UNIVERSITETET
I OSLO

Det medisinske fakultet

Institutt for helse og samfunn

MED1100

Samfunnsmedisin, Modul 1

Vår 2026

UNIVERSITETET I OSLO
DET MEDISINSKE FAKULTET
Institutt for helse og samfunn

MED1100
Samfunnsmedisin, Modul 1

KOPISAMLING

Dette materialet er vernet etter åndsverkloven og er helt eller delvis fremstilt etter avtale med Kopinor. Materialet kan benyttes av studenter som deltar i det aktuelle emnet, for egne studier, i ethvert format og på enhver plattform. Uten uttrykkelig samtykke fra rettighetshaver er annen eksemplarfremstilling og tilgjengeliggjøring bare tillatt når det er hjemlet i lov (kopiering til privat bruk, sitat o.l.) eller avtale med Kopinor (www.kopinor.no).
Opplag: 100 eksemplarer. Dato: 30. januar 2026.

9 028220 244906

En Kopisamling til bruk i undervisningen ved
Universitetet i Oslo.

Sammenstilt av Universitetet i Oslo.

Tekstene er kopiert etter tillatelse fra
opphavsrettshavere og copyrightholdere. Det må
ikke kopieres fra dette dokumentet i strid med
åndsverksloven eller med avtaler om kopiering
inngått med Kopinor, interesseorgan for
rettighetshavere til åndsverk.

Trykk og innbinding:
UiO - Grafisk Senter

Oversikt over læringsmål, undervisning og litteratur modul 1 – MED1100

Oppdatert vår 2026 av undervisningsleder Henrik Vogt

Kjære medisinerstudenter, velkommen til samfunnsmedisin i modul 1!

Nedenfor finner dere en oversikt over læringsutbyttene for modul 1 som også finnes på fakultets nettsider. Vi har her strukturert læringsutbyttene etter tema, og knyttet dem til relevant undervisning og anbefalt litteratur. Slik kan dere både finne frem til hvilken litteratur som hører til hvilken undervisning og hvilken undervisning som dekker hvilke læringsutbytter. Vi forventer at dere forbereder dere til gruppeundervisning og at dere som studenter leser litteratur i tillegg til å følge undervisning. Vi stiller også spørsmål som er dekket i litteraturen.

Vi anbefaler at dere kjøper to bøker i samfunnsmedisin. Disse brukes i modul 1, men også andre deler av studiet. Bøkene kan også lånes på biblioteket.

- **Krokstad og Walseth Hara: Sosialmedisin, velferd og trygd. 1. utgave (2022)**
- **Magnus og Bakketeig: Epidemiologi. 4. utgave (2013)**

Den øvrige anbefalte litteraturen skal du kunne lese og/eller laste ned i Leganto.

Du kan også kjøpe et trykket kompendium som inneholder de fleste av tekstene. Dere kan bestille som kompendium via trykkeriet.uio.no. Slik bestiller dere : 1) logg dere inn 2) velg Kompendier\Semester (ex V26)\Det medisinske fakultet\Institutt for Helse og Samfunn 3) Legg kompendium for modul 1 i handlekurven 4) under Koststed velg "kompendiebestilling" i nedtrykksmenyen 5) Trykk bekreft og send ordre.

Dette kompendiet vil også bli gjort tilgjengelig for dere som PDF i Leganto og/eller timeplanen, men skal ikke distribueres utover kullet.

Alle tekstene som er inkludert i kompendiet er kopiert i henhold til Kopinoravtalen. Dersom dere har spørsmål angående undervisningen eller litteraturen, send epost til undervisningsleder henrik.vogt@medisin.uio.no . Har du spørsmål om kompendiet, send gjerne mail til d.k.patel@medisin.uio.no

Læringsmål for samfunnsmedisin, modul 1

Ved fullført blokk skal studentene inneha en helhetlig forståelse av hvordan samfunnsmessige forhold i samspill med biologiske faktorer og helsetjenestens organisering påvirker befolkningens og planetens helse, både i et lokalt, nasjonalt og globalt perspektiv. De skal videre ha innledende kjennskap til epidemiologiske begreper og metoder, samt til grunnleggende prinsipper i folkehelsearbeid. De skal kunne drøfte samfunnsmedisinske aspekter ved et gitt helseproblem.

Læringsutbytter for samfunnsmedisin etter tema, modul 1:

Læringsutbytter knyttet til helse og sykdom som produkter av samspillet mellom biologiske og samfunnsmessige forhold, inkludert helsevesenet

- *gjøre rede for de viktigste årsakene til sosial ulikhet i helse (kunnskap)*
    - Undervisning
        - FOR: Ulikheter i helse
        - GR: Ulikheter i helse
        - GR: Fedme og klima?
    - Litteratur
        - På jakt etter sykdommers sosiale røtter. Kap1, i Sykdommers sosiale røtter, John Gunnar Mæland (redaktør), 1. utgave, Oslo, Gyldendal, 2020, Sidespenn 16 – 36
        - Livserfaringer og biologi - et biomolekylært perspektiv. Kap 10, del 1. Bokkapittel Getz, Linn og kolleger, i Sykdommers sosiale røtter, John Gunnar Mæland (redaktør), 1. utgave, Oslo, Gyldendal, 2020, Sidespenn 205 - 221
        - The sociology of health and illness (Utdrag). Nettleton, S. Avsnittet Explaining the links between class and health (s. 160-166). 2021.
        - Sosialmedisin, velferd og trygd (bokkapitler). Krokstad, S og Hara, KW. Kap 1,3 og 4. Kapittel 1: Sosiale helseforskjeller Kapittel 3: Uten sosialmedisin blir medisinen lett usosial Kapittel 4: Sosiale ulikheter i helsetjenesten og loven om den omvendte omsorg. 2022.
- *beskrive sammenhengen mellom bærekraftig utvikling og folkehelse, med spesielt fokus på klimaendringer og sosial bærekraft (kunnskap)*

    - Undervisning
        - GR: Fedme og klima i et globalt perspektiv
        - FOR: Ulikheter i helse
        - GR: Ulikheter i helse
    - Litteratur
        - The global obesity pandemic: shaped by global drivers and local environments (artikkel). Swinburn et al. s. 804-814. 2011.
        - The Global Syndemic of Obesity, Undernutrition, and Climate Change: The Lancet Commission report (utdrag fra artikkel). s 791-794 og 800-802. 2019.
- *reflektere over hvordan rasisme og diskriminering påvirker folks helse og kvaliteten på helsetjenestene (generell kompetanse)*
    - Undervisning
        - FOR: Ulikheter i helse
        - GR: Ulikheter i helse
    - Litteratur
        - Discrimination and Health. Magnus, J. Bokkapittel i Migrant Health: A Primary Care Perspective. s 55 – 63. 2019,
        - Livserfaringer og biologi - et biomolekylært perspektiv. Kap 10, del 1. Bokkapittel Getz, Linn og kolleger, i Sykdommers sosiale røtter, John Gunnar Mæland (redaktør), 1. utgave, Oslo, Gyldendal, 2020, Sidespenn 205 - 221
- *reflektere over hva som kan anses som et likt, likeverdig og rettferdig tilbud på helsetjenester (generell kompetanse)*
    - Undervisning
        - FOR: Ulikheter i helse
        - GR: Ulikheter i helse

Læringsutbytter knyttet til medisinsk vitenskapsteori, historie, filosofi og kritisk tenkning

- *gjøre rede for utfordringene som ligger i medikalisering, overbehandling, helseangst (kunnskap)*
    - *Undervisning*
        - FOR: Medikalisering og overbehandling
        - GR: Syk eller frisk
    - Litteratur
        - Medicalization. Conrad, P og Waggoner, M. Artikkel oppslagsverk. Sidespenn 1448 - 1452

        - Preventing overdiagnosis: how to stop harming the healthy. Moynihan, R; Doust, J; Henry, D. Artikkel fra BMJ. s 1-4
        - For mye helsetjenester. Nylenna, Keeping og Larsen. Kap 2.6 i Samfunnsmedisin på norsk., 2019. s 159-165.
- *drøfte begrepsparene helse/sykdom og normalt/patologisk (ferdighet)*
    - Undervisning
        - FOR: Medikalisering
        - GR: Syk eller frisk
    - Litteratur
        - Hva er sykdom? Hofmann, B. Bokkapittel i Sykdom og behandling av Bach-Gansmo, Ørn og Mjell. 2022. s 3-12
- *drøfte gjennom eksempler fra film og litteratur hvordan pasienters og legers erfaringer, samt legepasientforholdet fremstilles på ulike måter og hva vi kan lære av dette (ferdighet)*
    - Undervisning
        - GR: Medisinsk humaniora
    - Litteratur
        - Representations of Medicine, Illness and the Body. Bokkapittel i Medicine as Culture: Illness, Disease and the Body, av Lupton, D, London. 2012, s 52-78
- *reflektere over at det foreligger usikkerhet i det helsefaglige beslutningsgrunnlaget (generell kompetanse)*
    - Dette læringsmålet har ingen spesifikk undervisning, men undervisningen i epidemiologi er relevant. Undervisningen GR Syk eller frisk., FOR Medikalisering, GR Medisinsk humaniora er og relevant.
    - Litteratur
        - Tolerating Uncertainty — The Next Medical Revolution? Artikkel. Simpkin, AL og Schwartzstein, RM. Sidespenn 1713 – 1715

Læringsutbytte knyttet til legerolle og profesjonalitet

- reflektere over en profesjonsutøvers ansvar og rolle i samfunnet (generell kompetanse)
    - Undervisning:
        - FOR: Medikalisering og overbehandling
        - GR: Syk eller frisk
        - GR: Medisinsk humaniora
        - GR Fedme og klima
- ingen spesifikk litteratur, men referansene i punktene over er relevante.

Læringsutbytter knyttet til epidemiologi, folkehelse, forebygging og helsefremming

- *forklare de vanligste begreper for sykelighet, dødelighet og risiko for sykdom (kunnskap)*
    - Undervisning
        - FOR: Mål for forekomst av helse, sykdom og død
        - GR: Hvordan finne ut hvorfor noen blir syke
        - GR: Gruppearbeid – fedme og statistikk
        - GR: Syk eller frisk
    - Litteratur:
        - Epidemiologi (bok). Magnus og Bakketeig, 2013. Kap 7
- *gjøre rede for de vanligste samfunnsmedisinske metoder og forskningsdesign (kunnskap)*
    - Undervisning
        - FOR: Forskningsdesign
        - FOR: Mål for sammenheng mellom eksponering og sykdom
        - FOR: Mål for forekomst av helse, sykdom og død
        - GR: Hvordan finne ut hvorfor noen blir syke
        - GR: Gruppearbeid – fedme og statistikk
    - Litteratur
        - Epidemiologi (bok). Magnus og Bakketeig, 2013. Kap 1-8+10
        - På jakt etter sykdommers sosiale røtter. Kap1, i Sykdommers sosiale røtter, John Gunnar Mæland (redaktør), 1. utgave, Oslo, Gyldendal, 2020, Sidespenn 16 – 36
- *beskrive hovedlinjer i sykdomsutviklingen nasjonalt og globalt og redegjøre for utfordringer det skaper for helsetjenesten i dag (kunnskap)*
    - Undervisning
        - FOR: Morgendagens helseutfordringer
        - GR: Fedme og klima I et globalt perspektiv
    - Litteratur
        - Samfunnsmedisin på norsk. Helsetilstanden i Norge (utdrag). Kap 2.2 i Nylenna, Keeping og Larsen, 2019. s71 – 93.
        - World health statistics. Monitoring health for the SDGs, Sustainable Development Goals. Introduction and key messages s 1-4. World Health Organization, 2025
        - The global obesity pandemic: shaped by global drivers and local environments (artikkel). s. 804-814. Swinburn et al. 2011.
        - På jakt etter sykdommers sosiale røtter. Kap1.1 og 1.2, bokkapittel i Sykdommers sosiale røtter, John Gunnar

Mæland (redaktør), 1. utgave, Oslo, Gyldendal, 2020, Sidespenn 16 – 36

Læringsutbytte knyttet til helsetjenestens oppbygning, aktører og samhandling

- beskrive helsetjenestens oppbygning, rollen til ulike grupper av helsepersonell, samt samhandling mellom gruppene (kunnskap)
    - Undervisning:
        - FOR: Helsevesenets oppbygning
    - Litteratur
        - Sosialmedisin, velferd og trygd (bokkapitler) Kap 11 og 12. Kapittel 11: Veier til velferdsstaten og veien videre Kapittel 12: Den norske velferdsstaten- utdanning, helse, arbeid og trygd. Krokstad, S og Hara, KW, 2022.

Innhold

Mæland, John Gunnar	"Kap. 1: På jakt etter sykdommers sosiale røtter", i (Red) Mæland, John Gunnar	Side 1
	Sykdommers sosiale røtter 2020 ss. 16-37	
	ISBN: 9788205523944	
	Copyright (C) 2020 Gyldendal	
	Kopiert med hjemmel i Kopinor-avtalen	
Getz, Linn; Kirkengen, Anna Luise & Ulvestad, Elling	"Kap. 10: Kroppsliggjøring av erfaring og mening - belastende erfaringers innvirkning på menneskets biologi. Utdrag: Livserfaringer og biologi - et biomolekylært perspektiv", i (Red) Mæland, John Gunnar	Side 25
	Sykdommers sosiale røtter 2020 ss. 205-221	
	ISBN: 9788205523944	
	Copyright (C) 2020 Gyldendal	
	Kopiert med hjemmel i Kopinor-avtalen	
Magnus, Jeanette H.	"Ch. 5: Discrimination and Health", i (Red) Diaz, Esperanza & Kumar, Bernadette N.	Side 45
	Migrant Health: A Primary Care Perspective 2019 ss. 55-63	
	Copyright (C) 2019 CRC Press	
	Kopiert med hjemmel i Kopinor-avtalen	

Swinburn, Boyd A. et al	"The global obesity pandemic. shaped by global drivers and local environments", i *The Lancet* 378 (9793) 2011 ss. 804-814	Side 57

Copyright (C) 2011 Elsevier Ltd

Kopiert med hjemmel i Kopinor-avtalen

Swinburn, Boyd A. et al	"The Global Syndemic of Obesity, Undernutrition, and Climate Change. The Lancet Commission report", i *The Lancet. Vol 393, Issue 10173* 2019 ss. 791-794, 800-802	Side 71

Copyright (C) 2019 Elsevier Ltd

Kopiert med hjemmel i Kopinor-avtalen

Conrad, Peter & Waggoner, Miranda	"Medicalization", i (Red) Cockerham, William C.; Dingwall, Robert & Quah, Stella *The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society* 2014 ss. 1-4	Side 81

Copyright (C) 2014 Wiley Blackwell

Kopiert med hjemmel i Kopinor-avtalen

Henry, David; Doust, Jenny & Moynihan, Ray	"Preventing overdiagnosis: how to stop harming the healthy", i *The BMJ* 344 (7859) 2012 ss. 1-6	Side 87

Copyright (C) 2012 BMJ

Kopiert med hjemmel i Kopinor-avtalen

Larsen, Øyvind; Keeping, David & Nylenna, Magne	"Kap. 2.6: For mye helsetjenester", i *Samfunnsmedisin på norsk* 2019 ss. 159-165	Side 95

Copyright (C) 2019 Gyldendal

Kopiert med hjemmel i Kopinor-avtalen

Hofmann, Bjørn	"Kap. 1: Hva er sykdom?", i (Red) Ørn, Stein & Bach-Gansmo, Edvin *Sykdom og behandling* 2022 ss. 3-11 ISBN: 9788205542587	Side 105

Copyright (C) 2022 Gyldendal

Kopiert med hjemmel i Kopinor-avtalen

Lupton, Deborah	"Representations of Medicine, Illness and the Body", i (Red) Illness, Disease and the Body *Medicine as Culture* 2012 ss. 2-33	Side 117

Copyright (C) 2012 Medicine as Culture

Kopiert med hjemmel i Kopinor-avtalen

Simpkin, Arabella L. & Schwartzstein, Richard M	"Tolerating Uncertainty. The Next Medical Revolution?", i *The New England journal of medicine* 375 (18) 2016 ss. 1713-1715	Side 151

Copyright (C) 2016

Kopiert med hjemmel i Kopinor-avtalen

Larsen, Øyvind; Keeping, David & Nylenna, Magne	"Kap. 2.2: Helsetilstanden i Norge", i *Samfunnsmedisin på norsk* 2019 ss. 71-93	Side 157

Copyright (C) 2019 Gyldendal

Kopiert med hjemmel i Kopinor-avtalen

Introduction and key messages	"Introduction and key messages", i *World health statistics 2025. Monitoring health for the SDGs, Sustainable Development Goals* 2025 ss. 1-4	Side 183

Copyright (C) 2025

Ikke rettighetsbelagt

Mæland, John Gunnar

"Kap. 1: På jakt etter sykdommers sosiale røtter", i (Red) Mæland, John Gunnar

Sykdommers sosiale røtter 2020 ss. 16-37

ISBN: 9788205523944

Copyright (C) 2020 Gyldendal

Kopiert med hjemmel i Kopinor-avtalen

KAPITTEL 1

På jakt etter sykdommers sosiale røtter

John Gunnar Mæland

Moderne biomedisinsk forskning har hatt stor suksess. I dag forstår vi stadig mer av hvordan sykdom oppstår i kroppen – i vevene og organene, inne i cellene og etter hvert også på det molekylærbiologiske nivået. Denne dype og detaljerte innsikten åpner opp for tidligere og mer presis sykdomsdiagnose og for nye behandlingsmuligheter.

Men biomedisinsk forskning gir ikke fullgode svar på spørsmålet om *hvorfor* vi blir syke. Bak de aller fleste sykdommer ligger nemlig kaskader av påvirkninger som går gjennom flere steg. Noen av disse er biologiske og fysiologiske prosesser inne i kroppen, men mye er påvirkninger fra utsiden – fra det fysiske og det sosiale miljøet som omgir oss. Kunnskap om slike opprinnelige sykdomsårsaker (*root causes* på engelsk) og påfølgende årsakskjeder er avgjørende for mulighetene til å forebygge sykdom og for tidlig død.

Denne boken skal først og fremst handle om sosiale rotårsaker til sykdom og helsesvikt – en gammel tematikk som vi nå har fått mer konkret kunnskap om. Og vi skal knytte denne innsikten til ny forståelse av hvordan slike ytre påvirkninger «kommer under huden» og bidrar til sykdomsfremkallende forandringer i kroppen. Hensikten er å se kroppen og dens ulike komponenter i samspill med omgivelsene.

I jakten på sykdomsforklaringer er det lett å gå seg vill. «Alt henger sammen med alt» er en nærliggende ironisk kommentar. Men forskning har avdekket at sosiale faktorer er av større betydning for folks helse enn vi tidligere forestilte oss:

- folkehelsen følger en klar sosial gradient: jo høyere sosioøkonomisk status, desto bedre helse (kapittel 2)
- livshistorien vår innskrives i kroppen og former vår helse, også på lang sikt (kapittel 3, 4, 9 og 10)
- trygghet, stabilitet og positiv stimulans i de første leveårene er ofte avgjørende for framtidig psykisk og somatisk helse (kapittel 4 og 10)
- kontakt, samhandling og tillitsfulle relasjoner med andre mennesker har klare helsekonsekvenser (kapittel 5 og 6)
- diskriminering og utstøting fra fellesskapet er en stor helserisiko (kapittel 6 og 7)
- kjønnsforskjeller i helse er delvis sosialt betinget (kapittel 7)
- ressurs- og maktfordeling i samfunnet har betydning for folkehelsen, men også våre kollektive velferdsordninger, folkehelsearbeidet og fellesskapsfølelsen i samfunnet (kapittel 5, 8, 11 og 12)

Ennå er mye ukjent i dette sammensatte landskapet, men vi forstår stadig mer av årsakskjedene fra de sosiale omstendighetene til sykdom og helsesvikt. I dette innledningskapittelet skal vi diskutere noen overordnede modeller for slike sammenhenger og hvordan vi går fram for å avdekke disse. Utover i boken vil disse sosiale helsedeterminantene og de sykdomsskapende mekanismene bli nærmere beskrevet, før de praktiske implikasjonene for folkehelsearbeidet diskuteres.

1.1 Fra monokausale til sammensatte forklaringer på sykdom

Hvorfor blir vi syke? Dette spørsmålet har opptatt mennesker i årtusener. Lenge ble sykdom sett på som straff fra gudene (eller de underjordiske) når menneskene hadde satt seg ut over de regler som gjaldt. Selv om gresk medisin for 2400 år siden introduserte ideen om naturlige årsaker i kropp og omgivelser som forklaring på sykdom, forsvant denne innsikten i middelalderens Europa da kristendommen på ny brakte synd og moralsk svikt inn som sykdomsforklaringer. Først med renessansens gjenoppdagelse av antikkens ideer startet en mer rasjonell tilnærming til kropp og sykdom (Schiøtz 2017).

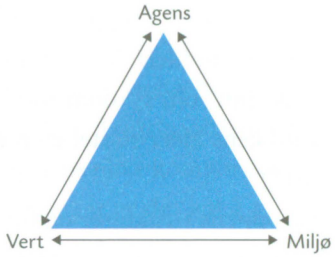
Selv om oppdagelsene av kroppens strukturer og funksjoner fra 1500-tallet av banet veien for en ny måte å forstå sykdom på – som spesifikke tilstander med sine egne årsaker – famlet man lenge i blinde om hva disse årsakene var. Det avgjørende

vendepunktet kom først i 1870-årene da det ble påvist at mange av datidens alvorlige sykdommer var forårsaket av spesifikke *agens*: bakterier og andre mikroorganismer.

Disse oppdagelsene løste tilsynelatende en gammel diskusjon om hva som forårsaker sykdommer og epidemier: *kontagium* (smittestoff) eller *miasme* (forurensning). Tilhengerne av miasmeteorien mente at sykdomsforklaringen lå i bedervet luft, som særlig rammet mennesker som enten gjennom arv eller livsbetingelser var predisponerte for sykdom. Kontagionistene mente det fantes smittestoff som kunne spres fra den syke til andre. Mens smitteteorien ledet til isolering (karantene) av syke personer som forebyggende tiltak, argumenterte tilhengerne av miasmeteorien for forbedring av det ytre miljø. De store sanitære reformene i Storbritannia på midten av 1800-tallet ble begrunnet i miasmeteorien, men også av statistiske data som viste hvordan sykdomsforekomst og dødelighet varierte geografisk og sosialt. Selv om miasmeteorien etter hvert led nederlag, førte forbedrete vann- og kloakkledninger, offentlig avfallshåndtering og økt bostandard til dramatiske forbedringer i folkehelsen. Dette viste at både smittestoff og omgivelser må inngå i en forståelse av hvordan sykdom oppstår og kan forhindres (Schiøtz 2017.)

Det ble etter hvert klart at slikt mikrobiell sykdomsagens var nødvendig, men ikke alltid tilstrekkelig årsak til at folk fikk den aktuelle sykdommen. Også egenskaper ved personen (*verten*), som for eksempel ernæringstilstand eller tidligere smitte av den aktuelle mikroorganismen, spilte en rolle for om mikroben utløste en sykdom eller ikke. Dessuten fant man at sykdomsrisikoen hang sammen med egenskaper ved omgivelsene (*miljøet*), som eksempelvis sanitære forhold, boligstandard og tilgang på helsetjenester.

Denne innsikten ble nedfelt i den velkjente *vert, agens og miljø-modellen* som har ligget til grunn for infeksjonsepidemiologisk forskning opp til våre dager (figur 1.1). Ifølge modellen kan de tre hjørnene i denne triaden gjensidig påvirke hverandre, og det er summen av sykdomsfremkallende krefter som avgjør om sykdom oppstår eller ikke. Men fokuset i folkehelsearbeidets første æra var utvilsomt på agenset, den spesifikke og nødvendige sykdomsårsaken.



Figur 1.1 Agens, vert, miljø-modellen.

Fra agens til multifaktorielle risikofaktorer

Vert, agens og miljø-modellen var så vellykket at den også ble lagt til grunn da man på 1900-tallet tok fatt på å forklare hvordan ikke-smittsomme sykdommer oppstår. For enkelte av disse tilstandene ble agens funnet å være enten for mye eller for lite av spesifikke kjemiske stoffer i kroppen, som vitaminer, mineraler eller hormoner. Men for mange sykdommer fant man ikke slike nødvendige årsaksfaktorer. I stedet ble sykdommen knyttet til ulike egenskaper ved individet eller miljøet som statistisk sett hang sammen med sykdomsrisikoen. Tanken om én spesifikk og nødvendig sykdomsårsak måtte byttes ut med en *multifaktoriell sykdomsmodell* der mange faktorer kunne samvirke i produksjonen av sykdom.

Jakten på slike risikofaktorer preget folkehelsearbeidets andre æra – som handlet om å bekjempe hjerte- og karsykdommer, kreft, sukkersyke, magesår, lungesykdommer og lignende kroniske sykdommer. Særlig drev den markerte økningen i hjerte- og karsykdommene i de industrialiserte landene etter siste verdenskrig årsakforskningen videre langs dette sporet. Da man fra 1960-årene av fikk større muligheter til å samle inn og analysere data fra store befolkningsgrupper, skjøt risikofaktorepidemiologien virkelig fart.

En utilsiktet konsekvens av denne nye epidemiologien var at man nå konsentrerte jakten på sykdomsforklaringer til egenskaper ved *individene*, enten det dreiet seg om fysiologiske mål (som høyde, vekt eller blodtrykk), biokjemiske målinger (som kolesterolnivå eller blodsukker) eller atferdsfaktorer (som tobakksrøyking, kostvaner eller fysisk aktivitet). Forhold utenfor individene ble enten ignorert eller forsøkt kontrollert for i de statistiske analysene av sammenhengene på individnivå.

Den individorienterte årsakforskningen, som fortsatt er helt dominerende, passer godt til den kliniske forskningsmodellen som gjerne går under navnet *biomedisinsk forskning*. Dette er naturvitenskapelig forskning basert på metodologisk reduksjonisme: at fenomener best forklares gjennom studier av enkeltdeler, og da fortrinnsvis på et så fundamentalt nivå som mulig. De mulige biomedisinske sykdomsårsakene dreier seg derfor om biologiske, kjemiske eller fysiske kroppsfaktorer som studeres i laboratorier eller gjennom teknologiske undersøkelser av deler av kroppen. Også her er det et viktig poeng å unngå påvirkning fra omgivelsene. Idealet er den objektive og kontrollerte studie av variasjon i biologisk funksjon som kan reproduseres under samme betingelser.

Fra risikofaktortenkning til salutogenese

På slutten av 1970-årene oppsto en reaksjon mot den snevre måten å lete etter sykdomsforklaringer på, og mot den ensidige naturvitenskapelige kunnskapsbasisen for forebyggende medisin (kapittel 10). Motreaksjonen var en del av 1970-årenes store frigjøringsprosjekter, som kvinnebevegelsen, forbrukerbevegelsen og frigjøringsarbeidet i de tidligere kolonimaktene i den tredje verden. Denne alternative måten å tenke og drive folkehelsearbeid på fikk etter hvert betegnelsen *helsefremmende arbeid* (engelsk: *health promotion*, se kapittel 11).

Tilhengerne av denne nye ideologien kritiserte risikofaktorepidemiologien for at den ignorerte de bakenforliggende forklaringene på folks levesett. Usunn livsstil var ikke noe folk valgte, men noe de ble påført av økonomiske og sosiale omstendigheter (Crawford 1977). De rådende risikofaktorforklaringene la skylden for dårlig helse urettmessig på enkeltpersoner som ofte var ofre for aggressiv markedsføring, usolidarisk prispolitikk og annen sosial urettferdighet. Man måtte altså gå bak individuell atferd og finne fram til «årsakene til årsakene», som var knyttet til folks livssituasjon og handlingsmuligheter.

I motsetning til sykdomsforebyggingens ekspertbaserte tilnærming, tok helsefremmende arbeid utgangspunkt i folks egne vurderinger og prioriteringer. Et nøkkelbegrep var *empowerment*, det å gi folk makt og kontroll over forhold som påvirket deres helse (kapittel 9). Helse ble forstått helhetlig, med vekt på helsens betydning for å kunne leve gode og produktive liv (Mæland 2009).

En viktig spore til denne dreiningen mot helsefremmende arbeid var Verdens helseorganisasjons program *Helse for alle*, først uttrykt i den såkalte *Alma Ata*-erklæringen fra 1978, som fremhevet viktigheten av en god primærhelsetjeneste. Her slo man også fast at folkehelsen berodde på innsats i alle samfunnssektorer og at folkehelsearbeidet krevde aktiv deltakelse fra folk og lokalsamfunn. Samfunnet måtte ta ansvar for å legge forholdene best mulig til rette for at alle fikk realisert sitt helsepotensial.

En annen inspirasjonskilde var *salutogenese*-begrepet som sosiologen *Aaron Antonovsky* (1923–1994) lanserte i boken *Health, Stress and Coping* (1979). Antonovsky argumenterte her for en ny måte å tenke forebygging på: I stedet for ensidig å fokusere på sykdomsfremkallende faktorer, måtte forebyggende helsearbeid dreie seg om å styrke folks helseressurser. Det innebar å forstå bedre hva som ga folk motstandskraft mot helsenedbrytende krefter.

Antonovsky lanserte sin egen teori om hva som best beskyttet mot negative påvirkninger: evnen til å oppfatte verden som rimelig forutsigbar og sammenhengende (*sense of coherence – SOC*). Dette er et psykologisk begrep (se kapittel 9 side 196), men Antonovsky pekte samtidig på ulike sosiale faktorer som han mente var helsefremmende, som sosial identitet og kulturelt fellesskap, materielle ressurser og sosial støtte. Han kalte dette generelle motstandsfaktorer (Antonovsky 1979, Antonovsky 2012).

En tredje inspirator var den amerikanske psykiater og indremedisiner *George L. Engel* (1913–1999), som i 1977 publiserte en oppsiktsvekkende artikkel i det anerkjente tidsskriftet *Science*, hvor han kritiserte den biomedisinske sykdomsmodellen og argumenterte for at den burde erstattes av en *biopsykososial* modell for sykdom. Denne artikkelen, som etter hvert ble en av de mest siterte vitenskapelige publikasjonene overhodet, traff nok en nerve i tiden, selv om Engel var nokså diffus om hva han la i sin alternative biopsykososiale modell (Engel 1977).

Med andre ord beveget man seg tidlig i 1980-årene fra en snever medisinsk sykdomsmodell til en bredere sosial helsemodell som basis for folkehelsearbeidet (Mæland 2016). Dette åpnet opp for å lete etter andre påvirkningsfaktorer for sykdom og helse enn individuelle biologiske og atferdsmessige karakteristika knyttet til sykdom. Den helsefremmende tenkningen la like stor vekt på ressurser for helse som på helseskadelige forhold. Begrepet *helsedeterminant* utfordret etter hvert risikofaktorbegrepet. Denne utviklingen kulminerte i det såkalte *Ottawa-charteret for helsefremmende arbeid* i 1986, som slo fast at de fundamentale forutsetningene for helse er fred, bolig, utdannelse, mat, inntekt, et stabilt økosystem, bærekraftige ressurser, sosial rettferdighet og likeverd (Helsedirektoratet 1987).

1.2 Sosialepidemiologi vokser frem

De første spirene til våre dagers sosialepidemiologi kan spores i 1960- og 1970-årene i USA gjennom en «befruktning» mellom epidemiologi og medisinsk sosiologi (Berkman & Kawachi 2014). En medvirkende årsak til den økte interessen for sosiale og samfunnsmessige helsedeterminanter var overraskende funn fra årsakforskningen på hjerte- og karsykdommer. I tillegg til individuelle risikofaktorer som høyt blodtrykk, høyt blodkolesterolnivå og tobakksrøyking, viste de epidemiologiske studiene sammenhenger mellom flere såkalte psykososiale faktorer og sykdomsrisiko. Dette handlet blant annet om stress, livshendelser, sosial status, etnisitet, sosial isolasjon og psykososialt arbeidsmiljø. At omgivelsene kunne bety mye for

risiko for hjerte- og karsykdom kom også frem i studier av befolkninger som flyttet fra en kultur til en annen, såkalte migrantstudier (kapittel 6).

Japanere (som i utgangspunktet hadde meget lav forekomst av hjerte- og karsykdommer) som flyttet til Hawaii, økte sin sykdomsrisiko, og japanere som emigrerte til USA, økte risikoen ytterligere (Marmot mfl. 1975).

Til dette kom at flere store longitudinelle befolkningsundersøkelser publisert i samme tidsrom fant klare sammenhenger mellom omfanget av folks sosiale nettverk og fremtidig dødsrisiko.

I 1965 ble nær 7000 personer bosatt i California nøye kartlagt om alle forhold som kunne ha betydning for deres fremtidige helse. Disse personene ble så fulgt i ni år. Det viste seg at risikoen for å dø hang tydelig sammen med hvor omfattende sosialt nettverk personene hadde i 1965. Man fant at følgende forhold hadde betydning for leveutsiktene: å være gift, å ha nære kontakter blant venner og familie, samt det å tilhøre et religiøst samfunn eller å være med i andre organisasjoner. Når forskerne slo disse opplysningene sammen i en indeks for sosialt nettverk, kom det frem et klart mønster for alle aldersgrupper og for begge kjønn: jo sterkere sosialt nettverk, jo lavere dødelighet (Berkman & Syme, 1979).

Disse sammenhengene gjaldt for alle de store dødsårsaksgruppene og lot seg ikke forklare ved andre faktorer (se kapittel 5). Sosial støtte syntes altså å virke helsebeskyttende i seg selv.

Sosiale ulikheter gjenoppdages

Den viktigste sporen til en sosialt orientert epidemiologi kom imidlertid da man i 1980- og 1990-årene gjenoppdaget sosiale ulikheter i helse. Denne gamle sosialmedisinske innsikten (rammetekst 1.1) var på denne tiden falt ut av oppmerksomheten. Utviklingen av velferdssamfunnet som jo nettopp tok sikte på å eliminere sosiale helseforskjeller, ble av de fleste ansett som så vellykket at man tok det for gitt at helseulikhetene var i ferd med å forsvinne. Sjokket var desto større da det viste seg at slike ulikheter ikke bare fortsatt var til stede, men at de også var økende (kapittel 2). Mønsteret er at helsetilstanden forverrer seg trinnvis jo lenger ned i det sosiale hierarkiet en gruppe befinner seg, det som kalles den sosiale *helsegradienten*. Sosial ulikhet i helse er altså ikke noe isolert fattigdomsproblem, men angår hele befolkningen.

Rammetekst 1.1 De fattige er de sykeste, og de sykeste er fattige

At fattige mennesker langt oftere enn andre ble rammet av sykdommer og tidlig død var en gammel observasjon. På 1800-tallet var den gjengse holdningen at fattige selv var ansvarlige for sin vanskjebne, bortsett fra hvis det forelå åpenbare grunner til armoden, som fysiske skavanker eller sinnssykdom. Fattigdom for øvrig skyldtes dovenskap og moralsk forfall (usedelighet). Å hjelpe fattige var bare å øke problemet, fordi økt levestandard ville føre til flere barn og flere fattige (Schiøtz 2017).

Noen reformvennlige leger, som den senere kjente tyske patologen *Rudolf Virchow* (1821–1902), anla et annet syn på denne sammenhengen. Etter å ha studert en tyfusepidemi i en fattig tysk provins i 1847–48, argumenterte Virchow for at det var selve fattigdommen som var den grunnleggende forklaringen på at sykdom florerte og at epidemier spredde seg i befolkningen. Dette skjedde, hevdet han, fordi sult og underernæring, elendige boforhold og mangel på penger økte sykdomsmottakeligheten og minsket muligheten for å få god legehjelp. Bare hvis staten grep inn for å bedre leveforholdene, kunne de fattige komme ut av den onde sirkelen mellom fattigdom og sykdom. «Medisin er en sosial vitenskap, og politikk er intet annet enn medisin i stort», sa Virchow, som regnes som en av grunnleggerne av sosialmedisin (Waitzkin 1981).

Den såkalte *Black-rapporten*[2] fra 1980 (Townsend & Davidsson 1982) avdekket sosiale klasseforskjeller i helse og dødelighet i Storbritannia av en dimensjon som ingen hadde forestilt seg (kapittel 3). Rapporten vakte stor oppsikt, selv om (eller kanskje nettopp fordi) den britiske regjeringen forsøkte å hindre at den ble kjent. I de skandinaviske landene gikk man likevel ut fra at klasseproblemet var langt mindre selv om det ble dokumentert betydelige forskjeller i dødelighet mellom Oslos bydeler. Men da en europeisk forskergruppe fant at Sverige og Norge hadde de største relative sosiale ulikhetene i helse av elleve europeiske land, ble dette en kraftig vekker for både forskere og politikere (Mackenbach mfl. 1997).

2 *Report of the Working Group on Inequalities in Health* fikk navn etter ekspertgruppens leder *Sir Douglas Black*.

Det var altså åpenbart at man omkring årtusenskiftet for alvor var kommet på sporet av sykdomsskapende faktorer i det sosiale miljøet. Dette ble starten på en ny type epidemiologisk forskning, som fikk betegnelsen *sosialepidemiologi*.

1.3 Hvordan skiller sosialepidemiologi seg fra tradisjonell epidemiologi?

En underliggende antagelse for sosialepidemiologi er at fordelingen av sykdom og helse i et samfunn er et uttrykk for hvordan sosiale fordeler og mangler er fordelt i dette samfunnet. Den epidemiologiske oppgaven er å finne hvilke sosiale og strukturelle forhold som spiller størst rolle for folks helse, og hvordan slike faktorer bidrar til sykdom hos enkeltpersoner, i utsatte grupper og i befolkningen (Honjo 2004).

Mens ordinær epidemiologi forsøker å holde sosiale bakgrunnsforhold konstante for å vurdere effekten av individuelle faktorer på helseutfall, er det nettopp forskjeller som følge av sosiale karakteristika som er den sosiale epidemiologiens kjerneområde. Sosialepidemiologi er altså opptatt av hvordan forhold som ligger utenfor og mellom individer i «det sosiale rom», påvirker folks helse i form av risiko for sykdom eller død.

Det sosiale rom

Metaforisk er det sosiale rom mer å ligne med et rom i et hus i en by i et samfunn – altså som kinesiske esker som omslutter hverandre (Susser & Susser 1996). Sosiale omgivelser spenner nivåmessig fra de nære mellommenneskelige forbindelsene i familie og vennekrets (*mikronivået*), via sosiale strukturer i nabolag og lokalsamfunn (*mesonivået*) til samfunnsmessige karakteristika av økonomisk, kulturell og politisk natur (*makronivået*).

Å gi oversikt over alle mulige sosiale helsedeterminanter er vanskelig, men figur 1.2 gir en indikasjon på hva dette handler om. I figuren er de viktigste påvirkningsfaktorene for helse fordelt lagvis med individet i sentrum. Hvert lag inneholder ulike elementer. Mellom disse lagene og elementene er det muligheter for gjensidige påvirkninger.



Figur 1.2 Påvirkningsfaktorer for helse. Etter Dahlgren og Whitehead 1991.

Sentralt i figuren finnes de upåvirkelige individuelle egenskapene ved individene: arvelige og biologiske faktorer, kjønn og alder. Samlet bestemmer disse forholdene personens sårbarhet eller robusthet overfor ytre påvirkninger, det vi kan kalle *individuell disposisjon* for sykdom og helse.

Det neste laget av påvirkningsfaktorer i modellen utgjøres av personens livsstil. Levevaner oppfattes tradisjonelt som uttrykk for individuelle valg, men gjenspeiler også langt på vei påvirkninger fra omgivelsene (se kapittel 9). Levevaner er derfor på samme tid uttrykk for personlige preferanser og for sosial påvirkning.

Det sosiale nettverket og nærmiljøet utgjør tredje lag i figuren. Omfang og kvalitet av samhandling med familiemedlemmer, venner, kolleger og andre personer er viktige for helsen. Likeledes har kollektive aspekter ved nærmiljøet som fellesskapsopplevelse, tillit, samhandling og identitet (ofte benevnt som *sosial kapital*) betydning for folks helseutvikling (se kapittel 5).

Det fjerde laget av ytre påvirkningsfaktorer handler om levekår. Dette består av en rekke delområder, hvorav noen dreier seg om materielle omstendigheter (bolig, ernæring, sanitærforhold) og andre om sosiale forhold, som utdanning, arbeid og helse- og velferdstjenester (kapittel 2, 8, 9 og 11).

Slike levekår bestemmes langt på vei av de *generelle økonomiske, kulturelle og politiske forholdene* i samfunnet, som danner modellens ytterste lag av påvirkningsfaktorer. Slike makrofaktorer har også betydning for mange av de mer individnære

helsedeterminantene, som kvaliteten på lokalsamfunn, sosial kapital og folks helserelaterte atferd (kapittel 8 og 11).

Påvirkning over tid

Det er en økende forståelse av at tidligere livsomstendigheter kan være bestemmende for individets handlinger, sosiale utvikling og helse. Sosialepidemiologi bygger derfor i økende grad på *livsløpsmodeller,* hvor påvirkninger spores etter lang tid innen individuelle livsløp og til og med over generasjoner. Denne tidsdimensjonen åpner opp for å studere mange ulike eksponeringer og deres felles betydning for senere helseutfall (se kapittel 3, 9 og 10).

En mye brukt metafor for sykdomsutvikling er bildet av mennesker som føres med vannet nedover en elv. Skjebnen til disse menneskene avhenger dels av forholdene langt nede i elven (*nedstrømsfaktorer, «down-stream»*), men først og fremst av det som fikk dem til å falle ut i elven høyere oppe (*oppstrømsfaktorer, «up-stream»*). Dette bildet illustrerer godt forskjellen på en primær forebyggende strategi (å hindre at folk faller ut i elven) og en behandlingsbasert strategi (å redde de druknende opp av vannet lenger nede). Elvemetaforen fremhever også tid og prosess i sykdomsutviklingen og at både distale (oppstrøms-) og proksimale (nedstrøms-) påvirkningsfaktorer må tas med i betraktning (kapittel 9).

Et økologisk perspektiv

Mens tradisjonell epidemiologi antar at fordelingen av sykdom i befolkningen utelukkende er et uttrykk for summen av individuelle risikoer, har sosialepidemiologi et økologisk perspektiv, nemlig at individene og deres biologi påvirkes av og samspiller med faktorer i omgivelsene på mange nivåer (Krieger 2014). Som enkeltmennesker tilhører vi ulike sosiale grupper, vi vokser opp i og tilbringer våre liv i ulike sosiale og kulturelle omgivelser, og vi er medlemmer av ulike samfunnsgrupper. Hver befolkningsgruppe har sin egen historie og kultur, som påvirker hvordan folk utsettes for spesifikke individuelle risikofaktorer (Honjo 2004). Et økologisk perspektiv som tar hensyn til tid, sted og sosial gruppe, er derfor nødvendig for å forstå hvordan sykdom og helse fordeler seg i befolkningen (rammetekst 1.2).

Rammetekst 1.2 Nancy Kriegers økososiale teori for sykdomsfordeling

Nancy Krieger er en ledende amerikansk sosialepidemiolog som har utviklet det hun kaller økososial teori for sykdomsfordeling (Krieger 2014). Krieger tar avstand både fra enkle kausale modeller av hvordan sykdom oppstår, og fra omfattende og vage «holistiske» sykdomsteorier. Hun mener også at det er feil å tenke dualistisk, som i kropp–sinn-, natur–kultur- eller biologi–samfunn-dikotomier. I stedet foreslår hun en integrativ og økologisk forståelse av sykdomsforekomst knyttet til historisk og plassmessig kontekst. Hennes teori omfatter det hun kaller *embodiment*, hvordan ytre hendelser og omgivelser «in-korpo-reres» eller kroppsliggjøres; de ulike mekanismene for hvordan ytre påvirkning «kommer under huden»; det kontinuerlige samspillet det er mellom eksponering, sårbarhet og motstandskraft over livsløpet og betydningen av *timing* og akkumulering av påvirkning. Endelig omfatter hennes teori plassering av ansvar og handlingsmulighet når det gjelder ulike helsepåvirkninger (se også kapittel 10).

Flernivåanalyser

Studier av befolkningshelse kan skje på makronivå, for eksempel ved å bruke nasjoner, regioner eller lokalsamfunn som analyseenheter, såkalte *økologiske* studier (kapittel 5 og 8). Data kan enten være karakteristika ved enhetene, som for eksempel inntektsfordeling, velferdspolitikk eller bosettingstetthet; eller aggregerte individdata, som dødelighet eller gjennomsnittlig selvrapportert helse. Noen ganger kan man studere helseutviklingen i et område over tid for å spore eventuelle effekter av politiske eller strukturelle endringer, såkalte *naturlige eksperimenter* (kapittel 5). Mer vanlig er det å se på mulige sammenhenger mellom sosiale eller samfunnsmessige karakteristika på den ene siden og helseutfall på den andre siden i et utvalg av geografiske enheter. Slike *økologiske korrelasjoner* gir ikke entydige svar på hvordan sammenhengene er på individnivå, men kan gi grunnlag for hypoteser som så kan utforskes gjennom andre studiedesign.

Noen ganger tilhører de uavhengige variablene ulike analytiske nivåer. En studie av utdanningssystemers virkninger på elevenes helse kan for eksempel tenkes å inkludere data på individnivå, på klassenivå, på skolenivå, på regionnivå og på nasjonalt nivå. Nyere statistiske metoder, såkalte *flernivåanalyser*, gir muligheter for å håndtere så kompleks informasjon i en felles analysemodell (kapittel 5).

Befolkningsorientert

I 1985 publiserte den britiske epidemiologen *Geoffrey Rose* (1926–1993) en banebrytende artikkel: «Sick individuals and sick populations» (Rose 1985). Her pekte Rose på at det er en fundamental forskjell mellom spørsmålet: *Hvem får sykdommen X*? og spørsmålet: *Hvorfor får så mange sykdommen X*? I jakten på sykdomsforklaringer blandes disse spørsmålene ofte sammen, men de to spørsmålene krever ulike forklaringer. Mens det første spørsmålet dreier seg om *individuell risiko* for sykdommen, dreier det andre spørsmålet seg om utbredelsen av sykdommen i samfunnet. For å besvare det første spørsmålet om individuell risiko, må man konsentrere seg om individuelle egenskaper. Det andre spørsmålet om sykdomsutbredelse må besvares ut fra egenskaper ved befolkningsgrupper og omgivelser, det som Rose kaller «the causes of the causes».

I boken *The Strategy of Preventive Medicine* (1992) utdyper Rose sine argumenter for en befolkningsorientert forebyggende modell. Han peker på at de aller fleste kjente individuelle risikofaktorene er kontinuerlige og graderte, uten klare sprang i risiko mellom én verdi og en nærliggende verdi. Sykdomsforekomsten i befolkningen henger nøye sammen med gjennomsnittsverdien for slike risikofaktorer. Selv små endringer i befolkningsgjennomsnittet påvirker fremtidig sykelighet og dødelighet langt mer enn man kan oppnå gjennom å lete etter og forsøke å påvirke de relativt få med høye nivåer av risikofaktoren. Rose argumenterte overbevisende for at befolkningens risikofaktornivå påvirkes best gjennom kollektive politiske og sosiale tiltak (kapittel 11). Lignende konklusjoner kom en samtidig kanadisk tenkegruppe frem til (Evans mfl., 1994).

Sosial påvirkning på tvers av sykdomsgrupper

Mens tradisjonell epidemiologi gjerne tar utgangspunkt i én bestemt type sykdom og leter etter sykdomsspesifikke årsaksforklaringer, er sosialepidemiologi mer opptatt av hvordan sosiale forhold bidrar til helse generelt. Sosialepidemiologi bruker derfor ofte *generiske* mål for helse, som total dødelighet, sykefravær, uførhet og egenvurdert helse. Et nyere studiefelt er *multimorbiditet*, dvs. tilstedeværelse av to eller flere kroniske sykdomstilstander samtidig (kapittel 2 og 10). Ved å bruke slike endepunkter bidrar sosialepidemiologi til å avdekke felles årsaksmekanismer til klinisk sett ulike sykdommer. Men selv om mange av de sosiale helsedeterminantene ser ut til å påvirke de fleste større sykdomsgruppene på samme måte (om enn ofte i ulik grad), er det også viktig å lete etter sykdommers sosiale særtrekk. Mange sosialepidemiologiske studier tar derfor utgangspunkt i enkeltsykdommer for å kartlegge spesifikke sosiale årsaksforhold (kapittel 3).

Teori styrer empiri

Mens tradisjonell epidemiologi er blitt kritisert for ukritisk bruk av multivariat statistikk i jakten på risikofaktorer (såkalt «black box»-epidemiologi), bruker sosialepidemiologi teori aktivt både for å fortolke funn, begrunne hvilke variabler som skal inngå i analysene, og lage modeller for hvordan den antatte sammenhengen er mellom dem. Sosialepidemiologi henter mye av sitt teoritilfang fra samfunnsvitenskapene, men er også inspirert av nyere økologisk teori og evolusjonsteori. Dette gjør sosialepidemiologi til et utpreget tverrfaglig forskningsfelt, hvor komplementær kompetanse er nødvendig.

1.4 Om å finne (sosiale) årsaker til sykdom og helsesvikt

Det er velkjent at en statistisk sammenheng mellom X og Y ikke nødvendigvis betyr at X forårsaker Y. Sammenhengen kan for eksempel skyldes at både X og Y påvirkes av en tredje faktor Z, som da gjerne kalles en *effektforveksler* eller *konfunder*. Hvis det ikke er sikker tidsforskjell mellom X og Y, kan sammenhengen også skyldes at Y påvirker X (*revers kausalitet*) eller at X og Y gjensidig/samtidig påvirker hverandre.

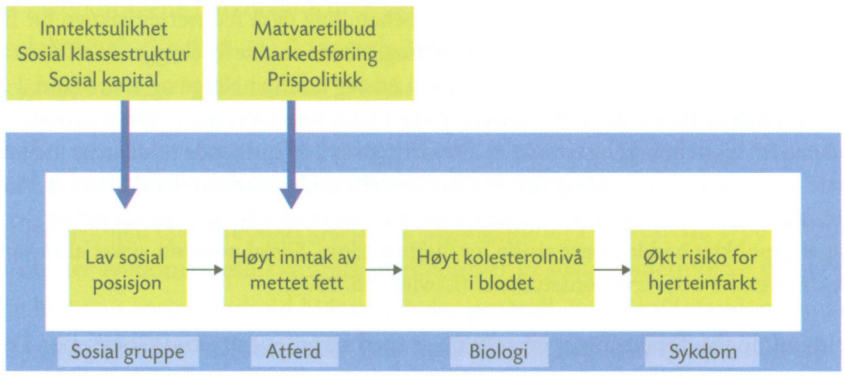
Tidlig på 1990-tallet ble det publisert flere studier som viste lavere risiko for hjerte- og karsykdommer blant kvinner som brukte østrogen-tabletter i overgangsalderen. Andre studier tydet på at slik hormonsubstitusjonsbehandling også reduserte risikoen for fremtidig osteoporose, depresjon og demensutvikling. Denne antatte forebyggende effekten førte til at antallet brukere av slike preparater på få år steg kraftig i Norge og andre land. I 2002 kom resultatene fra en stor randomisert studie i USA som tvert imot viste at østrogen *økte* risikoen for hjerteinfarkt og brystkreft. Forklaringen på de sprikende resultatene mellom de første observasjonsstudiene og den eksperimentelle undersøkelsen skyldes nok at kvinner som brukte hormonsubstitusjon, oftere kom fra høyere sosiale lag i samfunnet og dermed i utgangspunktet hadde lavere risiko for hjertesykdom. Sosial posisjon virket altså som en effektforveksler i observasjonsstudiene (Lawlor mfl. 2004).

Epidemiologisk forskning opererer videre med sammensatte årsaksbegreper. I noen tilfeller er en faktor en *nødvendig* årsak til en sykdom, i andre tilfeller kan forklaringen på en sykdom deles mellom flere *bidragende* årsaker. Selv når en nødvendig årsak er identifisert (som for eksempel tuberkelbakterien for å utvikle tuberkulose), er den, som vi har sett, sjelden tilstrekkelig i seg selv til å produsere sykdom. Som oftest vil en rekke bidragende årsaker, hvorav de aller fleste er ukjente, medvirke. Begrepet «årsaksnettverket» («the web of causation») brukes ofte for å karakterisere slike multikausale påvirkninger. De enkelte delårsakene kan enten virke *additivt* eller gjensidig forsterkende (*synergistisk*). I det siste tilfellet snakker vi gjerne om interaksjonseffekter.

Årsakskjeder

Sykdomsårsakene kan være forbundet i sekvensielle kjeder, slik at noen årsaker formidler *(medierer)* effekten av bakenforliggende faktorer. I slike årsakskjeder kan man noen ganger snakke om distale, intermediære og proksimale årsaksfaktorer. Med *proksimale* (nærliggende) årsaker menes faktorer som tidsmessig eller fysisk ligger nær opp til sykdomsmanifestasjon i en kausalkjede, altså faktorer som bidrar til å utløse sykdommen. Slike proksimale faktorer er vanligvis individuelle biologiske eller atferdsmessige karakteristika. Med distale determinanter menes enten faktorer som ligger utenfor individet, eller faktorer som opptrer tidligere i kausalkjeden («årsakene til årsakene» eller «rotårsakene»). Slike faktorer er gjerne sosiale eller strukturelle forhold.

Mange ulike forhold bidrar til å øke risikoen for et hjerteinfarkt. Høyt nivå av kolesterol i blodet er én slik proksimal individuell risikofaktor (figur 1.3). I tillegg til arvelige faktorer påvirkes kolesterolnivået av mengde og type fett i maten personen typisk spiser (Helsedirektoratet 2010). Høyt inntak av mettet fett er en intermediær risikofaktor for hjerteinfarkt. Folks kosthold påvirkes av hva andre i det nære nettverket spiser, og personer tilhørende lavere sosioøkonomiske grupper har høyt inntak av mettet fett. Endelig vet vi at sosiale ulikheter henger sammen med fordelingen av makt, ressurser og sosial kapital i samfunnet og at folks kosthold reflekterer markedsføring, matvaretilbud og priser. Slike samfunnsmessige forhold er distale risikofaktorer for hjerteinfarkt i denne årsakskjeden.



Figur 1.3 En kausalmodell for sammenhengen mellom sosial posisjon, matvaner og hjerteinfarkt. De individuelle risikofaktorene (inni rammen) påvirkes av samfunnsmessige forhold knyttet til sosial lagdeling og matvareforbruk.

Direkte og indirekte effekter

Effekten av en sosial helsedeterminant kan være enten *direkte* eller *indirekte*. En indirekte effekt medieres gjennom en mer proksimal faktor i en kausalkjede, altså en kjent individuell risikofaktor. Direkte effekter er slike som verken skyldes effekter av kjente, mer proksimale variabler eller effekter av bakenforliggende (distale) faktorer. De er altså effekten av den sosiale helsedeterminanten i seg selv.

Skillet mellom indirekte og direkte effekter av sosiale helsedeterminanter er viktig av flere grunner. For det første bidrar økt kunnskap her til bedre forståelse av hvordan sosiale faktorer utøver sin virkning på helse. Dernest kan denne kunnskapen brukes til en mer effektiv strategi for sykdomsforebygging. Hvis effekten av en sosial faktor hovedsakelig formidles gjennom kjente mellomliggende faktorer, vil gjerne disse by seg naturlig frem som kandidater for påvirkning. Hvis for eksempel mesteparten av sosial ulikhet i risiko for død skyldes ulikheter i tobakksrøyking, vil det være viktig å intensivere arbeidet for å redusere tobakksbruk blant de sosiale gruppene som røyker mest. Hvis sosial posisjon påvirker helseutsiktene direkte, må en strategi for å redusere sosiale helseforskjeller også omfatte arbeid for å motvirke sosial ulikhet (kapittel 11).

Eksponering, sårbarhet og beskyttelse

Sosiale faktorer kan påvirke sykdomsrisiko på flere måter (Krieger 2014, Diderichsen mfl. 2019). Sosiale omstendigheter kan virke sykdomsfremkallende gjennom mengde og varighet av helseskadelig *eksponering*, som når diskriminering påfører etniske minoriteter vedvarende stress og utrygghet (kapittel 6). Sosiale forhold kan videre gjøre folk *mer sårbare* for andre helserisikoer. Høyt alkoholforbruk gir større helseskader hos mennesker lavt på den sosiale rangstigen enn tilsvarende forbruk blant personer med høy status. I økonomiske nedgangstider er det gjerne ufaglærte og marginale arbeidstakere som først mister jobben, med de helsekonsekvenser dette gir.

Endelig kan noen sosiale faktorer gi *beskyttelse* mot ytre helsetrusler. Et velfungerende sosialt nettverk kan hjelpe folk til bedre å mestre sykdom og skader (buffereffekt, se kapittel 5). Høy sosial status kan gi folk lettere tilgang til avansert medisinsk behandling (kapittel 2).

Gjennom livet påvirkes organismen på mange ulike måter av de sosiale omstendighetene som et samspill mellom eksponering, sårbarhet og motstandskraft. Dette kan lettest forstås gjennom detaljerte individuelle livshistorier (se kapittel 10), men kan også avleses på befolkningsnivå ved å følge ulike sosiale gruppers helseutvikling over tid (kapittel 8 og 11).

Sosial årsak eller sosial seleksjon?

Til nå har vi tenkt på et helseutfall som siste ledd i en årsakskjede. For fenomenet død lar kausalmodellen seg avgrense på denne måten, men for andre indikatorer på helsesvikt må vi også ta i betraktning muligheten for *revers kausalitet*, det vil si at helsen påvirker personens sosiale situasjon. Dette kalles gjerne for en *seleksjonseffekt* eller sosial mobilitet ved at god helse bidrar til at personen avanserer i det sosiale hierarkiet, mens dårlig helse fører til sosial degradering. Selv om slik revers kausalitet er til stede i noen grad, tyder empirisk forskning på at dette langt fra er tilstrekkelig for å forklare de sosiale helseforskjellene man observerer (Dahl & Elstad 2009, se også kapittel 2).

1.5 Kroppsliggjøring av sosiale helsepåvirkninger

Forståelsen av hvordan ytre påvirkninger nedfeller seg i og påvirker menneskers biologi har økt vesentlig de siste 20–30 årene (Kubzansky mfl. 2014). Selv om stadig flere kunnskapsbiter kommer på plass, er det likevel ennå lenge til vi har et fullstendig bilde av hvordan sosiale forhold bidrar til sykdom og for tidlig død. Mange mekanismer og biologiske prosesser inngår, og kunnskapsutviklingen på dette området må derfor basere seg på bidrag fra en rekke spesialiserte forskningsfelt (se kapittel 9 og 10). Her skal vi trekke frem tre forskningsområder som har gitt økt innsikt i hvordan ytre omstendigheter nedfeller seg i kroppen:

Stressforskningen: fra homeostase til allostatisk belastning

Denne systemorienterte forskningen springer ut av ideen om organismens iboende tendens til å gjenopprette en indre balanse når den blir forstyrret, det såkalte homeostatiske prinsipp. Dette skjer gjennom negative tilbakekoblingssløyfer i kroppens biologiske og fysiologiske apparat.

I 1970-årene påviste stressforskere at langvarig aktivering av dette apparatet gjennom gjentatte eller kroniske belastninger kunne føre til mer vedvarende endringer i kroppen. Mange slike endringer er til nytte for organismen (adaptive endringer), mens andre øker risikoen for senere sykdom (maladaptive endringer).

Langvarig aktivering av den indre reguleringen av blodtrykket som følge av ytre stimulering kan for eksempel føre til vedvarende forhøyet blodtrykk, en risikofaktor for hjerte- og karsykdom.

Begrepet *allostase*, som betyr stabilitet gjennom endring, ble etter hvert knyttet til slike adaptive og maladaptive endringer. Ofte bruker begrepet *allostatisk (over) belastning (allostatic load/overload)*, som uttrykk for vedvarende og potensielt sykdomsskapende endringer i kroppens biologiske systemer som følge av gjentatte og kumulative ytre påvirkninger (McEwen 2000). Slike endringer skjer i det nevroendokrine systemet, det autonome nervesystemet, immunsystemet og det metabolske systemet. Disse kroppssystemene overlapper både perifert og i sentralnervesystemet, og allostatisk belastning er vist å påskynde aldring og sykdom i mange organer (kapittel 10).

Epigenetikken – hvor arv og miljø møtes

Et nytt forskningsfelt handler om de epigenetiske prosessene som regulerer om gener «slås på eller av» som svar på kjemiske signaler fra kroppen. Slike signaler er en del av den indre reguleringen, men kan også gjenspeile ytre påvirkning, som tilgang på ernæring eller graden av oppfyllelse av andre biologiske og psykologiske behov.

