

歴史統計を利用した天然痘伝播の定量的分析： 統計疫学モデルを動機とするデータベース 構築について

西浦 博[†]

(受付 2008年1月8日；改訂 2008年2月14日)

要 旨

稀にしか観察されない感染症の流行対策を計画する場合、過去の膨大な統計資料から疫学のおよび生物学的に重要な情報を抽出する作業が欠かせない。天然痘は人類が唯一根絶に成功した感染症であるが、根絶を逆手に取って天然痘ウイルスがバイオテロに用いられる可能性が危惧されている。本稿では、歴史統計を利用した具体的事例として、天然痘の伝播に関する定量的研究の方法について報告する。特に、天然痘の統計疫学モデルを動機としたメタデータのニーズと、対応するデータベースの構築を中心に議論する。単純な統計モデルの例として、検疫および隔離の最適化およびバイオテロ発生時の被害規模予測について、歴史統計の選定から意思決定に至るまでの研究手法を解説する。また、データベースの改善点および歴史統計を利用する上での疫学の問題についてまとめる。

キーワード：歴史統計，データベース，天然痘，疫学，モデル，バイオテロリズム。

1. はじめに

天然痘は人類が唯一根絶に成功した直接伝播する感染症である (Fenner et al., 1988)。しかし、根絶を理由に感受性宿主が増えたこと等を逆手にとって、天然痘ウイルスがバイオテロに用いられる可能性が危惧されている (Henderson et al., 1999)。21世紀以降も、その疫学や病理学及びバイオテロ対策の計画などに関して詳細な調査・研究が継続して行なわれてきた。

天然痘の予防医学は、歴史的に統計学的手法の開発と平行して発展してきた。その典型的史実の1つとして、天然痘ワクチン接種の効果に関する疫学的な評価手法開発の経緯が挙げられる。特に、今日の疫学における相対的リスクの概念は天然痘を中心とした感染症に対するワクチン接種効果の推定を動機に提案されたものである (Greenwood, 1935)。天然痘は他の感染症と比較すれば診断が比較的容易であり、非常に豊富なデータが古くから蓄積されてきた疾病の1つである。しかし、今日の統計学的諸手法の発展を待たずに天然痘が根絶されたことから、それらデータを統計学的あるいは数理疫学的に詳細に分析した研究が充分でなく、理論的・数理的に明らかにされていない基礎的事項に関する課題が未だ数多く残っている。同様のことは1918-19年に観察されたスペイン風邪の汎世界的流行(パンデミックインフルエンザ)を代表と

[†] ユトレヒト大学 理論疫学：Theoretical Epidemiology, University of Utrecht, Yalelaan 7, 3584 CL, Utrecht, The Netherlands; h.nishiura@uu.nl

して、他の感染症についても当てはまるだろう (Chowell and Nishiura, 2008).

天然痘の伝播を理論的に理解するための数理的アプローチとして、主に2つの方法が挙げられる。1つは数理疫学的な手法によるものであり、感染の個体群動態を基礎にした非線形モデルが主に用いられる。これは感染過程をボトムアップ方式で表現した生態学的モデルに基づくもので、主として流行メカニズムに関する定性的な理解を促すことに役立てられる (Koopman, 2004; 西浦・稲葉, 2006)。もう1つは、過去の観察データを利用した生物統計学および疫学的な基礎研究である。基礎的な統計疫学モデルを利用した後者の例として、歴史統計を用いたワクチン免疫持続期間の推定や (Nishiura and Eichner, 2006, 2007a; Nishiura et al., 2006)、妊産婦の致命割合とワクチン接種効果の推定 (Nishiura, 2006)、潜伏期間や感染性期間の推定 (Nishiura and Eichner, 2007b, 2008; Nishiura, 2008) などが挙げられる。

感染症の理論疫学(数理モデル)の特徴の1つとして、統計疫学モデルによって伝播の基礎的メカニズムが定量的に明らかにされない限り、現実的な伝播動態を適切に議論することが難しい点に注意しなければならない (Ferguson et al., 2003)。例えば、2001年9月11日の同時多発テロ事件以降、天然痘のバイオテロとその対策が数理モデルによって検討されてきたが (Halloran et al., 2002; Kaplan et al., 2002)、それらのモデルは伝播に関して(観察に基づかない)非現実的想定を数多く含んでいた。そのため、モデルを用いた研究毎で、被害規模予測値や予防接種戦略などの政策提言が大きく異なる結果を得た。後の研究で、歴史統計を利用して伝播メカニズムを定量的に詳しく分析することによって、天然痘伝播の多くが接触者追跡調査(contact tracing)と隔離のみで制御できること(必ずしも義務的な事前のワクチン集団接種が必要ないこと)が明らかにされた (Cooper, 2006; Eichner, 2003; Fraser et al., 2004; Klinkenberg et al., 2006)。つまり、伝播をボトムアップ方式にモデル化する以上、その内在性の感染過程を詳細に定量化した上でモデルを構築することは不可欠であり、その一助として歴史統計の分析が非常に有効であることが示された。

では、何のデータ(what)をどのように(how)整理すれば良いのだろうか。それは微生物学や公衆衛生学・疫学など現場のニーズに基づいて考えるとわかり易い。例えば、ワクチン接種に伴う副作用頻度について知りたい場合、他の疾病に関する疫学研究では系統的レビューが実施されることが多いが、こと天然痘に関しては現代的な統計資料が限られている。これまでは、個別の研究者が文献を整理することによって分析が行われてきた(例えば、Kretzschmar et al., 2006)。また、ワクチン接種効果に関する評価の1つとして独立性の検定を代表とする様々な仮説検定が行なわれることがあるが (Nishiura, 2006)、歴史統計を集積したところでランダムサンプリングが行なわれた天然痘の自然感染に関する研究成果が存在せず、明示的なメタアナリシスによってデータを分析することは容易でない (Petitti, 2000)。残念ながら、これらの問題を完全に解決する手段はないが、天然痘流行の歴史統計を蓄積したデータベースを構築し、可能な限りにメタデータを整理して系統立った分析を実現することが望ましいと考えられる。

本稿では歴史的な統計資料に基づいて天然痘の感染を定量化する手段について、著者の経験を基に事例を通じて議論する。メタデータの種類と量の特徴のために、データベースから情報を抽出する解析手法に関して特定のデータマイニング技法を用いる事例を議論しないが、統計疫学モデルを動機とするデータベースの構築と関連する諸問題について記録にとどめておきたい。特に、統計疫学モデルが必要とするマイクロスケール(個体レベル)データのニーズを理解することを通じて、欠かすことのできない歴史統計の種類・特徴を明らかにし、それに対応するデータベースを構築する際の疫学的問題について解説する。

2. 歴史研究におけるデータベース

まず最初に、これまで歴史研究の領域で利用されてきたデータベースの概要と特徴を紹介する。歴史学では、過去の史料の評価・発掘過程に欠かせない情報源の1つとして、データベースが頻繁に利用される。その収録内容によって分類すると、主に抄録・索引・目録型、画像型および素材型のデータベースが活用されてきた。抄録データベースは医学領域のPubMedに代表されるような文献のキーワード検索を目的とするが、歴史研究では目録型が好まれることも多い。歴史学にとって史料の特定と発見は最も重要な研究過程の1つであるため、その作業を簡便にする手段の1つとして様々な所蔵史料や古文書および文献目録に関する索引・目録型データベースが開発されてきた。また、非専門家も対象に公開されるデータベースとして、史料の画像型データベースが挙げられる。これは文字通り、検索によって画像そのものを再利用できるようにしたデータベースを意味するが、新聞記事を画像として蓄積したデータベースは典型的な例と言えるだろう(マイクロフィルムを含む)。また、最近では著作権保護期間が満了した図書が画像型データベースとして公開されることもある。歴史学では、史料に記された文章そのものを詳しく吟味するために文学的・社会的手法を用いることが多く(速水, 2006; 飯島, 2000)、これらのデータベースは歴史学の研究環境の向上に非常に有用である。

