Kan vi stole på resultater fra «liten N»?

Olav M. Kvalheim

Universitetet i Bergen

Plan for dette foredraget

- Hypotesetesting og p-verdier for å undersøke en variabel
- p-verdier når det er mange variabler
- Er økt populasjon (*N*) eller flere variabler (mønster) best for å oppnå økt signifikans?
- Eksempel

Hypotesetesting og p-verdier

Oppgave: Vi ønsker å finne ut om kvinner og menn i en region, f.eks. en kommune, har forskjellig for nivå av *det gode kolesterolet* (HDL) i blod.

I statistikkens verden formuleres oppgaven som en hypotese, *H1*, som må testes opp mot den alternative hypotesen *H0* (*nullhypotesen*) om at det ikke er en forskjell.

Hypotesetesting og p-verdier

Fremgangsmåte for å løse oppgaven: Vi samler inn og analyserer blodprøver fra *N* kvinner og *N* menn valgt tilfeldig i kommunen, men med en lik profil i alderssammensetning.

Hypotesetesting og p-verdier

Resultat av datainnsamling: Vi får en tabell med to kolonner, en for HDL i blod hos menn og en for kvinner (*N*=10):

Menn	Kvinner
1,1	1,9
1,6	1,5
1,3	1,3
0,9	1,8
1,7	1,4
1,4	1,6
1,5	1,5
1,1	1,0
1,2	1,7
1,3	1,4

Hypotesetesting og p-verdier

Analyse av data: Vi tallene i tabellen og får en sum for HDL i blod hos menn og en for kvinner:

Menn	Kvinner
1,1	1,9
1,6	1,5
1,3	1,3
0,9	1,8
1,7	1,4
1,4	1,6
1,5	1,5
1,1	1,0
1,2	1,7
1,3	1,4
13,1	**15,1**

Hypotesetesting og p-verdier

Analyse av data: Ved å dele summen på *N*=10 får vi et *gjennomsnitt* for HDL i blod hos menn og kvinner:

	Menn	Kvinner
Sum	13,1	15,1
Snitt	13,1 : 10 = **1,31**	15,1 : 10 = **1,51**

Hypotesetesting og p-verdier

Resultat av analyse: Kvinner har i gjennomsnitt 0,2 enheter mer av «det gode kolesterolet» HDL i blodet enn menn.

Hypotesetesting og p-verdier

Konklusjon: Kvinner har høyere nivåer av HDL enn menn!

Kan vi stole på dette?
Er dette et resultat som er «statistisk signifikant»?

Hypotesetesting og p-verdier

For å svare på dette må vi gjøre litt mer analyse.

Hypotesetesting og p-verdier

Mer analyse av data: Vi regner ut differansen mellom målt HDL og gjennomsnittet beregnet for menn (1,3) og kvinner (1,5):

Menn	Kvinner
-0,2	0,4
0,3	0
0	-0,2
-0,4	0,3
0,4	-0,1
0,1	0,1
0,2	0
-0,2	-0,5
-0,1	0,2
0	-0,1

Hypotesetesting og p-verdier

Mer analyse av data: Vi regner ut *variasjonen* i HDL for menn og kvinner ved å summere de kvadrerte avvikene fra gjennomsnittet for kvinner og menn:

Menn	Kvinner
0,04	0,16
0,09	0
0	0,04
0,16	0,09
0,16	0,01
0,01	0,01
0,04	0
0,04	0,25
0,01	0,04
0	0,01
0,55	**0,61**

Hypotesetesting og p-verdier

Mer analyse av data: Ved å dele variasjonen i HDL for menn og kvinner med *N*-1 = 9 (tap av en frihetsgrad siden vi har beregnet gjennomsnitt HDL) får vi *variansen* i HDL:

	Menn	Kvinner
Variasjon	0,55	0,61
Varians	0,55 : 9 = **0,06**	0,61 : 9 = **0,07**

Hypotesetesting og p-verdier

Mer analyse av data: Ved å ta kvadratroten av varians får vi *standard avvik* i HDL for menn og kvinner:

	Menn	Kvinner
Variasjon	0,55	0,61
Varians	0,55 : 9 = 0,06	0,61 : 9 = 0,07
Stand. Avvik	√0,06 = **0,24**	√0,07 = **0,26**

Standard avvik er et mål på *spredning* av målte HDL-nivåer i menn og kvinner!

