

23. august, 2019

# Tidsutvikling av utskrevet medisin i Norge

FARM3120 – Legemiddelbruk og folkehelse

Jon André Ottesen

## Introduksjon

Denne teksten har som hensikt å lede deg gjennom oppgaven og programmet *farm\_files.py* som skal anvendes i datalabben. Teksten er inndelt i to deler: den første delen omhandler litt generell teori samt informasjon om antiepileptisk medisin som kan knyttes inn i rapporten, og den andre delen fungerer som en dokumentasjon av programmet og dens anvendelse.

# 1 Antiepileptika, teori og programmet

## Introduksjon

Programmet som skal anvendes i denne oppgaven ble skrevet med hensikt på antiepileptisk medisin og bruken av dette i befolkningen. Det er derfor meget naturlig at dette står sentralt i oppgaven, men dette betyr **ikke** at den skal eksklusivt omhandle antiepileptika. Programmet kan anvendes på nesten alle dokumenter fra medisin registeret fra reseptregisteret men mer om dette finnes i seksjon [2](#).

## Generell teori

Hopp over

## Antiepileptika

Fertile og gravide kvinner gjennom svangerskapet utgjør populasjonen som undersøkes. En rekke ulike kvalitetsindikatorer benyttes for å se på legemiddelbruk ved ulike indikasjoner og gjennom tidsperioden 2004-2018. Hovedvekt ligger på legemidlet valproat ved epilepsi og bipolar lidelse, som har fått begrensninger for bruk hos fertile kvinner. Basert på aggregerende data vil tidstrender av valproat forskrivning hos fertile kvinner kunne visualiseres.

## 2 Oppstart

### Oppsett og de nødvendige datasett

Før vi starter å faktisk bruke programmet for å visualisere medikamentbruken i Norge de siste årene kreves det noen steg for at programmet skal fungere.

Disse er:

1. Datasett fra reseptregisteret<sup>1</sup> om medikamentene som skal undersøkes.

Når det gjelder det første punktet krever programmet at alle filene fra reseptregisteret **KUN GJELDER FOR ETT LEGEMIDDEL**. Dette betyr at det nødvendig å ha separate filer for f.eks valproat, lamotrigin, antiepileptika (den overordnende kategorien antiepileptika) osv. Når legemiddelet er valgt må **alt** merkes i kategoriene *Årstall* og *Alder* **unntatt** "Alle aldre". De resterende variabel gruppene *Kjønn* og *Bosted* kan man selv velge hva som skal merkes og være med videre. En bemerkning er at flere merkinger spesielt på kjønn gjør nettsiden lite responsiv. Jeg anbefaler derfor å velge "Mann" og "Kvinne" og alle stedene under *Bosted* for å unngå mye av dette problemet.

Den siste kategorien er *Måltall*. Som med valget av legemiddel godtar programmet kun at en av valgene er markert. Hovedsaklig er det tenkt at "Antall brukere" skal brukes, men hvis det er ønsket å bruke andre typer måletall er det nødvendig å lagre ett nytt datasett der denne variabelen er valgt. Programmet er i all hensikt skrevet med tanke på "Antall brukere", og det kan dermed være små feil i teksten visst på plottene men tallene er riktig.

Ett flowchart av prosessen av nedlasting av fra reseptregisteret som kompatible med programmet er gitt i figur 2.

2. Befolkningsdata fra SSB<sup>2</sup> eller reseptregisteret, jeg anbefaler reseptregisteret.

Angående punkt to om befolkningsdata anbefaler jeg på det sterkeste å bruke de dataene fra reseptregisteret under *Måltall* med navn "Befolkningsgrunnlag" istedenfor fra SSB. Ved nedlasting fra reseptregisteret velges de parameterne som datasettet/datasettene du har nedlastet tidligere bruker. Det er her **NØDVENDIG** å spesifisere et legemiddel men hvilket har ingen betydning<sup>3</sup>

---

<sup>1</sup><http://www.reseptregisteret.no/>

<sup>2</sup><https://www.ssb.no/statbank/table/07459/>

<sup>3</sup>Grunnen til dette er at programmet leser av spesifikke kolonner i csv-filen og uten et spesifisert legemiddel blir kolonnekombinasjonen forskjøvet en kolonne til venstre. Dette gjør at avlesningen ikke fungerer... Min feil ops.

