

Counterfactual Clinical Prediction Models could Help to Infer Individualized Treatment Effects in Randomized Controlled Trials

An Illustration with the International Stroke Trial

Presenter: Sureerat Suwatcharangkoon





RCT

Average treatment effect is commonly estimated

- assuming a homogeneous response to the treatment
- worsening outcomes in a minority of patients
- An individual treatment effect is not directly observed

The need for methods that can provide patient-level evidence about treatment effects

Evidence-based medicine targets the individual patient

Criteria to consider when applying the results of research studies to individual patients

	Healthcare	
Patient characteristics	characteristics	Outcome characteristics
 Biological factors (sex, comorbidities, race, age, severity of pathology) 	 Compliance of healthcare providers with treatment requirements 	Did the study measure an outcome of importance to the individual patient?
 Patient compliance with treatment requirements 	 Resources available for 'implementation (eg, availability of monitoring) Expertise of clinicians 	

Evid Based Med 2008;13(4):101-2.

The process of individualised EBM decision making

Sources of information						
Summary treatment effect from clinical studies	Results from subgroup analyses	N-of-1 RCT (limited to certain medical conditions and settings)				
Role of EBM:	Role of EBM:	Role of EBM:				
To assess the validity and applicability of study results, considering individual criteria:	To decide whether apparent differences are real	To help the clinic most bias-free str establish the trea	ian choose the udy design to tment effect in			
Patient: biological factors, socio- economic characteristics, compliance to recommendations		a chronic disorde	s suffering from r in which the are transient			
 Intervention/control: healthcare characteristics 						
Outcome: outcome characteristics						
Benefit: har	m ratios					
EBM tools: patient-specific number no needed to harm (NNH)	eeded to treat (NNT) or number					
Role of EBM:						
To effectively communicate individua the patient's baseline risk from variou guides, epidemiological studies, clinic	I risks and benefits by estimating s sources (clinical prediction cal experience)					
			,			
F	atient's values and preferences					
Role of EBM:						
To determine the extent to which the	patient wants to be involved in de	ecision-making.				
If shared decision-making is the goal,	EBM tools help to take patient pro	eferences and valu	es into account:			
 Decision aids 	Decision aids					
 Formal decision analysis 	Formal decision analysis					
Evidence-based individual treatment decision						

Evid Based Med 2008;13(5):130-1.



RCT

- Subgroup analyses
 - limited when many underlying characteristics are involved
 - prone to multiple testing -> risk of false-positive findings
- The Predictive Approaches to Treatment effect Heterogeneity (PATH)

THE LANCET

Articles

The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke

International Stroke Trial Collaborative Group*

	Patients randomised (n=19 435)						
Allocations							
Aspirin 300 + Heparin 12 500	Aspirin 300 + Heparin 5000	Aspirin 300 + No Heparin	No aspirin + Heparin 12 500	No aspirin + Heparin 5000	No aspirin + No Heparin		
2430	2432	4858	2426	2429	4860		

P: acute ischemic stroke with onset < 48 h previously

No (% randomised) with mortality follow-up

14 days

2430	2431	4858	2426	2429	4859
(100·0%)	(99·99%)	(100·0%)	(100·0%)	(100·0%)	(99·99%)
6 months					

2413	2410	4816	2411	2407	4828
(99.3%)	(99.1%)	(99-1%)	(99.4%)	(99.1%)	(99.3%)

O: death within 14 days and death or dependency at 6 months

Lancet. 1997 May 31;349(9065):1569-81.

1° outcome at 6 months in the aspirin group (62.2% vs. 63.5%, P = 0.07)