Hypotesetesting og p-verdier

Resultat av analyse:

	Snitt ± Stand. Avvik
Menn	1,31 ± 0,24
Kvinner	1,51 ± 0,26

Forutsetning: Målingene tilfeldig fordelt rundt snittverdiene. 68% av målingene fordele seg innenfor et stand. avvik over (+) og under (-) snittverdiene.

Hypotesetesting og p-verdier

1. Vi ønsket å teste om kvinner og menn har forskjellig nivå av «det gode kolesterolet» (HDL) i blod.
2. Vi har samlet inn materiale fra N kvinner og menn og vil teste en nullhypotese om at de er like.
3. Ved å utføre en test (f.eks. en såklat *t*-test når data er normalfordelt både for kvinner og menn) kan vi beregne en sannsynlighet *p* for at nullhypotesen er korrekt.
4. Dersom *p* blir mindre enn en valgt grenseverdi, ofte 0.05, forkastes nullhypotesen, og vi kan konkludere med at det er en signifikant forskjell mellom kvinner og menn i HDL-nivå.

Hva bestemmer signifikans (p-verdi)?

Forskjellen i middelverdi og spredning (standardavvik) av HDL-nivå for kvinner og menn.

Ved å øke antall observasjoner blir spredningen (standard avvik) mindre og signifikans av testen øker.
=> Signifikans øker med kvadratroten av antall observasjoner, $\sqrt{N}_{kvinner}$ og $\sqrt{N}_{menn}$.

Små forskjeller kan bli signifikante (liten *p*) dersom antall målinger er stort!

Hva bestemmer signifikans (p-verdi)?

Store forskjeller kan bli signifikante (liten *p*) selv om antall målinger er lite dersom det er (forholdsvis) stor forskjell i middelverdi eller liten spredning!

Hva om vi har flere variabler som skal testes for signifikans?

Vi ønsker å sammenligne om middelverdien for kvinner og menn er like for nivå i blod av TC, LDL, HDL og TG (nullhypotese).

Siden sannsynligheten for å finne forskjeller øker når antall tester øker, må vi korrigere for multippel testing.

Dette kan gjøre på to måter, Bonferroni eller «false discovery rate», FDR.

Korrigerte p-verdier for Bonferroni og FDR

For Bonferroni med 4 tester kreves p-verdi mindre eller lik 0.0125 for å forkaste nullhypotesen for hver enkelt variabel svarende til 0.05-nivå for en test.

For FDR sorteres p-verdiene i stigende rekkefølge. For variabelen med lavest p-verdi må p være mindre eller lik 0.0125 for at noen av variablene skal være signifikant på 0.05-nivå, men hvis første er under så er neste signifikant på 0.05-nivå dersom p er mindre eller lik 0.025 osv.

Hvilke konsekvenser har denne strategien for hypotesetesting?

