

Observational Health Data
Sciences and Informatics
(OHDSI): An International
Network for Open Science and
Data Analytics in Healthcare

Patrick Ryan, PhD
Janssen Research and Development
Columbia University Medical Center
9 May 2016



Odyssey (noun): \oh-d-si\

- 1. A long journey full of adventures
- 2. A series of experiences that give knowledge or understanding to someone



A journey to OHDSI



Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled, SECTION 1. SHORT TITLE. This Act may be cited as the "Food and Drug Administration Amendments Act of 2007".

to the safety of drugs, and for other purposes.

user-fee programs for prescription drugs and for medical devices, to enhance the postmarket authorities of the Food and Drug Administration with respect



What's the core problem?

We have lots of DATA we'd like to learn from...

....and very little EVIDENCE we can actually trust





Why large-scale analysis is needed in healthcare

All health outcomes of interest

	C E En	Ey Vi	Ex G (G Sa	Ti All Mic	d B	Fu G	H In	O Pr	o Vira	Ga li	nj Oc P	r S H	N Bo	Ir O	B In J	oi Mu O	Vi F	Ly N	им Р	Vi C	r In M	N N	IN Pe	S Vi	r Fo I	N P	r'C M	N S	Sin	j R Ur	Br M	PRE	Re Ut	in Pl	R R	SI U			Sk Vi Le	e H Th	CGI	N Sk	۷a ۱	٧e
n n n	9 9 1	P P 1	i de la composición dela composición de la composición de la composición dela composición dela composición dela composición de la composición de la composición dela com	Tr 1	1 1 11			79	44	, ,	- 1 -			1 /	n I	4 1-	ן לים די וויים	- 1 1		- 1	1.7	e de de	1 7	4 .	110	1 4 1		1 1 1	7 1			9 9	1 1	φ.	1 0		- 1 -	ar mand		T 1	1 1 11	P. I.		7 7	
e de la		100	10	11.				- 6					i.	- 3						100		100		d	1	41		a jan	6.33		1	σ.			9							all		1.	Ċ
- 900.0			1000			100		100						95			7 3					100		6	7	1					10	40.0			6			0.0							ď
- 3350			78.83		5.7.8	4.7	10.0		0.00	- 1			1	- 11	mis,	. 101	ж.,	d i		33		1	113	600	3.7	11.		1,000	1.1	160	1.47	200	1.1	6	Y	741 i	1	10,00	113	1.1		1	\$ 10	V. 11 -	d
10.4		45.4	0.60	100	d.								:	13	ing 3		40.0			1		1.						1.1			. 4	500	100		3.	94		1000		¥	i i	Ď.	į.		
ragina.			18	47.	in the second				- :				1	- 17			100		190	1		77				6.00					to all	200			10.04	0.0	1	ar Stadi	100	0.65		1	1	7.7	
100			" "						1				1	- ()											100	100		100	1.5	1	, I	5		1		200		100					- 1	100	d
66	4	100	0.49	М.	64. C			10						×.,	e Paris	an i					.1	100	150	la de	d i	0.0		46	10		440			100		40	4	3.60	10	es e	1	500		10	ò
46			Cáir	i lu		0.0			1				à i			. 150	.1				" "	Ti.		9	1	. :		10.		. 1	i L	£ 60			i.	å.		4,600		. 1		300	10	D.	ŀ
1000		130	444	- (1)	P\$13.	200			1.5	1			11	11) · · ·	Ψ.,	9.3				- 411	187	1	g troin		8 1		130	100	" (H	100			100	ge in	V 1	488	1	4		10.0	100	1.4	ï
	100	u tigaali	- Hiji	-41	99.91			- 1	200				!!	2.3	特拉拉	8,6	(an 3	11	150		- 1	144	1.4	\$ 600	H.	all is		144	di p		14.	170		1	2.4	46 17	100	拼料	(Spile)	131		131		1.0	
138	11.0	100	a ayaja	1.0	1299	1.		100		100			1	100	657	Ölti i	400			. 1		دودكل	100	ing:	r j	12		σÜ	d ķ.	eg s	ι÷.	00		3		£ 7.		daible	0.866			eb z	k .	2001	ij
