Overlevelsestavler

Anna-Vera Jørring Pallesen, Johan Sebastian Ohlendorff, Laust Hvas Mortensen and Thomas Alexander Gerds

1 Introduktion

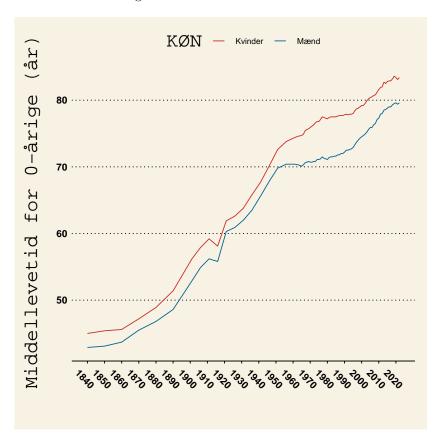
Overlevelsestavlen repræsenterer en matematisk model, der kvantificeres ved hjælp af konkrete demografiske data. Modellen genererer derefter en omfattende beskrivelse af dødelighedsforholdene i den specifikke befolkning. De forskellige mål for dødelighed konstrueres på baggrund af overlevelsestavlen. Dette gælder også for målene vedrørende forekomsten af vielser, skilsmisser, folkevandringer og i vis grad forskellige fertilitets- og reproduktionsmål. Overlevelsestavlen kunne derfor gennemgås på en ret abstrakt måde og fortolkes forskelligt, afhængigt af om den skal anvendes til at beskrive dødelighed, vielser, skilsmisser eller fertilitet. I det følgende vil vi dog fokusere på at opbygge modellen omkring målingen af befolkningens dødelighed for at gøre det lettere at forstå modellens umiddelbare anvendelighed.

1.1 Middellevetid

Hvor mange år kan en nyfødt i dag forvente at leve? Dette spørgsmål er umuligt at besvare korrekt, fordi svaret umiddelbart afhænger af, hvad der sker i fremtiden. Alligevel er middellevetid, altså den forventede gennemsnitlige levetid af en nyfødt, et demografisk værktøj som anvendes hyppigt til belysning af befolkningens nuværende dødelighedsniveau. Middellevetid bruges også som sammenligningsgrundlag på tværs af befolkninger og tid. Tallet angiver det gennemsnitlige antal år, som en nyfødt kan forvente at leve under den forudsætning, at de nuværende mortalitetsrater for alle grupperinger af køn og alderstrin holder sig på samme niveau i fremtiden. Med middellevetiden har man et relativt simpelt begreb, som gør det muligt at sammenligne forskellige befolkningers dødelighed. I praksis vil de nuværende dødshyppigheder formentlig ikke holde sig på et samme niveau i fremtiden.

Igennem mange år har der været en tendens til faldende mortalitetsrater, og der er meget, som tyder på, at det er en udvikling som fortsætter. Den konkrete fortolkning af middellevetiden for 0-årige som det gennemsnitlige antal år, som en nyfødt kan forventes at leve, vil derfor formentlig undervurdere den faktiske middellevetid. Men formålet med middellevetiden er heller ikke at forudsige præcist, hvor længe en nyfødt vil leve. Formålet er at have et simpelt begreb,

der kan sammenlignes på tværs af befolkninger og tid. Figur 1 viser udviklingen af middellevetid for 0-årige i Danmark siden 1840.



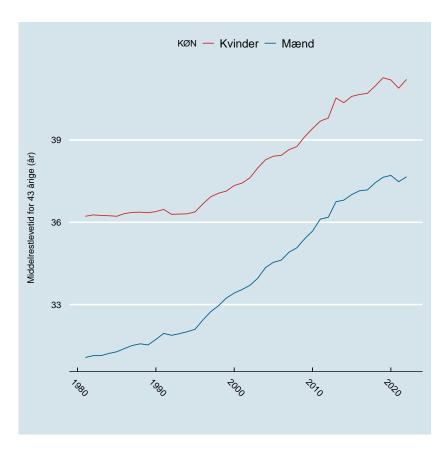
Figur 1: Udviklingen i middellevetid for 0-årige. Kilde: statistikbankens HISB7.

