloroquine and chloroquine, with or without azithromycin	10	1.70	4
	Predictors of treatment safety and effectiveness	10	2.70	4
	Safety for specific subpopulations (e.g., pregnancy)	10	2.90	3
Patterns of general outcomes for people with COVID-19 (e.g., death, time to disease resolution)		14	**1.86**	5
Can real world data support the evaluation of the performance characteristics of COVID-19 diagnostics?		5	**2.00**	2
Can data sources be used to help identify patients who can donate convalescent plasma?		5	**2.00**	1
Patterns of enrollment in COVID-19 clinical trials		8	**2.13**	1
Patterns of development of COVID-19 immunity across the US population		7	**2.29**	1

Table 2. Key Questions according to Timeframe. This gives a general sense as to whether evidence could be gathered to address each question in the short term vs. longer term priorities. Respondents were asked to determine whether each question could be answered <2 weeks (=1), 3-12 weeks (=2), or 3+months (=3). Key Questions are sorted by the average timeframe as identified by all respondents. Total numbers less than 2 may be an indicator that several organizations believe it is a short term analysis, those greater than two may be more appropriately grouped as a longer term initiative.

RANKED BY RESPONSES (as a potential indicator of feasibility)		Responses	Avg
Patterns of general outcomes for people with COVID-19 (e.g., death, time to disease resolution)		**14**	1.86
General epidemiology of COVID-19		**13**	1.38
Predictors of patients at risk for development of severe COVID-19 disease		**13**	1.54
Patterns of COVID-19 diagnostic testing and results		**11**	1.36
Patterns of enrollment in COVID-19 clinical trials		**8**	2.13
Can real world data provide initial understanding of safety and effectiveness of therapies used for COVID-19?		**8**	1.75
	In particular: safety of hydroxychloroquine and chloroquine, with or without azithromycin	10	1.70
	Predictors of treatment safety and effectiveness	10	2.70
	Safety for specific subpopulations (e.g., pregnancy)	10	2.90
Patterns of development of COVID-19 immunity across the US population		**7**	2.29
Are there data that could help identify an evolving COVID-19 hot spot before molecular testing results become available?		**7**	1.43
Can real world data support the evaluation of the performance characteristics of COVID-19 diagnostics?		**5**	2.00
What medications are doctors prescribing for COVID-19 in the real world?		**5**	1.40
	What treatments are being prescribed?	12	1.42
	Which patients are most likely to get which treatments?	11	1.45
	Treatment patterns for specific subpopulations (e.g., pregnant women, underlying COPD)	11	1.45
	Which treatments are being prescribed in the context of clinical trials?	6	2.17
Can data sources be used to help identify patients who can donate convalescent plasma?		**5**	2.00
Are there data that can inform risk and/or management of drug shortages (e.g., surge in demand, available drug supply)?		**4**	1.50

Table 3. Key Questions by Number of Responses. This gives a general sense of the feasibility of gathering evidence from various sources for each question. The greater number of responses that each question received may be a general indicator of the extent to which responders believe that obtaining the evidence necessary to address the question is readily available.

Table 4. **Additional Questions.** Respondents provided additional questions for consideration. These are from various vantage points (clinical, manufacturers, health systems). These may provide insight into additional priorities.

		ADDITIONAL QUESTIONS	SOURCE
1)		For patients who are receiving drugs associated with COVID treatment for other reasons (e.g. hydroxychloroquine for lupus or RA, is there a protective effect that can be identified?	Insurer
2)		Which patients are likely to develop mild/moderate disease? [https://covidwatch.dukehealth.org/]	Data Aggregators / Academic Med Centers
3)		Can remote monitoring be used to (a) detect and monitor outbreaks, and (b) conduct safer and more efficient clinical trials during an outbreak? [https://www.elektralabs.dev/covid-19]	Data Aggregators / Academic Med Centers
4)		How should patients with COVID-19 and concomitant chronic diseases requiring chronic therapy be managed (i.e. Benefit/risk of continuing/adapting/discontinuing chronic therapies)?	Pharmaceutical Company
5)		What is the impact of COVID19 on (observational) studies spanning the pandemic period but investigating non-COVID19 related outcomes:	Pharmaceutical Company
	5a)	Local/regional/national impact of COVID19 crisis on healthcare utilization for non-COVID19 diseases	Pharmaceutical Company
	5b)	Change in outcomes of non-COVID19 diseases (e.g. complications, emergency situations, fatality) due to altered healthcare	Pharmaceutical Company
	5c)	Short-term and long-term impact of COVID-19 on other co-existing illnesses	Pharmaceutical Company
	5d)	Change in Cause of Death statistics due to COVID19 (e.g. to assess potential impact on studies with mortality/survival endpoints)	Pharmaceutical Company
	5e)	Potential impact on national statistics and other data sources due to altered healthcare capabilities, e.g. potential differences in death certificates, hospital admission statistics, general coding practice in hospitals and other institutions	Pharmaceutical Company
	5f)	Identification of time periods during which COVID-19-related changes are relevant to inform stratification in pre / during / post COVID-19 period for ongoing studies	Pharmaceutical Company
6)		Long-term outcomes of COVID-19: Reinfection rates, potential seasonality patterns, long-term consequences after recovery of severe COVID19 infection, e.g. long-term lung damage / other organ damage in recovered individuals	Pharmaceutical Company