Måling av mange variabler => større N er nødvendig =>

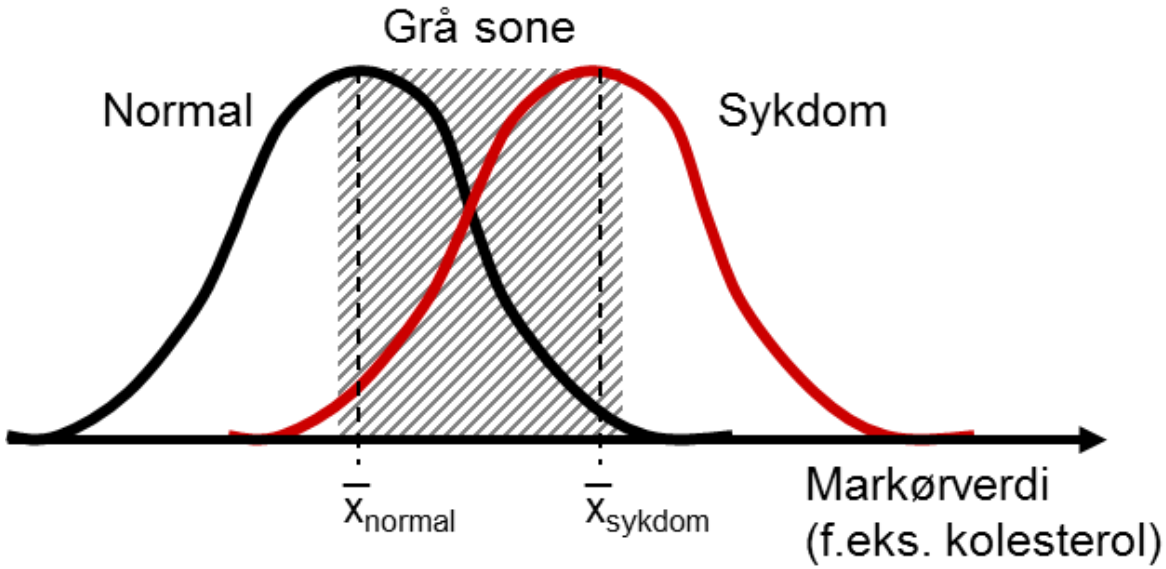
Fokus på en variabel om gangen siden testing av flere variabler krever sterk økning N; 2 variabler impliserer 4N, 3 variabler 9N osv. for å opprettholde samme signifikansnivå (p) forutsatt omtrent samme forskjell i middelverdi og spredning på de målte variablene.

Er det andre konsekvenser?

Man tar en datatabell med mange variabler og plotter en og en variabel om gangen inntil man (tilfeldigvis) finner en sammenheng.

Men: Man har da egentlig gjort en multippel testing uten å korrigere p!

Tradisjonelt – en variabel om gangen



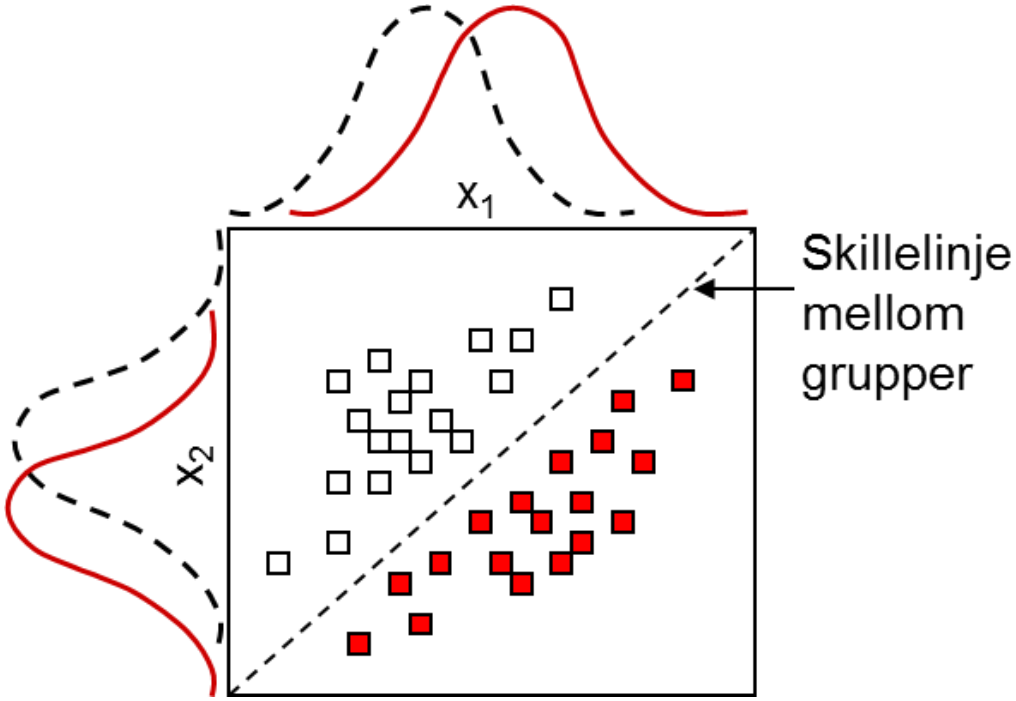
I tradisjonell «univariat» tilnærming brukes verdi på en variabel (markør) for å skille mellom to grupper individer. Grå sone viser område der det kan være vanskelig å si hvilken gruppe en person tilhører.

Finnes det en alternativ strategi?

