

# Propensity score, stratifisering og justering

Tron Anders Moger  
Avdeling for helseledelse og  
helseøkonomi/HELSAM  
Universitetet i Oslo

# Innhold:

- Introduksjon
- Definisjon av noen begreper ved hjelp av et kort eksempel: Konfundering, interaksjoner og stratifisering
- Basale metoder for å sammenligne utfall mellom populasjoner
  - Direkte og indirekte standardisering
- Mer avanserte metoder for å sammenligne utfall mellom populasjoner
  - Propensity score med eksempel
  - Direkte regresjonsmetoder med eksempel
- Oppsummerende tanker

# Introduksjon

- Vanlig situasjon: Vil sammenligne utfall (dødelighet, liggetid, kostnader osv) for pasienter fra litt ulike populasjoner:
  - Ulike sykehus
  - Ulike regioner
  - Ulike land
- Vil f.eks. finne ut hvem som gjør det best!
- Kan bruke registerdata til dette! (flere land enn Norge har registerdata)
- Men vil tallene være sammenlignbare? Hva hvis sykehus/region/land A har en mye eldre befolkning enn sykehus/region/land B?

# Introduksjon forts.

- Eller, kanskje pasientene ved sykehus A er «sykere» enn ved sykehus B, av helt andre grunner enn det du er interessert i å studere?
- Eller generell snitt-levealder i ulike land kan være forskjellig
- Hvordan skal man sikre seg at man sammenligner like pasienter fra de ulike populasjonene (også kalt risikojustering, risk adjustment)?
- Eller: Kan man få større innsikt i hvilke faktorer som forklarer hvorfor ujusterte estimater fra ulike populasjoner er forskjellige?

# Definisjon av noen begreper

# Illustrasjon av konfundering

- To sykehus: A og B
- Måler kostnad per pasient for å behandle en sykdom
- Av en eller annen grunn er menn i snitt 100000 kroner dyrere å behandle enn kvinner
- På sykehus A er 80% av pasientene menn
- På sykehus B er 20% av pasientene menn
- En analyse av kostnad vs sykehus vil si at effekten av sykehus er svært signifikant (Fy! Sykehus A bruker for mye penger på pasientene)
- Hvordan finne ut av dette?

# Illustrasjon av interaksjon

- To sykehus: A og B
- Måler kostnad per pasient for å behandle en sykdom
- På sykehus A er menn i snitt 100000 kroner dyrere å behandle enn kvinner
- På sykehus B er kvinner i snitt 100000 kroner dyrere å behandle enn menn
- Dette betyr at det er en interaksjon mellom kjønn og sykehus på behandlingstkosten
- Hvordan finne ut av dette?

# Illustrasjon av stratifisering

- En enkel måte å finne ut hvordan ting henger sammen på, er å gjøre stratifiserte analyser (men forutsetter at du har informasjon om kjønn i studien!)
- Stratifisering: Del inn dataene i grupper basert på en variabel, estimer så det du er ute etter for hver gruppe (stratum) separat.
- Første eksempel: Ser på forskjell i kostnader mellom sykehus A og B for menn og kvinner separat
- Da står man igjen med en forskjell mellom sykehus for menn, og en forskjell mellom sykehus for kvinner.
- Finner da kanskje at det ikke er noen forskjell i kostnader mellom sykehusene innen hvert kjønn.
- Altså skyldes hele den ujusterte forskjellen at kjønnsfordelingen var skjev mellom sykehusene



# Illustrasjon av stratifisering forts.

- Andre eksempel: Stratifiserte analyse kan også brukes til å finne ut av interaksjoner
- Ser på forskjell i kostnader mellom kjønnene for sykehus A og sykehus B separat
- Da vil man opplagt se at kjønnsforskjellen avhenger veldig av hvilket sykehus man er på
- Kan også finne at sykehusforskjellen avhenger av hvilket kjønn man har, siden interaksjoner alltid går begge veier.

# Alternative løsninger:

- Annen løsning, konfunderingseksemplet: Kan f.eks. gjøre en regresjonsanalyse av kostnad vs sykehus og kjønn. Den vil vise at sykehus ikke betyr noe som helst, det er kjønn som er viktig!
- Annen løsning, interaksjonseksemplet: Gjøre en regresjonsanalyse, og ta med interaksjonsledd i modellen. Da tillater man at effekten av sykehus på kostnadene avhenger av om man er mann eller kvinne og at effekten av kjønn avhenger av om man er på sykehus A eller B.

Basale metoder for å gjøre et utfall  
mer sammenlignbart over ulike  
populasjoner

# Introduksjon

- La oss si man fortsatt er i en ganske enkel situasjon: Vil sammenligne dødelighetsrater mellom to ulike land.
- Det eneste man har mulighet til å justere for, er alder (aldersfordelingen i et land er ofte kjent, men det meste annet kan være ukjent hvis man ikke har gode registre)
- Skiller mellom to ulike metoder: Direkte og indirekte standardisering
- Formålet er uansett å se hva dødeligheten hadde vært hvis aldersfordelingen var lik i landene

# Introduksjon forts.

- I begge metoder må man forholde seg til følgende:
- En referansepopulasjon med en kjent aldersfordeling eller aldersspesifikk dødelighet
- Hvor mange dødsfall man observerer i hvert land  
i motsetning til
- Hvor mange dødsfall man forventer hvis hvert land hadde aldersfordelingen/dødelighetsraten til referansepopulasjonen
- Får altså ut et «hypotetisk» estimat: Hva hadde dødeligheten i hvert land vært hvis aldersfordelingen/dødelighetsraten var lik og den samme som i referansepopulasjonen?