Slike epigenetiske endringer påvirker genuttrykket, altså i hvilken grad gener stimulerer til spesifikk proteinproduksjon i cellene. Det er slike prosesser som forklarer spesialiseringen av cellene under den føtale utviklingen og senere endringer i cellers funksjon og aktivitet. Det er vist at mange former for ytre påvirkning, som næringstilførsel, stress, infeksjoner og toksiner, fører til epigenetiske endringer (Cunliffe 2016). Hva mer er, slike epigenetiske adaptasjoner kan også skje i kjønnsceller, noe som kan være en mekanisme for rask tilpasning til ytre miljø gjennom overføring av ervervede endringer i arvematerialet til neste generasjon.

Epigenetiske endringer antas å gi kroppen mulighet for svært fleksible biologiske tilpasninger til endringer i det ytre miljøet. Selv om slike mekanismer er aktive hele livet, er det påvirkninger tidlig i livet som får de største konsekvensene for fremtidig funksjon og helse (kapittel 10).

Nevrovitenskapen: utvikling og plastisitet

Vårt syn på hjernen har endret seg mye de siste tiårene. Særlig i de første leveårene, men også senere i livet, påvirkes hjernens struktur og funksjon av ytre faktorer (McEwen & Gianaros 2010). Dette kalles *nevroplastisitet*. Slike endringer er ofte reversible og indikerer ikke nødvendigvis noen skade, men er et uttrykk for en dynamisk tilpasning til krav fra det ytre miljø. Samtidig er det vist at hjerneutviklingen

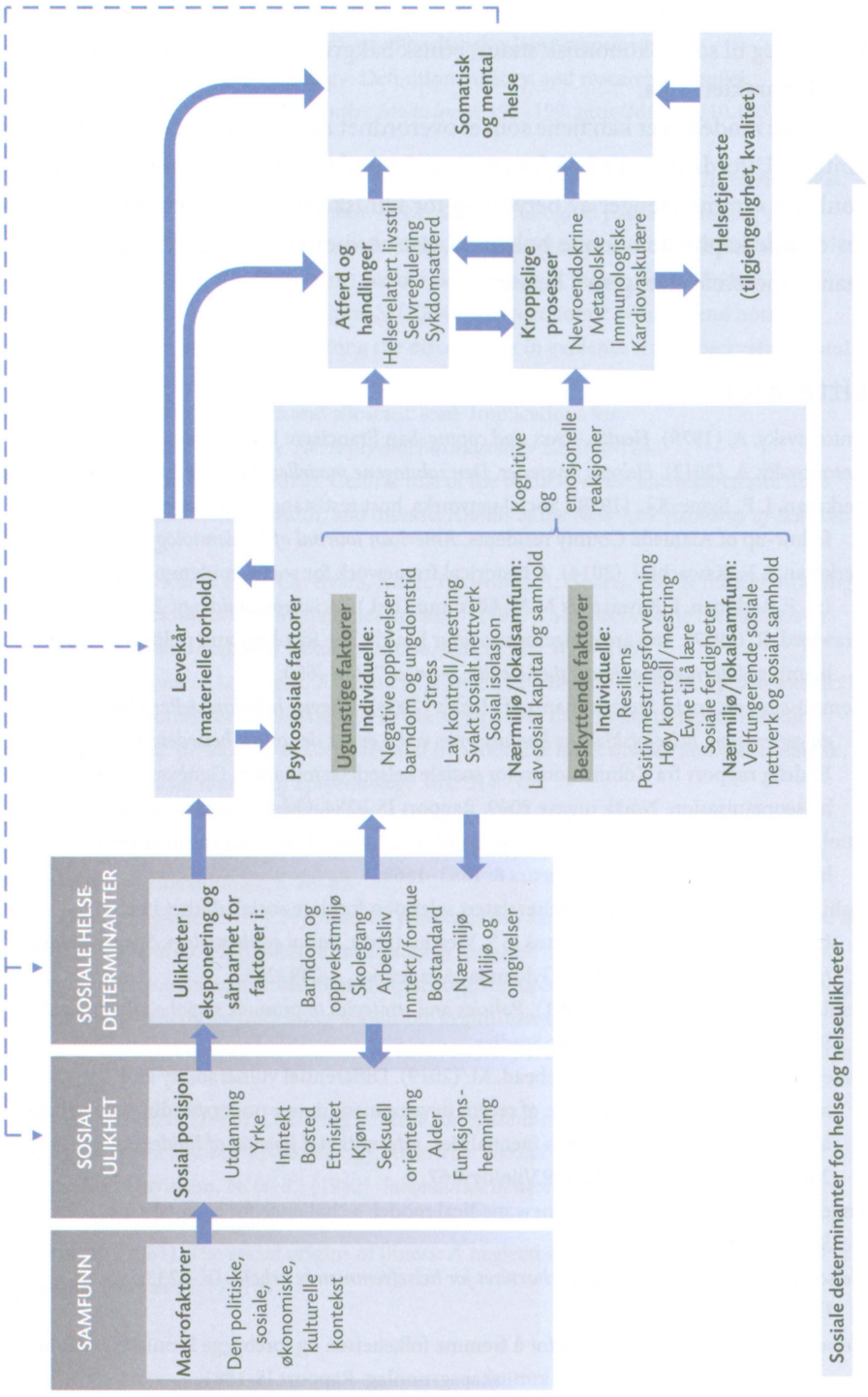
er spesielt sårbar tidlig i livet. Omsorgssvikt og traumatiske erfaringer i de første leveårene kan gi påvisbare strukturelle hjerneendringer, noe som kan bidra til å forklare den økte risikoen for senere mentale problemer og for kroppslig sykdom blant dem med slike erfaringer (kapittel 4 og 10).

En annen ny erkjennelse er kommet innen psykoendokrinologi, som følge av påvisningen av ulik hormonell innflytelse på hjernen. Vedvarende stressutløst hormonell aktivisering kan føre til skade av hjerneceller i hippocampus og amygdala, to hjerneområder som er viktige for regulering av affekt og hukommelse. Mens kortvarige truende situasjoner kan føre til adaptive endringer i form av læring og økt mestring, kan vedvarende stress gi kognitiv svekkelse gjennom ulike mekanismer, noen reversible og andre irreversible (kapittel 10).

1.6 Mot en sammenfattende biopsykososial sykdomsmodell

Den vitenskapelige utviklingen de siste tiårene har ført oss atskillig nærmere Georg L. Engels visjon om en biopsykososial sykdomsmodell (Engel 1977). Engel argumenterte for et systemperspektiv på sykdom, hvor sykdom forstås som resultat av både biologiske, psykologiske og sosiale faktorer og hvor sykdomsskapende faktorer på mange nivåer tas i betraktning. Den parallelle utviklingen de siste tiårene innen sosialepidemiologi, nevrovitenskap, stressforskning og molekylærbiologi har åpnet opp for en slik multidimensjonal og dynamisk forståelse av hvordan ytre forhold nedfeller seg i kroppen som sykdom (kapittel 10).

Det er vanskelig å sammenfatte så mange samvirkende faktorer i en felles årsaksmodell for sosiale helsepåvirkninger. Figur 1.4 viser likevel et slikt forsøk. Figuren bygger på det rammeverket som WHOs kommisjon for sosiale helsedeterminanter la til grunn for sitt arbeid (Commission on Social Determinants of Health 2008), men er videreutviklet av Public Health England (2017). Her knyttes samfunnsfaktorer på makronivå sammen med sosiale helsedeterminanter på mesonivå og videre til psykososiale og biologiske faktorer på mikronivå i påvirkningen av fysisk og mental helse. Figuren viser at sosiale faktorer på flere nivåer bidrar til sykdom og helsesvikt gjennom ulike kausale baner og mekanismer.



Figur 1.4 Et rammeverk for sosiale helsedeterminanter og sosial ulikhet i helse. Kilde: Public Health England 2017.

Modellen gir også et rammeverk for å forstå sosiale ulikheter i helse, enten disse knytter seg til sosioøkonomisk status, etnisk bakgrunn, kjønnsidentitet eller andre sosiale karakteristika.

Denne modellen er kan tjene som et overordnet rammeverk for det denne boken omtaler. Utfordringen er å gå fra en så omfattende og dynamisk modell til å finne konkrete sammenhenger av betydning for klinisk arbeid og folkehelsearbeid. De resterende kapitlene i denne boken vil vise at vi er kommet godt i gang, men at mange spørsmål står igjen å besvare.

Litteratur

Antonovsky, A. (1979). *Health, stress and coping*. San Francisco: Jossey-Bass.

Antonovsky, A. (2012). *Helsens mysterier. Den salutogene modellen*. Oslo: Gyldendal Akademisk.

Berkman, L.F., Syme, S.L. (1979). Social networks, host restistance and mortality: A nine-year follow-up of Alameda County residents. *American Journal of Epidemiology* 109: 186–204.

Berkman, L.F., Kawachi, I. (2014). A historical framework for social epidemiology. I: L.F. Berkman. I. Kawachi & M.M. Glymour (red.) *Social epidemiology*. 2. utgave, s. 1–16.

Crawford, R. (1977). You are dangerous to your health: The ideology and politics of victim blaming. *International Journal of Health Services* 7: 663–680.

Commission on Social Determinants of Health (2008). *Utjevne helseforskjeller i løpet av en generasjon: helseforskjellene kan utjevnes ved å endre de sosiale helsedeterminantene*. Endelig rapport fra Kommisjonen for sosiale helsedeterminanter. Genève: Verdens helseorganisasjon. Norsk utgave 2009, Rapport IS-0254. Oslo: Helsedirektoratet.

Cunliffe, V.T. (2016). The epigenetic impacts of social stress: how does social adversity become biologically embedded? *Epigenetics* 8: 1653–1669.

Dahl, E., Elstad, J.I. (2009). Kan helserelatert seleksjon forklare sosial ulikhet i helse? I: J.G Mæland, J.I Elstad, Ø. Næss & S. Westin. (red.). *Sosial epidemiologi. Sosiale årsaker til sykdom og helsesvikt*. Oslo: Gyldendal Akademisk, s. 249–265.

Dahlgren, G., Whitehead, M. (1991). *Policies and strategies to promote social equity in health*. Stockholm: Institute for Future Studies.

Diderichsen, F., Hallqvist, J., Whitehead, M. (2019). Differential vulnerability and susceptibility: how to make use of recent development in our understanding of mediation and interaction to tackle health inequalities. *International Journal of Epidemiology* 48: 268–274, https://doi.org/10.1093/ije/dyy167.

Engel, G.L. (1977). The need for a new medical model: a challenge for biomedicine. *Science* 196: 129–136.

Helsedirektoratet (1987). *Ottawa-charteret for helsefremmende arbeid*. IK-2233. Oslo: Helsedirektoratet.

Helsedirektoratet (2010). Kostråd for å fremme folkehelsen og forebygge kroniske sykdommer – Metodologi og vitenskapelig kunnskapsgrunnlag. Rapport IS-1881. Oslo: Helsedirektoratet.

https://helsedirektoratet.no/publikasjoner/kostrad-for-a-fremme-folkehelsen-og-forebygge-kroniske-sykdommer-metodologi-og-vitenskapelig-kunnskapsgrunnlag

Honjo, K. (2004). Social epidemiology: Definition, history, and research examples. *Environmental Health and Preventive Medicine 9*: 193–199. http://doi.org/10.1007/BF02898100

Krieger, N. (2014). *Epidemiology and the people´s health*. New York: Oxford University Press.

Kubzansky, L.D., Seeman, T.E., Glymour, M.M. (2014). Biological pathways linking social conditions and health: plausible mechanisms and emerging puzzles. I: L.F. Berkman, I. Kawachi & M.M. Glymour (red.) *Social epidemiology*. 2. utgave, s. 512–561.

Lawlor, D.A., Davey Smith, G., Ebrahim, S. (2004). Socioeconomic position and hormone replacement therapy use: Explaining the discrepancy in evidence from observational and randomized controlled trials. *American Journal of Public Health* 94: 2149–2154.

McEwen, B.S. (2000). Allostasis and allostatic load: Implications for neuropsychopharmacology. *Nevropsychopharmacology* 22: 108–124.

McEwen, B.S., Gianaros, P.J. (2010). Central role of the brain in stress and adaptation: links to socioeconomic status, health, and disease. *Annals of the New York Academy of Sciences* 1186: 190–222.

Mackenbach, J.P., Kunst, A.E., Cavelaars, A.E.J., Groenhof, F., Geurts, J.M. & the EU Working Group on Socioeconomic Inequalities in Health (1997). Socioeconomic inequalities in morbidity and mortality in western Europe. *Lancet* 349: 1655–1659.

Marmot, M.G., Syme, S.L., Kagan, A., Kato, H., Cohen, J.B., Belsky, J. (1975). Epidemiological studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California: prevalence of coronary and hypertensive heart disease and associate risk factors. *American Journal of Epidemiology* 102: 514–525.

Mæland, J.G. (2009). *Hva er HELSE*. Oslo: Universitetsforlaget.

Mæland, J.G. (2016). *Forebyggende helsearbeid. Folkehelsearbeid i teori og praksis*. 4. utgave. Oslo: Universitetsforlaget, s. 76–82.

Public Health England (2017). *Psychosocial pathways and health outcomes: Informing action on health inequalities*. PHE publications, gateway number: 2017209. London: Public Health England. http://www.instituteofhealthequity.org/resources-reports/psychosocial-pathways-and-health-outcomes-informing-action-on-health-inequalities

Schiøtz, A. (2017). *Viljen til liv. Medisin- og helsehistorie frå antikken til vår tid*. Oslo: Samlaget, s. 28–44.

Susser, M., Susser, E. (1996). Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *American Journal of Public Health* 86: 674–677.

Townsend, P., Davidson, N. (red.) (1982). *Inequalities in health. The Black Report*. Harmondsworth: Penguin Books.

Waitzkin, H. (1981). The social origins of illness: A neglected history. *International Journal of Health Services* 11: 77–103.

Getz, Linn; Kirkengen, Anna Luise & Ulvestad, Elling

"Kap. 10: Kroppsliggjøring av erfaring og mening - belastende erfaringers innvirkning på menneskets biologi. Utdrag: Livserfaringer og biologi - et biomolekylært perspektiv", i (Red) Mæland, John Gunnar

Sykdommers sosiale røtter 2020 ss. 205-221

ISBN: 9788205523944

Copyright (C) 2020 Gyldendal

Kopiert med hjemmel i Kopinor-avtalen

KAPITTEL 10

Kroppsliggjøring av erfaring og mening – belastende erfaringers innvirkning på menneskets biologi

Linn Getz, Anna Luise Kirkengen, Elling Ulvestad

Livserfaringer «innskrives» i menneskekroppen. De siste 20–30 årene har vi fått stadig mer detaljert innsikt i de involverte biologiske prosessene. Den nye kunnskapen stammer blant annet fra forskningsaktivitet i berøringsflatene mellom tidligere atskilte fagområder som psykologi, immunologi, endokrinologi, mikrobiologi, nevrovitenskap og genetikk. Derved belyses med naturvitenskapelige metoder spørsmål vi lenge har tenkt på som «humanistiske», for eksempel hvordan erfaringer av omsorg eller vanskjøtsel, tilhørighet eller ensomhet, rettferdighet eller urettferdighet kan knyttes til vedlikehold av helse eller utvikling av sykdom (Vogt mfl. 2014). Å forvalte den voksende innsikten på en god måte er en stor faglig utfordring (Tenketanken i Allmennmedisin 2013, Kirkengen 2018).

Tidligere i boka omtales flere epidemiologiske studier som dokumenterer *at* sosiale forhold innvirker på menneskets sykelighet – på gruppenivå. I dette kapittelet vil vi beskrive *hvordan* slike sammenhenger kan forklares biologisk – på individnivå. Med utgangspunkt i det enkelte individ er det stadig flere forskere som ser på interaksjon mellom ulike biologiske nivåer (molekyler, celler, organer) og deres samspill med fysiske, biologiske og sosiale miljøfaktorer. Kroppen framstår da som en kompleks, integrert helhet innbakt i den fysiske og biologiske *omverdenen* og i den sosiokulturelle *livsverden* av meningsfulle symbolsystemer, språk og verdier – slik disse gjennomleves og tolkes av det erfarende mennesket.

Kunnskapen på feltet vil utvilsomt bli nyansert og videreutviklet i kommende år, men det overordnede bildet framstår klart: Den menneskelige biologien – hele kompleksiteten av fysiologiske systemer og prosesser – er mettet med erfaring, på godt og vondt (Getz mfl. 2011).

Av historiske og pedagogiske grunner er dette kapittelet inndelt i tråd med det skillet *Charles Snow* i 1959 beskrev mellom «de to vitenskapelige kulturene» (se rammetekst 10.1). Første del tar utgangspunkt i kunnskap utviklet med *naturvitenskapelig* metode, mens andre del tar utgangspunkt i ideer og metoder fra *humaniora*. I tredje del skisserer vi noen medisinske implikasjoner av de to foregående delers utlegninger. Vi vil gjennomgående, og på en pragmatisk og bevisst jordnær måte, forsøke å vise hvorfor *René Descartes'* snart 400 år gamle oppdeling av mennesket i kropp og sinn var en villfarelse, og hvordan vi i praksis kan overskride skillet mellom kropp og sinn slik at kløften mellom de to kulturer forsvinner. Målet er å vise at dualismens kart er utilstrekkelig for å forstå sykdom: Skal vi kunne forstå en persons helse og sykdom, må vitenskapen klargjøre et overordnet, samlende bilde som gjør helheten til noe mer enn summen av delene. Ved det blir det også klart at kropp og sinn er hva de er i kraft av hverandre.

Rammetekst 10.1 Vitenskapsteoretiske skiller

Erkjennelsen av at mennesket må betraktes som en udelt helhet utfordrer flere vitenskapsteoretiske skiller. Ett slikt skille er *dualismen* mellom kropp og sinn, slik dette ble begrepsfestet av filosofen og naturvitenskapsmannen *René Descartes* (1596–1650). Studerer man mennesket langs disse to aksene, mente Descartes, ville man forstå menneskets væremåte bedre. Et annet historisk viktig skille går mellom *naturvitenskapenes* og *humanvitenskapenes* forklaringsmønstre, slik disse ble utlagt av filosofen og psykologen *Wilhelm Dilthey* (1833–1911). Dilthey skilte mellom epistemer (tankesystemer) for å *forklare* og epistemer for å *forstå*; forklarende tilnærminger baserer seg på objektiverende, naturvitenskapelige metoder, mens forstående tilnærminger anvender metoder som erkjenner subjektets egenart og subjektivitetens betydning for den enkeltes erkjennelse og erfaring (Dilthey 1985).

I ettertid ser vi at Descartes' todeling og vitenskapens epistemiske forgrening har gitt utilsiktede og til dels ganske skjebnesvangre følger: Ved Descartes' oppdeling ble mennesket omgjort til noe det ikke er, og ved vitenskapens forgrening ble det legitimt å utforske menneskets væremåte basert på to ulike

og til dels uforenlige perspektiver. Dermed bidro vitenskapen til å befeste, snarere enn å undergrave, Descartes' dualisme. Så befestet kom skillet til å bli at kjemiker og forfatter *Charles Percy Snow* (1905–1980) i 1959 kunne konstaterte – i boken *The Two Cultures* – at naturvitenskapelig og humanistisk forskning var blitt adskilt ved en dyp kløft preget av gjensidig mistillit og mangel på forståelse for hva kunnskapsutvikling innebærer (Snow 1993).

FØRSTE DEL: LIVSERFARINGER OG BIOLOGI – ET BIOMOLEKYLÆRT PERSPEKTIV

Vi vil nå presentere kunnskap som er utviklet med naturvitenskapelige metoder, men som samtidig er basert på en humanistisk forståelse av mennesket som relasjonelt, fortolkende og situert. Blant forskerne bak studiene vi bygger på, finnes fysiologer, leger og psykologer som, uten å bryte med prinsippene for god kvantitativ forskning, bidrar til å «dekonstruere» den kartesianske dualismen.

10.1 Stress og traumer virker inn på hele kroppen

Det velkjente, men samtidig tvetydige begrepet «stress» (kapittel 9) gir en egnet inngangsport til å forstå samspillet mellom livserfaringer og den menneskelige biologien. Det finnes flere definisjoner og modeller for hva stress er og hvordan det påvirker mennesket (rammetekst 10.2), men uansett modell tar tenkningen om stress utgangspunkt i det *erfarende* mennesket. Enhver erfaring «beveger» personen gjennom sansene. Sansningen og den tilhørende fortolkningen (persepsjonen) omsettes til samspill i kroppens fysiologiske nettverk, helt inn til subcellulært nivå. Dermed blir «erfaring i kroppen» et viktig vitenskapelig tema, både for epidemiologien og for basalfagene (Ulvestad 2018a, Ulvestad 2018b, se også andre del).

I vår videre omtale av stress tar vi utgangspunkt i en skjematisk, tredelt stressterminologi der stress kan kategoriseres som enten *positivt* (helsefremmende), *tolerabelt* (tålelig) eller *toksisk* (skadelig). Inndelingen viser til hvordan utfordrende erfaringer påvirker kroppens biologiske tilpasningssystemer, ikke til hva slags hendelser eller forhold som utløser stressresponsene. Modellen var opprinnelig mest opptatt av *barns* utvikling, men framvoksende forskning tilsier at tankegangen har relevans gjennom hele livsløpet.

Rammetekst 10.2 Stressforskning – et historisk riss

Det vi kaller moderne stressforskning, begynte med studier av kroppslige reaksjoner på akutt fare med utgangspunkt i *Claude Bernards* (1813–1878) modell for homeostase. *Walter Cannon* (1871–1945) introduserte begrepsparet «fight or flight». *Hans Seyle* (1907–1982) bygget videre på Cannons forskning og introduserte begrepet «stress». *John W. Mason* (1924–2014) integrerte emosjonelle påkjenninger i forståelsen av stress. *Peter Sterling* og *Josef Eyer* presenterte i 1988 en ny modell kalt «allostasis» som beskrev stressresponser som «stabilitet gjennom endring» (Sterling & Eyer 1988). *Bruce McEwen* og *Elliot Stellar* differensierte modellen for å kunne forklare virkningen av høy belastning over tid, kalt «allostatic load» (McEwen & Stellar 1993). Senere har McEwen begrepsfestet «allostatic overload» for å favne effektene av vedvarende, ufrakommelig overbelastning i form av fysiologisk dysregulering og helseskade. McEwen har senere i samarbeid med *Jack Shonkoff* og *Thomas Boyce* definert tre grader av stress, kalt «positive», «tolerable» og «toxic» stress (Shonkoff, Boyce & McEwen 2009). Toksisk stress dreier seg om ufrakommelig og destruktiv erfaring som er så uttalt og langvarig at hjernens arkitektur og funksjon endres – hvilket lar seg dokumentere med dagens teknologi.

Positivt stress

Positivt eller helsefremmende stress er en nødvendig del av sunn menneskelig utvikling. Her snakker vi om utfordringer som fremmer vekst og mestring. Fysiologisk kjennetegnes positivt stress bl.a. av kortvarig økning i hjertefrekvens og ledsagende svingninger i ulike hormoner. Hverdagserfaringer som typisk kan trigge slike responser hos et barn, kan være å starte i barnehage, få et sprøytestikk, eller miste mor og far av syne på et fremmed sted. Med støtte fra trygge voksne blir dette lærerike erfaringer som bidrar til vekst og utvikling.

Men ingen barn er like, og det som utgjør reelt stress for ett barn, kan oppleves spennende for et annet. Begreper som «sårbarhet», «resiliens» og «robusthet» får økende vitenskapelig oppmerksomhet, med tematisk spennvidde fra oppdragelsesstil til genetikk (Lupien mfl. 2018, se også kapittel 4 og 9). Det som er sikkert, er at alle barn trenger voksne som på en rimelig forutsigbar måte ivaretar deres individuelle behov for grunnleggende trygghet. Dernest er ingen barn tjent med overbeskyttelse.

Tolerabelt stress

Tolerabelt eller tålbart stress utfordrer kroppens tilpasningssystemer i langt høyere grad enn det positivt stress gjør, som følge av mer omfattende og ofte mer langvarige påkjenninger. Eksempler kan være tap av en nær pårørende, alvorlig sykdom, en svært skremmende hendelse eller en avgrenset periode med mobbing (og kombinasjoner av disse). Hvis den kraftige aktiveringen av kroppens alarmsystemer er tidsbegrenset og «bufres» av relasjonene til trygge voksne/andre, vil personen kunne bearbeide erfaringene og bevare god helse, altså «komme helskinnet gjennom det». Tolerabelt stress setter sine spor i personen, og kan komme til å påvirke senere livsvalg. For enkelte kan dyrekjøpte erfaringer bli en kilde til engasjement, mot og kreativitet, for andre kan det lede til unnvikelse og engstelse.

Toksisk stress

Toksisk eller skadelig stress innebærer at individet erfarer sterke, hyppige og/eller langvarige påkjenninger i fravær av tilstrekkelig støtte og hvile. I verste fall er det et barns omsorgsgivere som selv forårsaker belastningene eller bidrar sterkt til dem – enten gjennom aktiv handling eller ved å unnlate å ivareta barnets basale behov. Det kan dreie seg som fysisk, emosjonelt eller seksuelt misbruk, å bli utsatt for vedvarende mobbing, foreldre som ruser seg eller er mentalt syke, og/eller andre former for vanskjøtsel (Felitti mfl. 1998, se også kapittel 4). Noen ganger omfatter belastningen at en nær (omsorgsgiver) eller annen viktig person (for eksempel en lærer) forstår at noe er galt, men likevel velger å se en annen vei og tie. Når det gjelder overgrep og vanskjøtsel, gir det dermed mening å inkludere både «acts of commission» (utøvde handlinger) og «acts of omission» (unnlatte handlinger) (Greenfield 2010).

Over tid vil den kontinuerlige «alarmen» overvelde kroppens robusthet og tilpasningssystemer. Resultatet blir en fysiologisk dysregulering som innvirker både på hjernens arkitektur og kroppens øvrige organer og vev, med økt risiko for kognitive forstyrrelser og annen sykdomsutvikling, både på kort og lengre sikt (Danese & McEwen 2012, Tomasdottir mfl. 2015).

10.2 Kroppenes komplekse tilpasningsmekanismer

Ulike stressforskningsdisipliner forholder seg gjerne til egne modeller. De biomolekylære mekanismene er i utgangspunktet de samme, men perspektivene, terminologien og forskningsspørsmålene kan variere noe. Fagområdet *psykonevroimmunologi*

har de siste 30 år befattet seg med samspillet mellom mentale prosesser og observerbare prosesser i nerve- og immunsystemet (Irwin 2008). Hormonsystemet kan også inkluderes, og da betegnes disiplinen *psykonevroendokrinoimmunologi*. Slik kan man fortsette å bygge opp forståelsen i retning av en mer komplett (integrert) fysiologisk tilnærming (Goldstein 2013).

En anerkjent integrert stressmodell tar utgangspunkt i begrepene *allostase* og *allostatisk belastning* og *-overbelastning*. Modellen har en lengre forhistorie (rammetekst 10.2), men knyttes i dag til nevroendokrinologen *Bruce McEwen* (McEwen 1998, 2012, 2015, 2018, McEwen & Getz 2013, Goldstein 2013, Tomasdottir mfl. 2015, Tomasdottir 2017). Ordet «allostase» betyr «stabilitet gjennom endring» og henspiller på hvordan det fleksible og komplekse samspillet mellom nevrale, endokrine og immunologiske prosesser muliggjør kroppens tilpasning til livets utfordringer og påkjenninger (rammetekst 10.3).

Allostatiske prosesser er grunnleggende forutsetninger for at kroppen kan vedlikeholde et velfungerende biokjemisk og fysiologisk indre miljø (homeostase). All tilpasning «koster» noe i form av allostatisk belastning (*allostatic load*), men hvis påkjenningene er tålbare og personen får anledning til å restituere seg, er påkjenningene ikke helseskadelige. Med dette menes blant annet at erfaringene på lang sikt ikke framskynder de ellers normale aldringsprosessene i organismen, det allostaseforskerne gjerne kaller «wear and tear» (Danese & McEwen 2012).

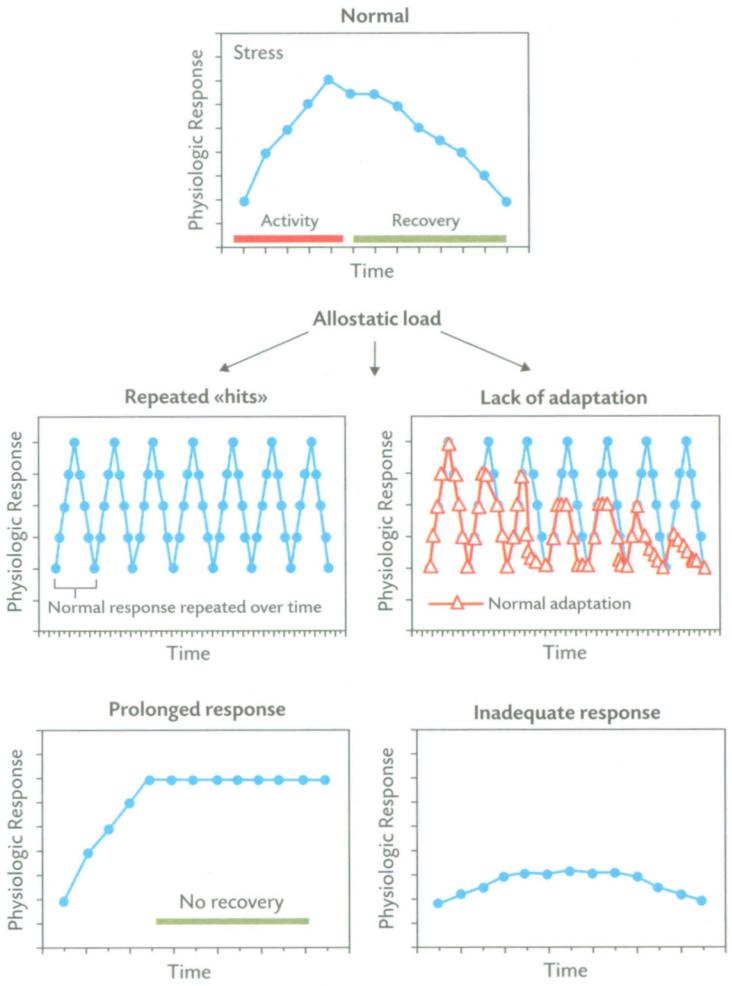
Rammetekst 10.3 Kompleksitet og nettverksteori

«Kompleksitet» er en term man stadig oftere støter på innen biologi og medisin/helsefag (se f.eks. Mazzocchi 2008, Antonsen 2019). Den menneskelige fysiologien er ett eksempel på et såkalt «komplekst adaptivt system». Den teknologiske revolusjonen innen biomedisinsk forskning etter 2000-årsskiftet har muliggjort nye tilnærminger til utforskning av kroppens biologi og ikke minst til omgivelsenes innvirkning på kroppens funksjonalitet. Mens forskerne tidligere måtte analysere et fenomen gjennom å utforske et fåtall molekylære og cellulære prosesser, kan de samme prosessene nå studeres ved bruk av storskala-analyseinstrumenter. Og med det har *nettverksteori* vunnet innpass i biomedisinen.

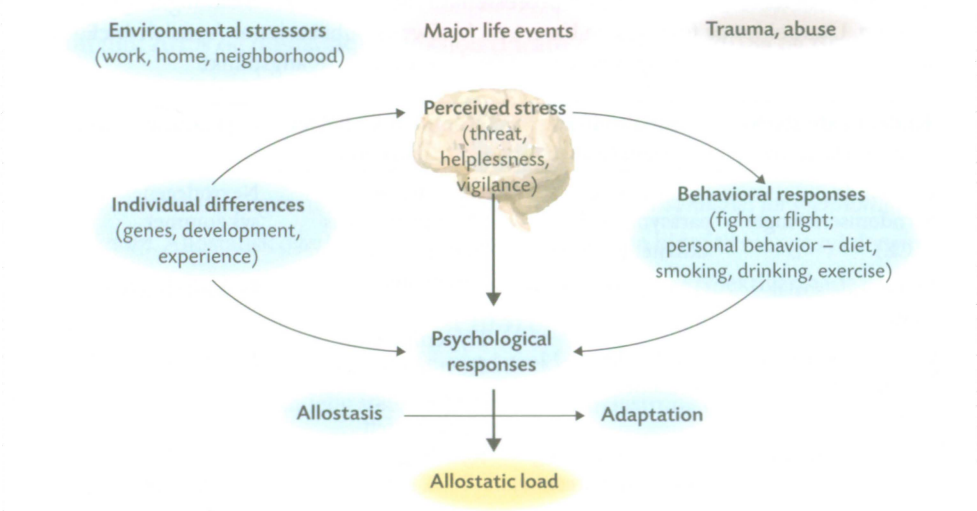
Nettverksteori tar høyde for at kroppens består av dynamiske systemer der de enkelte molekyler og celler innvirker på hverandre med ulik styrke og i ulike relasjoner alt etter beskaffenheten av systemenes omgivelser. Ordet «system» kan i en slik sammenheng referere til alt fra DNA via cellen, personen og familien til hele samfunn. Gjennom analyser basert på *kompleksitetsteori* kan man få forståelse for hvordan systemet makter å tilpasse seg ytre påvirkning, og hva som kreves for at systemet skal framstå som robust – det vil si at det vedlikeholder sin funksjonalitet på tross av påvirkningene. Allostasemodellen bygger på kompleksitetsteori, og robusthet er et sentralt element ved allostasemodellen.

Allostasemodellen illustrerer at alle typer utfordrende erfaringer, både de som tradisjonelt omtales som «mentale» og de som kalles «fysiske», involverer de samme biologiske systemene i menneskekroppen. I sentrum står personens hjerne, i gjensidig samspill med det autonome nervesystemet, hormonsystemene, immunsystemet og cellulære prosesser som regulerer blant annet mitokondrienes aktivitet og det genetiske uttrykket. Dette betyr at opplevd trygghet og kjærlighet, i samspill med inntak av sunn mat, variert fysisk aktivitet og et livskraftig ytre miljø innvirker på de samme biologiske systemene som det opplevelser av ensomhet, frykt, integritetskrenkelse og overveldende omsorgsbyrder gjør, men også som det støy, usunn kost og røyking bidrar til. En rekke forskningsmiljøer har de senere år anvendt allostasemodellen i forskning på sosial ulikhet og helse (se for eksempel Christensen mfl. 2018).

Allostatisk *belastning* er en del av livet selv – slik positivt (helsefremmende) stress og normal aldring er. Det som er mest relevant for oss, er allostatisk *over*belastning. Begrepet uttrykker hvordan *utålelige* nivåer av stress på lengre sikt vil «overbeskatte» de fysiologiske tilpasningsmekanismene og bane vei for sykdom (figur 10.1). Basalforskerne bak allostasebegrepet gir tre essensielle eksempler på slike overbelastende situasjoner: det å føle seg truet, det å oppleve hjelpeløshet og det å være i konstant alarmtilstand eller på vakt (figur 10.2).



Figur 10.1 Illustrasjon fra Bruce McEwens sentrale publikasjon (1998) som skjematisk introduserte ulike typer allostatisk belastning og stressrespons. Øverst skisseres en normal stressrespons som starter, vedlikeholdes en viss tid, og dernest avløses av hvile og restitusjon – på liknende måte som den naturlige døgnrytmen for binyrebark-hormoner. De øvrige bildene viser variasjoner av langvarig stress; repeterte korte responser uten (venstre) eller med (høyre) tilpasning; forlenget stress-respons uten restitusjon; og til sist manglende, eventuelt utmattet, evne til adekvat stressrespons.



Figur 10.2 Figur fra artikkelen «Brain on Stress: How the social environment gets under the skin» av stressforskeren Bruce McEwen (2012).

Ulike trinn i allostatisk overbelastning

Noe forenklet kan effekten av allostatisk belastning og overbelastning «avleses» som primære, sekundære og tertiære fenomener i en utviklingsprosess som harmonerer godt med medisinsk terminologi og klinisk praksis (tabell 10.1).

Det primære (initiale) trinnet i allostatisk overbelastning beskrives på nivå av *signalstoffer*. Det dreier seg om hormoner, cellulære stoffer og nevrotransmittorer som alle svinger mye i organismen. Måling av slike molekyler er krevende og skjer foreløpig mest innenfor rammene av laboratorieforskning. Av hormoner er det spesielt de som inngår i HPA-aksen (Hypothalamus-hypofyse-binyre-aksen) og binyrebarkhormonet kortisol som har vært studert. Andre forskningsmiljøer fokuserer på det autonome nervesystemet, inkludert binyremargen og produksjon av adrenalin og noradrenalin (Goldstein 2013).

Tabell 10.1 En forenklet oversikt over prosessen fra allostatisk overbelastning til sykdomsutvikling. Utarbeidet av forfatterne basert på McEwens publikasjoner (spesielt McEwen 2015) og annen basalforskning med samsvarende, teoretisk utgangspunkt.

Kilder til allostatisk overbelastning	Initiale effekter, signalstoffer	Kumulative effekter, markører	Sykdomsutvikling
Belastende barndomserfaringer (ACE)	Sympatisk og parasympatisk aktivitet (A, NA)	Hypertensjon Økt pulsfrekvens Redusert hjerterate-variabilitet	Nevrodegenerative sykdommer
Manglende sosial støtte	HPA-aksen, kortisol		Psykiske lidelser
Økonomisk utrygghet, deprivasjon, fattigdom	Metabolske hormoner	Hyperlipidemi Høyt blodsukker Vektøkning Ugunstig fettfordeling	Hjerte- og karsykdom
Tungt omsorgsansvar	Cytokiner (pro- og antiinflammatoriske)		Metabolske sykdommer
Tap og samlivsbrudd	Nevromodulatorer i CNS	C-reaktivt protein Fibrinogen, IL-6	Immunologiske sykdommer
Flukt, krig, tortur, traumer	Andre ...	Endret arkitektur i hjernen	Kreftsykdommer
Fysisk slit/overbelastning Inaktivitet		Telomerlengde Epigenetiske markører Mitokondriemarkører	Svangerskapskomplikasjoner, dødfødsler
Uheldig kosthold			Prematur aldring
Forurensning, røyking, medikamenter, rusmidler		Økt infeksjonstendens Patologisk celledeling	Annet
		Endret muskeltonus Økt smerteperseptivitet	
		Andre ...	

Det sekundære trinnet i allostatisk overbelastning beskrives gjennom markører som gjenspeiler høy allostatisk belastning på kroppen over lengre tid (kumulative effekter). Svingningene i disse markørene er relativt langsomme, og de lar seg lettere måle. Blant slike markører finner vi etablerte medisinske risikofaktorer som ofte kartlegges i klinisk sammenheng. Sentralt står blodsukkerstoffskiftet (langtidsblodsukker, insulinmetabolisme), hjerte- og karsystemet (blodtrykk, blodlipider, pulsfrekvens) og kroppsmål (vekt, fettfordeling). I eksperimentell sammenheng registreres også andre biomarkører, blant annet vagusnervens rolle (gjenspeilt i hjerterate-variabilitet), graden av systemisk inflammasjon (Williams mfl. 2019), antistoffnivå etter vaksinasjon

(Marketon & Glaser 2008, Gallagher mfl. 2008), telomerlengde (O'Donovan mfl. 2012; Zalli mfl. 2015) og mitokondriefunksjon (Picard & McEwen, 2018a og b).

Det tertiære trinnet i allostatisk overbelastning representeres av patologi i form av etablert *sykdom* (tabell 10.1). Vi hevder *ikke* at allostatisk overbelastning kan forklare all sykdom på individnivå. Poenget er at allostatisk overbelastning underminerer organismens robusthet og evne til selvhelbredelse, med *økt sannsynlighet for gjennombrudd av sykdom*. Allostatisk overbelastning kan på denne måten bidra til mange sykdommer. Dette innebærer at allostasemodellen gir nye vitenskapelige perspektiver, både på fenomenet «multimorbiditet» (rammetekst 10.4) og på den sosiale gradienten i sykdom. På bakgrunn av solid forskning kan vi nå trekke et direkte resonnement fra sosial deprivasjon via helseskadelig stress, allostatisk overbelastning og fysiologisk dysregulering til kronisk sykdom og multimorbiditet i samsvar med epidemiologiske studier (Tomasdottir 2017, Sturmberg mfl. 2017, Barnett mfl. 2012, se også kapittel 2–7 og 9).

Allostasemodellen er godt utviklet teoretisk, men ikke ferdig operasjonalisert med tanke på hvilke målinger som er best egnet til å avklare grad av allostase. Ny teknologi for storskalaanalysering av biologiske molekyler og celler vil sannsynligvis gi mulighet for mer sofistikerte målinger, både i øyeblikket og gjennom livsløpet. Det viktigste å formidle nå, er at allostase-begrepet kan bidra til en mer helhetlig, vitenskapelig forståelse av helse og sykdom i et sosialmedisinsk perspektiv.

Rammetekst 10.4 Allostatisk overbelastning og multimorbiditet

Vi definerer her *multimorbiditet* som samtidig forekommende, kroniske lidelser fra to eller flere sykdomsgrupper. Disse kan typisk omfatte hjerte- og karsykdommer, diabetes, inflammatoriske sykdommer, osteoporose, sykelig overvekt, demens, depresjon, angst, rusproblemer, smertesyndromer, kreftsykdommer, utmattelse, atferdsforstyrrelser og andre lidelser som i dag klassifiseres som nevrologiske eller psykiatriske.

Kronisk dysregulering av kroppens tilpasningssystemer som følge av allostatisk overbelastning kan forklare mye av den sosiale gradienten i multimorbiditet (figur 2.5, side 47). I et *intergenerasjonelt livsløpsperspektiv* finnes også en sammenheng mellom stress hos gravide (definert ut fra lav sosioøkonomisk status og/eller allostatisk belastning) og uheldige svangerskapsutfall i form av preeklampsi, premature fødsler, kompliserte forløsninger og lav fødselsvekt (Halfon mfl. 2014, se også kapittel 3). Stress i graviditeten er også forbundet med endret stressregulering hos det nyfødte barnet (Santos mfl. 2019; Stenz mfl. 2018). Vi kommer tilbake til dette under omtalen av epigenetikk senere i dette kapittelet.

10.3 Hjernen formes gjennom livsløpet

I sentrum av allostasemodellen står hjernen, menneskets mest komplekse og sårbare organ. Biomedisinen har gjennom tidene gitt inntrykk av at hjernen er et autonomt organ som foretar *egne* tolkninger. Men dette er feil – hjernen er et sosialt og lærende organ som utvikler seg i samspill med erfaringer gjennom hele personens livsløp. Om det ikke var slik, ville det ikke være mulig lære nye ting helt til vi dør. Denne formbarheten (plastisiteten) er av spesielt stor betydning tidlig i livet. Barneforsker *Jack P. Shonkoff* har sagt det slik (referert i Getz & Vogt 2013, se også kapittel 4):

> «Noen av de mest essensielle erfaringer som bidrar til å forme strukturen i den voksende barnehjernen, er 'gi- og ta-relasjonen' mellom barn og de viktige voksne i deres liv.»

Forskere arbeider nå med å identifisere utviklingsperioder der hjernen kan være spesielt formbar – hvilket på den ene siden vil innebære økt sårbarhet for stress, og på den annen side kan innebære muligheter for å kompensere for tidligere skjevutvikling. Begreper som «timing», «sensitive» og «kritiske» perioder står stadig mer sentralt i forskningen på sammenhenger mellom livserfaringer og biologi (Kuzawa & Thayer 2011; Halfon mfl. 2014, Lupien mfl. 2018, se også kapittel 3).

Psykiateren *Martin Teicher* ved Harvard leder et sentralt miljø innen den ovennevnte typen hjerneforskning. Hans gruppe har dokumentert hvordan vonde og vanskelige livserfaringer er assosiert med strukturelle endringer i hjernen (Teicher mfl. 2016; Teicher & Samson 2016; Ohashi mfl. 2017, Ohashi mfl. 2018). Dette er en kunnskapsfront der spørsmålene er mange og konklusjonene bør anses som tentative. De foreløpige funnene framstår som grensesprengende innen faget nevroradiologi, samtidig som de harmonerer godt med teorier, modeller og empiriske funn fra en rekke andre områder – slik vi allerede har omtalt.

I tråd med definisjonen som sier at skadelig (toksisk) stress påvirker hjernens struktur og koblingsmønster, konkluderer forskerne:

- Fysisk, seksuell og emosjonell mishandling i barndommen påvirker hjernens struktur og funksjon. Ut over posttraumatisk stresslidelse, er de radiologisk observerte endringene forbundet med økt risiko for søvnforstyrrelser, depresjon, angst, lærevansker, ADHD, rus- og avhengighetslidelser, personlighetsforstyrrelser, psykoser og alvorlige forløp av bipolar lidelse. Tidligere har man sett for seg en direkte årsakssammenheng *fra* «unormal» hjernestruktur *til* psykiske lidelser, mens vi nå må ta en tredje, konfunderende faktor i betraktning, nemlig at hjernen kan ha endret seg som følge av overlast, med både strukturendring og helseproblemer som konsekvens (Teicher mfl. 2016, Ohashi mfl. 2017, Ohashi mfl. 2018).

- Traumatiske erfaringer kan føre til sansespesifikke endringer i hjernen: Verbal mishandling er forbundet med endringer i hjernens hørselsbaner. Å ha bevitnet vold i nære relasjoner er forbundet med endringer i synsbanene. Seksuelt misbruk er forbundet med tynnere hjernebark i regioner som representerer kjønnsorganene, osv. (Teicher mfl. 2016).

Forskerne diskuterer hvorvidt de påviste endringene i hjernen kan anses som biologiske tilpasninger der individet på kort sikt oppnår en viss beskyttelse mot overveldende inntrykk. På lengre sikt vil imidlertid tilpasningene, som kan være beskyttelsesstrategier i situasjoner uten utvei eller valg, sammen med den generelle allostatiske overbelastningen individet har vært utsatt for, disponere for funksjonstap og helsesvikt (Teicher & Samson 2016).

10.4 Immunsystemet formes gjennom livsløpet

I likhet med hjernen og det endokrine systemet, er immunsystemet et adaptivt organ som formes etter omgivelsenes karakter. De tre systemenes respons henger sammen – aktivitet i de sentralnervøse og endokrine systemer innvirker på immunresponsens utvikling, og *vice versa*. Kjønnshormoner fører for eksempel til at menn og kvinner reagerer ulikt overfor smittestoffer, og en person vil kunne respondere annerledes i en stressende enn i en ustressende periode. Siden systemene henger sammen, og siden aktivitet i det ene systemet fører til aktivitet i de andre, vil utfallet av et bestemt stimulus – det være seg biologisk, psykologisk eller sosialt – kunne innvirke på alle systemene samtidig. Og siden den enkeltes erfaringsbakgrunn er unik, vil utfallet av et bestemt stimulus kunne bidra til helse hos ett individ og til sykdom hos et annet.

Et slikt fenomen sees for eksempel etter vaksinasjon, der et overveldende flertall får styrket sin immunologiske beredskap mens et fåtall personer vil kunne utvikle kronisk, immunbetinget sykdom.

Mye tyder nå på at komplekse og uforklarte sykdommer – for eksempel kronisk utmattelsessyndrom – kan forklares ved at belastninger skader organismens allostatiske beredskap. Og som vist i ACE-studien (kapittel 4 side 86), vil belastende barndomserfaringer kunne virke negativt inn på den voksne kroppenes evne til å reagere immunologisk (Dube mfl. 2009).

Immunsystemets responsdisposisjoner dannes allerede i tidlig barndom, som følge av kroppens interaksjon med omgivelsene. Det nyfødte barnets immunsystem er umodent og trenger stimuli fra mikrober på hud og slimhinner (mikrobiomet)

for å kunne utvikle et effektivt forsvar mot «farlig fremmed». I disse møtene med mikrober settes det i gang modningsprosesser som etter hvert gjør barnet immunkompetent overfor de fleste smittestoffer, noe som blant annet viser seg ved at eldre barn og voksne har lavere dødelighet som følge av infeksjoner enn spedbarn.

I møtet med smittestoffer er det ekstra viktig for organismen å respondere raskt og presist for om mulig å unngå alvorlig sykdom og død. Til dette formålet har evolusjonen frambragt et to-leddet forsvarssystem – det *medfødte* immunsystemet bestående av makrofager, granulocytter og *Natural killer* (NK)-celler som utøver en rask og noe upresis beskyttelse, og det *adaptive* immunsystemet som utøver en noe forsinket men svært presis beskyttelse.

Det adaptive immunsystemets presise beskyttelse oppnås gjennom modningsprosesser kroppens *lymfocytter* gjennomgår i tiden etter en infeksjon. Kroppen produserer et stort overskudd av lymfocytter, hvorav noen har reseptorer med utsøkt spesifisitet overfor bestemte mikrobielle strukturer. Når disse spesifikke lymfocyttene møter «sitt» antigen i en lymfeknute, vil de – på grunn av støttende omgivelser – modnes til immunkompetente lymfocytter. Hukommelse oppstår ved at en del av disse spesialiserte lymfocyttene vedvarer i organismen som hukommelsesceller. Dersom det samme smittestoffet invaderer organismen ved en senere anledning, vil hukommelsescellene bli aktivert og eliminere smittestoffet før det har gitt sykdom.

Modningen av immunsystemet kan dels tilskrives et program nedlagt i genene, og dels tilfeldige stimuli forårsaket av mikrobiell immunaktivering i tiden etter fødselen. Siden ulike barn møter ulike mikrobielle stimuli til ulike tider i utviklingsprosessen, vil immunsystemet utvikles unikt for hvert enkelt individ. Og derfor vil selv eneggede tvillinger etter hvert utvikle seg i forskjellig retning immunologisk sett. Blant annet derfor har immunresponsens styrke og retning lav arvelighet, sammenholdt med egenskaper som kroppshøyde, øyenfarge og blodtype. Dette kan forklare hvorfor autoimmun sykdom ofte utvikler seg i kun den ene av et enegget tvillingpar.

10.5 Genuttrykk formes gjennom livsløpet

Etter at vi nå har omtalt kroppens fysiologiske tilpasningssystemer generelt, og deretter mer spesifikt hjernens og immunsystemets tilpasninger til de erfaringer personen gjør gjennom livsløpet, retter vi nå blikket mot dypet av cellen, der arvestoffet DNA leses og fortolkes. I mange år har diskusjoner om arv, miljø og «genenes innflytelse» på sykdomsutvikling bølget fram og tilbake, både i forskning og

mediene. «Født sånn eller blitt sånn?» «Hvor arvelig er ADHD»? Denne typen spørsmål dukker stadig opp, og de kan lett bringe oss på avveie.

Fram til kartleggingen av det humane genom var ferdigstilt ca. år 2001, ble det satset hardt på å identifisere gener som disponerte for de store folkesykdommene. Tanken var at bestemte sykdommer kan kobles til genvarianter, og at kartlegging av variantene dermed ville være nyttig for forebygging og medikamentell behandling av sykdom. Visjonene ble ikke innfridd, men førte derimot til banebrytende innsikt i biologisk kompleksitet. Ideen om at gener er skjebne måtte vike for systembiologiske modeller som tilsier at reguleringsmekanismer og styring av gen-templatenes avlesning står sentralt.

Ifølge den oppdaterte forståelsen vil cellenes omgivelser ha avgjørende betydning for hvilke deler av DNA som kommer til uttrykk. Effektene ser vi tydelig under fosterutviklingen, der menneskets mer enn 200 celletyper utvikles fra den samme befruktede eggcellen. Cellene har det samme genom, men hvilke deler av DNA som kommer til uttrykk i den enkelte celle, avhenger av impulser fra cellens omgivelser.

Omgivelser må i denne sammenheng gis en bred fortolkning. Mens en celles omgivelser vil utgjøres av nærliggende celler og molekyler, vil et menneskes omgivelser, eller skal vi si omverden, utgjøres av et vidt spekter av ulike fenomener – biologiske, psykologiske og sosiale. Felles for alle disse er at de virker inn på enkeltceller. Men effektene vil ofte, og da særlig i sykdomssammenhenger, komme til uttrykk som endringer på kroppsnivå. Dette ser man hyppig hos eneggede tvillinger som fødes med nesten identisk genetisk utrustning. Disse «identiske» individene vil i ulik grad rammes av genetisk assosiert sykdom, for eksempel schizofreni, diabetes, brystkreft og leddgikt. Slike funn bidrar til en forståelse som tilsier at menneskets DNA avgrenser et rom av muligheter, men at utfallet ikke determineres av genene.

Epigenetisk regulering

Nøyaktig hvordan omgivelsene påvirker organismers livsløp, stod lenge uforklart. Men utover på 2000-tallet ble det stadig tydeligere at miljøpåvirkningen ble formidlet gjennom kjemiske markører – for eksempel metylgrupper – som festes til selve DNA-molekylet. Ved at DNA-sekvenser blir merket med molekylære «nøkler» som skrur gener «av» og «på» i et utall av mulige kombinasjoner, kan organismens genuttrykk endres i takt med endringer i miljøet. Siden endringene ikke betinges av genmutasjoner, men av genenes avlesning, betegnes slike endringer som epigenetiske (fra gresk: *epi* – over, rundt). Dyreforsøk har endatil vist at epigenetiske

reguleringer kan følge avkommet over i kommende generasjoner, uten at vi så langt overskuer implikasjonene av dette for mennesker (Halfon mfl. 2014, Stenz mfl. 2018, Chan mfl. 2018).

Betydningen av epigenetisk regulering kommer tydelig fram gjennom immunresponsens utvikling. Lenge trodde man at det medfødte immunsystemet ikke kunne opparbeide hukommelse mot smittestoffer, men nyere forskning har vist at også makrofager og NK-celler kan trenes opp til å «huske» tidligere møter med et smittestoff. Denne hukommelsen er tett koblet til epigenetiske mekanismer, dette i motsetning til hukommelse i det adaptive immunsystemet der genetiske endringer i lymfocyttreseptorene danner hukommelsens grunnlag.

Selv om epigenetisk regulering nødvendigvis utøves på molekylært nivå, vil effektene manifestere seg i og mellom celler og organer i organismen. Som eksempelet i rammetekst 10.5 illustrerer, vil til og med egenskaper vi vanligvis forbinder med et individs atferd og personlighet, kunne formes gjennom epigenetisk regulering.

Rammetekst 10.5 Epigenetisk innvirkning på stresstoleranse

Forskning utført i 1960-årene hadde vist at rotteungers atferd senere i livet betinges av mors atferd i tiden etter fødselen, og at omsorgsfulle rottemødre får barn med høyere stresstoleranse enn avvisende mødre. Med nyutviklet teknologi kunne forskere på 2000-tallet ta denne forskningen flere steg videre og blant annet vise at barna til omsorgsfulle rottemødre hadde høyere nivåer av kortisolreseptor i hjernen enn barn fra avvisende mødre (Meaney og Szyf 2005). Siden stressresponsen henger sammen med høye nivåer av kortisol, og siden høye nivåer av kortisolreseptor gir lavere stresseffekt av kortisol, kunne innvirkningen av morsatferd på stressresponsen nå kobles til ulik produksjon av kortisolreseptor i barnas hjerner. En viktig mekanisme for hvordan morsatferd kunne gi endret nivå av kortisolreseptor i hjernen ble også funnet: Genet som innvirker på produksjon av kortisolreseptor, viste seg å være ulikt metylert hos rottebarna – barn av omsorgsfulle mødre hadde lavere metyleringsgrad enn barn av avvisende mødre. Og siden lav grad av metylering gir økt genuttrykk og høyere produksjon av kortisolreseptor, hadde forskerne dokumentert at morsatferd innvirker på stressresponsen gjennom en epigenetisk regulatormekanisme.

For å undersøke om tilsvarende epigenetiske mekanismer kan gjøre seg gjeldende hos mennesker som hos rotter (beskrevet i rammetekst 10.5), undersøkte *Barry Lester* og medarbeidere hvordan fem måneder gamle barn tolererer stress (2018). Amming ble valgt som en grov indikator for «morsatferd», og barnets respons på stress ble knyttet til om de var blitt ammet eller ikke. Som for rotter kunne det vises at mors atferd hadde konsekvenser for genuttrykk og følgelig for barnets evne til å tolerere stress – barn født av ammende mødre hadde lavere metyleringsgrad enn barn født av ikke-ammende mødre, og de tolererte stress bedre. Studien til Lester og medarbeidere er utført på et fåtall barn, og resultatene må derfor fortolkes med varsomhet. Funnet er interessant med tanke på betydningen av god omsorg, men denne ene studien beviser *ikke* at amming som sådan er avgjørende for utvikling av god stresstoleranse hos barnet.

Sammenholder vi Lesters spedbarnsstudie med rottestudiene beskrevet i rammetekst 10.5, blir det klart at kunnskapen om epigenetiske reguleringsmekanismer kan ha stor betydning for tekningen om «genetisk risiko». Det som tidligere har vært ansett som negative «risikogener», kan i utgangspunktet være «potensialgener». Om genene utgjør en risiko eller kanskje også bidrar til positive egenskaper, kan avhenge av hvilke levde liv de blir en del av. Ofte er det derfor mer relevant å snakke om genetisk predisposisjon enn predestinasjon.

ANDRE DEL: HUMANISTISKE PERSPEKTIV PÅ DEN LEVDE KROPPEN

Vi introduserer nå et radikalt perspektivskifte, idet vi beveger oss over til humaniora. Det betyr blant annet at begrepsrammen for å observere *individer og hendelser* erstattes med en begrepsramme for å forstå *personer og erfaringer*. Og at kroppen, fra et vitenskapelig synspunkt, betraktes som bærer av historie.

10.6 Personen – mettet med erfaring

Begrepet erfaring er nær knyttet til personbegrepet (rammetekst 10.6). Den norske filosofen og etikeren *Arne Johan Vetlesen* sier: «uten subjekt ingen erfaring; uten erfaring intet objekt». Enhver erfaring er alltid *om noe* og *for noen*, for et *bestemt subjekt*, hvilket innebærer at erfaring hefter ved subjektet, ved *sitt* subjekt (Vetlesen 2009). Ingen kan gjøre eller fortolke en annens erfaring på den andres vegne. Hvert subjekt er et *selv*.

Magnus, Jeanette H. "Ch. 5: Discrimination and Health", i (Red) Diaz, Esperanza & Kumar, Bernadette N.

Migrant Health: A Primary Care Perspective 2019 ss. 55-63

Copyright (C) 2019 CRC Press

Kopiert med hjemmel i Kopinor-avtalen

CHAPTER 5

Discrimination and health

Jeanette H. Magnus

LEARNING OBJECTIVES

This chapter will:

- Present and discuss the relevance of discrimination to health inequity in migrant health
- Explore discrimination and the concept of implicit bias in primary health care

INTRODUCTION

Migration is one of the key drivers of diversity. Diversity and super-diversity imply differences (1). These differences often manifest in the unequal treatment of individuals or a socially defined group; in other words, unjust treatment (2). Racism and racial discrimination are increasingly receiving attention as determinants of racial/ethnic inequalities in health (3). The social and economic inequalities faced by ethnic minority groups are fundamental explanations for some of the disparity demonstrated between groups (4). We need to place ethnic inequalities in health within a wider social context. The lived experience as a minority is perhaps a more important determinant of differentials in mortality in multi-ethnic populations than biological markers of risk and behavioural factors, such as cigarette smoking or diet.

There is growing evidence of the experiences of discrimination and how these contribute to ethnic health inequalities (5). Discrimination can occur based on social identity characteristics such as age, gender, socioeconomic status, national origin, sexual orientation, gender identity, religious orientation, and disability status, in addition to race/ethnicity (6). Racial or race-based discrimination is behaviours, actions, situations, practices, legislation, or policies that result in avoidable and unfair inequalities in power, resources, and opportunities across

racial or ethnic groups (7). The harmful health effects of discrimination across a range of mental health outcomes including depression, psychological distress, anxiety, and well-being have been well documented (8). Perceived discrimination has also been linked to specific types of physical health problems, such as hypertension, self-reported poor health, and breast cancer, as well as for potential risk factors for disease, such as obesity, high blood pressure, and substance use (9). There are, however, few studies on the impact of discrimination on migrant health. In addition to the usual discriminatory characteristics such as age, gender, religion, race/ethnicity, or culture, migrants carry the additional characteristics of being a migrant, and therefore might be more prone to discrimination. In the following we explore the relevant literature with a special emphasis on the primary health care setting.

DISCRIMINATION: THE LAYERS AND COMPLEXITIES

In using the term 'racial discrimination', we include discrimination due to race, ethnicity, culture, and religion, acknowledging the overlapping nature of these categories within popular and academic discourse, rather than as an endorsement of 'race' as an essentialist biological category. While the inclusion of religion in such definitions is debated, we do so in recognition that religion is often conflated with ethnicity and culture in popular culture. Scholars are also increasingly describing the racialized nature of religious identity, thus making it difficult to disentangle. Priest and colleagues offer the following: 'Racism can be expressed through beliefs (e.g. negative and inaccurate stereotypes), emotions (e.g. fear/hatred) or behaviours/practices (e.g. unfair treatment), ranging from open threats and insults (including physical violence) to phenomena deeply embedded in social systems and structures' (10). Experiences of racial discrimination can be subtle, unintentional, unwitting, and even unconscious. In the health literature there is little recognition of racism and racial discrimination as an ideology of both inferiority and superiority (white supremacy) (7,11).