pipp x		c - 2 fee	a salah Maraji	40.1	ation and		i.	. 6		. : :			200	1.0	Lagrania	4	Tail.	1	1400	T .		6.5	in i		400	ja d		1,000		6.0	1.4	ab Ti	100	á.		ăţji		rungen.	1	140	- 0	plant as			ij
light to	100			814	billion.	100		1.65			3.7	11	50	1.7	h () ;	Sar :	30.00	100	Met.	11	175	100	-19	lague.	. 1	100		300	200	. !	110	24 3		2.	4 1	質用で	9 - 10 -	relation.	digital and	數据	- 4	自制的	det.	35.45	ľ
100		17.			riginia.	10.00		0.00	1,000	er jir	1.1	10.0	100	1 (1)	F 1 14	4 122 j	100.0	a=1		10	11	dere.	i ii ji	foliab	ui n	ëk i:	6.0	140	ly in	1 1	100	91 1	1.64	307	jii r	grjego	parties of	i pid	iti tur	fini n	- 4	(PP) - 1	9 49	9 1	þ
- 1932	2.1	700	- 13		1000			- 1	1.11	1	1	1			300	Ψ.	1		900		10	$T_{\rm eff}$		in the second		11.11	3	11()	41		110	4 .	4	100	1.1	4.0		4 6 9 8	17.	100		721	1.1	1.	ì,
- 100		1.5	.13	4.1	100												1								1	- 4		100				. "	1	3.	100	40.0		100		1				:	
وروالز	v.		- Lips	40	gar in			100				100	la .		i Regional	e earlie	gial,	100		1				ad a		6.4					100	100	6.5	1	100	april 1		e 14 e		r a	0.00	անինու		m . l	ŀ
F19-4() (g/		100	117	711	payers.			0.00	- 1	100		100	i i	- 24	militari	er o	Section		100	- i	0.00	ir 1		et in	1	60.00	100	gis.	1.3	6.00	a be	ii i e	1	100	No.	rich in	1	g Sijn	e e	5.5		oi in	j	194	i
10/4 /		100	- 4d.	L^{n-1}	(mails)	4		1						;	μ, ,	et i	"L - L				1	100		Date:	4	1.0		1416	e. L.	1	100	Ы.	:	А,	ie i	<u>Q. 1</u>	l l	1449		19	9	72.0	\$ E	1	į
- Jabaj P	0.000	7.0	u Isa	- 2 i	Delta .			1.4	1				1	0.1	p Gov	Υ.,	9100		0.00	- (100	h þ	g ton	100	10.0		700	1000	100	5 B (Шι.	I -	3	19. 3	97.1	1000	25 (4.89)	ur e	#Ut	- 4	911.0	1.23	6.1.1	ĺ
1 1 1 1 1	- i	100	94	(i) 1	1. 1. 11 1			- 4					ľ.	- 1	¥	14	4 C	100		- 1	١.,		1	li line.	4.0			100	Je i i		: 1 :	T			ü a	78 B		a or i		fig.		SIN.	1	444	i
4444		0.00		200	1						100		1		Kira i	Die.	0.03	i i		100		1	- 4	6.6	3			1	5.5	0.0	5.44	(4) ·	0.0		h .	強力	1	2035		100	- 1	400	:		1
righte light	100	0.0	1.0	100	Phr. 5	6		100					i -		igar i s		$H \rightarrow$	300	100	1	100	16		110	100	0.1		000		100	11	ar -		0.00		0.0	0.00	en makeli	DE N	# i		cirt 5	4		
10.55			. II	70.									100		ora de la compansión de l La compansión de la compa		300	1	1			10												1	4.0	in a		1000		0.00		4.		de.	
100		127	, 14	$-N_{\parallel}$	50.59		4.0	15		1			100				9.5		100	1.0	100	1.	100	9.0	100			100			(4)	34) B	100	1		9.	1	11144		147		22		10	i
13 Sam	400	141	- 31		500			19			1	1				47				- }	100	1000		100	7	di i	0.00	177	100	1	4	700	1		4.4	9.4		翻的	100	-1		, ili	100	10.00	ı
13.47		100	- 11	7.	100								17		K + J +		4.1	20	0.0	120		400	100	· .		2.0	100	455			100	57.5	1.0	44		11		177		0.00	1	ing ti	100	7. 1	
1000			SH	6.5	fig.			14					1	13	Pitas	11.0	6 P.		111	. 1		17	e glis	14	1	3.1	1	1385	1 12	e .	900	dra t	9.8	884		100	100	grown	10	prof.	3		5.74	r et	l
- 33	1.0	100	32	200	j.							1.			Post of		11	100	2	1		310	. ""		. 1	Ÿ	1	34.5	3.1	0.5	, (, ()	¥ .	1000	4	200	Inc. i	1.1	1327	И.,	100			200		ì
Opt 1	100	3.0	2010 a	100	par hiji				4.5		1		ŧ.				.''	10	11.	100		24	. 4	11		M 1.		10.25	1.0	- 1	175	i, .			8.0	75.5		1 1 1 1 K 10 1 1 1 1 K		11.		ali, 9			
