

すなわち、567サンプルが得られると推測される。

また、60歳以上の急性心筋梗塞入院患者数は、 $IHD60 \times 0.70 = 716.1$ 、すなわち、716サンプルが得られると推測される。

### (3) 実証分析結果をみるためのサンプル数の評価

60歳以上の入院患者が、IHDで1年あたり約1,000人、AMIで1年あたり約716人（推定値）というサンプル数は、個票データに基づくクロス・セクションの実証分析にとって、あるいはプールしたクロスセクション・データによる実証分析にとって、統計学的検定を行うには十分大きなサンプル数とは言えませんが、統計的実証分析に耐えうるサンプル数だと考えられる。

諸外国の論文を見ても、クロスセクション・データを用いた計量経済学的研究、統計的実証分析として発表された論文の中で、このようなサンプル数で推定結果を発表している論文は少なからず見いだすことが出来る。その意味では、「社会医療診療行為別調査」を用いた実証分析から「統計学的に有意」な結果が得られたとすれば、日米のサンプル数に相違があっても、国際比較研究の成果として公表することは可能ではないかと考えられる。例えば、両国それぞれの推定結果から、心臓カテーテルを患者が受けたかどうかというダミー変数の係数が死亡率に対しては negative に、診療点数や入院日数に対しては positive に、95%の有意性をもって得られた場合には、両国の結果の共通性について言及することが可能であると考えられる。

ただし、サンプル数が少ないため、回帰式の変数が異常値を示すものを取り除きなどの処理をした場合のサンプル数はさらに小さくなる可能性がある。また、サンプル数が小さければ小さいほど、あまり多くの説明変数を入れてしまうと、それだけ統計学的に有意な値を求めることが難しくなる傾向があるので、推計式の特定化に十分注意を払う必要がある。これらの問題を避けるためには、年齢階級を老人医療保健制度対象年齢や年金受給年齢に限定せず、日米で比較できる何らかの基準で合わせたより広い年齢層のサンプルを用いて実証分析することも、一つの有効な比較研究の方法であると考えられる。

NBERが使用可能なデータ・ベースを踏まえると、分析対象を広げて日米比較研究する方法には、次のような選択肢が考えられる。

- ① 60歳以上あるいは70歳以上の高齢者だけではなく、全年齢のサンプルを分析対象とする。この場合、NBER側は、メディケア・クレーム・データばかりではなく、HMO等から抽出された他のデータ・ソースとの比較も考慮する必要がある（NBER了解済み）。
- ② 入院・外来患者全てを分析の対象とする。NBERにはメディケアのPart B（外来患者のみ）のデータもあるので、これを活用することが出来る。
- ③ 傷病をIHDやAMIだけに限らず、NBERで現在分析が行われている、癌（肺癌、乳癌、前立腺癌、胃癌、大腸癌）、脳溢血、あるいは肝臓疾患等の実証分析結果と比較することを試みる。とりわけ、癌については、IHDやAMIよりも多いサンプル数を抽出することが出来ると思われる。

OECDにおけるTECHプロジェクトは、最終的にはAMIだけでなく、癌や脳溢血などの成人病について幅広く国際比較を行おうという意図の基に立ち上げられと伺ったことを勘案すると、「社会医療診療行為別調査」を使って、AMIのみならずAMI以外の傷病にも注目することは、TECHプロジェクトを側面から補強する実証分析を試験的に先行して行うという意義をもつと考えられる。

将来よりサンプル数の多いレセプト・データに基づく実証分析を行うファースト・ステージの実証分析を行うとしても、このように多様な拡張方法とTECHプロジェクトを何らかの形で細くすることが出来る可能性をもつ研究として、「社会医療診療行為別調査」に基づく実証分析は、重要な意義をもっていると考えられる。