Måling av mange variabler samtidig f. eks. med spørreskjema => Lav pris sammenlignet med å øke N.

Finnes det en strategi som kan gi høy signifikans med forholdsvis liten N, men mange variabler?

Bedre alternativ – mønster av flere variabler



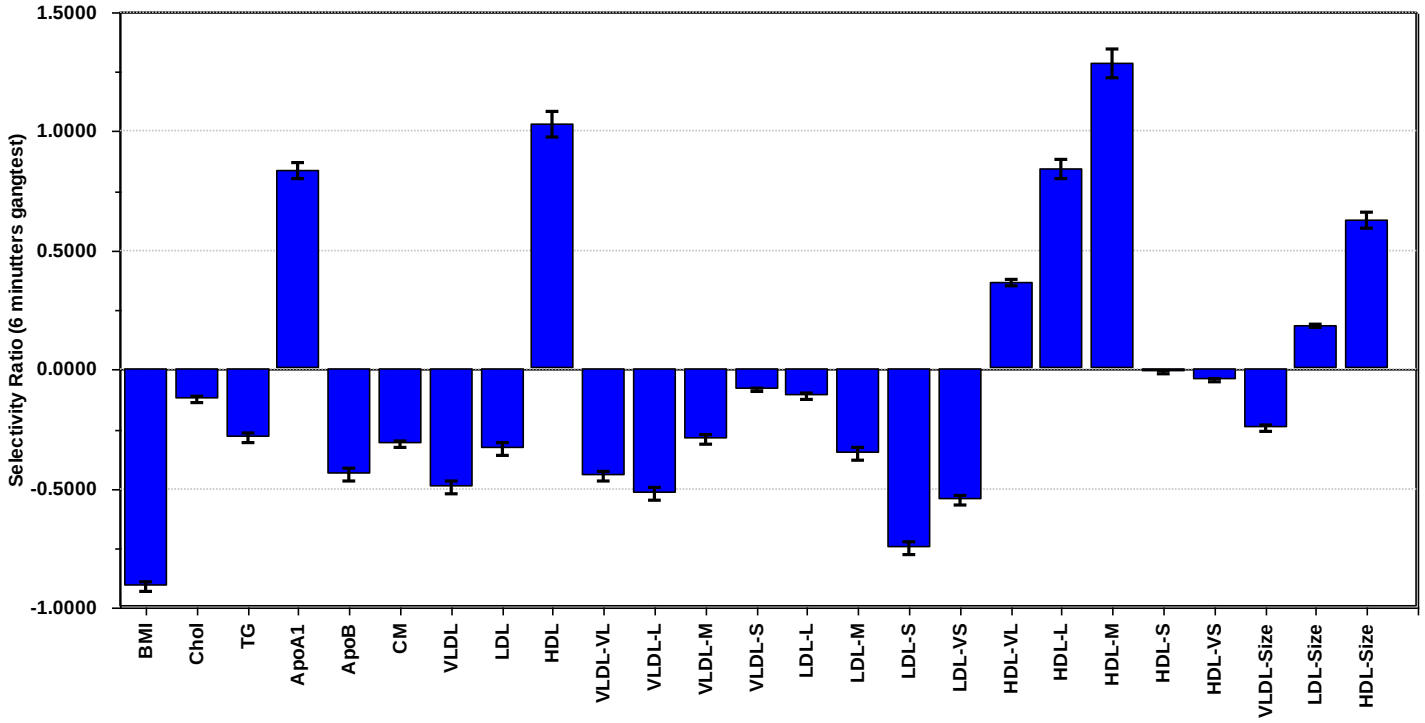
I «multivariat» tilnærming brukes flere variabler (markører) sammen for å gi et mønster som hjelper for å se forskjellene mellom de to gruppene av individer.

Bedre alternativ – mønster av flere variabler

Høy korrelasjon (sammenheng) mellom variablene betyr høy (statistisk) signifikans med lav N!