データベースに統計資料が収録されたものはどうだろうか。素材型データベースは、遺伝情報を代表として研究内容そのものに直結するデータが収録されたものを指すが、データ抽出のために様々な統計学的手法が用いられる点が特徴的である。しかし、歴史学の分野では素材型データベースが公開あるいは共有されている事例が非常に少ない。わが国における優れた例外として、経済史領域における長期経済統計データベースが挙げられる(藤野, 1978)。また、疾病統計に関する最近の数少ない試みとして、わが国の死因別死亡者数統計や法定伝染病統計を収録・公開した長期暦象データベース(平山・友部, 2004)や、歴史学的に第一級の研究書として活用される『日本疾病史』のデータベース化(浜野, 2004)などが挙げられる。これらは感染症流行に関する時系列の報告データを蓄積したものであり、主に政府刊行物などの2次データを対象に統計データが収録されている。天然痘の流行に関して、一部の例外を除けば1次データを手に入れることは極めて困難であり、ほとんどの研究では2次データを有効に活用して分析が行なわれる。

感染症流行に関する時系列の統計データが蓄積された場合、過去の疫学的動向を理解するのに役立つ記述統計としての資料価値に加えて、主に以下のような時系列関連の疫学的分析が可能である:

- 感染症流行の季節性と周期性
- 感染症流行に対する気候の影響
- 感染症流行に対する予防接種や衛生状態の影響
- 感染症流行と人口動態の関係
- 感染症を引き起こす病原体の進化動態分析

しかし、時系列データから得られる情報は非常に限られたものである。例えば、第1章で議論したように、バイオテロに関連した危機管理対策を動機に分析を行なう場合、蓄積された時系列のデータが研究ニーズに対応しないことも少なくないだろう。即ち、天然痘の伝播メカニズムを定量化する目的に応じた統計データとして、時系列の2次資料は部分的にのみ役立つものである。

そこで、疫学あるいは生物統計学における研究ニーズに応じた素材型データベースが求められる。「天然痘の伝播を定量化するための素材」あるいは「バイオテロ対策に直接的に役立つ

メタデータ」とはどのようなものか、その性質を理解するために、第3章で具体的な統計疫学モデリングの事例を考えていこう。

3. 歴史統計から抽出すべきデータの性質：統計疫学モデル

天然痘の伝播や流行拡大を適切に定量化するためには、詳細な感染・発病のメカニズムを理論的かつ数理的に理解しなければならない。具体的な例を挙げるならば、「感染した後にいつ症状が出現するのか」、「どの病期で2次感染が起こるのか」、「発症するとどれくらい重症なのか」、に関して明らかにすることは不可欠だろう。これら基礎的な伝播動態の機構は(公衆衛生対策が行なわれていない下で)天然痘の流行拡大に関する様相を知る上で最も重要な内在性の伝播過程(intrinsic transmission process)と呼ばれる。第1章で述べた通り、天然痘のバイオテロ対策の計画策定に数理モデルを用いる場合、様々な公衆衛生対策に関して(観察に基づかない)想定を施すよりも内在性の伝播過程を確実に捉えることの果たす役割が非常に大きい(Cooper, 2006; Eichner, 2003)。ここでは上述した3点に関する基礎的な統計疫学モデルを利用した議論を提示することによって、必要とされるデータの性質を明らかにする。具体的には、潜伏期間、感染性期間および致命割合の3つの推定問題を例として取りあげる。

3.1 天然痘の潜伏期間

感染症の潜伏期間(incubation period)とは、感染時刻と発症時刻の間隔を意味する(Nishiura, 2007)。これまでの天然痘対策の経験から、感染後最初の症状は10-14日間程度で観察されることが知られてきた(Mack, 1972)。潜伏期間は様々な臨床的・疫学的、及び生態学的用途に役立つが、中でも特に適切な検疫期間の決定をするために利用することが出来る(Farewell et al., 2005)。ここで検疫(quarantine)とは、暴露を経験した(感染した可能性のある)健康な者が発症しないかどうかを見極めるために行なわれる監視及び行動制限を指す。歴史的には、天然痘の検疫期間は14-16日間程度が適当とされていたが、これは専門家の意見や潜伏期間データの観察に基づく直感的示唆によるものであり、潜伏期間の確率分布を利用した明示的な分析によるものではなかった。

暴露を経験した個体の行動を制限する場合、これまでに観察された潜伏期間の最大値を検疫期間に定めれば検疫の有効性を保証することができるかも知れない。しかし、「観察された最大値」は常にサンプル数に依存しており(サンプル数が大きくなればなる程、新たな最大値を発見する確率が高くなる)、さらに個体の行動を制限する場合は法的および倫理的な観点から最短かつ効果的な検疫期間を定めることが望ましい。一方で、感染時刻が観察された歴史的な統計資料は限られている(例えば、暴露を1日のみ経験した症例データが限られている)。そのため、簡単な確率密度関数によって潜伏期間の頻度を統計モデルとして扱い、その分布の右裾を分析して「これ以降、更なる発症者を認める可能性が極めて低くなる」ような感染後時刻を検討することが望ましい。それには分位数(quantile)が有用であり、例えば潜伏期間の理論的分布における99パーセンタイルを検疫期間とすれば、それは「この感染後時刻までに99%の感染者が発症する」であろう理論的な潜伏期間を示唆する。

図1は暴露を1日しか経験しなかった(感染時刻が1日に限定できる)天然痘症例の集積データであり、131名の報告を利用した潜伏期間の推定例である(Nishiura, 2008)。平均値(標準偏差)および中央値は12.5(2.2)および12.0日である。ここで症例*i*の潜伏期間 t_i が対数正規分布に従うと想定すると、その密度関数 $f(t_i)$ は以下で与えられる:

$$(3.1) \quad f(t_i; \mu, \sigma^2) = \frac{1}{t_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(t_i) - \mu)^2}{2\sigma^2}\right).$$

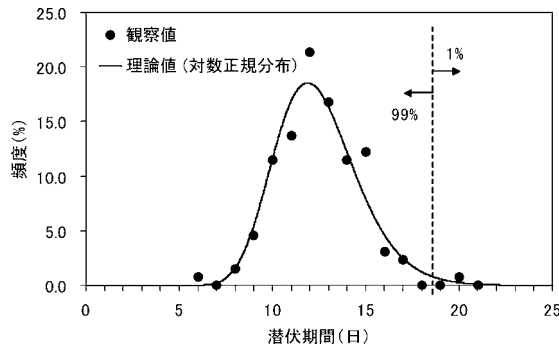


図 1. 対数正規分布を想定した天然痘の潜伏期間分布 ($n = 131$)。縦方向の点線は 99 パーセントイルを与える潜伏期間の最尤推定値 (Nishiura, 2008)。

ここで $\exp(\mu)$ および $CV = \sqrt{\exp(\sigma^2) - 1}$ は潜伏期間の平均値及び変動係数を与える。サンプル数 n 名の観察データに対して μ 及び σ を推定する尤度は

$$(3.2) \quad L(\mu, \sigma^2) = \prod_{i=1}^n f(t_i; \mu, \sigma^2)$$

によって与えられることは明らかである。式(3.2)を適用するだけで、99 パーセントイルを与える潜伏期間の最尤推定値は 18.6 日と推定され、その 95% 信頼区間は 16.8–22.2 日と推定された。このことから「少なくとも 23 日間の検疫を実施すれば、99% 以上の感染者がそれ以前に発症することが期待される」ことを理論的根拠として示すことが出来る。どの分位数を検疫期間とするかは、仮説検定における p 値の決定と同じく任意であるが、より明確に検疫期間を提案したい場合(例えば、さらに感染性などを考慮に入れて検疫期間を決定したい場合)、更なる理論的分析の材料として上記の結果を用いることが出来るだろう。