## 3. 2 「社会医療診療行為別調査」を用いる際の留意点

### (1) Sample Selection Biasの問題

「社会医療診療行為別調査」の場合、それが全数調査ではなく抽出による個票データであることに加えて、サンプル数が小さいことのために、実証結果の Sample Selection Bias（以下、バイアスと呼ぶ）が、米国のデータに比較するとかなり大きいおそれがあります。例えば、PTCAやCAGB(revascularization)を行う前段階として心臓カテーテルを受けた患者の方が、心臓カテーテルを受けなかった患者よりも年齢が低く健康状態が良い可能性が高い場合、あるいは、心臓カテーテルを受けた患者の方が、Teaching HospitalやResearch Hospitalといったようないわゆる「ハイ・テク技術」設備の整った病院に入院したために、revascularization以外の治療が積極的かつ効果的に行われた可能性が高いという場合には、心臓カテーテル自体が患者成績、医療費、QOLなどに対して与える効果を過大評

価(overestimate)してしまうおそれがある。

「社会医療診療行為別調査」におけるバイアスをいかに小さく修正していくかは確かに重要な課題です。この問題に対しては、McClellan博士やStaiger博士(Harvard大学)が用いている方法をこの実証分析に適用して対処することが出来ると考えられます。その方法は二つあり、一つは適切な操作変数(Instrumental Variable: IV)を見つけ出すことであり、もう一つはGMM(General Moment Method)を使う方法である。

米国の実証分析では、IV(操作変数)として、心臓カテーテルを年間4件以上使った病院(cath hospital)施設とそうでない医療施設(non-cath hospital)と患者の住所との距離の差をZipcodeから計算して、それをIVとして用いている。このような方法が採られる理由は、患者がcath hospitalの近くに住んでいるかどうかで、観察データを擬似的にランダムイズ(randomize)することが出来るからです。とうのも、もし病気が発生すれば住んでいる所でその人が患者となることは確かであるが、米国の労働移動・人口移動の実証研究の成果を踏まえると、個人が住所を定める際、cath hospitalに近いからそこに住むという理由を最も重視するとは考えにくいからである。

「社会医療診療行為別調査」では病院と患者の住所との実際の距離を推計することは困難ですが、操作変数法を適用することは出来ると考えられる。具体的には、「社会医療診療行為別調査」の使用申請で、患者一人一人の市町村コードの使用を許可してもらった上で、病床数の規模別データ等からその地域の「ハイ・テク」病院密度の推定値を求めることが出来れば、そうしたインデックスを操作変数(IV)として活用することが出来ると考えられる。

## (2) データの抽出期間が短期間で、抽出される医療施設が毎年変わる問題

データの抽出期間が毎年1ヶ月という短期間に限定され、抽出される月も変わる場合がある(ほとんどの場合は6-7月に集中しているようですが)ことに加えて、データ収集のために抽出される医療施設が毎年異なる問題は、米国のメディケア・クレーム・データと比較して、非常に実証分析の方法を難しくさせる問題だと言わざるを得ない。したがって、「社会医療診療行為別調査」を用いる場合には、出来る限りこの問題に慎重に対処するように努めなければならない。

データの抽出期間が毎年1ヶ月という短期間に限定され、抽出される月も変わるため、  
「社会医療診療行為別調査」個票データは、生存分析の観点からみてLeft-censoredとRight-censored両方の問題点をもっていることを意味する。そのうち、(回復の途中経過は分かるがそれよりも先が分からない、つまり未来が分からない) Right-censoredとみなされる患者については、サンプル数が更に縮小することを覚悟して、その月内に入院した患者のみをピック・アップして、このサンプルに対して生存分析を行うことが望ましいと考えられる。もちろん、日米比較に当たっては、まず、ハザード・レートの基になる確率分布についていくつかの分布を用いて複数の推計式の特定化を行い、日本側の複数の推定結果を求めて最も有意な結果を導く推定式を選択する必要がある。そして、このように特定化した推定式を米国データに適用した結果と日本側の結果を比較する必要がある。

Left-censoredの患者を含むデータ・セットで実証分析した結果の統計学的な有効性についての計量経済学的・統計学的理論付けはまだ十分には出来ていないのが現状である。この問題については、NBER側の研究補助金申請にもご助力下さっている計量経済学者の雨宮健先生に最近の理論的な展開とその適用可能性について後指導戴くことが望ましい。こうした根本的な問題の処理方法に加えて、既存の計量経済学・統計理論の範囲内で対処することも必要かと思われるので、次にそのひとつの方法を例示したい。