# Direkte standardisering

- Kan brukes når man har alders-spesifikke dødelighetsrater for hvert utvalg i studien
- Trenger å finne observert dødelighetsrate for hver aldersgruppe for hvert studieutvalg
- Finner forventet antall døde i hver aldergruppe for hvert studieutvalg ved å multiplisere de observerte dødelighetsratene med antall personer i hver aldersgruppe i referansepopulasjonen
- Finner aldersjusterte (eller standardiserte) rater ved å legge sammen forventet antall døde over alle aldersgrupper og dele på total størrelse av referansebefolkningen

# Direkte standardisering, eks:

Age group (years)	Population A			Population B		
	Population	Deaths	Age-specific death rate per 1,000	Population	Deaths	Age-specific death rate per 1,000
0–24	18,000	35	1.94	13,000	30	2.31
25–49	11,000	60	5.45	7,000	50	7.14
50–74	9,000	370	41.11	11,000	400	36.36
75 and above	3,000	250	83.33	4,000	380	95.00
Total	41,000	715		35,000	860	
Crude rate per 1000			17.44			24.57

Fra Naing 2000

# Direkte standardisering, eks:

	Population A			Population B	
Age group (years)	Reference population	Age-specific death rate per 1,000	Expected deaths	Age-specific death rate per 1,000	Expected deaths
0–24	11,000	1.94	21.34	2.31	25.41
25–49	17,000	5.45	92.65	7.14	121.38
50–74	20,000	41.11	822.20	36.36	727.20
75 and above	3,000	83.33	249.99	95.00	285.00
Total	51,000		1186.18		1158.99

Gir aldersjustert dødelighet lik:

Population A:  $1186.18/51,000=23.3$  per 1,000

Population B:  $1158.99/51,000=22.7$  per 1,000



# Indirekte standardisering

- Kan brukes når man ikke har alders-spesifikke dødelighetsrater for hvert utvalg i studien
- Trenger en referansepopulasjon med kjente aldersspesifikke dødelighetsrater
- Finner forventet antall døde i hver aldersgruppe for hvert studieutvalg ved å multiplisere dødelighetsratene fra referansen med antall personer i hver aldersgruppe i studieutvalgene
- Legg sammen forventet antall døde over alle aldersgrupper
- Beregn en ratio fra det observerte antall døde i hver studiepopulasjon delt på forventet antall døde (Kalles  $SMR = \text{Standardized Mortality Ratio}$ )
- Finn alderjustert (eller standardisert) dødelighetsrate for hver studiepopulasjon ved å multiplisere SMR med den ujusterte dødelighetsraten fra referansepopulasjonen

# Indirekte standardisering, eks:

Anta at: Observed deaths in population A=120  
 Observed deaths in population B= 30

Population A				Population B		
Age group	Population	Age-specific mortality rate per1000	Expected deaths	Population	Age-specific mortality rate per1000	Expected deaths
0–24	2000	4.0	8.0	1000	4.0	4.0
25–49	2500	7.0	17.5	1500	7.0	10.5
50–74	3500	10.0	35.0	2500	10.0	25.0
75+	4500	30.0	135.0	1000	30.0	30.0
Total	12500		195.5	6000		69.5

Aldersspesifikke dødelighetsrater fra referansepopulasjonen

Finner at SMR blir  
 Populasjon A:  $120/195.5=0.61$   
 Populasjon B:  $30/69.5=0.43$

Fra Naing 2000

# Hva velger man som referansepopulasjon?

- Mange muligheter finnes, kan bruke en samlet database fra alle sykehus/land i studien eller generelle nasjonale eller internasjonale tall.
- F.eks. har WHO en «world standard» populasjon, som kan finnes her:  
<http://seer.cancer.gov/stdpopulations/world.who.html>
- Finnes også skandinaviske standarder

# Utvidelser av standardiseringen

- Metodene for direkte og indirekte standardisering som vist, er ok hvis man bare vil justere for alder og f.eks. kjønn
- Registre har ofte mer informasjon, kan man bruke dette også?
- Snakker da om risikojustering, vil bruke mange variabler til å lage utvalg som er så like som mulig
- Øker sannsynligheten for at forskjeller man sitter igjen med mellom sykehus/land etc. skyldes behandling eller noe annet man er interessert i

# Matching

- Naiv, men intuitiv tanke:
- Hvorfor ikke bare lage utvalg som er like basert på bakgrunnsvariabler ved at man for hver person i et utvalg finner personer i de andre utvalgene som har de samme verdiene av bakgrunnsvariablene?
- Fungerer hvis det er få variabler man har mulighet til å matche på
- Hvis mange variabler: får for mange undergrupper og dermed for få observasjoner i de ulike undergruppene, selv med store registerdata!

# Propensity score

- Utviklet av Rubin og Rosenbaum i 1983.
- En variant av tradisjonell matching
- I stedet for å matche på hver eneste bakgrunnsvariabel enkeltvis, samler man alle bakgrunnsvariablene i et tall, kalt en indeks.
- Matcher observasjoner basert på dette ene tallet.
- Slipper å lage mange små undergrupper fordi man kan forholde seg til kun et tall som sier noe om hvor lik observasjonen er en annen observasjon.

# Propensity score – klassisk motivasjon

- Ble i utgangspunktet tenkt brukt for å lage noe som ca. kunne tilsvare randomiserte kliniske forsøk fra observasjonsstudier
- De som får behandling skal da være like som de som ikke får behandling på alle bakgrunnsvariabler man vil ta hensyn til
- Da sikrer man at man finner den kausale effekten av behandling på utfall, selv om det er en observasjonsstudie
- Men i vår situasjon er det ikke behandling som er poenget, men at personer fra ulike sykehus/land etc. skal være like

# Propensity score - Antagelser

- Skal helst ha inkludert alle bakgrunnsvariablene som påvirker både utfallet og seleksjonen inn i utvalgene (sykehus/region/land). Kalt "ignorability."
- Må eksistere individer som har tilnærmet lik verdi dvs. det må være noe overlapp av personer som er like der noen er i utvalg A og andre er i utvalg B. Til sammen kalles dette "strong ignorability."