Sammenheng mellom kardiorespiratorisk form og lipoproteinmønster i friske voksne

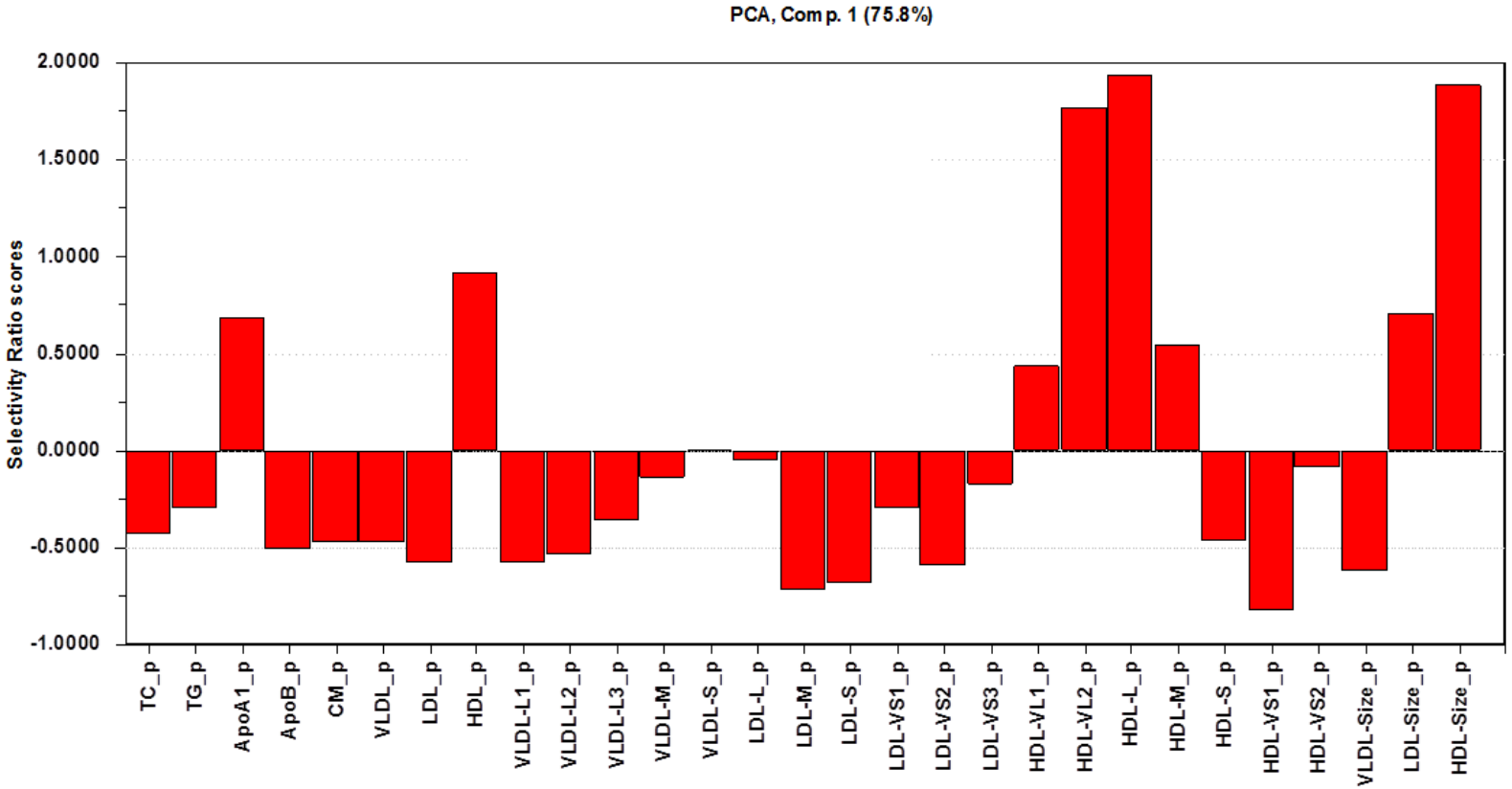
Lipoproteinmønster relatert til resultat av 6 minutters gangtest, voksne kvinner (N ca. 40)



Korrelasjon mellom målt og beregnet antall meter gått: R2Y=0.57

Fedme øker hastigheten på metabolsk aldring

Endringer i fettstoffskiftet etter fedmeoperasjon



Takk for oppmerksomheten!