IST Subgroup analyses

A Subgroup		Colline	Oddia reti Hoparin Letter	Hoparin Hoparin Wares
Trial assists	-	20001215		
No trial aspiris	3061/4515	306404825	_	
Dates (hours): 8 - 8	285408	3059428		
4- 6	796/1155	78371150	_	
7-12	1306/2081	1284/2023		
10-24	1725/2000	1087/2090		-
25-48	1861/0147	2009/3345	_	•
Fernale	3096/4440	3181-4126	_	
Unio	2867/5291	2581/0116		-
Apr - 75	2858/5477	2003/5414		-
75+	2194/1104	2059/4230		
tanta at count	-			
Asimp at unset	1738/2798	1756/2540		
				100
Drowwy or anotherious	2008/2009	1965/2237		
	40001400			
Sinue mythin	45547903	44947548		
Abiai Ebritation	1221/1548	1266/1634		
Restation RPh < 180	-	10110-17108		
(mm Hg) > 180	11141753	1114/1740		
Toron and a data data and a standard	-	un number		
Type of stroke: Total anterior	101002001	1974/2003		
Facility assesses	100220878	DADERDAR	-	
Lacunar	111462306	11142010	_	
Contraction of the second s	-	manual distances		
No leg deficit	104/2199	0670219		
			100	52
Pre-randomisation GT	100000448	1000/0401		
No CT	274082	343/642		
and so and a second	-	-	30.	
No Infanti seen	1769(3313	1705/0208		
			1.51	
mar diagnosits: Inchaemic stroks	5080/0001	0303/08/08		
Shocks have write and	2022428	200.000		
Not stroks	182905	105/289		
Deles manifolio di descale				
No prior aspirit of anyty	4473/7184	4471/7229		
No prior heperin	5010/9252	5805/9258	22	
	-		0.11	16 C
Pantow-up: Central Develoed	5021/8091 342/1080	120/1087	-	-
Paul presse Remainder of trial	5175/8142	5760/01/52		-
Predicted risk: 0 - 40 %	5517616	589*1980		
10 - 30 %	11331047	114702-004	-	
70 - 100 %	9526/4050	34854079		
			13	Data and a start
Overall	6063/	*252/	-	 -61% io 30
	182.67.1	(82.9%)		(7p > 0-1; M5)
 W1-2 viz- 105, mmi 				
		0.6	0.75 5	0 1.28 1.8

			Odds ratio and CJ.		
Subgroup	Aspiris	Cantrol	better	worse	
Hedium-does that heperin	1495(2413	1520/0411			
Low-dose trial heparin	15062410	1836/2401	-		
Any trial heperin	30024825	3051/4010		-	
No trial heparts	2068-1816	3064/4528		-	
Delay (hours): 0 - 3	282814	298/422			
4- 6	78521157	7981148			
7 - 12	1274/2006	1316/2079			
12-24	1701/2798	17212730		-	
23 - 40	10400.0000	Tank and			
Formale	31504530	3127/4436		-	
une .	20000100	2998/5210			
Age: < 75	2009/5478	2868/5415			
25+	31814161	3862/4038			
Another at mount	4300/6842	43545805			
Asleep at onset	1700/0797	17912841	_	-	
Enumerator or unsecondaria	1004/02/24	18040242			
Abort	4016/7485	4141/7404		- C.	
firm the her	ALMONT	45137514	_		
Atrial fibriliation	1056/1011	1231(1542			
Sector Do Bill of the	-	4000.000	_		
(mm Hg) = 180	1005/1740	1125/1758		-	
Research and the Restation in the		-	1.1.1.1.1		
Restarior cistulation	5501115	Subject to the			
Partial animity	Charlenges.	24447000			
Lacurar	1112/2008	11162308			
Les defait	1010/2001	R100/TEN			
No log deficit	901(2258	970/2216			
Pre-reactomization (77	1203-5420	TOTAL AND			
Post-randomination CT	1864/2716	10050773			
No CT	324414	299/090			
Industry Lance	DOTE DATE.	Bartonina a	-		
No interest many	1712/148	21/201218		-	
Pinal diagnosis: Ischeamic stroke	5291/0615	5435/8655			
Hasmonhagic stroka	232586	239/291 +			
Stroke type anknown	361/583	340/480			
Not stroks	1040311	109/012			
Prior aspirin (2 days)	12001048	1325/2025			
No prior aspirin	4425/7202	4516/7185	-		
Pales bernada (14 bernad	122224				
No prior hepatin	\$7559055	5005/6253			
Friday and Annual					
Densived	511/1050	550/1057		-	
Pilot phase Remainder of trial	SUSTANT AN	201490			
			-		
Predicted risk: 0 - 40%	818/1904	62101972			
55 - 75 %	114771830	11001800			
70 - 108 %	3484/4055	0821/4074			
				Odda metasta	
Overall	e000	*139L	-12-	53% 60 24	
	(82.2%)	(63-5%)		(2p = 0.67)	
· 305 st man 105 bein			475 14	1.85	
			310 14	140 14	



