

人口学講義「死亡の分析」

中澤 港（群馬大学大学院医学系研究科公衆衛生学教室・准教授）

2010年1月13日

目次

1	はじめに	3
2	ヒトの死とは？	3
3	死亡データの成立	3
4	死亡のプロセス	6
4.1	寿命とは何か？	6
4.2	ヒトはなぜ長寿なのか？	7
4.3	おばあちゃんの孫育て仮説	7
4.4	繁殖成功を犠牲にした長寿仮説	8
4.5	「課題と挑戦」仮説	8
5	死亡指標	8
5.1	年間死亡数から推計できる指標	9
5.2	遡及聞き取りデータからの推計法とその指標	13
5.3	生命表解析	13
5.4	生存時間解析	15
5.5	コホート分析	16
6	死亡の数学モデル	16
7	死亡データ分析の例	18
8	文献	28

人口学「死亡の分析」配布資料：中澤 港 <nminato@med.gunma-u.ac.jp>

2009年12月21日 第1版
2009年12月23日 第1.1版（生命表解析の記述を増やし，分析例のRコードを追加）
2009年12月24日 第1.2版（社会疫学の記述，及びGrevilleの補整についての脚注と文献追加）
2009年12月24日 第1.21版（Grevilleの補整の脚注で，余計な記述を削除）
2009年12月25日 第1.3版（demogRとEpiに脚注としてwebサイトURLを付記）
2009年12月29日 第1.4版（typo訂正と分析例のRによるモデル当てはめ追加）
2010年1月13日 第1.5版（粗死亡率・年齢調整死亡率の分母の説明を修正）

中澤 港（著）

内容に誤りは無いように努めていますが，もしお気づきの点がありましたらメールでご指摘ください。

1 はじめに

死亡は、さまざまな死因の集積による多面的現象である。このことは出生よりも広く認識されているが、同時に、年齢別死亡に何らかの規則性が存在することも古くから指摘されてきた。形式人口学における死亡の分析は、集団レベルでのそれを明らかにしようとする試みである。なかでも、生命表による死亡の分析は形式人口学のバックボーンとも言われる。

一方、実体人口学、例えば公衆衛生学や疫学の分野では、死亡は健康水準の最終的な指標である。乳児死亡率は対象集団の一般的な衛生状態や医療水準を反映するし、死因別死亡率を求めればその集団の死因構造において何が大きなリスク因子となっているのかを評価することができ、公衆衛生政策の立案の根拠となる情報を提供することができる。

人口転換にともなう死亡の高齢期への集中という（生存曲線の矩形化とも呼ばれる）現象を個人差の減少とみなせば、個人レベルの死亡現象と集団レベルの死亡現象がはっきりとつながっていることがわかるが、発生までの期間が長いために、個人レベルの現象としての死亡（発生した個体が死に至るまでの生物学的なプロセス）は、なかなか捉えにくい。

2 ヒトの死とは？

個々のヒトの死についてわかっていることは、誰でもいつかは必ず死ぬことや、個人個人で時期や原因が異なること（先天異常などで生まれつき死のリスクが高い人、直前までピンピンしていながら、突然、偶発的に交通事故や戦争で死ぬ人、百歳以上まで生き延びた後に老衰で死ぬ人、など）、以上2点から明らかなように、遺伝と環境の両方の影響を受けること、くらいであろう。

集団として考えるとどうだろうか。ある集団に属している人の死に方を表現する手段はあるだろうか。例えば、一般に、途上国の人は全体として先進国の人より若くして死んでしまう人が多いとはいえる。このことを、途上国は先進国より「死亡水準」が高いというためには、どうやって（即ち、どんな指標を使って）示せばいいのだろうか？

本稿では死亡データを分析して得られる代表的な指標について解説し、実際の計算例を示す。

3 死亡データの成立

日本では人口動態統計がしっかりしているので、年間の死因別死亡数が簡単に集計できるが、その仕組みは決して単純ではない。すべての死亡が死因とともに漏れなく届け出られ、登録されなくては、これは実現できないことなのである。医師法により、日本では、すべての死亡は届け出られることになっている。医師が看取った場合や、継続診療中の疾患で亡くなった後で死体を検査した（最終診察後24時間以内の場合は再度死体を検査しなくてもいい）場合は、医師が死亡診断書を発行する（医師法20条）。それ以外の死はほぼすべて「異状死」として扱われ、医師が死体を検査した後で死体検案書を発行し、24時間以内に所轄警察署に届け出なければならない（医師法21条）。

なお、東京都など監察医制度があるところでは死体検案は監察医によって行われる。

実は死因は怪しいという話

日本では死後の解剖が義務付けられていないので、死後に解剖が実施される割合は2%に過ぎず、死因の判定は主に医師が臨床診断として解剖せずに行うことになるが、臨床診断は誤っている場合が12-30%に及ぶという指摘が、作家の海堂尊（本職は病理医）によってなされている（海堂尊『死因不明社会』講談社ブルーバックス）。海堂は、この事実を踏まえ、死後にCT等の画像診断をして真の死因を明らかにすること（画像診断だけでは確定できない場合には解剖を行う根拠ともなりうる）を提唱している。これはオートプシー・イメージング (Ai) と呼ばれ、全国各地で徐々に採用する施設が増えてきている。

通常、死亡診断書または死体検案書は、A3横の死亡届書の右半分に作成される^{*1}。死因については、直接死因から遡って原死因（WHOの定義では、「直接に死亡を引き起こした一連の事象の起因となった疾病もしくは損傷」または「致命傷を負わせた事故もしくは暴力の状況」）まで記載することになっている。例えば、10年前から糖尿病を患っていた人が1ヵ月前に肺炎に罹り、回復することなく5日前に膿胸となり、2時間30分前に敗血症性ショックを起こして死に至った場合、直接死因は敗血症性ショックだが、原死因は肺炎となるため、死亡の原因を記載するI欄には、(ア)敗血症性ショック、(イ)膿胸、(ウ)肺炎と記載する。糖尿病は「直接には死因に関係しないがI欄の傷病経過に影響を及ぼした傷病名等」を記載するII欄に記載する。

届出人（優先順位は同居の親族、同居していない親族、同居者、家主、地主、家屋管理人、土地管理人、公設所の長の順）は、死亡届書の左半分に記入押印し、市町村役場（優先順位は、死亡者の本籍地、死亡地、届出人の現住所地の順）の戸籍課に提出する（ただし、通常、死亡届書の記入と提出は葬儀社により代行されている）。戸籍課では死亡者本人の戸籍の抹消と火埋葬許可証の交付を行い、死亡届書を1ヶ月保管した後、地方法務局へ送付する。

一方、人口動態調査令（厚生労働省令の1つ）により、厚生労働大臣は、人口動態調査票の用紙を保健所長経由で各市町村長に交付することになっている。市町村長は、上記死亡届書が提出されたらすぐに、人口動態調査票を作成し、人口動態調査票市町村送付票を添えて管轄保健所長に提出する（人口動態調査令施行細則）。保健所長はこれに基づき死亡小票を作成し3年間保存するとともに、1ヶ月ごとに取りまとめ（前月15日から当月14日まで）、人口動態調査票保健所送付票を添えて、当月25日までに都道府県知事（ただし政令指定都市や中核市に設置された保健所では市長）に提出する。都道府県知事（あるいは市長）は、これを取りまとめ、人口動態調査票都道府県送付票を添えて翌月5日までに厚生労働大臣に提出する。人口動態調査票を厚生労働大臣が保管する期間は翌年1年間だが、電子化されたデータは永久に保管することになっている。

上記のプロセスを経て、日本人すべての死亡データは厚生労働大臣のもとに集まり、人口動態統計として集計され報告されることになる^{*2}。そのおかげで、年齢別死因別死亡数といった数字を得

^{*1} <http://www.mhlw.go.jp/toukei/manual/index.html> として厚生労働省のwebサイトで記入マニュアルが公開されている。

^{*2} 人口動態統計では、原死因を死因として集計する。原死因は上記I欄のもっとも下に書かれているものとなるのが原則だが、それが呼吸不全、心不全など死因としてふさわしくないとICD-10でWHOが定めたものになっている場

ることができる。しかし、疫学研究などで個人ベースの詳細な死亡データが欲しいときは、死亡小票の目的外使用を申請し、総務大臣の許可を得てから、保健所で閲覧する必要があり、大変面倒である。平成 19 年の統計法改正を受けて学術会議から提言（次の枠内に一部引用）が出たが、いまだに改善されていない模様である。

学術会議分科会提言

(5) 日本版 National Death Index の創設と運用

疫学研究においては、一定集団について長期間観察を行うことによってはじめて結果が得られることが多く、人の生死情報がしばしば必須であり、人口動態統計との連結なしには科学的な結論が見出せない。これまで、人口動態統計使用に当たっては、厚生労働省を経由し、総務省に目的外使用に関する申請を行い、認可を得て情報を入手することとなっていた。申請から認可までに要した期間は、平均 11.8 か月（最短で 3 か月、最長は 23 か月）であり、申請書等の修正回数は平均 10.2 回（最少で 0 回～最多は 60 回）であった（日本疫学会将来構想検討委員会「政府統計の利活用に関する調査報告書」平成 19 年 11 月 30 日、<http://www.soc.nii.ac.jp/jes/news/pdf/20071206seifu.pdf>）。これに比し、欧米諸国では健康情報登録（疾病登録、患者登録等）が整備され、死亡データとのリンケージも容易である。例えば、米国では、National Death Index システムを用いて、氏名、性、生年月日、住所、社会保障番号などを指標とする死亡ファイルとの記録照合による追跡調査が容易に可能である。このシステムを利用する場合、利用申請書のフォーマットや、それを作成するための詳細なマニュアル等が整備されており、申請処理期間は約 2 か月であり、多くの医学研究がこのシステムを用いて行われ、大きな成果をあげている（National Health Statistics. National Death Index. <http://www.cdc.gov/nchs/ndi.htm>）。これまで我が国の死亡統計ファイルには氏名が入力されていなかったため、このような死亡ファイルとの記録照合は不可能であった。しかし、2003 年 1 月より厚生労働省が進めている人口動態オンライン報告システムでは、死亡小票氏名が入力されるようになり、このシステムを充実させて全国的な運用を図れば、日本版 National Death Index の基盤が整うこととなる。統計情報の高度利用の一環として、死亡統計ファイルとの記録照合の技術的および制度的な検討を早急に開始し、日本版 National Death Index システムの構築とそれを活用するための手順の標準化・マニュアル化を進めるべきである。これらのシステムの円滑な運用によって、我が国の保健医療分野における疫学研究が飛躍的に進展し、国民の健康と安全のための有益な種々の施策提案に結びつくことが期待される。