3.2 感染性期間

感染性期間 (infectious period) とは、感染個体が 2 次感染者を生み出す能力を有する期間である。感染性期間の情報が既知であれば、「流行を制御するために、感染者を(感染後あるいは発病後)いつの時点までに隔離することによって、2 次感染がどの程度予防できるか」など、隔離に関連する疑問に対して明示的に回答できる。ここで隔離 (isolation) とは、感染者を他者から分離することによって 2 次感染を防ぐことを目的とする行動制限と空間的配置を意味する。

感染性期間を明らかにする 1 つの手段は、感染した生体内において病原体の微生物量を生物学的に定量することである(例、ポリメラーゼ連鎖反応 (PCR) によるウイルス量の測定)。しかし、発病・診断の後でなければ感染個体の微生物量が分析されることがほとんどないため、その観察は発症後に限られる傾向がある。天然痘に関しては、根絶計画中の研究において発症者の「ウイルス陽性の期間」が検討されたが (Downie et al., 1961; Sakar et al., 1973)、発熱直後の病期についてごく少数のサンプル数しか得られなかった。さらに、最も重要な問題として、各病期における「ウイルス量」は実際の 2 次感染の相対頻度と必ずしも直接的に関係がある指標でなく、さらなる情報(例えば、接触の頻度、程度および様式など)が明らかでない限り病原体の体内量に基づく感染性を解釈することが難しい。

そこで、微生物学的手法でなく、疫学的情報のみを用いて同問題を部分的に明らかにする手段を検討してみよう。その一手段として、2 次感染の病期齢 (disease-age) に対する相対頻度を推定する方法が挙げられる (Nishiura and Eichner, 2007b)。ここで病期齢とは発症後の経過時

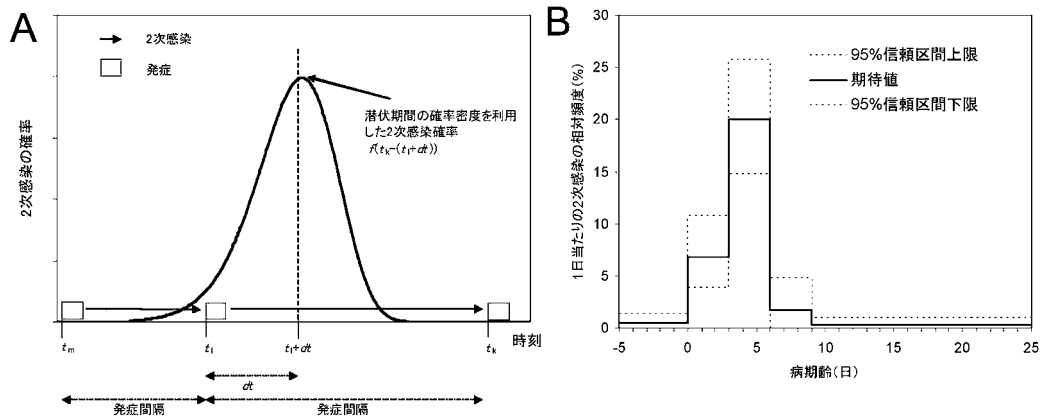


図 2. 天然痘の相対的感染性の推定. A. 伝播ネットワークを基にした 2 次感染の相対頻度推定に関する理論図. 発症間隔は, 感染源が 2 次感染を起こした病期齢と 2 次感染者の潜伏期間の和で与えられる (Nishiura and Eichner, 2007b). B. 相対的感染性の期待値とその信頼区間. 病期齢 0 は各感染者の発熱時刻を意味する.

間を意味し, 天然痘の場合は発熱時刻を 0 とした場合の発病後時刻を指す. 接触者追跡調査の結果に基づく伝播ネットワーク (誰が誰に 2 次感染を引き起こしたか) が明らかな歴史統計データを利用すれば相対的 2 次感染頻度を推定することができる. 伝播ネットワークに加えて各症例の発症時刻が記録されていれば, 感染源の発症から 2 次感染者の発症までを意味する発症間隔 (serial interval) の情報を抽出できるからである.

ここで発症間隔と潜伏期間の両方が既知の場合を考えよう (図 2A). 興味のある相対的感染性は「(感染源が) 発症後, 2 次感染を起こすまでの時間」を集積した情報である (これを病期齢 l 日目とする). 図 2A に示した通り, 発症間隔 s は l と (2 次感染者の) 潜伏期間 f の和で与えられる:

$$(3.3) \quad s = l + f.$$

ここで (3.3) の各々について確率密度関数を用いて分布を考えると

$$(3.4) \quad s(t) = \int_0^t l(t - \tau) f(\tau) d\tau$$

のような単純な畳み込みが与えられる. 即ち, 伝播ネットワークから $s(t)$ の観察データが抽出され, さらに 3.1 節で推定したように潜伏期間の密度関数 $f(\tau)$ が既知である場合, (3.4) から 2 次感染の相対頻度 $l(t - \tau)$ が逆計算できるのである. ここで, 観察データにおける隔離やワクチン接種の影響などを一旦は無視して 2 次感染の相対頻度を考える (同問題は (3.6) 以降に議論する). 長さ t_i の発症間隔を有する 2 次感染 i を合計 n 例だけ観察したとき, $l(t_i - \tau)$ を構成するパラメータを推定する尤度は

$$(3.5) \quad L = \prod_{i=1}^n s(t_i) \\ = \prod_{i=1}^n \int_0^{t_i} l(t_i - \tau) f(\tau) d\tau$$

で与えられる.

図 2B に推定された 2 次感染の相対頻度を示す (Nishiura and Eichner, 2007b). 2 次感染の相対頻度は以下で議論するように複数の内的・外的要因によって与えられるため, 単純かつ妥当なパラメトリック関数を想定することが容易でない. そのため, 図 2B では階段関数を用いており, パラメータの識別性及び生物学的なイベントに関する大まかな時刻 (発熱や発疹など) を反映した任意の 5 区間を想定している. 図 2B より, 発疹が出現する病期齢に該当する発症後 3-5 日目において 2 次感染が最も高頻度起こったこと (日別頻度 20.6 (95% 信頼区間: 15.1, 26.4) %) が推測される. 累積相対頻度を見た場合, 発症後 3 日目までに 23.7% の 2 次感染が発生したものと推定された. 累積頻度は感染者の隔離を遅くとも開始すべき目標日の設定に役立つ. 例えば, 隔離によって感染性宿主の 2 次感染が防止される効果が仮に 100% だとすると, 発症後 3 日未滿で全ての感染者を隔離すれば, 病期齢 3 日目以降の 76.3% の 2 次感染が効果的に予防されることが期待できるだろう. さらに 1 人当たりの感染者が生出す 2 次感染者数の平均値 (基本再生産数, R_0) がわかれば, 隔離が行なわれている下で生出される 2 次感染者数の期待値が計算できるだろう.

