

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 behandlingskostnaden
- 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 aldersgruppe 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
- Kommando i Stata: `stdize`

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=Standardized Mortality Ratio)
- Finn alderjustert (eller standardisert) dødelighetsrate for hver studiepopulasjon ved å multiplisere SMR med den ujusterte dødelighetsraten fra referansepopulasjonen
- Kommando i Stata: `istdize`

Indirekte standardisering, eks:


Anta at: Observed deaths in population A=120 Ujustert dødelighet: $120/12500=9.6$ per 1,000
Observed deaths in population B= 30 Ujustert dødelighet: $30/6000=5$ per 1,000

Population A

Population B

Age group	Population	Age-specific death rate per1000	Expected deaths	Population	Age-specific death rate per1000	Expected deaths
0–24	2,000	4.0	8.0	1,000	4.0	4.0
25–49	2,500	7.0	17.5	1,500	7.0	10.5
50–74	3,500	10.0	35.0	2,500	10.0	25.0
75+	4,500	30.0	135.0	1,000	30.0	30.0
Total	12,500		195.5	6,000		69.5

Aldersspesifikke dødelighetsrater fra referansepopulasjonen



Indirekte standardisering, eks:

- Finner at SMR blir
 - Populasjon A: $120/195.5=0.61$
 - Populasjon B: $30/69.5=0.43$
- Dvs. begge populasjoner har lavere dødelighet enn referansen
- Anta at ujustert dødelighet i referansen var 12.0 per 1,000
- Da blir aldersjustert dødelighet i populasjonene
 - Populasjon A: $0.61*12.0=7.3$ per 1,000
 - Populasjon B: $0.43*12.0=5.2$ per 1,000
- M.a.o. forskjellen mellom populasjonene er mindre etter justering for ulik aldersprofil

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). Kallt "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 eller probit 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 skjevheten 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` og `teffects` i Stata

Hva er vi interessert i å estimere?

- Mange mulige effektmål:
 - Average treatment effect on treated (ATT). Effekten av behandling på de som fikk den, det samme som defineres som «kausalt effekt» fra RCT'er!
 - Average treatment effect (ATE). Denne er definert som en gjennomsnittseffekt over alle i studien, dvs uavhengig av hvilken gruppe man var i
 - Forskjell mellom de to: hvis behandling ikke er relevant for alle i populasjonen, gir det mest mening med ATT. Hvis behandlingen derimot er mulig for alle, er kanskje ATE mer relevant
 - Finnes også Local Average Treatment Effects (LATE) og Marginal treatment effect (MTE)
- Men kan også være interessert i fordelingen av effektene: Hvor mange har positiv effekt av «behandling» og hvor stor er denne?

Enkle eksempler på kommandoer

- Er det bare to grupper av data, og du er i en situasjon hvor en to-utvalgs t-test hadde vært nok for å besvare forskningsspørsmålet (hvis du hadde hatt en RCT), trenger man ikke skrive masse kode for å gjøre analysen
- Eksempler fra Stata:
www.ssc.wisc.edu/sscc/pubs/stata_psmatch.htm
- Fire variabler: Behandlingsindikator t , kovariater x_1 and x_2 , utfall y . Sannsynlighet for behandling er korrelert med x_1 og x_2 , både x_1 og x_2 er korrelert med y . Mao: Kan ikke bare sammenligne y mellom gruppene

Eksempel:

```
. ttest y, by(t)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]		
Ujustert forskjell	0	667	-.4232434	.0634689	1.63917	-.5478667	-.2986201
	1	333	1.891074	.0876176	1.598872	1.718718	2.063429
combined	1,000	.3474242	.0619004	1.957462	.2259545	.4688939	
diff			-2.314317	.1090943		-2.528398	-2.100236

diff = mean(0) - mean(1) t = -21.2139
 Ho: diff = 0 degrees of freedom = 998

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000

```
. reg y x1 x2 t
```

Source	SS	df	MS	Number of obs	=	1,000
Model	2780.33314	3	926.777714	F(3, 996)	=	881.22
Residual	1047.49123	996	1.05169803	Prob > F	=	0.0000
Total	3827.82438	999	3.83165603	R-squared	=	0.7263
				Adj R-squared	=	0.7255
				Root MSE	=	1.0255


Justert forskjell er mye mindre

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x1	1.031167	.0346941	29.72	0.000	.9630853	1.099249
x2	.9927759	.0333297	29.79	0.000	.9273715	1.05818
t	.9791484	.0769067	12.73	0.000	.8282306	1.130066
_cons	.0591595	.0416008	1.42	0.155	-.0224758	.1407948

Eksempel:

Her brukes kommandoen med default-valg:

Vil gjøre 1-1 matching ved å velge nærmeste match med tilbakelegging uavhengig av hvor langt unna den er, probit regresjon for estimering av propensity score

 . psmatch2 t x1 x2, outcome(y)