PHEN	100		1	2:1) - T			3.6	1	1	$t \in \mathcal{T}$		16	- aj	19 1 1 2	100	900	100	100	1.	6.7%	killin		: 111	1	ilia e		100	2.15	100	Egir	į.	100	100	17.4	31.11	4 %	ig organ	80 B	030	- 1	ere i	100	de al	
gibyly in	1.0	100	dane		ly a de	100	100	100	2.6	1		1	ii -	(1)		erija i	\$ 10		(50.00	- 1		plan.	- 66	and t		er in		100	i, ir	÷т	ЫH	hier.	0.101	1	100	yl e	sylli	ak jedaj		4 L		删点	ļΩ	: 41	
- 3372	1. A	1.2	. 34	J_{ij}	len der Sen der			100		100	2	100	li.	- #		G1 7	17.3		6.			Jäsi	2.10	aretin Eust	4	77.4		1000 pt	de i c	10	Щ.	<u> 1</u> 2 - 17		2	\$ 0	120	Γ	1020		5 / 40 / 5 / 5		看的	100	- 64	ij
1970	Note t	o Maria	1000	2.1	(* 28)			14.	1	1	9.0	179	jir -	91	rr e ist	iely a	g int		-4-55		910	983	11	P arter	H-	By it	•		98 []	100	1180	ij - 1		ij,		(報)	1.5	4.000	ii.	902 S		10.0	10	870	
-	100	400	بالتهد	- i	0.00	100		0.00		100	- 5	0.00	0.0		in .		200	100		1	17	Acres (1 9	600	1	in in	4	- 100	Je fil	Ъ, .		i	0.00	11		÷.	ė.	or other	100	100	1 1	100		, in	
148.1	100		12	90	100		1	100		100	0		1	16		0.00	466					ind.	1	Çeri.		15.6	6	910	6.11	11	e i e	9. J	1	37	7.00	eger a	.1	organic		100		4	4 6		
- mát (s	100	4.7	: Ŋ.		1932			10	20	$i_{i} \in \mathcal{F}_{i}$			1.	3.3		45.00	\$15		0.0	1	100	إيرارا				4, 6		Medi-	种目	12 .	a je	h. i	of a	9	200	(8), E	į	2,76%		400		制力	1.1	840	į
2188.9	100	i.	: 44	1014	1.00								10		n co	ene i	11. 1		11.0	: [100	中盟		100		sie in		1.00		T.		lica i	CH.	4.5	3.0	piĝi. Sulos s	1.	8000		Fee.		施力		100	
186		. 11	1,3	-84	100		$C_{i,j}$	10	0.0		1		1	500	er i	100	400		14	10		1:53	j. 16		4.1	快点		F 1917		32)	Л	# 1	10	T.	# 4	uri i	:	4,656		677		蝉山	10	gmi	j
1960	100	198	- 11	46	100			- 1	10		1		500	- 6	p i e	100	pris.			. 1		4/4	:	a de		gr ib		1 (1)	Wei	47.3	T.	pion i	1	ļ.,	10	de e	1	ส์วิสต์ที่	40	M	100	die.	400	5 (8	ij
- 0.3		100	1 349		()			1					100	- 13	100	6	97.7			1		1.1		King .	1	A_{i+1}		1986	2.0	400	100	r to i		3	1	d _i	10	1.365	100	10		400	1		
- 63		0.00	1.00	(0.3	0.50									10	p 1 - 6	630	4				10	90				. 1		200		10%	tigo:	$\sigma_{ij}(s)$		1	1.	igi ya	100	Long					1.75		
- 19 <u>1</u> (100	4.4	1, 1	i serie. Na serie							10					J., 1					4.0	•	11	1	200		18.7%	¥1	10		are en				Barrio Sa	1,	20%		62.	1	nie Legacia			
gar a		3.7	- 34		floor:					9.7		- 17	1	- 11			\mathcal{H}^{\prime}	300			16	10.0		7.7	4.	5.	1	rii) min	7 1		r [20		2		Ş.		1000		ÄV.		200	100	9.3	
0000		110	- in di		herr's	300	0.00	9.	620			100	je –	- 19		- 1	71 io	100	1	1		jih p	1	i by	1	file of		1975	3.77		r ita	ř		40		7 e	į .	d ijiye	$\theta_{ij}^{(i)}(t)$	190		760	įπ,	la din	
_ ###	A 1	244	580	5:	his 17	100		117	· 14-	100	1	- 11	1	i de		ill som	duya	11.	2700	100	100	44, 0	1	mb.	. E -	il i		unfre.		3.3	idi.	F-11		4.5	3	A w	i i	0.56	ette.	p	4	405	5.00	200	
- P. 7	100	71	e grije	300	989	100		10	0.00				$\Gamma = 1$	1.5	ابت	إضرا	Y = 0	100				1000		174	1	65.1		1940	4.00	0.3	11	$\chi(t) =$		1	200	92	100	1.457	1.00	P		$(\mathbf{H}^{(i)}, t)$	0.3	$\mathcal{C} = \mathcal{I}$	I
447.3			: 4		in de							1.0			دن ۽ ۾	14		1	4			4.00	, n			1		a Aust			11.	_				4		166		93		30%	200	. 11	