# Propensity score - Praksis

- 1. Velg de variablene du ønsker at individene skal være like på og bruk disse til å lage en indeks for likhet.
- 2. Sammenlign observasjoner som har tilnærmet lik indeks for likhet
- Hvordan lage indeksen? Kjør logistisk regresjon hvis det kun er to populasjoner som skal sammenlignes, multinomisk logistisk regresjon hvis det er flere
- Utfallet blir da en kode for hver populasjon, og forklaringsvariablene er bakgrunnsvariablene man vil justere for
- Kan da beregne sannsynlighet fra modellen for at hver person tilhører hver populasjon

# Propensity score – Praksis forts.

- Etter at sannsynlighetene er estimert, kan man starte matchingen
- Mange muligheter, hvor en-til-en matching er enklest
- For hver person i et utvalg, finner man personer i de andre utvalgene som har likest verdi av indeksen
- Bør fungere for registerdata, da utvalgene er store
- Kan gjøres med tilbakelegging, vil ofte redusere skjeveheten sammenlignet med å matche uten tilbakelegging
- Gjør hovedanalysen på det matchede datasettet med metoder for matchede data
- Fins programmer for propensity score analyser, f.eks. `psmatch2` i Stata

# Propensity score - Svakheter

- Hvis datasettet er for lite, risikerer man at de matchede personene i praksis ikke er så like på bakgrunnsvariablene likevel. Hvordan bestemme hva som er likt nok?
- Gjøres matchingen med tilbakelegging, risikerer man at de samme personene brukes veldig ofte
- Hvordan estimere/fortolke usikkerheten i estimatene fra den endelige analysen?
- Teoretisk antar man at indeksene er faste størrelser men i praksis estimeres de
- Dette bidrar til usikkerhet, som er vanskelig å «ta med» i den endelige analysen

# Propensity score - praktisk eksempel

- Schreyögg et al. (2011): Vil sammenligne kostnader og within hospital dødelighet for hjerteinfarktpasienter i Tyskland og USA
- Vil også estimere effekten av ulike behandlinger
- Igjen, grunn til å tro at populasjonene er forskjellige mhp bakgrunnsvariabler
- Kjører da en propensity score matching vha logistisk regresjon, utfallet er om man er pasient i USA eller Tyskland, og bakgrunnsvariablene ser dere på neste slide
- Bruker en-til-en matching med tilbakelegging, men noen restriksjoner for å unngå å ta med matchede par som er en dårlig match

# Propensity score - praktisk eksempel forts.

VHA=Veterans Health Administration i USA

Characteristics	VHA	Germany
No. of cases in the sample	6598	585
No. of hospitals in the sample	130	18

Ser at datasettene er ikke veldig store i dette eksemplet. Propensity score metoden vil her bety at man ikke bruker mer enn max 585 av de amerikanske dataene i den videre analysen, siden man matchet 1:1

Parameter	Estimate	S.E.	Chi-Square	Pr > ChiSq
Intercept	-19.1	2.25	72.45	<0.0001
Age	0.55	0.07	64.05	<0.0001
Angina pectoris	0.52	0.32	2.60	0.1071
Cardiac dysrhythmias	-0.66	0.40	2.64	0.1042
Cancer	-5.58	2.35	5.62	0.0178
Cerebrovascular disease	-0.88	0.61	2.10	0.1471
Congestive heart failure	1.01	0.17	34.44	<0.0001
Chronic ischemic heart disease	1.64	0.79	4.25	0.0393
Acute renal failure	2.07	3.35	0.38	0.5369
Chronic renal failure	-8.22	3.44	5.70	0.0170
Diabetes	-0.96	0.15	40.34	<0.0001
Hypertensive heart disease	0.11	0.66	0.03	0.8683
Diabetes complications	-5.49	2.69	4.17	0.0412
Peripheral vascular disease	-7.63	3.87	3.89	0.0485
Pulmonary disease	-2.77	1.39	3.98	0.0460
Severe liver disease	-14.3	1459	0.00	0.9922
Shock	0.27	0.51	0.28	0.5962
<i>Interactions<sup>a</sup></i>				
Age*age	0.00	0.00	70.52	<0.0001
Age*acute renal failure	-0.08	0.05	2.43	0.1191
Age*chronic ischemic heart disease	-0.02	0.01	2.52	0.1127
Age*chronic renal failure	0.09	0.05	3.11	0.0778
Angina pectoris*acute renal failure	4.06	2.45	2.74	0.0977
Angina pectoris*cerebrovascular disease	3.34	1.39	5.75	0.0165
Angina pectoris*pulmonary disease	1.20	0.75	2.55	0.1105
Cancer*cardiac dysrhythmias	3.04	1.72	3.15	0.0759
Cancer*diabetes complications	5.68	2.30	6.09	0.0136
Cancer*peripheral vascular disease	7.33	2.24	10.74	0.0010
Cancer*pulmonary disease	4.05	1.61	6.32	0.0119
Diabetes*acute renal failure	2.74	1.38	3.97	0.0465
Diabetes*cardiac dysrhythmias	0.88	0.39	5.16	0.0231
Diabetes*chronic renal failure	-1.96	1.46	1.81	0.1786
Diabetes*congestive heart failure	0.60	0.32	3.54	0.0599
Diabetes*diabetes complications	2.45	1.52	2.60	0.1068
Diabetes*pulmonary disease	0.98	0.42	5.32	0.0210
Cardiac dysrhythmias*chronic ischemic heart disease	1.23	0.42	8.77	0.0031
Cardiac dysrhythmias*diabetes complications	2.06	1.05	3.84	0.0499
Chronic ischemic heart disease*diabetes complications	1.32	0.88	2.26	0.1325
Congestive heart failure*chronic renal failure	-1.87	0.95	3.85	0.0498
Congestive heart failure*diabetes complications	1.80	0.83	4.66	0.0309
Pulmonary disease*chronic renal failure	-3.38	2.22	2.33	0.1273
Pulmonary disease*diabetes complications	3.29	1.04	10.08	0.0015
Peripheral vascular disease*cerebrovascular disease	3.71	1.73	4.57	0.0325
Peripheral vascular disease*pulmonary disease	1.81	1.06	2.89	0.0892
Acute renal failure*chronic renal failure	7.38	1.47	25.26	<0.0001
Acute renal failure*diabetes complications	-1.60	1.21	1.73	0.1878
Renal failure chronic*diabetes complications	1.90	1.11	2.91	0.0879

