

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, vandringer 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 ret umuligt at besvare fordi svaret umiddelbart afhænger hvad der sker i fremtiden. Alligevel er middellevetid 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 forventes at leve under den forudsætning, at de nuværende mortalitetsrater for alle grupperinger af køn og alderstrin holder sig på det samme niveau i al fremtid. 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 konstant niveau i al fremtid.

I gennem 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 nyfødte vil leve. Formålet er at have et simpelt begreb, der kan sammenlignes på tværs af befolkninger og tid.

1.2 Andre dødelighedsmål

Middellevetiden er et mål som resulterer fra den såkaldte overlevelsestavle. En klassisk 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 .

2 Konstruktion af overlevelsestavle

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 $\ell(0)$ personer, som antages at være født på nøjagtig samme tidspunkt. Antallet af fiktive tabelpersoner $\ell(0)$ kaldes for 'radix' og radix sættes typisk til $\ell(0) = 100.000$.

2.1 Dekrementfunktion

Funktionen ℓ angiver hvor mange tabelpersoner er stadigvæk i live ved alder x og beskriver hvordan tabelbefolkningen reduceres på grund af dødsfald. Startværdien $\ell(0)$ angiver hvor mange tabelpersoner der er i tabelbefolkningen helt i begyndelsen hvor alder er lige med 0, og $\ell(30)$ angiver hvor mange tabelpersoner er i livet ved alder 30. Fordi ℓ er monoton faldende som funktion af alder, altså $\ell(x) \geq \ell(x+1)$, kalder man den for dekrementfunktion. For eksempel, $\ell(30) = 91.047$ betyder at ud af $\ell(0) = 100.000$ tabelpersoner 91.047 personer er stadigvæk i live ved alder 30. I dette eksempel er overlevelsessandsynligheden i tabelbefolkningen ved alder 30 lige med

$$o(30) = \frac{\ell(30)}{\ell(0)} = \frac{91.047}{100.000} = 91,1\%.$$

Overlevelsesfunktionen er defineret som $o(x) = \frac{\ell(x)}{\ell(0)}$. Under konstruktionen af overlevelsestavler er opgaven at beregne dekrementfunktionens værdier $\ell(x)$ for alle alderstrin $x = 1, 2, 3, \dots, x_{max}$ hvor x_{max} er det sidste alderstrin. Per konstruktion, der bliver forklaret i detaljer nedenfor, dør alle resterende tabelpersoner i det sidste alderstrin, det vil sige $\ell(x_{max} + 1) = 0$, og dermed også $o(x_{max} + 1) = 0$.

2.2 Dødshyppigheder

Dødshyppigheder ${}_kq_x$ beskriver sandsynligheden for at dø i aldersintervallet $(x, x+k]$ given overlevelse indtil alder x . Dødshyppigheder forbinder den ægte og åbne befolkning, som man interesserer sig for, med den tænkte lukkede tabelbefolkning der definerer overlevelsestavlen. Den underliggende ide er at mortalitetsraterne er det samme i disse to befolkninger for begge køn og alle alderstrin.

2.2.1 Approksimationsformel

For at beregne dødshyppighederne ${}_kq_x$ bruger vi data fra den ægte befolkning, typisk fra en kort kalenderperiode. Problemet er at den ægte befolkning er *åben*. Dødsfald bliver ikke registreret for personer som udvandrer. Desuden er det uklart hvordan personer som indvandrer skal bidrage til antal personer i starten af aldersintervallet. Aldersspecifikke summariske mortalitetsrater løser problemet ved at dividere antal dødsfald med antal risikotid hvor indvandrer og udvandrer bidrager med den tid de nu har været i befolkningen. Nøglen til en approksimation af dødshyppigheder baseret på mortalitetsrater er følgende centrale formel:

$${}_kq_x = \frac{k \cdot {}_kM_x}{1 + (k - {}_ka_x) \cdot {}_kM_x} \quad (1)$$

Formlen afhænger aldersspecifikke mortalitetsrater, længden af aldersintervallet k , og også en konstant ${}_ka_x$, som kaldes Chiang's a . Konstanten ${}_ka_x$ beskriver det gennemsnitlige levetid i aldersintervallet $(x, x + k]$ for personer, der døde mellem alder x og alder $x + k$. Dermed beskriver $k - {}_ka_x$ den gennemsnitlige tid som en person der døde i aldersintervallet var død.