1.2 Andre dødelighedsmål

Middellevetiden er måske det vigtigste mål, som resulterer af overlevelsestavlen. En overlevelsestavle beskriver også en række andre dødelighedsmål, såsom den forventede restlevetid fra alder x, sandsynligheden for at dø inden alder x og sandsynligheden for at være i live ved alder x. Figur 1 viser middelrestlevetiden for 43-årige i Danmark siden 1981.

1.3 Eksempel

Vi henter data fra statistikbankens register FOLK1a og FOD207 og beregner aldersspecifikke mortalitetsrater for kvinder i Danmark i 2019. Vi inddeler i 12 aldersintervaller, hvor det første interval har længde 1 år, det andet interval har



Figur 2: Udviklingen i middelrestlevetid for 43-årige. Kilde: statistikbankens HISB8.

længde 9 år, resten af intervallerne har længde 10 år og det sidste aldersinterval er fra 90 til 125 år.

```
# A tibble: 11 \times 6
   aldersinterval KØN
                             TID
                                       R
                                           Dod
                                                         М
                                   <dbl> <dbl>
   <fct>
                   <chr>
                            <dbl>
                                                    <dbl>
1 0
                   Kvinder
                            2019
                                   29448
                                             74 0.00251
 2 1-9
                                             24 0.0000889
                   Kvinder
                            2019 270111
 3 10-19
                   Kvinder
                            2019 332202
                                             32 0.0000963
 4 20-29
                   Kvinder
                            2019 383578
                                             73 0.000190
5 30-39
                   Kvinder
                            2019 336414
                                            128 0.000380
 6 40-49
                   Kvinder
                            2019 378914
                                           342 0.000903
7 50-59
                            2019 397594
                                          1160 0.00292
                   Kvinder
8 60-69
                   Kvinder
                            2019 336747
                                          2855 0.00848
9 70-79
                   Kvinder
                            2019 293474
                                          6016 0.0205
10 80-89
                   Kvinder
                             2019 129929
                                          8878 0.0683
11 90+
                                  32094
                                          6921 0.216
                   Kvinder
                            2019
```

Med disse tal fra den rigtige befolkning konstruerer vi overlevelsestavlen, som beskriver dødeligheden i en hypotetisk befolkning, der bliver født i 2019 og lever hele deres "liv" igennem alle alderstrin i 2019, hvor de bliver udsat for mortalitetsraterne fra 2019.

```
# A tibble: 11 \times 9
                                                           Τ
   Alder
              1
                    d
                                                  L
                           p
                                    q
                                           0
          <dbl> <dbl> <dbl>
                                                       <dbl> <dbl>
                                <dbl> <dbl>
                                              <dbl>
1 0
         100000
                  251 0.997 0.00251
                                     1
                                              99774 8334430. 83.3
                   80 0.999 0.000799 0.997 897385 8234656. 82.6
 2 1-9
          99749
 3 10-19
          99670
                   96 0.999 0.000963 0.997 996216 7337271. 73.6
 4 20-29
                  189 0.998 0.00190
                                      0.996 994789 6341056. 63.7
          99574
 5 30-39
          99384
                  377 0.996 0.00380
                                      0.994 991955 5346266. 53.8
 6 40-49
          99007
                  890 0.991 0.00899
                                      0.990 985620 4354311. 44.0
7 50-59
         98117 2821 0.971 0.0288
                                      0.981 967065 3368691. 34.3
```

8	60-69	95296	7751	0.919	0.0813	0.953	914204	2401626.	25.2
9	70-79	87545	16278	0.814	0.186	0.875	794062	1487422.	17.0
10	80-89	71267	36296	0.491	0.509	0.713	531192	693360.	9.73
11	90+	34971	34971	0	1	0.350	162168	162168.	4.64

Tabel 1: Forklaring af kolonner i en overlevelsestavle

Kolonne	Betydning
Alder	Aldersinterval
1	Dekrementfunktion: Antal tabelpersoner i starten af intervallet
d	Antal døde i intervallet
р	Sandsynlighed for at overleve i intervallet
q	Dødshyppighed: sandsynlighed for at dø i intervallet
0	Sandsynlighed for at overleve indtil starten af intervallet
L	Samlet risikotid i intervallet
T	Samletlevetid fra starten af intervallet
е	Middelrestlevetid (i første interval = middellevetid)

Fra overlevelsestavlen aflæser vi af kolonne e: under antagelsen af, at mortalitetsraterne i 2019 ikke ændrer sig i al fremtid vil man forvente, at en nyfødt pige lever 88,3 år og at en kvinde som er 30 år gammel kan forvente at leve 53,8 år.