```
Probit regression                               Number of obs   =       1,000
                                                LR chi2(2)      =       222.39
                                                Prob > chi2     =       0.0000
Log likelihood = -525.08947                    Pseudo R2      =       0.1748
```

t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
x1	.5305378	.0496371	10.69	0.000	.4332509	.6278247
x2	.4772386	.0465109	10.26	0.000	.3860789	.5683984
_cons	-.5090219	.0455374	-11.18	0.000	-.5982737	-.4197702

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
y	Unmatched	1.8910736	-.423243358	2.31431696	.109094342	21.21
	ATT	1.8910736	.871388246	1.01968536	.173034999	5.89

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support On suppor		Total
	Untreated	Treated	
Untreated	667	667	
Treated	333	333	
Total	1,000	1,000	

 Vanlig utfordring ved propensity score analyser

Eksempel: Balansering av gruppene

Ser at behandlings- og kontrollgruppene var veldig ulike i utgangspunktet, men er ikke det etter matching

```
. pstest, both
```

Variable	Unmatched Matched	Mean		%reduct		t-test		V(T) / V(C)
		Treated	Control	%bias	bias	t	p> t	
x1	U	.4327	-.23546	70.8		10.49	0.000	0.93
	M	.4327	.39773	3.7	94.8	0.49	0.624	1.04
x2	U	.40954	-.24135	66.3		9.81	0.000	0.92
	M	.40954	.44574	-3.7	94.4	-0.49	0.623	1.05

* if variance ratio outside [0.81; 1.24] for U and [0.81; 1.24] for M

Annen kommando:

Denne tar hensyn til at propensity scorene blir estimert, men ~~mangler mulighet for å vise hvor balansert det matchede datasettet er.~~ (kan få ut dette vha tebalance)