出典：提言「保健医療分野における政府統計・行政資料データの利活用について—国民の健康と安全確保のための基盤整備として」, pp.10-11, 平成 20 年（2008 年）8 月 28 日, 日本学術会議基礎医学委員会・健康・生活科学委員会合同パブリックヘルス科学分科会

<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-20-t62-6.pdf>

欧米先進諸国では概ね日本よりも信頼性の高い死亡データが得られる。地域がん登録の登録率も高いし、電子化も進んでいる。死因確定のための剖検割合も日本より高い国が多い。

途上国では事情は一変する。パプアニューギニアやソロモン諸島では、そもそも死亡事象を行政が把握するまで年単位の遅れが存在する場合さえあるし、死因も必ずしも医師によって診断されない。個人ベースの死因を把握しようと思ったら、綿密な聞き取り（ただしくロスチェックは必須）によって再構成する方が、公式統計よりもむしろ信頼性が高い場合さえある。

合は II 欄の傷病を原死因とする。

4 死亡のプロセス

個体に現実に起こる死亡のプロセスは、死亡をもたらす要因を最小限にとどめたとしても生物学の必然として起こる場合（一般に言う、天寿を全うする場合）の他に、物理化学的要因や生物学的要因などさまざまな要因からなる病気によって死がもたらされる場合や、あるいは交通事故や天災によって突然訪れる死もある。同じ病気に罹っても保健医療水準が高い先進諸国と、高度医療がほとんど受けられない途上国とでは死に至る可能性は異なることが普通である。

社会的な要因は、治療だけではなく、物理化学的要因や生物学的要因との接触、高脂肪食を中心とする食生活の変化まで含めて、病気をもたらすリスクにも影響する。近年では、社会疫学という研究分野がクローズアップされ、隣人がどれくらい信頼できるか、社会組織にどれくらい参加しているか、といった、ソーシャル・キャピタルと呼ばれる要因が健康状態に及ぼす影響が注目されている（このことは、2009年12月22日の厚生政策セミナー「長寿革命」でも、複数の演者が指摘していたが、例えば『社会格差と健康』（川上ら編，2006）などが参考になる）。

多くの社会では相対的に若齢では事故死が、高齢では病死が多いが、日本では中年層の自殺が多いことが目立つ。自殺も、かつてのブラジルで砂金掘りのガリンペイロたちが精錬のための水銀蒸気曝露によって中枢神経系に異常をきたし、結果として高い自殺率を示したような物理化学的要因もあるかもしれないが、社会的な要因の影響が強い死因である。

4.1 寿命とは何か？

一般に寿命といえば、事故がなかった場合に、そのものが機能停止するまでの時間をさすように思う。何らかのメカニズムで老化が進行すると、やがて機能停止に至ると考えると、寿命を考えることは老化を考えることと表裏一体である。老化がなぜ起こるかということについては、体細胞廃棄説など諸説紛々であり、結論は出ていない（杉本，古市，1998）。アロメトリーの成果によれば（たとえば Charnov, 1992; 本川, 1992），生まれてから死ぬまでに心臓がドクンと鼓動する回数や、呼吸する回数といったものは、哺乳動物にほぼ共通していることが知られており、代謝速度を基準に考えれば、寿命はどの哺乳動物でも一定と考えられることが指摘されている。テロメアの長さという点から考えると、体細胞の分裂回数で寿命が決まってくるとも考えられる。

いっぽう、人口学で寿命といえば、普通は平均寿命をさす。平均寿命とは、ゼロ歳平均余命を意味するテクニカルタームである。「平均」といっても死亡年齢の算術平均をとるわけではなく、同時に生まれたゼロ歳児が10万人いたとして、その人たちがある時点の年齢別死亡率に従って死んでいった場合に、平均何年生存するかという値である。言い換えると、平均寿命は年齢別死亡率を静止人口モデルで計算した値であり、計算には生命表を作ることが普通である。死因が何であるかには関係なく年齢別死亡率だけで計算されるため、戦争があったり大災害があったりすれば、平均寿命は短くなる。その意味で平均寿命は、遺伝と環境の両方の影響を受ける指標である。

4.2 ヒトはなぜ長寿なのか？

生物の世界を見渡すと、ヒトは少数の子どもを産んで大事に育てる、いわゆる K 戦略者である。K 戦略を Charnov (1992) の生活史戦略の観点に立って見直すと、寿命が長くなるような戦略と考えることもできる。

先にもあげた体細胞廃棄説とは、何十億もの体細胞すべてのエラー修復を長い間続けることは非効率なため、生殖細胞だけのエラー修復をして、そこから再び個体発生をして次世代を構築することで全体としてのエラー蓄積を避けることができるわけであるが、その代償として廃棄すべき体細胞にはエラーがたまっていくのは避けがたく、それが老化なのだ、という説である。この説で考えると、K 戦略は、生殖細胞よりも体細胞にエネルギーを振り分ける戦略と説明できる。必然的に寿命は長くなるはずである。

しかし、次世代の生産を終えた閉経後の生存が次世代の生存に寄与しないなら、閉経後の生存は淘汰を受けないことになり、遺伝的な影響はないことになる。したがって、同じように 50 歳で閉経するならば、寿命が 60 歳でも 100 歳でも子ども数とは関係しないわけで、閉経ぎりぎりまで体細胞がもつように振り分けをコントロールするのが最も適応的と思われる。しかし、現在の先進諸国では、どういうわけか閉経後もかなりの期間、生存するのが普通であり、日本を始めとして平均寿命が延びつづけている国がいくつもある。なぜ包括適度に寄与しないはずの生存期間が延長しつつあるのかは大きな謎であり、いくつかの仮説が提唱されている。

4.3 おばあちゃんの孫育て仮説

ヒトの場合は子どもが再生産年齢に達するまでの養育コストが大きいので、孫や曾孫を育てるという形による閉経後の家系の生存への寄与は無視できない、ということに着目したユタ大学のホークスは、チンパンジーと再生産期間はそれほど変わらないのにヒトの再生産完了後の生存年数が長いのはおばあちゃんが忙しい娘の子育てを手伝うためである、という説を唱えている (Hawkes et al., 1998; Alvarez, 2000)。彼らの説を簡単に紹介すると、以下のようにまとめられる。

長い閉経後の生存期間は、他のどんな霊長類にすらない、ヒト独自の特徴である。このパタンは、母子間の食物の共有という、年をとった女性が娘の出生力を強め、それによって老化に対抗する淘汰圧を大きくする行為を進化させてきたかもしれない。哺乳類の生活史についての Charnov (1992) の無次元の集合法則と組み合わせれば、この仮説は我々の成熟が遅いこと、離乳時の身体が小さいこと、そして（霊長類としては例外的といえるほど）高い出生力を説明する（次の表を参照； αM がどの哺乳動物でも一定というのが Charnov の法則）。このことは、過去のヒトの生存環境選択と社会組織に対して、またヒトの進化における学習能力の拡大と父親による食料調達的重要性を示唆するものである。

種	成体期間 ⁽¹⁾	成熟年齢	離乳年齢	α ⁽²⁾	αM	体重比 ⁽³⁾	娘出産率 (b)	αb
オランウータン	17.9	14.3	6.0	8.3	0.46	0.28	0.063	0.52
ゴリラ	13.9	9.3	3.0	6.3	0.45	0.21	0.126	0.79
チンパンジー	17.9	13.0	4.8	8.2	0.46	0.27	0.087	0.70
ヒト	32.9	17.3	2.8	14.5	0.44	0.21	0.142	2.05

(1) $1/M = 0.4 * \omega - 0.1$ ただし、 M は平均成人死亡率、 ω は理論上の最高年齢 (Charnov, 1992)

(2) 離乳から成熟するまでの期間

(3) 離乳時の体重の成人の体重に対する比

4.4 繁殖成功を犠牲にした長寿仮説

老化の原因としての体細胞廃棄説によれば、長寿に寄与する体細胞のエラーを嚴重にすることは生殖細胞が利用可能なリソースを減らすので、代償として子ども数が減ることは想像に難くない。実際ショウジョウバエなどでは以前よりよく知られていた。

Westendorp ら (1998) が、17 世紀から 19 世紀のイギリス貴族の家系データにより、死亡年齢別の子ども数を集計したところ、60 歳までは長生きするほど子ども数が多かったが、70 歳、80 歳と長生きになるほど子ども数が減っていた。このことはヒトでも体細胞のエラーを減らす傾向が強いほど（つまり DNA 修復などの体細胞維持作用が強いほど）、妊孕力が低いことを示唆する。これは、ヒトでも体細胞廃棄説が成り立っていることを間接的に支持する証拠である。

長生きする方がより蓄財できるとすれば、それが子どもの生存に有利に働くことによって結果的に包括適応度を上げるならば、長寿に寄与する遺伝子が自然淘汰によって残っていく可能性はある。もしそうならば、少子化社会は長寿の所産として必然なのかもしれない。