Kjønnsforskjeller og forandringer i lipoproteinmønster fra prepubertet til voksen og ved aldring

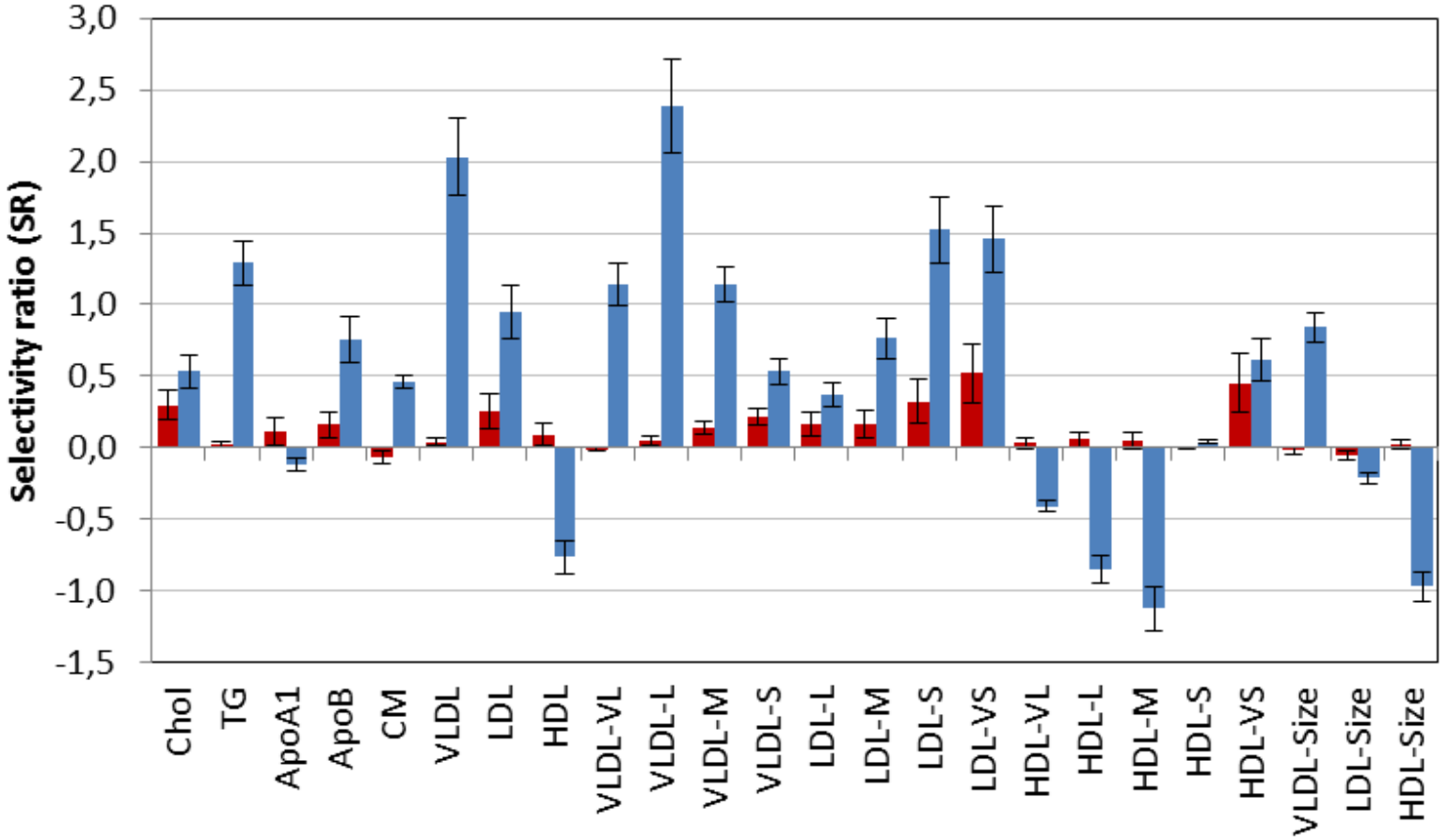
Rajalahti et al. (2016), *Metabolomics*, 12:51.
Rajalahti et al. (2016) *Metabolomics*, 12:81.