In addressing the challenges of explaining interrelated racially linked health disparities, systems theories offer a way to conceptualize racial discrimination as something greater than the sum of race-linked disparities across a set of subsystems. According to a systems perspective, a system is an entity that comprises a set of dynamically related components or subsystems. When looking at challenges related to discrimination, it is important to look at the smaller components of the system within the context of the larger system (7).

Racism and discrimination

Racism can occur at three levels (Figure 5.1): internalized (i.e. the incorporation of racist attitudes, beliefs, or ideologies into one's worldview), interpersonal (interactions between individuals), and systemic or institutionalized (production, control and access to labour, material, and symbolic resources within a society) (11). One way to understand the experience of discrimination is that it is a stressor that can broadly impact health (12). The most obvious is direct physical injury caused by racist violence. Another could be reduced uptake of healthy behaviours and/or increased adoption of unhealthy behaviours either directly as stress-coping, or indirectly via reduced self-regulation closely linked to other social determinants of health.

Although most stressful experiences do not increase vulnerability to illness, certain kinds of stressors – those that are uncontrollable and unpredictable – are particularly harmful to health, and these characteristics are common to discrimination experiences. Various models conceptualize discrimination as a social stressor that sets into motion a process of

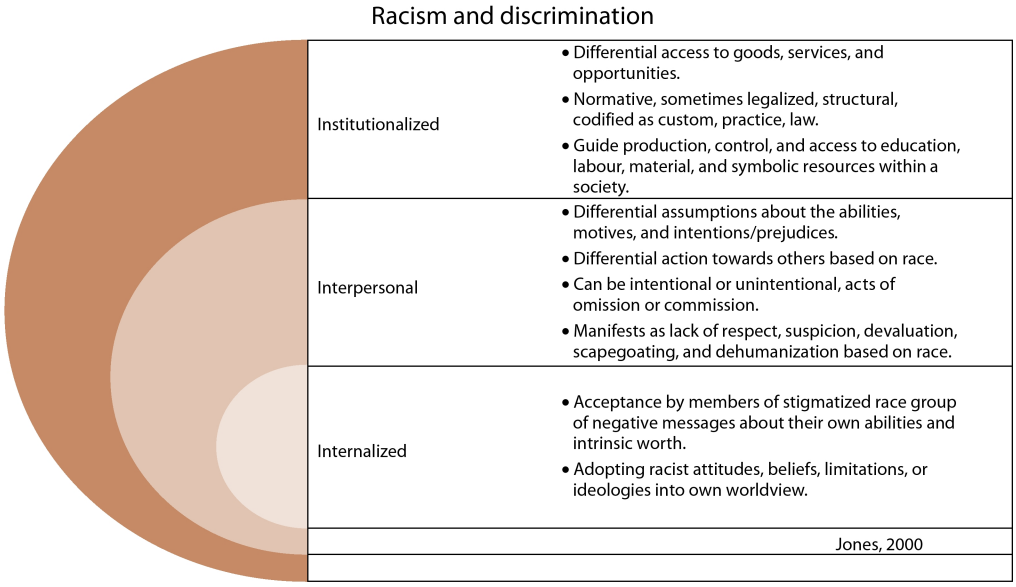


FIGURE 5.1 Levels of racism and discrimination.

physiological responses (e.g. elevated blood pressure, heart rate, cortisol secretions), and these heightened physiological responses can have downstream effects on health over time (13). Routine discrimination can become a chronic stressor that may erode an individual's protective resources and increase vulnerability to physical illness (14). As with other forms of cumulative stress, perceived discrimination may lead to chronic over- or underactivity of allostatic systems impacting allostatic load and other pathophysiological processes (15). Allostatic load (Chapter 2) is the sum of 'the wear and tear on the body' accumulating over time with exposure to chronic or repetitive stress. In addition, negative affective/cognitive and other pathopsychological processes alone or in combination with restricted access to social resources such as employment, housing, and education, and/or increased exposure to risk factors (such as unnecessary contact with the criminal justice system) can create and perpetuate ill health.

Discrimination also impacts children's health status (16). A direct effect of the maternal and family experiences of racial discrimination was found among women of Arab descent living in California, who were significantly more likely to give birth to a low-birth-weight or preterm baby in the 6-month period following September 11, 2001 than in the 6-month period before this date (17). This result is particularly noteworthy since Arab and Arab American women typically have very healthy pregnancies and consistently low rates of low-birth-weight or preterm babies. Similar increases were not noted among any other racial/ethnic group during the same time period. Additional studies are warranted to see how vicarious discrimination, direct or transmitted by various media, impacts migrant health.

Health providers and discrimination

Social psychology scholars have conceptualized prejudicial attitudes or bias as implicit and explicit (18). Explicit attitudes are thoughts and feelings that people deliberately think about

and can make conscious reports about. On the other hand, implicit attitudes often exist outside of conscious awareness, and thus are difficult to consciously acknowledge and control. These attitudes are often automatically activated and can influence human behaviour without conscious volition. Such notions are applied most often when people are busy, distracted, tired, and under pressure.

Despite the interest and importance of health care provider racism, we have limited knowledge of the extent and how to best capture it (19). Higher physician implicit bias has been linked to greater physician verbal dominance, giving less time to the patient's statements (20). Unintentional bias on the part of physicians can influence the way they treat patients from certain racial and ethnic groups (18). Most physicians are unaware that they hold such biases, which can unknowingly contribute to inequalities in health care delivery (21). People's ability to self-regulate the expression of bias in their behaviours depends on how aware they are of their racial bias, and how easy or hard the behaviours are to self-regulate (20). Virtually absent in the literature, however, is evidence-based information on how to reduce an individual health care provider's bias (21). Women, regardless of ethnicity, are more likely than men to experience biased interactions and treatment in care. Bias can exist on multiple social dimensions, and patients with multiple minority identities may be particularly affected. Interestingly, no study investigating implicit bias towards migrants was identified (18).

System discrimination

A race discrimination system is the product of both a system of race-linked disparities and a belief system that perpetuates them (7,22). It appears that bias and discrimination at a societal level create psychological environments that can also produce health care disparities among members of a stigmatized target group. A system's perspective gives understanding for how race-linked disparities affect the social and cultural milieu by creating a system that makes race salient in a way that distorts our perceptions and feelings (22). We need to come to a paradigm shift when discussing racial health disparity and focus on when discrimination is not at the root of health inequity. As soon as the concept of the 'others' is introduced we reinforce a system of discrimination. This will also impact how we teach science and train the future cadres of health professionals. Dr CP Jones, an American family physician, presented an allegory, 'A Gardener's Tale', in 2000 in the *American Journal of Public Health*, illustrating the concepts addressed previously. Readers are advised to take time to see presentations at TeDx from 2014, or 'Racism and the Social Determinants of Equity' at Beyond Flexner 2015, worth watching. Box 5.1 presents an abbreviated version of 'A Gardener's Tale' (11).

RACISM AND HEALTH IMPACTS

Exposure to interpersonal discrimination has been posited as an explanation, at least in part, for the stark and consistent racial disparities in health. A dose-response association between health outcomes and degrees of self-reported experience of racial discrimination has been observed (23). A rapidly growing body of literature links experiences of unfair treatment, discrimination, and racism to subsequent morbidity and even mortality. Reporting experiences of discrimination is independently related to the likelihood of reporting fair or poor health independent of socioeconomic status. Interestingly, young people are more likely to report race discrimination than older individuals, but the impact on health does not vary by age, gender, or socioeconomic status. The association between unfair treatment and

BOX 5.1 A GARDENER'S TALE

(Emerging from personal experience at their first house.) In spring, they discovered that only a few of the flower boxes had soil, so they bought potting soil to fill these, and planted marigolds in all of them. Weeks later some plants were doing great, others were struggling; it was clear that those in the old rocky soil were not on par with those in the new soil. Some seeds in each box were stronger and doing better than others, but the strongest flowers in the poor soil could barely keep up with the weakest in the good soil.

An analogy would be: what if a gardener decides to plant red and pink flowers, but likes red better, and plants them in the good soil? And when they do better, s/he would say, 'See? I knew red was better!' And, if the flowers were perennials and went to seed and regrew each year, this would perpetuate, if not worsen, the difference, the inequity.

And what if the gardener said, 'These pink flowers are going to do poorly anyway', and deadheaded the weakest, allowing them no chance at all? And in future generations if the gardener's children and grandchildren always grew up knowing that red flowers did better than pink?

What if the gardener does not redistribute the soil, then the flowers do not have an equal chance to grow? If s/he did, maybe generations of selection would take a few years to compensate for, or maybe because only the stronger pink seeds survived, they would do even better than the red if given a chance to have the same opportunity to grow in fertile soil. We cannot be sure until that opportunity is comprehensively and completely available.

well-being might be more pronounced for mental, as opposed to physical, health outcomes among migrants (24). Thus, with respect to the production of racial disparities in health, the chronicity rather than the severity of exposure to discrimination is often considered to be a stronger predictor of worse outcomes, either across subgroups or over time (5). Furthermore, everyday discrimination is more consistently associated with higher rates of morbidity and mortality than acute instances of unfair treatment. The extent to which exposure to unfair treatment is a key mechanism through which racial disparities in health are produced remains unclear (6). A recent systematic review assesses the limitations of current studies on racism and health service utilization demonstrating the need for more longitudinal studies (25).

There are hormonal consequences of experiencing or just visualizing racial discrimination that can be measured in a laboratory setting as a peripheral stress response, characterized by an increase in blood pressure and cortisol (13). Long-term exposure to discrimination and anxiety repeat episodes, perpetuate a vicious cycle, increasing the stress, and engages neural stress in regulatory circuits in the brain as well as inflammatory markers (12,13). Discrimination is fuelling processes related to mental and non-inflammatory chronic diseases, and science has just started exploring the complexity (26).

A life-course approach is also advocated to understand the influence of racial discrimination on children and young people across the lifespan regardless of the source of exposure (27). Childhood exposure to either direct and/or vicarious racial discrimination has been linked to poor child health, well-being, and development (10). Further exploration of the differential effects of racial discrimination experienced by caregivers, family, and peers, compared to experiences of racial discrimination by the children and young people themselves is

needed (28). Mustillo et al. demonstrated that perceived racial discrimination explains as much or more of the black/white disparity in poor birth outcomes (low birth weight and preterm birth) as maternal education, income, cigarette smoking, alcohol consumption, and depressive symptomatology taken together (29). The association between socioeconomic status (SES) and health is typically thought to be remarkably consistent, whether it is across time, place, or health outcome. However, more recent research has begun to call this seemingly evident 'truth' into question. As more evidence comes to light, it appears that the shape of the SES/health gradient might differ according to experience of racial discrimination.

Several studies have linked telomere length to health disparity. Telomeres are a stretch of DNA at each end of a chromosome that protects the protein-encoding part of the DNA. They become shorter each time a cell divides. When a telomere becomes too short, it can no longer protect the cell's DNA, leaving the cell at risk for serious damage. Leukocyte telomere length (LTL) is an indicator of biological ageing; shorter LTL is associated with earlier mortality. It also has been tied to many age-related health issues, such as heart disease, diabetes, dementia, Alzheimer's disease, and arthritis (30). Shorter telomeres have been associated with experiences of heightened psychosocial stress and exposure to adverse social conditions. Studies have found that LTL may exhibit accelerated shortening due not only to physiologic conditions, but also psychosocial stress, and therefore may be a pathway generating racial disparities in health. Experiencing everyday discrimination was associated with shortened telomere length among older black adults and associated with mental health (31). This gives us an understanding that discrimination might generate greater disease vulnerability and possibly premature death.

SOCIETAL IMPACT OF DISCRIMINATION

The societal impact of discrimination, as indicated by the findings of Elias and Paradies, is shown in a substantial loss in disability-adjusted life years (DALYs) due to discrimination. On average, Australia loses up to 3.02% of its GDP ($37.9 billion) per annum as a result of individuals being exposed to some form of racial discrimination. Another important cost component, not included in their study, was the pain and suffering involved due to illness that can be attributed to racial discrimination. These findings also indicate that some of the costs resulting from discrimination are avoidable if measures are taken by governmental and nongovernmental institutions to curb racial discrimination. Countries with racially and ethnically diverse populations can therefore realize substantial savings by enforcing effective antidiscrimination measures. Not addressing, preventing, or combatting discrimination is costly not only for the individual, but for society at large (32).

MIGRANT HEALTH AND DISCRIMINATION

Studies on ethnic discrimination and health controlling for various predictors of disease as well as relevant migration-specific factors are few (33). However, some of the laboratory biological hormone studies have been replicated, and the impact of racial discrimination on psychological and physiological stress responses confirmed in migrant populations (14).

A recent study from Canada highlights that limiting discrimination increases a sense of belonging, supporting good mental health of refugees and increasing the probability of successfully identifying with the national culture (34). Clustering of migrants and minorities in certain geographic areas is common across the globe. Part of this can be due to structural discrimination and legislations that confine the minorities to certain areas, part could be that living in areas

where the majority are like you reduces the daily risk of exposure to racism and discrimination. However, the Canadian study on refugees contradicted this, reporting no benefit on mental health if clustered in geographic areas (34). Although some new studies are being produced on the impact of discrimination on health, longitudinal studies are warranted that assess the effects of social contexts and discussions on the openness of the host society, exclusion and separation of migrant communities, and the respective effect on discrimination and distress (35).

DECREASING DISCRIMINATION AND IMPROVING MIGRANT HEALTH

Antonovsky's Salutogenesis and Sense of Coherence (SOC) as a global orientation in which the world is seen, to a greater or lesser extent, as comprehensible, manageable, and meaningful is worth exploring related to discrimination (36). According to Antonovsky, the meaningfulness component is strengthened through participation in shaping outcomes in socially valued contexts. 'When others decide everything for us — when they set the task, formulate the rules, and manage the outcome — and we have no say in the matter, we are reduced to being objects. A world thus experienced as being indifferent to what we do, comes to be seen as a world devoid of meaning' (36). One can speculate that living a life where you are surrounded by various forms of discrimination challenges you and diminishes your SOC. This is interestingly described in a recent study on differences in health and SOC between first- and second-generation migrants in Holland and contributes to the discussion on the healthy migrant effect of the first-generation migrants (37).

To reduce racial and ethnic disparities in health care, we must ascertain the prevalence of biased attitudes among health care providers and whether bias contributes to problems in patient-provider interactions and relationships, quality of care, continuity of care, treatment adherence, and patient health status. European health care and society at large should learn from the history of discrimination as well as the current system of racial discrimination in the United States, and avoid copying them. We must include this in education at all levels and also increase the awareness about implicit bias across the European health care system. Using an intersectionality approach, addressing political and social determinants of health is imperative when addressing migrant health (38). Interestingly, the rigor of methodological studies aimed at improving cultural competency in health care has been challenged (38). Most of the training interventions measured changes in professional attitudes towards the population of interest but did not measure the downstream effect of changing provider beliefs on the care delivered to patients. A limited number of studies have been published on interventions to improve migrant health.

We are socialized into discrimination of 'the other'. Small children are colour-blind until taught otherwise. Systems, group, and individual aspects all have to be taken into account when we now understand the complexity of health inequity and discrimination as a system of power. Health equity is a process not to be perceived as an attainment (39). 'Achieving health equity requires valuing all individuals and populations equally, recognizing and rectifying historical injustices, and providing resources according to need. It requires dismantling systems or structure inequity and putting in their place systems in which all people can know and develop their full potentials' (39, page S75).

SUMMARY

The increasing complexity of our societies and the challenges short- and long-term migration present to all aspects of society call for change. Change in how we think about health, about

what creates health, and how we as health care providers must participate in the shaping of a new paradigm. Discrimination, in all shapes and forms, fosters ill health and impacts the individual, young and old, the community, rich and poor, and the society at large, finanically and culturally. Addressing, preventing, and combatting discrimination is an important health and public health challenge for all, not only for migrant health.

REFERENCES

1. Vertovec S. Super-diversity and its implications. *Ethn Racial Stud*. 2007;30(6):1024–54.
2. Krieger N. Refiguring 'race': epidemiology, racialized biology, and biological expressions of race relations. *Int J Health Serv*. 2000;30(1):211–6.
3. Paradies Y, Ben J, Denson N, Elias A, Priest N, Pieterse A et al. Racism as a determinant of health: a systematic review and meta-analysis. *PLOS ONE*. 2015;10(9):e0138511.
4. Nazroo JY. The structuring of ethnic inequalities in health: economic position, racial discrimination, and racism. *Am J Public Health*. 2003;93(2):277–84.
5. Colen CG, Ramey DM, Cooksey EC, Williams DR. Racial disparities in health among nonpoor African Americans and Hispanics: the role of acute and chronic discrimination. *Soc Sci Med*. 2018;199:167—80.
6. Hicken M, Durkee M, Kravitz-Wurtz N, Jackson J. The role of racism in health inequalities: integrating approaches from across disciplines. *Soc Sci Med*. 2018;199(1):11–240.
7. Jones CP. Confronting institutionalized racism. *Phylon*. 2003;50(1):7–22.
8. Khan M, Ilcisin M, Saxton K. Multifactorial discrimination as a fundamental cause of mental health inequities. *Int J Equity Health*. 2017;16(1):43.
9. Pascoe EA, Smart Richman L. Perceived discrimination and health: a meta-analytic review. *Psychol Bull*. 2009;135(4):531–54.
10. Priest N, Paradies Y, Trenerry B, Truong M, Karlsen S, Kelly Y. A systematic review of studies examining the relationship between reported racism and health and wellbeing for children and young people. *Soc Sci Med*. 2013;95(1):115–27.
11. Jones CP. Levels of racism: a theoretic framework and a gardener's tale. *Am J Public Health*. 2000;90(8):1212–5.
12. Brody GH, Yu T, Beach SR. Resilience to adversity and the early origins of disease. *Dev Psychopathol*. 2016;28(4 Pt2):1347–65.
13. Berger M, Sarnyai Z. 'More than skin deep': Stress neurobiology and mental health consequences of racial discrimination. *Stress*. 2015;18(1):1–10.
14. Fischer S, Nater UM, Strahler J, Skoluda N, Dieterich L, Oezcan O, Mewes R. Psychobiological impact of ethnic discrimination in Turkish immigrants living in Germany. *Stress*. 2017;20(2):167–74.
15. Tomfohr LM, Pung MA, Dimsdale JE. Mediators of the relationship between race and allostatic load in African and White Americans. *Health Psychol*. 2016;35(4):322–32.
16. Bécares L, Nazroo J, Kelly Y. A longitudinal examination of maternal, family, and area-level experiences of racism on children's socioemotional development: patterns and possible explanations. *Soc Sci Med*. 2015;142:128–35.
17. Lauderdale DS. Birth outcomes for Arabic-named women in California before and after September 11. *Demography*. 2006;43(2):185–201.
18. Hall WJ, Chapman MV, Lee KM, Merino YM, Thomas TW, Payne BK et al. Implicit racial/ethnic bias among health care professionals and its influence on health care outcomes: a systematic review. *Am J Public Health*. 2015;105:e60–76.
19. Paradies Y, Truong M, Priest N. A systematic review of the extent and measurement of healthcare provider racism. *J Gen Intern Med*. 2014;29(2):364–87.
20. Hagiwara N, Penner LA, Gonzalez R, Eggly S, Dovidio JF, Gaertner SL et al. Racial attitudes, physician-patient talk time ratio, and adherence in racially discordant medical interactions. *Soc Sci Med*. 2013;87:123–31.
21. Van Ryn M. Avoiding unintended bias: strategies for providing more equitable health care. *Minn Med*. 2016;99(1):40–3.

22. Reskin B. The race discrimination system. *Annu Rev Sociol*. 2012;38(1):17–35.
23. Harris R, Tobias M, Jeffreys M, Waldegrave K, Karlsen S, Nazroo J. Racism and health: the relationship between experience of racial discrimination and health in New Zealand. *Soc Sci Med*. 2006;63:1428–41.
24. Hatch SL, Gazard B, Williams DR, Frissa S, Goodwin L, SELCoH Study Team, Hotopf M. Discrimination and common mental disorder among migrant and ethnic groups: findings from a South East London Community sample. *Soc Psychiatry Psychiatr Epidemiol*. 2016;51:689–701.
25. Ben J, Cormack D, Harris R, Paradies Y. Racism and health service utilisation: a systematic review and meta-analysis. *PLOS ONE*. 2017;12(12):e0189900.
26. Busse D, Yim IS, Campos B. Social context matters: ethnicity, discrimination and stress reactivity. *Psychoneuroendocrinology*. 2017;83:187–93.
27. Gee GC, Walsemann KM, Brondolo E. A life course perspective on how racism may be related to health inequities. *Am J Public Health*. 2012;102(5): 967–74.
28. Heard-Garris NJ, Cale M, Camaj L, Hamati MC, Dominguez TP. Transmitting trauma: a systematic review of vicarious racism and child health. *Soc Sci Med*. 2018;199:230–40.
29. Mustillo S, Krieger N, Gunderson EP, Sidney S, McCreath H, Kiefe CI. Self-reported experiences of racial discrimination and Black-White differences in preterm and low-birthweight deliveries: The CARDIA Study. *Am J Public Health*. 2004;94(12):2125–31.
30. Rizvi S, Raza ST, Mahdi F. Telomere length variations in aging and age-related diseases. *Curr Aging Sci*. 2014;7(2):161–7.
31. Chae DH, Nuru-Jeter AM, Adler NE, Brody GH, Lin J, Blackburn EH, Epel ES. Discrimination, racial bias, and telomere length in African-American men. *Am J Prev Med*. 2014;46(2):103–11.
32. Elias A, Paradies Y. Estimating the mental health costs of racial discrimination. *BMC Public Health*. 2016;16:1205–10.
33. Aichberger MC, Bromand Z, Rapp MA, Yesil R, Montesinos AH, Temur-Erman S et al. Perceived ethnic discrimination, acculturation, and psychological distress in women of Turkish origin in Germany. *Soc Psychiatry Psychiatr Epidemiol*. 2015;50:1691–700.
34. Beiser M, Hou F. Predictors of positive mental health among refugees: results from Canada's General Social Survey. *Transcult Psychiatry*. 2017;54(5–6):675–95.
35. Edge S, Newbold B. Discrimination and the health of immigrants and refugees: exploring Canada's evidence base and directions for future research in newcomer receiving countries. *J Immigr Minor Health*. 2013;15:141–8.
36. Antonovsky, A. *Unraveling the Mystery of Health*. San Francisco, CA: Jossey-Bass; 1987.
37. Slootjes J, Keuzenkamp S, Saharso S. The mechanisms behind the formation of a strong Sense of Coherence (SOC): the role of migration and integration. *Scand J Psychol*. 2017;58:571–80.
38. Truong M, Paradies Y, Priest N. Interventions to improve cultural competency in healthcare: a systematic review of reviews. *BMC Health Serv Res*. 2014;14:99.
39. Jones CP. Systems of power, axes of inequity. Parallels, intersections, braiding the strands. *Med Care*. 2014;52:S71–75.

Swinburn, Boyd A. et al "The global obesity pandemic. shaped by global drivers and local environments", i

The Lancet 378 (9793) 2011 ss. 804-814

Copyright (C) 2011 Elsevier Ltd

Kopiert med hjemmel i Kopinor-avtalen

Obesity 1

The global obesity pandemic: shaped by global drivers and local environments

Boyd A Swinburn, Gary Sacks, Kevin D Hall, Klim McPherson, Diane T Finegood, Marjory L Moodie, Steven L Gortmaker

Lancet 2011; 378: 804–14

See Editorial page 741

See Comment pages 743, 744, and 746

See Perspectives page 761

This is the first in a Series of four papers about obesity

WHO Collaborating Centre for Obesity Prevention (B A Swinburn MD, G Sacks PhD), **and Deakin Health Economics, Deakin Population Health** (M L Moodie DrPH), **Deakin University, Melbourne, VIC, Australia; National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Washington, DC, USA** (K D Hall PhD); **New College, University of Oxford, Oxford, UK** (Prof K McPherson PhD); **Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Vancouver, BC, Canada** (Prof D T Finegood PhD); **and Department of Society, Human Development and Health, Harvard School of Public Health, Harvard University, Boston, MA, USA** (Prof S L Gortmaker PhD)

Correspondence to: Prof Boyd A Swinburn, WHO Collaborating Centre for Obesity Prevention, Deakin University, Melbourne, VIC 3125, Australia **boyd.swinburn@deakin.edu.au**

The simultaneous increases in obesity in almost all countries seem to be driven mainly by changes in the global food system, which is producing more processed, affordable, and effectively marketed food than ever before. This passive overconsumption of energy leading to obesity is a predictable outcome of market economies predicated on consumption-based growth. The global food system drivers interact with local environmental factors to create a wide variation in obesity prevalence between populations. Within populations, the interactions between environmental and individual factors, including genetic makeup, explain variability in body size between individuals. However, even with this individual variation, the epidemic has predictable patterns in subpopulations. In low-income countries, obesity mostly affects middle-aged adults (especially women) from wealthy, urban environments; whereas in high-income countries it affects both sexes and all ages, but is disproportionately greater in disadvantaged groups. Unlike other major causes of preventable death and disability, such as tobacco use, injuries, and infectious diseases, there are no exemplar populations in which the obesity epidemic has been reversed by public health measures. This absence increases the urgency for evidence-creating policy action, with a priority on reduction of the supply-side drivers.

Introduction

As UN member states prepare to gather in New York in September, 2011, for the first High-Level Meeting of the UN General Assembly on non-communicable diseases (NCDs), the inexorable global rise of obesity will be the toughest challenge that they face. Many countries can serve as excellent exemplars for reduction of infectious diseases, injuries, and some of the risk factors for NCDs, such as smoking, high cholesterol, and hypertension. However, no country can act as a public health exemplar for reduction of obesity and type 2 diabetes. All countries are searching for answers about how to reverse the rising tide of adult and childhood obesity.

The 2004 WHO global strategy on diet, physical activity and health[1] provides an excellent overall guide for societal action. However, with few exceptions, governments have made very slow progress in the implementation of these strategies. The food and media industries have, by contrast, moved rapidly by making various national[2] and international[3] pledges, including self-regulatory codes of practice. Although independent assessment of the true effect of these pledges is needed, governments also need to meet their obligations for policy action and leadership, which are described in several authoritative reports.[1–5]

The aim of *The Lancet*'s Obesity Series is to state the case for action on obesity: what is the size and nature of the problem, what is driving its global increase, what will the future obesity burden be under a business-as-usual scenario, and what action is needed to reverse the epidemic? In this first report in the Series, we describe the obesity epidemic and explain the reasons for its concurrent rise across countries and the wide variation in obesity prevalence between countries. The interaction of these major determinants of obesity has important implications for the action needed to reverse the epidemic.

Key messages

- Changes in the global food system, including reductions in the time-cost of food, seem to be the major drivers of the rise of the global obesity epidemic during the past 3–4 decades, although substantial differences in national and local environments (especially sociocultural, economic, and transport environments) produce the wide variation in obesity prevalence recorded across populations.
- In the first half of the 20th century, increased mechanisation and motorisation were accompanied by corresponding decreases in food energy supply (indicative of consumption), thereby keeping obesity prevalence low. In many high-income countries, an energy balance flipping point seems to have occurred in the 1960s–70s, with an increasing food energy supply now pushing up energy intake and population weight.
- Adult obesity continues to increase almost universally, but in some childhood and adolescent populations the epidemic seems to be flattening or even decreasing.
- Present systems for monitoring population weight and nutrition are inadequate in almost all countries.
- Obesity is the result of people responding normally to the obesogenic environments they find themselves in. Support for individuals to counteract obesogenic environments will continue to be important, but the priority should be for policies to reverse the obesogenic nature of these environments.
- Governments have largely abdicated the responsibility for addressing obesity to individuals, the private sector, and non-governmental organisations, yet the obesity epidemic will not be reversed without government leadership, regulation, and investment in programmes, monitoring, and research.

The global rise in obesity prevalence

The rise of the obesity epidemic seemed to begin almost concurrently in most high-income countries in the 1970s and 1980s;[6] since then, most middle-income and many low-income countries have joined the global surge in obesity prevalence in adults and children.[7–9] By 2008, an estimated 1·46 billion adults globally were overweight (body-mass index [BMI] >25 kg/m²) and 502 million adults were obese (BMI >30 kg/m²).[9] Furthermore, an estimated 170 million children (aged <18 years) globally were classified as overweight or obese.[10] This estimate includes more than 25% of all children in some countries, more than double the proportions from the start of the epidemic (figure 1).

Analysis of the patterns of the obesity epidemic in the past four decades is limited by the absence of representative data from different countries.[11] Nevertheless, the pattern by which obesity prevalence rises in particular populations seems predictable. In low-income and middle-income countries, groups of high socioeconomic status in urban areas tend to be the first to have high obesity prevalence, but the burden of obesity shifts to low socioeconomic status groups and rural areas as a country's gross domestic product (GDP) increases.[14,15] In Brazil, one of the few middle-income countries with repeated cross-sectional surveys of BMI, this pattern was particularly evident for women, with obesity rates increasing rapidly in the lowest income groups.[16] The highest prevalences of overweight and obesity are in middle-age groups (45–59 years) throughout this transition.[12,17]

The global rise of obesity has serious health effects. Raised BMI is an established risk factor for diseases such as type 2 diabetes, cardiovascular diseases, and many cancers.[4,18,19] The disability attributable to obesity and its consequences was calculated in 2004 at more than 36 million disability-adjusted life-years,[18] with obesity accounting for between 2% and 6% of total health-care costs in many countries.[20] NCDs are now the dominant cause of preventable disease burden even in many low-income countries,[18,21] and obesity has overtaken tobacco as the largest preventable cause of disease burden in some regions.[22] Although the reduction in premature mortality and morbidity from cardiovascular diseases in high-income countries during the past 40 years has been substantial, there is serious concern that the rise of obesity and type 2 diabetes will slow or even reverse this trend.[23–25]

The increases in overweight and obesity in adults are widely projected to continue to heighten the burden of obesity-related morbidity and mortality in the coming decades.[1,26] However, encouraging reports are emerging from countries such as Sweden, Switzerland, France, and Australia that overweight and obesity prevalence in some childhood age groups might be flattening or even decreasing.[27] But, overall prevalence is still high. Crucially, very few countries have adequate monitoring systems in place, which is remarkable in view of the importance of this issue. Consequently, the frequency and standard of monitoring urgently needs to improve so that the progress of the global epidemic can be tracked and lessons from the experiences of different countries and population groups can be learnt.

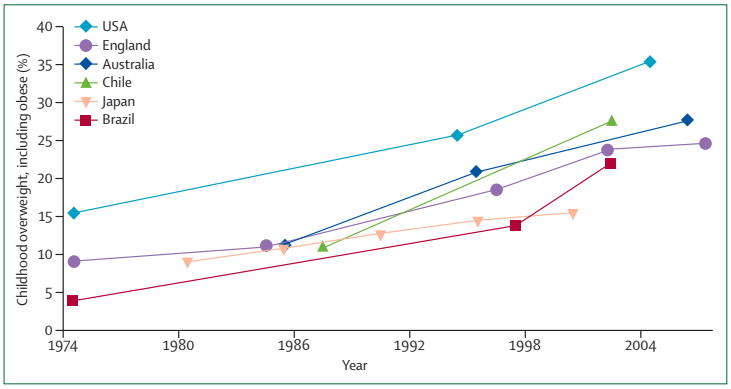


Figure 1: **Estimates of percentage of childhood population overweight, including obese (with use of International Obesity Taskforce cutoffs) in a selection of countries**
Based on data from Wang and Lobstein,[11] International Association for the Study of Obesity,[12] and Matsushita and colleagues.[13]

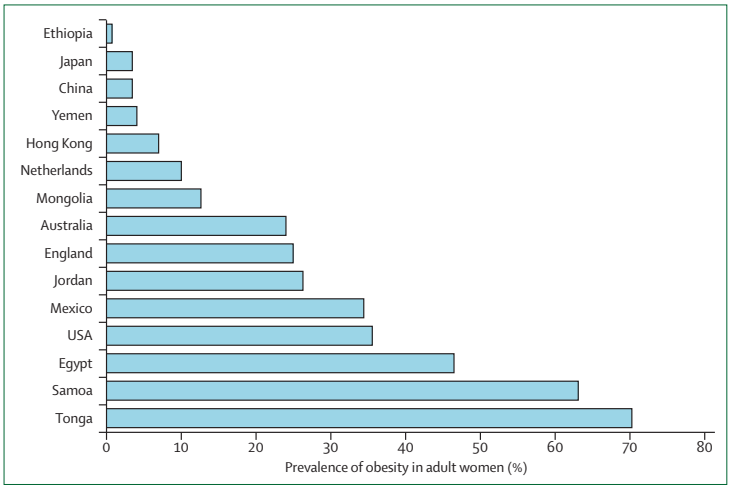


Figure 2: **Prevalence of obesity (body-mass index [BMI] >30 kg/m²) in adult women in a selection of countries in the 2000s**
Data from International Association for the Study of Obesity.[17]

The available data show very wide variations in obesity prevalence globally, particularly for women (figure 2). For some populations (eg, China), small body-frame sizes mean that a BMI cutoff point of 30 kg/m² for obesity will underestimate the amount of over-fatness and comorbidities, compared with other populations with larger frame sizes (eg, Tonga).[28] However, this definitional difficulty does not account for the 100-times differences between the populations (0·7% *vs* 70%). Many of the reasons for the variations across populations are intuitive. For example, Ethiopia does not have sufficient national

wealth for obesity to have manifested itself, and populations in Hong Kong and Jordan have had a greater exposure to obesogenic food environments than do their counterparts in China and Yemen. However, many complexities exist in understanding why some populations and subpopulations are more susceptible to the drivers of obesity than others, and how mediating factors affect different population groups.

Broad economic effects on obesity

The most obvious environmental precondition for a population to develop obesity is sufficient wealth. The relation between GDP and mean BMI is positive and linear up to a GDP of about US$5000 per person per year; at greater GDP, the relation with GDP and BMI is almost flat.[29] A degree of economic prosperity is thus an enabler for obesity, but the level of prosperity does not have to be high for obesity to manifest; in some low-income countries, such as Pacific Island nations, obesity prevalence is very high.[17] A return to national poverty is not a recommended approach to reduce obesity and type 2 diabetes but, as seen in Cuba and Nauru,[30,31] it can have that effect.

The economic transition towards increasing GDP brings with it several other transitions: demographic (younger to older population distribution, rural to urban); epidemiological or health (infectious diseases to NCDs); technological (low to high mechanisation and motorisation); and nutritional (traditional foods and cuisines to more processed energy-dense foods).[32] The pace of change of these transitions has increased substantially in recent decades; so many countries in transition are faced with double burdens of disease. For example, most countries that still have a substantial burden of undernutrition and its related diseases also have a substantial or emerging burden of overnutrition and its related NCDs. Both these conditions need to be addressed together for several important reasons: fetal and infant undernutrition followed by adult overnutrition has a double effect on the later burden of NCDs;[33] the underlying drivers within the food system (eg, food quality and food distribution) are often common to both disorders; and NCDs cannot be ignored even while efforts to reduce undernutrition continue.

In the same way as obesity is the result of people responding normally to the obesogenic environments that they find themselves in, so too do these obesogenic environments arise because businesses and governments are responding normally to the broader economic and political environments that they find themselves in. A central tenet of modern, market-based economies is the benefits of economic growth; and a parallel tenet of business and trade is the benefits of more liberalised, less regulated global markets. Economic growth is especially important for low-income countries to move them from poverty to economic prosperity; however, for high-income countries, higher levels of GDP do not bring greater happiness and wellbeing for their citizens but do bring greater consumption of all products.[34] The technological changes that are creating cheaper and more available food calories and the strong economic forces driving consumption will inevitably lead to overconsumption and obesity.[35–37]

In the broader view, obesity is similar to rising greenhouse gases and environmental degradation as yet another detrimental effect of individual and corporate overconsumption.[34] The pressure for market liberalisation means that regulatory approaches, although feasible, are difficult to achieve—as exemplified in the great reluctance of policymakers to regulate reductions in marketing of obesogenic foods and beverages, such as fast foods and sugar-sweetened drinks, to children. There are many reasons for government intervention to restrict marketing to children, including protection of the rights of children,[38] public demand for regulations,[39] and application of the precautionary principle of preventive action early, even before absolute proof is available.[40] Although obesity has been described as "a sign of commercial success but a market failure"[41] debate exists about whether market failure provides an additional argument for government intervention with respect to prevention of childhood obesity[41–45] (panel 1).

Panel 1: Is the market failing children?

A market fails when prices and the quantities bought and sold are no longer indicative of their costs and benefits to society.[46] Is the market failing children? The first of four reasons for market failure is when vulnerable individuals are not protected.[47] Children are clearly a vulnerable group that warrant societal protection, and this notion represents the strongest argument for government intervention. They are not mature, they do not have nutritional knowledge, are unable to perceive the risks of their behaviour, and their choices are readily affected by marketing.[47–49] The second reason is when consumers do not have the information necessary to make fully informed decisions about their food selection,[47,50] as is also clearly the case with children. However, generally, interventions to rectify information gaps seem to have modest effects.[46] The third reason for market failure is when people prioritise immediate gratification over potential long-term negative results, which is a hallmark of childhood. The final reason relates to spill-over effects (or externalities), when the costs of obesity are borne by society. Although yearly health-care costs to the taxpayer are higher for obese than for non-obese people, reduced life expectancy due to obesity makes it uncertain whether the life-time social costs are actually higher.[42,51] Externalities might arise at the family level through reduced household income or additional carer duties.[46] Thus, there is ample justification for protecting children's health from the predatory effects of markets, yet almost universally, governments are failing in this responsibility. The charge of so-called nannyism almost inevitably arises[42] in relation to regulatory interventions, yet for children, and even for adults, governments have a fundamental role in helping to make healthy choices the easy choices.[52]

Drivers of the obesity epidemic

We define a driver of the global obesity epidemic as an environmental factor that has changed substantially during the past 40 years (coinciding with the upswing of the epidemic), is global in nature (affecting almost all countries with enabling economic conditions), and is rapidly

transmissible (in view of the near simultaneous nature of the epidemic across countries). Some environmental determinants of obesity, such as the built environment, can have important effects on behaviours;[53] however, built environments have not changed simultaneously and universally to become more obesogenic during the past few decades. The built environment is thus unlikely to have been a major driver of the global epidemic, although the way in which people have responded to the built environment (eg, increased traffic congestion) has changed with time and might be important. The obvious possible drivers of the epidemic are in the food system:[35] the increased supply of cheap, palatable, energy-dense foods; improved distribution systems to make food much more accessible and convenient; and more persuasive and pervasive food marketing.[54]

Several studies have tested the hypothesis that increases in the food supply are the dominant drivers of the weight gain in populations.[55–57] Results from these investigations show that the rise in food energy supply was more than sufficient to explain the rise in obesity in the USA from the 1970s,[55,56] and most of the weight increase in the UK since the 1980s.[57] A related hypothesis is that the policies put in place in the USA and other countries to increase the food supply from the 1970s led to a situation in which the abundance of food in these countries began to push up population energy intake—a reversal of the previous situation in which energy intake was pulled down by decreases in physical activity (panel 2, figure 3).

Figure 4 shows the key drivers of the global obesity epidemic and presents an overview framework for understanding of population-level obesity determinants and solutions. Our framework has features in common with other frameworks for obesity determinants[26,65]—eg, the layered levels of determinants that recognise that the physiology of energy balance is proximally determined by behaviours and distally by environments. A further distinction is made between the obesogenic drivers within the food and physical activity environments, which are proximal determinants of behaviours, and more distal systemic drivers. Taxation regimes, regulation of the marketplace, and social and economic policies set the conditions under which businesses and individuals operate, and many of these conditions can have distal effects on obesity.[34,66] For example, independent of the overall wealth of a country or state, the higher the level of income inequality, the higher the prevalence of obesity.[67] These distal effects might convert to higher obesity prevalence through many pathways, such as through psychosocial and behavioural effects.[67]

Our framework (figure 4) recognises the importance of environmental conditions that operate on a population to accentuate or attenuate the effect that the drivers have on the trajectory of changes in obesity prevalence. These moderators or modulators, although important in affecting the slope of the rise of obesity, cannot be deemed drivers of the epidemic if they have not changed sufficiently and coincidentally with the onset of the obesity epidemic. For example, if the availability and promotion of cheap, energy-dense food increases globally throughout several decades, the traditionally high levels of active transport in the Netherlands will help to attenuate this obesogenic effect in the Dutch population; whereas the traditionally high levels of car transport in the USA will accentuate this obesogenic effect in the American population. Similarly, cultural body-size preferences could moderate the drivers such that increased food access will probably have a larger effect in

Panel 2: Energy balance flipping point

In most high-income countries, the energy expenditure needed for daily life has decreased since the beginning of the 20th century because of increasing mechanisation, urbanisation, motorisation, and computerisation. All else being equal, the expected result of an increasingly sedentary lifestyle would have been weight gain. Therefore, why did obesity prevalence not rise substantially until the 1970s?

A parallel reduction of food energy intake coupled with decreasing physical activity is one possible explanation for the low prevalence of obesity during the first three-quarters of the 20th century. Data from the US Department of Agriculture lend support to this explanation since they clearly show a reduction of per-person energy available in the food supply (food production plus imports minus exports and non-human use) from early in the 20th century until the 1960s[58] (figure 3), mainly because of reduced consumption of wheat products.[59,60] A 1948 editorial in the *Journal of the American Medical Association*[61] attributed the fall in US wheat consumption partly to the decreased energy demands resulting from reduction of hard labour. Presciently, the same editorial warned of a potential rise in obesity prevalence if consumption of sugars, sweets, and fats began to increase. Indeed, the 1970s saw a striking rise in the quantity of refined carbohydrates and fats in the US food supply,[62,63] which was paralleled by a sharp increase in the available calories (figure 3) and the onset of the obesity epidemic. As Cutler and colleagues[35] note, there was "a revolution in the mass preparation of food that is roughly comparable to the mass production revolution in manufactured goods that happened a century ago," that "lowered the time price of food consumption". The increased availability and marketing of cheap, readily available food was so great that food waste has progressively increased by about 50% since the 1970s.[55]

We postulate that an energy balance flipping point has occurred in most high-income countries in the past century with two distinct phases: the so-called move less, stay lean phase (1910–60), characterised by decreasing physical activity levels and energy intake, and a population that remained lean; and the subsequent so-called eat more, gain weight phase, characterised by increasing energy intake and a concomitant rise in population weight. In each of these phases, we postulate that energy balance was predominantly achieved by different mechanisms: decreasing energy intake (through appetite mechanisms) to match decreasing expenditure, followed by increasing energy expenditure (through increased weight and, thus, resting metabolic rate) to match increasing intake. Thus, decreases in physical activity in the first phase were probably able to pull down energy intake, because intake was being matched to expenditure. The result was that obesity rates did not increase during this period despite widespread uptake of mechanisation and motorisation. The second phase seems to have been ushered in by an energy balance flipping point, when energy intake rose because of environmental push factors (ie, increasingly available, cheap, tasty, highly promoted obesogenic foods). The concomitant rise in weight was the physiological mechanism for restoration of energy balance.[64] Food supply data from the US lends support to this flipping point hypothesis, but it needs to be tested in other countries and with more diverse datasets.

Tongan women, where large body size is a positive attribute,[68] than in Japanese women, where small body size is deemed ideal.[69]

The effects of the environmental moderators are shown in figure 4 as affecting the rise in obesity but are not depicted as part of the direct driver pathways.

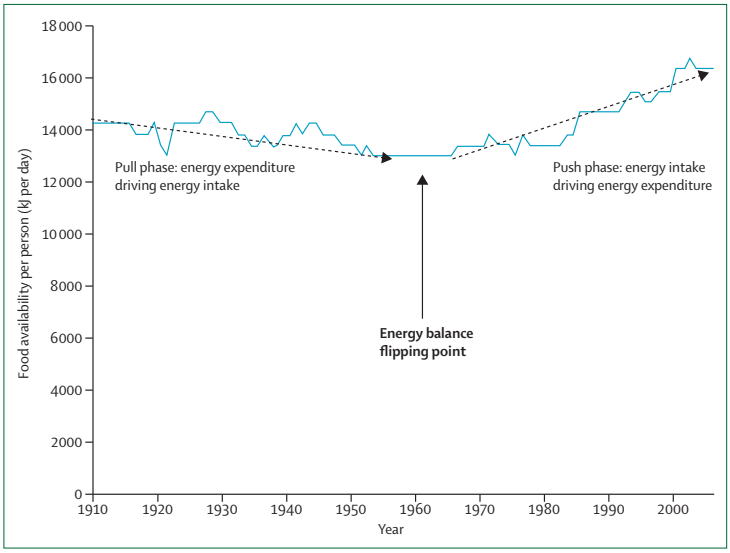


Figure 3: **Food availability for the USA, 1910–2006[58]**
There are two distinct phases: a decrease in food energy supply (postulated to be pulled down by reduced energy expenditure requirements for daily living), followed by an increase in food energy supply (postulated to be pushed up by increasing food access). An energy balance flipping point is proposed, marking the change in how the US population generally achieved energy balance.

However, we recognise that this scheme is simplified and that factors such as active transport environments and body-size preferences could change sufficiently with time to drive changes in obesity prevalence. Indeed, for some countries such as China, rapid urbanisation and motorisation of the population could be judged a major driver of the rise in obesity.[70] Although the interaction between all the environmental factors is undoubtedly complex and multifaceted (panel 3), the strong directional force of the systemic and environmental drivers is evidenced by the continued rise of the epidemic globally.

Other environmental and individual effects

The existing environments within a country (eg, the built environment, transport systems, active recreation opportunities, cuisines and food culture, and culture around body size) can greatly moderate or modulate the effects of the global obesity drivers on population BMI (figure 4). These effects can be powerful and help to explain much of the differences in obesity prevalence between populations. They clearly hold opportunities for interventions to make environments less obesogenic. Interventions such as increasing the price of unhealthy food and beverages[81–83] or decreasing the price of healthy foods[84] have received attention in recent years, but there has been very little research into understanding and changing the powerful sociocultural determinants of food choices, physical activity, and body-size perception.[85] These determinants should be a priority for research if the high prevalence of obesity in specific ethnic groups is to be addressed.[86]

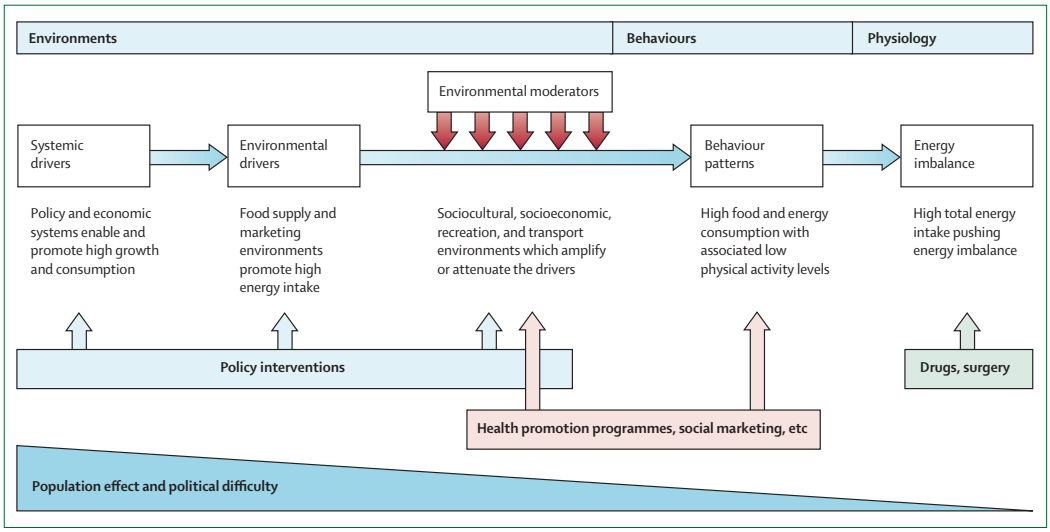


Figure 4: **A framework to categorise obesity determinants and solutions**
The more distal drivers are to the left and the environmental moderators that have an attenuating or accentuating effect are shown, along with some examples. The usual interventions for environmental change are policy based, whereas health promotion programmes can affect environments and behaviours. Drugs and surgery operate at the physiological level. The framework shows that the more upstream interventions that target the systemic drivers might have larger effects, but their political implementation is more difficult than health promotion programmes and medical services.

Within a given environment, BMI differs between individuals, which is indicative of individual moderators or modulators of energy balance. For example, too little sleep and impaired sleep quality have been shown to have important effects on weight gain.[87] Genetic effects are one of the most highly researched determinants of obesity. However, apart from the identification of a few rare, single genetic abnormalities, the quest to identify genes related to common obesity (obesity not caused by an underlying condition) has been disappointing. The potential epigenetic effects of behavioural and environmental factors on genetic expression are now receiving much attention (panel 4).

The role of individual choice in both the causes of and solutions to obesity has great appeal because of its simplicity. However, this notion is in fact highly contentious and is indicative of the tension between the simple and the complex that is ever present in debates about obesity. Undoubtedly the final decision to consume a particular food or beverage, or to exercise or not, is an individual decision. However, to negotiate the complexity of the environment and the choices it poses, many of these decisions are automatic or subconscious. Cohen[98] puts forward a strong argument that "excessive food consumption occurs in ways that defy personal insight or are below individual awareness", and provides plausible neurophysiological mechanisms for many automatic responses to physiological, environmental, and interpersonal cues in relation to eating and physical activity opportunities. Appetite physiologists label this response as "passive overconsumption".[99] Further increase in choice, especially through more processed products on supermarket shelves or more items on fast food menus, is unlikely to reduce overconsumption and could have the opposite effect. Furthermore, cognitive strategies to combat overconsumption, such as weight-loss diets, can be successful for some individuals but are unlikely to be population solutions. Even weight-loss diets that are supported by trial data, such as high protein or low glycaemic index diets,[100] might not be suitable as solutions for global obesity because of their detrimental effect on the environment (eg, high meat protein diets[101]) or staple food production (eg, rice).

Approaches to address the obesity epidemic

Approaches to address the obesity epidemic are broadly categorised within figure 4, with a detailed discussion of solutions presented in the report by Gortmaker and colleagues[75] in this Series.

The effects on population trajectories of obesity prevalence are likely to fundamentally differ between an intervention aimed at motivating behavioural changes (eg, health promotion programmes, social marketing, education) and policy interventions (in this context, meaning enforceable actions such as laws and regulations) that reverse the environmental drivers (eg, reducing the cost of healthy foods and increasing the costs of unhealthy foods). The interventions to motivate behavioural changes could be regarded as counteractions (ie, they counteract drivers of increasingly obesogenic environments by acting on some of their mediators) and they might have important obesity prevention effects, especially in children, if applied to a whole community.[102,103] However, sustainability and affordability are the two major continuing challenges, even for programmes with proven effectiveness.

Panel 3: Modelling for obesity

The Obesity System Map,[71] introduced by the UK Foresight Programme in 2007, is the most comprehensive compilation so far of the determinants of obesity and their inter-relations. The map clearly shows the enormous complexity of the causal relations relevant to obesity by linking physiology, individual behaviours, and environmental variables that stretch from local to global levels. In addition to the interdependence between factors and the importance of feedback loops, there are other characteristics of the obesity problem that are not shown. These characteristics include the heterogeneity of the factors that need to be involved in systems change, the diversity of environments where change needs to take place, the non-linear relations that govern the system's dynamics, and the sometimes random or quantum nature of behaviour change.[72] This complexity creates an apparently intractable or so-called wicked problem,[73] and the imperative to turn around the obesity epidemic clashes with its wicked nature. In aiming to establish and prioritise solutions, the challenge is to reduce the complexity of obesity enough so that it can be understood by researchers, policy makers, and the public without becoming overly simplistic. Some progress in this challenge lies in mathematical modelling applied to complex health and behavioural questions.[64,74,75]

In the past 5 years, several mathematical models have been developed by use of various modelling techniques.[76] As with other models, they can have descriptive, explanatory, or evaluative aims.[77] Figure 5 shows the aims for obesity modelling at the population level. Descriptive studies quantify the present burden and potential future trends of obesity. Explanatory studies analyse the causes of the rise in obesity prevalence with time and the variability across populations. Evaluative models assess the likely effect of interventions to reduce future prevalence.

Mathematical models are "a set of assumptions together with implications drawn from them by mathematical reasoning".[74] Compared with conceptual models, they have the valuable attributes of forcing theoretical precision, making assumptions transparent, promoting data analysis and hypothesis testing, and having many practical applications.[74] Since models match reality in important ways but are far simpler than reality (they draw attention to some aspects of reality but ignore others[78]), they can help us understand complex problems such as obesity and its solutions.[72] Indeed, mathematical modelling has been crucial to understand and to respond to other global threats, such as climate change[79] and infectious diseases.[80]

Irrespective of their goal, all mathematical models need a logical framework linking the pathways between each of the factors in the model, accompanied by tight specifications of the assumptions and quantitative estimates underpinning each of the links. Complex frameworks, such as the Foresight Obesity System Map, attempt to incorporate most individual and population-level determinants. However, substantial simplification of models can arise when only populations are included (determinants causing individual variability drop out), when questions are narrowed (determinants that have not changed with time can be dropped for questions about the rise of obesity), or when solutions are being modelled (factors not included in the intervention can be dropped). These methods can be expected to provide the core techniques to understand and to respond to complex non-communicable disorders such as obesity.[51,64,75]

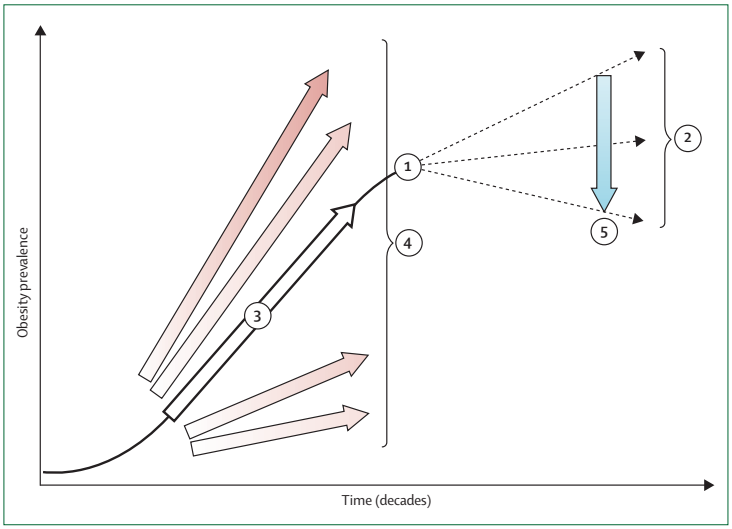


Figure 5: **A schematic diagram of the major uses of modelling of population prevalence of obesity**
The black line is the trajectory of increases in obesity prevalence with time, with the present burden (1) and projected future burdens (2) being descriptive uses of modelling. Explanatory uses of modelling include explaining the rise in obesity with time (3) and the differences in prevalence rates (4). Evaluative uses include assessment of the potential for solutions to reduce the future burden of obesity (5).

Panel 4: How much can genetics explain of the obesity epidemic?

One useful way to think about the relation of genes with obesity was expressed by George Bray when he said, "the genetic background loads the gun, but the environment pulls the trigger".[88] The rise of obesity prevalence throughout the past few decades clearly cannot be accounted for by population genetic changes. The heritability of body-mass index (BMI) is often cited as 40–70%,[89] yet large genome-wide association studies to identify common single nucleotide polymorphisms (SNPs) associated with BMI have been unable to explain more than a small proportion (<2%) of BMI variability.[90,91] Heritability is often misinterpreted as being the proportion of BMI variance caused by genetics; whereas heritability studies, rather than estimating cause, estimate the proportion of genetic variance and non-genetic variance that explains BMI variance in a given study population. The study populations are usually based on twin cohorts which, even if reared apart, live within very narrow environmental variances (rather than the wide variance of environments across countries shown in figure 2), which results in high heritability estimates. These estimates seriously understate the environmental effects and thus provide no indication of the potential to reduce obesity via environmental change.[92] Part of the missing heritability of obesity might be the result of small individual contributions of many SNPs that genome-wide association studies have been underpowered to detect, or some rare variants that are not identified by common SNPs. Much of the variability between individuals in bodyweight might be attributable to gene–environment or gene–behaviour interactions,[93] including interactions in the intrauterine environment.[94] An intriguing possibility mediating the gene–environment interaction is the potential contribution of epigenetic mechanisms that modify the expression of genes. For example, DNA methylation patterns can be affected by maternal diet; these epigenetic modifications can persist for decades,[95,96] and possibly be inherited by future generations.[97] Epigenetic contributions to obesity will need to be addressed through minimisation of the environmental triggers, rather than manipulation of the genetic guns, in the first instance.

Furthermore, such programmes do not address the underlying drivers of the epidemic.

Interventions that aim to reverse obesogenic drivers (and some of the environmental moderators) will almost all be policy-led—mainly government policy (eg, shifting agricultural polices to incorporate health outcomes, banning unhealthy food marketing to children, healthy public sector food service policies) but some could be food industry policies (eg, moving product formulation towards healthier compositions, self-regulation of marketing to children). Policy-led solutions that apply to environments and affect the whole population have several strengths compared with health education and promotion programmes.[104] They tend to be sustainable, affect the whole population (including those who are difficult to reach), become systemic (affect default behaviours), and reverse some of the environmental drivers. The degree of political difficulty for implementation of policy and regulatory interventions is typically much higher than that for programme-based and education-based interventions[105] (figure 4). Reasons for this reluctance to enact affordable and cost-effective policies include the powerful lobby force of the food (and allied) industries against government regulation of the food market[106,107] and public reluctance to change environments to which they have become accustomed (such as car access and cheap parking in cities, and high fat and sugar food choices in canteens). Nevertheless, the experience with trans-fatty acids in Denmark, where legislation was introduced to restrict their use in food production, is an example of a cost-effective government food policy that was successfully enacted for population health benefit.[108]

Policy interventions for obesity can only be realistically directed at the environment (making healthy choices easier) rather than the individual (compelling them to take the healthy choices). Unlike other public health issues for which enforceable policies can directly require specific behaviours (such as wearing a seat belt or not smoking in offices), there are no regulations that will require people to eat, or not eat, certain foods and to exercise,[66] with the possible exception of a few rules operating in school environments. For this reason, obesity prevention policies do not proscribe particular eating and physical activity behaviours and are thus much less intrusive of human liberties than many policies already in place to control other public health problems. The major strategies available to directly affect behaviours aim to increase the motivation to make healthy choices, and include social marketing, health education, and health promotion programmes.

Implications

In this report, we have provided an overview of the global obesity epidemic, describing the size and nature of the problem, discussing its drivers, and mapping out key contributing factors. We draw attention to the food

system, operating through the energy intake side of the energy balance equation, as the dominant driver of the rise in obesity, although many other environmental and individual factors modulate the effect of the drivers on obesity prevalence in populations and obesity presence in individuals. The highly complex global, national, and local food systems, although feeding the world's population, are nevertheless falling short by both promoting overnutrition (overweight or obesity) in an estimated 2 billion adults and children[9,10] and not reaching the further 1 billion people who are undernourished.[109] Both these forms of malnutrition are major contributors to preventable population disease burden and are thus substantial barriers to meeting the Millennium Development Goals.[110] With the global population expected to grow to 9 billion people by 2050,[111] and with increased disruption to food production from climate change—such as increased droughts, floods, hurricanes, and global temperatures[101]—the global food system will be placed under great stress which will probably result in more unequal distribution of calories. Policies that address the food system drivers of obesity will therefore be an important part of the much broader imperative to re-orient food systems towards health and development, particularly in vulnerable populations.

As argued in this report and elsewhere,[34] the economic priorities and policies that promote consumption-based growth, and the regulatory policies that promote market and trade liberalisation have produced many benefits but are now increasingly regarded as contributing to the global crises of overconsumption in general. Obesity is but one of these crises, as the private sector becomes ever more effective in its exploitation of basic human biological drives, desires, and weaknesses. Solutions to obesity and to improve health and development cannot be based on the existing framework (consumption-driven growth creating financially-defined prosperity) because this approach has helped to create the difficulties in the first place. Governments and international organisations such as the UN need to provide global leadership on these issues and not abdicate them to the private sector.[112,113] Actions that reduce the economic incentives for over consumption in general (eg, a price on carbon and subsidies on public transport) can have benefits for both human and environmental health.[34] Moreover, a new framework is needed that is based on broadly defined sustainable economics for prosperity (including economic, social, health, and environmental outcomes).[114]

Four decades after the onset of the global obesity epidemic, the awareness of the threat of obesity to population health and wellbeing can be seen in the plethora of national reports and strategic plans from many countries, yet their conversion to action remains largely unrealised (Australia has had three such unfulfilled national taskforce plans since 1997[115–117]). However, recognition in low-income and middle-income countries that the growing contribution of obesity to its burden of malnutrition is a threat to national development remains low. Some governments are making promising moves to convert the rhetoric into action. For example, in the mid to late 2000s, the Government in England developed a series of initiatives such as the Foresight report,[26] the cross government strategy for obesity,[118] restrictions on unhealthy food marketing to children,[119] and a yearly national child measurement programme.[118] Some other countries, notably Brazil, have made substantial steps in national monitoring programmes, restricting marketing to children, and improving school food; and the US White House Task Force on Childhood Obesity[120] holds much promise to bring about change, because of its high-level political commitment. International leadership from the UN and its agencies, and more national leadership from countries will be essential to solve this great challenge of the global obesity epidemic, which is why the UN High Level Meeting on NCDs in September, 2011, is so central to global population health and development.