4.5 「課題と挑戦」仮説

上記 2 つのような進化的な説明ではないが、Wilmoth J が提唱しているのが（注：2009 年 12 月 22 日、厚生政策セミナーでの基調講演による）、“Why mortality falls over time?” という問いに対して、あらゆる種類の Death に対しての human population の対応が、Recognition（認識）、Reaction（対策）を経て、Reduction（減少）へと至り、やがて別の原因による Death が主要死因になると、改めてこの“3R”プロセスが起これり、それが繰り返されることによって、死亡水準は永遠に低下し続けるという仮説である。Wilmoth 自身は“3R-theory of mortality decline”と呼び、堀内四郎は「課題と挑戦」仮説と訳している。

なるほどと思わされるが、この仮説の弱点は、地球上のリソースの有限性を考えていないところである。リソースが枯渇したらそれ以上の対策はとれないので、やはり寿命延長には限界があると考えるのが自然であろう。

5 死亡指標

いよいよ、さまざまな死亡データを分析して得られる、集団の死亡水準を示す指標値について説明する。どんな死亡指標を計算できるかは、研究方法に依存する。すべての死亡を記録しておき、

1年ごとにその数を層別に集計するのが基本であるが、歴史的コホートとして出生年ごとに集計する方法もあるし、前向きコホートとして集計することもある。

また、高齢者の方が若齢者より死にやすいのが普通なので、年齢構造の異なる集団間では、単純に死亡数だけで比較することには問題があり、年齢構造で補整^{*3}する方法がいくつか提案されている。

5.1 年間死亡数から推計できる指標

CDR Crude Death Rate の略。普通死亡率または粗死亡率という。率であるから、本来は延べ生存人年あたりの死亡数を求めるべきであり、それによって次元が(/年)となる。けれども、通常は単純化して1年間の人口変化が線型であると仮定し、分母は年央人口(7月1日現在の人口)で代替する。さらに、日本の人口動態統計では、

$$\frac{\text{年間総死亡数}}{\text{その年の10月1日人口}} \times 1000$$

と定義されている。10月1日人口を用いるのは、国勢調査人口が10月1日現在の人口だからである。一般に1年ごとに求めるが、小規模な集団の場合は、複数年のデータから平滑化して補整することもある。

ADR Age-specific Death Rate の略(Chamberlain, 2006)。Age-Specific Mortality Rate の略で ASMR と表記する場合もあるが同じものである。年齢別死亡率を指す。分母が年齢別の延べ生存人年(と考えると、実はいつの時点での年齢かという難しい問題があることがわかるが、ここではとりあえずそこまでは考えずにおく)であるが、その近似値として年齢別年央人口をとり、ADR の定義としては

$$\frac{\text{年齢別年間死亡数}}{\text{年齢別年央人口}}$$

となる。けれども、日本の人口動態統計では、粗死亡率と同じく、

$$\frac{\text{年齢別年間死亡数}}{10月1日の年齢別人口}$$

と定義している。通常は5歳階級だが、5歳未満は1歳階級にする場合が多い。年齢別死亡率は、後述する生命表解析にも、他の指標を計算する際にも基本となるので、きわめて重要である。

DSMR Directly Standardized Morality Rate の略。邦訳は直接法訂正死亡率、直接法年齢調整死亡率、直接標準化死亡率などいろいろあるが同じものを指す。日本の官庁統計で、たんに「年齢調整死亡率」と表記されている場合は、間接法ではなく、こちらを指す。

$$\frac{\sum \text{標準人口年齢別人口} \times \text{ADR}}{\text{標準人口}} \times 1000$$

^{*3} 補正と書くこともあるが、人口学でこの意味で使う場合は、補整と書くのが普通である。

と定義されている。標準人口は、比較したい集団全部に対して共通の年齢階級別人口を用いる。高齢者が多い集団は、ADR の高い年齢層が多いので、たとえ同じ年齢層で比較した場合の ADR が同じでも、若者が多い集団に比べて総死亡数が多くなるため、CDR は高くなってしまふ。人口構造におけるこうした年齢の偏りが CDR に及ぼす影響を調整するため、特定の人口構造を重みとした ADR の重み付き平均をとった値が、この DSMR である。対象とする集団の年齢別死亡率がわからないと計算できない。

SMR Standardized Mortality Ratio の略。日本語では標準化死亡比という。

$$\frac{\text{年間総死亡数}}{\sum \text{年齢別年央人口} \times \text{標準人口の年齢別死亡率}}$$

と定義される。対象集団の年齢別死亡数が未知でも計算できる。

ISMR Indirectly standardized mortality rate の略。間接法訂正死亡率，間接法年齢調整死亡率，間接標準化死亡率などと呼ばれる。

$$\frac{\text{CDR} \times \text{標準人口 CDR} \times \text{年央人口}}{\sum \text{年齢別年央人口} \times \text{標準人口 ADR}}$$

という式で表わされる（岡崎陽一「人口統計学」）。

しかし実は

$$\text{標準人口 CDR} \times \text{SMR}$$

と同値である（国際人口学会編，1994；Smith 1992）。

平均寿命 英語では Life expectancy という（Average life span という言い方もある）。より正確には、ゼロ歳平均余命と言った方がいい。前述の通り、死亡年齢の平均値ではなく、ある年の ADR に従ってゼロ歳人口 10 万人が死んだ場合に、平均して何年生きるかという値である。グラフィカルには、縦軸に生存率（ゼロ歳のときに 1.0）をとり、横軸に年齢をとって生存曲線を書いた場合の、曲線下面積として把握することができる。計算方法については、5.3「生命表解析」を参照されたい。

健康余命 英語では Health expectancy または Healthy life expectancy とよぶ。健康の定義も多様なため、これも多様な定義が可能だが、一般には、日常生活自立度が十分に高い状態で生存している人だけを生存と数えた場合の生命表解析により得られる。

HALE Health Adjusted Life Expectancy。QALY（Quality Adjusted Life Years）とも似た考え方で、健康状態を調整した余命の計算値。

DALE Disability Adjusted Life Expectancy。DALY（Disability Adjusted Life Years）と同じ考え方で、障害をもって生きる 1 年と障害なく生きる 1 年の価値を等しくないと考え、障害の程度によって調整した余命の計算値だが、一歩間違えば障害者差別につながりかねないので注意が必要である。

PMI Proportional Mortality Index あるいは Proportional Mortality Indicator の略。50 歳以上死亡数を総死亡数で除した値であり、国際比較によく用いられる。

死因別死亡率 特定死因による年間死亡数を年央人口で除して 100000 を掛けた値を指す。疾病による死亡の場合、死因が単独であることは少なく、複合死因である場合が多いことが問題である。通常、疾病の死因分類は、ICD (International Statistical Classification of Diseases and Related Health Problems; 国際疾病分類) によって行われる。現在のところ、第 10 版 (ICD-10) が最新である。死因分類が改訂されることは、死因別死亡率の経年変化を見る場合には注意しなければならない*⁴。日本では 1981 年以降一貫して悪性新生物死亡率がトップである。近年では悪性新生物死亡数は 34 万を超え、心疾患死亡数も 18 万を超えて、さらに増加を続けている。それに対して脳血管疾患死亡数は 12 万 6 千となり、わずかに減少している。

通常、疫学的転換の過程において、年齢調整した死因別死亡率の変化を追ってみると、まず感染症による死亡率が低下し、次いで心疾患による死亡率が低下し、最後にがんの死亡率の低下が始まる。がんによる死亡率が低下するのは、喫煙率の低下によって半分が説明できるという報告がある (Staetsky, 2009)。また、がんを胃がん、子宮頸がん、肝臓がんという感染性の病原体が関係するがんと、それ以外のがんに分けて、それぞれの年齢調整死亡率の年次推移をみるとパターンが異なる (Wilmoth ら、未発表、2009 年 12 月 22 日、厚生政策セミナー「長寿革命」での基調講演による)。

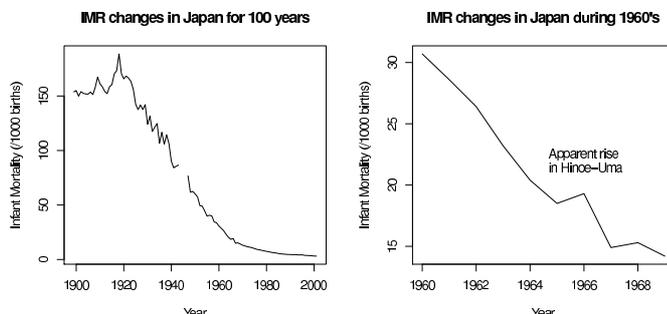
死因別死亡割合 特定死因による年間死亡数を年間総死亡数で除した値である。全死因のうち、どの死因による死亡が相対的に多いのかを示す。PMR (Proportional Mortality Ratio) ともいう。増減はその疾患の増減だけでなく、他の疾患の増減とも連動する。

YLL Years of Life Lost の略で、日本語では損失余命と呼ばれる。個々の死因について、生命表で、その死因による死亡がなかったら平均余命がどれだけ延びるかを計算して、その死因による損失余命と考えるのが、蒲生らの方法である (別に Graham らが開発した方法もある)。 <http://www.aist-riss.jp/software/riskcat/> から、その先のリスク計算も含めて計算できるソフト RiskCaT-LLE をダウンロードできる。

IMR Infant Mortality Rate の略で、乳児死亡率を意味する。出生 1000 当たりの、生後 1 年未満の死亡件数。分母に死産は含まないし人口でもないことに注意。たんに死亡水準の指標というよりも、とくに途上国で衛生状態や栄養状態を敏感に反映する指標なので、公衆衛生学的に重要である (生活文化水準を反映する指標ともいえる。日本はスウェーデンやスイスと並んで、世界で最も乳児死亡率が低い国の 1 つであり、日本での乳児の死因は、1979~1984 年のみ出産時外傷等が 1 位だが、その後は一貫して先天異常が 1 位である)。なお、分母と分子の期間がずれていることにも注意すべきである。つまり、実際の計算は、ある年の生後 1 年未満の死亡をその年の出生数で除して 1000 を掛けることによって行われるので、分子には前年に生まれた子供も含まれるし、分母には翌年亡くなる子供も含まれる。それが相殺されると見なして、ある年の指標値として使うのである。このことから考えると、例えば日