7)		Public health impact of the pandemic and lessons learned up to this point	Pharmaceutical Company
	7a)	Non-pneumonia morbidity and mortality in the general population resulting from paralysis of the healthcare system by the COVID-19 pandemic.	Pharmaceutical Company
	7b)	Prevention and mitigation of adverse effects of such pandemics in the future (e.g., improved outbreak investigation methods, early implementation of travel restrictions and quarantine measures, stockpiles of PPE, ventilators, development of remote learning and remote working capabilities, advancement of tele-medicine)	Pharmaceutical Company
8)		Immunity and vaccination	Pharmaceutical Company
	8a)	Duration of immunity, risk of re-infection, virulence on reinfection.	Pharmaceutical Company
	8b)	Effectiveness and safety of vaccine candidates, need for booster doses and/or seasonal re-vaccination.	Pharmaceutical Company
9)		Genetic characteristics of the virus, molecular and cellular pathogenesis	Pharmaceutical Company
	9a)	Genetic diversity of the virus, circulating sub-strains, implications for virulence and testing	Pharmaceutical Company
	9b)	Molecular and cellular pathogenesis, relative contribution of direct cytopathic effect versus cytokine storm (or other mechanisms) to adverse disease outcomes, pulmonary and extra-pulmonary effects	Pharmaceutical Company
10)		Can we detect asymptomatic spread using available data	Health System / Insurer
11)		Are we considering case definition?	Health System / Insurer
	11a)	And general epidemiology of presumptive or suspected COVID-19 [because of the severe lack of testing, we really need to understand more about the presumed infected and get better and better about identifying those people]	Health System / Insurer
	11b)	Identified either by health system intake (eg RN phone triage) or even through self-administered CDC screening questions on the web or in apps (This will help us start getting a pulse on what's happening in the community in the absence of testing)	Health System / Insurer
12)		Effectiveness of relying on coding versus syndromic and NLP-based case evaluation-Initial analyses planned in 2020 and applicable to future epi/pandemic scenarios	EMR

13)		What are the computable phenotype electronic algorithms for COVID-19 infection and resulting complications for administrative claims, electronic health record (EHR), and integrated administrative claims and EHR data environments?	Health System / Insurer
14)		What are the endpoints, treatment patterns, and underlying baseline comorbidity conditions among people with COVID-19? Are there baseline comorbidities or exposures with different rates of complications (e.g. ACEI, ARBs, COPD, Asthma).	Health System / Insurer
15)		What treatments are being used for patients with COVID-19, what are the baseline characteristics of those receiving them, and what is the comparative effectiveness and safety of the therapeutic interventions? What available treatment unexposed populations exist for comparison?	Health System / Insurer
16)		How can antibody testing inform epidemiology and patients that recover?	
17)		Safety of IL6 and JAK inhibitors-OHDSI has completed the design of a study to examine the safety of IL6 inhibitors and JAK inhibitors. Protocol and analysis code is posted: https://github.com/ohdsi-studies/Covid19EstimationIl6JakInhibitors. Analysis is being executed across OHDSI network, but prioritized to be completed after hydroxychloroquine safety and cauterization studies. Tentative target for completion: 17Apr2020.	Data Aggregators / Academic Med Centers
18)		Safety of protease inhibitors for HIV and Hepatitis C-OHDSI has begun to design a study to examine the safety of protease inhibitors in HIV and Hepatitis C. A protocol has been drafted but not yet finalized. A methodological challenge with examining the effects of these drugs is that they are often used in combination with other medicines, so separating out the ingredient-specific effect requires special attention.	Data Aggregators / Academic Med Centers
19)		ACE inhibitors and Angiotensin Receptor Blockers and risk of COVID-19 susceptibility and severity-OHDSI has completed the design of a series of studies to examine the effect of ACE inhibitors on COVID-19 susceptibility and severity. The protocol is posted: https://github.com/ohdsi-studies/Covid19EstimationRasInhibitors. The analysis code has been prepared and is currently completing testing. The analysis has been successfully performed within a data partner in South Korea, but was underpowered. These studies require sufficient sample of ACE inhibitor exposure amongst COVID-positive patients, so it will likely be multiple weeks before data can accumulate to satisfactorily answer the question.	Data Aggregators / Academic Med Centers
20)		What are the rates of endpoints, treatment patterns, and underlying conditions among people with COVID-19?	Health System / Insurer
21)		What are the potential electronic algorithms for COVID-19 infection and complications?	Health System / Insurer

22)		Do people with COVID-19 have different rates of complications depending on their use of drugs for other underlying conditions?	Health System / Insurer
23)		Are there marketed drugs that can lower or increase the risk of COVID-19 infection?	Health System / Insurer