2 Konstruktion af overlevelsestavler

Overlevelsestavler beskriver, hvordan en tænkt lukket fødselskohorte reduceres med stigende alder alene på grund af dødsfald. Fordi kohorten er lukket, er død den eneste mulige afgang fra kohorten. Der tages udgangspunkt i en fiktiv tabelbefolkning bestående af ℓ_0 personer, som antages at være født på nøjagtig samme tidspunkt. Antallet af fiktive tabelpersoner ℓ_0 kaldes for 'radix', og radix sættes typisk til $\ell_0 = 100.000$.

2.1 Dekrementfunktionen

Funktionen ℓ_x angiver hvor mange tabelpersoner stadigvæk er i live ved alder x og beskriver hvordan tabelbefolkningen reduceres på grund af dødsfald. Startværdien ℓ_0 angiver, hvor mange tabelpersoner, der er i tabelbefolkningen helt i begyndelsen, hvor alder er lig med 0, og ℓ_{30} angiver hvor mange tabelpersoner er i live ved alder 30. Fordi ℓ_x er monotont faldende som funktion af alder, det vil sige, at der gælder $\ell_x \geq \ell_{x+1}$, kalder man den for dekrementfunktionen. Af tabellen kan man aflæse, hvor mange personer forventes at overleve til en bestemt alder. For eksempel betyder $\ell_{30} = 99.345$, at 99.345 personer ud af $\ell_0 = 100.000$ tabelpersoner stadigvæk er i live ved alder 30. I dette eksempel er

overlevelsessandsynligheden i tabelbefolkningen ved alder 30 lig med

$$o(30) = \frac{\ell_{30}}{\ell_0} = \frac{99.345}{100.000} = 99,3\%,$$

eftersom overlevelsesfunktionen er defineret som

$$o_x = \frac{\ell_x}{\ell_0}.$$

Under konstruktionen af overlevelsestavler er opgaven at beregne dekrementfunktionens værdier ℓ_x for alle alderstrin $x=0,1,\ldots,x^{max}$ hvor x^{max} er det sidste alderstrin. Per konstruktion dør alle resterende tabelpersoner i det sidste alderstrin - det vil sige $\ell_{x^{max}+1}=0$ og dermed også $o_{x^{max}+1}=0$. Vi vil forklare hvorfor senere.

2.2 Dødshyppigheder

Dødshyppigheden $_kq_x$ beskriver for en person med eksakt alder x sandsynligheden for at dø inden alderen x+k. Dødshyppigheder forbinder den ægte, åbne befolkning, som man interesserer sig for, med den tænkte, lukkede tabelbefolkning, der definerer overlevelsestavlen. Man beregner dødshyppigheder baseret på aldersspecifikke mortalitetsrater, og den underliggende idé er, at mortalitetsraterne er ens i den ægte befolkning og i tabelbefolkningen for begge køn og alle alderstrin.

Bemærkning til notation:

Det er standardnotation i demografi at have index på begge sidder af symbolet ligesom i $_kq_x$. Her er index til højre starten af et aldersinterval og index til venstre er længden af aldersintervallet. Det er lidt forvirrende, fordi intervallet inkluderer startalder x:

Symbol	Start	Længden	Slut	Betydning
D_0	0	1	1	Antal døde i alder 0
$_4D_1$	1	4	4	Antal døde i alder 1, 2, 3, 4
$_5D_5$	5	5	9	Antal døde i alder 5, 6, 7, 8, 9

Vi ændrer nu også notationen for de aldersspecifikke mortalitetsrater. I Kapitel 2 har vi brugt M_x for mortalitetsraten i det x-te aldersinterval. Fra nu af bruger vi den mere præcise betegnelse $_kM_x$ for mortalitetsraten i det aldersinterval, som starter i alderen x og slutter i alderen x + k.

2.2.1 Approksimationsformlen

For at beregne dødssandsynligheder i den ægte befolkning vil man gerne dividere antal dødsfald i en kalenderperiode med antal personer i starten af perioden.