*⁴ 日本において、1995 年から心疾患死亡率が低下した最大の理由は、ICD-9 の死因簡単分類から ICD-10 の死因分類に変わったことによって死亡診断書の様式が変わり、「疾患の終末期の状態としての心不全、呼吸不全などは記載しない」とされたために、それまで心不全とされて心疾患死亡を押し上げていた死因が別の分類になったためである。

本の丙午のように、極端に出生数が少ない年があると、その年の乳児死亡率は過大評価される可能性がある*5。



新生児死亡率 生後4週未満の死亡を新生児死亡という。新生児死亡率は出生1000当たりの新生児死亡数である。

早期新生児死亡率 生後1週未満の死亡を早期新生児死亡という。早期新生児死亡率は出生1000当たりの早期新生児死亡数である。これは母体の健康状態の影響と、NICUの普及度など医療水準の影響が大きい。

周産期死亡率 周産期死亡については、1995年から新しい定義「妊娠満22週以後の死産と生後1週未満の早期新生児死亡を合わせた死亡」(旧定義は22週のところが28週)が採用されている。妊娠後期の死産と早期新生児死亡が母体の影響を受けやすいことと、途上国では早期新生児死亡が死産扱いされることが多いため、そこで誤分類があっても影響を受けない指標となる。現在の周産期死亡率は、ある年の出生数と妊娠22週以後の死産数の和を分母として、その年の早期新生児死亡数と妊娠22週以後の死産数の和を分子として、1000をかけた値である。最近の日本は旧基準では約3、新基準では約5である。

幼児死亡率 1～4歳の死亡を幼児死亡という(英語では Toddler Mortality)。2005年の日本の幼児死亡数を分子として、2005年日本の年央人口の1～4歳人口を分母として10万を掛けた値は25.4である(1999年には約33だった)。幼児死亡の死因は不慮の事故(欧米に比べて多い)や先天異常が多い。乳児死亡と幼児死亡をあわせて5歳未満死亡という概念も、一般的な衛生状態の指標としてよく使われる。

妊産婦死亡率 妊娠、分娩、産褥に直接関連する疾病や異常によって母性が死亡した場合を「妊産婦死亡」または「母性死亡」という。妊産婦死亡率は出産または出生10万当たりの妊産婦死亡数で表す。2006年日本は出生10万あたり4.9(1985年に15.8, 1999年には6.1)できわめて低い。主な死因は出血と妊娠高血圧症候群である(PIHと略記される。かつて妊娠中毒症と言われていたもの。2005年4月から変更。日本産婦人科学会の定義は「妊娠20週以降、分娩後12週まで高血圧が見られる場合、または高血圧に蛋白尿を伴う場合のいずれかで、かつこれらの症状が単なる妊娠の偶発合併症によるものではないものをいう」)。

*5 実際、その前後はずっと低下傾向にあるのに、1966年(丙午)の乳児死亡率はわずかに上昇している。これが見かけの上昇であることは、30年も前に指摘されている(近藤, 1980)。

5.2 遡及聞き取りデータからの推計法とその指標

計数データで推計できる指標は、遡及聞き取りデータからも推計できる場合が多い。集団サイズが小さい場合には、1年間では十分な代表性をもつ死亡データが得られない場合が多いので、遡及聞き取りによって何年分かをまとめて分析することも可能である。その他に、死亡年齢がわかる場合は、もっと精密な死力の分析が可能になる。

5.3 生命表解析

生命表解析の代表的なものは、ヒトの平均寿命を計算するときに行われている（官庁統計としても、まさしく生命表という形で発表されている^{*6}）。

平均寿命とは0歳平均余命のことだが、これは、ある時点での年齢別死亡率（後述する q_x ）に従って、ゼロ歳児10万人が死んでいったとすると、生まれてから平均してどれくらいの期間生存するのかという値である^{*7}。

一般に x 歳平均余命は、 x 歳以降の延べ生存期間の総和 T_x を x 歳時点の個体数 ℓ_x で割れば得られる。延べ生存期間の総和は、ちょうど x 歳に達した者が $x+1$ 歳に達しないで死亡する確率、すなわち年齢階級 $[x, x+1)$ における死亡率 q_x が変化しないとして、 $\ell_x(1 - q_x/2)$ によって x 歳から $x+1$ 歳まで生きた人口 L_x （開始時点の人口が決まっていれば死亡率も変化しないので x 歳の静止人口と呼ばれる）を求め、それを x 歳以降の全年齢について計算して和をとることで得られる。

ヒトの人口学では、通常、年齢別死亡率 m_x （ある年に x 歳で死亡した人数 d_x をその年の x 歳年央人口で割った値）から q_x を $q_x = m_x / (1 + m_x/2)$ として求めて生命表を計算するが^{*8}、ゼロ歳のところは短期間での死亡率の変化が大きいため、最初の1ヶ月は1週間ずつ、次いで2ヶ月、3ヶ月、6ヶ月と刻んで、1歳未満を7つの階級に分けて計算するのが普通である。また、年齢5歳階級で計算する生命表 (abridged life table) を作ることもあるが、その場合は死亡の線型性を仮定するのは無理なので別の補正を使う。よく用いられる Greville の方法は、

$${}_5q_x = \frac{{}_5m_x}{1/5 + {}_5m_x \left[1/2 + 5/12 \left[{}_5m_x - \frac{\ln({}_5m_{x+5}) - \ln({}_5m_x)}{5} \right] \right]}$$

である (Ng and Gentleman, 1995)。この式は、5歳階級での通常、年齢別死亡率の自然対数を

^{*6} <http://www.mhlw.go.jp/toukei/saikin/hw/life/life08/index.html> には、2009年7月16日に発表された、日本の平成20年簡易生命表が掲載されている。

^{*7} 誤解されることが多いが、死亡年齢の平均ではないので注意されたい。また、死亡年齢の中央値である半数死亡年齢は、ゼロ歳折半余命（または静止人口中位数年齢）という名前がある。ともに平均寿命とは別の指標値である。

^{*8} この式の意味は次の通りである。年央人口で x 歳だった集団 N_x の実際の年齢は、 x 歳以上 $x+1$ 歳未満である。もし死亡が一定速度で起こるなら、期首人口で x 歳だったのは $N_x + d_x/2$ 人となるはずである。この人たちが $x+1$ 歳になるまでに d_x 人死亡することになるので、死亡率 q_x は、

$$q_x = \frac{d_x}{N_x + d_x/2} = \frac{d_x/N_x}{N_x/N_x + d_x/2/N_x} = m_x / (1 + m_x/2)$$

となる。

とった値の傾き ($\frac{\ln({}_5m_{x+5}) - \ln({}_5m_x)}{5}$) が ${}_5m_x$ と等しければ、上の式と一致する*9。

生物一般について考えるときは、同時に生まれた複数個体（コホート）を追跡して年齢別生存数として l_x を直接求めてしまう方法（コホート生命表）とか、たんに年齢別個体数を l_x と見なししてしまう方法（静態生命表、偶然変動で高齢の個体数の方が多い場合があるので平滑化するのが普通）がよく行われる。

他には、Gompertz 曲線を使って補正したり*10、Siler モデルを使って補正する方法もある（非線形最小二乗法でパラメータ推定すればよい。R での非線形最小二乗法は、`nls()` 関数や `optim()` 関数で実行できる）。なお、Siler モデルの式は、

$$h(t) = a_1 \exp(-b_1 t) + a_2 + a_3 \exp(b_3 t)$$

である。

また一方、 l_x から

$$\mu_x = -\frac{1}{l_x} \frac{dl_x}{dx}$$

として求められる、ちょうど x 歳における瞬間の死亡率 μ_x は死力と呼ばれる。これは、次に述べる、個人ごとの生存年数をコックス回帰する場合のハザード関数に他ならない。

*9 Greville の式については、文献によって表記方法が若干異なるが、この、Ng and Gentleman (1995) による説明が一番わかりやすかった。オリジナルの Greville TNE (1943) Short methods of constructing abridged life tables. *Rec. Am. Inst. Actuar.*, 32: 29-43. が入手できないので確認できないのだが、館 (1963) によれば、

$${}_nq(x) = \frac{{}_nm(x)}{\frac{1}{n} + {}_n m(x) \left[\frac{1}{2} + \frac{n}{12} \{ {}_n m(x) - \ln c \} \right]}$$

で、 c は ${}_nm(x)$ が Gompertz 法則に従うものとして ${}_nm(x) = Bc^x$ としたことにより、US の経験から $\ln c$ の値は 0.080 ~ 0.104 で、通常 0.09 が用いられるとされている。一方、和田 (2006) では、

$${}_nq_x = \frac{{}_nm_x}{\frac{1}{n} + {}_n m_x \left[\frac{1}{2} + \frac{n}{12} \{ {}_n m_x - \log_e \left(\frac{{}_n m_x + n}{n} \right) \frac{1}{n} \} \right]}$$

と書かれている。これは Ng and Gentleman とまったく同値である。生命表解析の専門書である、Namboodiri and Suchindran (1987) によれば、

$${}_nq_x = \frac{{}_nM_x}{(1/n) + {}_n M_x \left[(1/2) + (n/12)({}_n M_x - k) \right]}$$

かつ k は生命表によってわずかに異なるかもしれない定数で、0.09 に等しいとしても大きな誤差はないだろうとされている。数理人口学者 Keyfitz による “Applied Mathematical Demography” では、

$$\int_0^n \mu(x+t) dt = {}_n m_x + \frac{n^3}{12} {}_n m_x^2 (\log_n m_x)'$$

が Greville の良く知られた結果であると書かれているが、 n^3 の部分が他の式と違っていて、何を意味するのかよくわからないが、それを除けば、どの本も違いがあるのは 1 カ所、そこをどの程度近似的な表現にしているかの違いなのではないかと思う。