2.2.2 Forklaring

I det følgende skal vi på en uformelt måde forklare formelen (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:

$$\text{Dødshyppighed} = \frac{\text{Antal dødsfald i aldersinterval}}{\text{Antal personer i starten}}.$$

Fordi aldersintervallet er k år lang gælder

$$\text{Antal personer i starten} = \frac{\text{Risikotid} + \text{Dødstid}}{k}.$$

Her er risikotid antal år som befolkningens personer har livet (i aldersintervallet) og dødstid antal år som befolkningens personer var døde. Med denne formel kan dødshyppigheden skrives som

$$\text{Dødshyppighed} = \frac{k \cdot \text{Antal dødsfald i aldersinterval}}{\text{Risikotid} + \text{Dødstid}}. \quad (2)$$

Vi sætter Chiang's a sådan at

$$\text{Dødstid i aldersinterval} = (k - {}_ka_x) \cdot {}_kD_x,$$

er en god approksimation af antal dødstid som alle personer der døde i aldersintervallet har været død. Hvis vi nu anvender formelen

$${}_kM_x = \frac{{}_kD_x}{{}_kR_x}$$

ser vi at den centrale formel (1) er faktisk lige med formel (2):

$$\begin{aligned}\frac{k \cdot {}_kM_x}{1 + (k - {}_ka_x) \cdot {}_kM_x} &= \frac{k \cdot \frac{{}_kD_x}{{}_kR_x}}{1 + (k - {}_ka_x) \cdot \frac{{}_kD_x}{{}_kR_x}} \\ &= \frac{k \cdot {}_kD_x}{{}_kR_x \cdot (1 + (k - {}_ka_x) \cdot \frac{{}_kD_x}{{}_kR_x})} \\ &= \frac{k \cdot {}_kD_x}{{}_kR_x + (k - {}_ka_x) \cdot {}_kD_x}.\end{aligned}$$

2.2.3 Chiang's a

Chiang's a skal approksimere det forventede antal år levet i intervallet af en person, som dør i intervallet.

5-års aldersintervaller

første leveår: ${}_ka_x[0]1 = 0,1$

aldersinterval 1-5 år: ${}_ka_x[1]4 = 4 \cdot 0,5 = 2$

sidste aldersinterval: $a_x^{max} = \frac{1}{{}_kM_x^{max}}$

alle andre intervaller: ${}_ka_x5 = 5 \cdot 0,5 = 2,5$ for $x = 2, \dots, (x^{max} - 1)$

10-års aldersintervaller

første leveår: ${}_ka_x[0]1 = 0,1$

aldersinterval 1-9 : ${}_ka_x[1]9 = 9 \cdot 0,5 = 4,5$

sidste aldersinterval: $a_x^{max} = \frac{1}{{}_kM_x^{max}}$

alle andre intervaller: ${}_ka_x10 = 10 \cdot 0,5 = 5$ for $x = 2, \dots, (x^{max} - 1)$

2.2.4 Beregning af antal dødsfald i tabelbefolkning

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 $x = 1$. Mere generel betegner vi med $d(x, x + k)$ antal tabelpersoner som dør mellem alder x og alder $x + k$. Dermed er $d(x, x + 1)$ antal personer som dør mellem alder x og alder $x + 1$. Sandsynligheden for at dø mellem to alderstrin (dødshyppighederne) er det centrale element ved konstruktionen af overlevelsestavler. Vi betegner med ${}_kq_x$ sandsynligheden for at dø mellem alder x og alder $x + k$ og har følgende formel for antal dødsfald i aldersintervallet $(x, x + k]$:

$$d(x, x + k) = {}_kq_x \ell(x)$$

2.3 Approximationsformler

Oplysningerne om dødedage bruges også til at beregne den gennemsnitlige tid, som en person, der døde, var død mellem de to fødselsdage. Her omregnes dødedage først til år ved at dividere med 365, hvorefter der divideres med antallet af døde. For de 0-årige drenge svarer de 44.656 dødedage fx til 122,34 dødeår. Ved at dividere de 122,34 dødeår med de 136 døde 0-årige drenge får man, at de 0-årige drenge, der døde mellem deres fødselsdag i 2008 og 2009, i gennemsnit var døde 0,90 år. At tallet er tæt på 1 indikerer, at de 0-årige, der dør, meget ofte dør kort tid efter fødslen. Falder dødsfaldene ligeligt fordelt mellem to fødselsdage vil det give et gennemsnitligt antal dødeår på 0,5. De 0-årige er et særtilfælde, og for de øvrige alderstrin med et forholdsvis stort antal døde, vil det gennemsnitlige antal dødeår ligge tæt på 0,5. For alderstrin med meget få døde kan der være store fluktuationer i det gennemsnitlige antal dødeår, hvilket dog netop pga. det lave antal døde stort set ikke har nogen betydning for de beregnede middellevetider.

Oplysningerne om dødshyppigheder og det gennemsnitlige antal dødeår for alle grupperinger af køn og alderstrin er de grundlæggende informationer, der er brug for i forbindelse med konstruktionen af dødelighedstavler. Målene for middellevetid er en del af dødelighedstavlerne. I dødelighedstavlerne tages der udgangspunkt i en fiktiv befolkning på 100.000 personer, som alle antages at være født på nøjagtig samme tidspunkt. Ud fra de aldersfordelte dødshyppigheder kan de korresponderende overlevelseshyppigheder beregnes, og man kan derved udregne, hvor mange af de 100.000 fra startbefolkningen, der er tilbage ved begyndelsen af hvert alderstrin. Det udregnes også hvor mange personår, som der i alt gennemleves fra alderstrin til alderstrin. Der anvendes en række forskellige bogstavbetegnelser i dødelighedstavlerne, som nedenfor er defineret, hvor x angiver alderstrin. Alle beregninger laves separat for både mænd og kvinder. q_x = sandsynlighed for at dø mellem alderstrin x og alderstrin $x+1$ (dødshyppighed). Denne sandsynlighed er beregnet uden for dødelighedstavlen som beskrevet ovenfor.