「社会医療診療行為別調査」の調査期間最終日までの入院日数を用いて、入院期間のインデックスを作ることが出来ます。例えば、入院日数のインデックスが5週間であれば、その患者はサンプリングの月の初めで入院1週間目であることが分かる。ただし、こうしたleft-censoredの患者についての問題は、入院期間(あるいは、調査月内に退院もしくは院内死亡が分かれば、outcomeもわかる)が分かるとしても、調査期間以前の治療内容が不明だということである。AMIに対する治療のゴールデン・タイムは通常病院到着後24-28時間以内と言われている。あるいは、一般的に"invasive"あるいは"aggressive"なケアと呼ばれるのは、病院到着後7日以内にその後のptcaやcabg等に引き続くであろう心臓カテーテルが行われるかどうか、ポイントになると言われている。「社会医療診療行為別調査」を用いた場合、抽出された月の1日に入院した患者は少なくとも30日という観察期間があるのに対して、その月の最後の日に入院した患者は1日しか観察することが出来ない。もしその月の最後の日に入院して翌日心臓カテーテルを行った患者がいたとすると、その患者に対していくらゴールデン・タイムにそのような治療が行われたとしても、観察できないという問題は避けられない。

したがって、米国のメディケア・クレーム・データと比較する場合には、入院後非常に短期間での患者の治療成績に注目した研究、例えば、1-day mortality、あるいはAMIゴールデンタイムに関する研究をテーマとし、このような研究を日本のデータのみならず米国データについても行い、その範囲

内で両者を比較する必要があると考えられる。(米国のデータの法がサンプル数をはるかに多いので、このような目的のためのサブサンプルを作ることは可能です。)

あるいは、個々の患者について治療のLikelihoodを推定し期待値を求めて(1st Stage)、実際の治療の有無ではなく、「治療」の期待値を医療の質を評価する回帰式の変数として用いる(2nd Stage)ことも考えられる。

最後に、抽出される医療施設が毎年異なることを考慮して、米国のメディケア・クレイム・データから構成したLongitudinalデータに基づく実証分析と比較できる結果を出すための推定方法について考察します。医療施設がもし変わらなければ、「社会医療診療行為別調査」の個票を、性別、住所(市町村コードなど)、世帯属性、傷病、など出来る限り細かい属性で分類し、これを生年別にソートして、ある特定の医療施設の患者となる人々の擬似パネルデータを作ることが出来ます。しかし、医療施設が毎年変わることから、Longitudinalデータに類似した擬似パネルデータも作れない問題を抱えていることを意味します。そこで、実証分析の枠組みから医療施設の枠組みをはずして、治療形態の推移(例えば、血栓溶解剤や $\beta$ 遮断薬等の薬物治療からptcaやcabg等の工事医療による治療への推移など)が患者成績に与える影響を検討することが、可能であると考えられる。

この場合もやはり、「社会医療診療行為別調査」に含まれる患者の性別、年齢、副疾患の有無、そしてどういった検査がなされたかどうか(検査結果ではありません)などの患者個人に関わる情報を出来る限り利用して、死亡率の期待値を推定し、実際に死亡したかどうかではなく、死亡のLikelihoodを回帰式の変数として用いることが考えられます。これは一種の「Weighted Regression」のやり方ですが、このような方法を適用することが、ここで指摘された諸問題に対処するために必要だと考えられる。

### 3. 3 医療の質を測定するための方法論の確立

治療効果を測定する上で、統計学上最も大きな問題が、バイアスの問題です。バイアスを避けるためのgold-standard methodとしてランダム化比較対照試験(clinical randomized trial)があるが、これはコストがかかり過ぎる上に、倫理的に非常に問題があり、患者・医師間でのインフォームド・コンセントが米国のように法制化されていないわが国においては、リスクの高過ぎる方法であると思われる。しかも、ランダム化比較対照試験は、しばしば「ランダム」ではないことが多いのが実情です。例えば、対象となった治療法の費用が高過ぎるため患者やその家族が比較対照しようとする治療を拒否したり、得てして重症者や高齢者等のリスクの高い人々を対象からはずすことがあるために、その結果が必ずしもすべての患者グループに対する結果の統計的な代表性をもっているとは必ずしも言えない場合がある。