Introducing OHDSI

- The Observational Health Data Sciences and Informatics (OHDSI) program is a multistakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics
- OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University



OHDSI's mission

To improve health, by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.



What evidence does OHDSI seek to generate from observational data?

- Clinical characterization
 - Natural history: Who are the patients who have diabetes? Among those patients, who takes metformin?
 - Quality improvement: what proportion of patients with diabetes experience disease-related complications?
- Population-level estimation
 - Safety surveillance: Does metformin cause lactic acidosis?
 - Comparative effectiveness: Does metformin cause lactic acidosis more than glyburide?
- Patient-level prediction
 - Precision medicine: Given everything you know about me and my medical history, if I start taking metformin, what is the chance that I am going to have lactic acidosis in the next year?
 - Disease interception: Given everything you know about me, what is the chance I will develop diabetes?



What is OHDSI's strategy to generate evidence?

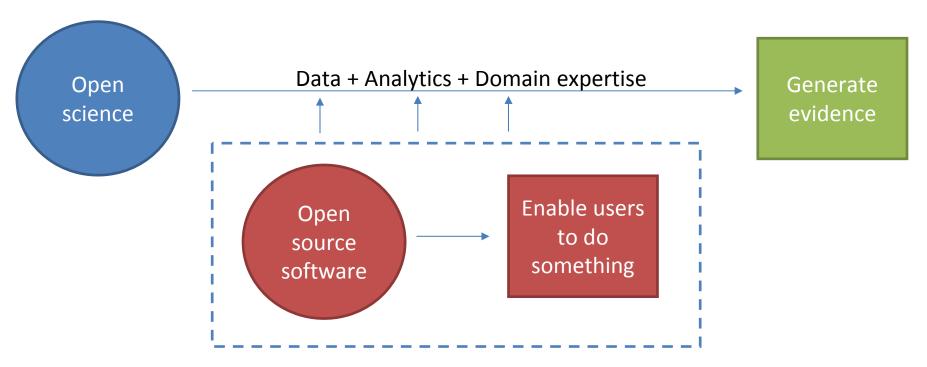
- Methodological research
 - Develop new approaches to observational data analysis
 - Evaluate the performance of new and existing methods
 - Establish empirically-based scientific best practices
- Open-source analytics development
 - Design tools for data transformation and standardization
 - Implement statistical methods for large-scale analytics
 - Build interactive visualization for evidence exploration

Clinical applications

- Identify clinically-relevant questions that require real-world evidence
- Execute research studies by applying scientific best practices through open-source tools across the OHDSI international data network
- Promote open-science strategies for transparent study design and evidence dissemination



OHDSI's approach to open science



- Open science is about sharing the journey to evidence generation
- Open-source software can be part of the journey, but it's not a final destination
- Open processes can enhance the journey through improved reproducibility of research and expanded adoption of scientific best practices



Standardizing workflows to enable reproducible research



Population-level estimation for comparative effectiveness research:

Is <intervention X> better than <intervention Y> in reducing the risk of <condition Z>?