```
. teffects psmatch (y) (t x1 x2, probit), atet
```

```
Treatment-effects estimation      Number of obs      =      1,000
Estimator      : propensity-score matching      Matches: requested =      1
Outcome model  : matching                      min =      1
Treatment model: probit                      max =      1
```

y		AI Robust		z	P> z	[95% Conf. Interval]
	Coef.	Std. Err.				
ATET						
t						
(1 vs 0)	1.019685	.1227801	8.30	0.000	.7790407	1.26033

Example: Alcohol and mortality

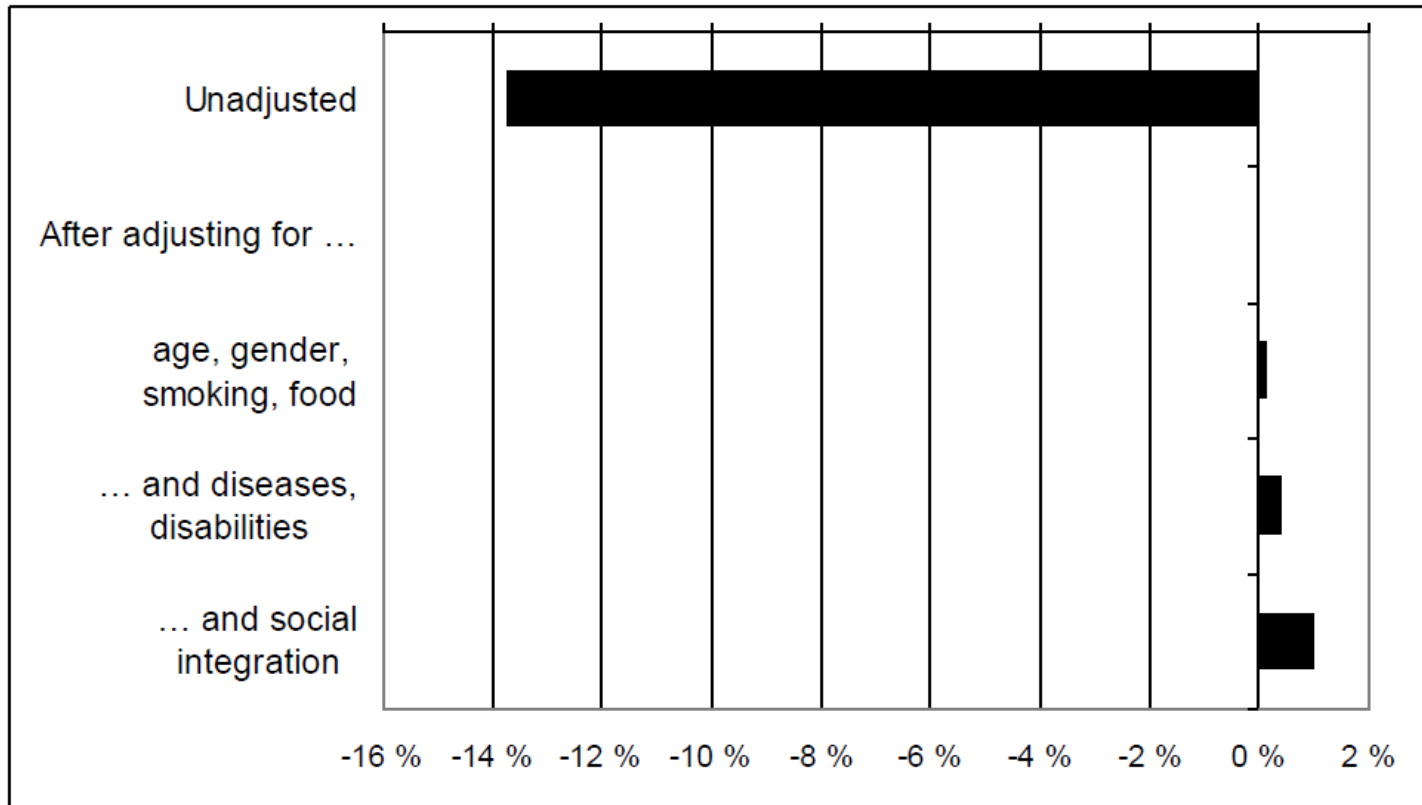
- Does moderate consumption of alcohol decrease mortality?
- Health and Lifestyle Survey (HALS), UK
 - 9 003 adults interviewed in 1984/5
 - 75% moderate drinkers
 - 25% abstainers or special occasion drinkers
 - Updated with information about mortality
 - In 2003
 - 35% of non-drinkers were dead
 - 20% of moderate drinkers
- So alcohol reduce mortality?
 - Problem I: Confounding variables
 - Problem II: Difficult to conduct experiment
- Try propensity score analysis

Output from pstest:

Table 1. Overview of important confounding variables, averages, and bias reduction

Variable	Sample	Average for drinkers	Average for non-drinkers	Percent difference	Reduction in bias (%)	p-value
Age	Unmatched	44,78	54,42	-56,5		0,00
	Matched	45,17	45,42	-1,4	97,4	0,21
Gender	Unmatched	0,53	0,25	60,0		0,00
	Matched	0,52	0,52	0,9	98,5	0,36
Income	Unmatched	5,58	4,45	45,4		0,00
	Matched	5,50	5,39	4,6	89,9	0,02
Smoking	Unmatched	0,35	0,27	18,4		0,00
	Matched	0,35	0,32	4,8	73,8	0,04
Disabilities	Unmatched	0,29	0,36	-13,0		0,00
	Matched	0,30	0,31	-3,4	73,5	0,23
Body Mass Index	Unmatched	24,50	25,08	-13,7		0,00
	Matched	24,54	24,52	0,6	95,3	0,89
Heart trouble	Unmatched	0,07	0,10	-10,6		0,00
	Matched	0,07	0,08	-3,0	72,2	0,24
Depression/ Nervous illness	Unmatched	0,18	0,22	-10,4		0,00
	Matched	0,18	0,17	1,8	82,7	0,67
High blood pressure	Unmatched	0,14	0,21	-18,7		0,00
	Matched	0,15	0,14	2,7	85,7	0,51

Example: Alcohol and mortality