<sup>a</sup>Interactions included into the model with  $p > 0.20$  are not reported: Age\*Diabetes complications, Age\*Peripheral vascular, Age\*Pulmonary disease, Angina pectoris\*Cancer, Angina pectoris\*Cardiac dysrhythmias disease, Angina pectoris\*Chronic renal failure, Cancer\*Chronic ischemic heart disease, Cancer\*Congestive heart failure, Diabetes\*Cerebrovascular disease, Diabetes\*Peripheral vascular disease, Cardiac dysrhythmias\*Cerebrovascular disease, Cardiac dysrhythmias\*Congestive heart failure, Cardiac dysrhythmias\*Peripheral vascular disease, Chronic ischemic heart disease\*Peripheral vascular disease, Congestive heart failure\*Acute renal failure, Congestive heart failure\*Cerebrovascular disease, Congestive heart failure\*Peripheral vascular disease, Congestive heart failure\*Pulmonary disease, Pulmonary disease\*Cerebrovascular disease, Peripheral vascular disease\*Diabetes Complications.

Table IV. Balance in measured baseline variables before and after matching

	Unmatched sample			Matched on propensity scores for co-morbidities (upper bound of predicted probabilities)			Matched on propensity scores for co-morbidities (predicted mean probabilities)			Matched of propensity scores for co-morbidities (lower bound of predicted probabilities)		
	VHA	Germany	$d_i$	VHA	Germany	$d_i$	VHA	Germany	$d_i$	VHA	Germany	$d_i$
No. of cases in the sample	6598	585		577	577		574	574		578	578	
<i>Patient characteristics</i>												
Age	65.71	62.31	35.18%	62.32	62.22	1.20%	62.77	62.18	7.09%	62.79	62.26	6.41%
Acute renal failure	0.03	0.04	-0.80%	0.03	0.04	-1.10%	0.03	0.04	-1.20%	0.02	0.04	-2.20%
Cancer	0.04	0.01	2.95%	0.03	0.01	2.50%	0.01	0.00	0.50%	0.01	0.01	0.00%
Chronic renal failure	0.08	0.05	3.36%	0.03	0.05	-1.80%	0.04	0.05	-1.00%	0.02	0.05	-2.80%
Congestive heart failure	0.09	0.18	-8.77%	0.16	0.18	-1.80%	0.16	0.18	-1.30%	0.15	0.18	-2.60%
Diabetes	0.25	0.16	8.47%	0.14	0.16	-2.60%	0.14	0.16	-1.80%	0.15	0.16	-1.50%
Diabetes complications	0.04	0.02	1.35%	0.02	0.02	0.30%	0.02	0.02	0.10%	0.00	0.02	-1.70%
Peripheral vascular disease	0.04	0.02	2.11%	0.02	0.02	0.50%	0.02	0.02	0.00%	0.01	0.02	-0.30%
Pulmonary disease	0.15	0.07	8.05%	0.07	0.06	0.70%	0.06	0.06	0.10%	0.07	0.06	0.70%
Severe liver disease	0.00	0.00	0.09%	0.00	0.00	0.18%	0.00	0.00	0.00%	0.00	0.00	0.00%
Angina pectoris	0.01	0.03	-1.81%	0.02	0.03	-0.60%	0.02	0.03	-0.40%	0.02	0.03	-0.30%
Cerebrovascular disease	0.02	0.01	0.91%	0.02	0.01	0.50%	0.02	0.01	0.30%	0.02	0.01	1.00%
Cardiac dysrhythmias	0.09	0.13	-4.01%	0.12	0.13	-1.50%	0.11	0.13	-2.20%	0.11	0.13	-2.10%
Chronic ischemic heart disease	0.66	0.80	-14.19%	0.78	0.80	-1.90%	0.79	0.80	-1.20%	0.79	0.80	-0.80%
Hypertensive heart disease	0.00	0.01	-0.13%	0.01	0.01	0.10%	0.00	0.00	0.00%	0.00	0.01	-0.20%
Shock	0.01	0.01	-0.19%	0.01	0.01	-0.30%	0.00	0.01	-0.50%	0.01	0.01	-0.30%

# Propensity score - praktisk eksempel

## forts.

- Ok, har laget matchede utvalg, hva nå?
- Lager en regresjonsmodell som inkluderer og estimerer effekten av ulike behandlinger og sykehusfaktorer på kostnader og dødelighet.
- Har også med komorbiditeter i denne modellen
- Problem: Kan ofte ha clustering på sykehus-nivå – Pasienter på samme sykehus er likere hverandre enn pasienter fra forskjellige sykehus, selv etter at man har justert for bakgrunnsvariablene
- Løsning: Bruker multilevel-modeller
- I tillegg må modellene ta hensyn til at det er matchede data, gjøres vha generalized estimating equations (GEE)



Table V. Joint regression estimates for both samples (presented as marginal effects)