REAGAN-UDALL FOUNDATION for the Food and Drug Administration

Data Elements Summary

I. Descriptions

a. Demographics

i. Consensus: age, gender, race, ethnicity

ii. Clarifications & Modifications:
   1. HCW - needs definition
   2. Occupation – more important might be work setting as proxy for exposure

iii. Additions: *Themes*: (1) Additional covariates: socio-economic status (important to understand health disparities); (2) Populations to capture: Sample general population

b. Location

i. Consensus: Country, State

ii. Clarifications & Modifications:
   1. City, Zipcode – census tract is preferred, but may not be available. Consider classification by urban vs. rural.
   2. Facility – clarify whether this refers to where pt initially tested/identified, or managed (which could be proxy for severity)
   3. Treatment setting – specify settings with higher/lower intensivist providers

iii. Additions: (1) **healthcare capacity** of region at time of diagnosis, including ventilator, PPE supply, beds, etc. (as a risk factor for higher morbidity/mortality); (2) Mobility data to link to COVID-19 cases to create spatial risk models (for hot spotting); (3) **Type of housing** (e.g. institutional, group home, long-term care, etc.)

c. Medical History

i. Consensus: Blood type, smoking, allergies, on oxygen, recent surgery, other procedures, hospitalization

ii. Clarifications & Modifications:
   1. Significant Comorbidities – only asthma, emphysema, COPD were highlighted – though more were initially included. *Themes*: (1) why not include all and calculate **risk score**?; (2) include **CVD, diabetes, cancer, lung/respiratory disease, autoimmune disease**
   2. Concomitant Meds – *Themes*: (1) include specific meds: **ACE, ARB**; NSAID, acetometaphin; **BP** med; antibiotic; statins; **immune modulating** drug; (2) debate on the need to capture drug vs. drug class; (3) BCG vaccine

d. Onset History

i. Consensus: date symptoms 1st reported, COVID-19 testing method, sample collection, sample collection date, COVID-19 DX, DX date, symptom changes, lab positive, recent travel history, known exposure, date of known exposure, characteristics of known exposure, admission from home, admission from institution (prison, nursing home, other group housing)

ii. Clarifications & Modifications:
   1. COVID-19 lab test date – needs definition (e.g. sample collection date, result date)
   2. Symptom duration – needs definition
   3. 1st symptom date – can only be estimated or patient reported
   4. Symptoms at onset – add GERD, conjunctivitis

(202) 849-2075 1900 L Street NW, Suite 835, Washington, DC. 20036 www.reaganudall.org

e. COVID Status and Progression

i. Consensus: Dates listed here are very important for epi parameters used in disease-transmission models. Vital signs, increased oxygen requirement, hospitalization, MEWS score, # days hospitalized, ICU course of disease, level of consciousness, coinfection, superinfection (y/n), superinfection (details), pneumonia, **ARDS**.

ii. Clarifications and Modifications:

1. Current symptoms – describe source (e.g. chief complaint, patient history, self-report vs. transcribed)
2. Temperature – describe temp (e.g. max in 2 hrs)
3. Oxygen saturation – important to know, also hospital vs. home device
4. Severity of COVID19 – how will this be collected? Needs definition.
5. Admission date – important for understanding epi
6. ICU admission – need type of mechanical ventilation (intubation vs. high-flow nasal canula)
7. COVID-19 complications – needs definition
8. Enrolled in COVID-19 clinical trial/specific COVID-19 trial/non-COVID clinical trial – not always identifiable in healthcare data; only clinical trials
9. Stigma – needs definition. Not usually captured in EHR. Not in claims.

iii. Additions – (1) Readmission rates and status at readmission as an outcome (2) Complications (e.g., respiratory failure, sepsis, multi-organ failure, **ARDS**, etc.)

f. Supportive Care

i. Consensus: Oxygen and date, Ventilation date, ECMO and date, Cytokines administered, pressors, CRRT

ii. Clarifications & Modifications:

1. On ventilation – include non-invasive positive pressure ventilation

iii. Additions: cardiovascular support, dialysis

g. COVID19 Specific Meds

i. Consensus: Medication, Medication Dosing Regimen, Adverse Event & date

ii. Clarifications & Modifications:

1. Medication Discontinuation Reason – mostly from clinical trials; not structured EHR data; not in claims

iii. Additions - tolerability

h. Other medications

i. Consensus: Medication, Medication Dosing Regimen, Adverse Event & date

ii. Clarifications & Modifications:

1. Medication Discontinuation reason – mostly from clinical trials; not structured EHR data; not in claims

i. Clinical Outcomes

i. Consensus: Time to clinical progression; Improvement on use of supplemental oxygen or mechanical ventilation; Amelioration of routine lab criteria and pulmonary function; Increase of neutralizing Ab titers & disappearance of SARS-CoV-2 RNA.

ii. Clarifications & Modifications:

1. Recovered from COVID-19 – needs definition
2. Time to clinical improvement – if based on ordinal scale, this is a clinical trial construct. Not often in RWD
3. Reduction of duration of hospital or ICU stay (in survivors) – unclear. Does this refer to shorter hospital stay in survivors vs. those who died (i.e. outcome=length of stay)? If so, there is potential immortal time bias because of classifying an exposure (i.e. death vs. survival) after the start of follow-up (i.e. Hosp/ICU admission). Seems confounded b/c the healthcare system can keep one alive through mechanical means for long periods of time – so time to death tends to be artificially elongated.