$p_x = 1 - q_x$ = sandsynlighed for at overleve fra alderstrin x til starten af alderstrin $x+1$ $l_{x+1} = l_x \times p_x$ = antal personer, som ved begyndelsen af alderstrinnet $x+1$ er tilbage i befolkningen. $l_0 = 100.000$ $l_1 = 100.000 \times p_0$ $l_2 = l_1 \times p_1$ og tilsvarende for alle øvrige alderstrin. $d_x = l_x \times q_x$ = antallet af døde mellem alderstrin x og $x+1$. Der er tale om et antal døde beregnet særligt til dødelighedstavlen med startbefolkningen på 100.000 0-årige, og det har derfor ikke noget med det faktiske antal døde i et givent år at gøre. d_{yearx} = det gennemsnitlige antal dødeår i tidsrummet mellem alderstrinene x og $x+1$ for personer, der døde mellem alderstrin x og $x+1$. Variablen er ligesom q_x beregnet uden for dødelighedstavlen. Metoden er beskrevet ovenfor i eksemplet med 0-årige drenge, der fik beregnet et gennemsnitligt antal dødeår på 0,90. $L_x = l_x - d_x \times d_{yearx}$ = antallet af personår gennemlevet mellem starten af alderstrin x og starten af alderstrin $x+1$. Der tages her hensyn til både antallet af døde og

hvor længe de døde i gennemsnit var døde i året. L_x udregnes som alder-

strinets startbefolkning fratrukket den tid (målt i personår) som personer, der døde på alderstrinet, var døde i året. L_x svarer også til det antal personer på alderstrin x , der i gennemsnit var til stede mellem alderstrinene x og $x+1$. L_x er mindre end l_x , hvis der er mindst ét dødsfald på alderstrin x . L_x er til gengæld større end l_{x+1} , hvis der er mindst ét dødsfald på alderstrin x , da den tid, som de døde lever på alderstrin x , også tæller med i den gennemlevede tid, mens de døde i sagens natur ikke er blandt næste alderstrins overlevende. Hvis dødsfald på et alderstrin er jævnt fordelt over året, er $l_{x+1} L_x = x_{x+1}$ en god approximation, og det er den formel, der tidligere har været anvendt. Med den nye metode, der udregner dødedage præcist, er approximationen ikke nødvendig.

akkumulerede antal personår gennemlevet fra starten af $t = 0$ alderstrin x til og med sidste alderstrin i dødelighedstavlen. Når x er 0, svarer T til alle de personår, som udgangspopulationen på 100.000 gennemlever, indtil den sidste af dem er død. Når x er 60, svarer T tilsvarende til alle de personår, som de, der overlevede frem til deres 60-års fødselsdag (dvs. 160), gennemlever fra de fyldte 60 år, til den sidste af dem er død. $e_x = T_x =$ det gennemsnitlige antal leveår fra starten af alderstrin x til og med sidste l_x alderstrin i dødelighedstavlen. Svarer til middellevetiden. Når x er 0, svarer e til alle de personår, som udgangspopulationen på 100.000 gennemlever, indtil den sidste af dem er død, divideret med udgangspopulation på 100.000. Det er især e_0 svarende til forventede levetid for en nyfødt 0-årig, der hyppigt anvendes, men e beregnes også for alle øvrige alderstrin. Af dødelighedstavlen for mænd, der er vist som eksempel, fremgår det, at middellevetiden for 0-årige er 76,52 år, mens den for 60-årige er 20,4 år. Tallet for 60-årige angiver det gennemsnitlige antal leveår, som en person, der er blevet 60 år, har tilbage at leve i regnet fra den dag, personen fylder 60 år. Fra alderstrinet 104 år og opefter er dødshyppigheden sat til 0,5, da de observerede tal for disse alderstrin er meget usikre pga. det meget lave antal personer på aldersstrinene. Det gennemsnitlige antal dødeår er også sat til 0,5 for alderstrin fra 104 år og efter. I praksis har dødshyppighederne for personer over 100 år meget lille betydning for middellevetiden for 0-årige.

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 og ind- og udvandringstidsdatoer. 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 på 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 dødedage. Personer vil placeres på det alderstrin, som svarer til det antal år, de fyldte i startåret, hvilket i eksemplet vil sige

1. En person, som fyldte 60 år 1. januar 2008 vil fx 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. I et Lexis skema svarer de anvendte grupper til såkaldte C-grupper, som omfatter personer, der er født samme år.

Til beregning af dødshyppigheden skal der kendes antallet af døde og antallet af personer med dødsrisiko. Antallet af døde fås simpelt ved at sammenlægge alle personer i en bestemt køns- og aldersgruppe, der døde mellem de to fødselsdage i 2008 og

1. For at få et mål for gruppen med dødsrisiko målt i personer lægges køns- og

aldersgruppens samlede antal levedage og dødedage sammen, hvorefter tallet divideres med 365.