	Unmatched samples	Matched on propensity scores for co-morbidities	Sample size VHA/GER
<i>1.<sup>a</sup>Note: Displayed marginal effects for costs are on the log scale; unmatched VHA patients are dropped in PSM models. S.E. in parenthesis. PSM, propensity score matching. ***p&lt;0.001.</i>			
<i>Coefficients for GER vs VHA with dependent Log (costs)</i>			
Based on unmatched samples	-1.402 (0.143)***		6598/585
Based on predicted mean values as PSM		-2.086 (0.300)***	574/574
Based on predicted lower bound estimate as PSM		-2.023 (0.321)***	578/578
Based on predicted upper bound estimate as PSM		-1.938 (0.305)***	577/577
<i>Coefficient for GER vs VHA with dependent hospital mortality</i>			
Based on unmatched samples	-0.057 (0.067)***		6598/585
Based on predicted mean values as PSM		-0.012 (0.018)	574/574
Based on predicted lower bound estimate as PSM		-0.032 (0.020)	578/578
Based on predicted upper bound estimate as PSM		0.001 (0.001)	577/577

Ser at Tyskland var litt billigere enn USA i umatchedede rådata, men enda billigere i matchede data

Ser også at dødeligheten var lavere i Tyskland enn i USA i umatchedede rådata, men ikke lavere i matchede data

# Propensity score - praktisk eksempel:

## Konklusjon

- Uten matching kan man bruke hele datasettet, får mindre usikkerhet men potensielt store skjevheter i estimatene siden man sammenligner populasjoner som er forskjellige
- Artikkelen finner store forskjeller i estimatene for forskjell mellom Tyskland og USA fra matchede til umatchede analyser, både for dødelighet og kostnader
- Konkluderer derfor med at matching kan ha betydning, selv om det selvfølgelig er vanskelig å vite «fasitsvaret»

# Alternativ til propensity score

- Kan legge bakgrunnsvariablene inn i hovedanalysen direkte, ved å bruke en regresjonsmetode
- Eksempler på vanlige regresjonsmetoder for ulike typer utfall:
  - Korttidsdød (opp til et år): Logistisk regresjon
  - Langtidsdød: Levetidsanalyse
  - Liggetid: Negativ binomisk regresjon (som Poisson regresjon, men tar hensyn til at liggetid ofte har større variasjon enn man får til med en Poisson modell)
  - Kostnader: Gamma regresjon, ofte med log-link. Gamma fordi kostnader er kontinuerlige, og veldig skjevfordelte

# Svakheter, regresjonsmetode

- For matche-metoder vil det ofte være opplagt hvis man har dårlig overlapp mellom datasettene på noen variabler (kan få beskjed om dette i Propensity score funksjoner f.eks.)
- For regresjonsmetoder må man være mer obs på det, må unngå å ekstrapolere til områder man har få observasjoner
- Gjelder antagelser om funksjonsform til regresjonsmodellen overalt, eller bare der man har mest data?

# Regresjonsmetode – praktisk eksempel

- Moger og Peltola (2014), eksempel fra EuroHOPE:
- Formålet er å sammenligne utfall som dødelighet, liggetid, kostnader, behandling osv mellom Norge, Sverige, Finland, Nederland, Ungarn, Skottland og deler av Italia for fem tilstander
- Bruker både ujusterte tall og diverse risikojusterte tall
- Viser her eksempel for hjerteinfarkt
- Utfall 90-dagers dødelighet og liggetid, første episode

## Descriptive statistics on background variables and outcomes used for AMI.

Variable	Data											
	Finland		Hungary		Italy		Norway		Sweden		Pooled	
	N (Avg.)	% (SD)	N (Avg.)	% (SD)	N (Avg.)	% (SD)	N (Avg.)	% (SD)	N (Avg.)	% (SD)	N (Avg.)	% (SD)
N	9102	100.0 %	14235	100.0 %	1563	100.0 %	10612	100.0 %	23768	100.0 %	59135	100.0 %
Male	5404	59.4 %	7907	55.5 %	1002	64.1 %	6562	61.8 %	14251	60.0 %	35126	59.4 %
Age	72.6	12.7	66.9	13.1	70.2	13.2	71.8	14.1	73.4	12.8	71.4	13.4
LoS previous year	5.1	10.9	4.8	10.5	1.0	5.0	5.6	13.0	4.6	12.7	4.8	11.8
<i>Age, classified</i>												
18 to 49	449	4.9 %	1429	10.0 %	122	7.8 %	772	7.3 %	1011	4.3 %	3783	6.4 %
50 to 54	473	5.2 %	1378	9.7 %	85	5.4 %	649	6.1 %	1002	4.2 %	3587	6.1 %
55 to 59	716	7.9 %	1437	10.1 %	157	10.0 %	856	8.1 %	1696	7.1 %	4862	8.2 %
60 to 64	808	8.9 %	1601	11.2 %	154	9.9 %	1134	10.7 %	2442	10.3 %	6139	10.4 %
65 to 69	883	9.7 %	1729	12.1 %	171	10.9 %	1033	9.7 %	2386	10.0 %	6202	10.5 %
70 to 74	1161	12.8 %	1853	13.0 %	208	13.3 %	1066	10.0 %	2729	11.5 %	7017	11.9 %
75 to 79	1429	15.7 %	1985	13.9 %	234	15.0 %	1271	12.0 %	3279	13.8 %	8198	13.9 %
80 to 84	1526	16.8 %	1618	11.4 %	218	13.9 %	1456	13.7 %	3936	16.6 %	8755	14.8 %
85 to 89	1098	12.1 %	794	5.6 %	136	8.7 %	1476	13.9 %	3416	14.4 %	6920	11.7 %
90 or older	559	6.1 %	266	1.9 %	78	5.0 %	899	8.5 %	1871	7.9 %	3672	6.2 %
<i>Measures to be risk-adjusted</i>												
90-day mortality	1753	19.3 %	3048	21.4 %	193	12.3 %	1230	11.6 %	3541	14.9 %	9765	16.5 %
LoS of first hospital episode	12.1	14.0	11.9	9.8	11.3	9.5	7.9	5.6	8.5	7.0	9.8	9.2