*10 <http://www.toukei.metro.tokyo.jp/seimei/2005/sm-gaiyou.htm> に示されている東京都の生命表は、高齢者の死亡を Gompertz 曲線で補正している。ここに書かれている解説は丁寧で参考になる。

5.4 生存時間解析

個人ごとの死亡までの間隔データ，つまり死亡年齢を用いれば生存時間解析を行うことができる。集団の死亡パターンは個人個人の frailty（生まれつきの脆弱さ），あるいは病気や死亡の特定原因への感受性の分布を反映して変わってくるので，これらの変数と死亡年齢の関連を分析することによって，死亡水準を，集団レベルの分析だけの場合よりも詳細に把握することができる。

古典的な統計手法としては，ロジスティック回帰分析やダミー変数を使った回帰分析があるが，死亡の場合は事象発生までの期間が長いために，これらの手法では観察期間中にエンドポイントが得られないとか，観察開始時が特定できないといった問題があった。これら打ち切りデータを分析から除外してしまうと，相対的に事象発生までの期間が長めのデータで除外が生じやすくなるため，推定される半数死亡年齢を過小評価してしまう危険がある。

この問題を解決したのがカプラン＝マイヤ法による生存曲線の推定やハザード解析である。カプラン＝マイヤ法は右側打ち切り（観察期間中に死亡が起こらなかつたり，観察途中で研究参加をやめてしまつたり，転居してしまつたりして，エンドポイントが不明なもの）をうまく考慮することができる，ノンパラメトリックな生存曲線の推定法であり，半数死亡年齢（死亡年齢の中央値）を推定することができる。具体的には，イベントが起こる可能性のある人口を分母，まだイベントが起こっていない人口を分子として，時間の長さを横軸にして，イベントが起こった場合は分子から 1 を引き，打ち切りが起こった場合は分母からも分子からも 1 を引くという形で，時間と人数の積和（Kaplan-Meier の積・極限推定量）を求め，まだイベントが起こっていない人口がちょうど全体の半分になるまでの時間をイベントが起こるまでの時間の中央値とする。

また，比例ハザード性を仮定したコックス回帰はきわめてロバストな（比例ハザード性だけを仮定し，死亡ハザードについて特定のパラメトリックな分布は仮定しないという意味で，いわばセミパラメトリックな）解析法である。若齢での危険因子と高齢での危険因子は異なっているため，経時的に変化する説明変数を扱える必要があることと，死亡の生物学的メカニズムを考えるべきであることから，加速モデルなど，病因論的なハザードモデルも開発されてきている。これらを総称して，生存時間解析と呼ぶ。R では survival ライブラリを使うと，これら各種の生存時間解析を実行できる^{*11}。

生存時間解析は，通常，フォローアップ研究または歴史人口研究（数百年前でも宗門人別帳やキリスト教の教区簿冊 [parish record] には信者個人の名前と結婚，出生，死亡，疾病罹患，家族構成など，ほとんどすべての人口動態イベントが記録されているので，そのデータを再構成することによって，個人ごとの人口動態イベントが起こるまでの時間は把握できる）から得られるデータに対して用いられる。

^{*11} <http://phi.med.gunma-u.ac.jp/swtips/survival.html> の解説を参照されたい。

5.5 コホート分析

生存時間解析同様，フォローアップ研究または歴史人口研究のデータによく用いられる解析法として，コホート単位で変化を追うことで，世代間差を評価する「コホート分析」がある。例えば，コホート生命表（世代生命表）というものも作られているし，疾病罹患や死亡といった人口動態イベントの発生には，年齢の効果，時期の効果，コホートの効果が別々に（あるいは相互に関連しながら）作用すると考えられることから，これら3要因の効果を分解して考える，Age-Period-Cohortモデル（APCモデルと略記される）やレキシスダイヤグラムもしばしば用いられる。Rでは，Epiライブラリに入っている `apc.fit()` や `apc.plot()` によって前者が，`plot.Lexis()` によって後者が実行できる（Carstensen, 2007）。

6 死亡の数学モデル

Graunt (1662) がロンドンの人口構造を推測するのに年齢依存の死亡スケジュールを仮定したのが，最初の数理的アプローチであった。次いで，DeMoivre (1725) が生存関数 $\ell(x)$ を年齢 x の一次関数として，

$$\ell(x) = \ell(0) \times (1 - x/a)$$

として定式化した。 $\ell(x)$ は，その集団のうち正確な年齢 x 歳まで生き延びる割合を意味し， a は生存する可能性がある最高年齢を意味する。この定式化は明らかに誤りだが，昔は（例えば明治生まれの日本人男性のコホート生命表に対しては）そんなに悪い近似でもなかった。

年齢別死亡秩序については，Graunt 以来，さまざまなモデルが立てられてきた。代表的なものとして，老化に伴う死亡増加と年齢とは無関係な偶発的死亡を定式化した Gompertz-Makeham モデル（既にも書いたが，生命表の高齢部分の補正にも用いられる），年齢を3区分してそれぞれの死因の違いを考慮した Thiele モデル，それを改良した5パラメータの Siler モデルや8パラメータの Mode-Busby モデル，さらに乳児期の死亡の説明力を高めた Helligman and Pollard モデル，ワイブル分布によって小児期の死亡逡減を表した Mode-Jacobson モデル（Gage and Mode, 1993），探索的に発見した3つのパラメータからなる簡単な関数でどんな死亡パターンも表現できることを示した Denny のモデル（Denny, 1997）^{*12}などがある。これらは，精緻化に伴って適合精度は向上したが，すべてアприオリに集団レベルでの年齢別死亡パターンを与えるものであった。もっとも，Siler モデル

$$h(t) = a_1 \exp(-b_1 t) + a_2 + a_3 \exp(b_3 t)$$

に関しては，Gage (1991) が解釈を与えている。Gage は，第1項は若年での感染症による死亡を

^{*12} <http://phi.med.gunma-u.ac.jp/demography/denny.html> を参照されたい。3つのパラメータを a, b, c とし

$$\ell(x) = \frac{1}{(1 + a(\frac{x}{105-x}))^3} + b\sqrt{e^{\frac{x}{105-x}} - 1} + c(1 - e^{-2x})$$

意味し、第2項は事故死と妊娠出産に関連した死亡を意味し、第3項は中高年での癌や心血管疾患での死亡を意味するとしている。また、衛生状態の改善と a_1 の低下が対応するなど、ある程度、パラメータが現実的意味をもつことも利点である。

これに代わるアプローチとしては、実際の生命表に基づいた Coale and Demeny の生命表^{*13}、基準となるモデル生命表から2つのパラメータでどんな生命表も表わせるとした Brass (1968) のロジットモデル^{*14}などが知られている。これらの死亡モデルは、パラメータを増やすほど実際の年齢別死亡曲線へのあてはまりは良くなるので、老年期における死亡率のばらつきの補正や短期間の死亡データの補正には使えるが、個人の死亡のメカニズムを表現したものではないため、(1) 年齢別死亡率の偶然によるばらつきを評価することが難しい、(2) 社会変化にともなう疾病構造の変化など、死亡構造へのダイナミックに変動する影響を取り込むことが難しい、(3) 死にやすさの個人差を無視している、という欠点があった。

そこで、最近注目されているのが、雪崩モデルである。これは、個人の死亡過程を定式化した死亡モデルである (Gavrilov and Gavrilova, 1991)。中身を簡単に説明する。

- (1) $s_0, s_1, s_2, \dots, s_n, \dots$ が $0, 1, 2, \dots, n, \dots$ 個の故障をもつ個人の人数とし、故障数と独立に速度 λ_0 で増加し、故障とは無関係に速度 μ_0 で死亡が起こって総人数から除かれる、
- (2) 故障数に応じて故障の増加速度が λ 増し、死亡率が μ 増す、と仮定する。これは、染色体損傷による死亡モデル (Le Bras, 1976) にバックグラウンド死力 μ_0 を加味した形である。
- (3) この仮定の下で各状態の変化速度についての微分方程式を解いて合計すると生存関数が得られ、そこから死力を導出すると、

$$\mu(x) = \mu_0 + \mu\lambda_0(1 - \exp(-(\lambda + \mu)x))/(\mu + \lambda \exp(-(\lambda + \mu)x))$$

となり、 $\lambda \ll \mu$ ならば年齢 x が小さいところでは年齢によらない死亡の成分と、加齢に伴う死亡増加成分の和という Gompertz-Makeham 型で近似できる。

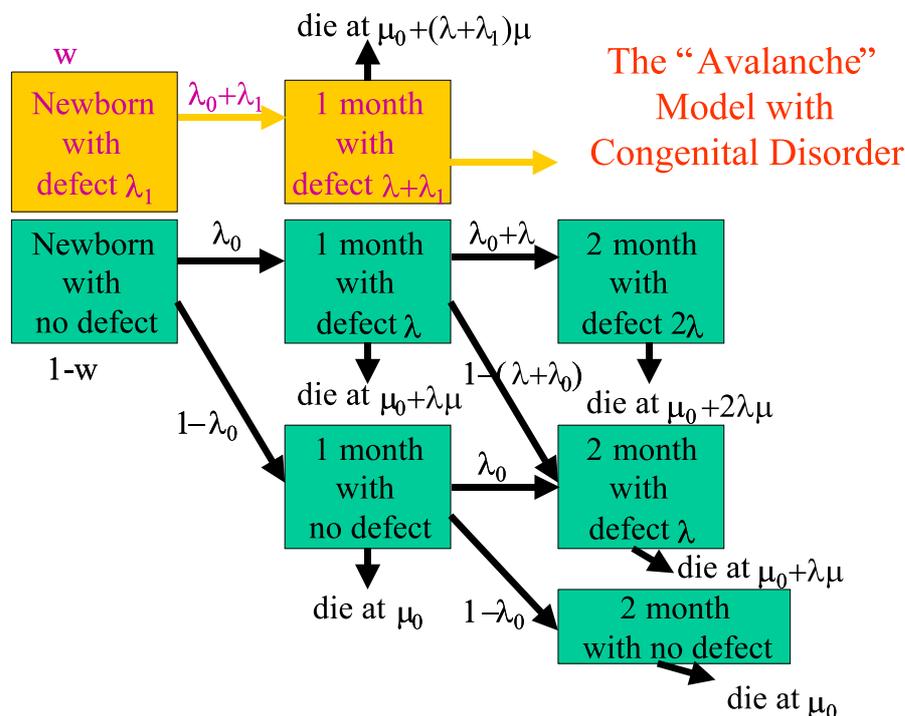
しかし、この形では乳幼児期の高死亡率とその逡減を表せないという大きな欠点があり、実際の死亡秩序への適合は高くなかった。雪崩モデルの改善には、出生時から故障を抱えている(つまり、先天異常をもつ)個人が集団内に存在するという frailty の不均質性を導入するとか(下図参照)、故障蓄積が死亡へと突き進む閾値を設定する、あるいは故障蓄積以外にも、事故や感染など突発的なリスクによって死亡してしまう可能性を考慮するといった方法がある (Mori and Nakazawa, 2003)。