このような理由から、治療法の有効性を検討するために、医療科学の世界ではランダム化比較対照試験を補完するものとして、観察データ(observational data)が頻繁に使われます。しかし、観察データを用いる治療法の硬化を実証分析する際の最も大きな問題が、比較対照実験でも問題になるバイアスの問題です。したがって、2節と3節で考察したように、患者の特性によるコントロールや統計学的方法によって、いかにバイアスの小さい結果を求めるかが実証分析での重要な課題になる。そのためには、バイアスがなくなるようにデータ・ソースを大規模化すること、例えば患者個人についてのカルテ・ベースでの詳細情報を収集することも一つの方法として考えられる。しかし、この方法は高いコストがかかります。米国において急性心筋梗塞患者に対する治療の質の向上を目指して立ち上げられたCooperative Cardiovascular Project(CCP)によって、1994年から1995年にかけてカルテ・ベースの情報が、全メディケア患者について収拾されましたが、患者一人につき\$100から\$150ドルの費用がかかっている。

そこで、現在、NBERをはじめとして全米の研究機関では、最小の情報(例えば、患者の性別、年齢、住所など)で観察対象の属性をコントロールしながら、できるだけ「公平」かつ「正確」な医療の質の効果に関する推定結果が得られるような統計的方法論・計量経済学的手法に関する研究が進められている。2節と3節で触れた推定上の留意点を含めた実証研究を実施することができれば、この研究は、主としてMark McClellan(Stanford University)とDouglas Staiger(Harvard University)を中心としたNBERの医療経済学研究グループで開発中の方法論と推定手法を、日本の個票データに適用する最初の試みになる。この方法論・推定方法の応用が成功すれば、日本においても、医療の質の測定をデータのバイアス問題に対処しながら行うことが可能になり、治療効果を高めるための可能な限り客観的な情報を、患者や医療サービス提供者および医療政策関係者に提供することが出来るようになると考えられる。そのためにも、第2節で検討した医療の質の経済的評価に適したデータ・ベースの構築が不可欠であると考えられる。

#### 4. 考察

治療法の有効性を検証する実証分析において留意しなければならない問題は、サンプル・セレクション・バイアスである。この問題に対処する方法には、バイアスがなくなるようにデータ・ベースを大規模化すること、及びサンプル・セレクション・バイアスを除去する推定方法の開発と計測がある。アメリカは、これら両方の分野で最先端の研究を実施している。前者の例には、急性心筋梗塞患者に対する治療の質の向上を目指して立ち上げられた Cooperative Cardiovascular Project (CCP) が、1994年から1995年にかけてカルテ・ベースの情報を全メディケア患者について収拾し、データ・ベースを作成した。後者に例には、Stanford大学とNBERの医療経済学研究グループが開発した操作変数法 (IV) やGMMに基づく推定方法を用いた急性心筋梗塞の治療効果に関する実証分析がある。これらの新しい推定方法を用いた医療の質の経済的評価の実証分析は、OECDによる医療の国際比較研究にも採用され、Stanford大学のMark McClellan、Harvard大学のDavid CutlerはOECDの医療経済研究班のアドバイザーを勤めている。また、高齢者の受診行動や健康状態は就業・引退行動と同時決定である側面があり、このような同時決定構造を持つ個人の行動を推定するための方法論は、Harvard大学のDavid Wiseが多くの研究業績を著している。

本研究では、世界水準の研究を達成するための評価システム作りに努めるために、医療の質の経済的評価のデータ・ベース開発と方法論について最も進んでいるアメリカのStanford大学とNBERの医療経済研究グループと共同研究をすることは有意義なことであると考えられる。