## Descriptive statistics on comorbidities used in risk adjustment of outcomes for AMI.

Variable	Data											
	Finland		Hungary		Italy		Norway		Sweden		Pooled	
	N	%	N	%	N	%	N	%	N	%	N	%
	<i>Comorbidities based on diagnoses and medication during 365 days prior to AMI</i>											
Hypertension	6467	71.1 %	12424	87.3 %	1346	86.1 %	6676	62.9 %	16293	68.6 %	42218	71.4 %
Coronary artery disease	1146	12.6 %	1971	13.8 %	119	7.6 %	1479	13.9 %	2711	11.4 %	7426	12.6 %
Atrial fibrillation	448	4.9 %	498	3.5 %	27	1.7 %	701	6.6 %	1377	5.8 %	3051	5.2 %
Cardiac insufficiency	538	5.9 %	894	6.3 %	34	2.2 %	643	6.1 %	1623	6.8 %	3732	6.3 %
Diabetes mellitus	2018	22.2 %	3618	25.4 %	347	22.2 %	1636	15.4 %	4530	19.1 %	11935	20.2 %
Atherosclerosis*	189	2.1 %	1091	7.7 %	12	0.8 %	193	1.8 %	240	1.0 %	1725	2.9 %
Cancer	255	2.8 %	344	2.4 %	29	1.9 %	235	2.2 %	730	3.1 %	1588	2.7 %
COPD and asthma	1379	15.2 %	2507	17.6 %	23	1.5 %	1762	16.6 %	3269	13.8 %	8940	15.1 %
Dementia*	453	5.0 %	185	1.3 %	9	0.6 %	421	4.0 %	579	2.4 %	1643	2.8 %
Depression	1073	11.8 %	1053	7.4 %	166	10.6 %	1330	12.5 %	3709	15.6 %	7223	12.2 %
Parkinson's disease	159	1.7 %	290	2.0 %	20	1.3 %	122	1.1 %	484	2.0 %	1069	1.8 %
Mental disorders	319	3.5 %	480	3.4 %	26	1.7 %	346	3.3 %	646	2.7 %	1804	3.1 %
Renal insufficiency*	60	0.7 %	345	2.4 %	22	1.4 %	324	3.1 %	413	1.7 %	1164	2.0 %
Alcoholism*	66	0.7 %	89	0.6 %	1	0.1 %	80	0.8 %	183	0.8 %	419	0.7 %
Stroke	232	2.5 %	565	4.0 %	26	1.7 %	375	3.5 %	643	2.7 %	1841	3.1 %
	<i>Comorbidities based on diagnoses during 365 days prior to AMI</i>											
Hypertension	570	6.3 %	2762	19.4 %	89	5.7 %	1143	10.8 %	2317	9.7 %	6881	11.6 %
Coronary artery disease	1146	12.6 %	1971	13.8 %	119	7.6 %	1479	13.9 %	2711	11.4 %	7426	12.6 %
Atrial fibrillation	448	4.9 %	498	3.5 %	27	1.7 %	701	6.6 %	1377	5.8 %	3051	5.2 %
Cardiac insufficiency	538	5.9 %	894	6.3 %	34	2.2 %	643	6.1 %	1623	6.8 %	3732	6.3 %
Diabetes mellitus	458	5.0 %	1316	9.2 %	42	2.7 %	619	5.8 %	1510	6.4 %	3945	6.7 %
Atherosclerosis*	189	2.1 %	1091	7.7 %	12	0.8 %	193	1.8 %	240	1.0 %	1725	2.9 %
Cancer	234	2.6 %	306	2.1 %	23	1.5 %	217	2.0 %	624	2.6 %	1404	2.4 %
COPD and asthma	260	2.9 %	314	2.2 %	23	1.5 %	539	5.1 %	806	3.4 %	1942	3.3 %
Dementia*	203	2.2 %	168	1.2 %	4	0.3 %	372	3.5 %	281	1.2 %	1028	1.7 %

# Bruker indirekte standardisering for å beregne risikojusterte tall

- Bruker observert/forventet-metoden som beskrevet i Ash *et al.* (2003)
- Estimerer regresjonskoeffisienter fra referansedata, så beregner ratioen mellom observert og forventet antall døde/LOS for hvert land
- For logistisk regresjonsmodell: Forventede sannsynligheter fra modellen estimeres fra:
$$p = \frac{\exp( B_0 + B_1 x_1 + B_2 x_2 + \dots + B_k x_k )}{1 + \exp( B_0 + B_1 x_1 + B_2 x_2 + \dots + B_k x_k )}$$
- Multipliserer ratioen med ujustert totaldødelighet/gjennomsnittlig LOS fra referansedataene for å finne risikojusterte tall
- Ved å se på disse og fordelingen av bakgrunnsvariablene, vil man også kunne se hvorfor noen land gjør det bedre enn andre



# Hvordan teknisk gjøre dette i Stata:

- Starter med regresjonskommando, f.eks.
  - `logit d365 kovariat1 kovariat2 kovariat3 osv.`
  - `d365` er ett-års dødelighet
- Må finne predikert dødelighet fra modellen:
  - `pred pred_d365`
- Så finne SMR for hvert land:
  - `ratio pred_smr: d365/pred_d365, over(land)`
  - Nå gjenstår kun å multiplisere punkttestimat, nedre og øvre konfidensgrense for `pred_smr` med ett-års dødelighet i referansepopulasjonen, så har man justert dødelighet for hvert land med konfidensintervall
- Logit kan byttes ut med annen relevant regresjonskommando

# Risikojusteringsmodeller

- Kjører tre modeller for hvert utfall:
  - M1: Justert for alder, kjønn
  - M2: Justert for alder, kjønn, LOS foregående år, komorbiditeter basert på primær og sekundærdiagnose i NPR i foregående år
  - M3: Samme som M2, men komorbiditetene var i tillegg basert på legemiddelkjøp i Reseptregisteret