Contributors

All authors jointly formulated the major ideas, and read and approved the final version of the report. BAS led the writing of the report. GS assisted in writing the report, and collated and included comments from all authors. KDH co-drafted the sections about genetics and the food supply and commented on the remainder of the paper. MLM co-drafted the section about market failure and commented on the remainder of the report. KM, DTF, and SLG provided comments on the report.

Conflicts of interest

We declare that we have no conflicts of interest.

Acknowledgments

This work was done under auspices of the Collaborative Obesity Modeling Network as part of the Envision Project, supported by the National Collaborative on Childhood Obesity Research, which coordinates childhood obesity research across the National Institutes of Health (NIH), Centers for Disease Control and Prevention (CDC), Department of Agriculture, and the Robert Wood Johnson Foundation (RWJF). This work was supported in part by grants from RWJF (grant numbers 260639, 61468 and 66284), CDC (U48/DP00064-00S1 and 1U48DP001946), including the Nutrition and Obesity Policy, Research and Evaluation Network, and the Office of Behavioral and Social Sciences Research of NIH. This work is solely the responsibility of the authors and does not represent official views of the Centers for Disease Control and Prevention or any of the other funders.

References

1 WHO. Global strategy on diet, physical activity and health. Geneva: World Health Organization, 2004. http://www.who.int/dietphysicalactivity/en/ (accessed June 23, 2011).

2 Peeler CL, Kolish ED, Enright M, Burke C. The children's food & beverage advertising initiative in action. A report on compliance and implementation during 2009. Arlington: Council of Better Business Bureaus, 2010.

3 European Union. EU Pledge 2009 Monitoring Report, September 09. Accenture Media Services, 2009.

4 World Cancer Research Fund and American Institute for Cancer Research. food, nutrition, physical activity, and the prevention of cancer: a global perspective, 2007. Washington, DC: American Institute for Cancer Research, 2007.

5 Koplan J, Liverman C, Kraak V, eds. Preventing childhood obesity: health in the balance. Washington, DC: Institute of Medicine of the National Academies, The National Academies Press, 2005.

6 Sassi F, Devaux M, Cecchini M, Rusticelli E. The obesity epidemic: analysis of past and projected future trends in selected OECD countries. Paris: Organisation for Economic Co-operation and Development (OECD), Directorate for Employment, Labour And Social Affairs, Health Committee, 2009.

7 WHO. Reducing risks, promoting healthy life. Geneva: World Health Organization, 2005.

8 WHO. Global strategy on diet, physical activity and health: a framework to monitor and evaluate implementation. Geneva: World Health Organization, 2008.

9 Finucane MM, Stevens GA, Cowan MJ, et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9·1 million participants. *Lancet* 2011; **377:** 557–67.

10 Lobstein T, Baur L, Uauy R, for the IASO International Obesity TaskForce. Obesity in children and young people: a crisis in public health. *Obes Rev* 2004; **5** (suppl 1)**:** 4–104.

11 Wang Y, Lobstein T. Worldwide trends in childhood overweight and obesity. *Int J Pediatr Obes* 2006; **1:** 11–25.

12 International Association for the Study of Obesity, International Obesity Taskforce. Global childhood overweight. http://www.iaso.org/site_media/uploads/Global_Childhood_Overweight_April_2011.pdf (accessed June 23, 2011).

13 Matsushita Y, Yoshiike N, Kaneda F, Yoshita K, Takimoto H. Trends in childhood obesity in Japan over the last 25 years from the national nutrition survey. *Obes Res* 2004; **12:** 205–14.

14 Monteiro CA, Moura EC, Conde WL, Popkin BM. Socioeconomic status and obesity in adult populations of developing countries: a review. *Bull World Health Organ* 2004; **82:** 940–46.

15 Mendez MA, Monteiro CA, Popkin BM. Overweight exceeds underweight among women in most developing countries. *Am J Clin Nutr* 2005; **81:** 714–21.

16 Monteiro CA, Conde WL, Popkin BM. Income-specific trends in obesity in Brazil: 1975–2003. *Am J Public Health* 2007; **97:** 1808–12.

17 International Association for the Study of Obesity, International Obesity Taskforce. Global obesity prevalence in adults. http://www.iaso.org/site_media/uploads/Prevalence_of_Adult_Obesity_May_2011_New.pdf. accessed 23/6/2011 (accessed June 23, 2011).

18 Ezzati M, Lopez A, Rodgers AD, Murray CJL, eds. Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors. Geneva: World Health Organization, 2004.

19 Australian Institute of Health and Welfare (AIHW) and National Heart Foundation of Australia. The relationship between overweight, obesity and cardiovascular disease. Canberra: AIHW (Cardiovascular Disease Series No. 23), 2004.

20 WHO. Prevention and control of noncommunicable diseases: implementation of the global strategy. Sixty-first World Health Assembly, Resolution WHA61.14. Geneva: World Health Organisation, 2007.

21 Lopez A, Mathers C, Ezzati M, Jamison D, Murray C. Global burden of disease and risk factors. Washington: The World Bank, 2006. http://www.dcp2.org/pubs/GBD (accessed June 23, 2011).

22 Hoad V, Somerford P, Katzenellenbogen J. High body mass index overtakes tobacco as the leading independent risk factor contributing to disease burden in Western Australia. *Aust NZ J Public Health* 2010; **34:** 214–15.

23 Stewart ST, Cutler DM, Rosen AB. Forecasting the effects of obesity and smoking on U.S. life expectancy. *N Engl J Med* 2009; **361:** 2252–60.

24 Olshansky SJ, Passaro DJ, Hershow RC, et al. A potential decline in life expectancy in the United States in the 21st century. *N Engl J Med* 2005; **352:** 1138–45.

25 Ford E, Ajani U, Croft J, Critchley J, Labarthe D, Kottke T. Explaining the decrease in U.S. deaths from coronary disease, 1980–2000. *N Engl J Med* 2007; **356:** 2388–98.

26 Butland B, Jebb S, Kopelman P, et al. Foresight. Tackling obesities: future choices—project report. London: Government Office for Science, 2007.

27 Rokholm B, Baker J, Sorensen T. The levelling off of the obesity epidemic since the year 1999—a review of evidence and perspectives. *Obes Rev* 2010; **11:** 835–46.

28 Rush E, Plank L, Chandu V, et al. Body size, body composition, and fat distribution: a comparison of young New Zealand men of European, Pacific Island, and Asian Indian ethnicities. *NZ Med J* 2004; **117:** U1203.

29 Lobstein T, Leach R. Foresight. Tackling obesities: future choices—international comparisons of obesity trends, determinants and responses—evidence review, adults. London: Government Office for Science, 2007.

30 Keke K, Phongsavan P, Dan L, et al, for Nauru Ministry of Health, WHO, and the Centre for Physical Activity and Health, University of Sydney. Nauru NCD Risk Factors STEPS Report. Suva, Fiji: World Health Organisation, 2007.

31 Franco M, Orduñez P, Caballero B, et al. Impact of energy intake, physical activity, and population-wide weight loss on cardiovascular disease and diabetes mortality in Cuba, 1980–2005. *Am J Epidemiol* 2007; **166:** 1374–80.

32 Popkin B. The nutrition transition and its health implications in lower-income countries. *Public Health Nutr* 1998; **1:** 5–21.

33 Barker DJ, Winter PD, Osmond C, Margetts B, Simmonds SJ. Weight in infancy and death from ischaemic heart disease. *Lancet* 1989; **334**: 577–80.

34 Egger G, Swinburn B. Planet obesity: how we're eating ourselves and the planet to death. Sydney: Allen and Unwin, 2010.

35 Cutler DM, Glaeser EL, Shapiro JM. Why have Americans become more obese? *J Econ Perspect* 2003; **17:** 93–118.

36 Philipson TJ, Posner RA. The long-run growth in obesity as a function of technological change. *Perspect Biol Med* 2003; **46** (suppl 3)**:** S87–107.

37 Finkelstein EA, Ruhm CJ, Kosa KM. Economic causes and consequences of obesity. *Annu Rev Public Health* 2005; **26**: 239–57.

38 Swinburn B, Sacks G, Lobstein T, et al. The 'Sydney Principles' for reducing the commercial promotion of foods and beverages to children. *Public Health Nutr* 2008; **11:** 881–86.

39 Morley B, Chapman K, Mehta K, King L, Swinburn B, Wakefield M. Parental awareness and attitudes about food advertising to children on Australian television. *Aust NZ J Public Health* 2008; **32:** 341–47.

40 Tickner J, Raffensperger C, Myers N. The precautionary principle in action: a handbook. Windsor, ND: Science and Environmental Health Network, 1999.

41 Moodie R, Swinburn B, Richardson J, Somaini B. Childhood obesity—a sign of commercial success, but a market failure. *Int J Pediatr Obes* 2006; **1:** 133–38.

42 McCormick B, Stone I, for the Corporate Analytical Team. Economic costs of obesity and the case for government intervention. *Obes Rev* 2007; **8** (suppl 1)**:** 161–64.

43 Jan S, Mooney GH. Childhood obesity, values and the market. *Int J Pediatr Obes* 2006; **1:** 131–32.

44 Anand P, Gray A. Obesity as market failure: could a 'deliberative economy' overcome the problems of paternalism? *Kyklos* 2009; **62:** 182–90.

45 Kuchler F, Golan E. Is there a role for government in reducing the prevalence of overweight and obesity? *Choices* 2004; **Fall**: 41–45.

46 Freebairn J. Taxation and obesity? *Aust Econ Rev* 2010; **43:** 54–62.

47 Mazzocchi M, Traill W, Shogren J. Fat economics: nutrition, health and economic policy. Oxford: Oxford University Press, 2009.

48 McCarthy M. The economics of obesity. *Lancet* 2004; **364:** 2169–70.

49 Institute of Medicine. Food marketing to children and youth. Threat or opportunity? Washington: National Academy of Sciences, 2006.

50 Sassi F. Obesity and the Economics of Prevention. Fit not fat. Paris: Organisation for Economic Co-operation and Development, 2010.

51 Wang Y, McPherson K, Marsh T, Gortmaker SL, Brown M. Health and economic burden of the projected obesity trends in the US and the UK. *Lancet* 2011; **378:** 815–25.

52 WHO. Ottawa Charter for Health Promotion. Ottawa: First International Conference on Health Promotion; 1986. Report No: WHO/HPR/HEP/95.1.

53 Frank LD, Andresen MA, Schmid TL. Obesity relationships with community design, physical activity, and time spent in cars. *Am J Prev Med* 2004; **27:** 87–96.

54 Kitchen P, Brignell J, Li T, Spickett-Jones G. The emergence of IMC: a theoretical perspective. *J Advertising Res* 2004; **March:** 19–30.

55 Hall KD, Guo J, Dore M, Chow CC. The progressive increase of food waste in America and its environmental impact. *PLoS ONE* 2009; **4:** e7940.

56 Swinburn B, Sacks G, Ravussin E. Increased food energy supply is more than sufficient to explain the US epidemic of obesity. *Am J Clin Nutr* 2009; **90:** 1453–56.

57 Scarborough P, Burg MR, Foster C, et al. Increased energy intake entirely accounts for increase in body weight in women but not in men in the UK between 1986 and 2000. *Br J Nutr* 2011; **105:** 1399–404.

58 Economic Research Service, US Department of Agriculture. Food availability (per capita) data system. Washington, DC: US Department of Agriculture. http://www.ers.usda.gov/Data/FoodConsumption (accessed June 23, 2011).

59 Gerrior S, Bente L, Hiza H. Nutrient content of the U.S. Food Supply. 1909–2000. Washington, DC: US Department of Agriculture, Center for Nutrition Policy and Promotion, 2004.

60 Putnam J. Major trends in the U.S. food supply, 1909–99. *Food Rev* 2000; **23:** 8–15.

61 Anonymous. Nutritional contributions of wheat. *JAMA* 1948; **138:** 972–73.

62 Putnam J, Allshouse J, Kantor L. US per capita food supply trends: more calories, refined carbohydrates, and fats. *Food Rev* 2002; **25:** 2–15.

63 Putnam J KL, Allshouse J. Per capita food supply trends: progress towards dietary guidelines. *Food Rev* 2000; **23:** 2–14.

64 Hall KD, Sacks G, Chandramohan D, et al. Quantifying the effect of energy imbalance on bodyweight change. *Lancet* 2011; **378:** 826–37.

65 Kumanyika S, Jeffery RW, Morabia A, Ritenbaugh C, Antipatis VJ, for the Public Health Approaches to the Prevention of Obesity Working Group of the International Obesity Taskforce. Obesity prevention: the case for action. *Int J Obes Relat Metab Disord* 2002; **26:** 425–36.

66 Sacks G, Swinburn B, Lawrence M. Obesity Policy Action framework and analysis grids for a comprehensive policy approach to reducing obesity. *Obes Rev* 2009; **10:** 76–86.

67 Wilkinson RG, Pickett KE. Income inequality and population health: a review and explanation of the evidence. *Soc Sci Med* 2006; **62:** 1768–84.

68 Mavoa HM, McCabe M. Sociocultural factors relating to Tongans' and Indigenous Fijians' patterns of eating, physical activity and body size. *Asia Pac J Clin Nutr* 2008; **17:** 375–84.

69 Hayashi F, Takimoto H, Yoshita K, Yoshiike N. Perceived body size and desire for thinness of young Japanese women: a population-based survey. *Br J Nutr* 2006; **96:** 1154–62.

70 Bell AC, Ge K, Popkin BM. The road to obesity or the path to prevention: motorized transportation and obesity in China. *Obes Res* 2002; **10:** 277–83.

71 Vandenbroeck IP, Goossens J, Clemens M. Foresight. Tackling obesities: future choices—building the Obesity System Map. London: Government Office for Science, 2007.

72 Finegood DT. The complex systems science of obesity. In: Cawley JH, ed. The Oxford handbook of the social science of obesity. Oxford, UK: Oxford University Press, 2011: 208–36.

73 Rittel H, Webber M. Dilemmas in a general theory of planning. *Policy Sci* 1973; **4:** 155–69.

74 Rodgers JL. The epistemology of mathematical and statistical modeling: a quiet methodological revolution. *Am Psychol* 2010; **65:** 1–12.

75 Gortmaker SL, Levy D, Carter R, et al. Changing the future of obesity: science, policy, and action. *Lancet* 2011; **378:** 838–47.

76 Levy DT, Mabry PL, Wang YC, et al. Simulation models of obesity: a review of the literature and implications for research and policy. *Obes Rev* 2010; **12:** 387–94.

77 Jacobs P, Rapoport J. Economics of health and medical care. 5th edn. Sudbury: Jones and Bartlett Publishers, 2004.

78 Pearl J. Causality: Models, reasoning, and inference. Cambridge: Cambridge University Press, 2000.

79 McMichael A, Campbell-Lendrum D, Corvalan C, et al, eds. Climate change and human health: risks and responses. Geneva: World Health Organization, 2003.

80 Weiss RA, McMichael AJ. Social and environmental risk factors in the emergence of infectious diseases. *Nat Med* 2004; **10** (suppl 12)**:** S70–76.

81 Sacks G, Veerman JL, Moodie M, Swinburn B. 'Traffic-light' nutrition labelling and 'junk-food' tax: a modelled comparison of cost-effectiveness for obesity prevention. *Int J Obes (Lond)* 2010; published online Nov 16. DOI:10.1038/ijo.2010.228.

82 Mytton O, Gray A, Rayner M, Rutter H. Could targeted food taxes improve health? *J Epidemiol Community Health* 2007; **61:** 689–94.

83 Brownell KD, Frieden TR. Ounces of prevention—the public policy case for taxes on sugared beverages. *N Engl J Med* 2009; **360**: 1805–08.

84 Ni Mhurchu C, Blakely T, Jiang Y, Eyles HC, Rodgers A. Effects of price discounts and tailored nutrition education on supermarket purchases: a randomized controlled trial. *Am J Clin Nutr* 2010; **91:** 736–47.

85 McCabe MP, Ricciardelli LA. Parent, peer and media influences on body image and strategies to both increase and decrease body size among adolescent boys and girls. *Adolescence* 2001; **36:** 225–40.

86 Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999–2004. *JAMA* 2006; **295:** 1549–55.

87 Chaput JP, Despres JP, Bouchard C, Tremblay A. The association between sleep duration and weight gain in adults: a 6-year prospective study from the Quebec Family Study. *Sleep* 2008; **31:** 517–23.

88 Bray GA. The epidemic of obesity and changes in food intake: the Fluoride Hypothesis. *Physiol Behav* 2004; **82:** 115–21.

89 Comuzzie AG, Allison DB. The search for human obesity genes. *Science* 1998; **280:** 1374–77.

90 Bogardus C. Missing heritability and GWAS utility. *Obesity (Silver Spring)* 2009; **17:** 209–10.

91 Hebebrand J, Volckmar AL, Knoll N, Hinney A. Chipping away the 'missing heritability': GIANT steps forward in the molecular elucidation of obesity—but still lots to go. *Obes Facts* 2010; **3:** 294–303.

92 Jencks C. Heredity, environment, and public policy reconsidered. *Am Sociol Rev* 1980; **45:** 723–36.

93 Bouchard C. Gene-environment interactions in the etiology of obesity: defining the fundamentals. *Obesity (Silver Spring)* 2008; **16** (suppl 3)**:** S5–10.

94 Yajnik CS. Fetal programming of diabetes: still so much to learn! *Diabetes Care* 2010; **33:** 1146–48.

95 Ahmed F. Epigenetics: Tales of adversity. *Nature* 2010; **468:** S20.

96 Stein AD, Kahn HS, Rundle A, Zybert PA, van der Pal–de Bruin K, Lumey L. Anthropometric measures in middle age after exposure to famine during gestation: evidence from the Dutch famine. *Am J Clin Nutr* 2007; **85:** 869–76.

97 Lange UC, Schneider R. What an epigenome remembers. *Bioessays* 2010; **32:** 659–68.

98 Cohen DA. Neurophysiological pathways to obesity: below awareness and beyond individual control. *Diabetes* 2008; **57:** 1768–73.

99 Blundell JE, King NA. Over-consumption as a cause of weight gain: behavioural-physiological interactions in the control of food intake (appetite). *Ciba Found Symp* 1996; **201:** 138–54.

100 Larsen TM, Dalskov SM, van Baak M, et al. Diets with high or low protein content and glycemic index for weight-loss maintenance. *N Engl J Med* 2010; **363:** 2102–13.

101 McMichael AJ, Powles JW, Butler CD, Uauy R. Food, livestock production, energy, climate change, and health. *Lancet* 2007; **370:** 1253–63.

102 Sanigorski AM, Bell AC, Kremer PJ, Cuttler R, Swinburn BA. Reducing unhealthy weight gain in children through community capacity-building: results of a quasi-experimental intervention program, Be Active Eat Well. *Int J Obes (Lond)* 2008; **32:** 1060–67.

103 de Silva-Sanigorski AM, Bell AC, Kremer P, et al. Reducing obesity in early childhood: results from Romp & Chomp, an Australian community-wide intervention program. *Am J Clin Nutr* 2010; **91:** 831–40.

104 Swinburn B, Egger G. Analyzing and influencing obesogenic environments. In: Bray G, Bouchard C, eds. Handbook of obesity: clinical applications. 3rd edn. New York: Informa Health Care, 2008: 177–93.

105 James P, Rigby N. Developing the political climate for action. In: Waters E, Swinburn B, Seidell J, Uauy R, eds. Preventing childhood obesity: evidence, policy and practice. Oxford, UK: Blackwell, 2010: 212–19.

106 Corporate Europe Observatory. A red light for consumer information: the food industry's €1-billion campaign to block health warnings on food. Brussels: Corporate Europe Observatory, 2010.

107 Brownell KD, Warner KE. The perils of ignoring history: big tobacco played dirty and millions died. How similar is Big Food? *Milbank Q* 2009; **87:** 259–94.

108 Stender S, Dyerberg J, Astrup A. High levels of industrially produced trans fat in popular fast foods. *N Engl J Med* 2006; **354:** 1650–52.

109 Food and Agriculture Organization of the United Nations. Global hunger declining, but still unacceptably high. Geneva, Switzerland: Food and Agriculture Organization of the United Nations, Economic and Social Development Department, 2010. http://www.fao.org/docrep/012/al390e/al390e00.pdf (accessed June 23, 2011).

110 Geneau R, Stuckler D, Stachenko S, et al. Raising the priority of preventing chronic diseases: a political process. *Lancet* 2010; **376:** 1689–98.

111 Department of Economic and Social Affairs, Population Division. World population prospects: the 2008 revision. Population newsletter. Geneva: United Nations Secretariat, 2009. http://www.un.org/esa/population/publications/popnews/Newsltr_87.pdf (accessed June 23, 2011).

112 Yach D, Khan M, Bradley D, Hargrove R, Kehoe S, Mensah G. The role and challenges of the food industry in addressing chronic disease. *Global Health* 2010; **6:** 10.

113 Ludwig DS, Nestle M. Can the food industry play a constructive role in the obesity epidemic? *JAMA* 2008; **300:** 1808–11.

114 Jackson T. Prosperity without growth? The transition to a sustainable economy. London: The Sustainable Development Commission, 2009. http://www.sd-commission.org.uk/data/files/publications/prosperity_without_growth_report.pdf (accessed June 23, 2011).

115 National Health and Medical Research Council (NHMRC). Economic issues in the prevention and treatment of overweight and obesity. Acting on Australia's weight: a strategic plan for the prevention of overweight and obesity. Canberra: NHMRC, 1997.

116 National Preventative Health Taskforce. Australia: the healthiest country by 2020—National Preventative Health Strategy—Overview. Canberra: Commonwealth of Australia, 2009. http://www.health.gov.au/internet/preventativehealth/publishing.nsf/Content/nphs-overview-toc (accessed June 23, 2011).

117 National Obesity Taskforce. Healthy Weight 2008 Australia's Future. Canberra: Department of Health and Ageing, 2003. http://www.health.gov.au/internet/healthyactive/publishing.nsf/content/healthy_weight08.pdf/$File/healthy_weight08.pdf (accessed June 23, 2011).

118 Cross-Government Obesity Unit, Department of Health and Department of Children, Schools and Families. Healthy weight, healthy lives: a cross-government strategy for England. London: HM Government, 2008. http://webarchive.nationalarchives.gov.uk/20100407220245/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_084024.pdf (accessed June 23, 2011).

119 United Kingdom Office of Communications. HFSS Advertising Restrictions - final review. London: Ofcom, 2010. http://stakeholders.ofcom.org.uk/binaries/research/tv-research/hfss-review-final.pdf (accessed June 23, 2011).

120 White House Task Force on Childhood Obesity. Solving the problem of childhood obesity within a generation. White House Task Force on Childhood Obesity, Report to the President. Washington DC, 2010. http://www.letsmove.gov/sites/letsmove.gov/files/TaskForce_on_Childhood_Obesity_May2010_FullReport.pdf (accessed June 23, 2011).

Swinburn, Boyd A. et al

"The Global Syndemic of Obesity, Undernutrition, and Climate Change. The Lancet Commission report", i

The Lancet. Vol 393, Issue 10173 2019 ss. 791-794, 800-802

Copyright (C) 2019 Elsevier Ltd

Kopiert med hjemmel i Kopinor-avtalen

THE LANCET







The Global Syndemic of Obesity, Undernutrition, and Climate Change: *The Lancet* Commission report



Boyd A Swinburn, Vivica I Kraak, Steven Allender, Vincent J Atkins, Phillip I Baker, Jessica R Bogard, Hannah Brinsden, Alejandro Calvillo, Olivier De Schutter, Raji Devarajan, Majid Ezzati, Sharon Friel, Shifalika Goenka, Ross A Hammond, Gerard Hastings, Corinna Hawkes, Mario Herrero, Peter S Hovmand, Mark Howden, Lindsay M Jaacks, Ariadne B Kapetanaki, Matt Kasman, Harriet V Kuhnlein, Shiriki K Kumanyika, Bagher Larijani, Tim Lobstein, Michael W Long, Victor K R Matsudo, Susanna D H Mills, Gareth Morgan, Alexandra Morshed, Patricia M Nece, An Pan, David W Patterson, Gary Sacks, Meera Shekar, Geoff L Simmons, Warren Smit, Ali Tootee, Stefanie Vandevijvere, Wilma E Waterlander, Luke Wolfenden, William H Dietz

Executive summary

Malnutrition in all its forms, including obesity, undernutrition, and other dietary risks, is the leading cause of poor health globally. In the near future, the health effects of climate change will considerably compound these health challenges. Climate change can be considered a pandemic because of its sweeping effects on the health of humans and the natural systems we depend on (ie, planetary health). These three pandemics—obesity, undernutrition, and climate change—represent The Global Syndemic that affects most people in every country and region worldwide. They constitute a syndemic, or synergy of epidemics, because they co-occur in time and place, interact with each other to produce complex sequelae, and share common underlying societal drivers. This Commission recommends comprehensive actions to address obesity within the context of The Global Syndemic, which represents the paramount health challenge for humans, the environment, and our planet in the 21st century.

The Global Syndemic

Although the Commission's mandate was to address obesity, a deliberative process led to reframing of the problem and expansion of the mandate to offer recommendations to collectively address the triple-burden challenges of The Global Syndemic. We reframed the problem of obesity as having four parts. First, the prevalence of obesity is increasing in every region of the world. No country has successfully reversed its epidemic because the systemic and institutional drivers of obesity remain largely unabated. Second, many evidence-based policy recommendations to halt and reverse obesity rates have been endorsed by Member States at successive World Health Assembly meetings over nearly three decades, but have not yet been translated into meaningful and measurable change. Such patchy progress is due to what the Commission calls policy inertia, a collective term for the combined effects of inadequate political leadership and governance to enact policies to respond to The Global Syndemic, strong opposition to those policies by powerful commercial interests, and a lack of demand for policy action by the public. Third, similar to the 2015 Paris Agreement on Climate Change, the enormous health and economic burdens caused by obesity are not seen as urgent enough to generate the public demand or political will to implement the recommendations of expert bodies for effective action. Finally, obesity has historically been considered in isolation from other major global challenges. Linking obesity with undernutrition and climate change into a single Global Syndemic framework focuses attention on the scale and urgency of addressing these combined challenges and emphasises the need for common solutions.

Syndemic drivers

The Commission applied a systems perspective to understand and address the underlying drivers of The Global Syndemic within the context of achieving the broad global outcomes of human health and wellbeing, ecological health and wellbeing, social equity, and economic prosperity. The major systems driving The Global Syndemic are food and agriculture, transportation, urban design, and land use. An analysis of the dynamics of these systems sheds light on the answers to some fundamental questions. Why do these

Lancet 2019; 393: 791–846

Published Online
January 27, 2019
http://dx.doi.org/10.1016/S0140-6736(18)32822-8

This online publication has been corrected. The corrected version first appeared at thelancet.com on February 21, 2019

See **Comment** pages 724, 726, and e30

School of Population Health, University of Auckland, Auckland, New Zealand (Prof B A Swinburn MD, S Vandevijvere PhD); **Department of Human Nutrition, Foods, and Exercise, Virginia Tech, Blacksburg, VA, USA** (V I Kraak PhD); **Global Obesity Centre, School of Health & Social Development, Deakin University, Geelong, VIC, Australia** (Prof B A Swinburn; Prof S Allender PhD); **Caribbean Community Secretariat, Bridgetown, Barbados** (V J Atkins); **Institute for Physical Activity and Nutrition** (P I Baker PhD) **and Global Obesity Centre, School of Health and Social Development** (G Sacks PhD), **Deakin University, Melbourne, VIC, Australia; Commonwealth Scientific and Industrial Research Organisation, Brisbane, QLD, Australia** (J R Bogard PhD, M Herrero PhD); **World Obesity Federation,**

Key messages

The pandemics of obesity, undernutrition, and climate change represent three of the gravest threats to human health and survival. These pandemics constitute The Global Syndemic, consistent with their clustering in time and place, interactions at biological, psychological, or social levels, and common, large-scale societal drivers and determinants. Their interactions and the forces that sustain them emphasise the potential for major beneficial effects on planetary health that double-duty or triple-duty actions, which simultaneously act on two or all three of these pandemics, will have. To mitigate The Global Syndemic, the Commission proposed the following nine broad recommendations, under which sit more than 20 actions:

- Think in Global Syndemic terms to create a focus on common systemic drivers that need common actions.
- Join up the silos of thinking and action to create platforms to work collaboratively on common systemic drivers and double-duty or triple-duty actions.
- Strengthen national and international governance levers to fully implement policy actions which have been agreed upon through international guidelines, resolutions and treaties.
- Strengthen municipal governance levers to mobilise action at the local level and create pressure for national action.
- Strengthen civil society engagement to encourage systemic change and pressure for policy action at all levels of government to address The Global Syndemic.
- Reduce the influence of large commercial interests in the public policy development process to enable governments to implement policies in the public interest to benefit the health of current and future generations, the environment, and the planet.
- Strengthen accountability systems for policy actions to address The Global Syndemic.
- Create sustainable and health-promoting business models for the 21st century to shift business outcomes from a short-term profit-only focus to sustainable, profitable models that explicitly include benefits to society and the environment.
- Focus research on The Global Syndemic determinants and actions to create an evidence base of systemic drivers and actions, including indigenous and traditional approaches to health and wellbeing.

London, UK (H Brinsden, T Lobstein PhD); **El Poder del Consumidor, Mexico City, Mexico** (A Calvillo); **Institute for Interdisciplinary Research in Legal Sciences, Catholic University of Louvain, Louvain-la-Neuve, Belgium** (Prof O De Schutter PhD); **Public Health Foundation of India, Centre for Chronic Disease Control, New Delhi, India** (R Devarajan, S Goenka PhD); **Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK** (Prof M Ezzati FMedSci); **School of Regulation and Global Governance** (Prof S Friel PhD) **and Climate Change Institute** (Prof M Howden PhD), **Australian National University, Canberra, ACT, Australia; Center on Social Dynamics & Policy, The Brookings Institution, Washington, DC, USA** (R A Hammond PhD, M Kasman PhD); **Public Health & Social Policy Department**

systems operate the way they do? Why do they need to change? Why are they so hard to change? What leverage points (or levers) are required to overcome policy inertia and address The Global Syndemic? The Commission identified five sets of feedback loops as the dominant dynamics underlying the answers to these questions. They include: (1) governance feedback loops that determine how political power translates into the policies and economic incentives and disincentives for companies to operate within; (2) business feedback loops that determine the dynamics for creating profitable goods and services, including the externalities associated with damage to human health, the environment, and the planet; (3) supply and demand feedback loops showing the relationships that determine current consumption practices; (4) ecological feedback loops that show the unsustainable environmental damage that the food and transportation systems impose on natural ecosystems; and (5) human health feedback loops that show the positive and negative effects that these systems have on human health. These interactions need to be elucidated and methods for reorienting these feedback systems prioritised to mitigate The Global Syndemic.

Double-duty or triple-duty actions

The common drivers of obesity, undernutrition, and climate change indicate that many systems-level interventions could serve as double-duty or triple-duty actions to change the trajectory of all three pandemics simultaneously. Although these actions could produce win-win, or even win-win-win, results, they are difficult to achieve. A seemingly simple example shows how challenging these actions can be. National dietary guidelines serve as a basis for the development of food and nutrition policies and public education to reduce obesity and undernutrition and could be extended to include sustainability by moving populations towards consuming largely plant-based diets. However, many countries' efforts to include environmental sustainability principles within their dietary guidelines failed due to pressure from strong food industry lobbies, especially the beef, dairy, sugar, and ultra-processed food and beverage industry sectors. Only a few countries (ie, Sweden, Germany, Qatar, and Brazil) have developed dietary guidelines that promote environmentally sustainable diets and eating patterns that ensure food security, improve diet quality, human health and wellbeing, social equity, and respond to climate change challenges.

The engagement of people, communities, and diverse groups is crucial for achieving these changes. Personal behaviours are heavily influenced by environments that are obesogenic, food insecure, and promote greenhouse-gas emissions. However, people can act as agents of change in their roles as elected officials, employers, parents, customers, and citizens and influence the societal norms and institutional policies of worksites, schools, food retailers, and communities to address The Global Syndemic. Across systems and institutions, people are decision makers who can vote for, advocate for, and communicate their preferences with other decision-makers about the policies and actions needed to address The Global Syndemic. Within the natural ecosystems, people travel, recreate, build, and work in ways that can preserve or restore the environment. Collective actions can generate the momentum for change. The Commission believes that the collective influence of individuals, civil society organisations, and the public can stimulate the reorientation of human systems to promote health, equity, economic prosperity, and sustainability.

Changing trends in obesity, undernutrition, and climate change

Historically, the most widespread form of malnutrition has been undernutrition, including wasting, stunting, and micronutrient deficiencies. The Global Hunger Index (1992–2017) showed substantial declines in under-5 child mortality in all regions of the world but less substantial declines in the prevalence of wasting and stunting among children. However, the rates of decline in undernutrition for children and adults are still too slow to meet the Sustainable Development Goal (SDG) targets by 2030.

In the past 40 years, the obesity pandemic has shifted the patterns of malnutrition. Starting in the early 1980s, rapid increases in the prevalence of overweight and obesity began in high-income countries. In 2015, obesity was estimated to affect 2 billion people worldwide. Obesity and its determinants are risk factors for three of the four leading causes of non-communicable diseases (NCDs) worldwide, including cardiovascular diseases, type 2 diabetes, and certain cancers.

Extensive research on the developmental origins of health and disease has shown that fetal and infant undernutrition are risk factors for obesity and its adverse consequences throughout the life course. Low-income and middle-income countries (LMICs) carry the greatest burdens of malnutrition. In LMICs, the prevalence of overweight in children less than 5 years of age is rising on the background of an already high prevalence of stunting (28%), wasting (8·8%), and underweight (17·4%). The prevalence of obesity among stunted children is 3% and is higher among children in middle-income countries than in lower-income countries.

The work of the Intergovernmental Panel on Climate Change (IPCC), three previous *Lancet* Commissions related to climate change and planetary health (2009–15), and the current *Lancet* Countdown, which is tracking progress on health and climate change from 2017 to 2030, have provided extensive and compelling projections on the major human health effects related to climate change. Chief among them are increasing food insecurity and undernutrition among vulnerable populations in many LMICs due to crop failures, reduced food production, extreme weather events that produce droughts and flooding, increased food-borne and other infectious diseases, and civil unrest. Severe food insecurity and hunger are associated with lower obesity prevalence, but mild to moderate food insecurity is paradoxically associated with higher obesity prevalence among vulnerable populations.

Wealthy countries already have higher burdens of obesity and larger carbon footprints compared with LMICs. Countries transitioning from lower to higher incomes experience rapid urbanisation and shifts towards motorised transportation with consequent lower physical activity, higher prevalence of obesity, and higher greenhouse-gas emissions. Changes in the dietary patterns of populations include increasing consumption of ultra-processed food and beverage products and beef and dairy products, whose production is associated with high greenhouse-gas emissions. Agricultural production is a leading source of greenhouse-gas emissions.

The economic burden of The Global Syndemic

The economic burden of The Global Syndemic is substantial and will have the greatest effect on the poorest of the 8·5 billion people who will inhabit the earth by 2030. The current costs of obesity are estimated at about $2 trillion annually from direct health-care costs and lost economic productivity. These costs represent 2·8% of the world's gross domestic product (GDP) and are roughly the equivalent of the costs of smoking or armed violence and war.

Economic losses attributable to undernutrition are equivalent to 11% of the GDP in Africa and Asia, or approximately $3·5 trillion annually. The World Bank estimates that an investment of $70 billion over 10 years is needed to achieve SDG targets related to undernutrition, and that achieving them would create an estimated $850 billion in economic return. The economic effects of climate change include, among others, the costs of environmental disasters (eg, drought and wildfires), changes in habitat (eg, biosecurity and sea-level rises), health effects (eg, hunger and diarrhoeal infections), industry stress in sectors such as agriculture and fisheries, and the costs of reducing greenhouse-gas emissions. Continued inaction towards the global mitigation of climate change is predicted to cost 5–10% of global GDP, whereas just 1% of the world's GDP could arrest the increase in climate change.

Actions to address The Global Syndemic

Many authoritative policy documents have proposed specific, evidence-informed policies to address each of the components of The Global Syndemic. Therefore, the Commission decided to focus on the common, enabling actions that would support the implementation of these policies across The Global Syndemic. A set of principles guided the Commission's recommendations to enable the implementation of existing recommended policies: be systemic in nature, address the underlying causes of The Global Syndemic and its policy inertia, forge synergies to promote health and equity, and create benefits through double-duty or triple-duty actions.

The Commission identified multiple levers to strengthen governance at the global, regional, national, and local levels. The Commission proposed the use of international human rights law and to apply the concept of a right to wellbeing, which encompasses the rights of children and the rights of all people to health, adequate food, culture, and healthy environments. Global intergovernmental organisations, such as the World Trade Organization, the World Economic Forum, the World Bank, and large philanthropic foundations and regional platforms, such as the European Union, Association of Southeastern Nations, and the Pacific Forum, should play much stronger roles to support national policies that address The Global Syndemic. Many states and municipalities are leading efforts to reduce greenhouse-gas emissions by incentivising less motorised travel and improving urban food systems. Civil society organisations can create a greater demand for national policy actions with increases in capacity and funding. Therefore, in addition to the World Bank's call for $70 billion for undernutrition and the Green Climate Fund of $100 billion for LMICs to address climate change, the Commission calls for $1 billion to support the

(R A Hammond), **Social System Design Lab** (Prof P S Hovmand PhD), **and Prevention Research Center** (A Morshed), **Brown School, Washington University in St Louis, St Louis, MO, USA; Institute for Social Marketing, University of Stirling, Stirling, UK** (Prof G Hastings PhD); **Centre for Food Policy, City University, University of London, London, UK** (Prof C Hawkes PhD); **Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA** (L M Jaacks PhD); **Department of Marketing and Enterprise, Hertfordshire Business School, University of Hertfordshire, Hatfield, UK** (A B Kapetanaki PhD); **Centre for Indigenous Peoples' Nutrition and Environment, McGill University, Montreal, QC, Canada** (Prof H V Kuhnlein PhD); **Dornsife School of Public Health, Drexel University, Philadelphia, PA, USA** (Prof S K Kumanyika PhD); **Endocrinology and Metabolism Research Center** (Prof B Larijani MD), **Diabetes Research Center** (A Tootee MD), **Tehran University of Medical Sciences, Tehran, Iran; Milken Institute School of Public Health, The George Washington University, Washington, DC, USA** (M W Long PhD, Prof W H Dietz MD); **Physical Fitness Research Laboratory of São Caetano do Sul, São Caetano do Sul, São Paulo, Brazil** (V K R Matsudo MD); **Institute of Health & Society, Newcastle University, Newcastle upon Tyne, UK** (S D H Mills PhD); **The Morgan Foundation, Wellington, New Zealand** (G Morgan PhD, G L Simmons); **Obesity Action Coalition, Tampa, FL, USA** (P M Nece JD); **School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China** (Prof A Pan PhD); **International Development Law Organization, The Hague, Netherlands** (D W Patterson); **Health, Nutrition, and Population Global Practice, The World Bank, Washington, DC, USA** (M Shekar PhD); **African Centre for Cities, University of Cape Town, Cape Town, South Africa** (W Smit PhD); **Scientific**

***Panel 1:* The *Lancet* Commission on Obesity**

The *Lancet* Commission on Obesity was formed following the publication of two *Lancet* Series on Obesity in 2011 and 2015. The Commission was under the auspices of *The Lancet*, the University of Auckland, George Washington University, and the World Obesity Federation. The Commission was comprised of 26 Commissioners and 17 Fellows from 14 countries. The disciplines and expertise of the Commissioners included global obesity, population health, nutrition (including undernutrition), food systems (including indigenous food systems), physical activity, political science and policy making, climate change, urban planning, epidemiology, consumer advocacy, human rights, international law, trade, health equity, social determinants, economics, marketing, agriculture, systems science, community interventions, implementation science, medicine, business, financing, and the experience of living with obesity.

The aims of the Commission were to:

- Identify the systemic commonalities in drivers and solutions across obesity, undernutrition, and climate change.
- Describe double-duty or triple-duty policies and actions to address The Global Syndemic, and ways to strengthen accountability systems for their implementation.

The Commission's work on The Global Syndemic came from two group model building sessions organised for the Commissioners, a review of existing conceptual and computational models, and three face-to-face meetings between February, 2016, and July, 2017. Additionally, consultation workshops were held around the world during 2017, to obtain feedback on the Commission's concepts. These workshops were hosted by the Australian National University, Canberra; Washington University, St Louis; The World Bank, Washington DC; Centre for Food Policy, City, University of London, UK; International Atomic Energy Agency, WHO, and UNICEF, Vienna, Austria; Endocrinology and Metabolism Research Institute of Tehran, University of Medical Sciences, Tehran, Iran; a satellite meeting at the International Congress on Obesity, Buenos Aires, Argentina; Huazhong University of Science and Technology, Wuhan, China; and the Center for Chronic Disease Control, Delhi, India.

Institute of Public Health (Sciensano), Brussels, Belgium (S Vandevijvere); **Department of Public Health Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands** (W E Waterlander PhD); **and School of Medicine and Public Health, The University of Newcastle, Newcastle, NSW, Australia** (L Wolfenden PhD)

Correspondence to: Prof Boyd Swinburn, Private Bag 92019, Auckland 1142, New Zealand
boyd.swinburn@auckland.ac.nz

For the ***Lancet* Commission on Obesity** see https://www.worldobesity.org/what-we-do/projects/lancet-commission-on-obesity

For the **EAT Forum** see https://eatforum.org/

efforts of civil society organisations to advocate for policy initiatives that mitigate The Global Syndemic.

A principal source of policy inertia related to addressing obesity and climate change is the power of vested interests by commercial actors whose engagement in policy often constitutes a conflict of interest that is at odds with the public good and planetary health. Countering this power to assure unbiased decision making requires strong processes to manage conflicts of interest. On the business side, new sustainable models are needed to shift outcomes from a profit-only model to a socially and environmentally viable profit model that incorporates the health of people and the environment. The fossil fuel and food industries that are responsible for driving The Global Syndemic receive more than $5 trillion in annual subsidies from governments. The Commission recommends that governments redirect these subsidies into more sustainable energy, agricultural, and food system practices. A Framework Convention on Food Systems would provide the global legal structure and direction for countries to act on improving their food systems so that they become engines for better health, environmental sustainability, greater equity, and ongoing prosperity.

Stronger accountability systems are needed to ensure that governments and private-sector actors respond adequately to The Global Syndemic. Upstream monitoring is needed to measure implementation of policies, examine the commercial, political, economic and sociocultural determinants of obesity, evaluate the impact of policies and actions, and establish mechanisms to hold governments and powerful private-sector actors to account for their actions.

Similarly, platforms for stakeholders to interact and secure funding, such as that provided by the EAT Forum for global food system transformation, are needed to allow collaborations of scientists, policy makers, and practitioners to co-create policy-relevant empirical, and modelling studies of The Global Syndemic and the effects of double-duty and triple-duty actions. Bringing indigenous and traditional knowledge to this effort will also be important because this knowledge is often based on principles of environmental stewardship, collective responsibilities, and the interconnectedness of people with their environments.

The challenges facing action on obesity, undernutrition, and climate change are closely aligned with each other. Bringing them together under the umbrella concept of The Global Syndemic creates the potential to strengthen the action and accountabilities for all three challenges. Our health, the health of our children and future generations, and the health of the planet will depend on the implementation of comprehensive and systems-oriented responses to The Global Syndemic.

Introduction

Obesity has risen inexorably worldwide in the past 4–5 decades and is now one of the largest contributors to poor health in most countries.[1] Despite nearly two decades of recommendations from authoritative national and international organisations, especially WHO, the implementation of effective obesity-prevention policies has been slow and inconsistent.[2] The Commission recognises that this patchy progress is intrinsic to the complexity of the obesity problem itself, and uses the collective term policy inertia to describe the combined effects of inadequate political leadership and governance to enact policies to respond to The Global Syndemic, strong opposition to those policies by powerful commercial interests, and a lack of demand for policy action by the public.[3] Although some high-income countries have experienced a plateau or slight decline in childhood obesity, no country has decreased the obesity epidemic across its population.

The *Lancet* Commission on Obesity (panel 1) developed a broader approach to obesity, on the basis of the concept that the obesity pandemic is one element of The Global Syndemic, which also includes undernutrition and climate change.

As originally defined, a syndemic is two or more diseases with three characteristics: they co-occur in time and place, they interact with each other at biological, psychological, or societal levels, and they share common underlying societal drivers.[4] Although the syndemic

achieve certain outcomes, such as economic prosperity. Due to the way that these systems have been designed, the inevitable overconsumption and inequitable distribution of resources has caused negative externalities and poor outcomes for the other three outcomes of social equity, human health and wellbeing, and ecological health and wellbeing. These global outcomes will be considered in more depth later in this report in relation to the different country contexts and their priorities for action.

Re-thinking obesity: The Global Syndemic and complex adaptive systems

The Global Syndemic

The original concept of a syndemic was largely applied to diseases at the individual level—two or more diseases clustering in time and place, interacting with each other and having common, societal determinants.[4,29] A subsequent extension of the concept used syndemics to describe health problems that synergistically affect population health in the context of economic and social inequalities.[30] To date, the main applications of the extended syndemic concept have been in relation to HIV/AIDS and its associations with substance abuse and violence,[4,31] the clustering of hepatitis C, alcohol abuse, and hepatocellular cancer,[31,32] and poverty, depression, and diabetes among low-income populations.[30]

The Commission proposes that the definition of syndemics should be further extended to the pandemics of obesity, undernutrition, and climate change. We consider climate change a pandemic because of its dynamic nature, its rapid rise, and its predicted catastrophic impact on human health. The interactions between these pandemics occur at both the individual and population levels (figure 2B). The Commission calls these three pandemics The Global Syndemic to emphasise the major global importance of this cluster of pandemics, which are now, and will be into the foreseeable future, the dominant causes of human and environmental (ie, planetary) ill-health. Recognition that these synergistic pandemics constitute a syndemic provides a more comprehensive view of their interactions, and promises common systemic actions that can unite previously disparate stakeholders.

Obesity, undernutrition, and climate change cluster in time and place

The prevalence of obesity has risen globally in the past four decades including an 8 times increase in girls to 5·6% and a 10 times increase in boys to 7·8% in 2016.[33] The rise in obesity prevalence in adults in the same period has also been relentless, increasing to 14·9% in women and 10·8% in men, in the same time period.[1] In 2015, excess bodyweight was estimated to affect 2 billion people worldwide, and accounted for approximately 4 million deaths and 120 million disability-adjusted life-years.[34] The estimated costs of obesity are about US$2 trillion annually,[35,36] representing 2·8% of the world's GDP.[36] The increase in the prevalence of obesity accounts for the rapid increase in diabetes, which now affects almost 9% of the world's population.[1]

The Global Burden of Disease data suggest that, by 2025, nearly 268 million children and adolescents in 200 countries will be overweight, 124 million will have obesity, and almost three-quarters (72·3%) of NCD-related illness and deaths will occur in LMICs.[34]

The prevalence of undernutrition has been declining for decades, although it is still highly prevalent in many LMICs. The Global Hunger Index (1992–2017) showed substantial declines in under-5 child mortality in all regions of the world but less substantial declines in the prevalence of wasting and stunting among children.[24] In 2008, stunting, severe wasting, and intrauterine growth retardation were estimated to account for 2·2 million deaths and 21% of disability-adjusted life-years in children under 5 years of age.[37] In 2018, the Global Nutrition Report found that 155 million children were stunted and 52 million children were wasted.[28] 2 billion people have a micronutrient deficiency, and 815 million people are chronically undernourished. Undernutrition disproportionately affects children and adults in low-income countries, particularly those in eastern and middle Africa and south-central Asia.[37] Because the prevalence of undernutrition has been declining, the prevalence of child and adolescent obesity might exceed moderate and severe undernutrition by 2022.[1] Estimates of the costs to the global economy from undernutrition, micronutrient deficiencies, and overweight are up to $3·5 trillion annually.[23]

Although malnutrition in all its forms is by far the largest cause of health loss in the world, it will be compounded by the health effects of climate change in the near future.[38] The health gains achieved in the past 50 years of global economic development could be reversed by 2050 due to the consequences of climate change.[39] Estimates of the future costs of climate change are 5–10% of the world's GDP, with costs in low-income countries in excess of 10% of their GDP.[40]

LMICs that produce the fewest greenhouse-gas emissions are more affected by climate change than those countries that produce the highest greenhouse-gas emissions. Furthermore, climate change will have a disproportionate effect on agricultural production and consequently human health in LMICs. The resultant population displacement might already account for increased global migration patterns in Africa and other regions.[39]

Obesity, undernutrition, and climate change interact with each other

Many interactions occur among the components of The Global Syndemic. The World Economic Forum's annual risk reports include the global risks of climate change, NCDs, food crises, failures of governance, and failures of urban planning.[41] The report's interconnections map shows the interdependency of these risks. The UN's

IPCC predicts that the biggest threat to health from substantial climate change will be undernutrition.[42,43] Phalkey and colleagues[44] provide an example of a causal loop diagram showing the systemic interconnections between climate change and food security (appendix p 3).

Climate change affects food systems in many ways. Small-scale, low-income farmers will likely be most affected, although environmental change will affect all producers exposed to storms, floods, drought, coastal erosion, warming oceans, and rising sea levels. A small degree of global warming might benefit some crops in some areas, taking into account the carbon fertilisation effects. However, global warming will lead to lower yields especially in tropical regions. Furthermore, climate change might reduce the protein and micronutrient content of plant foods.[43] Although not all crops are equally affected, elevated levels of carbon dioxide have been shown to decrease protein concentrations of wheat, barley, rice, and potato crops by 10–15% and soy by 1·4%.[45] These changes in nutrient value will further contribute to undernutrition, particularly in children. All of these changes will also increase the prices of basic food commodities,[46–48] and are expected to lead to increases in nutrient deficiencies and chronic undernutrition among the most food-insecure population groups.[49,50] Food insecurity could be exacerbated further by climate mitigation efforts associated with land sparing.[50]

Food production is one of the largest contributors to climate change. Agriculture directly contributes about 15–23% of all greenhouse-gas emissions, which is comparable to transportation. But when land conversion and the wider downstream food system processes, including food waste, are taken into account, the total contribution of food to emissions can be as high as 29%.[51,52] Livestock alone account for 12–19% of greenhouse-gas emissions.[53] The types of food produced have differing effects. Both meat and dairy products require more resources and generate larger emissions of methane than plant-based alternatives.[54,55] Additionally, non-seasonal fruits and vegetables produce substantial emissions when grown in greenhouses, preserved in a frozen state, or transported by air.[54] The energy required for the production, harvesting, transportation, and packaging of wasted foods also generates more than 3·3 billion tonnes of carbon dioxide annually, making food wastage the third top emitter after the USA and China.[56]

The globalisation of food systems has depended on cheap energy from fossil fuels for intensive large-scale agricultural production and long-haul transportation that has transformed the diets of the world's population. Increased urbanisation has exposed people in these environments to markets for mass-produced, processed food and beverage products.

The interactions between climate change and obesity are also numerous but less certain. Increasing ambient temperatures could contribute to obesity through reductions in physical activity.[57] Additionally, the effect of climate change on fruit and vegetable production will make these products more expensive, and might prompt shifts in the eating patterns of populations towards processed food and beverage products that are high in, fats, sugars, and sodium.[57] Increased food and agricultural production to meet the needs of a growing population with a high prevalence of obesity will increase the food system's greenhouse-gas emissions. Another mechanism by which obesity could contribute to climate change is through the increased costs of fossil fuels related to transporting populations with a high prevalence of obesity.[57] This added contribution to greenhouse gases attributed to obesity is very small relative to other greenhouse-gas emissions[54,58] and emphasising this pathway risks placing further blame on people with obesity—not only for their own condition but also for climate change. Because periodic, but not severe, food insecurity is associated with increased risks of obesity in high income countries,[59] increased food insecurity could theoretically increase the prevalence of obesity.

Obesity and undernutrition also interact. Undernutrition in early life is a predictor for later obesity. Biological and social mechanisms that explain this relationship include the contribution of fetal and infant undernutrition, food insecurity, and poor diet quality characterised by a low variety of healthy foods.[60] Many middle-income countries, especially in the Middle East and north Africa, are facing a double burden of undernutrition and overweight or obesity. For example, Iran, Morocco, Oman, Saudi Arabia, Syria, and Tunisia are all in the top tertile of countries for both the prevalence of adult female obesity (>27% with body-mass index [BMI] >30 kg/m²)[1] and female child and adolescent underweight (>18% with a BMI Z score <1 standard deviation; appendix p 4).[61] Indeed, the co-occurrence of stunting (low height-for-age) and obesity (high BMI for age) is not uncommon within the same country, village, family, and even individual.[23] LMICs carry the greatest triple-burdens of malnutrition. In LMICs, the prevalence of overweight in children less than 5 years of age is rising on the background of an already high prevalence of stunting (28%), wasting (8·8%) and underweight (17·4%).[36] The prevalence of obesity among stunted children is 3% and can be more than 10% in some middle-income countries.[62] Countries cannot afford to prioritise their nutrition policies to focus only on reducing undernutrition while costly obesity-related NCDs, such as type 2 diabetes, are overwhelming their national health systems.

Obesity, undernutrition, and climate change have common systems drivers

The shared societal determinants for obesity and climate change have previously been noted in the published literature.[63,64] Both are driven by the high consumption of cheap energy sources (foods and fossil fuels) and car-oriented transportation systems. The consumptogenic economic systems that promote excessive and

unsustainable consumption patterns value GDP growth and overlook its role in damaging the health of people, the environment, and the planet.[65] In economic systems in which the vested interests of powerful transnational corporations produce financial benefits that are maximally privatised, the social and environmental costs or externalities fall to consumers, taxpayers, ratepayers, and future generations. The major risks to society and economic development in the future are heavily neglected.[57]

The balance of power between actors within the governance mechanisms determines how the levers of power are used. Those levers include policies (eg, laws and regulations), economic incentives and disincentives, and societal norms and expectations. Although governance occurs at all levels, macro-level governance (usually national governments) creates the operating conditions for the major systems that drive The Global Syndemic—the food, transportation, land use, and urban design sectors. These macro systems, in turn, flow through the meso systems or settings in which people interact, such as schools, workplaces, retail outlets, and community spaces. They, in turn, influence the micro systems or social networks, such as families and social groups, affecting their behavioural patterns.

The power balance within the governance structures determines, for example, whether agricultural subsidies support monoculture crops and beef and dairy farming over more sustainable agricultural systems with fewer environmental effects and greater health benefits, whether funding for transportation infrastructure prioritises roads over mass transit and active commuting, whether periurban horticultural lands are zoned for housing and industry, whether economic policies promote consumption-driven growth, whether regulatory systems allow or constrain marketing of unhealthy food and beverage products to children or breastmilk substitutes to mothers, and whether cultural and religious codes of dress and behaviour, especially for girls and women, constrain their ability to be physically active and fully engaged with society. The sociocultural nature of some of the systems is important to note. A population's values, beliefs, attitudes, religious expectations, and social practices shape the types of foods people eat, how they use food for hospitality, the status attributed to particular foods, and their vulnerability and exposure to targeted commercial marketing that exploits these attitudes and values.

Key aspects of the political economy have been recognised as the deep drivers that shape the very nature of the systems creating The Global Syndemic. For example, economic power has become increasingly concentrated into fewer and fewer transnational corporations, and this is certainly true in the food sector.[66–68] According to the former Director General of WHO, this "market power readily translates into political power".[69] Specifically, the transnational corporations lobby for fewer regulations that apply to them (eg, no regulations on marketing unhealthy food to children or warning labels on processed foods), promote regulations that apply to other sectors (eg, trade and investment agreements that bind governments to protect corporate investment interests), resist or reject taxes that apply to their products (eg, taxes on sugary drinks and energy-dense, nutrient-poor foods), and lobby policy makers for subsidies that benefit their businesses (eg, agricultural and transportation subsidies).[68] The fossil fuel and food industries that are responsible for driving The Global Syndemic receive more than $5 trillion in annual subsidies from governments.[70]

Complex adaptive systems

Close examination of the contributors to The Global Syndemic reveals the role of complex adaptive systems operating at each of its levels (figure 2B). Systems, such as health systems, schools, or families are complex because the inter-relationships are multiple, change over time, and involve several interacting, reinforcing, and balancing causal feedback loops, as well as the fact that non-linear associations exist between causes and effects. Reinforcement of feedback loops leads to virtuous or vicious cycles, depending on the outcome, although balancing feedback loops counteract the directions of change that form the basis of homoeostasis and policy resistance in complex adaptive systems. For example, a new food launched into the market might accelerate profits for its manufacturer, which in turn supports more marketing and wider distribution that make more profits in a reinforcing feedback loop. However, the product's growth in sales and profits are not infinite because market saturation and competition act as balancing feedback loops that counteract the reinforcing feedback loop driving the initial growth.[71] Understanding the dynamics of the major feedback loops within a system is, therefore, crucial to identifying how to reorient the systems towards better outcomes.

The systems are also adaptive. Any change in one part of the system will lead to changes in the implicit and explicit rules of the actors in other parts of the system, generating new, emergent dynamics. Changing the reimbursement structure within a health system, changing governance structures in a school, or changing the structure of a family through marriage or separation, for example, can lead to the adaption of actors by changing the rules in how they respond to each other and their environment, and push a system toward a new equilibrium or system instability.

Taking a system dynamics approach to The Global Syndemic provides new insights into three critical questions: why are systems, including food systems, the way they are? Why do they need to change? Why are they so difficult to change?[72]

Five sets of feedback loops

The Commission considered that five crucial feedback loops (figure 2C) need to be assessed within the food,

Conrad, Peter & Waggoner, Miranda

"Medicalization", i (Red) Cockerham, William C.; Dingwall, Robert & Quah, Stella

The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society 2014 ss. 1-4

Copyright (C) 2014 Wiley Blackwell

Kopiert med hjemmel i Kopinor-avtalen

Medicalization

PETER CONRAD
Brandeis University, USA
MIRANDA WAGGONER
Princeton University, USA

Sociologists studying medicalization often cite Zola's (1972) paper as the first touchstone, followed by the writings of Conrad (1975; 1976) and Conrad and Schneider (1980), although the early medicalization work built on insights of Parsons (1951), Freidson (1970), and others. In the ensuing four decades, literally thousands of papers have been written on medicalization (e.g., see Conrad 2007; Clarke et al. 2010) by sociologists, anthropologists, historians, feminists, and medical scholars. Medicalization has become a common term in the social science and medical literature and to a degree in the public media.

While there may be numerous definitions of medicalization, in its simplest form it means "to make medical." A fuller depiction sees medicalization as the process by which previously non-medical problems become defined and treated as medical problems, usually as diseases or disorders. The emphasis in this perspective is on "process" and "definition."

There are several characteristics of medicalization. (1) The definitional issue is central to medicalization; that is, how a problem is defined is key to what is done about it. (2) There are degrees of medicalization: for example, some problems are virtually completely medicalized (e.g., schizophrenia, epilepsy), while others are only somewhat medicalized (e.g., Internet addiction, sexual addiction), with still others are contested or somewhere in between (e.g., obesity, opiate addiction). Thus medicalization is more of a continuum than a binary either/or distinction. (3) Medicalized categories can expand or contract. For example, attention deficit hyperactivity disorder (ADHD) was for many years seen as a disorder of children but is now seen as a lifespan disorder, including adult ADHD. Post-traumatic stress disorder (PTSD) emerged in the 1970s as a diagnostic category depicting Vietnam vets and has expanded to include survivors of violence, sexual abuse, or natural disasters; now it can even be applied to witnesses to violence or disaster. On the other hand, hysteria, which was a common medical category in the late nineteenth century, has almost disappeared as a diagnosis. (4) Physician involvement in medicalization is variable. With the medicalization of alcoholism, physicians were minimally involved; rather, the emergence of Alcoholics Anonymous was central to this process. With many other examples of medicalization, direct medical involvement is key. (5) Medicalization is bidirectional; that is, there can be medicalization as well as demedicalization. The most common example of demedicalization has been homosexuality, which due to the 1974 vote of the American Psychiatric Association and subsequent changes became "officially" demedicalized. There is no question, however, that there has been much more medicalization than demedicalization in the past four decades. One recent study (Conrad et al. 2010) has estimated that medicalized disorders cost about $77 billion or about 3.9 percent of US health care expenditures, more than is spent on all of US public health.