^{*13} 1966年発表, 1983年改訂。4つの異なるパタン [乳児, 子供, 成人, 高齢者の死亡割合パタンが異なる, 東・西・南・北] と25の異なる死亡水準をもつモデル生命表を立てた。しかし、このアプローチの致命的な弱点は、柔軟性がないことであった。あくまで、現実のデータに「近い」モデルを選ぶしかなかった。

^{*14} ターゲットとなる集団の生命表の l_x を $l_0(x)$ 、基準となる生命表の l_x を $l_s(x)$ と書き、2つのパラメータを a, b とすれば、常に

$$\frac{1}{2} \ln\left(\frac{1 - l_0(x)}{l_0(x)}\right) = a + \frac{b}{2} \ln\left(\frac{1 - l_s(x)}{l_s(x)}\right)$$

と表わせる。



平成 11 年日本人男性の生存曲線については，オリジナルの雪崩モデルではどうしてもフィットしないのだが，先天異常割合を 0.4% 入れるだけで，劇的にフィットが改善する。しかし 1891 年生まれ日本人男性コホートのデータについては，先天異常を 30% 入れても初期の激しい生存確率低下にフィットしなかったため，感染症の影響を組み込むことによって，漸くフィットさせることができた。しかし，それでも，中高年での生存曲線が直線的に低下するところにはフィットしなかったため，故障蓄積に対して線形に死亡が上昇するのではなく，S 字状の反応カーブを示すようにしたら，ほぼ完璧にモデルがデータにフィットした (Mori and Nakazawa, 2003)。この解析結果により，例えば 1920 年から 1995 年までの日本人男性の生存曲線にフィットさせたモデルの λ_0 の変化が，脳血管疾患による死亡が，戦後から高度経済成長期の塩分の高い食事，過重労働，過密な居住環境など（それらが λ_0 ）によって上昇し，その後，それらの改善によって低下したことと呼応していることが示された。

7 死亡データ分析の例

日本のデータを使って実際に計算してみる。Excel でもいいのだが，せっかくなので R を使ってみよう。Excel ではセルに式をコピーすることによって計算を行うが，R ではベクトル演算ができるので，式の書き方がより簡潔明瞭になる。

なお，R で人口学的な分析を行うためのライブラリとしては demogR^{*15}や Epi^{*16}（の一部）が

*15 <http://www.jstatsoft.org/v22/i10/paper> に詳細な説明がある。

*16 <http://staff.pubhealth.ku.dk/~bxc/Epi/> から関連文書へのリンクを辿れる。

あり, demogR では生命表分析をする関数や Coale and Demeny のモデル生命表, サンプルデータなどが提供され, Epi では既に触れたように Age-Period-Cohort 分析を行う関数やレキシスタイヤグラムを描く関数などが提供されている。

扱う死亡データは日本の人口動態統計からとる。

```
mortality.R(1)

# mortalityJ.R
# rev. 1.0 19 December 2009 (C) Minato Nakazawa <minato-nakazawa@umin.net>
# sample data definition and standardization
# References:
# http://www.stat.go.jp/data/nenkan/02.htm
# 平成 20 年日本統計年鑑 第 2 章 人口・世帯
# Mortality data of Japanese in Japan.
# 2006 年「国民衛生の動向」厚生統計協会 (S60modelpopJ = 昭和 60 年モデル人口)
#
S60modelpopJ <- c(8180,8338,8497,8655,8814,8972,9130,9289,9400,8651,7616,
  6581,5546,4511,3476,2441,1406,784)*1000
AC <- c(paste("[" ,0:16*5,"-",0:16*5+4,"]",sep=""),"[85-]")
# same as follows:
# AC <- c("[0-4]", "[5-9]", "[10-14]", "[15-19]", "[20-24]",
# "[25-29]", "[30-34]", "[35-39]", "[40-44]", "[45-49]",
# "[50-54]", "[55-59]", "[60-64]", "[65-69]", "[70-74]",
# "[75-79]", "[80-84]", "[85-]")
names(S60modelpopJ) <- AC

S60M <- c(6042, 1155, 1011, 3179, 3397, 3167, 4237, 7110, 10234, 15063,
  24347, 30747, 30884, 38240, 55100, 65593, 59125, 48786)
names(S60M) <- AC
H02M <- c(4532, 844, 760, 3204, 3466, 2916, 3264, 5449, 9769, 14218,
  20161, 32925, 42742, 42664, 51737, 69320, 67916, 67451)
names(H02M) <- AC
H07M <- c(3929, 752, 716, 2413, 3640, 3203, 3297, 4413, 8236, 15616,
  21905, 30491, 47188, 59828, 60927, 68504, 77924, 87750)
names(H07M) <- AC
H12M <- c(2933, 438, 493, 1721, 2875, 3271, 3749, 4621, 6840, 13141,
  24103, 31848, 42214, 60962, 76413, 73947, 73533, 102177)
names(H12M) <- AC
H17M <- c(2291, 409, 361, 1220, 2303, 2887, 3915, 4915, 6806, 10577,
  19546, 34233, 43403, 55261, 80198, 99338, 89502, 127261)
names(H17M) <- AC
```

mortality.R(2)

```
S60F <- c(4792, 636, 638, 1033, 1272, 1558, 2496, 4017, 5650, 7644,
11504, 14828, 19961, 26490, 40891, 55657, 64448, 80930)
names(S60F) <- AC
H02F <- c(3451, 533, 482, 1149, 1329, 1361, 1774, 3102, 5542, 7510,
10097, 14616, 19986, 27267, 38076, 58203, 71633, 110407)
names(H02F) <- AC
H07F <- c(3111, 483, 468, 949, 1447, 1393, 1832, 2426, 4578, 8520,
11041, 14241, 21122, 29261, 41516, 56924, 79939, 141519)
names(H07F) <- AC
H12F <- c(2336, 300, 251, 676, 1160, 1546, 1847, 2425, 3639, 6595,
11740, 14144, 18466, 28096, 40115, 57053, 73527, 171735)
names(H12F) <- AC
H17F <- c(1811, 246, 229, 582, 1067, 1283, 2037, 2554, 3432, 5177,
9418, 15346, 18855, 25568, 40627, 60024, 84683, 225778)
names(H17F) <- AC
S60P <-c(7459, 8532, 10042, 8980, 8201, 7823, 9054, 10738, 9135, 8237,
7933, 7000, 5406, 4193, 3563, 2493, 1433, 785)*1000
names(S60P) <- AC
H02P <- c(6493, 7467, 8527, 10007, 8800, 8071, 7788, 9004, 10658, 9018,
8088, 7725, 6745, 5104, 3818, 3018, 1833, 1122)*1000
names(H02P) <- AC
H07P <- c(5995, 6541, 7478, 8558, 9895, 8788, 8126, 7822, 9006, 10618,
8922, 7953, 7475, 6396, 4695, 3289, 2301, 1580)*1000
names(H07P) <- AC
H12P <- c(5904, 6022, 6547, 7488, 8421, 9790, 8777, 8115, 7800, 8916,
10442, 8734, 7736, 7106, 5901, 4151, 2615, 2233)*1000
names(H12P) <- AC
H17P <- c(5578, 5928, 6015, 6568, 7351, 8280, 9755, 8736, 8081, 7726,
8796, 10255, 8545, 7433, 6637, 5263, 3412, 2927)*1000
names(H17P) <- AC
```

ここで、男女を合計した死亡数を求めるには、ただベクトルの足し算をすればいい。年齢別死亡率 ASMR（上の説明では ADR としてある）を求めるには、年齢別の死亡数を年齢別の年央人口で割ればいい。粗死亡率は、死亡数合計を人口合計で割れば求められる。以上の R コードは次の枠内の通り。

mortality.R(3)

```
S60T <- S60M+S60F
HO2T <- HO2M+HO2F
HO7T <- HO7M+HO7F
H12T <- H12M+H12F
H17T <- H17M+H17F

S60ASMR <- S60T/S60P; S60CDR <- sum(S60T)/sum(S60P)
HO2ASMR <- HO2T/HO2P; HO2CDR <- sum(HO2T)/sum(HO2P)
HO7ASMR <- HO7T/HO7P; HO7CDR <- sum(HO7T)/sum(HO7P)
H12ASMR <- H12T/H12P; H12CDR <- sum(H12T)/sum(H12P)
H17ASMR <- H17T/H17P; H17CDR <- sum(H17T)/sum(H17P)

CDRs <- c(S60CDR, HO2CDR, HO7CDR, H12CDR, H17CDR)
```

直接法年齢調整死亡率は、次の枠内のように関数定義でき、年齢別死亡率を与えると計算できる。

mortality.R(4)