* Percentage difference in deaths among drinkers compared to non-drinkers 19 years after the interview.

Merknad:

- I alle disse standard-anvendelsene brukes altså bakgrunnsvariabler/confoundere kun for å estimere propensity-scoren, selve hovedanalysen er i praksis parallellt til en to-utvalgs t-test
- Gir sjelden mening å gjøre regresjonsanalyser på det matchede datasettet, akkurat som for RCT'er: Ved å gjøre regresjon på det matchede datasettet, innrømmer du at matchingen (eller randomiseringen hvis RCT) ikke var god nok
- Unntaket er selvsagt hvis du har et sett av variabler du matcher på (eks. individuelle faktorer) og et annet sett av variabler du vil måle effekten av (eks. sykehus/region-variabler)

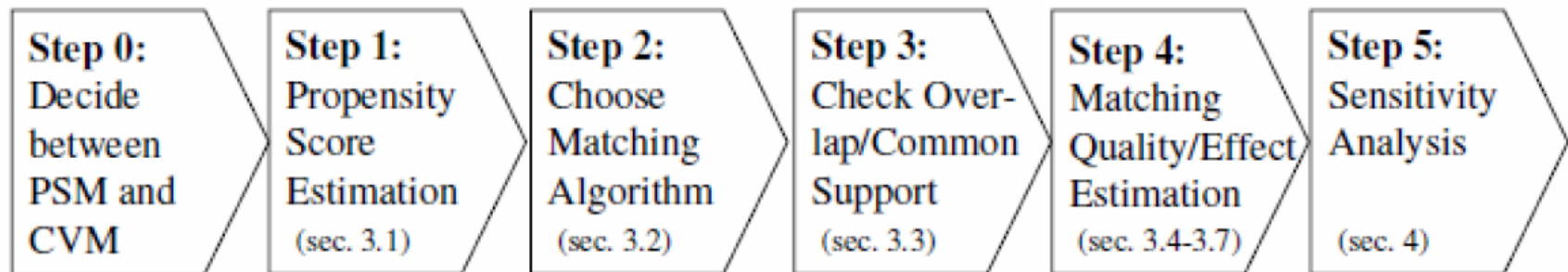
Propensity score - Svakheter

- 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

Konklusjon:

- Propensity score analyse er ikke en løsning på alt.
- Men videre muligheter utover det som finnes i standard programvare:
 - Flere enn to grupper som skal sammenlignes:
 - Enkleste er å gjøre gjentatte sammenligninger av to grupper, kan da bruke kommandoene vi har sett. Men må velge felles referanse for alle analyser slik at gruppene blir balansert på bakgrunnsfaktorer (behandling1 vs behandling2, beh1 vs beh3 osv)
 - Eller multinomisk logistisk regresjon for å beregne propensity scores, men begrensede ferdige kommandoer for dette
 - Usikkerheten i at propensity-scorene estimeres kan fikses ved bootstrap
 - Finnes test for hvor stor effekten av umålte confoundere må være for å ha betydning for konklusjonen (Stata f.eks. rbounds, mhbounds, sensatt)

Steg i propensity score analyse:



CVM=Vanlig regresjonsjustering

For større praktisk eksempel, se Schreyögg et al. (2011): Vil sammenligne kostnader og within hospital dødelighet for hjerteinfarktpasienter i Tyskland og USA

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

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 for hvert land
