



Manglende uttrykk for manglende data

SPRÅKSPALTEN

STIAN LYDERSEN

E-post: 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, Fakultet for medisin og helsevitenskap, Norges teknisk-naturvitenskapelige universitet (NTNU).

I datasett vil enkelte verdier kunne mangle. Data kan mangle helt tilfeldig, betinget tilfeldig eller ikke-tilfeldig. Dette bør gjenspeiles i de norske betegnelsene oversatt fra engelsk.

Som bidragsyter til Tidsskriftets spalte Medisin og tall blir man anmodet om å bruke norske fagtermer. Men i enkelte tilfeller finnes de ikke. Artikkelen om manglende data i dette nummeret av Tidsskriftet er et eksempel (1).

I hvilken grad mangler data tilfeldig? Dette kalles mekanisme for manglende data. Hvilken mekanisme som antas, er avgjørende for valg av analysemetode. De etablerte engelske termene for mekanismer for manglende data er vist i tabell 1. Det engelske uttrykket *missing at random*, forkortet MAR, kan misforstås (1, 2). Under antakelsen om *missing at random* kan sannsynligheten for manglende data være avhengig av observerte data. Dette kalles betinget sannsynlighet. En bedre betegnelse på engelsk kunne vært *missing conditionally at random*, men det ville fått samme forkortelse som *missing completely at random* – MCAR. Dessuten er termen *missing at random* godt etablert i engelsk faglitteratur (2). Andre betegnelser som er noe brukt, er *non-ignorable missing* og *ignorable missing* (2).

Tabell 1

Mekanismer for manglende data: I hvilken grad er sannsynligheten for manglende data avhengig av observerte eller uobserverte (dvs. manglende) data?

Engelsk term	Norsk term	Beskrivelse
<i>Missing completely at random (MCAR)</i>	<i>Mangler helt tilfeldig</i>	Sannsynligheten for manglende data avhenger verken av observerte eller uobserverte data
<i>Missing at random (MAR)</i>	<i>Mangler betinget tilfeldig</i>	Sannsynligheten for manglende data avhenger bare av observerte data
<i>Missing not at random (MNAR)</i>	<i>Mangler ikke-tilfeldig</i>	Sannsynligheten for manglende data avhenger av uobserverte data

Hvordan skal vi oversette disse termene til norsk? Jeg har ikke sett denne problemstillingen

omtalt i norske lærebøker, men så blir også metoder for å behandle manglende data sjelden omtalt i introduksjonslærebøker i statistikk. Jeg spurte noen norske statistikere, bl.a. Jan F. Bjørnstad i Statistisk sentralbyrå, som opplyser at han bruker følgende betegnelser i utvalgsundersøkelser: *tilfeldig frafall*, *stratifisert tilfeldig frafall* og *informativt frafall*. Men ellers ser det ikke ut til å være etablert noen norske oversettelser.

Nettsiden til International Statistical Institute (ISI) har en ordliste over statistiske termer på en rekke språk (3). Termene *missing completely at random* og *missing at random* inngår i denne ordlisten, men er ikke oversatt til norsk. De er oversatt til dansk som henholdsvis *fuldstændigt tilfældigt manglende* og *tilfældigt manglende*. Det danske *tilfældigt manglende* er like upresist som det engelske uttrykket. Oversettelsene til fransk, italiensk, nederlandsk, spansk, svensk og tysk har samme svakhet. I virkeligheten snakker vi om at data mangler tilfeldig *betinget* av data som er observert.

Som norske oversettelser foreslår jeg derfor *mangler betinget tilfeldig*, *mangler helt tilfeldig* og *mangler ikke-tilfeldig*. Norske forkortelser er neppe nødvendig ettersom de engelske er godt innarbeidet i statistisk faglitteratur og inngår i ISIs ordliste på flere språk.

LITTERATUR:

1. Lydersen S. Manglende data – sjelden helt tilfeldig. Tidsskr Nor Legeforen 2019; 139. doi: 10.4045/tidsskr.18.0809. [CrossRef]
2. Bjørnstad JF, Lydersen S. Missing data. I: Veierød M, Lydersen S, Laake P, red. Medical statistics in clinical and epidemiological research. Oslo: Gyldendal Akademisk, 2012: 429-61.
3. International Statistical Institute. The ISI glossary of statistical terms in a number of languages some of which use special characters. <http://isi.cbs.nl/glossary/> (29.11.2018).

Publisert: 8. februar 2019. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.18.0858
© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no