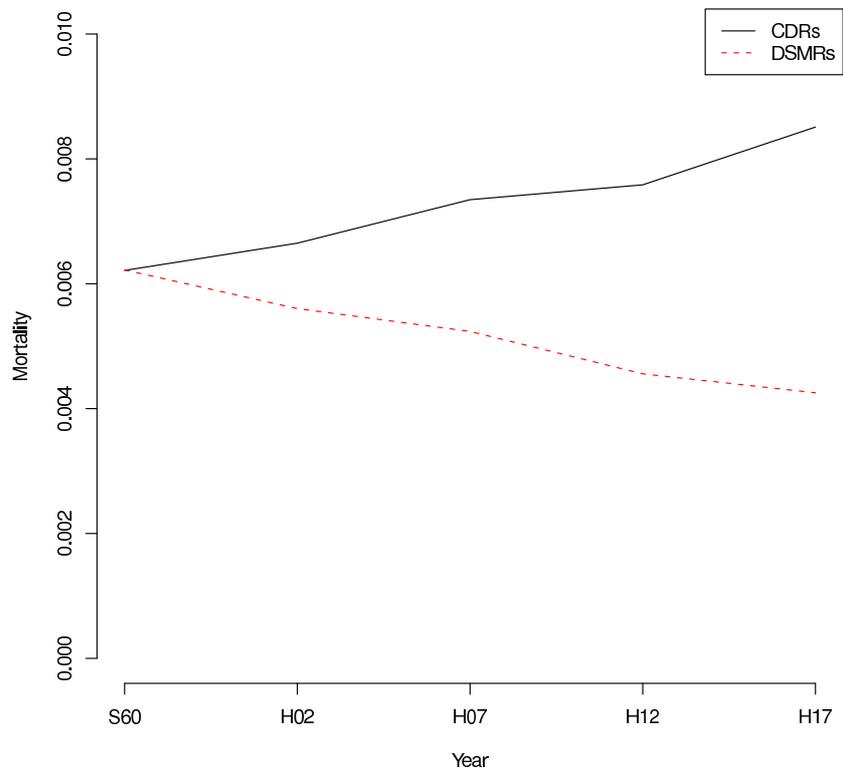
```
DSMR <- function(ASMR) {
  if (length(ASMR)!=18) { print("age class is inadequate."); NA } else {
    sum(ASMR*S60modelpopJ)/sum(S60modelpopJ)
  }
}

DSMRs <- c(DSMR(S60ASMR), DSMR(HO2ASMR), DSMR(HO7ASMR), DSMR(H12ASMR), DSMR(H17ASMR))
```

以上求めた粗死亡率と直接法年齢調整死亡率の年次変化をグラフにして pdf ファイルに出力するには、次のコードでできる。

mortality.R(5)

```
pdf("mortalityJ.pdf",width=8,height=8)
plot(1:5,CDRs,type="l",col="black",xlab="Year",axes=F,ylab="Mortality",
     ylim=c(0,0.01))
axis(1,1:5,c("S60","HO2","HO7","H12","H17"))
axis(2,seq(0,0.01,by=0.002))
lines(DSMRs,col="red",lty=2)
legend("topright",col=c("black","red"),lty=c(1,2),legend=c("CDRs","DSMRs"))
dev.off()
```



次に生命表を作ってみる。まず、通常の補間 (mode==1) と Greville の方法 (mode==2) で、生命表を計算する関数定義を示す。

```
lifetable.R(1)
# lifetable.R
# rev. 1.0 23 December 2009 (C) Minato Nakazawa <minato-nakazawa@umin.net>
# function definition to make a life table.
# included data is

lifetable <- function(mx,class=5,mode=1) {
  nc <- length(mx)
  if (length(mx)!=nc) exit
  qx <- numeric(nc)
  if (mode==1) { qx <- mx/(1+mx/2) } else {
    for (i in 1:(nc-1)) {
      qx[i] <- mx[i] /
        (1/class+mx[i]*(1/2+class/12*(mx[i]-(log(mx[i+1]))-log(mx[i]))/class))) / class
    }
    qx[nc] <- mx[nc] / (1/class+mx[nc]*(1/2+class/12*mx[nc])) / class
  }
  dx <- numeric(nc)
  lx <- numeric(nc)
  Lx <- numeric(nc)
  lx[1] <- 100000
  for (i in 1:(nc-1)) {
    dx[i] <- lx[i]*qx[i]*class
    lx[i+1] <- lx[i]-dx[i]
    Lx[i] <- (lx[i]+lx[i+1])/2*class
  }
  Tx <- cumsum(Lx[nc:1])[nc:1]
  ex <- Tx/lx
  data.frame(mx,qx,lx,Lx,Tx,ex)
}
```

lifetable.R(2)

```
qxJmH20 <- c(0.00266, 0.00038, 0.00027, 0.00019, 0.00014, 0.00012,
  0.00012, 0.00010, 0.00009, 0.00008, 0.00008, 0.00009, 0.00010,
  0.00012, 0.00016, 0.00020, 0.00027, 0.00034, 0.00040, 0.00047,
  0.00052, 0.00056, 0.00059, 0.00061, 0.00061, 0.00062, 0.00063,
  0.00064, 0.00067, 0.00069, 0.00071, 0.00072, 0.00075, 0.00078,
  0.00083, 0.00089, 0.00096, 0.00104, 0.00113, 0.00123, 0.00134,
  0.00146, 0.00158, 0.00170, 0.00184, 0.00201, 0.00223, 0.00250,
  0.00278, 0.00306, 0.00333, 0.00362, 0.00395, 0.00434, 0.00478,
  0.00527, 0.00578, 0.00636, 0.00700, 0.00765, 0.00833, 0.00907,
  0.00989, 0.01078, 0.01172, 0.01264, 0.01351, 0.01453, 0.01582,
  0.01741, 0.01925, 0.02130, 0.02359, 0.02620, 0.02935, 0.03294,
  0.03696, 0.04135, 0.04608, 0.05137, 0.05722, 0.06371, 0.07072,
  0.07819, 0.08653, 0.09543, 0.10549, 0.11693, 0.13037, 0.14353,
  0.15716, 0.17126, 0.18586, 0.20094, 0.21650, 0.23256, 0.24910,
  0.26612, 0.28362, 0.30158, 0.32000, 0.33886, 0.35814, 0.37782,
  0.39789, 1.00000)
names(qxJmH20) <- c(sprintf("%d",0:104),"105-")

qxJfH20 <- c(0.00247, 0.00032, 0.00023, 0.00016, 0.00011,
  0.00009, 0.00009, 0.00008, 0.00008, 0.00007, 0.00007, 0.00006,
  0.00006, 0.00007, 0.00009, 0.00012, 0.00015, 0.00018, 0.00021,
  0.00024, 0.00026, 0.00027, 0.00028, 0.00029, 0.00030, 0.00031,
  0.00031, 0.00031, 0.00033, 0.00035, 0.00037, 0.00040, 0.00042,
  0.00043, 0.00045, 0.00048, 0.00053, 0.00059, 0.00064, 0.00069,
  0.00073, 0.00079, 0.00085, 0.00092, 0.00100, 0.00109, 0.00119,
  0.00131, 0.00143, 0.00156, 0.00168, 0.00181, 0.00194, 0.00209,
  0.00225, 0.00242, 0.00259, 0.00280, 0.00306, 0.00333, 0.00358,
  0.00381, 0.00403, 0.00428, 0.00460, 0.00498, 0.00542, 0.00595,
  0.00659, 0.00734, 0.00821, 0.00922, 0.01036, 0.01161, 0.01300,
  0.01456, 0.01641, 0.01858, 0.02108, 0.02397, 0.02724, 0.03103,
  0.03558, 0.04088, 0.04712, 0.05405, 0.06168, 0.07024, 0.07997,
  0.09082, 0.10278, 0.11551, 0.12859, 0.14236, 0.15698, 0.17234,
  0.18848, 0.20540, 0.22312, 0.24165, 0.26099, 0.28115, 0.30211,
  0.32387, 0.34641, 1.00000)
names(qxJfH20) <- c(sprintf("%d",0:104),"105-")
```

上に示したのは、平成 20 年度日本人の男女別 q_x である。 q_x がわかっている場合は、そこから生命表を計算するのは容易であろう。

人口学「死亡の分析」配布資料：中澤 港 <nminato@med.gunma-u.ac.jp>

lifetable.R(3)

```
clifetable <- function(qx) {  
  nc <- length(qx)  
  lx <- numeric(nc)  
  dx <- numeric(nc)  
  Lx <- numeric(nc)  
  lx[1] <- 100000  
  for (i in 1:(nc-1)) {  
    dx[i] <- lx[i]*qx[i]  
    lx[i+1] <- lx[i]-dx[i]  
    Lx[i] <- (lx[i]+lx[i+1])/2  
  }  
  Tx <- cumsum(Lx[nc:1])[nc:1]  
  ex <- Tx/lx  
  data.frame(qx, lx, dx, Lx, Tx, ex)  
}  
  
clifetable(qxJmH20)
```