Counterfactual prediction models



Z; the treatment status: $Z_i = 1$; 'treated', and $Z_i = 0$; 'control' $Y_{(1)i}$ and $Y_{(0)i}$; the potential outcomes (or 'counterfactuals') *i*; a particular individual

Consistency
Consistency
Solution

$$Y_i = Z_i Y_{(1)i} + (1 - Z_i) Y_{(0)i}$$

$$Y_i : \text{the observed outcome}$$

$$Z_i = 1; 'treated', \text{and } Z_i = 0; 'control'$$

$$Y_{(1)i} \text{ and } Y_{(0)i} : \text{the potential outcomes (or 'counterfactuals')}$$

$$ATE = E(Y_{(1)} - Y_{(0)}) = E(Y_{(1)}) - E(Y_{(0)})$$
3

Theory 1. Definitions

$Xi \in \mathcal{X}$

$$ITE = E(Y_{(1)} - Y_{(0)}|X) = E(Y_{(1)}|X) - E(Y_{(0)}|X)$$

Mahidol University Wichten of the Lond

 $Xi \in \mathcal{X}$ denote the *baseline covariates*

ITE; *individualised treatment effect*











Calibration and Discrimination Performance of Predicted ITE

- ITE (individualised treatment effects) are <u>never</u> observed-but estimated
- Consistence between prediction and observation X
 - , but rather between different estimates



Discrimination Performance of Predicted ITE

• Use "c-index"



Calibration Performance of Predicted ITE

- Agreement between the predicted ITE and the corresponding "observed" ITE.
- 1. Stratify the validation sample according to quintiles of predicted ITE
- Compute the marginal average difference in observed outcome across the treatment groups





- Split the initial sample to generate a derivation sample and a validation sample (2:1)
 - enough outcomes to avoid overfitting in derivation (> 50 events/variable)
 - precisely quantify model performance during validation (> 200 events)



- 2. Fit separate logistic regressions, using 23 predictors (no variable selection), to predict the outcome to each treatment arm of the derivation sample
 - effect modification



3. Predict the probability of the counterfactual outcomes

 $\widehat{P}(Y_{(1)} = 1|X)$ $\widehat{P}(Y_{(0)} = 1|X)$



 The discrimination: calculate the discrimination (c-statistic) in the derivation and validation samples

5. The calibration:

- calculating the calibration (slope and intercept) in the validation sample
- using local regression curves
- 95% CI were calculated by bootstrapping (500 iterations)



Calculated the ITE (difference between the 2 counterfactual prognoses returned by the models)



Results

THE LANCET

Articles

The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke

International Stroke Trial Collaborative Group*

Outcome	Heparin vs no hep	arin		Aspirin vs no aspirin		
	Heparin	No heparin	Events prevented per 1000 (SD)	Aspirin	No aspirin	Events prevented per 1000 (SD)
No randomised	9717	9718		9720	9715	
No with 6 month data	9641 (99-2%)	9644 (99-2%)		9639 (99-2%)	9646 (99-3%)	-
Fully recovered, independent	1655 (17-2%)	1641 (17-0%)	-2 (5)	1694 (17.6%)	1602 (16-6%)	-10 (5)
Not recovered, but independent	1923 (19-9%)	1941 (20.1%)	2 (6)	1945 (20-2%)	1919 (19.9%)	-3 (6)
Dependent	3898 (40-4%)	3986 (41-3%)	9 (7)	3927 (40-7%)	3957 (41-0%)	3 (7)
Dead from any cause	2165 (22.5%)	2076 (21.5%)	-9 (6)	2073 (21-5%)	2168(22.5%)	10 (6)
Dead or dependent	6063 (62-9%)	6062 (62-9%)	0(7)†	6000 (61.2%)	6125(63-5%)	13 (7)‡

†After adjustment for prognosis predicted at baseline, the benefit from heparin was 0 (SD 6), NS. ‡After adjustment for baseline stroke severity, the benefit from aspirin was 14 (SD 6), (2p=0.03). Negative numbers; same conventions as in table 2.

*2p<0.05, **2p<0.01, ***2p<0.001, ****2p<0.0001.

Table 3: Outcome at 6 months

Lancet. 1997 May 31;349(9065):1569-81.