Lipoprotein-mønster for barn

- Små kjønnsforskjeller
- Gutter har høyere HDL konsentrasjon enn jenter
- Jenter har høyere VLDL (og TG) konsentrasjoner enn gutter

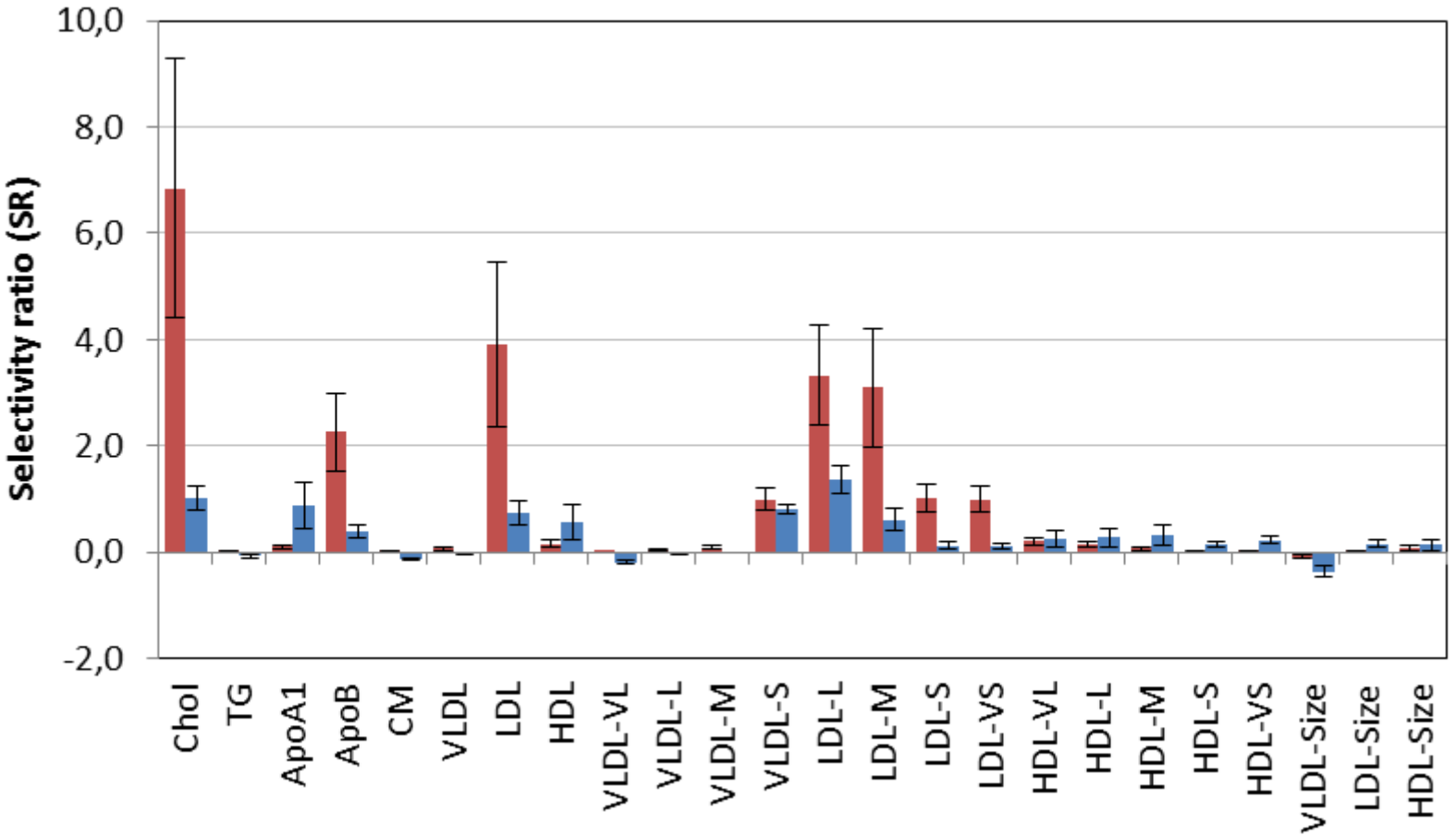
Forklaring: Gutter er mer fysisk aktive enn jenter

Forandring i lipoproteinmønster fra prepubertet til voksen



Blå => menn, rød => kvinner

Forandringer i lipoproteinmønster med aldring



Blå => menn, rød => kvinner