**En veldig viktig kommentar er at datasettet om befolkningen MÅ lagres med navnet 'Befolkning'.**

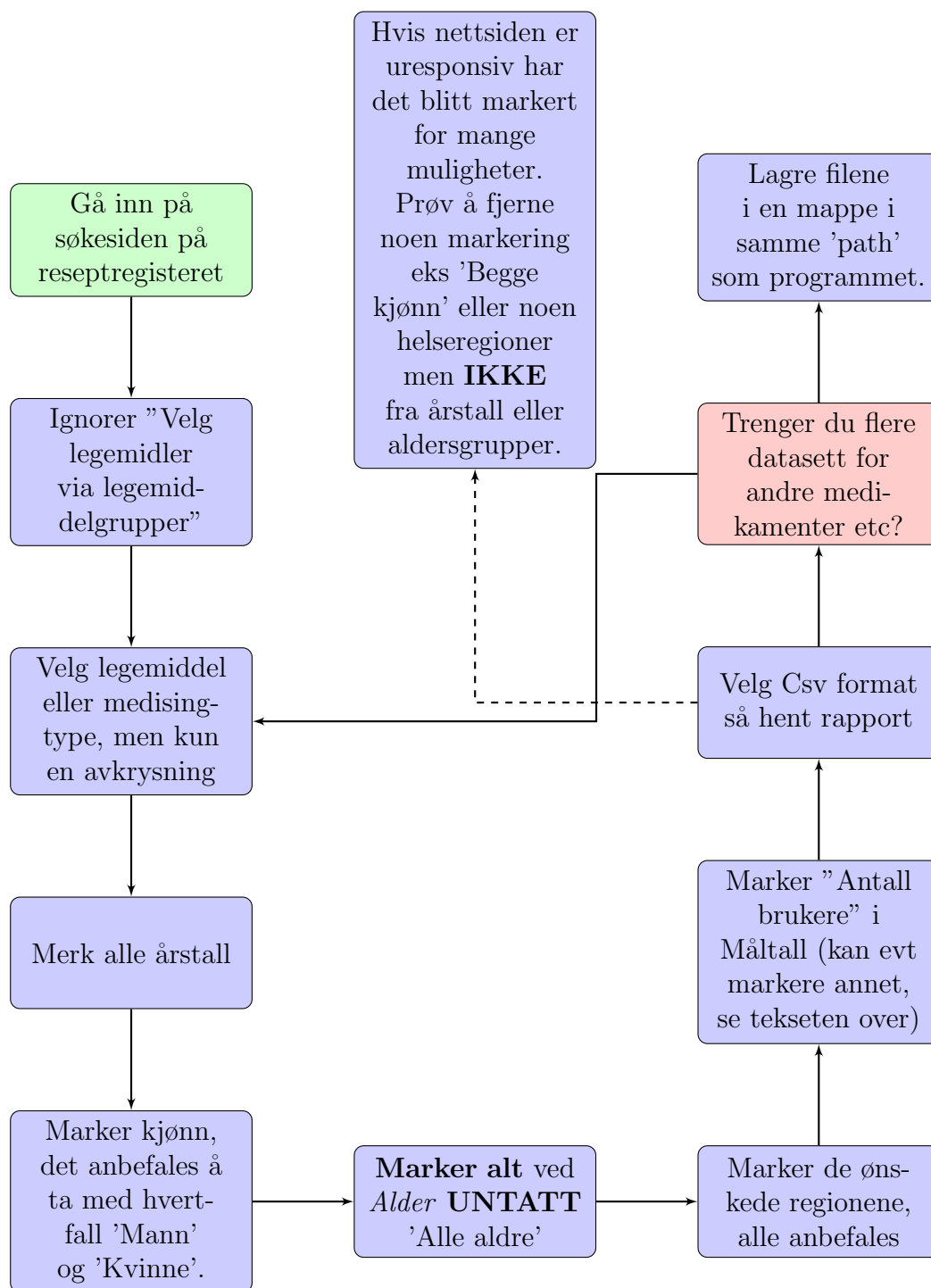
3. En meget spesifikk lagringsmetode og navnsetting for datasettene fra punkt (1) og (2).

Den siste delen omhandler lagring av datasettene. For at programmet skal kunne lese datasettene er det nødvendig å lagre disse i en eller flere mapper samme sted som programmet er lagret.

