
Justering av p-verdier ved multiple hypoteser

MEDISIN OG TALL

STIAN LYDERSEN

stian.lydersen@ntnu.no

Stian Lydersen er dr.ing. og professor i medisinsk statistikk ved Regionalt kunnskapssenter for barn og unge – psykisk helse og barnevern (RKBU Midt-Norge) ved Institutt for psykisk helse, NTNU. Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Det er ganske vanlig å undersøke flere hypoteser i én og samme studie, noe som øker sannsynligheten for å gjøre type I-feil. Dette kan håndteres på flere måter.

Det kan være ulike grunner til at en forsker vil teste flere hypoteser i samme studie, for eksempel at man ønsker å studere effekt på flere utfallsvariabler, sammenlikne flere enn to grupper eller gjøre separate analyser for undergrupper.

Ulike metoder for justering

La oss tenke oss en studie der man skal gjennomføre seks hypotesetester. Hvis man gjør alle testene ved et signifikansnivå på 5 %, vil man ved hver av dem ha en sannsynlighet på 5 % for å gjøre en type I-feil, dvs. feilaktig forkaste nullhypotesen (1). Sannsynligheten for å gjøre type I-feil i minst én av hypotesetestene, også kalt familievis feilrisiko (*familywise error rate*, FWER) (2), vil da være vesentlig høyere enn 5 %, i verste fall oppimot 30 %. Noen ganger er det ønskelig å kontrollere denne feilrisikoen slik at den ikke overstiger en forhåndsbestemt grense, for eksempel et signifikansnivå på 5 %.

Den enkleste metoden er en såkalt Bonferroni-korreksjon. Man multipliserer da p -verdiene med antall hypoteser, i dette eksempelet seks, før man sammenlikner med signifikansnivået. Bonferroni-korreksjon er imidlertid

svært konservativ, dvs. at den statistiske styrken, og dermed sannsynligheten for å påvise faktisk sanne hypoteser, blir sterkt redusert. Ved Šidák-korreksjon oppnår man bare en marginal forbedring. Alternative metoder etter økende statistisk styrke er Holms *step-down*-korreksjon, Hochbergs *step-up*-korreksjon og Hommel-korreksjon (3). Disse metodene gjelder under generelle forutsetninger og kan generelt anbefales.

I noen situasjoner undersøker man svært mange hypoteser. I genetikkstudier kan man for eksempel ha flere hundre tusen hypoteser. Da vil det i praksis være umulig å kontrollere familievis feilrisiko. I stedet nøyer man seg med å kontrollere falsk deteksjonsandel (*false discovery rate*, FDR) (2): Man tillater at en viss andel, vanligvis 5 %, av de hypotesene som vi markerer som sanne i én og samme studie, kan være falskt positive funn. Hvis man kontrollerer familievis feilrisiko, derimot, ville man ikke «akseptere» et eneste falskt positivt funn. Den vanligste metoden for å kontrollere falsk deteksjonsandel kalles Benjamini-Hochberg-korreksjon (4). Kontroll av falsk deteksjonsandel kan også være relevant i mindre studier med for eksempel så få som 8 til 16 hypotesetester, selv om fordelene er klarere ved et høyt antall hypotesetester (4).

La oss se på et eksempel der man har seks ujusterte p -verdier listet etter størrelse (tabell 1). Vi ser at metoder som gir høyere statistisk styrke, typisk vil gi lavere p -verdier, og at den laveste justerte p -verdien er den samme som man får ved Bonferroni-korreksjon, uansett metode. Den siste kolonnen med Benjamini-Hochberg-justerte p -verdier kontrollerer bare for falsk deteksjonsandel. Med kun seks hypotesetester vil man i praksis benytte en annen metode.

Tabell 1

Et eksempel med seks p -verdier, ujusterte og justert etter forskjellige korreksjonsmetoder.

Ujustert p -verdi	Bonferroni	Šidák	Holms <i>step-down</i>	Hochbergs <i>step-up</i>	Hommel	Benjamini-Hochberg
0,0003	0,0018	0,0018	0,0018	0,0018	0,0018	0,0018
0,009	0,054	0,053	0,045	0,042	0,028	0,021
0,013	0,078	0,076	0,052	0,042	0,039	0,021
0,014	0,084	0,081	0,052	0,042	0,042	0,021
0,04	0,24	0,22	0,08	0,08	0,06	0,048
0,06	0,36	0,31	0,08	0,08	0,06	0,06

Alltid justere?

Trenger vi alltid å justere for multiple hypoteser? Dette er et kontroversielt spørsmål. Epidemiologen Kenneth Rothman argumenterer mot å justere for multiplisitet i en del sammenhenger (5). Vi kan sette det på spissen: Tenk deg en forsker som studerer effekt av en behandling på tre utfallsvariabler. Trenger han å justere for multiplisitet hvis han splitter resultatene i tre publikasjoner, med bare én hypotese i hver? Eller kanskje han burde justere for alle hypotesene han har undersøkt i sin karriere?

Det finnes alternativer til justering. I en studie med flere utfallsvariabler er det vanlig å spesifisere hvilken som er den primære. Man gjennomfører hypotesetestene uten å justere, men legger «mindre vekt» på eventuelle funn på sekundære utfallsvariabler. I andre situasjoner kan det være aktuelt å velge en pragmatisk løsning som å sette signifikansnivået til 1 % istedenfor 5 %. Dette vil gi en viss beskyttelse mot falskt positive funn, men vanligvis uten å redusere statistisk styrke så mye som man ville ha gjort ved en formell justering.

Det finnes ingen allmenn konsensus om når, og eventuelt hvordan, man bør justere for multiple hypoteser. Men valg av fremgangsmåte må spesifiseres i protokollen eller analyseplanen på forhånd, for å unngå «fisking» etter signifikante funn.

LITTERATUR

1. Lydersen S. Type I-feil og type II-feil. Tidsskr Nor Legeforen 2021; 141. doi: 10.4045/tidsskr.21.0013. [PubMed][CrossRef]
2. Lydersen S. Justering av p-verdier på norsk. Tidsskr Nor Legeforen 2021; 141. doi: 10.4045/tidsskr.21.0360. [CrossRef]
3. Dmitrienko A, D'Agostino R. Traditional multiplicity adjustment methods in clinical trials. Stat Med 2013; 32: 5172–218. [PubMed][CrossRef]
4. Benjamini Y, Hochberg Y. Controlling the false discovery rate – A practical and powerful approach to multiple testing. J R Stat Soc B 1995; 57: 289–300. [CrossRef]
5. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990; 1: 43–6. [PubMed][CrossRef]

Publisert: 27. september 2021. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.21.0357
Opphavsrett: © Tidsskriftet 2026 Lastet ned fra tidsskriftet.no 27. mars 2026.