A few caveats about medicalization studies are necessary. (1) Some researchers focus on

The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society, First Edition.
Edited by William C. Cockerham, Robert Dingwall, and Stella R. Quah.
© 2014 John Wiley & Sons, Ltd. Published 2014 by John Wiley & Sons, Ltd.

whether medicalized problems are "really" diseases or whether they are socially invented medical categories. It is important for social scientists to remain agnostic on what is or is not a "real" disease. A better question to ask is: how did this problem become defined as a medical disorder and with what consequences? (2) Much of the popular and some of the academic writing on medicalization has been concerned with the "overmedicalization" of society. While considerable research on medicalization does suggest potential overmedicalization is an issue to be considered, medicalization studies focus on the "process" of medicalization and are not necessarily driven by concerns of overmedicalization. Sociologists study medicalization just like they study other social processes, such as industrialization, secularization, or urbanization, focusing on the origins and consequences of the phenomena. It is probably likely in some cases there has been an overmedicalization of some medical problems, while in others perhaps undermedicalization of a certain problem may be the case. (3) Occasionally critics will use the term "medicalization" as an epithet or pejorative, with the assumption that medicalization is by its nature a "bad" thing. But as we have emphasized, medicalization is a not a normative process; in itself it is neither bad nor good. Historically, for instance, there are many examples of beneficial medicalization such as epilepsy and, in important ways, childbirth. It is hard to imagine that most people would prefer to return to the era when epilepsy was seen as a "curse," a sign of being bewitched and a justification for institutionalization; nor would we want to revisit a childbirth era in which maternal mortality was common and infant mortality was a regular birth outcome. That is not to say that feminists and others have not importantly analyzed some of the consequences of what might be called the unreflective and extreme medicalization of modern childbirth.

Feminists and other scholars indeed have attended to gendered aspects of the medicalization process. Riska (2003; 2010) explains that the medicalization thesis can be divided into two gendered phases: in the 1970s, it was used in a gender-neutral manner; then, through the 1980s, feminists focused on the ways in which women's bodies were distinctly subjected to medicalization. Childbirth is a clear example of the medicalization of women's reproductive bodies. Over time, it moved from the social realm of the home and the midwife to the medical realm of the hospital and the obstetrician, such that today birth is primarily considered a medical event. The recent rise in cesarean sections reveals how medicalization may be both useful and problematic. While C-sections are a life-saving procedure in some cases of childbirth, they often now are performed without medical indication. As a result, C-section procedures in the United States rose 53 percent between 1996 and 2007, and now about one-third of all births are delivered by C-section in the US (Menacker and Hamilton 2010). The medicalization of women's reproductive lives has extended in recent years beyond pregnancy and childbirth with the rise of "pre-conception care," a novel medical belief which posits that adverse birth outcomes can be mitigated through medical surveillance and treatment prior to pregnancy (Waggoner 2011). This extends medical surveillance of the lives of virtually all women of childbearing age.

Whereas women's bodies and women's life experiences historically have been more medicalized than men's, men's bodies (especially aging men's bodies) have become subject to medicalization in the twenty-first century (Rosenfield and Faircloth 2006). The rise of Viagra and testosterone therapy, for instance, signifies the medical and pharmaceutical attention paid to medicalizing normal male life events. While problems with maintaining an

erection are certainly ages old, it was only with the rise of Viagra and the redefinition of the problem as erectile dysfunction that it became fundamentally a medical problem (Loe 2004).

Medicalization of all sorts of life problems is now common in contemporary culture. Some have argued (Clarke et al. 2003; 2010) that medicalization at the millennial turn is better described as "biomedicalization," a definition that highlights how biomedicine has transformed into a technoscientific enterprise. This is an inclusive and broad-based conception that includes many changes in health and health care beyond the impact of what we are calling medicalization. We can look at similar societal changes by contending that the engines of medicalization have shifted. In this framing, we see the process of medicalization is driven in contemporary society by commercial and pharmaceutical interests, genetic treatments, consumers, and managed care (Conrad 2005). For example, in the past few decades there has been an increase in what might be called medical markets for medicalized services, including a range of cosmetic surgeries, patient demands for new forms of medications, and treatment for problems that a few decades ago were not considered medical (e.g., infertility). In this sense, consumers have also become an important engine increasing medicalization.

While it is perhaps possible to evaluate each case of medicalization on its own merits, there are also several consequences that inhere to the wider phenomenon of medicalization irrespective of differences in the various examples or cases of medicalization (see Conrad 2007). These include (1) the pathologization of everything, or turning all human differences into medical diagnoses; (2) the rise of medical definitions of normality, that is, having pharmaceutical companies or physicians defining what are normal (and abnormal) conditions or thresholds of behavior; (3) a narrow focus on individuals rather than the social context, by turning social problems into clinical problems and largely ignoring social factors in causation or prevention; and (4) the expansion of medical social control, in terms of both medical surveillance of a range of bodily and behavioral measures and the use of pharmaceuticals, surgery, or other medical interventions to control or modify behaviors or body conditions.

It seems fair to say that medicalization has been increasing over the past four decades. While there are pockets of resistance to medicalization (e.g., fat acceptance groups, disability rights movement, backlash against medicating ADHD, etc.) these tend to be singular and insular. They might be seen as islands of resistance in a tidal wave of medicalization. Given this situation, it is probable that the trend toward medicalizing human problems will continue and likely expand into the twenty-first century. Specific areas that are likely to be medicalized include biomedical and genetic enhancements (e.g., genetic and cognitive enhancements), "risks" for illness or debility (e.g., especially newly measured genetic and physiological risks), anti-aging medicine (e.g., defining and treating more normal conditions of our rapidly aging population as illness or illness risk), and increasing consumer demands for more medical solutions to life's ills. Medicalization may be slowed by costs of treatment and the concerns and actions of organized critics of medicalized definitions and interventions.

SEE ALSO: Gender, Illness, and Social Construction; Medicine, Sociology of; Mental Illness, Medicalization of

REFERENCES

Clarke, A. E., Mamo, L., Fosket, J. R., Fishman, J. R., and Shim, J. K., eds. 2010. *Biomedicalization: Technoscience, Health, and Illness in the US.* Durham, NC: Duke University Press.

Clarke, A. E., Shim, J. K., Mamo, L., Fosket, J. R., and Fishman, J. R. 2003. "Biomedicalization: Technoscientific Transformations of Health, Illness, and US Biomedicine." *American Sociological Review* 68(2): 161–194.

Conrad, P. 1975. "The Discovery of Hyperkinesis: Notes on the Medicalization of Deviant Behavior." *Social Problems* 23(1): 12–21.

Conrad, P. 1976. *Identifying Hyperactive Children: The Medicalization of Deviant Behavior*. Lexington, MA: Heath.

Conrad, P. 2005. "The Shifting Engines of Medicalization." *Journal of Health and Social Behavior* 46(March): 3–14.

Conrad, P. 2007. *The Medicalization of Society: On the Transformation of Human Conditions into Treatable Disorders*. Baltimore, MD: Johns Hopkins University Press.

Conrad, P., and Schneider, J. W. 1980. *Deviance and Medicalization: From Badness to Sickness*. St. Louis, MO: Mosby.

Conrad, P., Mackie, T., and Mehrotra, A. 2010. "Estimating the Costs of Medicalization." *Social Science and Medicine* 70: 1943–1947.

Freidson, E. 1970. *Profession of Medicine*. New York: Harper and Row.

Loe, M. 2004. *The Rise of Viagra: How the Little Blue Pill Changed Sex in America*. New York: NYU Press.

Menacker, F., and Hamilton, B. E. 2010. *Recent Trends in Cesarean Delivery in the United States*. NCHS Data Brief no. 35. National Center for Health Statistics, Hyattsville, MD.

Parsons, T. 1951. *The Social System*. New York: Free Press.

Riska, E. 2003. "Gendering the Medicalization Thesis." In *Gender Perspectives on Health and Medicine: Key Themes*, edited by Marcia Texler Segal, Vasilikie Demos, and Jennie Jacobs Kronenfeld, 59–87. Oxford, UK: Elsevier.

Riska, E. 2010. "Gender and Medicalization and Biomedicalization Theories." In *Biomedicalization: Technoscience, Health, and Illness in the U.S.*, edited by A. E. Clarke, L. Mamo, J. R. Fosket, J. R. Fishman, and J. K. Shim, 147–170. Durham, NC: Duke University Press.

Rosenfield, D., and Faircloth, C. A., eds. 2006. *Medicalized Masculinities*. Philadelphia, PA: Temple University Press.

Waggoner, M. 2011. "Expanding the Reproductive Body: The Emergence of Preconception Care, 1980–2010." Unpublished doctoral dissertation, Brandeis University.

Zola, I. K. 1972. "Medicine as an Institution of Social Control." *Sociological Review* 20(4): 487–504.

Henry, David; Doust, Jenny & Moynihan, Ray

"Preventing overdiagnosis: how to stop harming the healthy", i *The BMJ* 344 (7859) 2012 ss. 1-6

Copyright (C) 2012 BMJ

Kopiert med hjemmel i Kopinor-avtalen

BMJ

BMJ 2012;344:e3502 doi: 10.1136/bmj.e3502 (Published 29 May 2012)

FEATURE

MEDICALISATION

Preventing overdiagnosis: how to stop harming the healthy

Evidence is mounting that medicine is harming healthy people through ever earlier detection and ever wider definition of disease. With the announcement of an international conference to improve understanding of the problem of overdiagnosis, **Ray Moynihan**, **Jenny Doust**, and **David Henry** examine its causes and explore solutions

Ray Moynihan *senior research fellow* [1], Jenny Doust *professor of clinical epidemiology*[2], David Henry *chief executive officer*[3]

[1]Bond University, Robina, Queensland, Australia; [2]Centre for Research in Evidence-Based Practice, Faculty of Health Sciences and Medicine, Bond University, Australia; [3]Institute for Clinical Evaluative Sciences, Toronto, Canada

Medicine's much hailed ability to help the sick is fast being challenged by its propensity to harm the healthy. A burgeoning scientific literature is fuelling public concerns that too many people are being overdosed,[1] overtreated,[2] and overdiagnosed.[3] Screening programmes are detecting early cancers that will never cause symptoms or death,[4] sensitive diagnostic technologies identify "abnormalities" so tiny they will remain benign,[5] while widening disease definitions mean people at ever lower risks receive permanent medical labels and lifelong treatments that will fail to benefit many of them.[3 6] With estimates that more than $200bn (£128bn; €160bn) may be wasted on unnecessary treatment every year in the United States,[7] the cumulative burden from overdiagnosis poses a significant threat to human health.

Narrowly defined, overdiagnosis occurs when people without symptoms are diagnosed with a disease that ultimately will not cause them to experience symptoms or early death.[3] More broadly defined, overdiagnosis refers to the related problems of overmedicalisation and subsequent overtreatment, diagnosis creep, shifting thresholds, and disease mongering, all processes helping to reclassify healthy people with mild problems or at low risk as sick.[8]

The downsides of overdiagnosis include the negative effects of unnecessary labelling, the harms of unneeded tests and therapies, and the opportunity cost of wasted resources that could be better used to treat or prevent genuine illness. The challenge is to articulate the nature and extent of the problem more widely, identify the patterns and drivers, and develop a suite of responses from the clinical to the cultural.

At the clinical level, a key aim is to better discriminate between benign "abnormalities" and those that will go on to cause harm. In terms of education and raising awareness among both the public and professionals, more honest information is needed about the risk of overdiagnosis, particularly related to screening. More deeply, mounting evidence that we're harming healthy people may force a questioning of our faith in ever-earlier detection, a renewal of the process of disease definition, and a fundamental shift in the systemic incentives driving dangerous excess.

Next year, an international scientific conference called Preventing Overdiagnosis aims to deepen understanding and awareness of the problem and its prevention. The conference will take place on 10-12 September 2013 in the United States, hosted by the Dartmouth Institute for Health Policy and Clinical Practice in partnership with the *BMJ*, the leading US consumer organisation Consumer Reports, and Bond University. The conference is timely, as growing concern about overdiagnosis is giving way to concerted action. The *Archives of Internal Medicine*'s feature "Less is More" now regularly augments the evidence base,[9] high level health policy groups in Europe are debating ways to tackle excess,[10] and the recently launched Choosing Wisely campaign warns about dozens of potentially unnecessary tests and treatments across nine specialties.[11]

Many factors—including the best of intentions—are driving overdiagnosis, but a key contributor is advances in technology. The literature suggests several broad and related pathways to overdiagnosis: screening detected overdiagnosis in people without symptoms; overdiagnosis resulting from use of increasingly sensitive tests in those with symptoms;

Correspondence to: R Moynihan raymoynihan@bond.edu.au

For personal use only: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

overdiagnosis made incidentally—"incidentalomas"; and overdiagnosis resulting from excessively widened disease definitions. These different pathways are not mutually exclusive, and a more rigorous classification of the different forms of overdiagnosis will be a focus of discussion at the 2013 scientific conference.

Screening detected overdiagnosis

This pathway to overdiagnosis occurs when a screening programme detects disease in a person without symptoms but the disease is in a form that will never cause that person symptoms or early death. Sometimes this form of disease is called pseudodisease. Contrary to popular notions that cancers are universally harmful and ultimately fatal, some cancers can regress, fail to progress, or grow so slowly that they will not cause harm before the individual dies from other causes.[5] As we will discuss below, there is now strong evidence from randomised trials and other studies comparing screened and unscreened populations that an important proportion of the cancer detected through some popular screening programmes may be pseudodisease.[4 12] Evidence from autopsy studies suggests a large reservoir of subclinical disease in the general population, including prostate, breast, and thyroid cancer, the bulk of which will never harm.[12] Similarly, screening the hearts of people without symptoms or at low risk may also lead to overdiagnosis of coronary atherosclerosis and subsequent unnecessary interventions.[13] Our understanding of the nature and extent of overdiagnosis and the amount of pseudodisease detected by screening remains limited but is evolving, and as Woolfe and Harris observed recently in *JAMA*, "concern about overdiagnosis is justified."[14]

Increasingly sensitive tests

People presenting to doctors with symptoms can also be overdiagnosed because changes in diagnostic technologies or methods have enabled the identification of less severe forms of diseases or disorders. It is becoming clearer that a substantial proportion of these earlier "abnormalities" will never progress, raising awkward questions about exactly when to use diagnostic labels and therapeutic approaches traditionally deployed against much more serious forms of disease.

Incidentalomas

Diagnostic scanning of the abdomen, pelvis, chest, head, and neck can reveal "incidental findings" in up to 40% of individuals being tested for other reasons.[15] Some of these are tumours, and most of these "incidentalomas" are benign. A very small number of people will benefit from early detection of an incidental malignant tumour, while others will suffer the anxiety and adverse effects of further investigation and treatment of an "abnormality" that would never have harmed them. As others have shown, the rapidly rising incidence for some cancers, set against relatively stable death rates, is a phenomenon suggestive of widespread overdiagnosis, whether from screening or the detection of incidentalomas (figure⇓).[12]

Excessively widened definitions

Another pathway to overdiagnosis is through disease boundaries being widened and treatment thresholds lowered to a point where a medical label and subsequent therapy may cause people more harm than good. Changing diagnostic criteria for many conditions are routinely increasing the numbers of people defined as sick,[16] causing virtually the entire older adult population to be classified as having at least one chronic condition.[17] This widening has happened both with asymptomatic conditions that carry a risk of an adverse event, such as osteoporosis, where treatments may do more harm than good for those at very low risk of fracture,[18] and for behavioural conditions such as female sexual dysfunction, where common difficulties have been reclassified as dysfunctions.[19]

Such changes in diagnostic criteria are commonly made by panels of health professionals with financial ties to companies that benefit directly from any expansion of the patient pool.[20] As definitions broaden and thresholds fall, people with smaller risks or milder problems are labelled, which means the potential benefits of treatment decline, raising the possibility that harms will outweigh benefits. As Welch and colleagues estimated in their 2011 book *Overdiagnosed*,[3] many people diagnosed and treated long term for near-normal cholesterol concentration or near-normal osteoporosis may be "overdiagnosed," in the sense that they would never have experienced the events their treatments are designed to prevent.

A related form of overdiagnosis occurs when people are diagnosed outside of already widened diagnostic criteria, as can occur when inappropriate manufacturers' norms exaggerate the incidence of abnormality,[21] when diagnostic methods wrongly label random or normal fluctuations in biomarkers as true abnormalities,[22] or when important qualifiers are left out of the process of diagnosis.[23]

Examples of overdiagnosis

The growing evidence on overdiagnosis suggests the problem may exist to varying extents across many conditions (box 1), including those for which underdiagnosis may simultaneously be a feature. For some conditions, the evidence remains tentative and speculative, for others it has become much more robust.

Breast cancer

Arguably the strongest evidence of overdiagnosis comes from studies of screening detected breast cancers, though estimates of its extent are wide ranging. A 2007 systematic review in *Lancet Oncology* found the proportion of overdiagnosis of invasive breast cancer among women in their 50s ranged from 1.7% to 54%.[24] An Australian study estimated the rate was at least 30%,[25] while a Norwegian study calculated 15-25%.[26] A 2009 systematic review in the *BMJ* concluded up to one third of all screening detected cancers may be overdiagnosed.[4] However, even with strong evidence from population based studies, it is currently impossible to discriminate between cancers that will harm and those that will not.

Thyroid cancer

While the chances of tests detecting a thyroid "abnormality" are high, the risk it will ever cause harm is low.[3 27] Analysis of rising incidence shows many of the newly diagnosed thyroid cancers are the smaller and less aggressive forms not requiring treatment,[28] which itself carries the risk of damaged nerves and long term medication.[3]

Gestational diabetes

A 2010 revision of the criteria defining gestational diabetes recommended a dramatic lowering of the diagnostic threshold, more than doubling the number of pregnant woman classified to almost 18%.[29] Proponents argue universal screening with the new definition will reduce health problems, including babies being "large for gestational age."[29] Critics, however, are calling

For personal use only: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

Box 1: Problems of overdiagnosis

Asthma—Canadian study suggests 30% of people with diagnosis may not have asthma, and 66% of those may not require medications[37]

Attention deficit hyperactivity disorder—Widened definitions have led to concerns about overdiagnosis; boys born at the end of the school year have 30% higher chance of diagnosis and 40% higher chance of medication than those born at the beginning of the year[46]

Breast cancer—Systematic review suggests up to a third of screening detected cancers may be overdiagnosed[4]

Chronic kidney disease—Controversial definition classifies 1 in 10 as having disease; concerns about overdiagnosis of many elderly people [23]

Gestational diabetes—Expanded definition classifies almost 1 in 5 pregnant women [31]

High blood pressure—Systematic review suggests possibility of substantial overdiagnosis[22]

High cholesterol—Estimates that up to 80% of people with near normal cholesterol treated for life may be overdiagnosed[3]

Lung cancer—25% or more of screening detected lung cancers may be overdiagnosed[56]

Osteoporosis—Expanded definitions may mean many treated low risk women experience net harm[18]

Prostate cancer—Risk that a cancer detected by prostate specific antigen testing is overdiagnosed may be over 60%[12]

Pulmonary embolism—Increased diagnostic sensitivity leads to detection of small emboli. Many may not require anticoagulant treatment [39]

Thyroid cancer—Much of the observed increase in incidence may be overdiagnosis[28]

for an urgent debate before the new expanded definition is more widely adopted, because they fear many women may be overmedicalised and overdiagnosed, that the screening test has poor reproducibility for mild cases, the evidence of benefit for the newly diagnosed pregnant women is weak, and the benefit modest at best.[30 31]

Chronic kidney disease

More than 10% of adults in the United States are now classified as having some form of chronic kidney disease.[32] A working definition launched as part of new clinical guidelines[33] asserts that an estimated glomerular filtration rate (eGFR) below 60 ml/min/1.73m^2 and sustained for three months or longer is deemed abnormal, a decision critics argue automatically creates the potential for overdiagnosis, particularly among elderly people.[34]

According to Winearls and Glassock in an article last year the new classification system is "like a fishing trawler" and "captures many more innocent subjects than it should."[23] They estimate that up to one third of people over 65 may meet the new criteria, yet of these, fewer than 1 in 1000 will develop end stage renal disease each year. They also point to major problems with the reliability and consistency of the eGFR test and express concern many older people are being labelled on the basis of a single and potentially inaccurate laboratory measure. Elsewhere they have argued that "the majority of those held to have CKD [chronic kidney disease] have no identifiable kidney disease" and they've highlighted attempts by some organisations to move away from the controversial new definition, raise the threshold for diagnosis, and dramatically reduce prevalence.[35] Responding to criticisms, proponents have defended the new definition as being "clear, simple, and useful."[36]

Asthma

Although asthma can be severe and may be underdiagnosed and undertreated, some studies suggest that there may also be substantial overdiagnosis. One large study in 2008 found that almost 30% of people diagnosed as having asthma did not have the condition, and almost 66% of those did not need drugs or asthma care during six months of follow-up.[37] The authors concluded, "A substantial proportion of people . . . may be overdiagnosed with asthma and may be prescribed asthma medications unnecessarily." In the same year a Dutch study found that of 1100 patients using inhaled corticosteroids, 30% may have been using the drugs without any clear indications.[38]

Pulmonary embolism

Doctors think of pulmonary embolism as a "not to be missed" diagnosis, because failure to detect it can have catastrophic consequences. Historically it was diagnosed only when the blockage was large enough to cause infarction of part of the lung or haemodynamic instability. In such patients, treatment with an anticoagulant or a thrombolytic agent was considered mandatory. Now, however, computed tomography (CT) pulmonary angiography can detect smaller clots, and there is uncertainty about whether treatment is always necessary.[39] Analysing trends before and after the widespread introduction of CT pulmonary angiography, Weiner and colleagues suggested that the almost doubling in incidence "reflects an epidemic of diagnostic testing that has created overdiagnosis," with much of the increase consisting of "clinically unimportant" cases that "would not have been fatal even if left undiagnosed and untreated."[40] An observational study is investigating the safety of not treating people with very small blood clots.[41]

Attention deficit hyperactivity disorder

Much has been written about expanding diagnostic definitions within mental illness and concerns about the dangers of overtreatment.[42] Debate has intensified with suggestions that current processes for defining disease may be contributing to the widespread overdiagnosis of conditions such as bipolar, autistic disorder, and attention deficit hyperactivity disorders.[43 44] One focus of concern is the possible overdiagnosis of children, who have no say in the appropriateness of a label that can permanently change their lives. This is particularly salient with attention deficit hyperactivity disorder.[45] A recent study of almost a million Canadian children found boys born in December (typically the youngest in their year) had a 30% higher chance of diagnosis and 40% higher chance of receiving medication than those born in January, with the authors concluding their findings "raise concerns about the potential harms of overdiagnosis and overprescribing."[46]

Drivers of overdiagnosis

The forces driving overdiagnosis are embedded deep within the culture of medicine and wider society, underscoring the challenges facing any attempt to combat them (box 2). A key driver is technological change itself. As Black described in 1998, the ability to detect smaller abnormalities axiomatically tends to increase the prevalence of any given disease.[5] In turn this leads to overestimation of the benefits of therapies, as milder forms of the disease are treated and improvements in health are

For personal use only: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

wrongly ascribed to treatment success, creating a "false feedback" loop fuelling a "cycle of increasing testing and treatment, which may eventually cause more harm than benefit."[5]

The industries that benefit from expanded markets for tests and treatments hold widereaching influence within the medical profession and wider society, through financial ties with professional and patient groups and funding of direct-to-consumer advertising, research foundations, disease awareness campaigns, and medical education.[8] Most importantly, the members of panels that write disease definitions or treatment thresholds often have financial ties to companies that stand to gain from expanded markets.[20] Similarly, health professionals and their associations may have an interest in maximising the patient pool within their specialty, and self-referrals by clinicians to diagnostic or therapeutic technologies in which they have a commercial interest may also drive unnecessary diagnosis.

Avoidance of litigation and the psychology of regret is another obvious driver as professionals can be punished for missing the early signs of disease yet don't generally face sanctions for overdiagnosing. Quality measures focused on doing more may also encourage overdiagnosis in order to meet targets for remuneration incentives.[47]

An intuitive belief in early detection, fed by deep faith in medical technology is arguably at the heart of the problem of overdiagnosis. Increasingly we've come to regard simply being "at risk" of future disease as being a disease in its own right. Starting with treatment of high blood pressure in the middle of the 20th century,[48] increasing proportions of the healthy population have been medicalised and medicated for growing numbers of symptomless conditions, based solely on their estimated risk of future events. Although the approach has reduced suffering and extended life for many, for those overdiagnosed it has needlessly turned the experience of life into a tangled web of chronic conditions. The cultural norm that "more is better" is confirmed by recent evidence suggesting patient satisfaction flows from increased access to tests and treatments, even though more care may be associated with greater harm.[49 50]

What can we do about overdiagnosis?

Building on existing knowledge and activity, the 2013 conference on overdiagnosis will provide a forum for learning more, increasing awareness, and developing ways to prevent the problem (www.preventingoverdiagnosis.net). Research on overdiagnosis is now recognised as part of the future scientific direction of the National Cancer Institute's division of cancer prevention in the United States.[51] The 2013 conference hopes to provide researchers working in this field with the chance to share and debate methods and further advance research agendas. As to education, the development of a range of curriculums and information packages could help raise awareness about the risks of overdiagnosis, particularly associated with screening.[52] In association with the *BMJ,* a series of articles about the potential for overdiagnosis within specific conditions is being planned. And at the level of clinical practice new protocols are being developed to bring more caution in treating incidentalomas.[3] Similarly, some are urging that we consider raising the thresholds that define "abnormal"—in breast cancer screening, for example—and evaluate methods of observing changes to some suspected pathologies over time, rather than intervening immediately.[53] As we've seen, early studies of how to safely undiagnose or de-prescribe are starting to emerge.

At a policy level, reform of the process of defining disease is urgently required, with one model coming from the National Institutes of Health in the United States, where people with financial or reputational conflicts of interest are disqualified from panel membership.[20] Dispassionate assessment of evidence may result in disease definitions being narrowed, as has been seen with the recent tentative proposals to raise thresholds for high blood pressure that could demedicalise up to 100 million people.[54] Processes for defining disease may also benefit from an attempt to synthesise the evidence from clinical medicine with literature on the wider social and environmental determinants of health. Other policy reforms could review the permanency of some diagnostic labels, address calls for increased independence in the design and running of scientific studies,[55] and adjust the structural and legal incentives driving overdiagnosis.

Concern about overdiagnosis does not preclude awareness that many people miss out on much needed healthcare. On the contrary, resources wasted on unnecessary care can be much better spent treating and preventing genuine illness. The challenge is to work out which is which, and to produce and disseminate evidence to help us all make more informed decisions about when a diagnosis might do us more good than harm.

We thank the participants at the April 2012 Coolangatta planning meeting on overdiagnosis: Paul Glasziou, Kirsten McCaffery, Melissa Sweet, Hilda Bastion, Andrew Wilson, Ian Scott, Suzanne Hill, Alexandra Barratt, Steve Woloshin, Lisa Schwartz, Fiona Godlee, and Rae Thomas. We also thank Julia Lowe.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: JD is supported by NHMRC project grant 511217; they have no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; all authors were at the April 2012 planning meeting for the 2013 conference and RM is undertaking a PhD on overdiagnosis.

Provenance and peer review: Commissioned; externally peer reviewed.

1 Abramson J. Overdosed America: the broken promise of American medicine. Harper, 2004.
2 Brownlee S. Overtreated: why too much medicine is making us sicker and poorer Bloomsbury, 2007.
3 Welch G, Schwartz L, Woloshin S. *Overdiagnosed: making people sick in pursuit of health.* Beacon Press, 2011.
4 Jørgensen K, Gøtzsche P. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;339:b2587.
5 Black W. Advances in radiology and the real versus apparent effects of early diagnosis. *Eur J Radiol* 1998;27:116–22.
6 Moynihan R, Heath I, Henry D. Selling sickness: the pharmaceutical industry and disease-mongering. *BMJ* 2002;324:886-91.
7 Berwick D, Hackbarth, A. Eliminating waste in US health care. *JAMA* 2012;307:1513-6.
8 Moynihan R, Cassels A. Selling sickness: how the world's biggest pharmaceutical companies are turning us all into patients. Nation Books, 2005.
9 Grady D, Redberg R. Less is more: how less health care can result in better health. *Arch Intern Med* 2010;170:749-50.
10 European Health Forum Gastein. Innovation and wellbeing: living longer—but are we living better? Press release, 8 September 2011. www.ehfg.org/826.html#c1950.
11 Choosing Wisely. US physician groups identify commonly used tests or procedures they say are often not necessary. Press release, 4 April 2012. http://choosingwisely.org/wp-content/uploads/2012/03/033012_Choosing-Wisely-National-Press-Rls-FINAL.pdf .
12 Welch G, Black W. Overdiagnosis in cancer. *JNCI* 2010;102:605-13.
13 Lauer M. Pseudodisease, the next great epidemic in coronary atherosclerosis? *Arch Intern Med* 2011;171:1268-9.
14 Woolfe S, Harris R. The harms of screening. *JAMA* 2012;307:565-6.
15 Orme NM, Fletcher JG, Siddiki HA, Harmsen WS, O'Byrne MM, Port JD, et al. Incidental findings in imaging research: evaluating incidence, benefit, and burden. *Arch Intern Med* 2010;170:1525-32.
16 Schwartz LM, Woloshin S. Changing disease definitions: implications for disease prevalence: analysis of the third national health and nutrition examination survey, 1988–1994. *Effect Clin Pract* 1999;2:76-85.
17 Kaplan R, Ong M. Rationale and public health implications of changing CHD risk factor definitions. *Annu Rev Public Health* 2007;28:321-44.
18 Herndon MB, Schwartz LM, Woloshin S, Welch G. Implications of expanding disease definitions: the case of osteoporosis. *Health Aff* 2007;26:1702-11.
19 Moynihan R. Merging of marketing and medical science: female sexual dysfunction. *BMJ* 2010;341:c5050.
20 Moynihan, R. A new deal on disease definition. *BMJ* 2011;342:d2548.

For personal use only: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

Drivers of overdiagnosis

- Technological changes detecting ever smaller "abnormalities"
- Commercial and professional vested interests
- Conflicted panels producing expanded disease definitions and writing guidelines
- Legal incentives that punish underdiagnosis but not overdiagnosis
- Health system incentives favouring more tests and treatments
- Cultural beliefs that more is better; faith in early detection unmodified by its risks

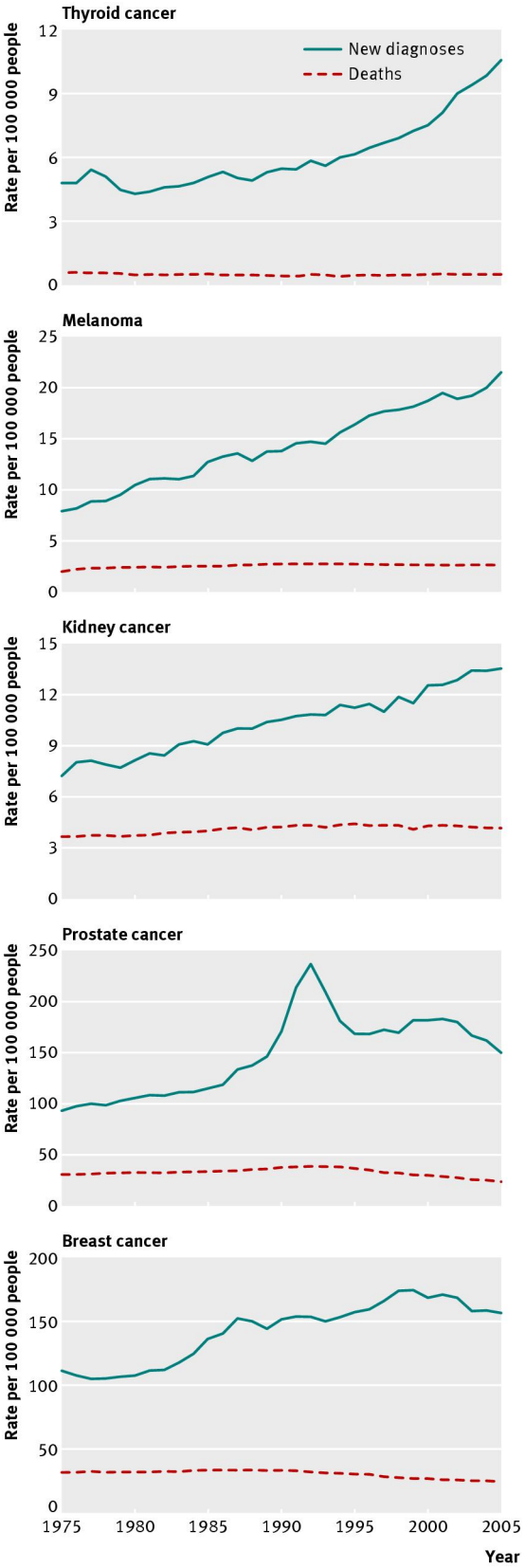
21 Ahmed A, Blake G, Rymer J, Fogelman I. Screening for osteopenia and osteoporosis: do the accepted normal ranges lead to overdiagnosis? *Osteoporos Int* 1997;7:432-8.
22 Hodgkinson J, Mant J, Martin U, Guo B, Hobbs F, Deeks J, et al. Relative effectiveness of clinic and home blood pressure monitoring compared with ambulatory blood pressure monitoring in diagnosis of hypertension: systematic review. *BMJ* 2011;342:d3621.
23 Winearls C, Glassock R. Classification of chronic kidney disease in the elderly: pitfalls and errors. *Nephron Clin Pract* 2011;119(suppl 1):c2-4.
24 Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. *Lancet Oncol* 2007;8:1129-38.
25 Morrell S, Barratt,A, Irwig L, Howard K, Biesheuvel C, Armstrong B. Estimates of overdiagnosis of invasive breast cancer associated with screening mammography *Cancer Causes Control* 2010;21:275-82.
26 Kalager M, Adami H, Bretthauer M, Tamimi R. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. *Ann Intern Med* 2012;156:491-9.
27 Tan G, Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. *Ann Intern Med* 1997;126:226-31.
28 Davies L, Welch G. Increasing incidence of thyroid cancer in the United States, 1973-2002. *JAMA* 2006;295:2164-7.
29 International Association of Diabetes and Pregnancy Study Groups. Recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care* 2010;33:676-82.
30 Ryan E. Diagnosing gestational diabetes. *Diabetologia* 2011;54:480-6.
31 Cundy T. Proposed new diagnostic criteria for gestational diabetes—a pause for thought? *Diabet Med* 2012;29:176-80.
32 Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;150:604-12.
33 National Kidney Foundation. Clinical practice guidelines for chronic kidney disease: evaluation, classification and stratification. *Am J Kidney Dis* 2002;39(suppl 1):s17-31.
34 Glassock R. Estimated glomerular filtration rate: time for a performance review? *Kidney Int* 2009;75:1001-3.
35 Winearls C, Haynes, R, Glassock R. CKD staging—evolution not revolution. *Nefrologia* 2010;30:493-500.
36 Eknoyan G. Chronic kidney disease definition and classification: the quest for refinements. *Kidney Int* 2007;72:1183-5.
37 Aaron S, Vandemheen K, Boulet L, McIvor R, FitzGerald J, Hernandez P, et al. Overdiagnosis of asthma in obese and nonobese adults. *CMAJ* 2008;179:1121-31.
38 Lucas A, Smeenk F, Smeele I, van Schayck C. Overtreatment with inhaled corticosteroids and diagnostic problems in primary care patients, an exploratory study. *Fam Pract* 2008;25:86-91.
39 Prasad V, Rho J, Cifu A. The diagnosis and treatment of pulmonary embolism. *Arch Intern Med* 2012 Apr [Epub ahead of print].
40 Weiner R, Schwartz LM, Woloshin S. Time trends in pulmonary embolism in the United States. *Arch Intern Med* 2011;171:831-7.
41 Carrier M. A study to evaluate the safety of withholding anticoagulation in patients with subsegmental PE who have a negative serial bilateral lower extremity ultrasound (SSPE). 2011. http://clinicaltrials.gov/ct2/show/NCT01455818.
42 Angell M. The epidemic of mental illness: why? *New York Review Books* 2011 Jun 23.
43 Frances A. The first draft of DSM-V. *BMJ* 2010;340:c1168.
44 Healy D. The latest mania: selling bipolar disorder. *PLoS Med* 2006;3:e185.
45 Thomas R. The diagnostic variability in attention deficit hyperactivity disorder. Presentation to Overdiagnosis Meeting, Coolangatta, 29-30 April 2012.
46 Morrow R, Garland E, Wright J, Maclure M, Taylor S, Dormuth C. Influence of relative age on diagnosis and treatment of attention-deficit/hyperactivity disorder in children. *CMAJ* 2012;184:755-62.
47 Heath I, Hippisley-Cox J, Smeeth L. Measuring performance and missing the point. *BMJ* 2007;335;1075-6.
48 Greene J. Prescribing by numbers; drugs and the definition of disease. Johns Hopkins University Press, 2007.
49 Fenton J, Jerant A, Bertakis K, Franks P. The cost of satisfaction. *Arch Intern Med* 2012;172:405-11.
50 Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending, part 2: health outcomes and satisfaction with care. *Ann Intern Med* 2003;138:288-98.
51 Kramer B. Current and future perspectives on cancer prevention research. National Cancer Advisory Board, 2012. http://deainfo.nci.nih.gov/advisory/ncab/161_0212/Kramer.pdf.
52 Schwartz L, Woloshin S, Fowler F, Welch G. Enthusiasm for cancer screening in the United States. *JAMA* 2004;291:71-8.
53 Elmore J, Fletcher S. Overdiagnosis in breast cancer screening: time to tackle an underappreciated harm. *Ann Intern Med* 2012;156:536-7.
54 Taylor B, Wilt T, Welch G. Impact of diastolic and systolic blood pressure on mortality: implications for the definition of "normal". *J Gen Intern Med* 2011;26:685-90.
55 Prasad V, Cifu A, Ioannidis J. Reversals of established medical practices. *JAMA* 2012;307:37-8.
56 Reich J. A critical appraisal of overdiagnosis: estimates of its magnitude and implications for lung cancer screening. *Thorax* 2008;63:377-83.

Cite this as: *BMJ* 2012;344:e3502

© BMJ Publishing Group Ltd 2012

For personal use only: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

Figure



Rates of new diagnosis and death for five types of cancer in the US, 1975-2005. Adapted from Welch and Black[12]

For personal use only: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

Larsen, Øyvind; Keeping, David & Nylenna, Magne

"Kap. 2.6: For mye helsetjenester", i

Samfunnsmedisin på norsk 2019 ss. 159-165

Copyright (C) 2019 Gyldendal

Kopiert med hjemmel i Kopinor-avtalen

2.6

For mye helsetjenester?

Helsetjenesten i Norge og de fleste sammenliknbare land har vokst gradvis de senere tiårene. Ikke bare har budsjettene og antall ansatte økt, det har også mulighetene for diagnostikk og behandling. Arbeids- og ansvarsområdet er utvidet, og ambisjonene er blitt større. Denne ekspansjonen har bivirkninger, og spørsmålet om hvor grensene bør gå for helsetjenestens virksomhet, diskuteres. Det at stadig flere menneskelige problemer forklares medisinsk og behandles av helsetjenesten, kalles gjerne *medikalisering*.

Tre utviklingstrekk ved helsetjenesten har vært tydelige:

- *Utvidelse av sykdomsbegrepet,* der tilstander som tidligere ble regnet som påregnelige plager og problemer, blir diagnoser, og der grensene for det normale blir smalere. Det gjelder både for biologiske variabler, som blodtrykk og kolesterolnivå, og for personlighetstrekk og mentale funksjoner. Innen psykisk helse er innsnevring av normalitet beskrevet som et økende problem (Frances 2013).
- *Flere og nye metoder for diagnostikk og behandling.* Avansert bildediagnostikk som CT og MR avdekker ofte uventede eller «unormale» funn som ikke nødvendigvis betyr sykdom. Kirurgisk og medikamentell behandling blir stadig mer potent. Kombinert med sterk intervensjonsiver og lite rom for «å vente og se» øker dette faren for bivirkninger og komplikasjoner.
- *Oppsøkende virksomhet,* det vil si jakt på mulige, fremtidige pasienter blant symptomfrie personer, enten systematisk (som ved screening) eller mer usystematisk kombinert med tilfeldige kontakter med helsetjenesten (ofte kalt villscreening).



Da American Heart Association i 2017 reduserte grensene for blodtrykksbehandling fra 140/90 mm Hg til 130/80 mm Hg, økte andelen voksne amerikanere med det som ble ansett som «hypertensjon», fra 32 % til 46 % (Bakris & Sorrentino 2018). (Foto Ø.L. 2019)

Alternativ medisin

Alternativ medisin er en fellesbetegnelse på medisinsk praksis som ikke baserer seg på naturvitenskapelig forskning, og som utøves utenfor den offentlige helsetjenesten. Betegnelser som naturmedisin og helhetsterapi brukes også mer eller mindre synonymt. Alternative behandlingsformer baseres ikke på dokumentert effekt i vitenskapelig forstand, anses ikke som kunnskapsbasert, og bygger ofte på teorier som ikke er etterprøvbare. *Komplementær medisin* betegner bruken av alternativ medisin som tillegg til kunnskapsbasert medisin. Ofte slås disse betegnelsene sammen til «komplementær og alternativ medisin».

De mest brukte behandlingsformene innen alternativ medisin er massasje, akupunktur og naprapati. Soneterapi, homøopati, osteopati og healing er andre eksempler på alternative behandlingsformer.

Nasjonalt forskningssenter innen komplementær og alternativ medisin (NAFKAM) er et offentlig forsknings- og informasjonssenter ved Universitetet i Tromsø som driver forskning på og opplysningsvirksomhet om alternativ behandling. Senteret kartlegger jevnlig bruken av alternativ medisin gjennom befolkningsundersøkelser. I 2016 angav 36 % av befolkningen at de hadde brukt alternativ behandling siste år. Det var en reduksjon fra 45 % i 2012 og 40 % i 2014. Flere kvinner enn menn bruker alternative behandlingsformer.[47]

47 http://nafkam.no/forskning-og-utvikling/bruk-av-alternativ-behandling-i-norge-2016

Ettersom alternativ medisin ikke er en del av den offentlige helsetjenesten, må slike tjenester finansieres privat og uten offentlig refusjon. NAFKAM har beregnet at det ble brukt over 4 milliarder kroner på alternativ behandling i Norge i 2016. Selv om færre oppsøker alternative behandlere enn før, brukes det stadig mer penger på slik behandling.

Uønsket variasjon i helsetjenesten

Oppmerksomheten omkring mulig overforbruk av helsetjenester begynte med påvisning av store geografiske variasjoner i bruk av tjenester (Wennberg 2010). Når slike forskjeller ikke lar seg forklare av sykdomsforekomst eller andre kjente variabler, kaller vi det *uønsket variasjon*. Uønsket variasjon aktualiserer spørsmålet om over- og underforbruk av ulike tjenester i både diagnostikk og behandling. For ni av 12 dagkirurgiske inngrep er det vist at boområdene i Norge med høyest forbruk har over dobbelt så høy operasjonsrate som dem med lavest forbruk. For meniskoperasjoner, skulderkirurgi og hemoroideoperasjoner er forholdstallet omtrent fire (SKDE 2015a). I én del av landet innlegges barn dobbelt så ofte i sykehus som i en annen del, uten noen kjent medisinsk forklaring. Barn med astma i ett boområde har nesten fire ganger så hyppig konsultasjon i spesialisthelsetjenesten som barn fra et annet boområde (SKDE 2015b). For eldre er det særlig bruken av polikliniske spesialisthelsetjenester som varierer. Noe av variasjonen kan skyldes ulike tradisjoner i arbeidsfordelingen mellom kommunehelsetjenesten og spesialisthelsetjenesten. Likevel er det uforklarlig at konsultasjonsraten i spesialisthelsetjenesten er fire til elleve ganger høyere ett sted sammenliknet med et annet for EKG, hjertesvikt og demens (SKDE 2017).

Slik variasjon kan skyldes underforbruk ett sted eller overforbruk et annet sted, eller en kombinasjon. Overforbruk kan defineres som helsetjenester som pasientene ikke har nytte av, eller som de ikke ville valgt om de hadde fått god veiledning. Underforbruk foreligger når pasienten ikke får nødvendig helsehjelp (Legeforeningen 2016). Problemet er at det er vanskelig å fastslå hva som er riktig forbruk av helsetjenester.

I tillegg til at uønsket variasjon medfører uheldig ressursbruk og feil økonomiske prioriteringer, kan både under- og overforbruk av helsetjenester være skadelig for den enkelte pasient.

Overdiagnostikk og overbehandling

Hensikten med å stille diagnoser og iverksette behandling er å bedre helsetilstanden til en person ved å reversere, stanse eller forsinke en sykdomsutvikling. Å diagnostisere eller behandle tilstander som ikke bør karakteriseres som sykdom, eller som ikke vil utvikle seg til sykdom, vil ikke bedre helsetilstanden hos noen. I verste fall blir resultatet motsatt. Det å bli definert som «pasient» kan være uheldig i seg selv, og både medisinske undersøkelses- og behandlingsmetoder kan ha direkte skadelige følger (Welch 2011). Leger har en tendens til å overvurdere nytten ved medisinsk behandling, og å undervurdere ulempene (Hoffmann & Del Mar 2017).

Med *overdiagnostikk* menes medisinske funn og diagnoser som ikke gir noen nytte eller fordel for pasienten (Carter et al. 2015). Overdiagnostikk kan føre til unødvendig behandling, som igjen kan gi bivirkninger og komplikasjoner. Nye diagnostiske metoder, som avansert bildediagnostikk og biomarkører, vil kunne påvise avvik fra det som regnes som normalt, uten at det verken gir symptomer eller noen gang vil utvikle seg til sykdom (Hofmann & Welch 2017).

Det utarbeides jevnlig internasjonale oversikter over undersøkelser og behandlingsmetoder som ikke har dokumentert nytte, som brukes for ofte (Morgan et al. 2018).

Risikant

Nyordet *risikant* ble lansert av samfunnsmedisineren Jostein Holmen (f. 1947) i 1994 som betegnelse på en person som føler seg frisk, men som får påvist én eller flere risikofaktorer for sykdom og eventuelt tidlig død (Holmen 1994a).

Risikoen kan som regel reduseres enten ved livsstilsendring eller ved hjelp av legemidler, og slike intervensjoner iverksettes på stadig videre indikasjoner (Holmen 1994b). Risiko for sykdom har etter hvert fått mye oppmerksomhet i helsetjenesten, og er i ferd med å bli en diagnose i seg selv (Swensen 2000).

Behandling av risiko for sykdom hos folk uten symptomer reiser mange spørsmål: Hvor stor må faren for og konsekvensene av en mulig framtidig sykdom være for å forsvare de menneskelige og økonomiske kostnadene ved risikoreduserende tiltak? Hvor mange ulike medisinske risikofaktor er det rimelig å teste en person og en befolkning for? Gir risikostrategien flere gode leveår for befolkningen, eller fører den til en uheldig medikalisering?

Screening

Screening er en spesiell form for diagnostisering der målet er å identifisere personer med en risikofaktor for sykdom, eller en sykdom som ennå ikke har gitt symptomer, med håp om å bedre prognosen for den eller de sykdommer som er gjenstand for undersøkelsen.

Systematisk screening er masseundersøkelser av hele eller deler av befolkningen, fortrinnsvis med henblikk på alvorlige sykdommer med høy dødelighet. Et godt eksempel er blodprøver av nyfødte, som screenes for 25 ulike medfødte sykdommer (2019). Det er to nasjonale screeningprogrammer for kreft i Norge: screening for brystkreft (kvinner 50–69 år) og screening for livmorhalskreft (kvinner 25–69 år). I tillegg pågår et pilotprosjekt for tarmkreftscreening.

Villscreening eller sykdomsjakt (case-finding) er betegnelsen på usystematisk bruk av diagnostiske tester på individnivå uten spesifikk indikasjon.

Det er to hovedutfordringer ved screeningundersøkelser:

- For det første må det være sikkert at tidlig behandling gir bedre resultat enn om man venter til symptomer oppstår.
- Dernest må testene som benyttes, være trygge og ha tilstrekkelig sensitivitet (evne til å identifisere de syke) og spesifisitet (evne til å frikjenne de friske) til at den prediktive verdien av screeningen blir høy. Både falskt positive og falskt negative prøvesvar skaper problemer.

I tillegg må kostnadene ved undersøkelsen stå i forhold til gevinsten.

Screeningundersøker er kontroversielle, og effekten av mange screeningundersøkelser er omdiskutert. Det er for eksempel ingen dokumentasjon for at generelle helseundersøkelser hos voksne reduserer sykelighet eller dødelighet (Krogsbøll et al. 2012).

Sensitivitet, spesifisitet og prediktiv verdi

Tabell 8 Ingen diagnostiske tester er ideelle i den forstand at de med 100 % sikkerhet identifiserer de som er syke, og like sikkert frikjenner de friske. Dette er en stor utfordring ved screeningundersøkelser.

Sensitiviteten («sensisykiteten») er testens evne til å finne de syke – sannsynligheten for at en syk person får positivt testresultat. Sensitiviteten uttrykkes som andelen sant positive av alle med sykdom, a/a+c.

Spesifisiteten («spesifriskiteten») er testens evne til å finne de friske – sannsynligheten for at en frisk person får et negativt testresultat. Spesifisiteten uttrykkes som andelen med sant negative av alle uten sykdom, d/b+d.

		Sykdom	
		Ja	Nei
Testresultat	Positivt	Sant positiv (a)	Falskt positiv (b)
	Negativt	Falskt negativ (c)	Sant negativ (d)

Lav sensitivitet øker faren for at noen med sykdom blir oversett, mens lav spesifisitet betyr at mange friske får beskjed om at de kanskje er syke. Sensitivitet og spesifisitet er konkurrerende størrelser, særlig når det brukes kontinuerlige biologiske variabler, for eksempel PSA (prostataspesifikt antigen)-nivå i blodet for å finne prostatakreft. Jo lavere grensen settes, desto høyere blir sensitiviteten (og desto færre oversette krefttilfeller), men desto lavere blir spesifisiteten (og desto flere friske blir miskjent som syke) (Tretli & Weiderpass 2007).

Testens *prediktive verdi* (utsagnskraft) er et uttrykk for hvor sannsynlig det er at testsvaret som er funnet, er korrekt. Positiv prediktiv verdi er sannsynligheten for at den undersøkte virkelig er syk, om testen var positiv, a/a+b. Negativ prediktiv verdi er sannsynligheten for at den undersøkte virkelig er frisk, om testen var negativ, d/c+d. Testers prediktive verdi avhenger helt og holdent av hvor hyppig den aktuelle sykdommen forekommer i den undersøkte befolkningen, det vil si prevalensen av sykdommen. Jo sjeldnere en sykdom forekommer, desto lavere blir den positive prediktive verdien av en test, selv om sensitivitet og spesifisitet er den samme. Selv med både sensitivitet og en spesifisitet på 99 % vil en test som benyttes i en befolkning der bare én av tusen (en promille) i virkeligheten er syke, ende opp med ti falskt positive testsvar for hver syk person som identifiseres (Thelle & Laake 2008).

Å gjøre kloke valg

Bekymringen for at noen helsetjenester kan gjøre mer skade enn gagn, har ført til flere internasjonale initiativer for å bevisstgjøre helsepersonell og pasienter. *Too much medicine*[48] er en britisk kampanje for å framheve helsefarene ved overdiagnostikk og ressursbruken knyttet til unødvendig behandling. *Choosing wisely*[49] er en amerikansk kampanje for å stimulere helsepersonell og pasienter til å velge undersøkelser og behandling som er kunnskapsbaserte, nødvendige og ufarlige.

I Norge lanserte Legeforeningen i samarbeid med Norsk Pasientforening og andre helseprofesjonsforeninger i 2018 kampanjen *Gjør kloke valg*[50] for å redusere utredning og behandling som pasienter har lite eller ingen nytte av. Målet er ikke økonomiske innsparinger, men å unngå unødvendige og potensielt skadelige tiltak. I en erkjennelse av at pasienter ofte blir testet, utredet og behandlet for mye, inviteres pasientene blant annet til å stille legen fire spørsmål:

- Er denne testen/prosedyren/behandlingen nødvendig?
- Hva er bivirkningen(e)?
- Finnes det alternativer?
- Hva skjer om jeg ikke tar testen/prosedyren/behandlingen?

Samtidig utfordres Legeforeningens ulike fagmedisinske foreninger til å identifisere undersøkelser og behandlinger som kan utsette pasienten for belastning eller skade, og som bør reduseres eller avvikles.

48 https://www.bmj.com/too-much-medicine
49 http://www.choosingwisely.org/
50 https://beta.legeforeningen.no/kloke-valg/

Hofmann, Bjørn

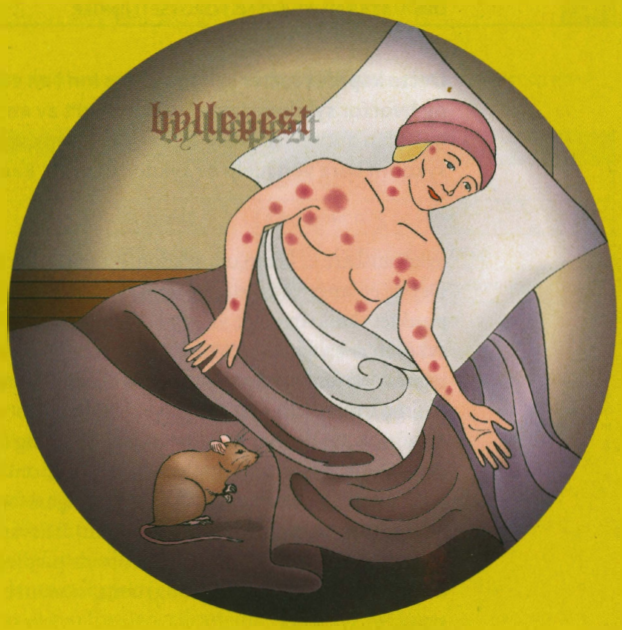
"Kap. 1: Hva er sykdom?", i (Red) Ørn, Stein & Bach-Gansmo, Edvin

Sykdom og behandling 2022 ss. 3-11

ISBN: 9788205542587

Copyright (C) 2022 Gyldendal

Kopiert med hjemmel i Kopinor-avtalen



Hva er sykdom?

Bjørn Hofmann

Dette kapitlet setter sykdomslære inn i en større sammenheng. Det drøfter hva sykdom er, og viser hvordan og hvorfor sykdomslæren er dominert av en naturvitenskapelig forståelse av sykdom. Kapitlet viser også hvordan andre aspekter ved sykdom er vesentlig for å hjelpe lidende mennesker. Like lite som vi kan forstå sykdom bare ved hjelp av sykdomslæren, kan vi forstå sykdom uten den.

Sykdomslære er læren om sykdom – men hva er sykdom? Det er et langt vanskeligere spørsmål enn en i utgangspunktet skulle tro. Slår vi opp i et leksikon, kan vi lese at sykdom er en «fellesbetegnelse på tilstander som kjennetegnes ved forstyrrelser i kroppens normale funksjoner» (snl.no), men samtidig står det at «begrepet er vanskelig å definere, og det har hatt en tendens til stadig å bli utvidet». Dette resulterer i opphetede debatter om hvorvidt ulike tilstander er sykdom, slik som amalgamforgiftning, el-allergi og myalgisk encefalopati (ME). Tilsvarende har vi sett heftige diskusjoner om hvorvidt medieoffersyndromet (MOS), kjønnsinkongruens, sjenanse, sorg og sjalusi er sykdom.

Når vi sliter med å avgjøre om ulike tilstander er sykdom eller ikke, skyldes det at sykdomsbegrepet ikke er entydig og klart. Det betyr ikke nødvendigvis at sykdomslæren er vag og uklar, men at det er viktig å se den i perspektiv. Sykdom avhenger av tid, sted og hvilket syn vi har på verden.

Sykdom i tiden

Malaria oppfattes i dag som en sykdom som skyldes parasitter som overføres fra mygg. Men navnet kommer fra *mala aria* (italiensk), som betyr dårlig luft. En trodde altså at malaria skyldtes luftforurensning. Tilsvarende tenker vi oss i dag at bakterien *Helicobacter pylori* er en viktig årsak til magesår, men for ikke lenge siden var magesår karakterisert av feil i syrebalansen i magen forbundet med stress.

På 1600-tallet var *nostalgia* (hjemlengsel) en sykdom, med symptomer som apati og feber, som hadde dårlig prognose, og som ikke sjelden fikk dødelig utgang. I 1850-årene ble det beskrevet en sykdom i de amerikanske sørstatene som kaltes *drapetomani*. Betegnelsen kommer av gresk og betyr «rømmesyke». Tilstanden forekom hos slaver og besto i at de hadde en umåtelig trang til å rømme. Behandlingen var hard og ikke reseptbelagt. I 1919 skrev den danske patologiprofessoren Carl Julius Salomonsen en avhandling om en smittsom sinnssykdom som bredte seg blant modernistiske kunstnere. Han hevdet at kubismens forvridde og (for ham) heslige bilder var uttrykk for en alvorlig sinnslidelse, som han ga navnet *dysmorfisme*.

Verken nostalgi, drapetomani eller dysmorfisme blir oppfattet som sykdommer i dag, men historiene om dem kan lære oss at sykdommer kommer og sykdommer går. I tillegg skifter sykdommenes forklaringer: I antikken trodde en at epilepsi skyldtes et angrep av demoner, derav navnet «den hellige sykdom». I dag oppfattes den som en kronisk tilstand med nevrologisk opphav. Det vil si at det ikke bare er klassifiseringen av sykdommer som endrer seg, men også sykdomslæren.

La oss se litt nærmere på en av vår tids store folkesykdommer, diabetes. Betegnelsen stammer fra jonisk gresk og betyr «å renne gjennom en sil». I antikken ble tilstanden kjennetegnet ved stor økning i urinmengden og ved at urinen hadde en søtlig smak. Sykdommen «smelter kjøtt og bein til urin», og pasientenes liv er «kort, avskyelig og smertefullt, og innen det er gått lenge, utånder de» (Aretaeus fra Kappadokia, 81–138). På 1100-tallet hadde en begynt å analysere innholdet i urinen for å stille diagnoser, og på 1500-tallet kokte en urin og separerte et hvitt pulver, som en mente var salt. Diabetes ble derfor oppfattet som en endring i blodet, der saltet forårsaket tørst i nyrene som førte til den overdrevne urinlatingen.

På 1600-tallet mente enkelte at diabetes skyldtes overdrevne mengder av mat, alkohol og sex, og at det derfor var en sykdom i magen. Samtidig fikk også den søtlige smaken på urinen fornyet oppmerksomhet, og tilstanden fikk navnet *diabetes mellitus* (*mellitus* betyr honningsøt). På slutten av 1700-tallet ble det utviklet en test for sykdommen der det ble brukt gjær for å påvise sukkerinnholdet i urinen.

På 1800-tallet skjedde det mye som endret forståelsen av sukkersyke. Diabetes ble periodevis oppfattet som en sykdom i nyrene, blæra, blodet, leveren og nervesystemet. Den berømte franske legen Claude Bernard (1813–1878) mente for eksempel at diabetes måtte være en sykdom i leveren, fordi det var der han fant at druesukkeret ble lagret i kroppen. Etter hvert ble det påvist at sukkersyke, eller det vi i dag kjenner som diabetes mellitus type 1, var en sykdom i bukspyttkjertelen (pankreas). En klarte blant annet å framkalle diabetes hos hunder ved å fjerne pankreas (i 1898). Framskritt innen laboratoriemedisinen gjorde at sykdommen ble lokalisert til langerhansøyene i pankreas, og i 1921 ble et av hormonene som styrer omdanningen av sukker i kroppen, funnet. Det fikk navnet insulin (*insula* betyr øy).

I dag oppfattes diabetes type 1 som en autoimmun destruksjon av de insulinproduserende cellene i langerhansøyene i pankreas. Årsakene til denne destruksjonen er ukjent, men det dreier seg trolig om en kombinasjon av genetiske faktorer og utløsende miljømessige forhold, som virusinfeksjon, belastninger i omgivelsene eller liknende.

Diabetes har altså vært oppfattet som en sykdom i blodet, blæra, nyrene, leveren og nervesystemet og som et utslag av livsførselen. Noe tilsvarende gjelder selvsagt mange andre sykdommer. Det betyr at sykdomslæren har endret seg opp gjennom historien. De forklaringene vi har på sykdom i dag, er derfor neppe de som vi kommer til å ha om femti år.

Fra antikken og fram til langt inn på 1800-tallet var sykdomslæren dominert av væskelæren (humoralpatologi), der sykdom var avvik fra en balanse (homøostase) i de fire kroppsvæskene (blod, slim og gul og svart galle). Kroppens balanse var påvirket av ytre faktorer, som klima (tørt, vått, varmt, kaldt), og dominans av enkelte kroppsvæsker ga ulike sykdomstegn og sinnsstemninger, som vist i tabell 1.1. Eksempler på ulike teorier om sykdom opp gjennom tidene er vist i tabell 1.2.