For bedre å forstå hvordan datasettene skal lagres skal jeg forklare tankegangen bak lagringen. Som nevnt over må alle datasettene lagres i mapper samme sted som programmet. Poenget med å lagre datasettene i mapper er for å kunne klassifisere legemiddelet etter legemiddeltypen. Tankegangen er nemlig slik at man laster ned ett eller flere datasett om f.eks flere legemidler under klassifikasjonen antiepileptika som valproat og lamotrigin. Denne mappen navngis så etter legemiddeltypen som her ville vært antiepileptika. Videre ville man også lagt inn et datasett om legemiddeltypen med samme navn som mappen. Dette gjør at programmet ser hva som er data for legemiddeltypen og hva som er data for legemiddelet. Endepunktet ville vært mappenavn; Antiepileptika, filnavn; Antiepileptika, Valproat, Lamotrigin osv. Videre skal man da kunne lage en ny mappe om f.eks diabetes eller astma og bruke de som separate tilfeller.

Det er ikke nødvendig med ett datasett navngitt etter legemiddeltypen men uten er det flere funksjoner som ikke vil fungere. Disse er spesifisert nærmere nede i seksjon 3 om anvendelse. Dette betyr at du kan ha en mappe med forskjellige datasett hvor legemiddelene ikke har noe medisinsk korrelasjon, men det er interessant å utregne forskjellige forhold mellom disse eller man kun trenger enkelt plot og grafer.

**En siste kommentar er at navnet datasettene blir lagret med er navnet du vises med i plottene. Jeg anbefaler dermed på det sterkeste å navngi filene noe meningsfullt som eks navnet på legemiddelet.**



Figur 1: Et flowchart om nedlasting av filer fra reseptregisteret.

### 3 Anvendelse

Før vi kan stare å anvende programmet er det først nødvendig å åpne Jupyter Notebook. Dette gjøres ved å først åpne kommandovinduet for så å navigere seg frem til mappen med programmet *farm\_files.py*. Dette gjøres med bruk av kommandoen *cd* (change directory) og med 'the path of the folder'<sup>4</sup>/miljøet der filen er lagret. Når kommandovinduet er navigert seg til filmappen med programmet skriver man inn 'Jupyter Notebook' og den åpnes. I jupyter notebook lager man så en ny notebook under *new* → *Python 3*.

Nå som alt er klart kan vi endelig starte å bruke programmet. Det første som **MÅ** gjøres er å importere de nødvendige pakkene og programmet inn i notebook.

```
1 from farm_files import visualization as visualization #Importere programmet for↵
bruk i jupyter
2 from IPython.display import HTML #En pakke for å se videoer i jupyter
```

Det neste steget er så å initialisere programmet for de forskjellige mappene med datasett.

```
1 antiep = visualization('Antiepileptika') #Lage et tilfelle knyttet til mappen ↵
Antiepileptika.
2 respir = visualization('R') #Ett tilfelle knyttet til mappen om R fra ↵
kategorien 'Respirasjonsorganer' på reseptregisteret.
```

Datasettene tilknyttet til R er **KUN** ment for å illustrere hvordan programmet fungerer.

Nå er programmet ferdig initialisert og resten av denne seksjonen vil omhandle kjøreeksampler av programmet. Før dette vil jeg rette oppmerksomheten din til tabell ... hvor alle metodene i programmet er listet opp samt argumentene de godtar. Videre i tabell .... er alle variablene listet opp med hvilke input som godtas.

Vi starter med etter min mening den metoden *tabell* som bør blir kalt først.

```
1 antiep.tabell() #Gir en tabell av variablene for kjønn, legemidler og steder i ↵
for Antiepileptika.
2 respir.tabell() #Gir en tabell av variablene for kjønn, legemidler og steder i ↵
for R.
```

Denne er nyttig spesielt siden man kan kopiere variablene og se hvilke variabler som er tilgjengelig.