Mahidol University With Land

Table 1. Baseline characteristics at randomization and outcomes

	Derivatio	on sample	Validation sample		
Variable	Aspirin 6,260 (49.7%)	Control 6, 338 (50.3%)	Aspirin 3, 460 (50.6%)	Control 3 377 (49.4%)	
Age (y)	74 (65–80)	74 (65–81)	73 (65-80)	73 (65-80)	
Delay (h)	18 (9-28)	19 (9-29)	20 (10-30)	20 (9-30)	
Systolic blood pressure (mmHg)	160 (140-180)	160 (140-180)	160 (140-180)	160 (140-180)	
Male sex	3,278 (52.4%)	3,358 (53.0%)	1,875 (54.2%)	1,896 (56.1%)	
Computerized tomography (CT)	4,175 (66.7%)	4,228 (66.7%)	2,316 (66.9%)	2,305 (68.3%)	
Infarct visible at CT	2,036 (32.5%)	2,146 (33.9%)	1,140 (32.9%)	1,093 (32.4%)	
Atrial fibrillation	1,092 (17.4%)	1,081 (17.1%)	530 (15.3%)	466 (13.8%)	
Missing value	278 (4.4%)	279 (4.4%)	215 (6.2%)	212 (6.3%)	
Aspirin within previous 3 d	1,317 (21.0%)	1,340 (21.1%)	644 (18.6%)	639 (18.9%)	
Missing value	278 (4.4%)	279 (4.4%)	215 (6.2%)	212 (6.3%)	
Not assessable	89 (1 4%)	84 (1.3%)	34 (1.0%)	40 (1 2%)	
No	1.679 (26.8%)	1.658 (26.2%)	888 (25.7%)	864 (25.6%)	
Ves	4 492 (71 8%)	4 596 (72 5%)	2 538 (73 3%)	2 473 (73 2%)	
Arm/hand deficit	4,452 (11.070)	4,000 (72.070)	2,000 (70.070)	2,475 (75.270)	
Not assessable	39 (0.6%)	43 (0.7%)	16 (0.5%)	25 (0.7%)	
No	872 (13.9%)	870 (13.7%)	476 (13.7%)	449 (13.3%)	
Yes	5,349 (85,5%)	5,425 (85,6%)	2,968 (85.8%)	2,903 (86.0%)	
Leg/foot deficit					
Not assessable	94 (1.5%)	77 (1.2%)	39 (1.1%)	45 (1.3%)	
No	1,469 (23.5%)	1,473 (23.2%)	803 (23.2%)	757 (22.4%)	
Yes	4,697 (75.0%)	4,788 (75.6%)	2,618 (75.7%)	2,575 (76.3%)	
Dysphasia					
Not assessable	190 (2.9%)	220 (3.5%)	91 (2.6%)	83 (2.5%)	
No	3,250 (53.2%)	3,348 (52.8%)	1,922 (55.6%)	1,822 (53.9%)	
Yes	2,820 (43.9%)	2,770 (43.7%)	1,447 (41.8%)	1,472 (43.6%)	
Hemianopia					
Not assessable	1,391 (22.2%)	1,375 (21.7%)	596 (17.2%)	583 (17.2%)	
No	3,896 (62.2%)	3,949 (62.3%)	2,301 (66.5%)	2,248 (66.6%)	
Maa	973 (15.6%)	1.014 (16.0%)	563 (16.3%)	546 (16.2%)	