通常は年齢別人口と年齢別死亡数から始まるので、例えば、昭和 60 年の年齢別死亡率から、Greville の方法で補正した生命表を求めるには、次の枠内を打てば、その次の枠内の結果が得られる。

lifetable.R(4)

```
lifetable(S60ASMR, class=5, mode=2)
```

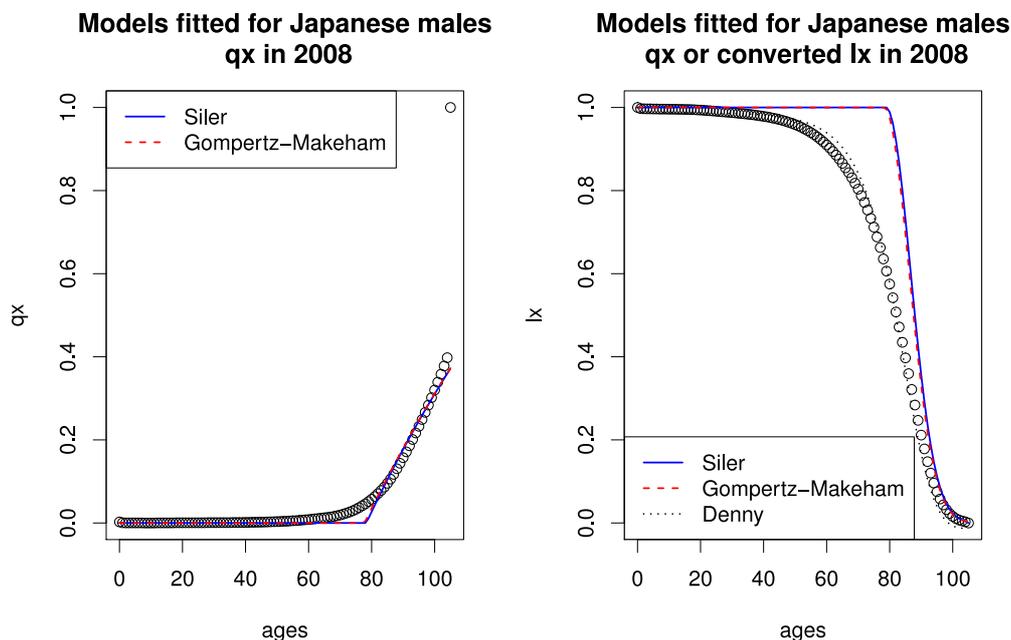
	mx	qx	lx	Lx	Tx	ex
[0-4]	0.0014524735	0.0014455260	100000.00	498193.1	7602349.2	76.023492
[5-9]	0.0002099156	0.0002098010	99277.24	496125.8	7104156.1	71.558761
[10-14]	0.0001642103	0.0001641547	99173.09	495662.0	6608030.2	66.631280
[15-19]	0.0004690423	0.0004685105	99091.70	494878.2	6112368.3	61.683960
[20-24]	0.0005693208	0.0005685192	98859.57	493595.3	5617490.1	56.822927
[25-29]	0.0006039882	0.0006031087	98578.55	492149.6	5123894.8	51.977786
[30-34]	0.0007436492	0.0007423446	98281.28	490494.4	4631745.2	47.127440
[35-39]	0.0010362265	0.0010337771	97916.49	488317.2	4141250.8	42.293701
[40-44]	0.0017388068	0.0017318457	97410.37	484943.1	3652933.6	37.500459
[45-49]	0.0027567075	0.0027393406	96566.87	479527.7	3167990.5	32.806183
[50-54]	0.0045192235	0.0044715872	95244.22	470897.5	2688462.8	28.227043
[55-59]	0.0065107143	0.0064121767	93114.76	458110.4	2217565.3	23.815401
[60-64]	0.0094052904	0.0092050297	90129.42	440276.5	1759454.9	19.521427
[65-69]	0.0154376342	0.0149082601	85981.20	413883.1	1319178.3	15.342637
[70-74]	0.0269410609	0.0253626487	79572.05	372633.3	905295.2	11.377050
[75-79]	0.0486361813	0.0436233035	69481.26	309518.8	532661.9	7.666268
[80-84]	0.0862337753	0.0714027853	54326.25	223143.2	223143.2	4.107465
[85-]	0.1652433121	0.1124108423	34931.02	0.0	0.0	0.000000

R でモデルを当てはめるには、`optim()` 関数を使うが、初期値の与え方が難しい。また、どうしても絶対値が大きいところに残差が引きずられるので、Siler モデルや Gompertz-Makeham モデルで普通に $q(x)$ で当てはめをすると高齢のところの当てはまりは良いが若いところの当てはまりが悪くなる。当てはまりを良くするには、年齢区分別に項を分けて `optim()` を適用すべきであろう。

Denny のモデルは $\ell(x)$ で当てはめをするので生存曲線を描かせたときの当てはまりは非常に良いけれども、残差の絶対値が大きいため、なかなか収束しない（実はこのパラメータでも収束していない）。雪崩モデルはシミュレーションしないと当てはめができないので R ではおそらく莫大な時間がかかると思われ、まだ試していない。

lifetable.R(5)

```
# H20 日本男性への Siler , Gompertz-Makeham , Denny の当てはめと描画
clx <- function(qx) {
  nc <- length(qx); lx <- numeric(nc); dx <- numeric(nc)
  lx[1] <- 1
  for (i in 1:(nc-1)) { dx[i] <- lx[i]*qx[i]; lx[i+1] <- lx[i]-dx[i] }
  lx[nc] <- 0
  lx
}
Siler <- function(a1,b1,a2,a3,b3,t) {
  rval <- a1*exp(-b1*t)+a2+a3*exp(b3*t); ifelse(rval<0,0,rval) }
fSilerM <- function(x) {
  sum((Siler(x[1],x[2],x[3],x[4],x[5],ages[1:105])-qxJmH20[1:105])^2) }
(rS <- optim(rep(0,5),fSilerM))
GompertzMakeham <- function(A,B,C,t) {
  rval <- A + B*C^t; ifelse(rval<0,0,rval) }
fGompertzMakehamM <- function(x) {
  sum((GompertzMakeham(x[1],x[2],x[3],ages[1:105])-qxJmH20[1:105])^2) }
(rG <- optim(rep(0.1,3),fGompertzMakehamM))
Denny <- function(a,b,c,t) {
  1/(1+a*(t/(105-t))^3)+b*sqrt(exp(t/(105-t))-1)+c*(1-exp(-2*t)) }
fDennyM <- function(x) {
  sum((Denny(x[1],x[2],x[3],ages[1:105])-clx(qxJmH20[1:105]))^2) }
(rD <- optim(rep(0,3),fDennyM))
layout(t(1:2))
plot(ages,qxJmH20,
  main="Models fitted for Japanese males\n qx in 2008",ylab="qx")
lines(ages,Siler(rS$par[1],rS$par[2],rS$par[3],rS$par[4],rS$par[5],ages),
  col="blue",lty=1,lwd=2)
lines(ages,GompertzMakeham(rG$par[1],rG$par[2],rG$par[3],ages),
  col="red",lty=2,lwd=2)
legend("topleft",lty=1:2,lwd=2,col=c("blue","red"),
  legend=c("Siler","Gompertz-Makeham"))
plot(ages,clx(qxJmH20),
  main="Models fitted for Japanese males\n qx or converted lx in 2008",ylab="lx")
lines(ages,clx(Siler(rS$par[1],rS$par[2],rS$par[3],rS$par[4],rS$par[5],ages)),
  col="blue",lty=1,lwd=2)
lines(ages,clx(GompertzMakeham(rG$par[1],rG$par[2],rG$par[3],ages)),
  col="red",lty=2,lwd=2)
lines(ages,Denny(rD$par[1],rD$par[2],rD$par[3],ages),col="black",lty=3,lwd=2)
legend("bottomleft",lty=1:3,lwd=2,col=c("blue","red","black"),
  legend=c("Siler","Gompertz-Makeham","Denny"))
```



8 文献

- Alvarez HP (2000) Grandmother hypothesis and primate life histories. *American Journal of Physical Anthropology*, 113: 435-450.
- Carstensen B (2007) Age-period-cohort models for the Lexis diagram. *Statistics in Medicine*, 26(15): 3018-3045.
- CDC: Deaths and Mortality. [<http://www.cdc.gov/nchs/fastats/deaths.htm>]
- Chamberlain AT (2006) *Demography in Archaeology*. Cambridge University Press.
- Charnov EL (1992) *Life History Invariants*. Oxford University Press.
- Denny C (1997) A model of the probability of survival from birth. *Mathematical and Computer Modelling*, 26: 69-78.
- Gage TB (1991) Causes of death and the components of mortality: Testing the biological interpretations of a competing hazards model. *American Journal of Human Biology*, 3(3): 289-300.
- Gage TB, Mode CJ (1993) Some laws of mortality: How well do they fit? *Human Biology*, 65: 445-461.
- Gavrilov LA, Gavrilova NS (1991) *The Biology of Life Span: A Quantitative Approach*. Harwood Academic Publisher, New York.
- Hawkes KJF, O'Connell NG, Blurton-Jones HA, Charnov EL (1998) Grandmothering, menopause, and the evolution of human life histories. *Proceedings of National Academy of Sciences, USA*, 95: 1336-1339.

- 本川達雄 (1992) ゾウの時間・ネズミの時間. 中公新書
- 川上憲人, 小林廉毅, 橋本英樹 [編] (2006) 『社会格差と健康：社会疫学からのアプローチ』, 東京大学出版会.
- Keyfitz N, Caswell H (2005) *Applied Mathematical Demography. 3rd Ed.*, Springer Science+Business Media, Inc., New York.
- 国際人口学会編, 日本人口学会訳 (1994) 『人口学用語辞典』, 厚生統計協会.
- 近藤良 (1980) 1966年(ひのえうま)における先天異常乳児死亡率の見かけ上の増加について. 先天異常, 20: 7-16.
- Mori Y, Nakazawa M (2003) A new simple etiological model of human death. *The Journal of Population Studies (Jinko-Gaku-Kenkyu)*, 33: 27-39.
- Ng E, Gentleman JF (1995) The impact of estimation method and population adjustment on Canadian life table estimates. *Health Reports*, 7(3): 15-22.
- Smith DP (1992) *Formal Demography*. Plenum Press.
- Staetsky L (2009) Diverging trends in female old-age mortality: A reappraisal. *Demographic Research*, 21: 30.
[<http://www.demographic-research.org/volumes/vol21/30/21-30.pdf>]
- 杉本正信, 古市泰宏 (1998) 老化と遺伝子. 東京化学同人
- 舘稔 (1968) 『人口分析の方法』, 古今書院.
- 和田光平 (2006) 『Excel で学ぶ人口統計学』, オーム社.
- Westendorp RGJ, Kirkwood TBL (1998) Human longevity at the cost of reproductive success. *Nature*, 396: 743-746.