- For logistisk regresjonsmodell: Forventede sannsynligheter fra modellen estimeres fra:
$$p = \frac{\exp(B_0 + B_1x_1 + B_2x_2 + \dots + B_kx_k)}{1 + \exp(B_0 + B_1x_1 + B_2x_2 + \dots + B_kx_k)}$$
- Multipliserer ratioen med ujustert totaldødelighet 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 d90 kovariat1 kovariat2 kovariat3 osv.`
 - d90 er 90-dagers dødelighet
- Må finne predikert dødelighet fra modellen:
 - `pred pred_d90`
- Så finne SMR for hvert land:
 - `ratio pred_smr: d90/pred_d90, 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

Alternativ måte for risikojusterte estimer:

- Kan beregne marginale effekter:
 - Starter med regresjonskommando, f.eks.
 - `logit d90 i.country kovariat1 kovariat2 kovariat3 osv.`
 - d90 er 90-dagers dødelighet
 - Så gi kommandoen: `margins country`
 - Logit kan byttes ut med annen relevant regresjonskommando
- Margins beregner: Hva ville dødeligheten vært, gitt verdiene av andre variabler i modellen, hvis alle pasientene var norske? Hva hvis alle var finske? Altså separere ut selvstendig effekt av variabelen man angir i margins.
 - Margins krever at man kan ha alle data i en database (sikkert ikke noe problem for dere)
 - Metoden på foregående slide kan brukes uavhengig av dette (relevant for EuroHOPE-eksemplet)

Tema i eksempel – utfordringer i internasjonale sammenligninger:

- 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

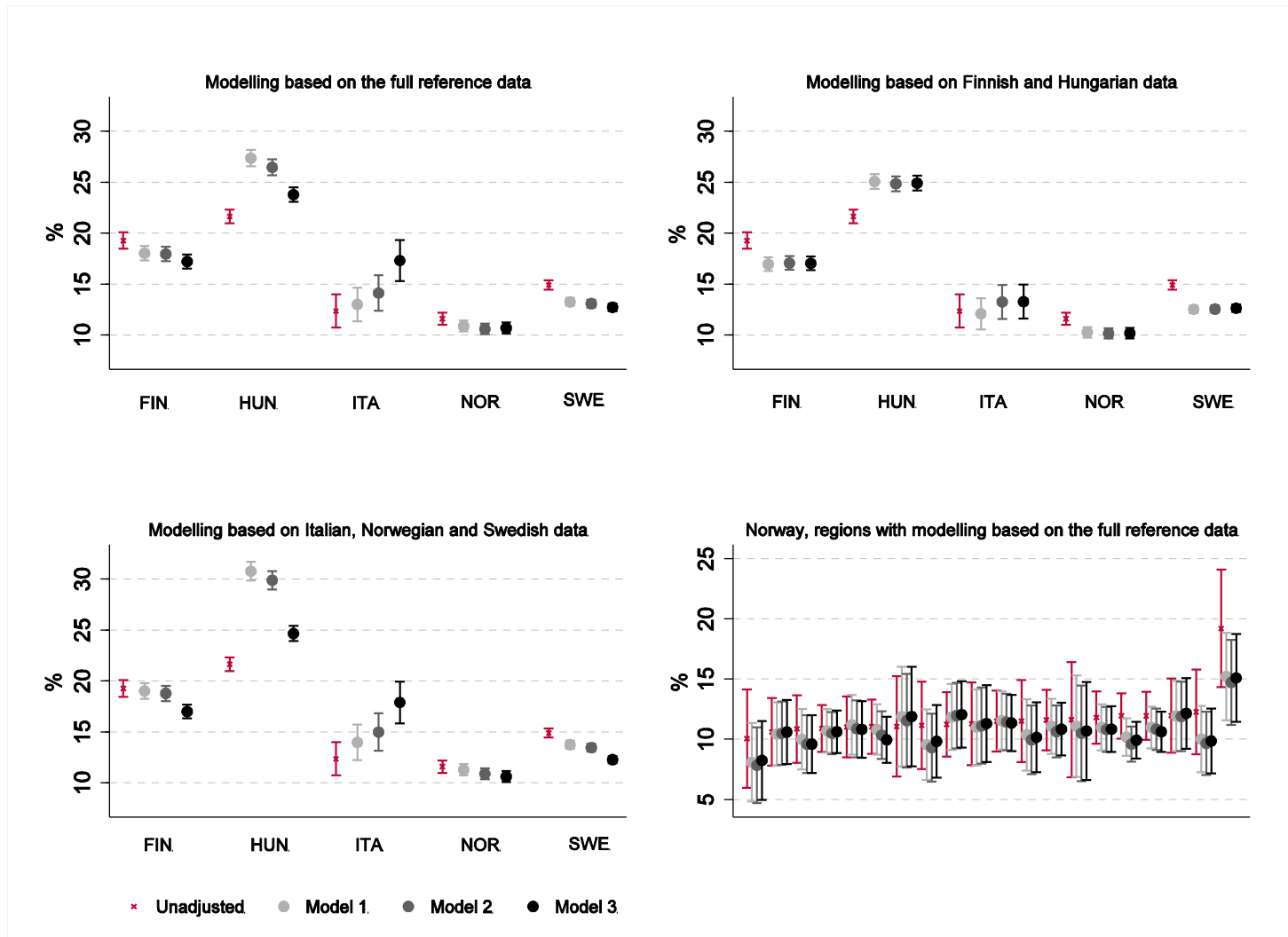
Risikojusteringsmodeller

- 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

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.



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 utfallet vist her logistisk) 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.
- Caliendo M, Kopeinig S: Some Practical Guidance for the implementation of Propensity Score Matching. *Journal of Economic Surveys* 2008: <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- 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.