4. Improvement on symptoms – include improvement on weakness on standing, sore throat resolution, improved taste or smell, improved lightheadedness.
5. Potential candidate for donation of convalescent plasma – based on recovered status and ability to send updates to patient

iii. Additions – For pregnant populations: pregnancy and birth outcomes, including COVID-19 infection status of newborn.

j. Lab/Imaging

i. Consensus: All listed but COVID-19 serologies, CBC and CRP

ii. Clarifications & Modifications:

1. CBC – full WBC differentials if possible
2. CRP – also include CPK
3. COVID-19 serology – distinguish across current infection, immunity, and strains.

iii. Additions – *Themes*: (1) Expand testing in populations: a) Repeated serology to understand resolved asymptomatic infection; b) Random sample general population and/or those with mild symptoms to understand true prevalence (2) Capture COVID19 testing policy in place for region at time of diagnosis: everybody tested, broad testing in population / contact persons of COVID19-positive persons tested / only tested if contact with COVID19 positive person and symptoms / only tested if admitted to hospital / no tests done asymptomatic people from general population.

***Highlighted** data elements were reported by >1 respondent.

II. Feasibility

Respondents

A total of 26 organizations responded: 7 Health Systems, including a large clinical network; 6 insurers, 5 Data Aggregators; 2 Technology Companies; 1 registry, 3 EMR, and 2 pharmaceuticals companies

14 of 26 contained EHR data; and 14 contained claims data (some had both). All but 1 had inpatient and outpatient data. One system was outpatient only.

Data Availability

Generally, all organizations could answer priority data elements in the following categories: **Demographic, Location, and Medication History** (though smoking status may not be well populated).

Information on priority data elements related to **Onset of Symptoms** will be difficult to capture in RWD – this is better suited for Clinical Trials. **Clinical Trial information** is not easily identified outside of clinical trials. **Reason for discontinuing medication** is not usually captured or easy to define in EHR data; and is unavailable in claims. Finally, **laboratory results and images** are often not available in claims – only reimbursement claims for such procedures are available. EHR data will contain such data.

Lag Times

Certain questions depend on more real-time data. Questions addressing epidemiologic parameters and death need more timely data. Generally speaking, refresh times for insurers tend to be more delayed than health systems.

Limitations/Benefits of Claims vs. EHR

Limitations of claims systems include missing lab results, misclassification of inpatient procedure dates because of roll-up to admission or discharge dates. Outpatient dates are fine. Since billing codes do not necessarily map directly to clinical constructs, some validation may be necessary. For medications -

dispensing information is available, but not prescription - making adherence questions challenging to answer.

A major benefit of claims systems is the ability to derive a denominator among health plan enrollees. This population is generally less biased than the underlying population that present at EHR systems. This availability of enrollment start and stop dates allows for more accurate calculation of person-time.

EHR systems have deeper data, including lab results, physician notes, pharmacy prescriptions and fills (for integrated systems), images, etc. However, unless integrated, these systems don't have a denominator. Though work-arounds can be created to identify a base-population who sought care for at a minimum level – this will tend to bias towards a sicker population; or women. This may lend to biased estimates.

III. Major Additions

More than 1 respondent felt it was important to identify and sample cohort of general population to understand asymptomatic infection.

Table 5. Data Element Descriptions. Provides any suggested modifications by data element as well as any questions received. Core data elements indicated by pink highlight. This could help to further clarify or revise the core set of data elements.

CATEGORY	Data Element	ORGANIZATION COMMENTS							
		Org 1	Org 2	Org 3	Org 4	Org 5	Org 6	Org 7	Org 8
Demographic Info			Socioeconomic status would be helpful too						
	• Patient ID			MRN Number					
	• Age								
	• Gender								
	• Race								
	• Ethnicity								
	• Occupation								
	• HCW (y/n)								Y: possibly identify work setting as a variable for gauging intensity of exposure and its impact on viral load/intensity of
Location									
	• Country								
	• State								
	• City					I think we will need more granular data than city. In fact, census tract would be better, but is not available at point of care			
	• Zip Code		Zip codes are not very useful when thinking in geo-coded units. County would be much more useful to link to other data sources e.g. Census, county-level mobility patterns					(useful for census and linkage)	
	• Facility			For Inpatient					
	• Treatment setting (e.g., hospital, clinic, inpatient, outpatient)			Inpatient / Outpatient				Is this the treatment setting where initial testing occurs or the setting where the patient is managed? Otherwise it could be: inpatient (severe), inpatient, observation, outpt: urgent care outpt: office or clinic	Should this include Home Care and SNF or are these lumped into outpatient for home care and inpatient for SNF?
Medical History									
	• Blood type								
	• Smoking history (current smoker, former smoker, never smoker, vaping)								
	• Allergies								
	• Significant comorbidities (e.g., asthma, emphysema, COPD, cancer, diabetes, hypertension, cardiovascular disease, chronic renal disease,		Even for the 'early questions/analysis', we'd want to get as many comorbidities identified In addition, we could use these to calculate some measure of 'severity' of people who have multiple conditions.		(Cancer completed treatment, Cancer active treatment, Cystic Fibrosis, HIV/AIDS, IBD, Pulmonary Fibrosis, Theumatoid Arthritis, Diabetes Type 1,		Not sure why you wouldn't capture these data elements. Many are already in the EHR. It is true that EHR has an			