しかし, ここで推定された 2 次感染の相対頻度に関する解釈は慎重に行なわれなければならない. その理由に関して, 以下に再生方程式 (renewal equation) を用いて考えてみよう. 時刻 t における新規感染者数を $I(t)$ とすると, それは時刻 t および病期齢 σ における感染者 1 人当たりの瞬間的な 2 次感染率を意味する再生産核 $A(t, \sigma)$ と時刻 t よりも前に感染した者 $I(t - \sigma)$ によって与えられる (Murray, 2002; 稲葉, 2002; 瀬野, 2005):

$$(3.6) \quad I(t) = \int_0^{\infty} A(t, \sigma) I(t - \sigma) d\sigma.$$

ここで興味の対象である 2 次感染の情報は $A(t, \sigma)$ に含まれている. 例えば, 質量作用の法則 (the law of mass action) に従って感染動態が説明されると仮定した場合, 予防接種やその他の公衆衛生対策のない場合における再生産核 $A_0(t, \sigma)$ は

$$(3.7) \quad A_0(t, \sigma) = S(t) \beta(\sigma) \Gamma(\sigma)$$

のように分解できる. ここで $S(t)$ は時刻 t における感受性宿主の密度であり, $\beta(\sigma)$ および $\Gamma(\sigma)$ はそれぞれ病期齢 σ における感染率および感染性宿主の生残率である. 興味の対象である相対的な感染性は $\beta(\sigma) \Gamma(\sigma)$ を用いて与えられるだろう. しかし, 上述した統計モデル (3.3) - (3.5) は流行時刻に関係なく集積されたデータを用いて分析しているため, 感受性宿主の密度減少が大きく再生産核に影響するような事例においては, 本来は感受性宿主減少の影響を無視できないことに注意しなければならない. さらに, 流行途中でワクチンの緊急集団接種などが施されることによって生じる時刻に対する 2 次感染の相対的減少効果を $H_1(t)$ とし, 同様に各個体が発病後に隔離されるなど病期齢に対する 2 次感染の相対的減少を $H_2(\sigma)$ とすると, 様々な公衆衛生対策を加味した (3.6) の再生産核は

$$(3.8) \quad A(t, \sigma) = H_1(t) H_2(\sigma) A_0(t, \sigma)$$

で与えられると考えることが妥当であろう. ここで仮に 1 つの流行データのみを基に発症間隔の情報を抽出したとすると, 再生産核を時刻 t に対して積分したものが (3.4) における $l(\sigma)$ であるから

$$(3.9) \quad \begin{aligned} l(\sigma) &= \frac{\int_0^{\infty} A(t, \sigma) dt}{\int_0^{\infty} \int_0^{\infty} A(t, \sigma) dt d\sigma} \\ &= \frac{\beta(\sigma) \Gamma(\sigma) H_2(\sigma) \int_0^{\infty} S(t) H_1(t) dt}{\int_0^{\infty} \int_0^{\infty} A(t, \sigma) dt d\sigma} \end{aligned}$$

が推定された 2 次感染頻度を反映している。上記のような変数分離の想定が妥当である場合、隔離政策が 2 次感染相対頻度の推定値に大きな影響を及ぼすことがわかる。例えば、図 2B のデータソースは隔離の実施を無視してデータを集積した上で分析した結果であり、病期齢の後期(特に、発疹出現後)における 2 次感染の頻度が低く見積もられた可能性が高い。つまり、(3.9)において特に $H_2(\sigma)$ が別途に推定できない限り、たとえ 2 次感染の相対頻度が推定できたとしても、真の「相対的感染性」を推定することが極めて難しい。同様の問題は歴史統計を利用した推定作業において常に注意すべき事項であり、与えられる情報の理論的解釈やモデル想定の欠陥について細心の注意を払わなければならない。

3.3 致命割合

致命割合(case fatality)とは発症者の間で死亡する者の割合を指す。致命割合を詳細に知ることは、バイオテロ発生時の被害規模(死亡者数)を予測する上で欠かせない。また、被害規模だけでなく、流行時の入院者数や埋葬数などに関する社会的負荷を見積もる場合においても致命割合が必要である。天然痘の根絶計画の途中、20 世紀中の限られた文献資料を用いて致命割合が系統的にレビューされたことがあるが(Shafa, 1972)、国・地域別で検討した致命割合の分散が非常に大きく推定された。致命割合が高いことで知られる天然痘(*variola major*)を対象を限った場合、最も低い地域の粗推定値が 10% 程度であるのに対し、東パキスタン(現バングラデシュ)やインドでは各々 26 および 36% と高い推定値が得られた。その毒力(virulence)の違いを決定する 1 つの要因として、最近になってウイルスの遺伝情報と致命割合の関係が分析されたが(Esposito et al., 2006)、致命割合は遺伝情報以外にも様々な要因によって修飾を受ける。それにも関わらず、これまでの数理疫学研究(例えば大規模シミュレーションを利用した政策提言)では、頻繁にワクチン未接種者の致命割合を一律 30% と雑に想定する傾向が見られた。

統計学的に致命割合を書き換えるならば、それは感染・発症が与えられた場合の死亡の条件付き確率、 $\Pr(\text{death}|\text{onset})$ 、である。この確率は病原体および宿主の特徴によって大きく変動する。ウイルスの遺伝情報は病原体の特徴の 1 つである。一方で、最も重要な宿主側の情報として、「感染を経験する年齢」と「過去のワクチン接種歴」の 2 つが挙げられる。例えば後者に関して言えば、過去にワクチン接種歴がある者が、接種日から長い年月が経過していようとも天然痘の重症化・死亡から免れる可能性が高いことが明らかにされている(Nishiura and Eichner, 2006; Nishiura et al., 2006)。ここでは単純な議論のみを扱うこととして、ワクチン未接種者における致命割合について(前者の年齢別の致命割合について)考えよう。図 3 はワクチン未接種者の中で年齢 a における致命割合、 $c(a)$ 、を歴史統計から抽出し(Barry, 1889; Rutten, 1997)、以下の単純なモデルを外挿したものである($\alpha, \beta, \gamma, \delta$ はパラメータ)。

$$(3.10) \quad c(a) = \alpha \exp(-\beta a) + \gamma(1 - \exp(-\delta a))^2$$

成人の発症者は小児や高齢者よりも死亡する確率が低いため、年齢に依存する致命割合は一般的に(3.10)で与えられるような U 字型のパターンを呈する。年齢別の致命割合の多くは年齢を離散化して報告されることが多いので、ここで年齢階級別の報告データを考えよう。発症者の年齢群 i における年齢中央値を a_i とし、その年齢群では M_i 名の死亡と N_i 名の生存が観察された場合、上述の 4 つのパラメータを推定する尤度は

$$(3.11) \quad L(\alpha, \beta, \gamma, \delta) = \prod_i c(a_i)^{M_i} (1 - c(a_i))^{N_i}$$

で与えられる。

年齢だけでなく、致命割合は宿主に見られる他の生物学的背景にも大きく修飾される。最も典型的な例として、天然痘が重症化する頻度が著しく高い妊産婦の感染が挙げられる(Rao et

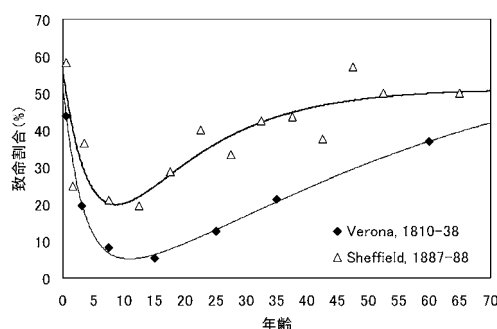


図 3. 年齢別の天然痘の致命割合推定. ヴェローナ(イタリア, 1810-38)およびシェフィールド(英国, 1887-88)におけるワクチン未接種者の年齢別の天然痘致命割合 (Barry, 1889; Dietz and Heesterbeek, 2002; Rutten, 1997). 通常, 急性感染症の致命割合は乳幼児と高齢者の 2 者で高い U 字型のパターンを呈する.

al., 1963). 図 3 から明らかな通り, 健全な成人の致命割合は低いにも関わらず, 妊産婦の感染者は 3 割強が死亡に至ることが知られている. 系統的に歴史統計を分析した結果, 妊産婦を除いた成人の粗致命割合は 12.7 (95%信頼区間: 11.2, 14.3) %であるのに対し, 妊産婦のそれは 34.3 (95%信頼区間: 31.4, 37.1) %と高いことが明らかにされている (Nishiura, 2006). その他の既往・基礎疾患(糖尿病や癌など)を有する者において死亡確率が高いだろうことも容易に想像される. こういった単純な粗推定値であろうとも, 妊産婦とその他の致命割合を分離して推定・比較するためには, 個体レベルの属性を示すデータ(例. 妊娠やワクチン接種歴)が豊富であることが必要なことに気付かれない.