	Derivatio	on sample	Validatio	n sample
Variable	Aspirin 6,260 (49.7%)	Control 6, 338 (50.3%)	Aspirin 3, 460 (50.6%)	Control 3 377 (49.4%)
Visuospatial disorder		-,,		,
Not assessable	1,181 (18.9%)	1,192 (18.8%)	534 (15.4%)	541 (16.0%)
No	4,037 (64.5%)	4,076 (64.3%)	2,379 (68.8%)	2,317 (68.6%)
Yes	1,042 (16.6%)	1,070 (16.9%)	547 (15.8%)	519 (15.4%)
Brainstem/cerebellar signs				
Not assessable	571 (9.1%)	584 (9.2%)	226 (6.5%)	211 (6.3%)
No	4,983 (79.6%)	5,049 (79.7%)	2,865 (82.8%)	2,807 (83.1%)
Yes	706 (11.3%)	705 (11.1%)	369 (10.7%)	359 (10.6%)
Other deficit				
Not assessable	419 (6.7%)	423 (6.7%)	214 (6.2%)	193 (5.7%)
No	5,455 (87.1%)	5,502 (86.8%)	3,026 (87.4%)	2,984 (88.4%)
Yes	386 (6.2%)	413 (6.5%)	220 (6.4%)	200 (5.9%)
Consciousness				
Fully alert	4,742 (75.7%)	4,803 (75.8%)	2,721 (78.7%)	2,655 (78.6%)
Drowsy	1,437 (23.0%)	1,447 (22.8%)	690 (19.9%)	680 (20.1%)
Unconscious	81 (1.3%)	88 (1.4%)	49 (1.4%)	42 (1.3%)
Stroke type				
PACS	2,538 (40.5%)	2,568 (40.5%)	1,382 (39.9%)	1,367 (40.5%)
TACS	1,546 (24.7%)	1,539 (24.3%)	781 (22.6%)	772 (22.8%)
LACS	1,428 (22.8%)	1,474 (23.3%)	898 (26.0%)	857 (25.4%)
POCS	733 (11.7%)	735 (11.6%)	388 (11.2%)	372 (11.0%)
Other	15 (0.2%)	22 (0.3%)	11 (0.3%)	9 (0.3%)
Region				
Europe	5,243 (83.8%)	5,309 (83.8%)	2,876 (86.0%)	2,804 (86.0%)
North America	96 (1.5%)	94 (1.5%)	28 (0.8%)	30 (0.9%)
South America	205 (3.3%)	213 (3.4%)	142 (4.3%)	133 (4.1%)
Africa	33 (0.5%)	32 (0.5%)	2 (0.1%)	2 (0.1%)
Middle East	107 (1.7%)	107 (1.7%)	93 (2.8%)	93 (2.8%)
North Asia	44 (0.7%)	45 (0.7%)	18 (0.5%)	17 (0.5%)
South Asia	112 (1.8%)	117 (1.8%)	81 (2.4%)	79 (2.4%)
Oceania	420 (6.7%)	421 (6.6%)	105 (3.1)	104 (3.2%)
Death/dependency at 6 mo	3,896 (62.2%)	4,027 (63.5%)	2,104 (60.8%)	2,098 (62.1%)
Missing value	43 (0.7%)	42 (0.07%)	38 (1.1%)	27 (0.8%)

Medians (interquartile ranges) and counts (proportions) are reported for continuous and binary or categorical variables, respectively.