	chronic liver disease, immunocompromised condition, neurologic/neuro developmental disability, other chronic disease, currently pregnant)	Count of concomitant meds could also be used to get at 'severity'.	? history of thrombosis or increased disposition to thrombosis	Heart Arrhythmia, Kidney disease, Stroke, Asthma, Cardiomyopathy, Diabetis Type 2, Heart Disease, Lung Disease, TB, Autoimmune Disease, Congestive heart Failure, COPD, Hypertension, MS)	CAD	incomplete list, so worth while asking pt, but if you get it from EHR it saves time asking pt. *CRW - I think he means that it can be captured from EHR using labs or DX codes and not necessarily thru patient history.		Add hx of thrombosis	
	• Concomitant medications (detailed list or by drug category)	Recent publications described the how SARS-CoV-2 binds to the human receptor angiotensin-converting enzyme 2 (ACE2). Are there any drugs currently being used for other conditions that might interfere with this binding or that modulate ACE2 that we should be certain to include in analysis?	NSAIDs/acetaminophen could be useful as specific items since they are non-Rx and there have been questions about whether they increase risk of severe disease	ACE or ARB, Immune modulating Drug, Antibiotic, Statin, Blood Pressure Medication, Hydroxychloroquine		Need drug, not class		Use drug category mapped to ATC or RxNorm for appropriate classification	
	• On oxygen (Y/N)								Continuous or prn
	• Recent surgery			Text Field "Describe other medical conditions"					
	• Other procedures			Text Field "Describe other medical conditions"					
	• Hospitalization within past 30/60/120 days			Text Field "Describe other medical conditions"					
Onset History									
	• First symptom date							Estimated	
	• Date of first consultation/symptom first reported								
	• Symptoms at disease onset (e.g., fever > 100.4, subjective fever, chills, cough, sore throat, shortness of breath, nasal congestion, sputum production, chest pain, diarrhea, nausea and vomiting, headache, muscle pain, abdominal pain, loss of taste, loss of smell and malaise)				conjunctivitis			Add GERD	
	• COVID-19 lab test date			Date Sample Taken				Nasal	
	• Testing Method for COVID-19 diagnosis								
	• Sample collection (NP Swab, OP Swab, Sputum, Other)								
	• Date of Sample Collection			Date Positive Test Result					
	• COVID-19 diagnosis date								
	• Symptom duration (before diagnosis or hospitalization)								
	• Symptom change since last reported date (improved/worsened)							How would this be measured? Could measure symptoms at regular interval, as in at 7 days what are the symptoms?	
	• Laboratory positive (i.e., viral nucleic acid test positive) (Y/N/pending)								
	• COVID-10 diagnosis								
	• Recent Travel History (location, date of arrival, date of return)								
	• Known Exposure to COVID-19 (Y/N)								
	• Date of Known Exposure to COVID-19								
	• Characterization of Known Exposure (Household/Community/Healthcare/Occupational/Travel)								
	• Admission from community (home) (Y/N)								
	• Admission from institution (prison, nursing home, etc) (Y/N)								
COVID Status and Progression									

	• Current symptoms (e.g., fever > 100.4, subjective fever, chills, cough, sore throat, shortness of breath, nasal congestion, sputum production, chest pain, diarrhea, nausea and vomiting, headache, muscle pain, abdominal pain, loss of taste, loss of smell and malaise)						Our practices are identifying their patients for screening for Covid. If found to be positive they are managed by a separate medical team (outside of US Oncology/McKesson). We capture their treatment and outcomes from progress notes, hospital notes which are typically provided to the oncology practice. However we may not have all the details described in the COVID Status and Progression section since our practices are not directly managing the Covid patients.		
	• Vital signs (Blood pressure, heart rate, respiratory rate)								
	• Temperature			Max in 2 hours					does the method of measurement matter (i.e. oral, axillary, temporal)
	• Increased Oxygen Requirement (Y/N, or PaO2/FiO2 ratio or SpO2/Fi)2)								
	• Oxygen Saturation					Definitely important; would also like to receive this from home SpO2 devices since we will need to monitor more people from home either pre- or post-hospitalization			are their specific metrics for what constitutes each category? If not the assessment may be too subjective
	• Severity of COVID-19 (mild, moderate, severe)		Would this information be routinely collected in a standardized way?	None, Mild, Moderate, Severe, Critical					
	• Hospitalization (Y/N)								
	• MEWS Score at Hospital admission								
	• Date of admission					Needed. Timing of exposures, symptoms, and hospitalizations are important for understanding disease and epi			
	• Days hospitalized								
	• Date of hospital discharge								
	• ICU (Y/N)								
	• ICU Admission Reason (respiratory failure requiring mechanical ventilation, shock identified by the use of vasopressor therapy and elevated lactate levels (>2 mmol/L) despite adequate fluid resuscitation, failure of other organs requiring admission to the intensive care unit)		Type of mechanical ventilation - intubation vs high-flow nasal cannula (HFNC) may be useful information to capture						
	• Days in ICU								
	• Date of ICU Admission		Prioritize (dates can give insights into important epi parameters used in classical disease transmission models; also can derive outcomes of interest like severity of disease requiring hospitalization)						
	• Date of ICU Discharge								
	• APACHE II Score at ICU Admission								
	• SOFA Score at ICU Admission								
	• ICU course of disease								