Den neste metoden er *medisinforbruk* som gir et søylediagram av måletallene (of-

---

<sup>4</sup>I mangel på en bra norsk oversettelse.

test 'Antall brukere') for alle datasettene i hele filmappen. Den kan også gi forholdet legemiddel/legemiddeltipe for alle filene i datasettet (unntatt datasettet for legemiddeltypen).

```
1 antiep.medisinforbruk(gender = 'Mann', region = 'Helseregion Midt-Norge', ←  
  age_start = 9, age_end= 20, period_start = 1955, period_end = 2019) #Et ←  
  eksempel der antallet brukere plottes.  
2 respir.medisinforbruk(gender = 'Kvinne', region = 'Hele landet', age_start = 15, ←  
  age_end= 49, period_start = 2004, period_end = 2018, ratio = True) #Et ←  
  eksempel med utregning av forholdet mellom legemiddeltypen/legemiddelet.
```

Den neste metoden er `medisinforbruk.i` og er tilnærmet lik den forrige, men her plottes kun en spesifisert legemidler.

```
1 antiep.medisinforbruk(drug = 'Valproat', gender = 'Kvinne', region = 'Østfold', ←  
  age_start = 25, age_end= 92, period_start = 2009, period_end = 2015, ratio = ←  
  True)  
2 respir.medisinforbruk(drug = ['R01', 'R02'], gender = 'Kvinne', region = 'Hele ←  
  landet', age_start = 15, age_end= 49, period_start = 2004, period_end = ←  
  2018) #Legg merke til at når flere alternativer velger må de lagres i en ←  
  liste.
```

Den neste metoden er `forhold_medisin` som plotter forholdet mellom valgte legemidler. Legg merke til at jeg ikke har spesifisert kjønn, alder, region eller årstall. Programmet bruker istedenfor standardverdiene som er spesifisert i tabell ...

```
1 antiep.forhold_medisin(teller = ['Lamotrigin', 'Levetiracetam', 'Oxkarbazepin'], ←  
  nevner = 'Valproat') #Her brukes alle standardverdiene  
2 respir.forhold_medisin(teller = ['R01', 'R02'], nevner = ['R03', 'R05'], gender ←  
  = 'Mann') #Her endres kun variabelen 'gender' resten av parameterne er gitt ←  
  av standardverdiene.
```

Den neste metoden er `kake_medisinforbruk` som plotter et sektordiagram av alle datasettene i filmappen. I denne metode må kun ett årstall spesifiseres i motsetning til de andre metodene.

```
1 antiep.kake_medisinforbruk(year = 2015, region = 'Oslo') #Sektordiagram for ←  
  2015 i Oslo, resten av verdiene er standard  
2 respir.kake_medisinforbruk(year = 2020, save_fig = 'eksempel') #Sektordiagram ←  
  for 2020, resten er standardverdier og plottet lagres som en png fil med ←  
  navnet eksempel.png
```

Den neste metoden er `medisinforbruk.tidsutvilgning` som enten viser en video eller gir ett plott med datapunkter. Ett par bemerkninger angående videovisningen: for at animasjonen skal fungere kan kun koden tilhørende animasjonen være i en

cellen, i cellen til animasjonen vil det også være ett plott som bare må ignoreres<sup>5</sup>.

```
1 HTML(antiep.medisinforbruk_tidsutviling(drug = 'Valproat')) #Merk at for at ←  
    videoen skal vises i jupyter krves det bruk av pakken H1ML som ble importert←  
    i starten av jupyter.
```

```
1 respir.medisinforbruk_tidsutviling(drug = 'R01', year = 2018, ratio = True) #←  
    For stillbilder må argumentet 'year' brukes, ved stillbilder skal ikke H1ML ←  
    brukes.
```

Den nest siste metoden er [medisiner\\_og\\_befolkning](#) og her sammenlignes medisinbruken med antall sykdomstilfeller basert på befolkningstallene og prevalens i prosent.

```
1 antiep.medisiner_og_befolkning(prevalens = 2.5, sykdom = 'Epilepsi', age_start = ←  
    25) #Merk at prevalensen gis i prosent.  
2 respir.medisiner_og_befolkning(prevalens = [2.5, 1.5], save_fig = 'eksempel2') ←  
    #Hvis legemiddelet gjelder for 2 eller flere sykdommer gis prevalensen som ←  
    en liste av procenter.
```

Den siste metoden er [medisiner\\_og\\_befolkning2](#)<sup>6</sup> og er nesten helt lik den forrige metoden men her på ett mer individuelt nivå.

```
1 antiep.medisiner_og_befolkning2(prevalens = 2.5, drug = ['Valproat', 'Lamotrigin←  
    '], ratio = True, label = 'Noe greier') #Forholdet som nå vises er brukere ←  
    av det spesifiserte legemiddelet over antall med en sykdom basert på ←  
    prevalensen og befolkningsdataene.  
2 respir.medisiner_og_befolkning2(prevalens = [2.5, 1.5], drug = 'R05', gender = '←  
    Mann')
```

---

<sup>5</sup>Jeg fant ikke ut hvordan man fjerner den.