4. 統計疫学モデルを動機とするデータベース構築

第 3 章で述べた統計疫学モデルから得られる情報は, 天然痘のバイオテロ対策において公衆衛生上の意思決定に広く役立てられる. 潜伏期間の分布は最も適切な検疫期間の決定に欠かせない. また, 感染性期間のモデルと推定値は隔離を開始すべき発症後時刻やその効果の推定に役立つであろうし, 致命割合の推定とそれを特徴付ける要因の統計モデルは被害規模の予測に有用であろう. それら疫学的及び生物統計学的な基礎的情報の分析に必要とされる統計データは少なくとも以下の情報を含まなければならない:

- 感染個体の感染時刻と発症時刻
- 伝播ネットワークで繋がった個体間の発症間隔
- 感染個体の転帰(生存あるいは死亡)及び, その背景要因としての人口動態変数(性や年齢など)や既往疾患(妊娠や喘息など)

著者はこれまでに, 以上のような統計疫学モデルの動機に基づく必要情報を歴史的な 2 次データから蓄積・抽出したデータベースを開発・構築してきた. 上述のようなデータは, 政府刊行物などで広く公表された歴史統計資料から得ることが難しい. そのため, 世界各地の流行の報告書や論文について言語に関係なく収集し, 優れた文献を精選しながら手作業でデータを蓄積する地道な努力を要する. 2007 年 12 月末現在で入力完了した文献数は 3940 編で, その入力作業は著者ら関連研究者による個人レベルの努力で実現することができた. しかし, 入力のクロスチェックが終了した文献は未だ 500 編に満たない. これは個別の素材型データベースを構築する歴史研究の難しさを指し示すものである.

対象とする文献は主に19世紀以降の出版物が多い。それは生物統計学が発展した社会的な歴史背景に主な理由を認める。エドワード・ジェンナーによる牛痘(人類初の天然痘ワクチン)の研究成果が発表されたのは18世紀最後であるが、当時にワクチン接種の効果を評価する手段は個別の症例報告を通じた議論に基づいていた(“minute inquiry into individual cases”)。つまり、少数の例外を除いて、19世紀前半までのワクチン接種効果の議論は症例報告か大まかな感染者数の記述のみに基づいており、症例データを収集して分析的に吟味したものが少なかった(Simon, 1887)。しかし、1850年前後に英国を含む欧州各国で義務的な予防接種が法制化され、その効果を証明する手段として(法制化を正当化する手段として)天然痘の統計データを蓄積・分析する必要性が自然に生じたのである。しかし、その社会史を逆に考えれば、19世紀以降に観察・報告された流行データのほとんどは、必ずと言って良いほどワクチン接種が実施された下で観察された自然感染を反映している。ワクチン接種による集団免疫や様々な生物学的効果を排除するためには18世紀以前のデータも非常に貴重である。ワクチン接種の個人・集団レベルの影響を除いた分析を可能にするために、本データベースは16-18世紀の観察データも追加して収録した。

文献の検索手段(2次データを収集する手段)として、最近の系統的レビューに見られるような抄録データベースのみに基づく方法を取ることが難しい。Web of ScienceやMEDLINEなどの文献データベースは1950年以降に発表された文献を主な対象としているからである。それ以前の医学文献を対象としたデータベースとして、米国国立医学図書館によるIndexCat(<http://indexcat.nlm.nih.gov/>)が公開されている。また、その他にも英語以外の言語を含めて、歴史的な医学文献が収録された抄録データベースが多数公開されている。しかし、これらのデータベースは製作した研究機関の所蔵史料の範囲内で構築されており、それらが対象とする言語範囲が狭い傾向がある。つまり、史料が単一の抄録データベースに網羅的に整理されていることは非常に稀である。そのため、文献データベースの検索に加えて、天然痘の専門書籍や各論文で参照された引用・参考文献を全て追跡して2次データを入手することが欠かせない。原則として、詳しい関連文献の引用・参照が習慣化されていなかった19世紀前半に至るまで(それよりも古い引用文献がなくなるまで)引用文献を繰り返し追跡すること以外に同目的を効果的に達成する手段がない。その作業は言語に関わらず実施されなければならない。現在までに、翻訳可能な言語として、英語・日本語に加えて東西の欧州各国の言語、アラビア語、中国語の文献に関して入手可能な資料が整理された。残念ながら、アフリカ地域や東南アジア地域の文献について分析・追跡する機能的な手段が現時点で得られていない。南米地域はスペイン語・ポルトガル語で執筆された文献が多く、それらの出版物については追跡中である。

4.1 データベース概要

文献によって掲載される疫学的情報が異なり、統計疫学モデルの用途別で必要データが異なることから、本データベースの種類としては、多くの素材型データベースが採用している関係モデル(relational model)を利用した関係データベースが有用である(Codd, 1990)。対象とするマイクロデータは政府刊行物に掲載された時系列データを含むが、主に接触者追跡調査や小規模人口における流行の詳細な観察記録および日別の発症者数を記録した短期的流行の報告書に焦点を当てて文献が整理されている。多くの2次データは各流行の基礎的情報(場所、時刻、感染者・死亡者の総数)の報告であるか、あるいは何らかの天然痘に関する医学的分析(例、妊娠と天然痘)が施された原著論文が多い。

データベースにおける複数の関係変数と属性の設計例を表1に示す。太線は候補キーかつ主キーであり、下線が外部キーである。外部キーは他の候補キーのいずれかに合致するよう設計されている。文献変数も含め、各変数の属性は感染個体別に纏められている。全ての個体別

表 1. 天然痘疫学の歴史統計データベース：関係変数と属性の設計. 各項目(1)-(6)は関係変数で, ()内は属性を表す. 太字は主キー (primary key). 下線は外部キー (foreign key) で整合性制約に用いられる.

関係変数	属 性
(1) 流行	(<u>流行ID</u> , <u>個体ID</u> , 性, 年齢, 流行年, 国名, 地域名, 種類)
(2) 予防接種・基礎疾患	(ワクチン <u>接種歴</u> , <u>個体ID</u> , <u>流行ID</u> , 年齢, ワクチン接種回数, 予防接種後の経過時間, 妊娠, 免疫不全)
(3) 感染・発病	(<u>個体ID</u> , <u>流行ID</u> , <u>ワクチン接種歴</u> , 感染日, 発熱日, 発疹出現日, 感染源個体ID, 発症間隔, 2次感染数)
(4) 診断	(<u>個体ID</u> , <u>流行ID</u> , <u>診断方法記載</u> , 臨床/ウイルス)
(5) 転帰	(<u>個体ID</u> , <u>流行ID</u> , <u>隔離開始日</u> , 入院日, 生存・死亡)
(6) 文献	(<u>流行ID</u> , <u>個体ID</u> , 著者, 著書・雑誌名, 出版年, 出版地)

表 2. 天然痘疫学のデータベースにおける流行変数の例. 各行は感染個体を表す. 最下段のインド(1959年)における症例のように, 各個体について空白の情報(上記例では年齢が欠測値)が数多く見られる. 「種類」は天然痘の分類で致命割合が高い(variola major)か低い(variola minor)かの違いを指す.