	• Level of consciousness								
	• Coinfection (Y/N)			Text field available					
	• Coinfection (detail)								
	• Superinfection (Y/N)								
	• Superinfection (details)								
	• Did patient develop pneumonia (Y/N)								
	• Did patient develop ARDS (Y/N)								
	• COVID-19 complications								
	• Enrolled in a COVID-19 clinical trial (Y/N)								
	• Specific COVID-19 clinical trial								
	• Enrolled in a non-COVID-19 specific clinical trial (Y/N)								
	• Stigma (CDC has had some concerns about stigma among those infected)								CDC has had some concerns about stigma among those infected) – this might be best collected as a PRO with stigma defined so its clear what it means to the reporter
Supportive Care/Standard of Care									
	• O2 (Y/N)								
	• Date								
	• On ventilation (Y/N)				non-invasive positive pressure ventilation				
	• Date								
	• ECMO								
	• Date								
	• Other cytokines (e.g., IFN-gamma, TNF-alpha, IL-1, GM-CSF)								
	• Pressors								
	• CRRT								
COVID-19 Specific Medications								Does this mean COVID Treatment? Would list by category (e.g. antivirals, antimalarials, interleukins, cytokines, antibiotics, steroids, tyrosine kinases, ARBs)	
	• Medication								
	• Medication dosing regimen (Dosage, Start and End Date)								
	• Medication discontinuation reason			Notes Section					
	• Adverse Event			Notes Section					
	• Date of Adverse Event								
Other Medications								What is this measuring?	
	• Medication								
	• Medication dosing regimens (Dosage, Start and End Date)								
	• Medication discontinuation reason								
	• Adverse Event								
	• Date of Adverse Event								
Clinical Outcomes									
	• Recovered from COVID-19 (Y/N)					Need precise definition			Care setting of recovery (i.e. home, hospital, SNF
	· Time to clinical improvement (e.g., defined as the time from baseline – e.g., diagnosis or admission – to an improvement of two points on a 7-category ordinal scale[1] or live discharge from hospital, whichever came first)								
	• Time to clinical progression (e.g., defined as death, mechanical ventilation, or ICU administration)								
	• Improvement on use of supplemental oxygen or mechanical ventilation (e.g., shorted duration)								
	• Reduction of pulmonary lesions on chest CT examinations								

	• Amelioration of routine laboratory criteria and pulmonary function (e.g., lymphocytopenia, parameters indicative of inflammation and/or liver dysfunction such as C-reactive protein, ALT/AST, total bilirubin, and SaO2)								
	• Increase of neutralizing antibody titers and disappearance of SARS-CoV-2 RNA								
	• Reduction of duration of hospital or ICU stay (in survivors)								
	• Improvement on symptoms (e.g., fever, cough, shortness of breath)			Shortness of Breath, Weakness on Standing, Sore Throat, Loss of Taste or sell, Lightheadedness, Dry Cough, GI, Fever)					
	• Potential candidate for donation of convalescent plasma (Y/N)			Based on recovered status and ability to send updates to patient					
	• Death (Y/N)								Setting of death (i.e. home, hospital, SNF
	• Date of death								
Lab/Imaging									
	• Imaging performed (X-Ray, CT, Ultrasound/Echo)								
	• Imaging Dates								
	• Imaging (e.g., CT) results								
	• Lung Involvement (Y/N)								
	• CBC with neutrophil and lymphocyte		Full WBC differential if possible						
	• ALC								
	• ANC								
	• AEC								
	• Hgb								
	• PLT								
	• Creatine								
	• T Bili								
	• AST								
	• ALT								
	• PT								
	• APTT		INR						
	• D dimer								
	• LDH								
	• TNI								
	• BNP								
	• CRP		CPK						
	• Il6								
	• GFR/Creatinine Clearance								
	• Ferritin (lab)								
	• Troponin/CKMB								
	• NT-proBNP								
	• Metabolic Panel								
	• Culture tests/lab microbiology (e.g., MRSA)								
	• SARS-CoV-2 viral load/titer (PCR, serology)								
	• SARS-CoV-2 RNA Detection-Qualitative (e.g., detected or not detected)								
	• SARS-CoV-2 RNA Level-Quantitative (e.g., Ct value, or copies/mL)								
	• SARS-CoV-2 Virus Level (e.g. PFU/mL, TCID50)								
	• Virus amino acid substitution data, including viral target protein, specific amino acid position and substitution, and frequency in population (this is very high level and no doubt we would have a request for more detailed resistance data)								
	• SARS-CoV-2 IgM, IgG or Neutralizing Antibody (e.g., ELISA titer or PRNT50 titer)								
	• Serum cytokine/chemokine level (need analyte name somewhere)								
	• Other labs								

ADDITIONAL DATA ELEMENTS									
	Clinical Outcomes								
			Time to extubation						
			Length of time on ventilation						

Table 6. Data Elements Feasibility. Provides a general sense of organizations capable of collecting core data elements (pink), as well as additional data elements and/or entire categories. This includes the source of data (EHR, Claims, or Both). This provides a landscape of numerous organizations (15+) that could be leveraged to collect similar core data elements and enable robust, rapid analyses.