Etter middelalderen, da åndelige størrelser som synd og straff gjorde seg gjeldende i sykdomslæren, ble ulike naturvitenskapelige tilnærminger stadig mer framtredende. Diabeteseksemplet viser hvordan sykdomslæren endret seg over tid. Opp gjennom historien har sykdomslærens grunnelementer variert fra organer (Giovanni Battista Morgagni, 1682–1771), via vev (Xavier Bichat, 1771–1802) til celler (Rudolf Virchow, 1821–1902) og senere til cellenes bestanddeler, for eksempel DNA. Sykdomslæren har også satt søkelyset på eksterne faktorer som påvirker mennesket, slik som luftforurensning (miasmeteorien), bakterier og virus.

I nyere tid har sykdomslæren hovedsakelig fulgt den naturvitenskapelige utviklingen. Derfor vil også denne boka handle om naturvitenskapelige tilnærminger til sykdom. Det fins selvsagt andre og viktige forståelser av sykdom, som vi skal komme tilbake til nedenfor. Men først er det viktig å merke seg at heller ikke en naturvitenskapelig forståelse av sykdom er absolutt og entydig: Sykdom varierer ikke bare i tid, men også med sted.

Tabell 1.1 Sammenhengen mellom den antikke sykdomslærens grunnelementer og sinnsstemninger.

Kroppsvæske	Kvaliteter	Sinnsstemning
Blod	Varmt og vått	Sangvinsk: lystig og optimistisk
Slim	Kaldt og vått	Flegmatisk: rolig og fattet
Gul galle	Varmt og tørt	Kolerisk: oppfarende og hissig
Svart galle	Kaldt og tørt	Melankolsk: tungsindig og sturende

Sykdom og sted

Laktoseintoleranse er en tilstand som gir magesmerter og diaré etter måltider som inneholder melk eller melkeprodukter. Tilstanden skyldes mangel på enzymet laktase, som spalter melkesukker i tarmen. Laktoseintoleranse er lett å påvise ved klare symptomer dersom det inntas melkeprodukter. En kan også måle manglende laktase i tarmslimhinnen. 75 % av verdens befolkning har laktoseintoleranse, men svært få av disse har noen plager, ganske enkelt fordi de ikke drikker melk eller spiser melkeprodukter. Det betyr at den samme tilstanden oppfattes som plagsom og uønsket i områder der det drikkes melk, slik som i Vest-Europa og enkelte afrikanske land, mens den ikke oppfattes slik i resten av verden.

Noe tilsvarende gjelder for sigdcelleanemi. Dersom en er bærer av et gen for sigdcelleanemi (uten å ha sykdommen), endres hemoglobinet i blodet. Samtidig beskytter tilstanden mot malaria. Det betyr at å ha et gen for sigdcelleanemi kan oppfattes som en sykdom i områder uten malaria, men som en fordelaktig (beskyttende) tilstand i områder med malaria.

De samme fysiologiske, biokjemiske og molekylærbiologiske forholdene kan altså oppfattes som sykdom noen steder, men ikke andre steder. Selv med en naturvitenskapelig forståelse av sykdom må vi derfor ta hensyn til geografiske og kulturelle forhold.

Sykdom som avvik fra normal funksjon

Som vi så innledningsvis, defineres sykdom i leksikon som avvik fra normal funksjon. Slike standarddefinisjoner av sykdom har sin opprinnelse i antikke forestillinger om at det som er friskt, er velfungerende og i balanse

Tabell 1.2 Noen teorier om sykdom opp gjennom historien.

Teori	Tidsperiode	Sykdomsoppfatning	Referanse
Demonisk	Alle tider	Onde ånder	Antikkens Egypt
Vitalisme	2000 f.Kr.–	Vitalt prinsipp	Antikkens Egypt, Hellas
Humoralpatologi, væskelære	400 f.Kr.–1800 e.Kr.	Ubalanse i de fire kroppsvæsker	Hippokrates
Miasmeteori	500 f.Kr.–1880 e.Kr.	Forurensning gjennom luft	India, antikkens Hellas
Stimulus-respons-teori	1850–	Respons på stimuli	Florence Nightingale
Bakterieteori	1850–	Bakterier	Louis Pasteur, Robert Koch
Neohumoralisme (nyvæsketeori)	1620–	Anomalier i blodet	William Harvey
Endret anatomi	1750–	Endring i organer	Giovanni Morgagni
	1800–	Vevsforandringer	Xavier Bichat
	1850–	Celleforandringer	Rudolph Virchow
Genetikk	1980–	Genetiske årsaker	Frederick Sanger, Kary Banks Mullis
Risikomodell for sykdom	1950–	Økt risiko	Forebyggingsbevegelsen
Bio-psyko-sosial modell	1980–	Komplekst sykdomsbegrep	George L. Engel
Systembiologi	2000–	Interaksjon mellom gener, proteiner og miljø	Ludwig von Bertalanffy

(homøostase). Sykdom er avvik fra en normal balanse, og målet for behandlingen er å styrke organenes funksjon og gjenopprette balansen. Sykdomslæren beskriver avvik i kroppens bestanddeler og deres funksjoner og prosesser: Kroppstemperatur på over 37 °C er hypertermi, og kroppstemperatur under 37 °C er hypotermi.

Hva mener vi så med «normal funksjon»? Med «normal» menes det som er statistisk normalt, og med funksjon menes det som bidrar til overlevelse og reproduksjon. Eksemplet med laktoseintoleranse viser at begrepene ikke er så enkle som de kan virke ved første øyekast. Størstedelen av verdens befolkning har laktoseintoleranse, og det er unormalt å tåle laktose. Som melkedrikkende nordboere oppfatter vi oss likevel ikke som syke. Å tåle laktose reduserer ikke våre evner til overlevelse eller reproduksjon. Et annet eksempel er at toppidrettsutøvere presterer langt over det som er normalt, uten at vi vanligvis omtaler dem som syke.

Hva betyr det at sykdom defineres ut fra det som ikke bidrar til overlevelse og reproduksjon? La oss spørre mer konkret: Hva betyr det at hjertets funksjon er å pumpe blod? Ett svar kan være at hjertet pumper blod for å transportere oksygen fra lungene ut til vevet og karbondioksid tilbake. Mennesker som har oksygen- og karbondioksidtransport utenfor det som er statistisk «normalt», vil likevel kunne oppfatte seg som friske. Vi må derfor fortsette å spørre: Hva er formålet med å opprettholde oksygen- og karbondioksidtransporten? Standardsvaret (i dag) er at formålet er å opprettholde forbrenningen. Igjen kan vi spørre: Hva er hensikten med å opprettholde forbrenningen? Slik kan vi fortsette spørsmålsstillingen helt til vi kommer til spørsmålet «Hva er menneskets funksjon?».

Naturvitenskapens svar på dette spørsmålet er overlevelse og reproduksjon. Med andre ord: *Sykdom er en intern tilstand i en organisme som hindrer organismen i å utøve en naturlig funksjon, det vil si det som er et artstypisk bidrag til organismens overlevelse og reproduksjon*. Men heller ikke denne definisjonen er uproblematisk, for den kan gjøre homoseksualitet til sykdom, i den grad homoseksualitet ikke bidrar til reproduksjon.

Sykdomslæren baserer seg altså på en naturvitenskapelig oppfatning av sykdom som avvik fra normal funksjon. En slik definisjon av sykdom kan være problematisk og må ses i perspektiv.



ARNE LARSEN er forskalingssnekker og 47 år. Han oppsøker lege fordi han har store smerter i nakken og ryggen. Legen undersøker ham grundig og tar ulike tester og prøver, men hun kan ikke finne noe galt. Arne er bekymret for at det kan være kreft, og legen sender ham til røntgenundersøkelse for sikkerhets skyld, men resultatet avkrefter at Arne har kreft. Legen sier: «Jeg har undersøkt deg grundig, og du har ikke kreft.» Arne mener at det gjør så vondt at han ikke kan arbeide. Legen har ikke funnet noen objektive tegn på sykdom, men hun ser at Arne har smerter, og sykmelder ham i en uke.

Sykdommer i perspektiv

På 1800-tallet døde mange kvinner i barselfeber. Den ungarsk-østerrikske legen Ignaz Semmelweis (1818–1865) hevdet i 1840-årene at barselfeber skyldtes smitte som legene førte med seg fra obduksjonssalen. For dette ble han ledd av og uthengt. Selv da han gjorde eksperimenter som viste at dødeligheten gikk kraftig ned når personalet vasket hendene, ble han ikke trodd. Vi må huske at dette var før bakteriologiens innmarsj med Louis Pasteur og Robert Koch. Teorien til Semmelweis om en usynlig «likgift» som ble ført fra obduksjonssalen og opp på fødestuene, virket som ren galskap på samtiden. Semmelweis var så rystet av motstanden mot hans nye viten at han skrev arge brev til kolleger, der han beskyldte dem for å være mordere som ikke vasket seg. Han ble oppfattet som gal og lagt inn i sinnssykeasyl, der han døde i 1865.

Fortellingen om Semmelweis viser oss med all tydelighet hvor standhaftige våre oppfatninger og teorier om sykdom er. Datidens sykdomslære var ikke klar for «likgift». Samtidig lærer den oss at våre (naturvitenskapelige) oppfatninger av sykdom endrer seg over tid, og at hvert perspektiv har sine begrensninger. Også i dag kan vi finne teorier om «likgift» som vi vil mene er den rene galskap, men som i framtiden kanskje vil være pensum i sykdomslære. Ett eksempel er da de australske forskerne Barry Marshall og Robin Warren mente at magesår skyldes en bakterie (*Helicobacter pylori*) i 1984. De ble utledd. Men i 2005 fikk de Nobelprisen i fysiologi for oppdagelsen.

I tillegg til at vi kan ha ulike naturvitenskapelige perspektiver på sykdom, fins det andre og viktige perspektiver som sykdomslæren ikke fanger opp. Eksemplet om Arne Larsen kan illustrere dette.

I eksemplet finner legen ingen konkret sykdom hos Arne, som ikke spiller syk for å få en sykmelding. Sykdomslæren dekker derfor ikke Arnes tilfelle. Nye framskritt innen bildediagnostikk eller laboratoriemedisin kan kanskje finne forklaringen en gang i framtiden, men dagens sykdomslære kommer her til kort. Men selv om Arne ikke har en sykdom som lar seg identifisere, føler han seg åpenbart syk. Slik er det for mange mennesker: De føler seg syke, men vi finner ingen sykdom. Betyr det at de ikke er syke? Hvis vi mener at de ikke er syke, hevder vi samtidig at folk med aids ikke var syke før hiv-viruset ble oppdaget. Det er altså en viktig forskjell på det å ha en sykdomsopplevelse og det å ha en sykdom. På engelsk skilles det gjerne mellom sykdomsopplevelse, *illness*, som beskriver den enkeltes opplevelse av sykdom, og *disease*, som er sykdommen slik den oppdages og observeres av helsepersonell.

I tillegg ser vi at legen sykmelder Arne fordi hun mener at han ikke kan arbeide i den tilstanden han er i. Det betyr at Arne er syk for sin arbeidsgiver og for sine kolleger. Han endrer samfunnsrolle fra å være arbeidsaktiv til å innta en *sykerolle*. På engelsk kalles det samfunnsmessige perspektivet på sykdom *sickness*. Arne

er altså syk fra et personlig perspektiv (illness) og i et samfunnsperspektiv (sickness), men han er ikke syk i et medisinsk perspektiv (disease) (figur 1.1 a).

Denne boka handler om sykdom i et medisinsk perspektiv (disease). Sykdomslære er her læren om sykdommene i naturvitenskapelig forstand. Det betyr ikke at de andre perspektivene ikke er viktige for å forstå sykdom. Andre fag, som samfunnsmedisin, epidemiologi, antropologi og sosiologi, undersøker samfunnsmessige perspektiver på sykdom (sickness), mens sykepleie, psykologi, filosofi og skjønnlitteratur er opptatt av den personlige opplevelsen av å være syk (illness).

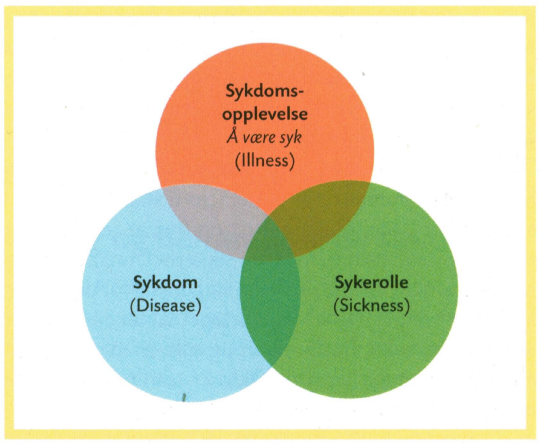
Dessuten er perspektivene ikke upåvirket av hverandre. Tilstander som tidligere bare ble oppfattet som sykdomsopplevelse, *illness*, slik som fibromyalgi, har etter mye forskning blitt akseptert som *disease*. Motsvarende har tilstander som har vært oppfattet som *sickness*, slik som hyperaktivitet, blitt oppfattet som *disease* (ADHD). Og enkelte tilstander som har blitt oppfattet som *sickness* og *disease*, slik som homoseksualitet og drapetomani, blir ikke lenger oppfattet som sykdom fra noen av perspektivene.

Denne boka handler altså bare om ett av flere viktige perspektiver på sykdom, *disease*. Sykdomslæren ser sykdom i lys av moderne naturvitenskap. Men til de ulike perspektivene svarer også ulike sykdomsmodeller.

Sykdomsmodeller

Avhengig av hvilke faktorer eller perspektiver vi vektlegger, benyttes ulike sykdomsmodeller. Sykdomslæren bygger på en biomedisinsk sykdomsmodell, som legger vekt på biologiske, fysiologiske, molekylærbiologiske og miljømessige forhold. En psykologisk sykdomsmodell vektlegger hvordan mentale faktorer virker inn på sykdom, mens en sosial sykdomsmodell vektlegger samfunnsmessige faktorer.

Men sosiale, mentale og biologiske forhold virker ikke isolert.

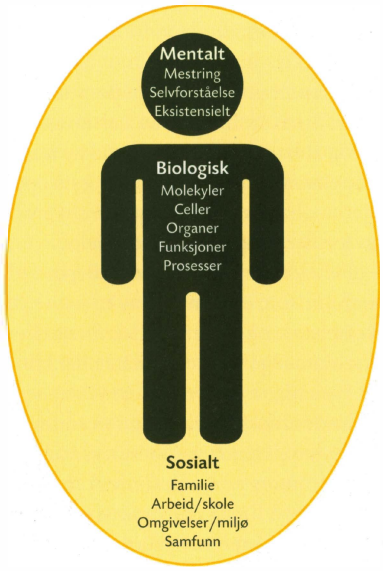


Figur 1.1a Ulike perspektiver på sykdom

EVA BERG er en yrkesaktiv kvinne på 41 år. Hun er samboer med Hallvard (42), og paret har tre barn (2, 5 og 8 år). Eva og Hallvard har gode og utfordrende jobber. I det siste har Eva følt at hun ikke klarer å komme igjennom alle oppgavene hun har blitt pålagt og pålegger seg selv. Paret har nettopp kjøpt hus og har en stram, men ryddig økonomi. Eva er svært pliktoppfyllende både på arbeidet og hjemme. Hun påtar seg omsorg for hus og hjem, med matlaging og stell av barn, og oppfølging med kjøring til ulike aktiviteter. De har stor hjelp av Hallvards mor, som bor like i nærheten, men Eva føler også at svigermoren har sine krav og oppfatninger. Eva har oppgaver fra tidlig morgen til sent på kveld alle dager i uka, og det blir lite tid til mosjon, til å slappe av eller til bare å gjøre det hun har lyst til. I den siste tiden har hun hatt vekselvis løs og fast mage. Hun går til legen, som tar en rekke prøver.

Legen konstaterer irritabel tarm-syndrom hos Eva. Tilstanden skyldes tilsynelatende både sosiale og mentale forhold. Mange bruker derfor den biopsykososiale sykdomsmodellen (figur 1.1 b), som legger vekt på samspillet mellom biologiske, mentale og sosiale faktorer.

Når sykdomslæren begrenser seg til den biomedisinske sykdomsmodellen, utelates viktige aspekter ved sykdommer og deres årsaker. Sykdom omfatter langt mer enn bare biologiske og fysiologiske forhold. Sykdom er også et eksistensielt, sosialt og kulturelt fenomen. Selv om disse aspektene blir ivaretatt av andre fag, er det viktig å ha dem i minnet hele tiden. Sykdomslære, slik den blir framstilt i denne boka, dreier seg om de organer, funksjoner og prosesser som lar seg beskrive med naturvitenskapelige metoder. Det er viktig å huske at disse organene, funksjonene og prosessene er påvirket av mentale, sosiale og eksistensielle eller åndelige forhold. Sykdomslæren belyser derfor bare ett av flere viktige aspekter i forståelsen av sykdom.



Figur 1.1b Den biopsykososiale sykdomsmodellen

Helsefag uten sykdomslære?

Noen vil si at de ikke trenger sykdomslære, fordi de er opptatt av den enkeltes opplevelse av sykdom (illness) og av hele mennesket. Men selv om det er gode grunner til å hevde at mennesket er langt mer enn summen av sine organer, funksjoner og prosesser, og at sykdomsopplevelsen er sentral, er det vanskelig å avsondre seg fra sykdomslæren. Like lite som vi kan si at sykdomslære er alt vi trenger for å håndtere menneskers sykdom, kan vi si at vi kan klare oss uten den. I dag er det ikke mulig å forstå hva sykdom er, uten å kjenne de naturvitenskapelige aspektene. En kan ikke si at «jeg er opptatt av å pleie den syke og trenger derfor ikke å vite så mye om sykdomslære». God pleie forutsetter også at en forstår det biologiske og fysiologiske grunnlaget for pasientens sykdom.

KURT RUDOLFSEN, 77 år, er innlagt med hjertesvikt som innleggelsesdiagnose. Han har i tillegg noe nedsatt nyrefunksjon og har den siste tiden blitt tiltakende slapp og tungpustet. Til tider er han uklar, han er hoven i beina og klager over smerter øverst i mageregionen. Blodtrykket er 170/100 mm Hg, pulsfrekvensen 90 per minutt og temperaturen 36,8 °C. Nattevakten forteller at Rudolfsen har sovet lite i løpet av natten, hun har flere ganger funnet ham oppe inne på rommet og vurdert ham som uklar. Selv sier han at han må på do. Rudolfsen var i ferd med å gå ut av senga, han er blek, kaldsvett og «plukkete» og kaver en del med sengetøy og morgenkåpe og sier han må på do. Da det er tid for mat, vil Rudolfsen ha «skikkelig kost». Da sykepleieren kommer inn på rommet etter måltidet, er Rudolfsen på nytt på vei ut av senga. Han er nå helt våt og kald av svette, han har surklende respirasjon og svarer ikke på spørsmål, men sier han igjen må på do. Han er cyanotisk på leppene og marmorert på beina. Pulsfrekvensen er nå 120 per minutt.

Dersom en i denne situasjonen hjelper Rudolfsen på do og til å spise, uten å innse at han har lungeødem og er hypoksisk, og at han er i ferd med å dø av hjertesvikt på grunn av de ekstra belastningene med å stå opp og spise, er en neppe en god sykepleier. Å kunne sin sykdomslære blir her en forutsetning for å kunne hjelpe.

I tillegg er sykdomslære en innfallsport til å forstå dagens samfunn, som er dominert av en naturvitenskapelig, reduksjonistisk og materialistisk forståelse av mennesket. Den biomedisinske tenkemåten har blitt en del av vår kultur. Pasienter forklarer sin sykdom med celleforandringer, hormonforstyrrelser og serotoninnivåer. Det er få ting som har påvirket vår kultur, vår tenkemåte, vår forståelseshorisont og vårt samfunn så mye som naturvitenskapene. For å forstå det moderne mennesket må en forstå naturvitenskapene. Sykdomslære er derfor ikke bare læren om sykdommenes biomedisinske grunnlag, men også en kulturlære.

Sykdomsforløp

Sykdom forløper i flere faser, hvor hver fase påvirker pasienten på forskjellig måte. Når en forebygger og behandler sykdom, er det viktig å ta hensyn til hele sykdomsforløpet og pasientens evne, både fysisk og psykisk, til å håndtere sykdommen (figur 1.2).

Mange sykdommer har en fase før de oppdages. Eksempelvis starter sykdomsprosessen ofte før personen blir oppmerksom på den og tar kontakt med helsetjenesten, eller før sykdomsprosessen oppdages av helsepersonell. Før sykdommen bryter ut, kan enkelte også fornemme tidlige tegn eller symptomer som kommer før typiske manifestasjoner av sykdom. Slike tegn kalles gjerne prodromer.

Ved en del sykdommer kan sykdomsutviklingen hindres eller forsinkes gjennom forebyggende tiltak der sykdomsfremmende faktorer reduseres (f.eks. røyking, fysisk inaktivitet og feilernæring). Samfunnet kan også bidra til å redusere sykdomsutløsende faktorer (f.eks. helseskadelige kjemikalier og stråling) gjennom lover og forskrifter samt andre offentlige tiltak.

Sykdom kan forebygges igjennom reduksjon av risikofaktorer og ved å oppdage og behandle forstadier. Noen risikofaktorer og forstadier kan oppdages igjennom generelle helsekontroller og masseundersøkelser. Sjansene for sykdom øker med økt forekomst av risikofaktorer, inklusive sykdom blant nære familiemedlemmer. De siste årene har genetisk kartlegging og veiledning til å identifisere personer med behov for tidlig forebyggende behandling og mer persontilpasset behandling blitt stadig mer brukt.

Når personen blir oppmerksom på sykdom (sykdomsopplevelse) og tar kontakt med helsetjenesten, kan helsepersonell stille en diagnose, igangsette målrettet behandling, helbrede eller begrense og forutsi sykdomsforløpet (prognose). Å få en diagnose gir en forklaring på situasjonen og peker på et behandlings- og sykdomsforløp; for mange pasienter er det godt å få denne oppklaringen. Men diagnosen i seg selv kan også oppleves belastende for både pasient og pårørende, i form av frykt og angst, men også stigmatisering og diskriminering. Helsepersonell har en svært viktig oppgave med å formidle diagnosen, behandlingsstrategiene og sykdomsforløpet på en forståelig måte for pasienten og pårørende.

Valg av behandling av en sykdom vil være avhengig både av sykdommen og av pasientens tilstand. I en akuttsituasjon er behandlingen først og fremst rettet mot å ivareta pasientens umiddelbare sikkerhet, men samtidig må en ta stilling til hva som er mest gunstig for det videre sykdomsforløpet. Det videre forløpet vil ikke bare være avhengig av selve sykdommen og pasienten, men også andre sykdommer og risikofaktorer som pasienten har eller utvikler i løpet av sykdomsforløpet.

Den videre behandlingen må ta hensyn til den samlede summen av gevinster og ulemper for pasienten. Jo

mer en vet om sykdommen, behandlingsalternativene og pasientens grunnleggende tilstand, desto bedre kan en tilrettelegge for et optimalt behandlingsforløp. I mange tilfeller vil pasienten bli helt frisk enten med eller uten behandling. I andre tilfeller vil sykdommen og de komplikasjonene som sykdommen skaper, påvirke det videre sykdomsforløpet. Noen ganger blir pasienten aldri frisk og må fortsette med behandling resten av livet (kronisk sykdom).

Noen sykdommer vil være så belastende at pasienten vil trenge tid før kroppen klarer å komme tilbake til et normalnivå, selv om pasienten blir frisk av sykdommen. I disse tilfellene vil rehabilitering være viktig for pasientens videre livskvalitet. Uansett hvor en befinner seg i behandlingsforløpet, er det viktig at pasientens perspektiv ivaretas, og at en forsøker å gi pasienten den innsikten som er nødvendig for det best mulige sykdomsforløpet.

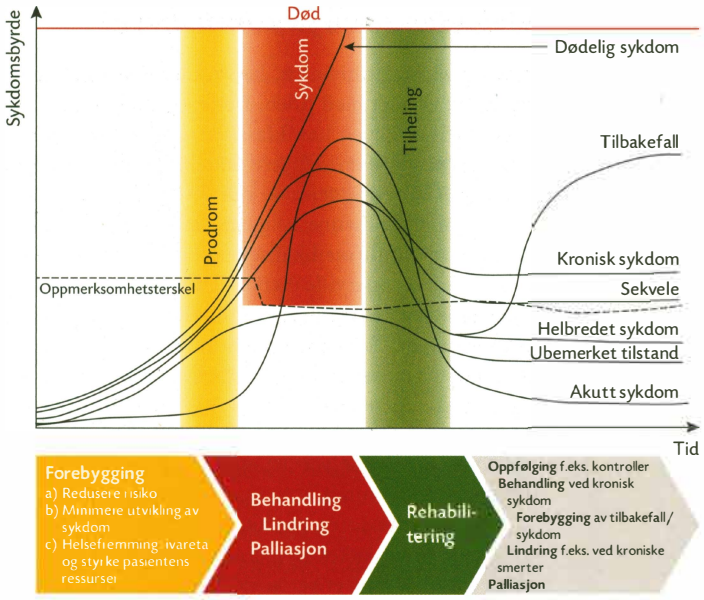
Klassifisering av sykdom

Tallet på sykdommer har endret seg drastisk opp gjennom historien. Et oversiktsverk fra 1763 beskrev 2400 sykdommer med stort og smått. I dag har vi omfattende systemer for klassifisering av sykdom som rommer flere ganger så mange. Bare verdens helseorganisasjons ICD-11 (WHO, International Classification of Disease and Health Related Problems), som er rettet mot spesialisthelsetjenesten, har over 55 000 sykdomskoder. I tillegg har vi ICPC-2 (International Classification of Primary Care) utgitt av World Organization of Family Doctors (WONCA), DSM-V (Diagnostic and Statistical Manual of Mental Disorders) utgitt av American Psychiatric Association og ICF (International Classification of Functioning, Disability and Health), som også er utgitt av WHO. I tillegg har sykepleie sitt eget kodeverk for sykepleiediagnoser, International Classification for Nursing Practice (ICNP). Dette er et verktøy for dokumentasjon av sykepleiediagnoser, -problemer, -tiltak og -mål i ett og samme system. Det er dekkende for alle deler av sykepleiepraksis, på tvers av spesialiteter, språk, land og kulturer.

Klassifiseringssystemene er pragmatiske kataloger som skal hjelpe oss til å holde orden på sykdommene. De er ofte bygd opp rundt ulike organgrupper (f.eks. sirkulasjonssystemet), organer (øyet), funksjoner (fordøyelse), prosesser (infeksjoner) eller tilstander (graviditet). Slik svarer de godt til sykdomslærens innhold.

Måling og behandling av sykdom

Vår naturvitenskapelige oppfatning av sykdom er også preget av et sterkt krav om objektivitet og etterprøvbarhet. Sykdom skal kunne måles uavhengig av den sykes opplevelser, og også fritt for fagpersonens feiltolkninger. Mennesker opplever smerter og symptomer svært forskjellig, og helsepersonell tolker dem på ulike måter. Sykdomslæren søker bak den enkelte pasients opplevelse og bak den enkelte pleiers oppfatning av sykdommen.



Figur 1.2 Sykdomsforløp med eksempler: alvorlig sykdom som fører til død, akutt sykdom, sykdom som behandles og helbredes, kronisk sykdom, sykdom med tilbakefall og sykdom med et forløp som ikke er merkbart for personen. Hvert forløp oppleves ulikt og stiller ulike krav til helsepersonell. Alminnelige oppgaver for helsepersonell i de ulike fasene er angitt nederst i figuren (Bjørn Hofmann og Stein Ørn 2022).

Kravet om objektiv målbarhet gjør seg gjeldende i både diagnostikk, palliasjon og behandling. En krever at behandlingen er effektiv, uavhengig av både pasientens og behandlerens oppfatninger og preferanser. Det oppnås blant annet gjennom strengt fastsatte rutiner for utprøving av legemidler. Da ekskluderes (eller kontrolleres det for) en av de viktigste effektene i helsefagene, nemlig den enkeltes forventning – placeboeffekten.

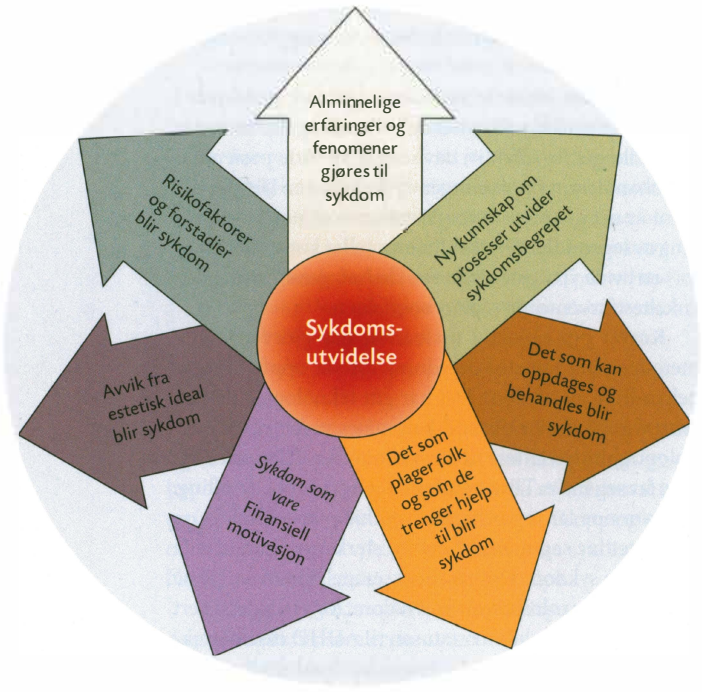
Kravet om objektiv målbarhet har gjort at målemetodene har fått stor betydning for vår oppfatning og behandling av sykdom. Det har gitt teknologien en helt sentral rolle: At en tilstand kan måles ved hjelp av teknologi, gjør den lettere til en sykdom enn tilstander som ikke lar seg måle. Tilsvarende er det også for behandling: Dersom noe lar seg behandle ved hjelp av teknologi og resultatet lar seg måle, er det en sterk kandidat til å få status som sykdom. Det var for eksempel få som tenkte på at ufrivillig barnløshet var en sykdom, før vi fikk assistert befruktning, og sykdomsstatusen til ADHD økte betraktelig da en kunne behandle tilstanden med ritalin.

Hva vi oppfatter som sykdom og ikke som sykdom, påvirkes med andre ord av hva vi kan gjøre, og hva vi kan måle objektivt (med teknologi). Sykdomslæren utvikles derfor i takt med den teknologiske utviklingen. På samme måte som molekylærbiologi og genetikk har bidratt med mye ny kunnskap om sykdommer i de senere årene, vil kanskje maskinlæring eller andre teknologier endre vår kunnskap om sykdom i framtiden.

Utvidelse av sykdomsbegrepet

Som vi så innledningsvis, har sykdomsbegrepet blitt vesentlig utvidet med tiden. Dette er ofte av det gode, eksempelvis når en kan klassifisere og behandle nye sykdommer som en tidligere ikke kunne forklare eller behandle. På den annen side kan det være problematisk der alminnelige menneskelige tilstander, som sorg, blir gjort til sykdom (medikalisering), eller der tilstander som personen ikke ville merket fordi de ville dødd med dem og ikke av dem, blir gjort til sykdom og behandlet (overdiagnostikk og overbehandling). Slike utfordringer oppstår når sykdomsbegrepet utvides ut over det som gagner personen. Figur 1.3 gir en oversikt over ulike former for sykdomsutvidelse. Med utvidelse av sykdomsbegrepet følger det også en ekspansjon av hva sykdomslæren omfatter.

Når risikofaktorer, prediktorer eller forstadier blir gjort til sykdom, kalles det gjerne en begrepsmessig utvidelse av sykdom. Det ser vi eksempelvis ved tilstander som får forstavelsen «pre», som ved pre-diabetes og predemens, men også når risikofaktorer blir sykdom, som ved hypertensjon.



Figur 1.3 Ulike måter som sykdomsbegrepet har blitt utvidet på.

Når ny kunnskap utvider sykdomsbegrepet, kalles dette gjerne for kunnskapsmessig (epistemisk) utvidelse. Det er av det gode når vi får ny kunnskap som gjør oss i stand til å hjelpe på nye og bedre måter, men er av det onde når det resulterer i overdiagnostikk og overbehandling.

Pragmatisk utvidelse skjer når noe gjøres til sykdom blant annet fordi det kan behandles. Kjønnsinkongruens kan kanskje være et eksempel på det.

Etisk utvidelse er når det som plager folk, blir gjort til sykdom, som ADHD, mens estetisk utvidelse er når eksempelvis utstående ører behandles som sykdom. Sosial (økonomisk) utvidelse skjer når noe gjøres til sykdom fordi noen kan profittere på det, eksempelvis ved salg av medikamenter, som viagra og testosteronbehandling.

Å gjøre alminnelige atferdsformer eller livserfaringer, som sjenerthet og sorg, til sykdom kalles gjerne medikalisering og representerer en vesentlig utvidelse av sykdomsbegrepet.

Sykdommens årsaker

Sykdommer klassifiseres og forstås ut fra de bakenforliggende årsakene. Nå kan det være mange årsaker til at mennesker blir syke. Eksemplet med Eva på side 7 viser at sosiale faktorer er viktige. Tilsvarende kan både åndelige (eksistensielle) og kulturelle aspekter gjøre seg gjeldende. Sykdomslæren legger vekt på årsaker, prosesser og mekanismer i et naturvitenskapelig perspektiv. Det gjøres blant annet ut fra kravet om objektiv målbarhet og etterprøvbarhet. Vi søker de naturvitenskapelige årsakene til sykdom for å finne effektive måter å gripe inn og hjelpe mennesker på.

Samtidig må vi ikke glemme at sykdomslæren dermed avgrenser seg fra andre og viktige årsaksfaktorer til sykdom. Noen ganger skyldes de målbare årsakene og sykdomsprosessene helt andre forhold (sosiale, kulturelle eller eksistensielle/åndelige). Dessuten har de fleste tilfeller av sykdom en lang rekke mer eller mindre sammenvevde årsaker. Tilfellene der én årsak (bakterie) gir én sykdom (tuberkulose), hører med til unntakene. I dag ser vi at genetiske og epigenetiske faktorer spiller inn sammen med livsstil og andre miljøfaktorer. Nyere forskning viser også at menneskers opplevelser og livserfaring ikke bare gjenspeiles i psyken, men også i nervesystemet og immunsystemet og i cellene. Slik viten visker ut det tradisjonelt skarpe skillet mellom naturvitenskapelige og humanvitenskapelige oppfatninger av sykdom. Dersom vi ser oss blinde på sykdommens naturvitenskapelige årsaksforhold, kan vi ende med å drive symptombehandling og ikke sykdomsbehandling.

KAPITTEL 1 Oppsummering av kapitlet, www.37grader.no

KAPITTEL 1 LITTERATUR OG NETTADRESSER

Referanser

Album D. Sykdommers og medisinske spesialiteters prestisje. Tidsskr Nor Lægeforen 1991; 111(17): 2127–33.

Brülde B, Tengland P-A. Hälsa och sjukdom – en begreppslig utredning. Lund: Studentlitteratur, 2003.

Campbell EJ, Scadding JG, Roberts RS. The concept of disease. BMJ 1979; 2(6193): 757–62.

Helman CG. Culture, health and illness. London: Hodder Arnold, 2007.

Hofman, B. Perspectives on human malady: Disease, illness and sickness. The Routledge Companion to Philosophy of Medicine. London: Routledge, 2016.

Hofmann B. Acknowledging and addressing the many ethical aspects of disease. Patient education and counseling 2021. doi.org/10.1016/j.pec.2021.09.015

Hofmann B. Suffering: Harm to Bodies, Minds, and Persons. I: Thomas Schramme og Steven Edwards, red. Handbook of the Philosophy of Medicine. Dordrecht: Springer, 2017.

Hofmann B. Trenger helsevesenet begrepet sykdom? Tidsskr Nor Lægeforen 2003; 123(4): 487–90.

Hofmann B. Simplified models of the relationship between health and disease. Theor Med Bioeth 2005; 26(5): 355–77.

Hofmann B. Hva er sykdom? Oslo: Gyldendal Akademisk, 2008/2014.

Hofmann B. Teknologi skaper sykdom: Om teknologi, sykdom og verdier. Tidsskrift for forskning i sygdom og samfund 2009; 11: 5–12.

Hofmann B. Disease. I: ten Have, H. Encyclopedia of Global Bioethics. Dordrecht: Springer, 2015.

Hofmann B. Expanding disease and undermining the ethos of medicine. Eur J Epidemiol 2019; 34(7): 613–9.

Humber JM, Almeder RF, red. What is disease? Totowa, New Jersey: Humana Press, 1997.

Kim K. The Social Construction of Disease: From Scrapie to Prion. London: Routledge, 2007.

Lea M, Hofmann BM. Dediagnosing – a novel framework for making people less ill. European Journal of Internal Medicine 2021. doi.org/10.1016/j.ejim.2021.07.011

Smith R. In search of «non-disease». BMJ 2002; 324(7342): 883–5.

Svenaeus F. Sjukdomens mening: Det medicinska mötets fenomenologi och hermeneutik. Stockholm: Natur och Kultur, 2003.

Søvik O. Sykdomsbegrepet i historisk og dagsaktuell sammenheng. Tidsskr Nor Lægeforen 2001; 121(22): 2624–7.

Nettsteder (www)

globalfamilydoctor.com/groups/workingparties/wicc.aspx (International Classification of Primary Care (ICPC-2) utgitt av World Organization of Family Doctors (WONCA)).

icn.ch/what-we-do/projects/ehealth-icnptm/about-icnp (WHO, International Classification for Nursing Practice (ICNP)).

psychiatry.org/psychiatrists/practice/dsm (American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders (DSM-5)).

snl.no/sykdom (Store norske leksikon).

who.int (WHO, International Classification of Disease and Health Related Problems (ICD-11)).

who.int (World Health Organization (WHO), International Classification of Functioning, Disability and Health (ICF)).

Til fordypning

Aronowitz RA. When do symptoms become a disease? Ann Intern Med 2001; 134(9): 803–8.

Boorse C. On the Distinction Between Disease and Illness. Philosophy and Public Affairs 1975; 5: 49–68.

Canguilhem G. The Normal and the Pathological. New York: Zone Books, 1991.

Caplan AL, Engelhardt HT, McCartney JJ, red. Concepts of Health and Disease: Interdisciplinary perspectives. Reading, Massachusetts: Addison-Wesley, 1981.

Garson J. What Biological Functions Are and Why They Matter. Cambridge: Cambridge University Press, 2019.

Hofmann B. Complexity of the concept of disease as shown through rival theoretical frameworks. Theoretical Medicine and Bioethics 2001; 22(3): 211–37.

Reznek L. The nature of disease. New York: Routledge & Kegan Paul, 1987.

Schwartz, PH. Risk and disease. Perspect Biol Med 2008; 51(3): 320–34.

Sontag S. Illness as Metaphor. New York: Farrar, Straus and Giroux, 1978.

Lupton, Deborah

"Representations of Medicine, Illness and the Body", i (Red) Illness, Disease and the Body

Medicine as Culture 2012 ss. 2-33

Copyright (C) 2012 Medicine as Culture

Kopiert med hjemmel i Kopinor-avtalen

© Deborah Lupton 2012

Representations of Medicine, Illness and the Body

Introduction

Mass-media texts such as novels, television programmes, news-media accounts and artistic works are important forums for the portrayal of medicine, illness and disease. By their use of language or visual imagery they make an important contribution to the construction of people's ideas and beliefs about illness, disease and doctors. So too, texts directed at the medical profession, such as their professional journals, medical textbooks and magazines, contribute to and reproduce doctors' views of their patients. This chapter focuses on the dominant ways in which medicine, disease, death and illness are portrayed in such texts. The discussion begins by looking at the representation of disease, death and medical practice in literature, followed by an account of the ways in which popular culture such as television drama and documentary and news articles portray these phenomena. The chapter also looks at the use of metaphor as an important linguistic device for conveying meaning about medicine, health and illness, with detailed discussion of the metaphor of the machinery of the body, the military metaphor, metaphors of the immune system, cancer metaphors and metaphors related to gender. The chapter ends with a description of dominant visual images that have appeared in popular and medical texts, and of artists' portrayals of the experiences of illness, disease and death.

Disease, Death and Medical Practice in Literature

Writers who have suffered serious illnesses often incorporate their experiences into their work, using dramatic imagery in their attempts to convey the meaning of the illness as they experienced it. Such accounts of the effects of illness and disease upon the body, both fictional and autobiographical, provide insights into the ways in which illness is experienced which other, more prosaic, forms of expression sometimes fail adequately to achieve. Indeed, it has been suggested that medical students and practising physicians should be encouraged to read literary accounts of illness, pain and death so that they learn to treat their patients with empathy and compassion and become more sensitive to the human dimension of the illness experience and the complex moral implications of medical practice (Pellegrino, 1980).

The terrors of physical decay, pain, suffering and death are the very stuff of drama. As a result, descriptions of ageing, illness and disease have received the attention of novelists, poets and playwrights for centuries, especially leprosy, the plague, tuberculosis, cancer, syphilis and mental disorders (Horst and Daemmrich, 1987). Indeed, several renowned literary figures either were doctors themselves (for example, Rabelais, Chekhov, Maugham, Keats, William Carlos Williams) or were the sons of doctors (including Proust, Flaubert and Hemingway) (Peschel, 1980). Even in the 19th century, as epidemic disease was in regression in Europe, paradoxically the sick body was expressed more and more frequently in literary works (Herzlich and Pierret, 1987). Disease used as a recurring motif or theme in literature has focused attention 'on the intractable, capricious, and mysterious forces beyond human control or comprehension' (Horst and Daemmrich, 1987: 90).

Disease and death have traditionally been used as devices in novels to reveal the psychological development of characters and to structure events (Meyers, 1985). The functions of disease in literary texts have included:

- **illness as a sign of divine power or providence (common in biblical writings and the** *Iliad);*
- **illness or epidemics as the test of the moral fibre of the afflicted individual and society, exposing their true nature (for example, Camus'** *The Plague,* **1947);**
- **disease as a recurrent metaphor for moral or social decay (Ibsen's** *Ghosts,* **1881);**
- **as a vision of collective social disaster;**
- **as a sign of the individual's inability to escape from a destined fate;**
- **as a catalyst for artistic or intellectual genius and as a sign of emotional, intellectual or moral curiosity or superiority (Beecher Stowe's** *Uncle Tom's Cabin,* **1852 and O'Neill's** *Long Day's Journey into Night,* **1956);**
- **as a means of redemption for the fallen or the outcast (Dumas'** *Lady of the Camellias,* **1848);**
- **as a way of heightening the awareness of death, calling up questions of mortality and life's complexity (Joyce's** *The Dead,* **1914);**
- **and as an alien, incomprehensible force penetrating human life and destroying it (Solzhenitsyn's** *Cancer Ward,* **1968) (Horst and Daemmrich, 1987).**

In novels, '[d]isease has always been a great mystery: a visitation, a curse, a judgement' (Meyers, 1985: 2). For example, in Oscar Wilde's short novel *The Picture of Dorian Gray,* first published in 1891, the decadence of the protagonist is reflected in his painted visage, which, as the years go by, becomes more and more frightful in its external signs of old age, decay and corruption while the man himself remains youthful and handsome. The Gothic novel emerging in the Romantic era (spanning the late 18th century to the 1830s) portrayed the fearsomeness of nature wrought upon humanity, depicting diseases and the bodily symptoms of illness or

physical decay as evil demons and monsters to evoke horror and to symbolize the internal sufferings or evil of characters. The Victorian period of literature following was obsessed with death and the supernatural. One example is the writings of Edgar Allan Poe, which centred on the horror of tortures and the physical wages of sin, including his short story 'Masque of the Red Death' (1840) in which the protagonists fall prey to a plague-like illness after daring to trivialize and mock death. Mary Shelley's *Frankenstein* (1818) dealt with the moral issue of humans usurping God as creator of the human body, while Bram Stoker's *Dracula* (1897) focused on themes to do with the sexual body, gender, death and immortality with a particular emphasis on the symbolic nature of blood. It is telling that both these latter literary works remain central in contemporary popular culture as horror stories, demonstrating the resonance of their themes for audiences even today. The success of popular vampire fiction and related dramatized versions, such as the *Twilight* and *True Blood* series of books and films demonstrates this contemporary fascination with blood, death and the erotic.

Representations of illness and disease in literature have closely followed changes in medical discourses on the body. During the Romantic period the creativity of the writer or artist was believed to emerge from suffering due to travails such as serious illness. More than any other illness, tuberculosis, a common cause of death in the 19th century, was believed to stimulate the creative impulse, fuelled by the early deaths from that disease of literary figures such as Chatterton, Keats, Percy Shelley and Byron and the success of the romantic and tragic operas *La Traviata* and *La Bohème* (Meyers, 1985). With the advent of scientific medicine, naturalist writers in Europe began to incorporate clinical descriptions of illnesses and neuroses into their novels (Herzlich and Pierret, 1987). By the 20th century, novels were beginning to include clinical discussions of illness and disease to represent not only dissolution, decay and physical corruption but metaphysical despair, loneliness, alienation and self-doubt. In such writing, the medical treatment foisted upon the patient becomes torture to be endured by sheer strength of will, while the existence of illness forces characters to confront their mortality, the threat of nothingness (Meyers, 1985). Mann's *The Magic Mountain* (dealing with tuberculosis), *The Black Swan* (uterine cancer) and *Death in Venice* (a cholera epidemic) and Solzhenitsyn's *Cancer Ward* are examples of such works.

It is not only disease and death that have provided inspiration for novelists. Literary representations of doctors as they go about their day-to-day work contribute to sociocultural analyses of medical practice from the perspective of the medical professional rather than that of the patient. In his novel *Saturday* (2005) Ian McEwan presents a day in the life of a London neurosurgeon, including detailed descriptions of his working life and his state of mind while performing surgery. As a novelist, McEwan is able to portray the interior emotions, thought processes and psychological states of the neurosurgeon, and demonstrates how his protagonist's

medical identity extends into his everyday life, shaping his thoughts and his interactions with others. Such artistic representations of medical practice and health and illness are able to provide the rich kind of 'sensory detail and mundane trivialities' (Harris, 2008: 512) which may offer enlightening insights into the ontology of medical practice for readers of the work.

Medicine, Illness and Death in Popular Culture

Representations of medicine, illness and death are not limited to cultural texts such as literature and opera, but also routinely feature in the popular mass media. Fictitious portrayals of doctors and other health professionals and patients have been common in popular fiction and film and particularly television dramas and soap operas since the 1950s. Successful and often long-running television series since that time have included the American *Marcus Welby MD, Dr Kildare, M*A*S*H, St Elsewhere, Doogie Howser MD, Chicago Hope, ER, House, Scrubs, Nurse Jackie* and *Grey's Anatomy* and the British *Dr Finlay's Casebook, Casualty, Cardiac Arrest, Peak Practice, Holby City* and *Doc Martin.* In soap operas, sickness, injury and death have provided vital themes for plots, especially accidents and violent events, suicides, homicides, psychiatric disorders, pregnancy-related conditions, heart disease and infectious disease (Cassata et al., 1979). Medical breakthroughs and the personalized stories of people suffering from disease regularly make newspaper headlines and provide the basis of television and radio news and documentaries.

In most fictional television programmes, sickness is treated, usually competently and successfully, by doctors in hospitals using the latest technology and fast-acting drugs (Turow and Coe, 1985). The status of the doctor is usually indicated by his or her white coat, signifying authority, the objectivity and power of laboratory science and hygienic purity, and the stethoscope, ultimate symbol of medical technology and the ability of the doctor to gain access to the patient's body to hear or see bodily functions denied patients themselves, while retaining a certain distance between doctor and patient (Krantzler, 1986). Until quite recently, members of the medical profession were most often represented as omnipotent figures who had the power to save lives against all the odds, while nurses were depicted either as maternal, nurturing figures, dispensing kindly care and assisting doctors while validating their superiority, or as promiscuous sex objects (Karpf, 1988; Lupton and Chapman, 1991; Turow, 1989). For example, analyses of the depiction of doctors on American television dramas have found that they were generally portrayed as successful, benevolent, knowledgeable and authoritative, with almost mystical powers to dominate and control the lives of others (McLaughlin, 1975; Turow,

1989).

Doctors as characters in television dramas have tended to dominate over other professional occupations, and to be shown as more successful, peaceable and fair-minded than most other characters (Gerbner et al., 1981). They have routinely been depicted as 'saving patients from themselves', by forcing patients to confront truths or show courage in the face of pain or death (Turow, 1989: 146). So too in the news media, doctors are routinely presented as superhuman figures, the secular equivalent of clergy, and medicine is portrayed as the avenue by which miracles may be wrought (Lupton and Chapman, 1991; Lupton and McLean, 1998).

Documentary representations have similarly been very positive in their portrayals of members of the medical profession and of scientific medicine generally. *RPA,* a 'reality' docudrama made in Australia and focusing on the experiences of real patients and staff in Royal Prince Alfred Hospital, Sydney, especially underlines the power and benevolence of the medical practitioners working at that hospital. In scenario after scenario, grateful and passive patients are cured of their illness and are shown thanking their doctors profusely before leaving the hospital. While members of the medical staff often appear realistically awkward in their face-to-face dealings with their patients, this very popular series rarely shows them making mistakes or a procedure going wrong. The fantasy that doctors will do the right thing all the time and that they can cure any malady is strongly perpetuated in this series.

In popular media representations, medical technology, in particular, is singled out as the apotheosis of medical magic. The trappings of technology – the complex machinery, the flashing lights, the 'blips' of the monitors, the graph measuring the strength and regularity of the pulse – dwarf the figure of the patient, shown in medical dramas and news and documentary features as lying passively on the bed, connected to the machinery by various tubes and wires, rendered part-human, part-machine.

Despite the long-term dominance of positive portrayals of medical practitioners and medicine itself in popular culture, in the later decades of the 20th century and into the 21st century there have been a growing number of portrayals of doctors and scientific medicine as fallible. News coverage of doctors commonly reports cases of professional negligence, sexual harassment and assault of patients by doctors, fraudulent behaviour and other scandals (Bradbury et al., 1995; Lupton and McLean, 1998). There have even been several highly publicized cases reported in the news media about doctors who have deliberately killed their patients. The English GP Harold Shipman was one of most notorious in recent times. He was convicted in 2000 of 15 counts of murder of patients from his Manchester practice and is suspected of murdering more of his patients with lethal heroin injections than could be proven in court. One official report claimed that he was responsible for

at least 215 deaths and probably many more (BBC News World Edition, 2003).

Television dramas have begun to represent doctors as all-too-human characters, capable of making mistakes or lapses of judgement under pressure and coping with complicated private lives which sometimes affect their work. For example, some physician characters in *ER* were portrayed as thoroughly unpleasant and unlikeable, although technically very competent. *Chicago Hope* commonly dramatized the ethical and legal issues around the practice of medicine in a large hospital, and the Australian medico-legal drama, *MDA,* focused entirely upon stories involving cases brought against doctors by their patients. The main doctor character around which the *House* series focuses is a very flawed character: eccentric, temperamental and obsessive with an unfortunate manner with others, including his patients. One of the two doctors who feature in the series *Nurse Jackie,* Dr Cooper, is portrayed as silly, vain, self-obsessed and childish, a figure of contempt for those he works with, to the extent that it is almost impossible to believe that he is a qualified doctor. He is contrasted with the eponymous character, an experienced nurse who, although she has her flaws (including an addiction to pain-killers, which she filches from the hospital), is wise, compassionate, kind and more competent at medical practice than Dr Cooper.

Nevertheless, in both news and fictional portrayals of doctors and scientific medicine, positive representations continue to outweigh those that are negative. Doctors remain the voice of authority in news stories on medical issues, receiving privileged access to frame debates in the news media. Medical negligence, malpractice and other misdoings on the part of doctors are newsworthy, but so too are continuing reports about the skills, expertise and dedication of members of the medical profession and the advances of medicine (Hodgetts and Chamberlain, 1999; Hodgetts et al., 2007; Lupton and McLean, 1998). While the doctor characters in ER, *House, Nurse Jackie* and several other medical dramas are now commonly presented as beset with human failings, the majority of television programmes tend to continue to represent them as competent in most cases and above all as highly dedicated to their profession and concerned about their patients' welfare and well-being.

Illness and Metaphor

Literary and popular-culture accounts of the illness experience have provided fertile ground for scholars of rhetoric, especially those interested in the use of metaphor. Individuals who have experienced illness or disease have often reverted to metaphorical discourse to conceptualize and articulate their experiences to them-

selves and to others. In the context of illness, metaphors enable people to render indefinite physical sensations such as pain more concrete (van der Geest and Whyte, 1989). This reliance is partly due to the lack of linguistic expressions for effectively conveying bodily feelings. As Virginia Woolf, in her essay 'On Being Ill' (first published in 1930), commented:

> English, which can express the thoughts of Hamlet and the tragedy of Lear, has no words for the shiver and headache. The merest schoolgirl, when she falls in love, has Shakespeare or Keats to speak her mind for her; but let a sufferer try to describe a pain in his *[sic]* head to a doctor and language at once runs dry. There is nothing ready made for him. He is forced to coin words himself, and, taking his pain in one hand, and a lump of pure sound in the other (as perhaps the people of Babel did in the beginning), so as to crush them together that a brand new word in the end drops out. (Reprinted in Enright, 1989: 19)

The frequent use of metaphor in the medical context is not surprising, for metaphor is used in all areas of verbal communication as an epistemological device, serving to conceptualize the world, define notions of reality and construct subjectivity (Clatts and Mutchler, 1989; Lakoff and Johnson, 1981). Metaphor works by association, comparing two non-associated entities with each other by centring on the ways in which they resemble each other. In doing so, the metaphor shapes perception, identity and experience, going beyond the original association by evoking a host of multiple meanings. The establishment of metaphorical associations both relies upon pre-established knowledge and belief systems to establish meaning and extends these meaning systems in other directions.

There is a reflexive relationship between metaphorical discourse applied to illness and disease: just as other concepts or things are used to describe disease, so is disease used as a metaphor. But metaphorical representations are not politically neutral; in fact, metaphors are commonly used in ideological struggles around a contested site of meaning, a linguistic strategy used to persuade the acceptance of one meaning over another. Metaphor works to 'naturalize' the social, turning that which is problematic into the obvious (Kress, 1985). For example, disease metaphors are most commonly used to describe disorder, chaos or corruption, as when describing communism as 'a cancer on society', or describing a psychopathic murderer as 'sick'. The word 'invalid' is also suggestive of loss of integrity:

> In English, we use the same word to describe an expired passport, an indefensible argument, an illegitimate legal document, and a person disabled by disease. We call each of them *invalid.* To be an invalid, then, is to be an invalidated person, a human being stamped *not valid* by the invisible but

invincible hand of public opinion. (Szasz, reprinted in Enright, 1989: 1454, emphasis in the original)

Metaphors describing the body, illness and disease are strongly associated with the commonplace objects of everyday life. Pouchelle examined the treatise on surgery written between 1306 and 1320 by Henri de Mondeville, surgeon to the French King, Philip the Fair, and the 'unacknowledged father of French surgery' (1990: 2). She traced the dominant metaphorical systems evident in de Mondeville's medical writings. Architectural metaphors were often used to describe the body; the body as house, as fortress, as edifice, cell, cage, box, coffer, with the womb as 'ultimate dwelling place, intended to shelter the foetus' (1990: 134). The eyes were represented as sentinels, with a porter (the duodenum) guarding the lower door of the stomach against the arrows of flatulence and needles of pain caused by bile, while the stomach itself was the kitchen of the household. The orifices and openings of the body were variously described as doors, passageways and bridges, with cracks and holes in the body's fortress serving to expose its occupant to the outside world. While the escape of bodily secretions was considered normal and necessary, an accidental or surgical wound provided a dangerous outlet for vital warmth and spirits and allowed contact with the outside pestilential air; the body was viewed as being in constant threat from outside.

Other conceptions of the body in de Mondeville's treatise described diseases as 'nature' invading the body with her mineral, vegetable and animal forms, or represented the body as a tool, or a workshop full of tools. The choice of metaphorical systems in this medieval medical text suggests the terror felt against the 'animality' of the flesh at that time. The view of the body as a potential for 'wilderness' against which all good Christians should guard is evinced by the metaphors of disease as the invasion of nature, and of the body as fragile, imperfectly protected by a skin riddled with holes that allow easy penetration. They imply a subsequent need to conceptualize the body as an ordered, reasoned, 'civilized' system that is not of nature. The medieval doctor-surgeon was charged with the responsibility for controlling the traffic between the inside and outside of the body, protecting culture, as represented by the body as dwelling or fortress, against the incursions of untamed nature (Pouchelle, 1990).

In 17th-century England, religion and health and illness were closely intertwined and religious metaphors were common in depicting the body. Writers represented the body as God's workmanship, a temple to be kept pure, clean and in good order. Medical metaphors were also employed to describe the religious experience: the sinner was described as a sick man, while sin was said to be 'cured' or 'healed' by Christ the 'healer' of the soul. The words 'health' and 'salvation' were exchangeable (Wear, 1985). Illness was commonly spiritualized by the Puritans as a corrective or punishment of God the Father, and often seen in providential terms,

© Deborah Lupton 2012

in which guilt and a sense of sin were prominent: 'Health, a good night's sleep or recovery from illness were noted as a sign of God's favour, but the onset of illness stirred up anxiety, self-doubt and guilt' (Wear, 1985: 71). However, even devout Puritans often used medical or physically based explanations of illness as well as those relying on religious belief, although medical remedies were often couched as 'God's gifts' (Wear, 1985).

The use of language when referring to disease and those at risk of contracting it has a direct effect upon the way in which ill people deal with their condition and the ways in which others treat them. Susan Sontag has published two influential essays which consider contemporary metaphors of illness and disease and their use to make moral judgements about the ill. In her introduction to the first essay, *Illness as Metaphor,* first published in 1978, she commented that, 'it is hardly possible to take up one's residence in the kingdom of the ill unprejudiced by the lurid metaphors with which it has been landscaped' (Sontag, 1989: 3–4). Although Sontag's work has been the subject of criticism for her belief that metaphorical associations can and should be stripped from diseases (notice her own use of metaphor when writing about the illness experience in the sentence quoted above), she was one of the first modern critics to cogently argue that disease is rendered meaningful through the use of metaphor.

In *Illness as Metaphor* Sontag focused upon two diseases, cancer and tuberculosis. She examined the ways in which these diseases, and later HIV/AIDS, have achieved an extraordinary cultural resonance in western society over the past two centuries. Sontag drew upon the understanding of metaphor not only as a rhetorical device, which compares unlike objects and concepts to emphasize their relatedness, but as a vital epistemological device, by which we understand the world. Writing from the perspective of one who had been diagnosed and treated for cancer, Sontag commented upon the stigmatizing and discriminatory effect of metaphorical discourse. In comparing tuberculosis and cancer, Sontag described the former as being mythologized as a disease of romance and passion, a sign of 'inward burning' or ardour, conceptualized as disintegration, transparency, hyper-activity alternating with elegant languidness, leading to a noble and often lyrical death. By contrast, cancer is a disease of uncontrolled, abnormal growth that invades the body, 'a demonic pregnancy' (1989: 14), a disease of industrial society and also the wages of repression of feeling. Cancer is thus a shameful disease, a metaphor for evil. Sontag argued that the stigmatizing nature of the metaphors that surround diseases such as cancer has a direct effect on practices, for people feel inhibited and frightened to seek diagnosis and early treatment.

In her essay on the metaphors of HIV/AIDS (1989), Sontag saw HIV/AIDS as attracting the same metaphors of invasion, but also those of pollution, implying invasion from outside rather than from within. The virus that

causes the immunodeficiency characterized by HIV/AIDS is represented as an invader of the body, removing its defences and allowing the opportunistic infections associated with the onset of HIV/AIDS to colonize and overwhelm the body: 'Cancer makes cells proliferate; in AIDS, cells die' (Sontag, 1989: 107). Thus both cancer and HIV/ AIDS are represented as punishments for living unhealthy lives, for taking health risks, for excesses of diet and lifestyle, weaknesses of will, self-indulgence and addiction (1989: 113).

The comparing of HIV/AIDS with plague inspires centuries-old fears of incipient illness, spread mysteriously and striking down large numbers of people almost without warning. Plagues are 'invariably regarded as judgements on society', and the use of the plague metaphor in describing another illness imputes such moralism (Sontag, 1989: 242). Other metaphors commonly used to describe HIV/AIDS have invoked the detective story, the wrath of God, apocalyptic visions, Gothic horror, silent assassins, tombstones and the grim reaper to 'make sense' of the new disease and its impact on society (Clatts and Mutchler, 1989; Lupton, 1993a, 1994a; Rhodes and Shaughnessy, 1990; Sontag, 1989; Watney, 1987; Williamson, 1989). These choices of metaphor to give meaning to HIV/AIDS throw light upon the stigmatized nature of the disease and the discriminatory attitudes towards people living with HIV/AIDS and those who are believed to be at high risk of HIV infection.