<sup>6</sup>Jeg klarte helt ærlig ikke å finne på et bedre navn.



### 3.1 Metoder, argumenter og beskrivelse

Metode	Argumenter	Beskrivelse
<code>visualization</code>	<code>folder_name</code> <code>plot_type</code>	
<code>tabell</code>	————	En tabell som printer alle variablene for <code>gender</code> , <code>region</code> og <code>drug</code> samt hvilke årstall som det finnes data for.
<code>medisinforbruk</code>	<code>gender</code> <code>region</code> <code>age_start</code> <code>age_end</code> <code>period_start</code> <code>period_end</code> <code>ratio</code> <code>save_fig</code> <code>label</code>	Denne metoden kan gi to mulige plott. Enten ett plott av alle dataene for hver enkelt datasett i filmappen, eller ett plott av forholdet mellom dataene for legemidlene over dataene for legemiddeltypen. Dvs måldataene for hver enkelt fil delt på måldataene til filen med navnet til filmappen på y-aksen. På x-aksen plottes det året dataene gjelder.
<code>medisinforbruk_i</code>	<code>drug</code> <code>gender</code> <code>region</code> <code>age_start</code> <code>age_end</code> <code>period_start</code> <code>period_end</code> <code>ratio</code> <code>save_fig</code> <code>label</code>	Plotter dataene for enten en eller flere valgte legemidler eller forholdet mellom dataene for legemiddelet/legemiddelene over dataene for filen for legemiddeltypen(filen med navn likt som filmappen) på y-aksen og år på x-aksen.
<code>forhold_medisin</code>	<code>teller</code> <code>nevner</code> <code>gender</code> <code>region</code> <code>age_start</code> <code>age_end</code> <code>period_start</code> <code>period_end</code> <code>ratio</code> <code>save_fig</code> <code>label</code>	Plotter forholdet mellom summen av legemidlene spesifisert i telleren over summen av legemidlene spesifisert i nevneren på y-aksen og året datene gjelder på x-aksen.

Metode	Argumenter	Beskrivelse
<code>kake_medisinforbruk</code>	<pre> gender region age_start age_end year save_fig </pre>	Plotter et sektordiagram basert på dataene fra alle datasettene for et spesifisert år. Hvis filen med samme navn som filmappen eksisterer og hvis dataen for denne filen er numrisk større enn for summen av de andre anses dette tallet for 100% i sektordiagrammet. De resterende dataene er da X% av denne mengden. Hvis tallet er numrisk mindre eller filen eksisterer anses summen fra de andre filene som 100%.
<code>medisinforbruk_tidsutviling</code>	<pre> drug gender region ratio year save_fig </pre>	Hovedsaklig ment som en animasjon av utvikling av måledataene for et spesifisert legemiddel gjennom de årene med data. Her plottes måledataene på y-aksen og aldergruppen på x-aksen. Animasjonen kan også vise animere forholdet mellom legemiddelet og filmappe-filen (legemiddeltypen). Kan også gi ett stillplott for et spesifisert år. Hvis metoden brukes i IPython(jupyter notebook) må kommandoen <code>HTML</code> fra IPython anvendes.
<code>medisiner_og_befolkning</code>	<pre> prevalens sykdom gender region age_start age_end period.start period.end ratio save_fig label </pre>	Plotter antallet personer med sykdom X basert på prevalensen gitt i prosent med dataene fra alle datasettene i filmappen. Kan også plotte forholdet mellom dataene og antall personer med sykdom X på y-aksen. På x-aksen vises korresponderende år.
<code>medisiner_og_befolkning2</code>	<pre> prevalens drug gender region age_start age_end period.start period.end ratio save_fig label </pre>	Plotter antall personer med sykdom X basert på prevalensen gitt i prosent med dataene fra de spesifiserte datasettene. Kan også plotte forholdet mellom de spesifiserte datasettene og antall personer med sykdom X. På x-aksen vises korresponderende år for dataene.

Argument	Input	Beskrivelse
folder_name		
plot_type		
gender		
region		
age_start		
age_end		
period.start		
period.end		
ratio		
save_fig		
label		
drug		
teller		
nevner		
prevalens		
sykdom		