Problemet er, at den ægte befolkning er åben: Dødsfald bliver ikke registreret for personer som udvandrer i perioden, og både udvandrere og indvandrere i perioden bidrager ikke med risikotid til hele perioden. Ideen er derfor at tilnærme dødshyppighederne baseret på mortalitetsrater. Aldersspecifikke mortalitetsrater kan beregnes på de registrerede data, ved at dividere antal dødsfald i befolkningen med risikotiden, hvor indvandrere og udvandrere kun bidrager med den tid de har været i befolkningen (se Kapitel 1 og 2). Nøglen til en tilnærmelse af dødshyppighederne baseret på mortalitetsrater er følgende centrale formel for overlevelsestavlen:

$$_{k}q_{x} = \frac{k \cdot _{k}M_{x}}{1 + (k - _{k}a_{x}) \cdot _{k}M_{x}}$$
(K3.1)

Formlen afhænger aldersspecifikke mortalitetsrater $_kM_x$, længden af aldersintervallet k og også en konstant $_ka_x$, som kaldes Chiang's a. Konstanten $_ka_x$ beskriver den gennemsnitlige levetid i aldersintervallet for personer, der døde mellem alderen x og aldeen x+k. Dermed beskriver $(k-_ka_x)$ den gennemsnitlige tid som en person der døde i aldersintervallet var død. Hvis vi for eksempel ser på et aldersinterval mellem 70 og 79 år og en person døde i alder 74, så har den person været i live i 4 år (70, 71, 72, 73) og død i 6 år (74, 75, 76, 77, 78, 79). En person som døde i alder 78 har været i live i 8 år og død i 2 år, og så videre. Værdien af $_{10}a_{70}$ skal afspejle det gennemsnitlige antal år, som personer der døde i denne aldersgruppe, var i live. For de fleste intervaller vil man antage at gennemsnittet ligger i midten, altså i eksemplet vil man vælge $_{10}a_{70} = 5$.

2.2.2 Chiang's a

For at beregne dødshyppigheder med den centrale formel (K3.1) har vi brug for at specificere Chiang's a for alle aldersintervaller. Chiang's a skal tilnærme det forventede antal år levet i intervallet af en person, som dør i intervallet. Hvis Chiang's a opfylder dette, kan vi tilnærme den samlede dødstid, som alle personer der døde i aldersintervallet har været døde:

Samlede dødstid i aldersintervallet = $(k - {}_k a_x) \cdot {}_k D_x$, k = Antal år i aldersintervallet ${}_k D_x = \text{Antal døde i aldersintervallet}$ ${}_k a_x = \text{Gennemsnitlige antal dødsår i intervallet}$ $\{x, x+1, \dots, x+k\} = \text{År i intervallet}$.

Hvis vi antager at dødstider er lige fordelt i aldersintervallet, altså at det er lige sandsynligt at dø i starten som det er at dø i slutningen af aldersintervallet, er det rimeligt at vælge

$$_k a_x = \frac{k}{2}.$$

Det første og sidste aldersinterval vil dog altid kræve særlige værdier af $_ka_x$. I det første leveår er dødstiderne meget skævt fordelt over året - de fleste dødstider

inden 1-års fødselsdagen ligger kort efter fødslen. Derfor sætter vi $_1a_0=0,1.$ For det sidste interval x^{max} vælger vi

$$_{\infty}a_{x^{max}} = \frac{1}{_{\infty}M_{x^{max}}},\tag{K3.2}$$

så dødshyppigheden i det sidste interval bliver 1, og det betyder, at alle tabelpersoner dør i det sidste aldersinterval, dvs. $_{\infty}q_{x^{max}}=1$. ved formel (K3.1).

Tabel 2: Tabellen viser hvordan vi vælger Chiang's a for 1-års, 5-års og 10-års aldersintervaller.

	5-års aldersintervaller	10-års aldersintervaller
Første leveår	$_{1}a_{0}=0,1$	$_{1}a_{0}=0,1$
Aldersinterval 1-5 år		$_9a_1 = 9 \cdot 0, 5 = 4, 5$
Alle andre intervaller	$_k a_5 = 5 \cdot 0, 5 = 2,5$	$_k a_{10} = 10 \cdot 0, 5 = 5$
Sidste aldersinterval	$a_{x^{max}} = \frac{1}{\infty M_{x^{max}}}$	$a_{x^{max}} = \frac{1}{\infty M_{x^{max}}}$