DATA ELEMENTS	Category	Tech Co		Health Plan / Health System Network										Health Plan / Health System Network	EMR Developer			Data Aggregators (AMCs+Pharma+CROs)					Health System	Health Plan	Pharma		Registry
	Data Source		EHR	CLAIMS	EHR	EHR	Claims	Claims	Medicare Claims	Medicaid Claims	EHR	Integrated EHR +Claims	EHR	EHR, Claims	EHR	EHR	EHR	EHR, Claims	EHR, RX	Claims, EHR	EHR, Claims	EHR	EHR	Claims, EHR	Clin Trials, Claims	Clin Trials, Claims	Claims
	Clinical Setting	inpt/ outpt		inpt/ outpt	inpt/ outpt	inpt/ outpt	inpt/ outpt	inpt/ outpt	inpt/ outpt	inpt/ outpt	inpt/ outpt	inpt/ outpt	inpt/ outpt	inpt/ outpt				inpt/ outpt	outpt		inpt/ outpt	inpt/ outpt	inpt/ outpt	inpt/ outpt			inpt/ outpt
	Freq of refresh			1 yr	monthly	monthly	1 yr	1 yr	1 yr	1 yr	monthly	monthly	monthly									daily					
	Pop Size							118+M													314+M	75+M		EMR= 105M, Claims+ EMR= 20M			
Demographic Info								Y	Y	Y	Y	Y	Y	Y		Y		Y		Y							
	Patient ID	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	Age	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	Gender	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	Race	N		Y	Y	Y	Y	SOME	Y	Y	Y	Y	Y		Y												
	Ethnicity	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	Occupation	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		N												
	HCW	N		DERIVED	N	N	DERIVED	DERIVED	DERIVED	DERIVED	N	N	N		Y												
	(Height / weight)	Y		SOME	Y	Y	SOME	SOME	SOME	SOME	Y	Y	Y														
Location								Y	Y	Y	Y	Y	Y	Y		Y		Y		Y							
	Country	All US to date		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	State	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	City	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	Zip Code	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		N												
	Facility	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		N												
	Treatment Setting	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
Medical History								Y	Y	Y	Y	Y	Y	Y		Y		Y		Y							
	Blood Type	N		SOME	Y	Y	SOME	SOME	SOME	SOME	Y	Y	Y														
	Smoking History	Y		SOME	Y	Y	SOME	SOME	SOME	SOME	Y	Y	Y		Y												
	Significant comorbidities	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	Concomitant Medications	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	On Oxygen	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	Recent Surgery	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	Other Procedures	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	Hospitalization within past 30/60/90 days	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
Onset History								Y	Y	Y	Y	Y	Y	Y		Y		Y		Y							
	First symptom date	Y		N	Y	Y	N	N	N	N	Y	Y	Y		Y												
	Date of first consultation / symptom first reported	N		N	Y	Y	N	N	N	N	Y	Y	Y														
	Symptoms at disease onset			N	N	N	N	N	N	N	N	N	N		Y												
	COVID-19 lab test date	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	Testing Method for COVID-19	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	Sample Collection Method	N																									
	COVID-19 Diagnosis date	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	Symptom duration (before diagnosis or hospitalization)	Derived		N	N	N	N	N	N	N	N	N	N		Y												
	Laboratory positive	Y		N	Y	Y	N	N	N	N	Y	Y	Y		Y												
	COVID-19 Diagnosis	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	Recent Travel History (location, date of arrival, date of return)														Y												
	Characterization of known exposure	N		N	N	N	N	N	N	N	N	N	N		Y												
	Admission from community (home)	N		N	Y	Y	N	N	N	N	Y	Y	Y														
	Admission from institution	N		N	Y	Y	N	N	N	N	Y	Y	Y														
COVID Status and Progression														Y		Y		Y		Y							
	Current Symptoms							Y	Y	Y	Y	Y	Y														
	Vital signs	N		SOME	Y	Y	SOME	SOME	SOME	SOME	Y	Y	Y														
	Temperature	Y		SOME	Y	Y	SOME	SOME	SOME	SOME	Y	Y	Y														
	Increased Oxygen Requirement (Y/N, or PaO2/FiO2 ratio or SpO2/Fi)2																										
	Oxygen Saturation	N		SOME	Y	Y	SOME	SOME	SOME	SOME	Y	Y	Y														
	Severity of COVID-19 (mild, moderate, severe)	Y		N	DERIVED	DERIVED	N	N	N	N	DERIVED	DERIVED	DERIVED														
	Hospitalization (Y/N)	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	MEWS Score at Hospital admission	N		N	Y	Y	N	N	N	N	Y	Y	Y														
	Date of Admission	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	Days Hospitalized	DERIVED		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	Date of Hospital discharge	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	ICU (Y/N)	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	ICU Admission Reason	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	Days in ICU	N		DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED														
	Date of ICU Admission	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	Date of ICU Discharge	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	APACHE II Score ar ICU Admission	N																									
	SOFA Score at ICU Admission	N																									
	ICU course of disease	N																									
	Lost of consciousnes	N																									
	Coinfection (Y/N)	N			Y	Y					Y	Y	Y														
	Coninfection detail	N			Y	Y					Y	Y	Y														
	Superinfection (Y/N)	N																									
	Superinfection (details)	N																									
	Did the patient develop pneumonia (Y/N)	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	Did patient develop ARDS (Y/N)	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	COVID-19 complications	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	Enrolled in a COVID-19 clinical trial	N		N	N	N	N	N	N	N	N	N	N														
	Specific COID-19 clinical trial	N		N	N	N	N	N	N	N	N	N	N														
	Enrolled in a non-COVID-19 specific clinical trial	N		N	N	N	N	N	N	N	N	N	N														
	Stigma	N		N	N	N	N	N	N	N	N	N	N														
Supportive Care / Standard of Care								Y	Y	Y	Y	Y	Y	Y		Y		Y		Y							
	O2 Y/N	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	Date	N																									
	On Ventilation Y/N	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y												
	Date	N																									
	ECMO	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														
	Date	N																									
	Other cytokines (e.g., IFN-gamma, TNF-alpha, IL-1, GM-CSF)kines	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y														