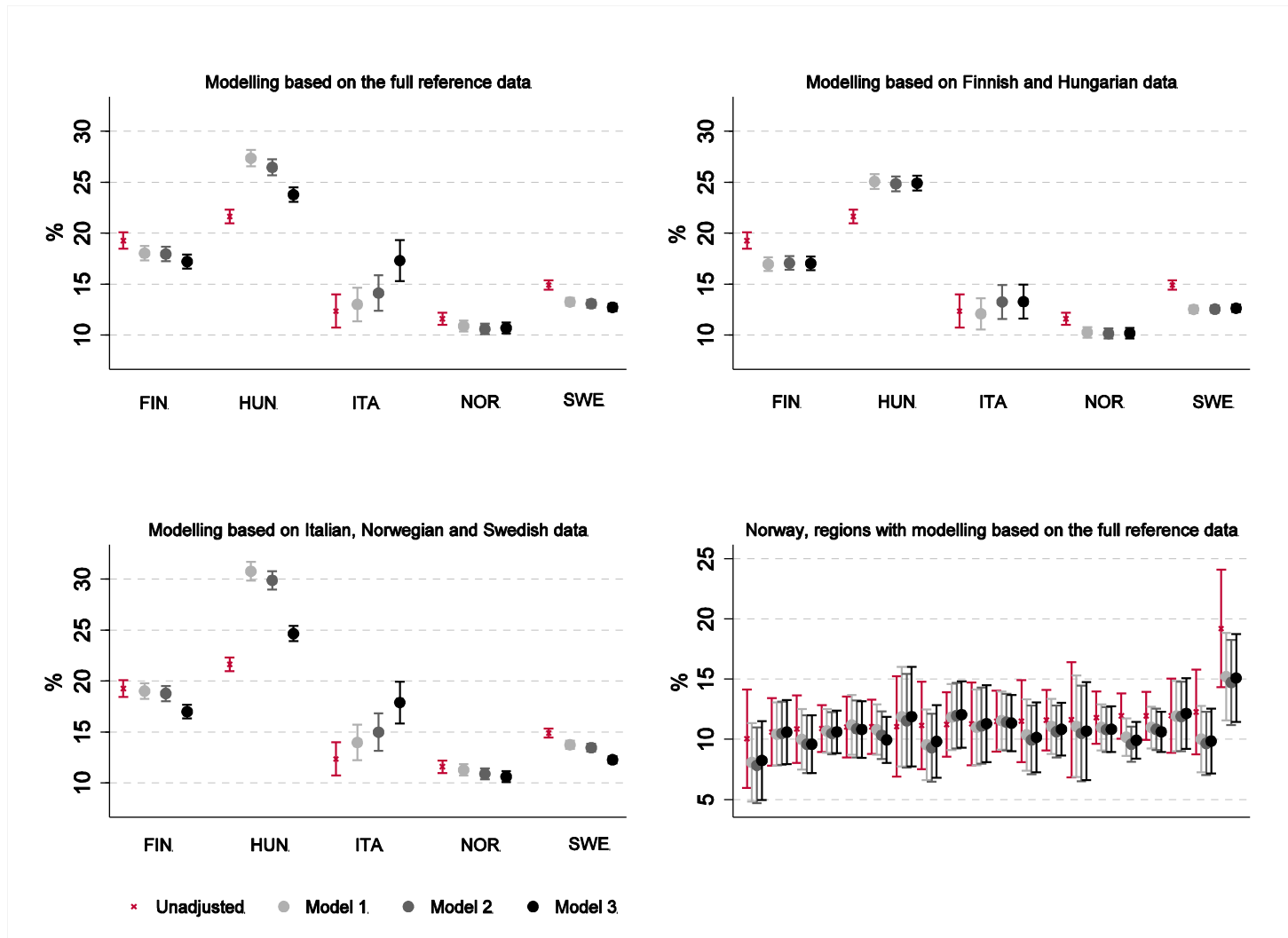
# Tilleggsproblemer i EuroHOPE:

- Ikke alle land kunne dele dataene. Dette gjaldt Nederland og Skottland
- Referansedataene er da fra de resterende landene
- Kan derfor ikke engang sjekke om det er interaksjoner mellom land og bakgrunnsvariablene
- I hvilken grad avhenger resultatene av hva man bruker som referanse?
- Gjør analysen på bare de landene som kan dele data for å studere dette

Unadjusted and risk adjusted 90-day mortality in five countries.