Generate evidence

Database summary

Cohort definition

Cohort summary

Compare cohorts

Exposure-outcome Effect estimation & Compare databases calibration summary

Defined inputs:

- Target exposure
- Comparator group
- Outcome
- •Time-at-risk
- Model specification



Consistent outputs:

- analysis specifications for transparency and reproducibility (protocol + source code)
- only aggregate summary statistics (no patient-level data)
- model diagnostics to evaluate accuracy
- results as evidence to be disseminated
 - static for reporting (e.g. via publication)
 - interactive for exploration (e.g. via app)



OHDSI community in action



OHDSI Collaborators:

- •>140 researchers in academia, industry, government, health systems
- •>20 countries
- •Multi-disciplinary expertise: epidemiology, statistics, medical informatics, computer sciences machine learning, clinical sciences

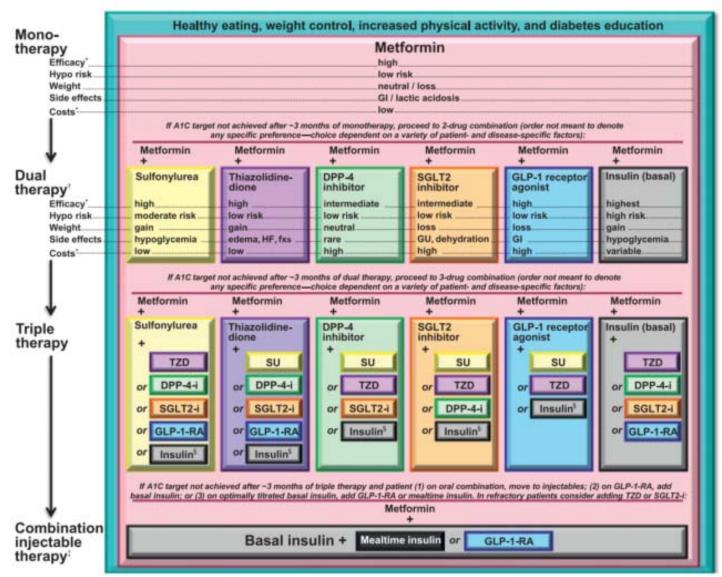
Ask clinical question

Develop standardized analytics

Community:



ADA T2DM Guidelines, 2015



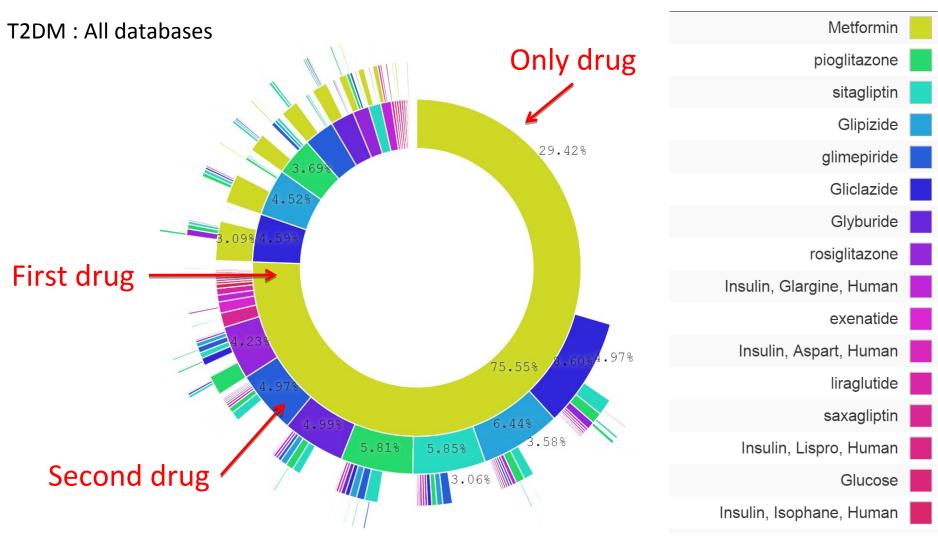


OHDSI participating data partners

Code	Name	Description	Size (M)
AUSOM	Ajou University School of Medicine	South Korea; inpatient hospital EHR	2
CCAE	MarketScan Commercial Claims and Encounters	US private-payer claims	119
CPRD	UK Clinical Practice Research Datalink	UK; EHR from general practice	11
CUMC	Columbia University Medical Center	US; inpatient EHR	4
GE	GE Centricity	US; outpatient EHR	33
INPC	Regenstrief Institute, Indiana Network for Patient Care	US; integrated health exchange	15
JMDC	Japan Medical Data Center	Japan; private-payer claims	3
MDCD	MarketScan Medicaid Multi-State	US; public-payer claims	17
MDCR	MarketScan Medicare Supplemental and Coordination of Benefits	US; private and public-payer claims	9
OPTUM	Optum ClinFormatics	US; private-payer claims	40
STRIDE	Stanford Translational Research Integrated Database Environment	US; inpatient EHR	2
НКИ	Hong Kong University	Hong Kong; EHR	1

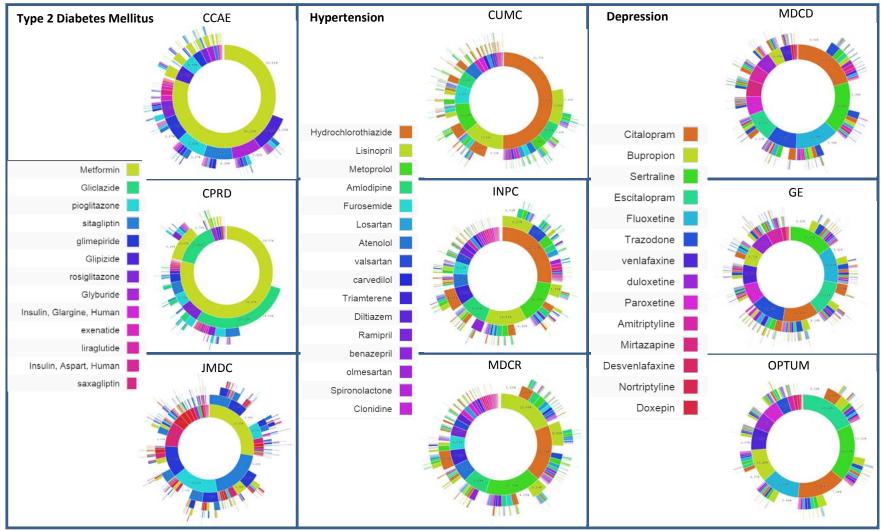


Treatment pathways for diabetes





Population-level heterogeneity









Indication

Synagis® (palivizumab) is a prescription medication that is used to help prevent a serious lung disease caused by respiratory syncytial virus (RSV) in children at high risk for severe lung disease from RSV.

Select Safety Information

Common side effects of Synagis include fever and rash. Other possible side effects include skin reactions around the area where the shot was given (like redness, swelling, warmth, or discomfort).

Please see complete Important Safety Information on pages 22-24 and accompanying full Prescribing Information, including Patient Information.

Doctor X: "This paper says there's side effects, but I've never seen them happen"



SYNAGIS- palivizumab injection, solution MedImmune, LLC

HIGHLIGHTS OF PRESCRIBING INFORMATION

These highlights do not include all the information needed to use SYNAGIS safely and effectively. See full prescribing information for SYNAGIS.

SYNAGIS® (palivizumab) injection, for intramuscular use

Initial U.S. Approval: 1998

------ INDICATIONS AND USAGE

Synagis is a respiratory syncytial virus (RSV) F protein inhibitor monoclonal antibody indicated for the prevention of serious lower respiratory tract disease caused by RSV in children at high risk of RSV disease.

- Safety and efficacy were established in children with bronchopulmonary dysplasia (BPD), infants with a history of
 premature birth (less than or equal to 35 weeks gestational age), and children with hemodynamically significant
 congenital heart disease (CHD).
- The safety and efficacy of Synagis have not been established for treatment of RSV disease. (1)

----- DOSAGE AND ADMINISTRATION ------

15 mg per kg of body weight, administered intramuscularly prior to commencement of the RSV season and remaining doses administered monthly throughout the RSV season. (2.1)

Children undergoing cardio-pulmonary bypass should receive an additional dose of Synagis as soon as possible after the cardio-pulmonary bypass procedure (even if sooner than a month from the previous dose). Thereafter, doses should be administered monthly as scheduled. (2.1, 12.3)

----- DOSAGE FORMS AND STRENGTHS

Single-dose liquid solution vials: 50 mg per 0.5 mL and 100 mg per 1 mL. (3)