2.2.3 Forklaring af den centrale formel

I det følgende skal vi på en uformel måde forklare formel (K3.1). Hvis den ægte befolkning var lukket, altså uden forekomst af ind- og udvandring, ville man kunne beregne dødshyppighederne simpelt som antal dødsfald i aldersintervallet divideret med antal personer i starten af aldersintervallet:

$$\label{eq:definition} D \emptyset ds hyppighed = \frac{Antal\ d \emptyset ds fald\ i\ alders intervallet}{Antal\ personer\ i\ starten}.$$

Hvis aldersintervallet er over k år gælder

$$\text{Antal personer i starten} = \frac{\text{Risikotid} + \text{D} \emptyset \text{dstid}}{k}.$$

Her er risikotiden det samlede antal år, som befolkningens personer har levet (i aldersintervallet), og dødstiden er tilsvarende det samlede antal år, som befolkningens personer var døde. Med denne formel kan dødshyppigheden skrives som

$$\label{eq:definition} D \emptyset d s hyppighed = \frac{k \cdot \text{Antal d} \emptyset d s f ald i alders interval}{\text{Risikotid} + D \emptyset d s t i d}. \tag{K3.3}$$

Vi sætter Chiang's a sådan at

Dødstid i aldersinterval =
$$(k - {}_k a_x) \cdot {}_k D_x$$
,

er en god tilnærmelse af den samlede dødstid, som alle personer der døde i aldersintervallet har været døde (c.f., afsnit 2.2.2). Hvis vi nu anvender formlen for den aldersspecifikke mortalitetsrate fra Kapitel 2,

$$_{k}M_{x}=rac{_{k}D_{x}}{_{k}R_{x}},$$

ser vi at den centrale formel (K3.1) faktisk er lig med formel (K3.3):

$$\frac{k \cdot {}_k M_x}{1 + (k - {}_k a_x) \cdot {}_k M_x} = \frac{k \cdot {}_k \frac{D_x}{kR_x}}{1 + (k - {}_k a_x) \cdot {}_k \frac{D_x}{kR_x}}$$
$$= \frac{k \cdot {}_k D_x}{kR_x \cdot (1 + (k - {}_k a_x) \cdot {}_k \frac{D_x}{kR_x})}$$
$$= \frac{k \cdot {}_k D_x}{kR_x + (k - {}_k a_x) \cdot {}_k D_x}.$$

2.2.4 Beregningen af antal dødsfald og overlevelser

Vi fortsætter nu konstruktionen af overlevelsestavlen. Vi starter med en radix af ℓ_0 tabelpersoner. For at beregne antal tabelpersoner som overlever indtil det første alderstrin, x=1, skal vi beregne hvor mange tabelpersoner dør mellem alder x=0 og alder x=1. For at beregne hvor mange tabelpersoner der overlever alder x+k skal vi beregne hvor mange af de resterende ℓ_x tabelpersoner der dør i aldersintervallet. Vi betegner med $_kd_x$ antal tabelpersoner som dør mellem alder x og alder x+k. Dermed er $_1d_x$ antal tabelpersoner som dør mellem alder x og alder x+1. Sandsynligheden for at dø mellem to alderstrin (dødshyppighederne) er det centrale element ved konstruktionen af overlevelsestavlen. Vi beregner antal dødsfald i aldersintervallet ved at gange antal tabelpersoner i starten af intervallet med dødshyppigheden:

$$_k d_x = _k q_x \cdot \ell_x. \tag{K3.4}$$

Det er vigtig at skelne mellem antal døde $_kD_x$ i den ægte befolkning og antal døde $_kd_x$ i tabelbefolkningen. Med formel (K3.4) er det en let sag at beregne, hvor mange tabelpersoner er i live i starten af det næste aldersinterval:

$$\ell_{x+k} = \ell_x - {}_k d_x.$$

Alternativt kan vi starte med at beregne dekrementfunktionen baseret på dødshyppigheden

$$\ell_{x+k} = \ell_x \cdot (1 - q_x).$$

Bagefter er det simpelt at beregne antal dødsfald som

$$_k d_x = l_x - l_{x+k}.$$

Med disse formler kan vi konstruere overlevelsestavlen vigtigste kolonner (ℓ_0 og $_k d_x$). Vi beskriver nu de vigtigste dødelighedsmål som overlevelsestavlen viser.