流行ID	個体ID	性	年齢	流行年	国名	地域名	種類
00001	00001	F	28	1870	Germany	Hamburg	Major
00312	00031	F	18	1878	USA	Philadelphia	Major
02492	00013	F	18	1893	Netherlands	Rotterdam	Major
03330	00002	F	16	1913	Australia	New South Wales	Major
03411	00013	F		1959	India	Madras	Major

データは手作業で入力した. データソースと含まれるデータ種類が多岐に渡るため, データベースは最も単純な関係データベースの構造に従って設計されている. 上述の通り, データソースの文献が対象とする研究課題が各々で異なり, その結果として情報を持たない空白の属性(欠測値)が多数存在する. 表 2 は流行変数が含む個別データを例示している. データベースが最も有効な点として, 表 2 に示された個体データが常に文献や他の変数と結合できることが挙げられる. これはメタデータのファイルフォーマットの典型的形式に従っており, 文献変数がヘッダ部のメタ情報に対応し, その他がデータ部に対応する.

最も特徴的な点は, 全ての属性に関して完全な情報を与える感染個体が文献上で存在しないことである. 例えそうだとした場合, 研究テーマとして属性ひとつひとつが解決すべき事象に直結することから, 不完全データであろうとも欠測値に関係なくデータベース上に登録している. 後に必要情報に応じて結合することで, 各研究の分析が行なわれる. また, もう一つの特徴として, データが2次資料から抽出されているため, 感染個体が同一資料内で識別できないことがほとんどであることが挙げられる. 例えば, 文献(論文)Aの中で, 仮に図1・表1および表2の3箇所収録すべきデータが記載されていたとしても, 各図表内に示される個体が図表間で識別できる(繋がっている)ことが少ない. そのため, データベース上に入力される各流行の感染個体数は流行中に実際に観察された総感染者数よりも必ず多くなる.

4.2 データベースの問題点

以上のように, 天然痘の疫学データベースの構造は従来の関係データベースの設計に従って構築され, 簡単な入出力作業が可能な仕組みに作られている. ただし, 歴史的な文献に含まれる2次データを個別に利用している限り, 欠測値が多い問題を避けることができない. また, 以下のようなデータベース自体の問題が挙げられる:

- (1) 感受性宿主に関する情報が得られない

- (2) 時系列データを格納するのに有用でない
- (3) 定義の曖昧さを明示的に測る尺度がない
- (4) 隔離・検疫など個体レベルで実施された感染対策が定量し難い
- (5) オブジェクト化することが難しい

このデータベースは「感染個体」のみに関する情報が収録されており、残念ながら感受性宿主に関する情報が得られない。歴史的に、数少ない血清疫学調査を除いて、感染しなかった者（あるいは発症しなかった者）が観察された研究は極めて稀である。しかし、感染症の流行を理論的に理解する上で、感受性宿主の行動やワクチン接種状況・集団免疫レベルなどに関する情報を得ることは不可欠であり、それらが欠如している場合は感受性宿主の動態を簡略化したモデルを新たに定義する必要性が生じるだけでなく、抽出できる情報量が極端に減ってしまう。

例えば、流行曲線を利用すれば各時刻の実効再生産数（1人当たりの感染者が時刻 t で生み出した2次感染者数の平均値、 $R(t)$ ）を推定することが可能である。実効再生産数は(3.6)の再生方程式より

$$(4.1) \quad R(t) = \int_0^{\infty} A(t, \sigma) d\sigma$$

のように数理的に定義される。しかし(3.7)や(3.8)に従って(4.1)を書き換えると

$$(4.2) \quad R(t) = H_1(t)S(t) \int_0^{\infty} H_2(\sigma)\beta(\sigma)\Gamma(\sigma) d\sigma$$

であり、 $R(t)$ の時刻に対する変化が「感受性宿主数 $S(t)$ の変化」(例. 接触を回避する行動)によるものなのか、公衆衛生対策の影響によるのか(例. 流行途中のワクチン接種など $H_1(t)$ への影響)等、分離して明らかにすることができない。

また、2番目の時系列データの格納問題は、感染個体に関する各イベントを基礎にデータベースが構築されていることに直接の原因がある。時系列データの多くは発症時刻(あるいは診断・報告時刻や死亡時刻)のみを与えるため、データベース上で言えば「感染・発病」変数の属性である発熱日あるいは発疹出現日を除いた他の情報が一切得られない(他は全て欠測値として扱う)。3点目として、それぞれの記載の定義に関する問題が挙げられる。今日の医学文献と異なり、多くの史料には診断基準や各疫学的観察の定義が与えられていない。例えば、文献内で「発症日」と書かれている場合に、それが発熱日を指すのか、あるいは発疹出現日を意味するのか、に関して区別できないことがある。同様の様々な疫学的・臨床医学的な定義が曖昧であることは稀ではない。4番目の問題として、隔離・検疫など個体を対象に感染対策が行なわれた場合、その「程度」を測る尺度を与えることが非常に難しい。例えば、隔離を開始した時刻が仮に与えられているデータであっても、その隔離がどの程度厳格なものであったのかは必ずしも明らかでない。5番目として、データベースが扱う文献の種類が多岐に渡るため、データベースのオブジェクト化が容易でないことが挙げられる。明示的なオブジェクト化を実施せずに、統計疫学モデルという目的・ニーズを明確にした上で個体別の情報を蓄積したデータベースを設計したが、それを更に多目的に発展させるアイデアに関しては未だ開発の余地がある。

また、天然痘の歴史統計を利用する上で、その2次資料が有する疫学的問題に注意を払うことが重要である。例えば、統計疫学モデルを動機にデータを抽出する場合、以下の点を認識しなければならない:

- (1) 統一の診断基準がない
- (2) Time event の定義
- (3) ワクチンの種類と起源

(4) 毒力(virulence)および病原性(pathogenicity)