------CONTRAINDICATIONS ------

Previous significant hypersensitivity reaction to Synagis. (4)

------ WARNINGS AND PRECAUTIONS ------

- Anaphylaxis and anaphylactic shock (including fatal cases), and other severe acute hypersensitivity reactions have been reported. Permanently discontinue Synagis and administer appropriate medications if such reactions occur. (5.1)
- As with any intramuscular injection, Synagis should be given with caution to children with thrombocytopenia or any coagulation disorder. (5.2)
- Palivizumab may interfere with immunological-based RSV diagnostic tests such as some antigen detection-based assays. (5.3, 12.4)

----- ADVERSE REACTIONS ------

Adverse reactions occurring greater than or equal to 10% and at least 1% more frequently than placebo are fever and rash. (6.1)



Pediatric patients across US observational databases

source	age group (at database entry)	persons	avg years of observation
CCAE	1. newborn 0 to 27d	3,360,896	2.23
MDCD	1. newborn 0 to 27d	1,862,651	1.74
Optum	1. newborn 0 to 27d	1,473,940	1.82
CCAE	2. infant and toddler 28d to 23mo	3,275,604	2.34
MDCD	2. infant and toddler 28d to 23mo	1,379,760	1.70
Optum	2. infant and toddler 28d to 23mo	963,770	1.97
CCAE	3. children 2 to 11 yo	14,904,293	2.46
MDCD	3. children 2 to 11 yo	4,037,836	1.77
Optum	3. children 2 to 11 yo	4,951,888	2.18
CCAE	4. adolescants 12 to 18 yo	12,218,224	2.41
MDCD	4. adolescants 12 to 18 yo	2,565,515	1.57
Optum	4. adolescants 12 to 18 yo	3,805,609	2.23



Exploring palivizumab exposure and hypersensitivity in observational data

data		persons		•	risk (events / 1000 persons);
source	age group (at time of exposure)	_	_		95%CI
CCAE	1. newborn 0 to 27d	1839	4829	0	0 (0 - 2.59)
MDCD	1. newborn 0 to 27d	381	760	0	0 (0 - 12.41)
OPTUM	1. newborn 0 to 27d	2610	5916	1	0.38 (0 - 2.45)
CCAE	2. infant and toddler 28d to 23mo	42843	106320	27	0.63 (0.43 - 0.92)
MDCD	2. infant and toddler 28d to 23mo	19910	41196	17	0.85 (0.53 - 1.38)
OPTUM	2. infant and toddler 28d to 23mo	22365	48632	11	0.49 (0.27 - 0.9)
CCAE	3. children 2 to 11 yo	544	1525	1	1.84 (0 - 11.68)
MDCD	3. children 2 to 11 yo	265	706	0	0 (0 - 17.78)
OPTUM	3. children 2 to 11 yo	204	574	0	0 (0 - 23.01)
CCAE	4. adolescants 12 to 18 yo	38	93	0	0 (0 - 115.33)
MDCD	4. adolescants 12 to 18 yo	33	28	0	0 (0 - 131.21)
OPTUM	4. adolescants 12 to 18 yo	11	44	0	0 (0 - 334.22)

Back of the envelope:

Assuming CCAE+MDCD+OPTUM represents 10% of US and exposures are evenly distributed across ~1000 NICUs, doctor would have seen ~50 newborns with exposure... even if the true event rate was 1%, there's >60% chance they'd never see



OHDSI: what does it mean to me?

Methodological research

Open-source analytics development

Clinical applications

Observational data management

Where is there reliable data about the health of children?

Clinical characterization

Who are the children who are exposed to palivizumab?

Population-level estimation

Does palivizumab cause anaphylaxis in newborns?

Patient-level prediction

Will my daughter be the one to develop anaphylaxis?



Concluding thoughts

- Observational databases can be a useful tool for generating evidence to important clinical questions in...
 - Clinical characterization
 - Population-level estimation
 - Patient-level prediction
- ...but ensuring that evidence is reliable requires developing scientific best practices, and transparent and reproducible processes to conduct analyses across the research enterprise
- An open science community allows all stakeholders to contribute to and benefit from a shared solution...anyone can get involved...that mean's YOU!
- Every patient, caregiver, parent and child deserves to know what is known (and what remains uncertain) from the real-world experience of others in order to inform their medical decisionmaking



Join the journey



Interested in OHDSI?

Questions or comments?

ryan@ohdsi.org