2.2.5 Beregning af middelrestlevetid og middellevetid

Vi betegner med $_kL_x$ den samlede gennemlevede tid i tabelbefolkningen i alderen mellem x og x + k. Da dødsfald er eneste afgangsårsag i tabelbefolkningen har

vi

$$_kL_x=$$
 bidrag fra overlevende + bidrag fra døde
$$=k\cdot\ell_{x+k}+_ka_x\cdot_kd_x$$

$$={}_ka_x\cdot\ell_x+(k-_ka_x)\cdot\ell_{x+k}.$$

Vi skal nu beregne den forventede restlevetid for en x-årig tabelperson. For en nyfødt er x=0 og dermed bliver den forventede middelrestlevetid til den forventede levetid, som betegnes med middellevetid. Lad T_x angive den samlede levetid i tabelbefolkningen efter x-års fødselsdagen, specielt er T_0 den samlede levetid i tabelbefolkningen. Vi beregner

$$T_x = {}_k L_x + \dots + {}_k L_{x^{max}}$$

= ${}_k a_x \cdot \ell_x + (k - {}_k a_x) \cdot \ell_{x+k} + \dots + {}_\infty a_{x^{max}} \cdot \ell_{x^{max}}.$

I tabelbefolkningen overlever ℓ_x personer til deres x-års fødselsdag, så den gennemsnitlige levetid efter x-års fødselsdagen bliver

$$e_x = \frac{T_x}{\ell_x} = \text{gennemsnitlige restlevetid.}$$
 (K3.5)

Denne kvotient kaldes den forventede restlevetid eller middelrestlevetid for en x-årig tabelperson. På tilsvarende vis bliver middellevetid beregnet som

$$e_0 = \frac{T_0}{\ell_0} = \text{middellevetid.}$$
 (K3.6)

2.2.6 Fortolkning

Når man fortolker middellevetid og middelrestlevetid er det vigtigt at huske og fremhæve at beregningen bygger på en hypotetisk tabelbefolkning som lever hele deres liv i en kort kalenderperiode. Danmarks Statistik forklarer middelrestlevetiden sådan¹:

Middelrestlevetiden er det gennemsnitlige antal år, som personer på en given fødselsdag har tilbage at leve i, hvis deres dødelighed fremover (alder for alder) svarer til det niveau, som er konstateret i den aktuelle periode.

2.3 Overlevelsestavle med 1-års intervaller

2.4 Danmark statistik

Med etableringen af den personstatistiske database har Danmarks Statistik fået nye muligheder for at beregne dødshyppighederne mere korrekt, idet databasen for alle personer i Danmark indeholder eksakt information om eventuel dødsdato

 $^{^{1} \}verb|https://www.dst.dk/da/Statistik/emner/borgere/befolkning/middellevetid|$

og ind- og udvandringsdatoer. Der kan således for hver enkelt person udregnes nøjagtigt, hvor mange dage personen i en årsperiode har været i Danmark og hvor mange af dagene i årsperioden, personen har været død. Den søgte dødshyppighed skal præcist angive sandsynligheden for at dø i et bestemt alderstrin, dvs. mellem to fødselsdage. For at opnå denne hyppighed laves der en særlig beregning for hver enkelt person fra fødselsdag til fødselsdag i en periode, der omfatter to kalenderår. I offentliggørelsen af middellevetid fra 19. marts 2010 er det kalenderårene 2008 og 2009, der ligger til grund for beregningerne. For alle personer, der var i den danske befolkning på et eller andet tidspunkt mellem deres fødselsdag i 2008 og i 2009, er der lavet en beregning for antallet af dage, personen var i Danmark og antallet af dage personen var død i perioden mellem de to fødselsdage. For personer, der ikke dør mellem to fødselsdage, vil antallet af dage som død naturligvis være 0. Efterfølgende laves der en sammenlægning for personer med samme køn og alderstrin for at få det samlede antal levedage og dødsdage. Personer vil placeres på det alderstrin, som svarer til det antal år, de fyldte i startåret, hvilket i eksemplet vil sige 2008. En person, som fyldte 60 år 1. januar 2008 vil f.eks. tilhøre de 60-årige. Det samme vil en person, der fyldte 60 år 31. december 2008. Der kan altså i yderste konsekvens være næsten et års forskel mellem den periode, som personer på samme alderstrin følges.

Man kan se de nyeste beregninger fra Danmark Statistik her: Middelevetid, og finde dokumentation for beregningerne her: Dokumentation.