The Machinery of the Body

Since the industrial revolution, the mechanical metaphor has been frequently used in discourses on the body. As described above, even pre-industrial revolution metaphors adopted the imagery of the body as a workshop full of instruments and tools, including the lungs as blacksmith's bellows. Such conceptualizations have proved long-lasting in the ways that the body is represented in both medical and lay forums. Even when the mysteries of the human body were finally revealed by medical dissection, ages-old metaphorical understandings of the body's interior were not unduly disturbed, because such pre-established beliefs served to shape the perceptions of the newly exposed interior (Pouchelle, 1990).

The French philosopher Rousseau wrote of his body the following: 'After having read a little physiology, I began the study of anatomy, and took a survey of the number and working of the individual parts which composed my bodily machine. Twenty times a day I was prepared to feel the whole out of gear' (cited in Herzlich and Pierret, 1987: 91). Today we speak about 'recharging our batteries', 'blowing a fuse', of feeling 'run down'. Clockwork metaphors – 'I'm all wound up', 'What makes him tick?' – and hydraulic metaphors – 'bottled-up emotions', 'overflow of feelings' – are also commonly used when talking about bodily processes or human

© Deborah Lupton 2012

behaviours (Seddon, 1993: 183). These expressions adopt the language of technology in conceiving of the internal workings of the body as a combustion engine, a battery-driven machine or a plumbing system. Indeed, the concept of the heart as a 'pump', which can 'break' and which contributes to low or high blood 'pressure', has been so naturalized as to be rendered a dead metaphor, or an expression which we no longer recognize as a metaphor (Seddon, 1993). However, the heart is itself a powerful metaphor of emotion, especially romantic love, and heart problems are also often said to be linked to emotional rather than physical causes using older concepts of heart function; one is 'heartsick', 'heartsore', 'heavy' or 'light' of heart and may die of a 'broken heart' caused by emotional distress (Helman, 1985; Seddon, 1993).

The mechanical metaphor includes the idea that individual parts of the body, like parts of a car or plumbing system, may 'fail' or stop working, and can sometimes be replaced. The metaphor has the effect of separating mind and body, of valorizing medical techniques which focus upon locating a specific problem in a part of the body and treating only that part, and devaluing healing relationships which rely upon spirituality, personal contact, intimacy and trust (Helman, 1985; Martin, 1987; Scheper-Hughes and Lock, 1987; Stein, 1990; Turner, 1996). Hence the importance of the technological imperative in biomedicine: the dependence upon the use of machinery to fix machinery. The routine employment of organ transplants and artificial organs or parts such as pacemakers, plastic joints and hearing aids in high-technology medicine both is supported by and reinforces this image (Stein, 1990).

It was a logical conceptual leap in the late 20th century to imagine the body as a computerized system. There developed a symbiotic metaphorical relationship between computers and humans, in which computers were anthropomorphized (for example, they are said to have a 'memory'), while humans have increasingly been portrayed as 'organic computers' (Berman, 1989: 7). Indeed, the notion of 'biofeedback' originated in computer terminology (Stein, 1990). As a result of the computer metaphor, argues Haraway, bodies have become cyborgs and disease has become viewed as 'a subspecies of information malfunction or communications pathology' (1989: 15). Such a metaphor serves to restore order, rationality and predictability into our visions of the body (Montgomery, 1991).

One example of the machine/computer metaphor as applied to the human brain is an article published in the British popular science magazine *New Scientist* (Young and Concar, 1992) as part of the series entitled 'The secret life of the brain'. In the article were references to 'how the brain stores information', 'memory's hardware', 'the internal filing system', 'the machinery that enables us to retain a sequence of digits, letters

or words', 'electrical impulses', 'the brain's cellular architecture', 'fine-tuning connections between neurons' and 'memory networks'. The article then discussed attempts to perfect silicon chips that mimic the behaviour of the human brain. By this imagery there occurs 'the transformation of the human subject into an object, a repository, or else a collision site, for various types of detectable and usable information' (Montgomery, 1991).

The bio-information metaphor may be linked to the discovery of the genetic material DNA in cell nuclei in 1953 and the later introduction of the term 'genetic code'. The DNA molecule was described by Watson and Crick (the scientists who discovered DNA) as a 'template', a 'mechanism for self-duplication' and a 'copying process' for important information (Montgomery, 1991). With the discovery of DNA and RNA, the human body began to be represented in medical and scientific texts as a complex 'chemical factory' manufacturing 'chemical building blocks' using 'blueprints' (Nelkin and Tancredi, 1989). The current enthusiasm towards identifying or 'mapping' the structure and sequence of the genetic material in the human genome presents an image of the body and mind 'as machine-like "systems" that can be visualized on a computer screen and understood simply by deciphering a code' (1989: 15). The use of the word 'map', and the visual depiction of the alternating dark and light bands of chromosomes in identified genes like an aerial photograph of the earth's continents, invoke a related cartographic metaphor (Lippman, 1992) in which the human genome is portrayed as undiscovered and uncharted territory awaiting exploration and documentation of its secrets.

These metaphorical systems reduce individuals to their DNA codes. They are reworkings of the mechanical metaphor in representing the body as being comprised of a multitude of tiny interchangeable parts, rendering the body amenable to objectification and technological tinkering in the interests of developing the 'perfect' human. Sickness becomes viewed as the product of biological destiny, located within the individual and therefore requiring the intervention of technology to 'correct' the faulty code, hence drawing attention away from the examination of the social context of illness such as poverty, racism and sexism (Conrad, 1999; Lippman, 1992).

The Military Metaphor

The language of warfare is extremely common in modern medical and public-health discourses that deal with cancer, infectious diseases and other illnesses such as HIV/ AIDS. The immune system, for example, is commonly described as mounting a 'defence' or 'siege' against the 'invasion' of 'alien' bodies or tumours which

© Deborah Lupton 2012

are 'fought', 'attacked' or 'killed' by white-blood cells, drugs or surgical procedures. Such metaphors are common not only in literature dealing with medicine and individual treatment, but in public-health campaigns directed at large populations.

Sontag traces the use of the military metaphor in public-health education to campaigns mounted in the First World War to educate people about syphilis, followed after the war by educational programmes against tuberculosis. She cites one example of a campaign conducted in Italy in the 1920s, using a poster called *'Guerra alle Mosche'* ('War against Flies') which depicted the flies as 'enemy aircraft dropping bombs of death on an innocent population' (Sontag, 1989: 98). On the bombs were inscribed the words *'microbi'* (microbes), *'germi della tisi'* (the germs of tuberculosis) and *'malattia'* (illness). Sontag (1989) asserts that the military metaphor has such resonance in western society's discourses on illness and disease because it appeals to the need to mobilize against an emergency, to make sacrifices, to do everything possible to counter a threat to life. Montgomery (1991) goes even further into history for the origin of the military metaphor, commenting that images portraying disease in the form of an 'attacker' armed with spear or quiver go back centuries to the Middle Ages. He argues that contemporary images of violence and warfare tend to 'connect disease in its connotative charge with the greatest evils and atrocities of the twentieth century: fascism, Stalinism, even the Holocaust' (Montgomery, 1991: 349).

In the modern era, a dominant image around which notions of the internal workings of the body have been organized is that of the germ theory of disease. Illness is no longer predominantly conceived as an evocation of evil incurred by the wrath of God, but as a microscopic invader, intent on entering the body and causing trouble. 'Germs' are commonly believed to have motivation and evil intentions: 'There are no "good" Germs or "normal" Germs; all Germs are bad' (Helman, 1978: 119). Among lay people, bacteria and viruses are conceived of as:

> … living, invisible, malevolent entities [which] have no free existence in nature, but exist only in or among people. They are thought of as occurring in a cloud of tiny particles, or as a tiny, invisible, single 'insect'. They traverse the spaces between people by travelling in the air, or in the breath. Germs causing gastrointestinal symptoms are seen as more 'insect-like' ('Bugs'), and are larger in size than those Germs causing other symptoms. Germs have personalities; these are expressed in, and can be recognized by, the various symptoms they cause. (Helman, 1978: 118)

By Edwardian times, the germ theory of disease had become central to medical and lay discourses on illness,

centring on anxieties over pollution, purity and boundary maintenance. An example is the constant use of such words and phrases in the media and political forums as 'social purity', 'mothers of the race', 'the imperial race', 'British stock', 'degeneration' and 'purification of the race'. This usage may be interpreted as emerging from concerns about maintaining the boundaries between inside and out and protecting against internal pollution, decay and degeneration (Wright, 1988). This was evident from the phraseology of a hygiene reader, E. Hood's *Fighting Dirt: the World's Greatest Warfare* (1916), in which it was asserted that:

> While the boy and girl yearn to handle a sword and go to the wars, deadly enemies surround them, lurk in their clothes, cling to their flesh, penetrate into their mouths, and only wait a favourable moment to attack their bodies in force, as soldiers have attempted the capture of some great fortress, such as Gibraltar. (cited in Wright, 1988)

Military metaphors continually emerge in contemporary popular-media accounts of medical science and immunology. For example, an article in the American *Time* magazine (31 August 1992) entitled 'Attack of the superbugs' described the increasing resistance of bacteria to drugs such as penicillin, using the following rhetoric:

> In the battle against old scourges, magic bullets are losing their power, and invisible legions of drug-resistant microbes are again on the march … medicine is no longer confident of winning the battle … Using marvellous powers of mutation, some strains of bacteria are transforming themselves into new breeds of superbugs that are invulnerable to some or all. (Nash, 1992: 46–7)

In such texts, bacteria are anthropomorphized into wily aggressors, deliberately changing themselves to elude detection and attack from their human foes. Even the scientists interviewed for the article used colourful metaphors of war to describe their efforts; language that perhaps gives their endeavours legitimacy and a sense of great importance, positioning them as the 'generals' in the battle against disease. The military metaphor has been particularly dominant in representations of HIV/ AIDS. For example, in an article on HIV/ AIDS published in *New Scientist* (Brown, 1992), such phrases as 'HIV can reproduce only by hijacking its host's genetic material … the virus is winning … a moving target … why is the war against the virus proving so difficult? … the virus attacks the very cells that should defend the body against it … know your enemy' prevail.

Yet the use of such metaphorical systems is not confined to scientific, medical or media representations of

HIV and AIDS. Autobiographical accounts of HIV/AIDS draw upon similar imagery to describe personal experiences of illness, suggesting the difficulty of resisting dominant military metaphors even for those who are ill. As Dreuihle vividly described in his account of his experience as a person who is HIV positive:

> My personal war began two years ago when I was mobilized by AIDS. All the pleasures of peacetime and my carefree life were suddenly banished, as if an orchestra had stopped playing to let the theatre manager announce that war had just been declared, that Pearl Harbor had been bombed. (1988: 6–7)

Medical practitioners themselves may draw upon a complex series of metaphor systems to describe their work and patients. Stein (1990), as a medical anthropologist working in an American hospital, engaged in participant-observation of physicians treating their patients. From this position he identified a series of common metaphors within western medicine, including those of: military/war; sports; technology/engineering/mechanical; family; life-cycle; religion; economics/business; freedom; masculinity/femininity; and time. These metaphor systems were evident in the informal and formal discussions between doctors, talk between doctors and patients, at medical seminars and conferences, in drug advertisements and articles in medical journals and in representations of medicine and disease in the popular media. Doctors working in the hospital repeatedly described themselves as being 'on the front line', in need of 'getting aggressive' with patients and using 'shotgun therapy' or 'magic bullets'. They commonly described working in the emergency room as being 'in the trenches'.

The doctors in this study saw themselves as battling not only against disease, but also against the masculine world of the medical system and other physicians, attempting to master their work as 'winners', never allowing themselves to lose control or to fail in the clinical setting. In this context, the patient became the feminized, passive object, to be poked, prodded and 'attacked' as the focus of the battle for medical male supremacy (regardless of the gender of the patient or physician). Hence such statements made by doctors as 'He tackled the real difficult patient. He handled it aggressively. But he got shot down', and 'In the clinic I hit her with everything (antibiotics), but she still came in the hospital door', which combine military and sporting metaphors (Stein, 1990: 72).

The Body against Itself: Metaphors of the Immune System

The discourses of science as an heroic quest, nuclear exterminism, space adventure, extra-terrestrialism, exotic invaders and military high-technology are dominant in popular accounts of the immune system (Haraway, 1989). The anthropology of immunology was the subject of an analysis that drew upon major mass-media articles and popular monographs on the immune system published in the United States in the 1980s for the source of data (Martin, 1990a). The dominant imagery used was that of 'the body as nation state at war over its external borders, containing internal surveillance systems to monitor foreign intruders' (1990a: 410). Another study by the same researcher found that the dominant metaphorical systems she had identified as recurring in popular medical and media texts on the immune system were emerging in lay people's descriptions of their bodies. They commonly used violent imagery such as 'the policing system', 'the defence system', 'like an army' to articulate their understandings of the processes of their immune system, and often compared the strength or effectiveness of their immune response to that of others, evincing beliefs that one can 'train' or 'build up' one's immune system (Martin, 2000).

Such phrases demonstrate the pervasiveness of value judgements made about relative superiority in which some groups or individuals are said to have 'better', 'stronger' or 'tougher' immune responses than others. Martin (1990a, 2000) views this discourse as being part of a metaphorical 'currency' of health which is a response to an increasing focus of media and medical literature over the past decade on the immune system and its link with such diseases and disorders as HIV/AIDS, allergies, cancer, heart disease, arthritis, diabetes and chronic fatigue system. She sees this change in popular body imagery as shifting from the centuries-old notion of the body as defended from disease at its external borders, like a fortress, to a body defended by a complex internal system.

Under the ideology of the 'currency' of health, the discursive way in which the 'deficient' immune systems of those people who have an auto-immune system or immunodeficiency disease (such as systemic lupus erythmatosus [lupus] or HIV/ AIDS) are described becomes emblematic of their imputed moral deficiencies. People whose immune systems are 'inferior' become members of a new stigmatized and victimized underclass, subject to 'Body McCarthyism' where 'political discourse is reduced to the purity of your bodily fluids' (Kroker, 1992: 325).

One example of the use of discriminatory language is an article entitled 'Self Destruction' published in a week-

© Deborah Lupton 2012

ly Sydney community newspaper (*Western Suburbs Courier*, 15 January 1992) as part of a series of regular medical columns by Professor John Dwyer, an immunologist known largely for his work in HIV/AIDS. In his article, Dwyer outlined the symptoms, causes, risk factors and treatment of lupus, using revealing terminology in his attempts to describe the immunological intricacies of the disease to a lay audience. The title of the article itself implies blaming the victim and punishment: the illness is 'self' destruction, not externally imposed destruction; it is the body turning against itself. Dwyer gives a detailed explanation of the immune system and the reasons for auto-immunity: 'that unfortunate set of circumstances which sees our normally impeccable immune system, designed to attack "foreignness", make a mistake and instead attack "self" – a biological form of self-destruction. How can such a terrible thing happen?'

These sentences are revealing of the discourses of military strategy, xenophobia, purity and blame which recur in medical and media representations of the immune system. In this account, the onset of auto-immunity is described as an 'unfortunate set of circumstances' and a 'terrible thing'. The purity, normality and balance of the body have been overturned, the 'impeccable' status lost. Instead, the body has made a 'mistake' and turned against itself rather than against 'foreignness'. Later in the article, when Dwyer is describing the treatment for auto-immune diseases such as lupus, he uses the phrase 'The punishment must fit the crime', in which there is a curious slippage between the criminal (the disease) and the patient's body. The 'punishment' is the treatment, and therefore is administered not only to the condition, but to the patient's body.

The metaphors used to describe the immune system thus construct and maintain the boundaries and binary oppositions between inside and outside, the normal and the pathological. Popular medical and scientific discourses surrounding the immune system seek to mark out difference. Like other uses of military metaphors to describe disease causation and control, both scientifically and in lay discourses, they reflect wider divisions, anxieties and concerns in society. While military imagery may overtly connote decisive action and the refusal to 'give in' to the disease, at a deeper level of meaning this discourse serves to draw boundaries between Self and Other by representing the body as a nation state which is vulnerable to attacks by foreign invaders, invoking and resolving anxieties to do with xenophobia, invasion, control and contamination. The ideological work performed by such imagery is to make violent destruction seem ordinary, prosaic and domesticated, and xenophobia acceptable (Haraway, 1989; Martin, 1990a). In the arena of public health, military imagery also serves to justify the intervention of the state in the everyday habits of its citizens, granting power to external agencies to control individuals' bodies in the interests of the health of the body politic.

Cancer Metaphors

Like HIV/AIDS, cancer is a disease that has occasioned a constellation of metaphorical systems, largely due to its severity, mystery and evasion of medical solutions. Indeed, Herzlich and Pierret assert that for the populations of western societies, 'cancer is THE illness of our time' (1987: 55). As they note, the name of the disease is itself a metaphor, derived from the Latin word for crab, and denoting a swollen protuberance like the legs of a crab. In medieval times, cancer was commonly represented as a hungry animal, 'gnawing' at the body (Pouchelle, 1990: 169). Even today, 'cancer is also fraught with phantasms of rot invading the body, animals that gnaw and destroy it' (Herzlich and Pierret, 1987: 56). This metaphor was vividly evident in the view of one Frenchman, who commented that, 'I have the feeling that I am in a pre-cancerous state, in the condition that favours its hatching' (Herzlich and Pierret, 1987: 65).

In modern times, notions of moral culpability have surrounded people suffering with cancer. This disease in modern western societies is often viewed as brought upon oneself through irresponsibility, either by indulging in injudicious diets or tobacco consumption, or by suppressing negative or angry thoughts (Sontag, 1989). A French study analysed letters sent by cancer patients in response to a call by the French minister of health to instigate a public debate on cancer. It was found that the way in which cancer patients dealt with the dominant discourses on their disease were often contradictory; on the one hand, the victorious and hopeful 'winning the battle against cancer' discourse, often expressed in advertisements for cancer agencies and benefits; on the other, the 'cancer equals death' discourse that was also common (Pinell, 1987). An homology between social disorder and biological disorder appeared in many letters that referred to poor working conditions or low socioeconomic status as a cause of their cancer: for example, one woman noted that, 'I am convinced that there will be no cancer in a better world' (Pinell, 1987: 32).

Such interpretations pointed to a cause of cancer which was located externally to the people with the disease, and thus for which they could not personally be blamed. Letter-writers wrote of the stigmatization they had experienced. As one person wrote; 'Friends avoid you, not knowing what to say, and it's a mistake, one should speak about it; it is less frightening' (Pinell, 1987: 36). Cancer was often referred to by cancer patients as evil, sometimes with a capital E, or as 'the thing', 'the scourge', 'an invisible enemy, whose weapons we know nothing about', or 'that contemporary plague, which has taken other forms in the history of the world' (1987: 37–8).

Like other medical writing aimed at lay audiences, popular media accounts of cancer often incorporate metaphors in the attempt to 'simplify' explanations. For example, in a self-help book entitled *Your Cancer, Your Life,* Dr Trish Reynolds describes cancer cells as disorderly and out-of-control: 'Cancer growths are made up of cells which belong to our body but which have stopped behaving in a co-operative and orderly fashion' (Reynolds, 1987: 26). Later in the book she observes that the multiplication of cancer cells 'has no purpose – it is not in order to replace or repair. The cells do not differentiate into useful cells. They keep growing to the detriment of the body and regardless of normal control mechanisms. Unlike normal body cells we can think of cancer cells as uncooperative, disobedient, and independent' (1987: 27) and further comments that '[n]ormal cells exist peacefully side by side with their neighbours. Cancer cells damage and destroy them' (1987: 33). The overriding metaphor used here is that of the cellular dimension of the human body as a society, peopled by 'good' and 'useful' cells that serve a function, but with social order threatened by deviant cells that refuse to obey societal laws. Cancer is therefore portrayed as a challenge to rationality.

Military discourse is often linked to other discourses in representations of cancer. Thus, for example, mass-media accounts of people's experiences of illnesses such as cancer may incorporate military language, bringing together 'battle' and 'warrior' with such words as 'God', 'Jesus', 'faith' and 'miracle' to represent cancer as a 'challenge to the spirit' (Hoffman-Goetz, 1999). A study of American-press accounts of men with testicular cancer found that articles frequently drew upon both sporting and military discourses (Clarke and Robinson, 1999). Similarly, a Canadian analysis of newspaper reports of prostate cancer found descriptions redolent of the battlefield, including representations of men who had died from the disease as 'fallen heroes' who had nobly fought cancer and idealized portrayals of stereotypically masculine traits of stoicism, competitiveness and valour (Halpin et al., 2009). Indeed, sporting metaphors are common in media representations of various types of cancers for both men and women, depicting their 'heroic struggle' against the disease (Seale, 2001).

As mentioned above, a dominant discourse surrounding cancer in modern western societies is that of hope. This discourse is related to military and sporting metaphors and discourses, for it postulates that 'winning' the battle against cancer is intimately linked to having a positive attitude to getting better. To despair, to lose hope, are frowned upon as strategies of dealing with diseases such as cancer. People with cancer are lauded if they appear to be brave, never allowing themselves to 'give in' to the disease. There is an emphasis on 'will' which argues that 'if one has enough hope, one may will a change in the course of the disease in the body', an emphasis which is related to notions of individualism, fighting spirit and the power of thought to shape one's life course which are particularly evident in American society (Good et al., 1990: 61). This discourse supports

the current practice of informing a patient of his or her diagnosis in the United States, a practice which is not followed in Japan or Italy where it is believed best to withhold or mask the diagnosis of cancer in order better to protect the patient (Good et al., 1990).

The discourse of hope can be problematic when, despite the patient's and the doctor's best efforts to remain optimistic, the illness worsens, for the patient can then be held culpable for not being positive enough. When this metaphor dominates, cancer victims are expected to fight the sickness, providing them with the 'glories of a war hero' even if they do not actually get better. In a study including interviews with cancer patients and oncologists, both patients and their doctors constantly referred to the disease as an enemy and an aggressor, and perceived themselves as allies against cancer (Erwin, 1987). As these data suggest, in western anglophone societies, cowardice, giving in, fear and 'denial' are not considered socially acceptable ways of dealing with the disease. Rather, optimism, a coping attitude, cheerfulness and strength are required. Dying bravely becomes a victory. Similarly, the physician must respond with military-like leadership, authority, precision, confidence, lack of emotion and an aura of control.

The discourse of hope is often evident in popular representations of people with cancer, especially if they are well-known. Women's magazines are particularly fond of personalizing stories about famous women's experiences of breast cancer in terms of their triumph over the scourge. An example is an account of ex-First Lady Nancy Reagan's experience with breast cancer. The article noted that after her recovery from mastectomy, Reagan 'raised her arms in the air like a boxer giving a victory salute' (New *Idea* magazine, 5 December 1987). Another example is an article printed in the *Australian Women's Weekly* (Olivar, 1993) in which singer/actress Olivia Newton-John was interviewed about her (then) recent diagnosis and treatment for breast cancer. The article featured a number of photographs of Newton-John with her husband and child, smiling brightly and looking attractive and healthy. Titled 'My Fight for Life', the article comments that 'positive thinking is proving to be Olivia's greatest weapon in fighting her cancer, and, if attitude can cure, she has every reason to be optimistic … As she chats, you would never guess she was suffering from a life-threatening illness.' When describing her feelings after being told her diagnosis, Newton-John is quoted as saying, 'I don't consider myself ill. I had cancer and I'm having treatment for it, but papers write that you're this, you're that, you're depressed, you're really sick. That makes me cross. I need people telling me that I'm doing well and giving me support, not the other way round.' Such cases are held up as a public example of how personal courage and the constant maintenance of a positive hopeful attitude, in conjunction with medical knowledge and skill, can save people with cancer from death.

© Deborah Lupton 2012

Metaphor, Gender and Medicine

In her analysis of the discourses around medicine, science and gender between the 18th and 20th centuries, Jordanova (1989) suggests that the biomedical sciences in 18th- and 19th-century France and Britain were characterized by a discourse which linked nature with women and culture with men. The Enlightenment privileged an epistemology that was rooted in empirical information derived from the senses. The methods of science and medicine were considered to avoid superstition and dogma, and were addressed to investigating phenomena surrounding sexuality. In so doing, she argues, the discourses of science and medicine employed sexual metaphors to describe their endeavours: 'for example by designating nature as a woman to be unveiled, unclothed and penetrated by masculine science' (1989: 24).

For Jordanova, the wax anatomical models of women used in the medical context, complete with eyelashes, eyebrows and pubic hair as well as detachable organs, are representative of and contribute to a constellation of practices which constituted central discourses on gender, medicine and science in the Enlightenment. She views these wax models as ultimately erotic, in that they allowed penetration beneath the surface of the skin by the successive removal or 'unveiling' of layers. The metaphor of 'unveiling' signifies the feminine mystique rendered open to the masculine gaze, allowing possession and mastery.

Jordanova uses other texts – medical textbooks and illustrations of the 18th century, the writings of the 19th-century French historian Jules Michelet, a statue of a woman made in 1899 entitled 'Nature Unveiling Herself before Science', paintings of dissections of female corpses, the Fritz Lang film released in 1927, *Metropolis,* and contemporary medical advertisements – to inquire into the historical linkages between science, medicine and the constitution of gender and sexuality. The woman/ nature metaphor was characterized by ambivalence: women were represented as softer, more sensitive and more vulnerable than men, but also tougher and more tenacious of life. The breast was of particular interest to practitioners concerned with moral philosophy and ethics, as it symbolized women's role as mothers. The women-as-nature and men-as-culture dichotomies became key terms in struggles around child-care and midwifery in the period. Women as mothers and as midwives were represented as irrational and irresponsible, their knowledge based on tradition rather than experience, needing the guidance of men who possessed scientific knowledge.

A study of contemporary visual and print media's representations of scientific explanations of human reproduction (Martin, 1991, 1992) found that while both male and female reproductive organs were metaphorically

© Deborah Lupton 2012

represented as 'systems for the production of valuable things', menstruation was described as the death of tissue, implying wastefulness, chaos and 'production gone awry' (Martin, 1992: 411). The processes of male reproduction, by contrast, were portrayed as awe-inspiring, productive in the capitalist sense, successful. This representation was evident in a medical textbook in which the following sentence appears: 'Whereas the female *sheds* only a single gamete each month, the seminiferous tubules *produce* hundreds of millions of sperm each day' (Martin, 1991: 486, emphasis in the original). Ova were anthropomorphized in a feminine manner, depicted as passive, holy, lying hidden in wait: 'a dormant bride awaiting her mate's magic kiss, which instils the spirit that brings her to life' (1991: 490); while sperm were described using active masculine words connoting strength and aggression: they 'deliver' their genes to the egg and 'burrow through the egg coat' and 'penetrate' it (1991: 489).

Such metaphors appear to pervade popular descriptions of human fertilization in the western mass media. For example, in a segment of the Australian popular science television programme *Quantum* (7 April 1993), the ovum was described as coyly preparing itself for its romantic interlude: 'patiently waiting' for 'that ultimate encounter – with the sperm'. Sperm production was described as 'a *very*flamboyant affair … one of the processes where you get these spectacular volumes of cells produced … The sinuous passageways of the testes are a vast production line, an impressive three metres long when uncoiled.' It is notable that the 'journey' undertaken by the sperm through the vagina up to the fallopian tubes was portrayed as dangerous, the sperm struggling to survive in a 'treacherous' environment, competing amongst themselves to win the battle against the vagina dentata: 'they must escape the acid secretions of the vagina and then negotiate the mucous-covered folds and recesses of the uterus … By the time [the sperm] actually reach the egg, only a couple of hardy survivors remain.' The union of the sperm with the ovum was described as a violent, almost rape-like attack: the sperm were said to wear 'a protective cap, a kind of chemical crash helmet' which helped them to 'hurl themselves against the cell wall [of the ovum] to batter a tunnel right through'. Yet the ovum enjoyed the encounter: 'at the moment of contact between egg and sperm, the sparks fly. A wave of electric excitation sweeps across the egg membrane … as if to celebrate the happy union, the cell almost immediately begins to split in two.' The subtextual ideologies of these descriptions are transparently drawing upon stereotypical concepts of masculine and feminine sexuality in describing the behaviour of cells.

It seems that such stereotyping is difficult to avoid. Even when it became recognized amongst biophysicists that the ovum plays an active role in fertilization, metaphors of entrapment and aggression depicting the ovum as 'femme fatale' replaced the passive metaphors in medical texts. These portrayals are revealing of the di-

chotomies constructed in medical and scientific discourses between masculine and feminine. The imagery also has implications for debates over human reproduction. By ascribing personhood to gametes, this discourse can play an important role in legitimizing medical scrutiny of reproductive processes, and has implications for legal and other initiatives concerning abortion, foetal surgery, genetic screening, assisted reproductive technologies and contraception (Martin, 1991).

The Iconography of Illness, Disease and Death

Over the centuries, the visual dimensions of disease, death and the body have provided much inspiration for artistic works. Vivid descriptions from historical accounts of illness such as syphilis and the plague represented the ill body as a monstrous sight, an 'exteriorized horror' of rotting flesh, in which the presence of disease was immediately visible to onlookers and sufferers themselves (Herzlich and Pierret, 1987). Given the hideous visual nature of illness and disease, it is not surprising that attempts to portray death and the fear of death recur frequently in medieval and Renaissance art, particularly the image of the 'Dance of Death' during the years of the Great Plague that devastated Europe in the late 16th century. In the 15th and 16th centuries, death was commonly portrayed as a skeleton, hovering about and ready to pounce, riding on horseback, or gleefully celebrating his victories over humanity in a dance. Depictions sought to convey the ubiquity and arbitrary nature of death: 'in Hans Holbein's woodcuts of the dance of death (1538) we see death snatch at one moment a healthy child from the hearth, and at another a robust woman from a nunnery' (Prior and Bloor, 1993: 353). To illustrate the idea that death was caused by chance, a stroke of fate, dice and playing cards were often depicted in paintings and engravings of death, and time pieces were used to denote the inevitability of mortality (Prior and Bloor, 1993).

By the Renaissance, the need to understand death and disease was expressed in illustrations and paintings which sought to portray the inside of the body, often placing the anatomically detailed form in a naturalistic setting so as to imply morality. Anatomical engravings of the 16th century often had a hint of sadism as well as a feeling of compassion for the exposed bodies, their skin flayed to reveal the muscular and venous systems below. By the 17th century, the machine metaphor had received expression in portrayals of the human body in fanciful illustrations which represented heads and bodies as comprised of wheels, cogs, bellows, pistons, metal plates and other industrial objects (Lucie-Smith, 1975).

© Deborah Lupton 2012

Over the history of western medicine, visual representations of illness have played an important political role in the categorizing of the Other, those deemed abnormal or dangerous to a society. For example, portraits of patients were commissioned by doctors in the early 19th century in the attempt to construct a 'pathological identity' by documenting physiognomies of deviant types institutionalized for insanity or 'crimes against nature' (Marshall, 1990: 22). By the late 19th century, with the development of photographic techniques, many doctors began to document the manifestation of illness in their patients using photography. Patients were routinely depicted facing the camera, displaying the marks of their illness. The disempowered – male homosexuals and other sexual 'deviants', prostitutes, criminals, drunkards, women hysterics, sufferers of sexually transmitted diseases (especially syphilis), the poor and the mentally ill – were subjected to such recording of their bodies in the interests of the clinical gaze, effectively extending the surveillance techniques of the medical profession (Fox and Karp, 1988; Marshall, 1990). This enterprise of detection, labelling and categorization conducted by 19th-century medics, psychologists and sexologists can be compared to voyages of botanical discovery in the 17th and 18th centuries when a new world of exotic flora and fauna was discovered by the early colonialists exploring the 'dark continent' – the 'Other' of western civilization (Marshall, 1990).

The practice of documenting pathology by photographing the diseased body remains central to medical literature, especially medical journals, textbooks and magazines. While patients' faces are perhaps not shown as freely now as they were in the late 19th century, parts of their bodies are routinely depicted, often in full colour and gory detail, for the interest of medical practitioners and researchers, including close-up shots of diseased or deformed genitalia. The news media also routinely publish or broadcast images of the disempowered diseased body: the starving, emaciated child in Africa, the equally thin gay man dying with HIV/AIDS and marked by the lesions of Kaposi's sarcoma, the brutally exposed, pulsing internal organs of a person undergoing surgery. Mass-media health-promotion campaigns directed at frightening people of the dangers of smoking use vivid, disgusting images of the bodies of smokers ravaged by disease caused by their habit – clogged lungs and arteries, gangrenous limbs, blood coughed into a handkerchief – and in the attempt to persuade viewers to adopt healthy eating and exercise habits have often portrayed semi-naked fat bodies in unflattering images as examples of what will happen if people allow themselves to put on weight.

The iconography of medical advertising is a revealing insight into the ideologies and mythologies surrounding the doctor-patient relationship in the context of the biomedical system of healing. I undertook a study examining the representation of patients in Australian medical advertising in magazines directed towards the medical profession readership (Lupton, 1993b). I found that common to several medical advertisements was the pictorial dismemberment of the human body, featuring 'sawn-off' bits of bodies. Body parts were sometimes

transformed and distorted through graphic-design techniques. The mechanical model of the body received overt expression in a number of striking visuals which departed from any naturalistic representation of body parts by using symbolism which rendered organs, limbs or biological processes as parts of machinery. Indeed, some advertisements avoided *any* direct reference to the shape or appearance of body organs, using highly symbolic imagery to connote body processes. For example, an advertisement for a stomach-ulcer treatment used a photograph of a large brass switch mounted on a metallic surface, as if part of industrial machinery. Next to the switch was shown a begrimed gauge with clock-like face. The switch was labelled with the word 'ACID' in big brass letters, followed by the engraved words 'ON' and 'OFF'. The switch lever was set at 'OFF', implying the effect of taking the drug upon the patient's stomach is like a mechanical switch which is able simply to 'shut off' acid production by cells in the gastro-intestinal tract.

Few of the advertisements examined showed a doctor actually touching a patient, evidence of the diminishing importance of human contact in the medical encounter. The effect of such fragmented representations is to dehumanize the doctor-patient relationship by rendering patients' bodies as distinct from their individuality. When only part of the patient's body is depicted without a face, the result is dehumanizing and promotes anonymity.

In his collection of essays, Gilman (1988) undertook a wide-ranging enquiry into how disease and illness have been represented in visual media over the centuries. The sources he used to trace representations included: the illustrations used by the ancient Greeks and the 19th-century Chinese in their medical literature; drawings of sexual intercourse by Leonardo da Vinci; medieval paintings and engravings of madness; photographs of mental patients from the 19th century; a German opera libretto dramatizing homosexuality; the paintings of Blake, Goya, Delacroix, Vincent van Gogh and others depicting asylums; and recent mass-media images of people with HIV/AIDS. Gilman's thesis is that the iconography of illness is an indication of the ways in which society deals with and conceptualizes disease: 'The portrait of the sufferer, the portrait of the patient, is therefore the image of the disease anthropomorphized' (1988: 2). Gilman argues, like Sontag, that the construction of stereotypes of an illness and the patient suffering from it through images and language influences the way in which the illness and the patient are treated by others, and is internalized by patients, shaping their experience of the illness. His tracing of common representations over centuries of western development demonstrates the resonance of certain images even today.

These images result from society's need to organize a chaotic, frightening world, to draw (literally) the boundaries between ourselves and the diseased Other so as to provide control over the randomness of disease

© Deborah Lupton 2012

(Gilman, 1988). By portraying the diseased Other in art, it is labelled and stereotyped for ease of identification, providing 'a basic reference against which all images of disease, whether literary or visual, are measured' (1988: 3). Images of disease remain relatively stable over centuries, influenced only slightly by shifting popular or medical views.

There are two levels on which the visual imagery of disease is coded: first, on the level of the social construction of categories of disease; and second, on the level of the internalization of such images in individuals or groups who are labelled as 'at risk' (Gilman, 1988). Thus Gilman examines myths of madness in Renaissance paintings as well as the art produced by those who have been labelled insane. The most recent icons of disease examined are American newspaper and magazine photographs and drawings of people living with HIV/AIDS. Gilman draws a direct historical parallel between such representations and the centuries-old iconography of syphilis, demonstrating how the stigma of sexually transmitted disease was similarly appropriated by visual imagery. By doing so, he places representations of HIV/AIDS in their historical context and casts light upon society's response to people living with HIV/AIDS.

While the female body has been subjected to 'an excess of visibility' in medical and popular-media representations, the male homosexual body has been virtually invisible. It was not until the advent of HIV/AIDS that gay men began to be represented visually, albeit in a manner that tended to objectify and stigmatize them, linking death and homosexuality as almost inseparable (Marshall, 1990). An analysis of an exhibition of photographs that included images of people with HIV/AIDS which was shown by the Museum of Modern Art in New York argued that such visual images serve not only to invade the privacy of those pictured, and to give the impression that HIV/AIDS is an inevitable death sentence, but also personalize the disease in such a way as to obscure the social conditions of the epidemic (Crimp, 1992). The photographs overwhelmingly portrayed people with HIV/AIDS as 'ravaged, disfigured, and debilitated by the syndrome; they are generally alone, desperate, but resigned to their "inevitable" deaths' (1992: 118). The argument here is similar to that of Gilman's: visual imagery of disease serves to safely define the Other, to create a stereotype, and in doing so dehumanizes the sufferer.

Iconographic representations of health, illness and disease have become more important as western societies have moved towards mass communication via the visual media, including online technologies. As noted earlier, medical 'breakthroughs', surgical techniques and the use of medical technology often receive the attention of high-rating television infotainment programmes, and often feature on the front page of newspapers, news magazines, women's magazines and popular science magazines. Computer graphics and electro-mi-

croscopes have realized in three dimensions and full colour the T cell and HIV, and intimate colour photographs of the foetus in the womb and the moment of human fertilization have become the stuff of coffee-table art books and prime-time television documentaries. Computer apps, initially designed for medical students but available for anyone to purchase, provide detailed images of the parts of the human body.

Such images, some of which allow access into the very interior of the body using the technology of the endoscope (a tiny television camera), are seductive, extending the clinical gaze into the colourful landscape of the body's interior (Vasseleu, 1991). These images vividly render the internal body as a form of 'inner space', colonized and busy with the violent activities of foreign bodies engaged in war-like activities, destabilizing the sense of perspective people have about their bodies as a self-contained, knowable part of Self (Haraway, 1992).

There seems little doubt that the scientific technologies which produce images of reproductive processes contribute to the constitution of individuals' perception of their bodies. Of concern is the way in which iconographic images of the interior of the body have been used to serve political purposes, especially in the area of struggles over abortion, assisted conception technologies and prenatal screening. Anti-abortion groups have often used the power of the visual image of the foetus to sway opinion to their argument by depicting the foetus as an independent being, unconnected to the mother except by the umbilical cord, floating in undefined space like an astronaut, and deserving of human rights as an autonomous human. Similarly, the ultrasound image allows obstetricians to take charge of the pregnancy by commandeering the first images of the foetus, applying surveillance techniques out of control of the pregnant woman, and constructing a 'patient' out of the foetal images (Petchesky, 1987). While anti-abortion ideology and medical writings contend that knowledge of the foetus, including the visual representations offered by ultrasound, contributes to mother-infant bonding, some critics argue that the focus of such screening on the foetus conversely tends to separate the woman from the foetus by portraying it as an individual identity, with its own rights and autonomy (Ginsburg, 1990; Rapp, 1990; Rowland, 1992).

Countering the Oppressive Image: Cultural Activism

With the advent of the HIV/AIDS epidemic there has been an increased interest on the part of cultural commentators in deconstructing representations of illness and sexual 'deviance'. A number of scholars who have criticized the discursive and iconographic formations around HIV/AIDS and people living with HIV/AIDS have

gone beyond criticism to shaping resistance. In an attempt to challenge oppressive stereotypes, people with HIV/AIDS have worked to resist the dominant media iconography of HIV/AIDS. The AIDS Coalition to Unleash Power (ACT UP) groups in several western countries, including the United States, Britain and Australia, have staged protests and 'media events' using clever performance art and graphics. Groups such as the London-based 'AIDS and Photography Group' have produced cultural artefacts and staged exhibitions which challenge the way HIV/AIDS has been represented in the media. Crimp has termed this work 'cultural activism' and claims that 'art does have the power to save lives, and it is this very power that must be recognized, fostered and supported in every way possible' (Crimp, 1989: 3).

People with disabilities and their advocates have also been active in attempting to overcome the negative discourses and meanings surrounding disability. Drawing attention to language use has been a central part of their activism, including contesting the words used in relation to disability. Some activists, for example, argue that 'impairment' is a better and more empowering term than 'disability', a term which itself had replaced the more stigmatizing 'handicap' (although there is by no means agreement on this issue within the disability politics movement). Many disability activists also assert that the individual who has been labelled as 'disabled' should not be described and viewed as 'disabled' or 'dependent', but rather that the social conditions and meanings surrounding them create their disability and dependence. Such people therefore experience a form of social oppression that serves to isolate them, render them deviant and impair their functioning (Shakespeare, 1994; Williams and Busby, 2000). Thus, for example, a person who cannot use her legs and is reliant upon a wheelchair is rendered 'disabled' by the lack of facilities – such as ramps and lifts – available to her to move around freely. Activists have focused attention on these issues to highlight the ways in which language and discourse are inextricably intertwined with practice and experience.

Some people who are overweight or obese have also begun to seek to counter what they see as discriminatory attitudes to fat people. Given the prevailing focus in the popular media and in medical discourses on the health risks of obesity, and on a more symbolic level the accompanying expressions of contempt and revulsion for fat bodies, 'fat activists', as some dub themselves, have sought to challenge these meanings. Other terms for these activists include the fat or size acceptance movement, the fat liberation movement and fat power. Members of these groups contest the taken-for-granted assumptions espoused by medical and popular discourses that obesity is a cause of ill-health and that losing weight will lead to better health. Fat activists claim that such discourses unfairly single out and stigmatize overweight and obese people and that emphasis is erroneously placed upon body weight in medical and public-health writings.

© Deborah Lupton 2012

The 'fatosphere' of fat activists now comprises many groups, several books and an untold number of internet blogs (with names such as 'The Big Fat Blog' and 'The F Word') and forums, all of which seek to challenge negative meanings around fat embodiment and instead replace these with positive, celebratory meanings (Meleo-Erwin, 2011). Cultural fat-activist events have included fat people joining together to perform belly dancing (Fat Chance Belly Dance), cheer leading (Fat Action All Star Spirit Squad) or burlesque dancing (Big Burlesque), in the attempt to reclaim pride in their bodies and to present their bodies to others as desirable and sexually attractive.

Some artists have used their experiences of serious illness in their work. The late British photographer Jo Spence used the creative process to challenge the mythologies of illness. Drawing upon her experience as a breast-cancer patient, Spence and her colleague Rosy Martin invented a technique called phototherapy, a synthesis of photography and aspects of psychoanalytic theory. The technique, as employed by Spence, used photography to document the diseased body and her experiences of biomedical treatment. It was concerned with the politics of representation, struggle over discourse and the creation of identity (Martin, 1997; Radley and Bell, 2007). Spence's work engages with discourses surrounding health, disease, the body, pornography, religion, fashion, ageing, death and motherhood, reflecting upon the loss of dignity and power accompanying illness and medical intervention. Her own body formed the subject of her photography where she deliberately exposed images of age, ugliness and pain. Spence used her work to empower people with disease, to resist medical dominance of the body and to render problematic the functioning of the health-care system (Spence, 1986).

In collaboration with Rosy Martin, Spence re-enacted her emotions as she underwent hospitalization for cancer: her feelings of powerlessness, the reversion to infantilism in the face of medical authority, the need to reassert control over her own body. Spence's photographs were displayed in two exhibitions: 'The Picture of Health?' (1986), which positioned health and ill-health as major political issues and critiqued the loss of dignity and power that accompanies disease and the way a cancer patient is silenced and professionally 'managed'; and 'Narratives of Disease: Ritualized Procedures' (1990), which dealt with Spence's repressed anger about her health and her ageing and surgically scarred body.

Other artists have drawn attention to the ways in which damaged women's bodies are marginalized and represented as deviant in wider society. The photographer Matuschka, for example, documented her own body in a series of self-portraits after she had undergone a mastectomy. She uses the term 'invasive' to describe both the breast cancer she experienced and the ways in which she has sought to create an awareness of the

© Deborah Lupton 2012

politics of breast cancer through her artistic and activist efforts. A self-portrait that appeared on the front page of the *New York Times Magazine* in 1993 inspired controversy because of its open depiction of her chest following surgical removal of one of her breasts. Matuschka's intention was to represent her body as heroic, beautiful, strong and dignified, countering negative representations and resisting pity (Malchiodi, 1997).

Such 'works of illness' (Radley and Bell, 2007: 386) provide a cultural space for people with illness and disease to challenge medical representations and discourses, to be seen as an active, sentient individual rather than as a passive product of medicine. More controversially, perhaps, a book and related exhibition of photographs by the well-known celebrity photographer Annie Leibovitz included images of her lover, Susan Sontag (the writer on illness as metaphor referred to earlier in this chapter) in the late stages of dying from cancer. Images of Sontag in this exhibition, which travelled to several countries from 2008, showed her in such situations as lying drugged and ill in a hospital bed, receiving chemotherapy and looking extremely wasted and sick. The exhibition also included photographs of Sontag after her death.

Such images could not be interpreted as explicitly having a political agenda, or seeking to counter medicalized or dehumanizing discourses on the ill and dying body or to take control over and reframe one's own experience of disease and surgery, as Spence and Matuschka attempted in their photographic works. Nonetheless, in the context of an exhibition otherwise featuring glossy images of celebrities, such photos are a stark memento mori, a reminder that bodily decay and death happen to even the most photogenic, famous, well-groomed body. They also may contribute to some extent to a de-mystification of dying and death for those viewers who may not yet have personally encountered these phenomena.

The French multi-media performing artist Orlan addresses notions of how women may assert power over their bodies using medical techniques – specifically, cosmetic surgery. In her long-term performance piece entitled 'The Re-Incarnation of Saint Orlan', Orlan underwent a series of cosmetic surgery procedures and video-taped each one. She instructed the surgeons to give her facial characteristics from classical paintings, including Mona Lisa's forehead, the chin of Venus in Botticelli's *Birth of Venus,* the nose of Gerard's *Psyche* and the lips of Moreau's *Europa.* These operations were later shown in a video installation, with a glass vessel containing the pieces of flesh cut from her face and Orlan herself present as a 'living work of art'. Orlan claims that she is not attempting to achieve an ideal of beauty through her operations (she chose the characteristics above because of their cultural meaning, not their beauty). Rather, she uses the idea of the plasticity of appearance as it may be achieved via medical procedures as a way of demonstrating the malleability of identity as it is experienced through embodiment. Surgical alteration has allowed her to transform her physical

© Deborah Lupton 2012

appearance in ways that she has dictated, to create a 'self-portrait' using new technologies. In doing so, she explores the pressures that are placed upon women to conform to the conventional norms of beauty (Davis, 1997).

Engaging in artistic and creative pursuits has the potential of serving several purposes for people who are ill and their supporters. Expressing their feelings and experiences creatively allows a cathartic release of the fears, confusion and anxieties surrounding serious illness, and thus fulfils a personal need. But when the work is received into the public domain, when it is exhibited, published, transmitted or screened, the personal experience of disease fulfils a collective goal of drawing attention to the ways in which linguistic and visual representations of disease constitute practices. It is rarely that the oppressed voice is heard, that the disabled or ill person has a say in the public domain. By exposing the underlying ideological bases of metaphorical and iconographic images of illness, and the linguistic practices around illness and disease, the politics of representation are uncovered and a space emerges for passiveness to give way to active agency and the opportunity to express one's views. The potential to resist oppressive and stigmatizing discourses and practices is liberated, and artistic or cultural criticism endeavours become activism.

Conclusion

The linguistic and visual representations of medicine, illness, disease and the body in elite and popular culture and medico-scientific texts are influential in the construction of both lay and medical knowledge and experiences of these phenomena. The metaphorical systems describing illness, disease and the body are important linguistic choices which are revealing of deeper societal anxieties about the control and health of the body politic as well as that of the body corporeal. Similarly, iconographic representations of the ill body are inherently political, seeking to categorize and control deviancy, valorize normality and promote medicine as wondrous and ever-progressive. As shown, common ways of conceptualizing illness or the threat of illness often incorporate imagery associated with war, fear, violence, heroism, religion, xenophobia, contamination, gender roles, vilification and control. Awareness of these latent meanings as expressed in elite, scientific and popular texts is vital for scholars and students in the humanities and social sciences who are interested in medicine as culture, and provide the basis of efforts on the part of cultural activists to resist or subvert stigmatizing or disempowering representations of the ill.

© Deborah Lupton 2012

Discussion Questions

1 List some of the major metaphors and images that have been used to conceptualize the body and illness, from past eras to today. How and why have these changed? What metaphors and images have stayed consistent over the centuries, and why?

2 Identify some other metaphors or visual ways of representing the body that are not included in the chapter.

3 Find a newspaper or magazine article about an illness or disease, and examine the language used to describe it, including broader discourses and more specific metaphors. Analyse any images used to illustrate the article for the meanings they convey about the illness or disease.

- cancer
- metaphor
- disease
- illness
- HIV/AIDS
- immune system
- patients

https://doi.org/10.4135/9781446254530

Simpkin, Arabella L. & Schwartzstein, Richard M

"Tolerating Uncertainty. The Next Medical Revolution?", i *The New England journal of medicine* 375 (18) 2016 ss. 1713-1715

Copyright (C) 2016

Kopiert med hjemmel i Kopinor-avtalen

BECOMING A PHYSICIAN

Tolerating Uncertainty — The Next Medical Revolution?

Arabella L. Simpkin, B.M., B.Ch., M.M.Sc, and Richard M. Schwartzstein, M.D.

"At once it struck me what quality went to form a Man of Achievement . . . when a man is capable of being in uncertainties, mysteries, doubts, without any irritable reaching after fact and reason."

— *John Keats, December 1817*[1]

These words penned by John Keats, who was a physician as well as a poet, remind us of the human struggle to live in a gray-scale space where uncertainty is rife — a space that is neither black nor white. Our quest for certainty is central to human psychology, however, and it both guides and misguides us.

Although physicians are rationally aware when uncertainty exists, the culture of medicine evinces a deep-rooted unwillingness to acknowledge and embrace it. Embodied in our teaching, our case-based learning curricula, and our research is the notion that we must unify a constellation of signs, symptoms, and test results into a solution. We demand a differential diagnosis after being presented with few facts and exhort our trainees to "put your money down" on a solution to the problem at hand despite the powerful effect of cognitive biases under these conditions. Too often, we focus on transforming a patient's gray-scale narrative into a black-and-white diagnosis that can be neatly categorized and labeled. The unintended consequence — an obsession with finding the right answer, at the risk of oversimplifying the richly iterative and evolutionary nature of clinical reasoning — is the very antithesis of humanistic, individualized patient-centered care.

We believe that a shift toward the acknowledgment and acceptance of uncertainty is essential — for us as physicians, for our patients, and for our health care system as a whole. Only if such a revolution occurs will we thrive in the coming medical era.

In medicine today, uncertainty is generally suppressed and ignored, consciously and subconsciously. Its suppression makes intuitive sense: being uncertain instills a sense of vulnerability in us — a sense of fear about what lies ahead. It is unsettling and makes us crave black-and-white zones, to escape this gray-scale space. Our protocols and checklists emphasize the black-and-white aspects of medicine. Doctors often fear that by expressing uncertainty, they will project ignorance to patients and colleagues, so they internalize and mask it. We are still strongly influenced by a rationalist tradition that seeks to provide a world of apparent security.

Yet the reality is that doctors continually have to make decisions on the basis of imperfect data and limited knowledge, which leads to diagnostic uncertainty, coupled with the uncertainty that arises from unpredictable patient responses to treatment and from health care outcomes that are far from binary. Key elements for survival in the medical profession would seem, intuitively, to be a tolerance for uncertainty and a curiosity about the unknown. Have we created a culture that ignores and denies that requirement? Could our intolerance of uncertainty, in turn, be contributing to the accelerating rates of burnout and the rising cost of health care? For there is no doubt that absolute truth and certainty are hard to come by in clinical medicine.

Great tensions are created by the conflict between the quest for certainty and the reality of uncertainty. Doctors' maladaptive responses to uncertainty are known to contribute to work-related stress.[2] Physicians' difficulty in accepting uncertainty has also been associated with detrimental effects on patients, including excessive ordering of tests that carry risks of false positive results or



iatrogenic injury and withholding of information from patients.[3] In addition, by attempting to achieve a sense of certainty too soon, we risk premature closure in our decision-making process, thereby allowing our hidden assumptions and unconscious biases to have more weight than they should, with increased potential for diagnostic error.

Our need to tolerate uncertainty has never been more urgent. Technology is advancing at lightning speed, and we are now able, at the touch of a button, to get instant access to a plethora of services and products. In our experience, many current medical students, the digital natives, seek structure, efficiency, and predictability; they insist on knowing "the right answer" and are frustrated when one cannot be supplied. This attitude no doubt increases the likelihood that they will perceive uncertainty as a threat. Given the growth of access to information online and electronically, students can spend less time at the bedside in the gray-scale world of medicine and more time in front of a screen absorbing processed and general information rather than immediate and idiosyncratic realities. Their online experience may reinforce their sense of a black-and-white world where certainty is readily achievable — the antithesis of the perspective they will need to thrive in 21st-century medicine.

We believe that cultivating a tolerance of uncertainty, and addressing the barriers to this goal for physicians, patients, and the health care system, will require a revolutionary change in medicine's cultural attitude and approach to uncertainty. Our curricula (formal, informal, and hidden), assessments, and evaluations will need to be modified to emphasize reasoning, the possibility of more than one right answer, and consideration of our patients' values. Educators can start by asking questions that focus on "how" and "why," not "what" — stimulating discussion that embraces the gray-scale aspects of human health and illness, aspects that cannot be neatly categorized, and encouraging students' curiosity to explore and capacity to sit comfortably with uncertainty, acknowledging that certainty is not always the end goal.

We can speak about "hypotheses" rather than "diagnoses," thereby changing the expectations of both patients and physicians and facilitating a shift in culture.

Our curricula should recognize diagnosis as dynamic and evolving — an iterative process that accounts for multiple, changing perspectives. We can speak about "hypotheses" rather than "diagnoses," thereby changing the expectations of both patients and physicians and facilitating a shift in culture. This shift may entail discussing uncertainty directly with patients, intentionally reflecting on its origins — subjectivity in the illness narrative, diagnostic sensitivity and specificity, unpredictability of treatment outcomes, and our own hidden assumptions and unconscious biases, to name a few. We can then teach physicians specifically how to communicate scientific uncertainty, which is essential if patients are to truly share in decision making, and we can reduce everyone's discomfort by reframing uncertainty as a surmountable challenge rather than as a threat.

In keeping with these curricular goals, our assessments of students can reflect the gray-scale environment, shifting away from the black-and-white multiple-choice questions that are all too common in our exams and that inculcate in students the belief that there is always a right answer. We need to focus on evaluating clinical reasoning and the demonstration of tolerance for uncertainty.

As we move further into the 21st century, it seems clear that technology will perform the routine tasks of medicine for which algorithms can be developed. Our value as physicians will lie in the gray-scale space, where we will have to support patients who are living with uncertainty — work that is essential to strong and meaningful doctor–patient relationships. It is therefore critical that we focus on thriving in this space and changing our professional culture to allow for uncertainty. As faculty, we will have to model for our students the practice of medicine in which it is all right to be uncertain — perhaps reminding ourselves of Osler's maxim that "medicine is a science of uncertainty and an art of probability."[4] Ironically, only un-

certainty is a sure thing. Certainty is an illusion.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the Division of General Internal Medicine, Massachusetts General Hospital (A.L.S.), Harvard Medical School (A.L.S., R.M.S.), and the Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center (R.M.S.) — all in Boston.

1. Forman MB. The letters of John Keats. Oxford, United Kingdom: Oxford University Press, 1931.
2. Logan RL, Scott PJ. Uncertainty in clinical practice: implications for quality and costs of health care. Lancet 1996;347:595-8.
3. Kassirer JP. Our stubborn quest for diagnostic certainty: a cause of excessive testing. N Engl J Med 1989;320:1489-91.
4. Bean RB, Bean WB. Sir William Osler: aphorisms from his bedside teachings and writings. New York: Henry Schuman, 1950.

DOI: 10.1056/NEJMp1606402
Copyright © 2016 Massachusetts Medical Society.

Larsen, Øyvind; Keeping, David & Nylenna, Magne

"Kap. 2.2: Helsetilstanden i Norge", i

Samfunnsmedisin på norsk 2019 ss. 71-93

Copyright (C) 2019 Gyldendal

Kopiert med hjemmel i Kopinor-avtalen

2.2

Helsetilstanden i Norge

Sykdom og død er uunngåelige hendelser i et menneskeliv. Kunnskap om hvilke sykdommer som rammer en befolkning, er en forutsetning både for forebyggende innsats og for planlegging av helsetjenester. Det kan være store forskjeller mellom de sykdommer, skader og ulykker «folk dør av», og de sykdommer og tilstander «folk lever med». Sykdomsmønsteret varierer mellom land og befolkningsgrupper, og endrer seg over tid. Historisk er infeksjonssykdommer i vår del av verden langt på vei erstattet av kroniske sykdommer. Økende levealder medfører nye helseproblemer.

De langsomme endringene

Mange av endringene i samfunnet skjer så langsomt at utviklingen må følges over flere generasjoner for at vi skal kunne kartlegge sammenhenger i det som skjer.

Betegnelsen *demografisk transisjon* brukes om overgangen fra høy dødelighet kombinert med en tilsvarende høy fødselsrate, til en forholdsvis stabil og lav dødelighet og en tilsvarende lav fødselsrate. Først synker dødeligheten, deretter synker fødselsraten. Denne overgangen er vanligvis relatert til industrialisering og teknologisk utvikling. I Norge var både fødsels- og dødsraten høy fram til dødeligheten begynte å synke tidlig på 1800-tallet. Fødselsraten avtok først femti år senere.

For å knytte endringer i sykdom og helse til den demografiske utviklingen ble *epidemiologisk transisjon* introdusert som begrep av egypteren Abdel Rahim Omran (1925–1999) i 1971 (Omran 1971). Omran beskrev tre basale faser i en epidemiologiske transisjon:

- en fase med pest og hungersnød, der dødeligheten er høy og levealderen meget lav

- en fase med færre pandemier, der levealderen øker jevnt og befolkningsøkningen etter hvert blir sterk
- en fase preget av degenerative og menneskeskapte sykdommer, der dødeligheten fortsetter å avta og til slutt nærmer seg stabilitet på et relativt lavt nivå

Transisjonsbegrepet kan også brukes om andre utviklingstrekk i samfunnet som har helsemessige konsekvenser. Vi kan snakke om en *hygienisk transisjon* i Norge i det 20. århundret, og kanskje også om en *velferdstransisjon*, der det ikke er åpenbart at utviklingen går i en bestemt retning (Larsen 2014a).

Helse og sykdom

Helsetjenesten har, navnet til tross, alltid vært mer opptatt av sykdom enn av helse. Det er allmenn enighet om at de to begrepene ikke automatisk er motparter. Iblant brukes heller betegnelsen uhelse som det motsatte av helse (Gjersvik & Hem 2014). Mange mennesker beskriver helsen sin som god selv om de har mange diagnoser og slik sett også mange og kanskje endatil alvorlige sykdommer rent medisinsk sett.

Det er gjort mange forsøk på å definere helse. Alle kjenner definisjonen til Verdens helseorganisasjon fra 1946, som sier at helse er «en tilstand av fullstendig fysisk, mentalt og sosialt velvære og ikke bare fravær av sykdom og lyte». Inkluderingen av sosialt velvære var en ny og radikal utvidelse av helsebegrepet som hadde åpenbare politiske føringer. Helse er evne til å mestre og tilpasse seg livets mange vanskeligheter, sa professor Peter F. Hjort (1924–2011). I boka *Hva er helse* presenterer John Gunnar Mæland flere tilnærminger til begrepet helse (Mæland 2009). Alle tilnærmingene bygger på at helse er et flertydig begrep. En av definisjonene som presenteres, bygger på en tredeling av helsebegrepet:

- helse som fravær av sykdom, altså som fravær av symptomer og plager
- helse som likevekt, det vil si en tilstand av harmoni og velvære
- helse som reservoar, som en ressurs

Samfunnsmedisineren Per Fugelli og antropologen Benedicte Ingstad undersøkte i perioden 1997–2000 helseoppfatningen i Norge ved å gjennomføre 80 intervjuer rundt om i landet. De konkluderte med tre kjennetegn ved det folkelige perspektivet på helse (Fugelli & Ingstad 2001):

- helhet, noe som er «vevd inn i alle sider av livet og samfunnet»
- pragmatisme, det vil si at helse er et relativt fenomen som vurderes ulikt i ulike faser av livet

- individualisme, altså at helse er et personlig fenomen som varierer fra individ til individ

De dokumenterte at det er mye annet enn det snevert medisinske som går inn i nordmenns oppfatning av hva god helse er. Det utkrystalliserte seg ti punkter, hvorav bare ett er fravær av sykdom, slik man tradisjonelt har tenkt. Forfatterne fant at nordmenn opplever helse som trivsel, likevekt, helhet, tilpasning, funksjon, å bevege seg mot mål, å ha livskraft, motstandskraft, fravær av tanker om helse – i tillegg til fravær av sykdom. Et særtrekk ved den norske helseoppfatningen er dessuten forholdet til naturen, og i så måte skiller vi oss kanskje fra andre land (Fugelli & Ingstad 2009).