1つ目は流行の背景の1つとして、どのように天然痘の診断が行なわれたか、に関する問題を指す。例えば、天然痘根絶計画の開始以前に、ウイルス学的に天然痘の確定診断が行なわれた症例は極めて少ない。多くの診断は発疹の特徴や接触歴の調査結果などに基づく臨床的な根拠に基づいて下されていたことに気付かなくてはならない。これは歴史統計を扱う上で最も重要な問題の1つであり、例えば19世紀後半までに天然痘と診断された者に水痘の症例が含まれていたことが過去に議論されている(Simon, 1887)。また、臨床的な重症度の診断基準に関しても同様の問題を認め、現に重症度の分類は時代と共に何度も訂正・改善されてきたことを詳細に理解しておかなければならないだろう(Ricketts and Byles, 1908)。2番目は上述の定義の問題に関係するが、感染・発症あるいは死亡を代表として各出来事の定義自体が曖昧であることがほとんどである。例えば、感染日が記録されている場合、それは本来は直視下で観察できない情報であるため、報告書内で「どのような方法で」感染日を特定したのかに関する説明が欠かせない。また、上述の通り、発症は発熱と発疹のどちらか、あるいは診断日や報告日を指すのか等が明確に区別されていないことが多い。同様に、例えば「回復日」が記録されていることがあるが、それは解熱日なのか発疹部位の乾燥・消失なのか、定義が不明な場合がほとんどである。3点目の問題として、ワクチンの質が時代とともに改善しており、その種類も大きく異なることが挙げられる。特に、20世紀前半から中盤にかけて天然痘ワクチンの品質が飛躍的に向上したと考えられている。ワクチンの種類によって大幅に効果が異なるかも知れないが、未だそれは疫学的に明らかにされていない。天然痘ワクチンとして利用されるワクチニアウイルスが「いつ、どこで」天然痘ワクチンとして利用されはじめたのかさえ議論の余地が残されており(Baxby, 1977)、根絶計画以前のワクチン接種効果をワクチンの種類の別で明示的に分析する手立てが今のところ見当たらない。4番目の問題は、天然痘ウイルスの毒力や病原性の違いである。天然痘は致命割合が高い *variola major* と比較的弱毒の *variola minor* の2種類に区別されることが経験的に知られている(Chapin, 1926)。そのウイルス学的な違いはこれまで不明であったが、最近になって毒力の違いの一部が遺伝情報に還元されることが明らかにされた(Esposito et al., 2006)。しかし、その他の詳細は未だ明らかでなく、これは問題点というよりも、むしろ遺伝情報を利用した統計学的分析によって今後明らかにしなければならない研究課題と言える。

これらの問題は統計疫学モデルを構築する上で共通しており、基礎的情報を抽出する上で最も重要な注意点と考えられる。文献毎に得られる情報が異なるため、常に「どの情報が揃った文献データを研究対象とし、どれを対象外とするか」について、疫学的ニーズを根拠に、除外定義を明確にしつつ、データベース内のデータ種類・量と照らし合わせて各課題に臨むことが必要とされる。また、診断基準の問題に代表されるように、報告された天然痘症例の情報が妥当であることが正当化できる範囲の文献以外を対象に含む場合、その研究の価値が著しく損なわれてしまう可能性があることにも注意しなければならない。

5. 考察

本稿では、天然痘の研究を事例として、統計疫学モデルを動機とした素材型の歴史統計データベースの構築とそれに認める問題点についてまとめた。歴史統計を基にして、生物統計学領域におけるモデルを多数応用できるデータベースを構築することは古くて新しい課題である。残念ながら、未だデータマイニング技法が応用できるほどデータ量が豊富でない。また、手作業を要するために大部分の入力データに関するクロスチェックが終了しておらず、簡便なユーザーインターフェースの設計にも至っていないことから、未だデータベースの公開に至っていない。

これは今後の重要な課題の1つである。

今回紹介したような感染症の統計疫学モデルを用いた研究では、データの内容と種類・量に応じて統計モデルおよび数理モデルの構造と想定を柔軟に制限・緩和することによって分析が行われてきた。しかし、興味の対象によっては、詳細な観察に基づく単一の流行報告だけを分析することによって、他の統計資料では明らかにすることが出来ない疑問に回答することも少なくない。時には数多くのデータソースに何らかの根拠を求めるよりも、単一の流行に関して豊富な調査結果がまとめられた記録を基に統計モデルを構築して推定作業を実施することがある。例えば、最も頑健と考えられる天然痘の基本再生産数の推定値は、ナイジェリアのアバカリキという小規模集団において観察された流行記録に基づく (Eichner and Dietz, 2003)。同報告では、接触調査結果に加えて感受性宿主(かつ流行中に発症を経験しなかった者)に関する詳細な情報を含むものであったことから、同一データを利用して何度も繰り返し天然痘の基本再生産数が推定・議論されている (Becker, 1989)。1つの報告が含むデータの質と量が極めて豊富であることが明らかであれば、必ずしも統計疫学モデルはメタデータの蓄積に頼る必要がない場合もある。

特定研究を動機にしたデータベースを構築する以上、地道な文献の収集や手作業によるデータ入力を通じたデータ蓄積の努力が欠かせない。このデータベース構築はドイツにおける共同研究プロジェクトとして開始した。当初は著者らが個人レベルで構築したデータベースを欧州研究プロジェクトにおける研究者間で共有する目的の下で発展させ、更に系統的にデータ・収録内容を整理したデータベースとしてまとめた。今後、公開までに少なくとも関連専門家間でデータベースを共有の上、インターフェースを改善しつつ収録内容を拡張しなければならない。また、同様のアイデアはインフルエンザや麻疹などの数理疫学研究を目的として応用されつつある。これらの感染症は未だ観察されることもあるのだから、新たな感染個体の観察が直接に可能な場合は、過去の文献に含まれる情報の種類や内容に妥協せずに新たなフィールド調査を行なうことによってモデルを構築・分析していくことも必要だろう。さらに、間もなく根絶が期待できるだろうポリオを代表として、同様のアイデアが更なる他の感染症についても応用・発展できるものと思われる。

最後に、「そこで何が起こっていたのか」を明らかにする手段は、観察データを分析することだけではない。特に、生の観察データを扱う限り、人類学および社会学的視点から流行の文化的及び社会的背景を理解することが欠かせない。分析に要する背景を理解する目的だけでなく、データベース構築や運営作業を含め、歴史学および文化人類学等の専門家と連携して研究を進めることが重要である。感染症の歴史統計学や歴史人口学は歴史学の研究領域のみで議論されることも少なくなく、未だ明らかにされていない統計学的・疫学および生物学的な課題が山積している。感染症の歴史統計データベースがその発展の一助になることを願いたい。

謝 辞

本報告の研究内容の一部は井上科学振興財団による研究補助(井上研究奨励金)を受けて実施された。

参 考 文 献

- Barry, F. W. (1889). *Report on an Epidemic of Small-pox at Sheffield, during 1887-88*, Local Government Board, London.
- Baxby, D. (1977). The origins of vaccinia virus, *Journal of Infectious Diseases*, **136**, 453-455.