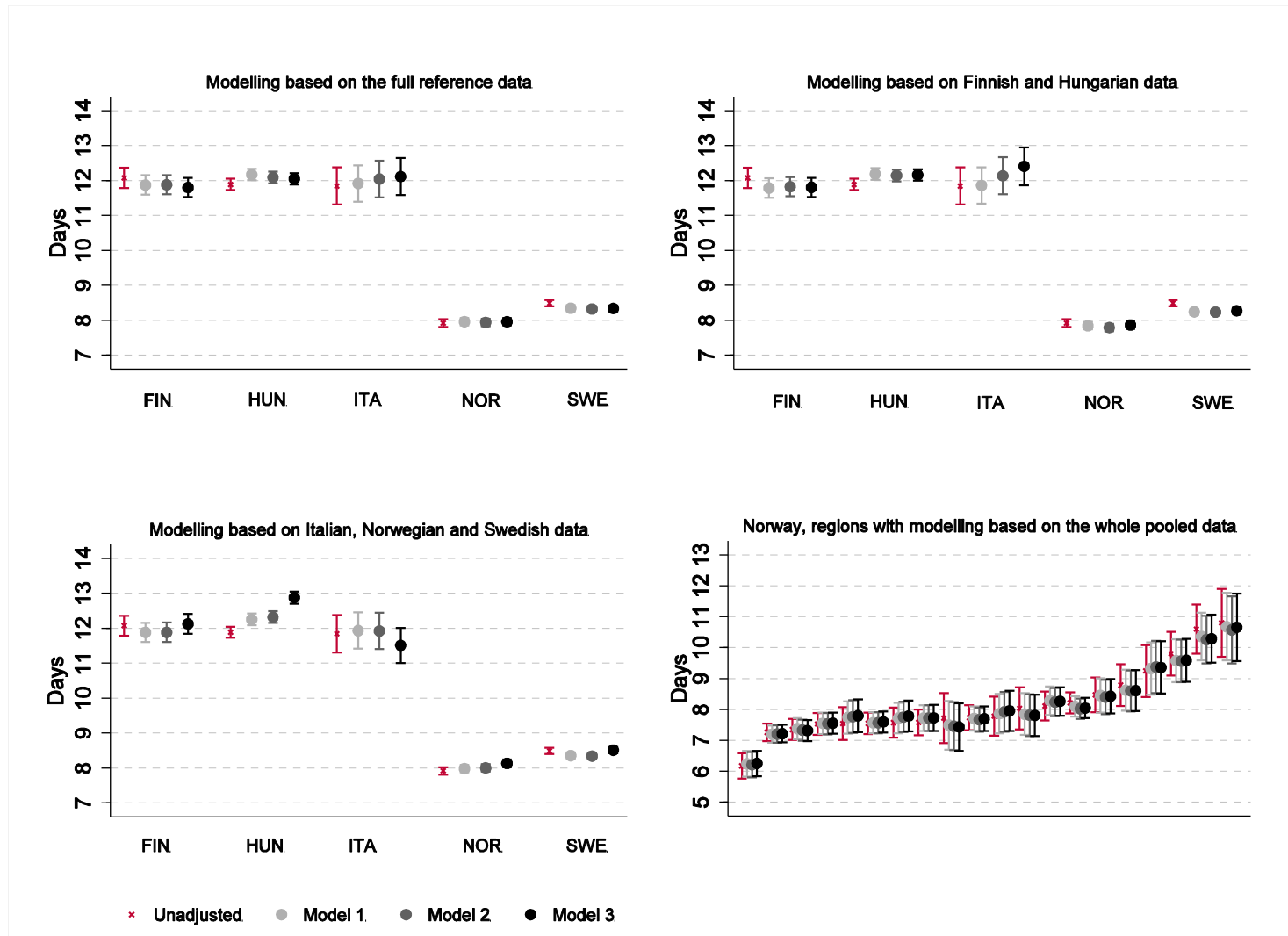
Upper left: 90-day mortality proportions with 95% confidence intervals for each country, full data used as reference in adjustment.

Upper right: 90-day mortality proportions for each country, data of Finland and Hungary used as reference in adjustment. Lower left: 90-day mortality proportions for each country, data of Norway, Sweden and Italy used as reference in adjustment. Lower right: Regional 90-day mortality proportions in Norway. M1 = sex/age adjusted, M2 = sex/age/comorbidity w/o medication adjusted, M3 =sex/age/comorbidity w medication adjusted.



## Unadjusted and risk adjusted length of stay in five countries.

Upper left: Average length of first hospital episode with 95% confidence intervals for each country, full data used as reference in adjustment. Upper right: 90-day mortality proportions for each country, data of Finland and Hungary used as reference in adjustment. Lower left: 90-day mortality proportions for each country, data of Norway, Sweden and Italy data as reference in adjustment. Lower right: Regional 90-day mortality proportions in Norway. M1 = sex/age adjusted, M2 = sex/age/comorbidity w/o medication adjusted, M3 = sex/age/comorbidity w medication adjusted.



# Svakheter i eksempelet

- Mange svakheter som følge av at ikke alle land kan dele dataene sine
- Ville ofte brukt en form for multi-level (for utfallene vist her logistisk/negativ binomisk) modell med en random effekt som tar hensyn til clustering av pasienter på samme sykehus
- Dvs. at man modellerer at pasienter på samme sykehus er likere hverandre enn pasienter på ulike sykehus, også etter at man har justert for bakgrunnsvariablene
- Får ikke engang estimert hvor forskjellig effekten av bakgrunnsvariablene er for alle land, men kun for de landene som kan dele data

# Oppsummerende tanker

- Mange spennende muligheter ved å bruke registerdata i forskning
- Har veldig store datasett, ikke noe problem å få signifikante p-verdier, blir mer viktig å se på estimatene og bedømme om forskjeller er klinisk relevante eller ei
- Har mulighet til å løfte sammenligninger opp på et internasjonalt nivå
- Men, tross alt begrensninger på hvor mye informasjon variablene i registrene faktisk klarer å fange opp
- Hvis internasjonal studie: alt må defineres likt på tvers av land, fører til grovere kategorisering av variablene
- Fortsatt mye byråkrati involvert i å få tilgang til data og koble data fra ulike registre – dette er en tidkrevende prosess

# Referanser

- Ash AS, Schwartz M, Peköz EA: Comparing outcomes across providers. In *Risk adjustment for measuring health care outcomes*. 3rd ed. Edited by Iezzoni LI. Chicago: Health Administration Press; 2003:297-333.
- Moger TA, Peltola M: Risk adjustment of health-care performance measures in a multinational register-based study: A pragmatic approach to a complicated topic . SAGE Open Medicine (2014) vol 2. doi:10.1177/2050312114526589
- Naing NN: Easy Way to Learn Standardization : Direct and Indirect Methods. Malays J Med Sci. (2000) 7: 10–15.
- Rosenbaum PR, Rubin DB: The central role of the propensity score in observational studies for causal effects. Biometrika (1983) 70: 41-55.
- Schreyögg J, Stargardt T, Tiemann O. Costs and quality of hospitals in different health care systems: a multi-level approach with propensity score matching. Health Economics (2011) 20: 85-100.