Mahidol University Wicalom of the Land

Table 2. Models with and without aspirin predicting death or dependency at 6 M

	With aspirin	R.	Without aspir	in		With aspirin	Ú.	Without aspiri	in
Variable	Odds ratio (95% CI)	Р	Odds ratio (95% CI)	Р	Variable	Odds ratio (95% CI)	Р	Odds ratio (95% CI)	Р
Intercept	0.08 (0.03-0.20)		0.12 (0.05-0.32)		Visuospatial disorder (reference: No)		< 0.001		< 0.001
Age (v)	1.03 (1.02-1.04)	< 0.001	1.03 (1.02-1.04)	< 0.001	Not assessable	1.57 (1.22-2.03)		1.69 (1.31-2.18)	
(Age)'	1.03 (1.01-1.04)		1.03 (1.01-1.04)		Yes	1.59 (1.28-1.99)		1.79 (1.44–2.23)	
Delay (h)	1.00 (1.00-1.01)	0.061	1.00 (1.00-1.01)	0.001	Brainstem/cerebellar signs (reference:		0.019		0.414
Systolic blood pressure (mmHg)	1.00 (0.99-1.00)	0.001	1.00 (0.99-1.00)	0.003	Not assessable	1.20 (0.85-1.69)		1.10 (0.80-1.50)	
(Systolic blood pressure)'	1.00 (0.99-1.01)		1.00 (1.00-1.01)		Yes	2 87 (0 89-9 26)		1 98 (0 78-5 07)	
Male sex	0.76 (0.67-0.86)	< 0.001	0.79 (0.70-0.90)	< 0.001	Other deficit (reference: No)	2.07 (0.05 5.20)	0.001	1.50 (0.70 0.07)	0.233
Computerized tomography (CT)	0.55 (0.47-0.64)	< 0.001	0.55 (0.47-0.64)	< 0.001	Not assessable	1.57 (1.05-2.34)		0.75 (0.53-1.06)	
Infarct visible at CT	1.47 (1.26-1.73)	< 0.001	1.51 (1.30-1.76)	< 0.001	Yes	1.60 (1.21-2.13)		1.07 (0.82-1.40)	
Atrial fibrillation	1.19 (0.99-1.43)	0.046	1.28 (1.06-1.54)	0.005	Consciousness (reference: Fully alert)		< 0.001		< 0.001
Aspirin within previous 3 d	1.20 (1.03-1.40)	0.094	1.28 (1.10-1.48)	0.005	Drowsy	2.84 (2.31-3.49)		2.73 (2.22-3.36)	
Face deficit (reference: No)		< 0.001		< 0.001	Unconscious	8.98 (2.05-39.39)		11.57 (3.38-39.67)	
Not assessable	1.13 (0.55-2.32)		0.87 (0.43-1.78)		Stroke type (reference: PACS)		<0.001		< 0.001
Yes	1.24 (1.07-1.44)		1.18 (1.02-1.36)		TACS	1.14 (0.86-1.50)		1.08 (0.82-1.42)	
Arm/hand deficit (reference: No)		< 0.001		< 0.001	LACS	0.93 (0.76-1.14)		0.88 (0.72-1.08)	
Not assessable	0.57 (0.20-1.58)		1.03 (0.32-3.32)		POCS	0.32 (0.10-1.03)		0.45 (0.18-1.13)	
Yes	1.42 (1.13-1.79)		1.41 (1.12-1.76)		Other	0.81 (0.21-3.20)		0.92 (0.33-2.57)	
Leg/foot deficit (reference: No)		< 0.001		< 0.001	Region (reference: Europe)		< 0.001		< 0.001
Not assessable	1.93 (0.96-3.86)		2.10 (0.89-4.98)		North America	0.38 (0.23-0.65)		0.81 (0.49-1.32)	
Yes	2.21 (1.84-2.64)		1.97 (1.65-2.35)		South America	0.52 (0.37-0.72)		0.62 (0.45-0.85)	
Dysphasia (reference: No)		0.002		0.397	Africa	0.27 (0.11-0.67)		0.46 (0.20-1.08)	
Not assessable	2.36 (1.12-4.97)		1.18 (0.72-1.94)		South Asia	0.93 (0.59–1.47)		0.65 (0.42-1.02)	
Yes	1.14 (0.96-1.35)		1.20 (1.02-1.43)		Oceania	0.66 (0.51-0.84)		0.58 (0.46-0.74)	
Hemianopia (reference: No)		< 0.001		< 0.001	A restricted cubic spline with three knot	ts was used to describe the effects	s of age (knots at 56	5, 74 and 85 years) and systolic	blood pressure
Not assessable	1.53 (1.16-2.01)		1.41 (1.08-1.85)		(knots at 130, 160 and 200 mmHg).	ulation and any TACC total on			DOCC
Yes	1.70 (1.30-2.22)		1.66 (1.27-2.15)		rior circulation syndrome	curation syndrome; TAUS, total an	terior circulation syl	nurome; LACS, lacunar syndrome	;; FUUS, poste-







Calibration of the predicted ITE on an absolute risk difference scale



The red dashed line refers to the ideal calibration.









Discussion

Subgroup analysis	 one variable- at-a-time analyses 					
	 may lead to false-positive findings 					
"multivariable" subgroup	 rely on multivariable models using disease risk scores 					
analysis	unable to properly define thresholds					
	 increase the risk of false-positive findings 					
Counterfactual prediction	no need to define thresholds					
models	ITE is directly inferred from comparing the counterfactual risks of outcome					



Limitations

- 1. Model development
 - Models are transparently reported as stated for diagnostic and prognostic research
- 2. External validation
 - Refine inclusion criteria for secondary trials
- 3. Impact analysis
 - Require more meticulous practices than usual



Limitations

- Need for large RCTs
- Further studies are needed to explore the robustness of this approach against model misspecification

• ITE

- evidence inferred in (fine) groups of patients sharing similar characteristics
- uncertainty due to the gap between groups and individuals



What is new?

Key findings?

We illustrate how clinical prediction models used under a counterfactual framework could allow the inference of individualized treatment effects



What is new?

What this adds to what was known?

Counterfactual prediction models return, given a patient, the predicted risks of outcome under different scenarios (e.g. patient risk of outcome under treatment vs. patient risk of outcome under control)



What is new?

What is the implication and what should change now?

The comparison of counterfactual predicted risks may help refine clinical therapeutic decision-making at the patient level, as shown in this illustration.