- Becker, N. G. (1989). *Analysis of Infectious Disease Data*, Chapman and Hall, New York.
- Chapin, C. V. (1926). Changes in type of infectious disease as shown by the history of smallpox in the United States, 1895–1912, *Journal of Infectious Diseases*, **13**, 171–196.
- Chowell, G. and Nishiura, H. (2008). Quantifying the transmission potential of pandemic influenza, *Physics of Life Reviews*, **5**, 50–77.
- Codd, E. F. (1990). *The Relational Model for Database Management: Version 2*, Addison Wesley Publishing Company, Massachusetts.
- Cooper, B. (2006). Poxy models and rash decisions, *Proceedings of the National Academy of Sciences USA*, **103**, 12221–12222.
- Dietz, K. and Heesterbeek, J. A. (2002). Daniel Bernoulli's epidemiological model revisited, *Mathematical Biosciences*, **180**, 1–21.
- Downie, A. W., St Vincent, L., Meiklejohn, G., Ratnakannan, N. R., Rao, A. R., Krishnan, G. N. and Kempe, C. H. (1961). Studies on the virus content of mouth washings in the acute phase of smallpox, *Bulletin of World Health Organization*, **25**, 49–53.
- Eichner, M. (2003). Case isolation and contact tracing can prevent the spread of smallpox, *American Journal of Epidemiology*, **158**, 118–128.
- Eichner, M. and Dietz, K. (2003). Transmission potential of smallpox: Estimates based on detailed data from an outbreak, *American Journal of Epidemiology*, **158**, 110–117.
- Esposito, J. J., Sammons, S. A., Frace, A. M., Osborne, J. D., Olsen-Rasmussen, M., Zhang, M., Govil, D., Damon, I. K., Kline, R., Laker, M., Li, Y., Smith, G. L., Meyer, H., Leduc, J. W. and Wohlhueter, R. M. (2006). Genome sequence diversity and clues to the evolution of variola (smallpox) virus, *Science*, **313**, 807–812.
- Farewell, V. T., Herzberg, A. M., James, K. W., Ho, L. M. and Leung, G. M. (2005). SARS incubation and quarantine times: When is an exposed individual known to be disease free?, *Statistics in Medicine*, **24**, 3431–3435.
- Fenner, F., Henderson, D. A., Arita, I. and Ladnyi, I. D. (1988). *Smallpox and Its Eradication*, World Health Organization, Geneva.
- Ferguson, N. M., Keeling, M. J., Edmunds, W. J., Gani, R., Grenfell, B. T., Anderson, R. M. and Leach, S. (2003). Planning for smallpox outbreaks, *Nature*, **425**, 681–685.
- Fraser, C., Riley, S., Anderson, R. M. and Ferguson, N. M. (2004). Factors that make an infectious disease outbreak controllable, *Proceedings of the National Academy of Sciences USA*, **101**, 6146–6151.
- 藤野正三郎 (1978). 長期経済統計 (LTES) データベースの研究. 経済統計データ・バンクの基礎研究, 一橋大学経済研究所日本経済統計文献センター, 東京.
- Greenwood, M. (1935). *Epidemics and Crowd Diseases. An Introduction to the Study of Epidemiology*, Macmillan, New York.
- Halloran, M. E., Longini, I. M., Nizam, A. and Yang, Y. (2002). Containing bioterrorist smallpox, *Science*, **298**, 1428–1432.
- 浜野 潔 (2004). 『日本疾病史』データベース化の試み, 關西大學經濟論集, **54**, 433–444.
- 速水 融 (2006). 『日本を襲ったスペイン・インフルエンザ—人類とウイルスの第一次世界戦争』, 藤原書店, 東京.
- Henderson, D. A., Inglesby, T. V., Bartlett, J. G., Ascher, M. S., Eitzen, E., Jahrling, P. B., Hauer, J., Layton, M., McDade, J., Osterholm, M. T., O'Toole, T., Parker, G., Perl, T., Russell, P. K. and Tonat, K. (1999). Smallpox as a biological weapon: Medical and public health management. Working Group on Civilian Biodefense, *The Journal of the American Medical Association*, **281**, 2127–2137.
- 平山 勉, 友部謙一 (2004). Rekishow Authoring Tool の可能性—歴史研究のためのデータベースと

- の関連を中心に—, *MediaNet*, **11**, 34–37.
- 飯島 渉(2000). 『ベストと近代中国. 衛生の「制度化」と社会変容』, 研文出版, 東京.
- 稲葉 寿(2002). 『数理人口学』, 東京大学出版会, 東京.
- Kaplan, E. H., Craft, D. L. and Wein, L. M. (2002). Emergency response to a smallpox attack: The case for mass vaccination, *Proceedings of the National Academy of Sciences USA*, **99**, 10935–10940.
- Klinkenberg, D., Fraser, C. and Heesterbeek, H. (2006). The effectiveness of contact tracing in emerging epidemics, *PLoS ONE*, **1**, e12.
- Koopman, J. (2004). Modeling infection transmission, *Annual Review of Public Health*, **25**, 303–326.
- Kretzschmar, M., Wallinga, J., Teunis, P., Xing, S. and Mikolajczyk, R. (2006). Frequency of adverse events after vaccination with different vaccinia strains, *PLoS Medicine*, **3**, e272.
- Mack, T. M. (1972). Smallpox in Europe, 1950–1971, *Journal of Infectious Diseases*, **125**, 161–169.
- Murray, J. D. (2002). *Mathematical Biology: I. An Introduction*, 2nd ed., Springer, New York.
- Nishiura, H. (2006). Smallpox during pregnancy and maternal outcomes, *Emerging Infectious Diseases*, **12**, 1119–1121.
- Nishiura, H. (2007). Early efforts in modeling the incubation period of infectious diseases with an acute course of illness, *Emerging Themes in Epidemiology*, **4**, 2.
- Nishiura, H. (2008). Determination of the appropriate quarantine period following smallpox exposure: An objective approach using the incubation period distribution, *International Journal of Hygiene and Environmental Health* (in press, doi: 10.1016/j.ijheh.2007.10.003).
- Nishiura, H. and Eichner, M. (2006). Estimation of the duration of vaccine-induced residual protection against severe and fatal smallpox based on secondary vaccination failure, *Infection*, **34**, 241–246.
- Nishiura, H. and Eichner, M. (2007a). Predicting residual immunity against smallpox, *Journal of Infectious Diseases*, **195**, 160–161.
- Nishiura, H. and Eichner, M. (2007b). Infectiousness of smallpox relative to disease age: Estimates based on transmission network and incubation period, *Epidemiology and Infection*, **135**, 1145–1150.
- Nishiura, H. and Eichner, M. (2008). Interpreting the epidemiology of postexposure vaccination against smallpox, *International Journal of Hygiene and Environmental Health*, **211**, 219–226.
- 西浦 博, 稲葉 寿(2006). 感染症流行の予測: 感染症数理モデルにおける定量的課題, *統計数理*, **54**, 461–480.
- Nishiura, H., Schwehm, M. and Eichner, M. (2006). Still protected against smallpox? Estimation of the duration of vaccine-induced immunity against smallpox, *Epidemiology*, **17**, 576–581.
- Petitti, D. B. (2000). *Meta Analysis, Decision Analysis and Cost-effectiveness Analysis: Methods for Quantitative Synthesis in Medicine*, Oxford University Press, New York.
- Rao, A. R., Prahlad, I., Swaminathan, M. and Lakshmi, A. (1963). Pregnancy and smallpox, *Journal of Indian Medical Association*, **40**, 353–363.
- Ricketts, T. F. and Byles, J. B. (1908). *The Diagnosis of Smallpox*, Cassell, London.
- Rutten, W. (1997). *De Vreselijkste Aller Harpijen. Pokkenepidemieën en pokkenbestrijding in Nederland in de achttiende en negentiende eeuw: een social-historische en historisch-demografische studie*, Agricultural University Wageningen, Wageningen.
- Sakar, J. K., Mitra, A. C., Mukherjee, M. K., De, S. K. and Mazumdar, D. G. (1973). Virus excretion in smallpox. 1. Excretion in the throat, urine, and conjunctiva of patients, *Bulletin of World Health Organization*, **48**, 517–522.
- 瀬野裕美(2005). 『数理生物学—個体群動態の数理モデリング入門—』, 共立出版, 東京.
- Shafa, E. (1972). Case fatality ratios in smallpox, WHO Surveillance Epidemiology, No. 72–35, World

Health Organization, Geneva.

Simon, J. (1887). History and practice of vaccination, *Public Health Reports*, **1**, 167–317.

Quantifying Smallpox Transmission Using Historical Data: A Database for Statistical Modeling

Hiroshi Nishiura

Theoretical Epidemiology, University of Utrecht

Although smallpox is the only disease to have been eradicated worldwide, the threat of bioterrorism has led to debate on potential countermeasures in the event of an attack. Because of its global eradication, we have to use historical records to quantify important biological and epidemiologic characteristics in order to optimize intervention. This article reports quantitative modeling of the transmission and spread of smallpox using historical data. In particular, technical issues on database construction specifically aimed at statistical modeling are summarized. As typical examples, I briefly discuss how smallpox spreads in the absence of intervention, focusing on the optimization of quarantine and isolation and on disaster size estimation in the event of an attack. Critically important aspects in extracting key information from the database of smallpox epidemics are summarized, and I mention potential pitfalls in utilizing the historical data.