

En enkel rapport om utfyllingsgrad av skjema i NNRR

Kevin Thon

29 mai 2017

Bakgrunn/innledning

Det er ingen valideringsregler eller logikk i MRS-løsningen som forhindrer at skjema kan leveres med ingen eller minimalt av feltene utfylt. Dette skal endres i kommende versjoner av registeret, i første omgang for behandlerskjemaet 1b. For nå er det interessant å gjøre opp status for NNRR sine registreringer for 2016.

Denne rapporten tar for seg datakvaliteten i NNRR. Den fokuserer på utfyllingsgrad og ser på følgende problemstillinger:

- Antall tomme skjemaer
- Andel besvart av spørsmålene
- Sammenheng mellom tomme pasientskjema og behandlerskjema

Metode

NNRR består av tre skjemaer:

- Pasientskjema før behandling - 1a
- Behandlerskjema - 1b
- Pasientskjema etter behandling - 2

Skjemaene har felter som i MRS-løsningen er delt inn i følgende datatyper:

- Dato/Tid
- Enum
- Numerisk (heltall)
- Numerisk (flyttall)
- Text
- Avkrysning
- Id (Guid)

De tre NNRR-skjemaene har (ifølge kodebok hentet fra MRS-løsningen) følgende fordeling av datatyper.

Table 1: Antallet felter på NNRR sine skjema etter felttype

	Avkrysning	Dato/Tid	Enum	Id (Guid)	Numerisk (flyttall)	Numerisk (heltall)	Text
Skjema 1a	41	4	91	1	4	27	4
Skjema 2	2	3	44	1	3	8	1
Skjema 1b	63	4	14	0	0	3	8

Denne rapporten vil fokusere på tre datatyper, nemlig Avkrysning, Enum og Numerisk (heltall) siden disse utgjør mesteparten av feltene. Her følger en kort forklaring til datatypene og hvorfor jeg har valgt å telle på den måten jeg gjør.

Avkrysning

Dette er boolske variabler (True/False) med default-verdi “False”. Avkrysningsboksene er typisk organisert under overskrifter hvor én eller flere kryss forventes satt. Det finnes derfor ingen fasit på hvor mange bokser som skal være avkrysset ved en komplett utfylling. En mer grundig gjennomgang enn hva det er tid til her kan telle hver overskrift som har minst en avkrysning under seg, samt ta høyde for at en del felter bare skal fylles ut når man har krysset av for et annet. For nå vil det presenteres en fordeling av antall utfylte variabler.

Enum

Dette er kategoriske variabler hvor ett alternativ velges fra en liste. I de fleste tilfeller er manglende verdier kodet som -1 men i noen tilfeller kodes de med 0. For å komplisere bildet ytterligere er ikke verdiene i datadump konsistent med verdiene i kodebok, og ofte gis etikettene fremfor kodingen i datadumpen. Det er ikke mulig å automatisere prosessen med å finne de ikke utfylte variablene i disse tilfellene.

I presentasjon av utfyllingsgrad brukes kun de Enum variablene som er kodet numerisk i datadumpen, hvilket gjør at en del variabler faller bort fra vurderingen. I tillegg må det presiseres at ikke alle registrerte nødvendigvis skal svare på alle spørsmål. F.eks. finnes det nakkespesifikke verktøy som ikke skal benyttes av de med utelukkende ryggsmarter.

Numerisk (heltall)

Disse variablene er som regel én av to typer:

- Velg et antall (f.eks. barn) eller tall på en skala (f.eks. smerter fra 0-10).
- Koding for et svaralternativ som skal inngå i beregningen av en score.

Skjema 1b

Tomme skjema

Skjema 1b er behandlerskjemaet og består i stor grad av avkrysningsbokser. Med hele 63 avkrysningsbokser antar jeg at skjema som ikke har et eneste kryss er tomme også på andre felter. Avdelingsvis fordeler de helt tomme skjema 1b seg slik:

	Tomme	Totalt
Haukeland	195	875
St. Olavs	14	835
UNN	13	350

Mangelfullt utfylte skjema

Vi ser nå på utfyllingsgraden totalt for Avkrysninger. Første rad angir antallet avkrysninger mens andre rad gir antallet skjema med gitt antall avkrysninger. Konsistent med tabellen over er det 222 skjema med ingen avkrysninger, men det er verd å merke seg at det finnes alle mulige mellomnivåer av utfylling. Det er grunn til å være skeptisk til datakvaliteten når det kan virke noe vilkårlig hvor mye av skjemaene fylles ut. Det vil formodentlig bedre seg når MRS-løsningen gis et sett med regler og valideringer som gjør at skjema ikke kan lagres uten at felter er utfylt.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
222	3	16	54	180	346	403	364	298	241	199	178	102	103	58	29	11	2

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----

For skjema 1b finnes det 14 variabler av type Enum. Av disse er det bare syv som er kodet numerisk og dermed mulig å automatisk beregne utfyllingsgrad. Denne er oppsummert i tabellen under, hvor første rad angir antall utfylte felter og andre rad angir hvor mange skjema som har gitt antall utfylte Enum-felter.

0	1	2	3	4	5	6
261	149	630	1018	505	203	43

På skjema 1b er det bare tre numeriske variabler (“WorkingPercent”, “SickLeave_percent”, “Pension_percent”) og disse skal kun fylles ut ved avkrysnig av gitte felter. Det gir derfor ikke mening å se på utfyllingsgraden i denne sammenhengen.

Skjema 1a

Tomme skjema

Skjema 1a består for det meste av variabler av type Enum, så vi begynner med å se på utfyllingsgraden av disse feltene. For skjema 1a er alle feltene av Enum-type kodet numerisk hvilket gjør jobben enklere. Skjemaet har tabellen under oppsummerer