Å høre til i naturen og å ferdes i naturen, henger sammen med god helse hos nordmenn flest. Sammenhengen mellom natur og helse ser ut til å være et særnorsk fenomen, og kan også forklare at færre her i landet enn andre steder ser på helse som en vare som kan kjøpes for penger (Fugelli & Ingstad 2001). (Fra Kvaløya, Troms, foto Ø.L. 2016)

Disease, illness, sickness

«Illness is what you have when you go to the doctor. Disease is what you have when you leave the doctor.» Ian McWhinney (1926–2012)

Sykdomsbegrepet er mangfoldig og har flere dimensjoner. Internasjonalt skilles det mellom tre ulike sykdomsbegreper, nemlig *disease*, *illness* og *sickness*. I norsk terminologi foreslås det slike utdypninger av disse begrepene (Hofmann & Wilkinson 2016):

- disease, den profesjonelle, medisinske klassifisering av sykdom, fortrinnsvis knyttet til en diagnose. Personen har en «sykdom»;
- illness, den subjektive opplevelse av sykdom. Man kjenner seg syk;
- sickness, den sosiale sykerollen. Den er for eksempel knyttet til sykmelding.

Hva som oppfattes og karakteriseres som sykdom, varierer både over tid og mellom ulike kulturer. Man kan være «litt syk» og «veldig syk». Denne graderingen kan skje etter ulike akser: den subjektive, basert på hvordan man føler seg, den mer objektive, basert på den medisinske prognosen, og basert på funksjonsnivået, hvordan man fungerer i dagligliv eller arbeidsliv. Med tiden har den subjektive helseopplevelse fått økende betydning.

Leksikalsk kan sykdom defineres som «fellesbetegnelse på tilstander som kjennetegnes ved forstyrrelser i kroppens normale organiske eller mentale funksjoner og forandrer dem på en skadelig måte» (Store medisinske leksikon). Men hva er egentlig normale funksjoner? Og hvor store skal «forstyrrelser» være før de kan kalles sykdom? For mange medisinske tilstander er det ingen entydig grense mellom det «normale» og det «patologiske». Normalitet kan defineres både som det ønskelige, det optimale, og som det vanlige, det statistisk gjennomsnittlige.

I dag avgrenses som oftest sykdom av diagnosesystemer der klassifiseringen skjer på bakgrunn av (subjektive) symptomer og (objektive) tegn (Faiz 2018). Mens leger fram til midten av 1800-tallet var mer opptatt av å klassifisere sykdom basert på symptomer, er interessen i dag først og fremst knyttet til etiologi (sykdomsårsak). Hovedhensikten med å stille en diagnose er å gi grunnlaget for å velge den best egnede behandlingen. Behandling av årsaken til en sykdom er som regel det mest effektive. At det finnes en behandling for en uønsket tilstand, at noe kan gjøres med den, kan i seg selv være grunn til at tilstanden oppfattes som sykdom. Et eksempel på dette er ufrivillig barnløshet. Muligheten for assistert befruktning førte til at ufrivillig barnløshet av mange ble oppfattet som en sykdom (Hofmann 2014). Det

finnes imidlertid også mange andre, ikke behandlingsorienterte, motiver for å stille en diagnose: Diagnoser kan gi ulike ytelser og rettigheter, diagnoser kan brukes til prioriteringsformål, og diagnoser kan gi fritak for plikter og straffansvar.

Med økende oppmerksomhet omkring risikofaktorer for sykdom, for eksempel blodtrykk og kolesterolnivå, problematiseres oppfatningen av sykdom som avvik fra en norm, «det normale». Samtidig utviskes forskjellen mellom å ha en sykdom og å ha økt risiko for å få en sykdom. Det normale defineres ofte som det vanlige, hva som er statistisk normalt. Et slikt utgangspunkt betyr for eksempel at normalvekten hos voksne varierer over tid og mellom ulike land og befolkningsgrupper. Men det normale kan også defineres som det ønskelige. Når det ønskelige endres, endres også det normale. Med en innsnevring av normalitet for både fysiologiske og mentale forhold økes sykeligheten i befolkningen. Det viser det relative i sykdomsbegrepet.

Helseparadokset

I 1988 introduserte den amerikanske psykiateren Arthur Barsky (f. 1943) begrepet *helseparadokset* (Barsky 1988). Til tross for økende levealder og en betydelig bedring i den objektive helsen hos folk flest, observerte han en subjektiv forverring og en økende bekymring for egen helse hos stadig flere personer. Han identifiserte fire hovedgrunner til dette:

- Bedre behandling fører til at flere lever med kroniske sykdommer.
- En økende bevissthet om sykdom og helse i samfunnet gjør den enkelte mer oppmerksom på sine symptomer.
- Skremselspropaganda skaper angst og bekymring.
- Medikaliseringen av dagliglivet gir folk urealistiske forventninger om et liv uten plager og problemer.

Denne utviklingen har fortsatt, og gjør det vanskelig å skille objektive helsekriterier fra subjektive helsebekymringer.

I helsetjenesten knyttes sykdom til diagnoser. En tilstand som har en diagnose, blir en sykdom. Slik får diagnoselistene og diagnosesystemene en stor betydning for sykdomsoppfatningen. Flere diagnosesystemer brukes i dagens norske helsetjeneste. I spesialisthelsetjenesten brukes først og fremst International Classification of Diseases (ICD), et internasjonalt kodeverk til å klassifisere og registrere sykdommer og helseproblemer. ICD forvaltes av Verdens helseorganisasjon, og ble opprinnelig etablert som et system for å klassifisere dødsårsaker på slutten av 1800-tallet.

Etter flere revisjoner var sjette utgave (ICD-6), som ble publisert i 1949, den første som var egnet til detaljert diagnoseklassifisering. Senere er det gjort flere revisjoner, og fra 1990-tallet benyttes en norsk versjon av ICD-10 her i landet. Verdens helseorganisasjon har siden år 2000 arbeidet med nok en revisjon, og den første utgaven av ICD-11 ble publisert i juni 2018 for å gi medlemslandene mulighet til å begynne planleggingen av innføring av en ny versjon. Som et supplement til ICD for psykisk helse brukes Diagnostic and Statistical Manual of Mental Disorders (DSM), som utarbeides av the American Psychiatric Association. I primærhelsetjenesten brukes International Classification of Primary Care (ICPC) til å klassifisere diagnoser. Dette systemet, som også inkluderer koder for kontaktårsaker og flere uklare helseproblemer enn ICD, ble utarbeidet av World Organization of Family Doctors (Wonca) på 1980-tallet, og dagens versjon, ICPC-2, ble publisert rundt årtusenskiftet. Diagnosekoder etter ICPC benyttes blant annet ved sykmeldinger og legeerklæringer til NAV.

Noen av diagnosene i de forskjellige klassifiserings- og kodingssystemene har klare kriterier, men for de fleste bygger klassifiseringen på skjønnsmessige vurderinger. Selv for tilstander med internasjonale, standardiserte kriterier basert på biomarkører, for eksempel akutt hjerteinfarkt, er ikke diagnostikken uproblematisk (Atar 2012).

Er sorg en sykdom?

Diagnosen *nostalgi* ble skapt på 1600-tallet av Johannes Hofer (1669–1752,) og ble lenge betraktet som en alvorlig sykdom. «Motesykdom» brukes som betegnelse på tilstander som i en tidsperiode oppfattes som sykdommer, og gjerne sprer seg raskt i avgrensede befolkningsgrupper, men som senere blir borte. Eksempler er *nevrasteni* omkring forrige århundreskifte, *colitis* på begynnelsen av 1900-tallet, og kanskje også *fibromyalgi* for noen tiår siden (Lillestøl & Bondevik 2013).

Homofili var en psykisk diagnose her i landet til 1977. Nå er det store variasjoner mellom ulike land i holdningen til *ADHD*, i bruken av denne diagnosen og i legemiddelbehandling av den.

Dette viser hvordan sykdomsoppfatning endrer seg i tid og rom. I vår tid diskuteres det om *sorg* og *ensomhet*, som alltid har vært, og fortsatt er, tunge opplevelser, også er medisinske problemer og kanskje endatil burde være diagnoser. Sorg er en normal, i betydningen vanlig og forventet, reaksjon på tap. I dag er «forlenget sorg-reaksjon» på vei inn i diagnoselistene. Det lages kriterier for sorgdiagnosen og for forskjellen mellom sorg og depresjon (Kristensen et al. 2017).

Noe av grunnen til diskusjonen om sykdomskriterier er at diagnoser er en forutsetning for tildeling av økonomiske goder, for eksempel en sykmelding. Gjennom livet åpner diagnoser dører til ulike tjenester, fra spesialundervisning i skolen til parkeringsplasser for funksjonshemmede, fra trygdeytelser til institusjonsplasser.

Helse og sykdom er relative begreper, og den subjektive opplevelsen og mestringen av helseproblemer vil alltid være vanskelig å klassifisere.

Sykdomsbyrde

I befolkningssammenheng er det behov for begreper og metoder som gjør det mulig å sammenlikne informasjon om sykdom og skade på globalt og nasjonalt nivå, og å følge utviklingen. Den vanligste måten å måle sykdomsbyrde på er å studere dødeligheten; hva dør folk av? Kombineres dette med alder ved død, kan antall tapte leveår beregnes for hver enkelt sykdom. En annen måte å måle sykdomsbyrde på er å studere antall leveår med helsetap. Det som nå regnes som den beste målemetoden, er å beregne summen av tapte leveår som følge av tidlig død og antall leveår med sykdom. Ved å kombinere dødelighet med helsetap, kan antall tapte gode leveår beregnes ved hjelp av *DALY (Disability-Adjusted Life Year)*, helsetapsjusterte leveår. DALY er summen av tapte leveår på grunn av «for tidlig» død, og helsetap fra ikke-dødelige sykdommer.

Global Burden of Disease er et omfattende internasjonalt sykdomsbyrdeprosjekt som ble etablert på 1990-tallet, og som fra 2007 koordineres fra *Institute for Health Metrics and Evaluation* ved University of Washington, i Seattle, USA. Formålet med prosjektet er å kartlegge sykdomsbyrden og utviklingen av denne på verdensbasis (Murray & Lopez 2017). Sykdomsbyrden baseres på data for død og helsetap fra 195 land og territorier knyttet til over 300 sykdommer og omtrent 80 risikofaktorer. Prosjektet benytter fire hovedmål på sykdomsbyrde: antall dødsfall, tapte leveår, ikke-dødelig helsetap og helsetapsjusterte leveår (DALY).

I Norge har *Folkehelseinstituttet* ansvar for å holde oversikt over helsetilstanden i befolkningen. I 2015 ble *Senter for sykdomsbyrde* etablert i Folkehelseinstituttet. Dette senteret skal gi en overordnet analyse av forekomsten av dødsårsaker og sykdommer, beregne sykdomsbyrden, som helsetap og tapte leveår, og belyse hvilke risikofaktorer som har størst betydning for sykdomsbyrden i Norge. *Senter for sykdomsbyrde* samarbeider nært med Global Burden of Disease og publiserer jevnlige rapporter og oversikter.

DALY (disability-adjusted life year) – helsetapsjusterte leveår

DALY ble introdusert på 1990-tallet av Verdensbanken og Verdens helseorganisasjon som et mål for tapte gode leveår i en befolkning. Dette målet brukes både til å følge utviklingen i en befolkning og til internasjonale sammenlikninger.

Beregningen bygger på et antatt helsetap for ulike tilstander eller diagnoser som er estimert av et ekspertpanel og fastsatt på en skala fra 0 (perfekt helse) til 1 (død). Slik kan helsetapet for sykdommer i en befolkning beskrives. For et bestemt år teller man opp antall dødsfall og anslår hvor mange leveår som gikk tapt i forbindelse med disse dødsfallene, basert på alder ved død. Dessuten anslår man hvor mange mennesker som samme året levde med ulike sykdommer. Hvert av disse levde årene multipliseres med det estimerte helsetapet som ekspertpanelet har fastsatt for den aktuelle sykdommen eller skaden.

«En DALY» kan betraktes som et tapt sykdomsfritt år. For en bestemt sykdom eller tilstand kalkuleres DALY ved å addere antall tapte leveår på grunn av for tidlig død og antall år tapt på grunn av funksjonsnedsettelse (beregnet som relativt helsetap multiplisert med sykdomsvarighet).

Summen av antall DALY i en befolkning er et mål på gapet mellom den aktuelle helsetilstanden og en ideell tilstand der hele befolkningen lever uten sykdom og funksjonshemning til maksimal forventet levealder.[25]

Såkalt «aldersvekting» av DALY, der sykdomsfrie leveår hos yngre får større vekt enn hos eldre, har vært omdiskutert, og Verdens helseorganisasjon avviser nå slik diskontering.

De ti viktigste årsakene til helsetapsjusterte leveår (DALY) i Norge forklarer 43 % av sykdomsbyrden i landet. Selv om tilstanden ikke er dødelig, er korsrygg- og nakkesmerter på grunn av utbredelsen den desidert viktigste årsak til helsetapsjusterte leveår (DALY) i Norge. Korsrygg- og nakkeskader forårsaker alene over 10 % av DALY. Iskemisk hjertesykdom kommer på andreplass og forårsaker 7,1 % av DALY, mens de neste åtte årsakene er omtrent like store, på rundt 3 % hver.

25 http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/

Tabell 2 De ti viktigste årsakene til helsetapsjusterte leveår (DALY) i Norge i 2015 (Knudsen et al. 2017).

		Antall DALY	Andel (%)
1	Korsrygg- og nakkesmerter	126 517	10,8
2	Iskemisk hjertesykdom	83 568	7,1
3	Kreft i luftrør, lunge og bronkie	40 657	3,5
4	Alzheimers sykdom og andre former for demens	40 373	3,4
5	Sykdommer i sanseorgan	39 921	3,4
6	Depressive lidelser	38 271	3,3
7	Karsykdom i hjernen	36 890	3,1
8	Kronisk obstruktiv lungesykdom	35 016	3,0
9	Angstlidelser	32 434	2,8
10	Diabetes mellitus	31 586	2,7

Den samlede sykdomsbyrden i Norge målt ved hjelp av DALY, har falt de seneste tiårene til tross for at befolkningen er blitt både større og eldre. Dette skyldes først og fremst en sterk reduksjon i hjerte- og kardødelighet. Samtidig er betydningen av Alzheimers sykdom og andre demensformer økt.

Insidens og prevalens

Utbredelsen av sykdom beskrives i form av insidens og prevalens.

Insidens er hyppigheten av en sykdom oppgitt som antall nye tilfeller av sykdommen i en befolkningsgruppe per tidsenhet, for eksempel per år.

Prevalens er forekomsten av en sykdom oppgitt som den andelen av en gitt befolkning som har sykdommen på et visst tidspunkt (*punktprevalens*). Betegnelsen *livstidsprevalens* brukes iblant om den andelen av en gitt befolkning som noen gang i sin levetid får en bestemt sykdom.

Akutte, kortvarige sykdommer har gjerne høy insidens og lav prevalens, mens kroniske sykdommer kan ha lav insidens og høy prevalens.

Hva dør vi av?

Dødsalder og dødsårsaker har tradisjonelt vært regnet som de viktigste indikatorene på helsetilstanden i en befolkning. Allerede i kirkebøkene fra 1600-tallet finnes slik informasjon, men det var først fra midten av 1800-tallet at det ble utarbeidet landsdekkende statistikk basert på informasjon fra legene. Den første «Beretningen

om Sundhedstilstanden og Medicinalforholdene i Norge» gjelder for 1853. I 1925 fikk Statistisk sentralbyrå ansvaret for medisinalstatistikken, og i 2002 overtok Folkehelseinstituttet ansvaret for dødsårsaksregisteret (Gjertsen 2002). Dødsårsaksregisteret reguleres av en egen forskrift (forskrift om innsamling og behandling av helseopplysninger i Dødsårsaksregisteret – dødsårsaksregisterforskriften), og alle leger har plikt til å melde personopplysninger om den avdøde og opplysninger om dødsårsak på et spesielt skjema. Klassifiseringen og kodingen av dødsårsaker er standardisert av Verdens helseorganisasjon, og som medlem har Norge forpliktet seg til å utarbeide dødsårsaksstatistikk i overensstemmelse med International Classification of Diseases (ICD) og regler for koding av underliggende dødsårsak. Med underliggende dødsårsak menes den sykdom eller skade som startet rekken av de sykelige tilstander som ledet direkte til døden, eller de ytre omstendigheter ved den ulykken eller voldshandlingen som var årsak til den dødelige skaden (Gjertsen 2000). Slik kan registeret også benyttes til internasjonale sammenlikninger.

Letalitet og mortalitet

Flere begreper brukes for å beskrive dødelighet. Letalitet beskriver hvor ofte en bestemt sykdom eller diagnose fører til døden, og er et uttrykk for en sykdoms alvorlighetsgrad. Mortalitet er egentlig insidensen av død, og beskriver altså dødshyppigheten i en befolkning. Mortalitet kan angis for enkeltårsaker (årsaksspesifikk mortalitet) eller samlet (totalmortalitet). Sjeldne sykdommer med høy letalitet påvirker mortaliteten i befolkningen lite.

Endringene i dødelighet og dødsårsaker viser utviklingen både i samfunnet og i helsetjenesten. Fra de første dødsårsaksstatistikkene ble publisert midt på 1800-tallet, er dødeligheten i Norge kraftig redusert. Forventet levealder ved fødsel har økt med godt over 30 år fra de første dødsårsaksstatistikkene til i dag, og samtidig er også dødsårsakene endret. Mens det for hundre år siden var infeksjonssykdommer, særlig tuberkulose, som krevde flest liv, er det nå hjerte- og karsykdom og kreft som er de dominerende dødsårsaker (Pedersen AG 2007). Særlig dramatisk har reduksjonen i spedbarnsdødeligheten vært. I 1901–1905 døde 80 barn per 1 000 levendefødte i løpet av første leveår (Pedersen 2003). I 2016 var spedbarnsdødeligheten 2,2 per 1 000 levendefødte barn.[26]

26 https://www.ssb.no/befolkning/statistikker/dode/aar/2017-03-09

Dødsraten – antall dødsfall per 100 000 nordmenn – er redusert de siste årene, og avspeiler at vi gjennomsnittlig lever lenger. I 2017 døde vel 40 000 personer i Norge, og for første gang var det flere som døde av kreft enn av hjerte- og karsykdommer.[27] Dødeligheten av hjerte- og karsykdommer har sunket kraftig siden 1970-tallet. Det skyldes både endrede levevaner, først og fremst mindre røyking, og bedre behandling. Omtrent hvert femte dødsfall skjer hos personer under 70 år, og under 5 % før 50-årsalder. Etter kreft og hjerte- og karsykdom følger lungesykdommer og demens som de viktigste dødsårsakene.

Tabell 3 Dødsårsaker i Norge 2017 (kilde: Dødsårsaksregisteret)

Alle dødsfall	**40 678**
Dødsfall av sykdommer	**37 441**
Infeksiøse inkl. parasittære sykdommer	910
Svulster	11 123
Sykdommer i blod og bloddannende organer, forstyrrelser i immunsystem	147
Indresekretoriske og ernærings-/stoffskifte sykdommer	1 071
Psykiske lidelser og atferdsmessige forstyrrelser	3 067
Sykdommer i nervesystemet og sanseorganene	2 220
Sykdommer i sirkulasjonsorganene (hjerte- og karsykdommer)	10 370
Sykdommer i åndedrettsorganene	4 615
Sykdommer i fordøyelsesorganene	1 296
Sykdommer i hud og underhud	118
Sykdommer i skjelett, muskler og bindevev	264
Sykdommer i urin- og kjønnsorganer	887
Komplikasjoner i svangerskap, fødsel og barsel	-
Visse tilstander som oppstår i perinatalperioden	62
Medfødte misdannelser og kromosomanomalier	135
Symptomer og ubestemte tilstander	1 156
Skader og forgiftninger (voldsomme dødsfall)	**2 478**
Mangler dødsmelding	**759**

27 https://www.fhi.no/nyheter/2018/dodsarsakene-2017/

Ikke-smittsomme sykdommer

Ikke-smittsomme sykdommer (non-communicable diseases, NCD) er en fellesbetegnelse på ikke-infeksiøse sykdommer som er knyttet til livsstil og levevaner. Hjerte- og karsykdommer, kreft, kroniske lungesykdommer og diabetes står på verdensbasis for 70 % av alle dødsfall.[28] I Norge skyldes to tredeler av alle dødsfall før 75 års alder disse sykdommene (Helsedirektoratet 2013a).

Selv om disse sykdommene også kan oppstå hos enkeltpersoner med de mest helseriktige levevaner, er de epidemiologisk knyttet til risikofaktorer som røyking, lav fysisk aktivitet, overvekt og alkoholkonsum. Derfor ble og blir sykdommene ofte kalt livsstilssykdommer. De fleste ikke-smittsomme sykdommer forekommer hyppigere hos personer med lav enn hos personer med høy sosioøkonomisk status (basert på utdanning, inntekt og yrke). Det betyr at det største potensialet for å redusere sykelighet og dødelighet av disse sykdommene ligger i forebyggende tiltak som er effektive også i grupper med lavere sosioøkonomisk status (Helsedirektoratet 2013a). I en samfunnsmedisinsk sammenheng er ikke-smittsomme sykdommer særlig interessante fordi forebyggingspotensialet er stort gjennom grupperettede tiltak.

Det er definert konkrete mål for arbeidet mot ikke-smittsomme sykdommer, blant annet at for tidlig død (det vil si under 70 år) av slike sykdommer skal reduseres med 25 % innen år 2025. Verdens helseorganisasjon har utarbeidet 25 indikatorer på utviklingen.[29]

Også psykiske lidelser er inkludert i begrepet ikke-smittsomme sykdommer i Verdens helseorganisasjons europeiske handlingsplan, blant annet fordi det antas å være store synergieffekter i det forebyggende arbeidet.

Hva lever vi med?

Sykdommer som ikke medfører død, men som likevel kan være svært belastende å leve med, har fått langt mindre oppmerksomhet enn dødsårsaker i oversikter over helsetilstanden i en befolkning. Det skyldes selvsagt at kartleggingen er mye vanskeligere. Det finnes ingen komplette registre over forekomsten av alle sykdommer, og i tillegg til at samme sykdom kan ramme ulikt hos ulike personer, er også den individuelle mestringsevnen forskjellig.

28 http://www.who.int/ncds/en/
29 https://www.fhi.no/nettpub/ncd/sammendrag/oversikt-indikatorer2/

Sykdomsbyrdeprosjektet forsøker gjennom sine estimater av antatt helsetap ved ulike sykdommer og tilstander å gjøre kvantifiseringer og sammenlikninger mulig. Men fortsatt er den reelle forekomsten av mange sykdommer ukjent, og anslagene av helsetap-omfanget blir tilsvarende usikre. Det som er sikkert, er at bedre medisinske behandlingsmetoder og en aldrende befolkning har ført til at stadig flere mennesker lever med kroniske sykdommer av både ikke-dødelig og potensielt dødelig karakter.

I motsetning til dødeligheten som domineres av sykdommer som først og fremst forekommer i høy alder, som hjerte- og karsykdom og kreft, preges det som karakteriseres som ikke-dødelig helsetap, av sykdommer som rammer i yngre alder. De to viktigste sykdomsgruppene for ikke-dødelig helsetap var i 2015 muskel- og skjelettsykdom og psykiske lidelser, som forårsaket henholdsvis 25,9 % og 20,2 % av helsetapet. Psykiske lidelser var den viktigste årsaken fra ungdomstid til ung voksen alder (fra 10 til 34 år), mens muskel- og skjelettsykdom, først og fremst korsrygg- og nakkesmerter, dominerte fra 35 års alder og oppover (Senter for sykdomsbyrde 2017). Muskel- og skjelettlidelser, som i denne forbindelse inkluderer både revmatiske sykdommer (som leddgikt og artrose) og osteoporose, er den vanligste årsak til sykefravær og uføretrygd hos voksne i yrkesaktiv alder. Et særnorsk fenomen i denne sykdomskategorien er for øvrig en høy forekomst av hoftebrudd, særlig hos kvinner.

Tabell 4 De ti viktigste årsakene til ikke-dødelig helsetap i Norge i 2015 for begge kjønn og alle aldre samlet (Senter for sykdomsbyrde 2017).

		Andel (%)
1	Korsrygg- og nakkesmerter	20,3
2	Sykdommer i sanseorganene	6,4
3	Depressive lidelser	6,2
4	Angstlidelser	5,2
5	Sykdommer i hud og underhud	4,7
6	Migrene	4,3
7	Diabetes mellitus	3,7
8	Tann- og munnsykdommer	3,5
9	Andre muskel- og skjelettsykdommer	2,9
10	Fallskader	2,8

Vi lever altså lenger enn før, men for hvert vunnet leveår vil en del av det året, gjennomsnittlig 1–2 måneder, være preget av sykdom. Dette medfører et gradvis økende behov for helsetjenester, eventuelt andre tiltak som kan bedre mestringsevnen.

Sykelighet

Studiet av sykelighet har flere dimensjoner (Larsen 2012):

- sykelighetens ulike former: naturlig sykelighet, opplevd sykelighet, diagnostisert sykelighet, omformet sykelighet (i statistikker o.l.), håndtert sykelighet
- sykelighetens alvorlighetsgrad: symptomer, funn, funksjonsnivå, dødelighet
- sykelighetens forvaltning: pasienter, befolkning, helsepersonell, helsemyndigheter, politikere

Barn og unge

Barn er ikke små voksne, men har særtrekk og følsomhet som gjør at de reagerer annerledes på påkjenninger. Tilstander i barndommen kan få følger for et langt liv. Selv om det fortsatt er barnefattigdom i Norge, er det lenge siden sult og underernæring var de største helseproblemene. Nå er utfordringene først og fremst knyttet til psykososiale forhold.

Det anslås at 15–20 % av barn og unge under 18 år har nedsatt funksjon på grunn av psykiske vansker, og at ca. 8 % av alle mellom 3 og 18 år på et gitt tidspunkt har en psykisk lidelse. For de fleste er symptomene forbigående (Mykletun et al. 2009). Skolefrafall er et økende problem, særlig blant gutter.

Mye tyder på at skikkelighetstrenden som har preget ungdom de siste 10–15 årene, er i ferd med å snu. I Oslo er det nå en økning i problematferd hos de unge og en økning i cannabisbruken. Spørreundersøkelsen «Ung i Oslo 2018» blant elever i ungdomsskole og videregående skole viser at de fleste har det bra med seg selv, med foreldrene sine, på skolen og i fritiden. Men samtidig er det en økning i andelen av ungdommene som har blitt utsatt for vold og mobbing, og som har ulike helseplager. Endringene forekommer på tvers av sosioøkonomiske skiller og i alle områder av byen (Bakken 2018).

Den fysiske helsen hos barn og unge er gjennomgående god. Oppslutningen om barnevaksinasjonsprogrammet er høy, og alvorlige smittsomme sykdommer er sjeldne. Astma, allergi og atopisk eksem er de vanligste kroniske sykdommene og forekommer hos omtrent en firedel av barn i oppveksten.

Folkehelserapporten viser at forekomsten av type 1-diabetes hos barn under 15 år er høyere i Norge enn i de fleste andre land. Omtrent 200 barn under 18 år får kreft hvert år, hyppigst blodkreft og lymfekreft. Omtrent en firedel av alle skader som behandles i helsetjenesten, rammer barn under 18 år (Folkehelseinstituttet 2018).

Eldres helse

Et av de sikreste trekkene ved framskrivingen av den norske befolkningen er at vi vil få flere eldre. De fleste over 67 år rapporterer god helse, men økende alder er den største risikofaktoren for mange sykdommer. En aldrende befolkning vil derfor medføre mer sykelighet. Eldre mennesker har også flere samtidige helseproblemer enn yngre. Likevel er det usikkert hvilke endringer i befolkningens helsetilstand de demografiske endringene vil føre med seg på lengre sikt.

Den normale aldringsprosessen medfører ulike former for funksjonsnedsettelse, og det kan være vanskelig å skille «normal aldring» fra sykdommer. Hudforandringer er et synlig eksempel på hva som skjer med årene. Muskelkraften blir svakere og utholdenheten dårligere. Mye går langsommere. Også kognitive funksjoner endres. Oppmerksomhetsevnen reduseres, hukommelsen svekkes, og reaksjonstiden forlenges. Den svenske sosiologen og gerontologen Lars Tornstam (1943–2016) har utviklet en teori om psykologisk aldring, *gerotranscendens*, der det naturlige perspektivskiftet med årene står sentralt (Tornstam 2005). Mennesket har en tilpasningsevne som kan gjøre enhver alder til «den beste». Tornstam påpeker at god aldring ikke er det samme som å fortsette med middelalderens travle liv. En gerotranscendert person får et redefinert selvbilde, blir mer selektiv i valget av sosiale aktiviteter og får en ny forståelse av fundamentale, eksistensielle spørsmål. I høy alder blir man mindre interessert i materielle ting og får et større behov for meditativ tilbaketrekking, en slags god, kontemplativ ensomhet. Men dessverre stiger også risikoen for alvorlig kognitiv svikt med alderen.

Forekomsten av demenssykdom øker sterkt i høy alder, og kvinner rammes oftere enn menn, særlig etter 80 års alder (tabell 5). Etter 90 års alder har nesten halvparten av alle kvinner demens. Det er beregnet at gjennomsnittlig årlig risiko for å få demens mellom 60 og 64 år er 0,34 % per år. Dette øker til 3 % i aldersgruppen 80–84 år, og etter passerte 95 år er årlig risiko for å rammes av demens rundt 16 %. Antall personer med demens i Norge er anslått til mellom 80 000 og 100 000, og det betyr at det er omtrent 300 000 nære pårørende til personer med demens (Reneflot et al. 2018). Omtrent 80 % av beboere i sykehjem har demens i en eller annen form.

Tabell 5 Alders- og kjønnsspesifikke anslag av prosentvis demensforekomst i Vest-Europa (WHO 2012).

Kjønn	Alder							
	60–64	65–69	70–74	75–79	80–84	85–89	90+	60+
Menn	1,4	2,3	3,7	6,3	10,6	17,4	33,4	
Kvinner	1,9	3,0	5,0	8,6	14,8	24,7	48,3	
Alle	1,6	2,6	4,3	7,4	12,9	21,7	43,1	7,3

Også forekomsten av hjerte- og karsykdom øker med økende alder, og utviklingen av kreftsykdom er uløselig knyttet til aldring. «Blir vi gamle nok, får vi antakelig kreft alle sammen», kan man si (Kyte 2013).

Multisykdom hos eldre, det vil si mange samtidige diagnoser, avspeiles i legemiddelbruken. Data fra Reseptregisteret viser at 90 % av alle over 65 år fikk minst ett legemiddel på resept i 2016, og to tredeler av disse brukte fem eller flere ulike legemidler i løpet av året (Folkehelseinstituttet 2018). Samtidig bruk av mange legemidler (polyfarmasi) hos eldre er en økende utfordring.

Mer oppdatert informasjon om eldres helse finnes i kapitlet *Helse hos eldre* i Folkehelserapporten fra Folkehelseinstituttet: https://www.fhi.no/hn/eldre/.

Psykisk helse

Psykiske lidelser, inkludert rusbrukslidelser, er den viktigste årsaken til helsetap hos barn og yngre voksne. I motsetning til de fleste somatiske lidelser, debuterer vanligvis psykiske lidelser før 30 års alder (Folkehelseinstituttet 2018).

En gjennomgang av psykisk helse i Norge fra Folkehelseinstituttet konkluderte i 2018 med at utbredelsen av psykiske lidelser i den voksne befolkningen er stabil. Oversikten påpeker en bekymringsfull økning i andelen unge jenter som rapporterer et høyt nivå av psykiske plager, og som oppsøker helsetjenesten for sine plager. Det anslås at omtrent 20 % av den voksne befolkningen i Norge vil ha en psykisk lidelse i løpet av ett år. De vanligste psykiske lidelsene blant voksne er angstlidelser, depresjon og rusbrukslidelser. Depressive lidelser etterfulgt av angstlidelser er vanligste konsultasjonsgrunn for psykiske lidelser hos fastlege/legevakt i primærhelsetjenesten (Reneflot et al. 2018).

Alkohol er det vanligste rusmiddelet i Norge og det som fremkaller flest rusrelaterte lidelser. Sammenliknet med andre land har Norge en høy forekomst av overdosedødsfall, internasjonalt definert som «et dødsfall som skjer umiddelbart eller kort tid etter at en person har inntatt ett eller flere rusmidler og/eller medikamenter

og når dødsfallet har direkte sammenheng med dette inntaket». Helsedirektoratet har utarbeidet en veileder for lokalt forebyggende arbeid til kommuner med overdoseutfordringer.[30]

Pasienter med alvorlige psykiske lidelser har en betydelig overdødelighet. Forventet levetid er 15–20 år kortere for mennesker med alvorlige psykiske lidelser enn i befolkningen for øvrig, og denne forskjellen er ikke redusert de seneste tiårene. Overdødeligheten sees i alle psykiatriske diagnosegrupper og skyldes i hovedsak somatiske sykdommer som hjerte- og karsykdommer, diabetes, kreft og infeksjoner (Lien et al. 2015).

Ulykker

I samfunnsmedisinen har ulykker og skader særlig stor interesse fordi forebyggingspotensialet er så åpenbart, og fordi ulykkesforebygging involverer mange samfunnsområder. Det skilles mellom skader som skjer utilsiktet (ulykkesskader), og skader som påføres med hensikt (voldsskader og villet egenskade). Skadeomfanget ved ulykker varierer fra skrubbsår til død, men en klassifiseringsgrense kan settes ved skader som fører til legebesøk. Det er beregnet at ca. 12 % av befolkningen behandles av lege for skader i løpet av ett år, noe som utgjør vel 600 000 personer. Mange ulykker får store konsekvenser. Det anslås at det hvert år dør ca. 2 500 mennesker på grunn av skader, mens 300 000 behandles på sykehus (Folkehelseinstituttet 2018).

I dødsårsaksstatistikken brukes betegnelsen *voldsomme dødsfall* om alle dødsfall forårsaket av ytre årsaker. Det omfatter alle former for ulykker, samt drap og selvmord. Ulykker krever mange liv og er den viktigste dødsårsaken for nordmenn under 45 år. Dødsfall på grunn av ulykker er først og fremst knyttet til fall, forgiftningsulykker (inkludert overdoser) og trafikkulykker. Dødeligheten på grunn av ulykker er kraftig redusert de senere tiårene.

Få helseskader påvirkes så mye av endringer i omgivelsene som de som skyldes ulykker (Pedersen AG 2007). Dårlig sikring av arbeidsplasser og nærmiljøet førte lenge til det vi i dag vil karakterisere som unngåelige ulykker. Så sent som på 1950-tallet døde årlig rundt 100 barn under 15 år av drukning i brønner i nærheten av hjemmet. Med økende biltrafikk etter den andre verdenskrig økte trafikkulykkene og trafikkdødsfallene kraftig. Systematisk og målrettet innsats har gitt en formidabel reduksjon av både skader og dødsfall i trafikken, til tross for den store

30 https://helsedirektoratet.no/folkehelse/psykisk-helse-og-rus/overdose-veiviser-i-lokalt-forebyggende-arbeid

veksten i antall biler, kjørelengde og kompleksitet i trafikken. Konkrete tiltaksplaner utarbeidet av alle berørte etater og institusjoner, er et viktig virkemiddel (Statens vegvesen 2018). Trafikksikkerhetsarbeidet er bygget på *nullvisjonen*, en visjon om et transportsystem der ingen blir drept eller hardt skadd. Dette arbeidet er et godt eksempel på hvordan ulike tilnærminger kan kombineres i skadeforebygging i sin alminnelighet: grundig kartlegging og overvåkning, lovreguleringer, fysiske tiltak og teknologiske forbedringer, informasjon og opplæring.

Trafikksikkerhet – en suksesshistorie

Norge er nå det mest trafikksikre landet i Europa. På 1970-tallet omkom omtrent 500 mennesker årlig i trafikken. På 1980-tallet ble dette redusert til ca. 400, og nedgangen fortsatte til ca. 300 dødsfall årlig i 1990-årene. Selv om nedgangen etter hvert har gått saktere, har den fortsatt. I 2018 omkom 108 personer i trafikken. Samtidig har trafikken økt kraftig. De siste 20 årene har antall biler økt med 60 %, og både kjørelengden og hastighetene har økt.

Reduksjonen er et resultat av tiltak på mange områder. Bedre veier, med midtdelere på utsatte steder som det viktigste, er avgjørende. Sikrere biler er en annen faktor, likeså mer trafikkontroll. Sannsynligvis er også bilistene og andre trafikanter blitt mer kyndige, årvåkne og disiplinerte.

Også forekomsten av arbeidsulykker synker. Arbeidslivet har endret seg radikalt. Færre arbeider i fysisk krevende yrker, og sikkerhetstiltakene er blitt mye bedre. Et eksempel er reduksjonen i antall drukningsulykker blant sjøfolk og fiskere. Drukning forekommer nå hyppigere ved fall fra land eller fritidsbåter og i forbindelse med bading enn i arbeid. Mens det på 1970-tallet var over hundre arbeidsulykker med døden til følge hvert år, var dette i 2016 redusert til 45 dødsfall. For alle næringer samlet ble det i 2016 registrert 3,7 arbeidsulykker med langvarig fravær per 1 000 ansatte.[31] Sysselsatte i industri, bygg- og anlegg, landbruk og transport er mer utsatt for arbeidsulykker enn andre.

Selvmord

Selvmord utgjør en stor del av det som karakteriseres som voldsomme dødsfall. Det registreres årlig ca. 600 selvmord i Norge, men registreringen er usikker, og

31 https://www.ssb.no/helse/statistikker/arbulykker/aar

mange ukjente dødsårsaker kan være selvmord. Henging er den vanligste selvmordsmetoden.[32] Omtrent dobbelt så mange menn som kvinner begår selvmord. Fra slutten av 1960-årene til slutten av 1980-årene ble selvmordsraten fordoblet hos både menn og kvinner. Senere falt hyppigheten av selvmord blant menn, mens hyppigheten av selvmord blant kvinner har vært konstant de siste 20 årene. I 2015 var det 15 selvmord per 100 000 menn og 7 per 100 000 kvinner. Fortsatt er selvmord den viktigste enkeltårsaken til død i aldersgruppen 15–49 år for både kvinner og menn. Som for mange andre dødsårsaker synker selvmordsraten med økende utdanning.

Sosial ulikhet i helse

Helsetilstanden er ikke likt fordelt i befolkningen. Variasjonen med alder er åpenbar. Jo eldre, desto større sannsynlighet for sykdom og helseplager. Men ulikheten som følger andre kategorier, er like stor. Forskjellen i levealder for menn varierer opptil åtte år mellom bydeler i Oslo (Folkehelseinstituttet 2018). Slike forskjeller er ikke biologisk betinget, og må skyldes sosiale og økonomiske forhold.

Særlig sterk er sammenhengen mellom helsetilstand og utdanning, yrke og inntekt. Personer med lengst utdanning lever gjennomsnittlig fem, seks år lenger enn de som har kortest utdanning her i landet. Det er beregnet at om alle hadde hatt den samme lave dødeligheten som de best utdannede i Norge, kunne 15 000 dødsfall blant kvinner og 28 000 dødsfall blant menn vært unngått over en tiårsperiode (Dahl et al. 2014).

En analyse av dødeligheten i Norge på 1970-, 1980- og 1990-tallet viser at sosiale ulikheter i dødelighet økte i denne perioden. Både utdanning og inntekt så ut til å bety mer for menns helse enn for kvinners. Basert på dødelighetsutviklingen har det vært en bedring i helsen fra 1970-tallet til 1990-tallet, men bedringen har vært sterkere for dem med høy utdanning og høy inntekt enn for dem med lav utdanning og lav inntekt. Slik er det blitt større helseforskjeller mellom ulike sosiale grupper (Zahl et al. 2003). En gjennomgang av dødeligheten i Norge fra 2005 til 2015 viste økende forskjell i forventet levealder sammenholdt med inntekt i løpet av perioden. Forskjellen i forventet levealder var 8,4 år for kvinner og 13,8 år for menn om man sammenlikner de (1 %) som tjener aller mest med de (1 %) som tjener aller minst (Kinge et al 2019).

32 https://www.fhi.no/fp/psykiskhelse/selvmord/selvmord-og-selvmordsforsok---fakta/

En mindre påaktet, men likevel stor forskjell i helse er knyttet til det som formelt kalles sivilstatus, altså om man er gift eller ugift. Det har lenge vært kjent at de som aldri har vært gift, har høyere dødelighet enn gifte. Dette gjelder både kvinner og menn. Det bekymringsfulle er at mens forventet levealder hos gifte har steget betydelig de seneste 30 årene, har ikke ugifte hatt tilsvarende utvikling. Kombineres sivilstatus med utdanning, blir forskjellene dramatiske. På 1970-tallet hadde en ugift 50 år gammel mann med kun grunnskole knapt to år kortere forventet gjenstående levetid enn en gift mann. 30 år senere var forskjellen blitt nesten seks år. For kvinner har forskjellen økt enda mer, fra vel ett år til over seks år. Sammenliknes 50-årige ugifte personer med grunnskole med gifte personer med høyere utdanning som endatil har høyt utdannede ektefeller, er forskjellen i gjennomsnittlig forventet gjenstående levetid nesten ti år (Kravdal 2017, Kravdal et al. 2018).

Den systematiske sammenhengen mellom sosial status på den ene side og sykelighet og levealder på den annen side, betegnes som *sosial ulikhet i helse*. Ikke bare har den sosiale statusen konsekvenser for helsen. Dårlig helse kan også gi problemer med utdanning og arbeid og slik sett forverre den sosioøkonomiske situasjonen og skape en vond sirkel. Troen på at sosiale ulikheter i helse var mindre i Norge enn i andre land, ble avkreftet av en internasjonal studie før tusenårsskiftet (Mackenbach et al. 1997, Mackenbach et al. 2008).

Sosial ulikhet i helse måles ved å studere sammenhengen mellom helse og sosioøkonomiske faktorer som *utdanning*, *yrke* og *inntekt*. Disse faktorene innebærer både metodologiske styrker og svakheter. Utdanning er en mye brukt variabel fordi utdanningslengden er lett å kvantifisere, og fordi sammenhengen mellom utdanning og helseutfall er sterk. Yrkers betydning for sosial status kan være uklar og varierer over tid. Klassifisering av yrker er dessuten mer komplisert enn klassifisering av utdanning, selv om det finnes forenklede, aggregerte yrkeskategoriseringer (Skrede 1971, Erikson & Goldthorpe 2002). Skattbar inntekt er en viktig og målbar variabel, men ikke alltid dekkende for økonomisk evne, både fordi utgifter og forpliktelser varierer sterkt, og fordi noe inntekt kanskje ikke blir rapportert til skattemyndighetene.

Ulikheter i helse finnes ikke bare mellom de rikeste og de fattigste eller mellom dem med mest og minst utdanning. Ulikheter danner et jevnt trappetrinnsmønster gjennom hele befolkningen, og beskrives som en *gradientutfordring*. Personer med universitets- og høgskoleutdanning kommer bedre ut enn de med videregående skole. Og personer med videregående skole har bedre helse enn de med bare grunnskoleutdanning (Sosial- og helsedirektoratet 2005a).

Tidligere tiders tilnærming bygget på at ulikheter i helse i all hovedsak skyldtes materielle forskjeller mellom fattige og rike. Dagens faglige perspektiv er mer sammensatt, og tar utgangspunkt i at sosiale *determinanter* er hovedforklaringen på ulikhetene. Årsakene til ulikheter i helse er mangfoldige, og fortsatt delvis uutforsket. Det handler både om oppvekstvilkår og levekår, om materielle og immaterielle forutsetninger, om påkjenninger og mestringsevne, om helseatferd og om bruk av helsetjenester. Det er mange samfunnsforhold som bestemmer hvordan helsen formes og utvikles gjennom livet. Dessverre er det dem med færrest ressurser som oftest også utsettes for de største påkjenningene, og det gir til dels store helseforskjeller mellom ulike grupper. I dagens norske samfunn ser betydningen av utdanning ut til å bli stadig viktigere.

Levevaner som røyking, kosthold og fysisk aktivitet henger nøye sammen med utdannings- og inntektsnivå. Jo lengre utdanning og jo høyere inntekt, desto mer helsegunstige levevaner. Mens bare 5 % av menn mellom 25 og 74 år med universitets- eller høgskoleutdanning var dagligrøykere i 2017, var andelen 25 % i gruppen med bare grunnskoleutdanning (Folkehelseinstituttet 2018). Det er beregnet at forskjellene i levevaner forklarer 94 % av forskjellene i totaldødelighet hos menn i Norge og 59 % av forskjellene hos kvinner (Hoffmann et al. 2013).

De sosiale helseforskjellene avspeiles i forekomsten av en lang rekke sykdommer og plager. Flere personer med lav enn med høy sosioøkonomisk status rapporterer dårlig helse, langvarig smerte og muskel- og skjelettsykdommer. Personer med grunnskoleutdanning har tre ganger høyere risiko for kronisk obstruktiv lungesykdom (kols) enn personer med universitetsutdanning, og forekomsten av type 2-diabetes i denne gruppen er også høyere.

«The inverse care law»

I en legendarisk artikkel i *The Lancet* i 1971 introduserte den britiske allmennlegen Julian Tudor Hart (1927–2018) begrepet «the inverse care law», som vi kan oversette med «loven om den omvendte omsorg» (Hart 1971). I sin praksis som lege hadde han observert at folk med dårligst helse og levekår, og slik sett også med størst behov for helse- og omsorgstjenester, var de som hadde det dårligste tjenestetilbudet. Tudor Hart viste at vanlige markedsmekanismer er uegnet for å regulere tilgangen til helsetjenester, og påpekte betydningen av en god offentlig helsetjeneste som tilpasses behovet.

Tudor Hart visste at dårlige levekår var enda viktigere enn kvaliteten av helsetjenester som årsak til den høye sykeligheten og dødeligheten i Wales, der han arbeidet. Som allmennlege var det lite han kunne gjøre med fattigdom og sosial nød, men han så betydningen av god kjennskap til samfunnsforholdene og effekten av målrettet innsats der behovet var størst. Helsetjenesten burde kompensere for ulikhetene i levekår, ikke forsterke dem.

Misforholdet mellom behovet for og tilgangen til helsetjenester er særlig stort i lav- og middelinntektsland, og et sterkt argument for universelle helsetjenester (Nambiar & Mander 2017). Men også i land med gode offentlige helse- og omsorgstjenester er det fortsatt vanskelig å ivareta de mest sårbare gruppene og å prioritere tjenestetilbudet til dem som trenger det mest (Marmot 2018).

Det diskuteres i hvilken grad og på hvilken måte sosial ulikhet i bruk av helsetjenester har betydning for ulikheten i helse og sykelighet. Utfordringen ved undersøkelser av dette er at dårlig helse i seg selv medfører et økt behov for helsetjenester, og at det er vanskelig å kontrollere for dette behovet. Dersom bruken av helsetjenester bare var bestemt av behovet, skulle man forvente at personer med lav utdanning bruker helsetjenester i større grad enn personer med høy utdanning fordi de har dårligere helse. Slik er det også, særlig for allmennlegetjenester og sykehusinnleggelser (Lunde et al. 2017).

Sammenhengen mellom inntektsnivå og bruk av helsetjenester er tydeligere enn sammenhengen mellom utdanningsnivå og helsetjenestebruk. Det er lavinntektsgrupper som hyppigst blir innlagt i sykehus. Med unntak for dem med aller lavest inntekt, er det ingen forskjell knyttet til inntekt for bruken av allmennlege eller fysioterapeut. Men det er en klar sammenheng med inntekt for bruk av legespesialister og tannlegetjenester (Lunde et al. 2017). Det bekreftes av Helseundersøkelsen i Nord-Trøndelag (HUNT), som viser at sosioøkonomiske forskjeller i bruk av allmennlege og sykehusinnleggelser er små, mens bruken av polikliniske spesialisthelsetjenester øker både med økende inntekt og med økende utdanning (Vikum et al. 2013). Dette samsvarer med internasjonal forskning (Hanratty et al. 2007).

En studie av kritisk syke og døende sykehuspasienter viser et høyere forbruk av spesialiserte helsetjenester hos pasienter med høyere utdanning enn hos pasienter med bare grunnskole, selv etter korrigering for medisinsk behov (Elstad 2018). Bruken av private sykehus, innenfor rammen av fritt sykehusvalg, øker også med økende inntekt og økende utdanning (Holom et al. 2018). Personer med høy utdanning

vaksinerer seg mot influensa i større grad enn personer med lav utdanning – også om det korrigeres for helsetilstand (Lunde et al. 2017).

Alt i alt ser det altså ut til å være systematiske forskjeller i bruken av helsetjenester knyttet til sosioøkonomiske forhold, uten at man med sikkerhet kan si hvilke konsekvenser disse forskjellene har for utfallsvariabler som helsetilstand og dødelighet.

Sosial ulikhet i helse er en helsepolitisk utfordring i mange land, og en av grunnene til at folkehelsemålet i Norge ikke bare er flest mulig gode leveår for alle, men også reduserte sosiale helseforskjeller i befolkningen. Så langt ser ikke Norge ut til å lykkes spesielt godt i dette arbeidet (Mackenbach et al. 2016).

«Det går nok over»

«I Midtbygda søker de lege for bagateller. I Skogsbygda må de nærmest være dødssyke før de kontakter helsetjenesten.»

Sosial ulikhet i helse skyldes ikke bare forskjeller i økonomi og utdanning, men også tradisjoner og kultur. Fastleger rundt om i landet observerer geografiske særpreg, som riktignok kan samvariere med sosioøkonomiske forhold, men som er viktige i seg selv. Mens helsebevisstheten enkelte steder kan være vel stor, er det i andre miljøer fortsatt skjebnebaserte holdninger til sykdom. Dette avspeiler seg ikke bare i levevaner, men også i bruk av helsetjenester. Leger med god lokalkunnskap vet at det kan skjule seg store forskjeller bak telefonsamtaler med samme ordlyd fra ulike deler og miljøer i samme kommune.

Introduction and key messages

"Introduction and key messages", i

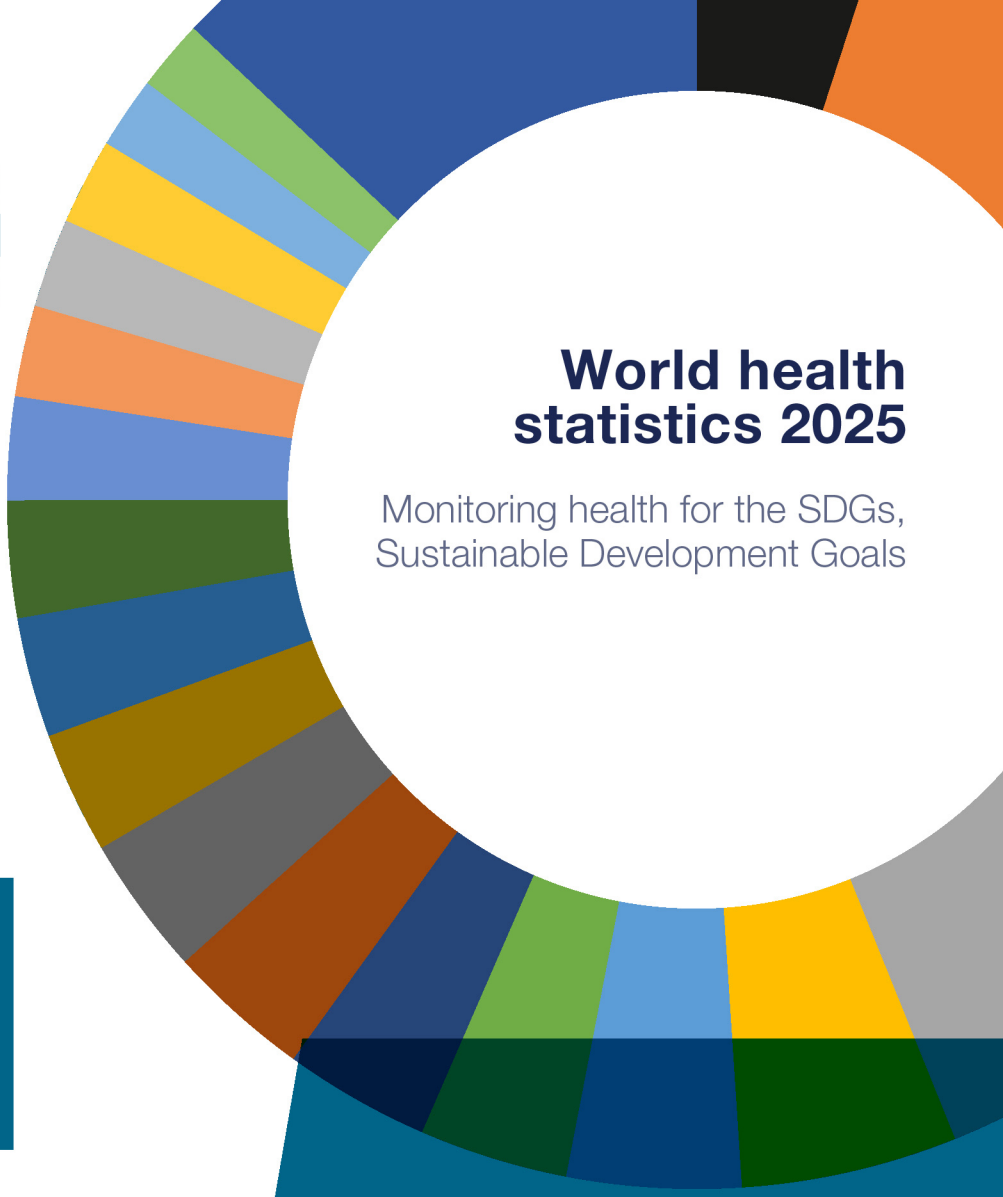
World health statistics 2025. Monitoring health for the SDGs, Sustainable Development Goals 2025 ss. 1-4

Copyright (C) 2025

Ikke rettighetsbelagt

World health statistics 2025

Monitoring health for the SDGs, Sustainable Development Goals



2025

World Health Organization

Introduction

The *World health statistics* report is the annual compilation of health and health-related indicators, which has been published by the World Health Organization (WHO) since 2005.

The 2025 edition consists of four chapters and the accompanying annexes. Chapter 1 presents an in-depth analysis of global and regional estimates of life expectancy, healthy life expectancy and progress in the reduction of premature mortality. Chapter 2 reviews the status of the health-related Sustainable Development Goal (SDG) indicators, covering age- and cause-specific mortality, infectious diseases, risk factors for health, and universal health coverage (UHC) and health systems. A summary of global and regional trends towards the achievement of selected health-related SDG indicators is further presented in Annex 1. Chapter 3 presents updated projections on WHO's Triple Billion targets, based on the most recent available data. It also describes WHO's transition from its thirteenth general programme of work (GPW13) to the fourteenth general programme of work (GPW14), aligning with emerging global health challenges and priorities. Chapter 4 looks at inequalities in immunization as an important global health issue.

The information presented in *World health statistics 2025* is based on data available from global monitoring as of April 2025. These data have been compiled primarily from publications and databases produced and maintained by WHO, other United Nations entities (such as UNICEF, UNAIDS, and the Population Division in the Department of Economic and Social Affairs of the United Nations), United Nations Inter-Agency bodies of which WHO is a member, and other international organizations. A summary of the methodology for each chapter is described in Annex 2. Tables of health statistics by country and area, WHO region and globally can be accessed via https://www.who.int/data/gho/publications/world-health-statistics.

Key messages

The major contributors to the 5.4-year pre-pandemic increase in global healthy life expectancy (HALE) at birth between 2000 (58.1 years) and 2019 (63.5 years) were mortality reduction from communicable and perinatal conditions among ages under 5 years, and from noncommunicable diseases (NCDs) among those 30 years and older. However, worsening morbidity due to diabetes among ages 30 years and older has led to a 0.14-year loss in HALE. Each WHO region exhibits distinctive patterns given their unique compositions of mortality and morbidity.

Between 2019 and 2021, the HALE loss can be almost entirely explained by mortality both directly and indirectly attributable to coronavirus disease (COVID-19) among those aged 30 years and older, both globally and in most WHO regions. The pandemic also caused increased morbidity from anxiety disorders and depressive disorders, each responsible for a 0.06-year (or 3-week) loss in global HALE in 2019–2021, and together effectively wiped out 80% of the positive contribution to HALE (0.15 years, or 8 weeks) by mortality decline from all NCDs combined in that two-year period.

The contributors to the HALE difference between males and females show a mixed pattern. While lower female mortality from injuries among 5–69-year-olds and from NCDs including ischaemic heart disease and stroke among females of 30 years and older gives a lead in HALE compared with males, mortality in maternal conditions and breast cancers, and higher female morbidity from conditions including back and neck pain, gynaecological diseases, migraine, depressive disorders and anxiety disorders have effectively eliminated a considerable share of the female HALE lead.

While lower mortality from communicable and perinatal causes explains a significant amount of the HALE lead that populations in high-income countries (HICs) had against those in low-income countries (LICs) and lower-middle-income countries (LMICs), higher mortality from tracheal, bronchus and lung cancers and drug use disorders, and higher morbidity from falls and back and neck pain has cost HICs some loss in HALE compared with other income groups.

Premature mortality was in decline globally and in all WHO regions from 2000 to the onset of the COVID-19 pandemic. Despite the overall progress prior to the pandemic, stagnation was seen after 2015. Globally, the pace of decline is projected to slow between now and 2050, compared with the rates observed in 2000–2019. Only about one third of countries are projected to have accelerated progress in 2019–2050 compared with 2000–2019, mostly located in the Region of the Americas and the Western Pacific Region.

As the world moves closer to 2030, overall progress is insufficient to meet the health-related SDG and other global targets. Declines were observed in mortality from causes addressed by the SDG indicators, including maternal mortality, child and neonatal mortality, premature NCD mortality, injury mortality, and mortality attributed to unsafe water, sanitation and hygiene (WASH) and air pollution. However, the progress for all these indicators is either insufficient or stalled, and is currently off-track for achieving their respective global targets.

There has been mixed progress in the global fight against infectious diseases, with declines in HIV and tuberculosis (TB) incidence rates, as well as in the number of people requiring interventions against neglected tropical diseases (NTDs), while malaria incidence rates have increased since 2015 and hepatitis B and antimicrobial resistance (AMR) continue to pose challenges.

Similarly, the world is off-track for meeting most of the global targets in reducing the prevalence or exposure to health risk factors such as malnutrition, tobacco use, unsafe WASH, air pollution and violence against women and girls, despite improvement in some of these areas.

However, the world is on-track to reach the 20% reduction in total per-capita alcohol consumption by 2030, although progress differs by region.

While gains in UHC service coverage index (SDG indicator 3.8.1) were observed globally over the past two decades, progress has slowed in recent years. In 2019, some 344 million people were pushed or further pushed into extreme poverty by out-of-pocket (OOP) health spending, while 13.5% of the global population spent more than 10% of their household budget on OOP payments for health.

The estimated global shortage of health workers of 15.4 million in 2020 decreased to 14.7 million in 2023. The projected 2030 shortage of 11.1 million shows slow progress in closing the gap, with the WHO African and Eastern Mediterranean regions projected to bear nearly 70% of the shortage.

Progress towards achieving the Triple Billion targets remains uneven. The expansion of UHC is insufficient, with only about 431 million additional people gaining access to essential health services without incurring

financial hardship by 2024 compared with the 2018 baseline set in the WHO Thirteenth General Programme of Work, and only a projected 500 million by 2025 – just half of the targeted one billion. Meanwhile, close to 637 million additional individuals are expected to be better protected from health emergencies by 2024. This is expected to increase to only 697 million in 2025 – a substantial gain that nonetheless remains a little over 30% short of the one-billion target. On the plus side, an estimated 1.35 billion more people will experience healthier lives by 2024, rising to 1.5 billion by 2025, thus exceeding the original target of one billion. However, this progress remains insufficient to put the world on track to achieve the health-related SDGs by 2030. Moreover, recent interruptions in international aid threaten to disrupt health services and systems, disproportionately impacting countries and communities with the greatest health-care needs. Safeguarding the gains made towards the Triple Billion targets remains of paramount importance for the global community in the years ahead.

Within-country inequalities in childhood immunization related to economic status and education level of the mother persist. Positively, inequalities in the last decade have diminished, compared with those in the previous decade, in LICs and LMICs. For instance, economic-related inequality in DTP3 coverage and zero-dose prevalence have reduced by more than half across LICs. Eliminating economic-related inequality across 88 LICs and middle-income countries is associated with improving the national average DTP3 coverage across these countries by 10 percentage points and halving zero-dose prevalence. Achieving equity in immunization coverage requires focused attention on urban poor areas, remote/rural areas, conflict areas, and gender-related inequities and barriers.