	Pressors	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y																
	CRRT	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y																
COVID-19 Specific Medications								Y	Y	Y	Y	Y	Y	Y		Y		Y		Y									
	Medication	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y														
	Medication dosing regimen (Dosing, Start and End Date)	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y														
	Medication discontinuation reason	N		N	N	N	N	N	N	N	N	N	N		Y														
	Adverse Event	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y														
	Date of Adverse Event	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y														
Other Medication								Y	Y	Y	Y	Y	Y	Y		Y		Y		Y									
	Medication	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y														
	Medication dosing regimens (Dosage, Start and End Date)	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y																
	Medication discontinuation reason	N		N	N	N	N	N	N	N	N	N	N																
	Adverse Event	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y																
	Date of Adverse Event	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y																
Clinical Outcomes								Y	Y	Y	Y	Y	Y	Y		Y		Y		Y									
	Recovered from COVID-19	Y		DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED																
	Time to clinical improvement	DERIVED		DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED																
	Time to clinical progression	N		DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED																
	Improvement on use of supplemental oxygen or mechanical ventilation	DERIVED		DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED	DERIVED																
	Reduction of pulmonary lesions on chest CT examinat	N		N	N	N	N	N	N	N	N	N	N																
	Amelioration of routine laboratory criteria and pulmo	N		N	N	N	N	N	N	N	N	N	N																
	Increase of neutralizing antibody titers and disappear	N		N	y	y	N	N	N	N	N	N	N																
	Improvement on symptoms (e.g., fever, cough, shortr	Y		N	N	N	N	N	N	N	N	N	N																
	Potential candidate for donation of convalescent plasma (Y/N)	Derived (Possibly)																											
	Death	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y														
	Date of Death	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y														
Lab / Imaging								Y	Y	Y	Y	Y	Y	Y		Y		Y		Y									
	Imaging performed (X-Ray, CT, Ultrasound/Echo)	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y																
	Imaging Dates	N		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y																
	Imaging results	N		N	Y	Y	N	N	N	N	Y	Y	N																
	Lung Involvement Y/N	N		N	Y	Y	N	N	N	N	Y	Y	N																
	CBC with neutrophil and lymphocyte	N		N	Y	Y	N	N	N	N	Y	Y	N																
	ALC	N		N	Y	Y	N	N	N	N	Y	Y	N																
	ANC	N		N	Y	Y	N	N	N	N	Y	Y	N																
	AEC	N		N	Y	Y	N	N	N	N	Y	Y	N																
	Hgb	N		N	Y	Y	N	N	N	N	Y	Y	N																
	PLT	N		N	Y	Y	N	N	N	N	Y	Y	N																
	Creatine	N		N	Y	Y	N	N	N	N	Y	Y	N																
	T Bili	N		N	Y	Y	N	N	N	N	Y	Y	N																
	AST	N		N	Y	Y	N	N	N	N	Y	Y	N																
	ALT	N		N	Y	Y	N	N	N	N	Y	Y	N																
	PT	N		N	Y	Y	N	N	N	N	Y	Y	N																
	APTT	N		N	Y	Y	N	N	N	N	Y	Y	N																
	D dimer	N		N	Y	Y	N	N	N	N	Y	Y	N																
	LDH	N		N	Y	Y	N	N	N	N	Y	Y	N																
	TNI	N		N	Y	Y	N	N	N	N	Y	Y	N																
	BNP	N		N	Y	Y	N	N	N	N	Y	Y	N																
	CRP	N		N	Y	Y	N	N	N	N	Y	Y	N																
	IL6	N		N	Y	Y	N	N	N	N	Y	Y	N																
	GFR	N		N	Y	Y	N	N	N	N	Y	Y	N																
	Ferritin	N		N	Y	Y	N	N	N	N	Y	Y	N																
	Troponin/CKMB	N		N	Y	Y	N	N	N	N	Y	Y	N																
	NT-proBNP	N		N	Y	Y	N	N	N	N	Y	Y	N																
	Metabolic Panel	N		N	Y	Y	N	N	N	N	Y	Y	N																
	Culture tests/lab microbiology	N		N	Y	Y	N	N	N	N	Y	Y	N																

		LIMITATIONS OF DATA			**BENEFITS OF DATA**
		Claims	Limitations on inpatient dates b/c of roll-up. Outpatient dates are fine. Billing code NE clinical construct (some validation necessary).Medication - dispensing info available, not prescription (making adherence qs challenging).	**Claims**	Has a denominator (enrolled) and can better calculate p-t. More representative population number.
		EHR	Unless integrated, don't have a denominator. Sicker people will show up or women. May lend to biased estimates.	**EHR**	richer data source: lab results, physician notes, images, etc.