要点をまとめれば、わが国では、アメリカのメディケア・クレーム・データのような高齢者の患者データ・ベースが構築されていないため、国民医療費の増加要因とされる老人医療費を抑制するための効率的な方法を科学的な証拠に基づいて提示することが困難であった。また、「社会医療診療行為別調査」などの官庁統計は一時点のクロス・セクション・データであるため、医療の成果評価に用いられる生存分析の観点からみて、Left-censoredとRight-censored両方の問題点をもっている。また、レセプト縦覧データだけでは、患者特性が識別できないため医療成果が元々の患者属性によるのか医療の質によるのかが識別できない。また、医療政策における情報技術基盤の整備は、高度情報社会医療情報システムや医療機関内情報統合システムの試みがある。しかし、焦眉の課題となっている医療費の適正化に資する患者データ・ベース及び解析プログラムの開発は、アメリカに比べて相当程度遅れをとっている。

したがって、プライバシーの保護に配慮しつつ、複数月のレセプト縦覧データを接続しながら患者属性と診療行為、医療支出等の情報を含む高齢者の患者データ・ベースを構築することは、医療資源の効率的配分を実現するための情報基盤整備と医療の成果評価における技術的課題の解決にとって、欠くことのできない国際共同研究である。

# Definition of AMI Cohort

- Cohort Based on Longitudinal or Non-Longitudinal Data?
    - Tier I: Analysis based on individual hospital admissions only; no or little longitudinal patient data
    - Tier II: Analysis based on data with unique patient identifiers, so that longitudinal analysis and linkage to other datasets is possible
  - Definition of AMI Cohort
    - ICD9 410, ICD9-CM: 410 unless 5th digit = “2” (equivalent codes for ICD8, ICD10)
    - Length of stay  $\geq 3$  days unless transferred or died
    - No cases that were admitted as transfers from another hospital (such cases should be counted in the first hospital); no cases that were “elective” or “booked” admissions (probably not new AMIs)
    - No hospitalization for AMI in previous 1 year (this condition applies to “Tier II” countries with longitudinal data only)
- 

## Construction of Independent Variables

- Patient characteristics (“X” variables)
  - created for each patient
  - demographic variables: age, gender, and interactions
  - geographic variables: region of country (state, province)
  - comorbidity variables (if possible, based on secondary diagnoses at time of first AMI hospitalization): Charlson comorbidity variables
- Health system characteristics (“Z” variables)
  - same for all patients in a particular year or time period
  - variables summarizing economic and regulatory incentives influencing treatment choice
  - more details in future mailing; see Table 2 in *Health Affairs* article

# Construction of Dependent Variables

- Tier I: Primary data are from individual hospital admissions, not linked over time
    - In-hospital treatment (within 1 and 7 days of admission, and during hospitalization)
    - In-hospital length of stay and costs (if available)
    - In-hospital mortality (within 1 and 7 days, and during hospitalization)
    - Transfer to another acute hospital
  - Tier II: Primary data on individual patients are linked over time, and possibly linked to other datasets
    - Treatment (within 1, 7, 30, 90, and 365 days, and during initial hospital stay)
    - Total length of stay, and if possible, length of stay divided into days in regular unit and days in intensive care unit (ICU) or coronary care unit (CCU) (within 30, 90, and 365 days, and during initial hospital stay)
    - Mortality (within 1, 7, 30, 90, and 365 days, and during initial hospital stay)
    - Readmission with cardiac complications (31-60 days, 61-365 days)
    - Costs (if available; within 30, 90, and 365 days)
- 

## Treatment Variables

- “Required” Treatment Variables: Must be created by all research teams
  - Cardiac catheterization (angiogram)
  - Coronary angioplasty (PTCA)
  - Coronary artery bypass graft (CABG) surgery
- “Optional” Treatment Variables: Will be included in analysis of some countries
  - Drug Therapies (thrombolytics, ACE inhibitors, aspirin/antiplatelet drugs, beta blockers, etc.)
  - Intracoronary stent
  - Repeat coronary angioplasty (2 or more angioplasty procedures in time interval)

## Optional Variables - Not “Required” (for Supplemental Studies)

- Will be included in supplemental studies that involve some countries, with adequate data (participation is optional)
  - More details on construction of these variables are provided below, and will be included in future mailings
  - Cost variables - additional dependent (Y) variables
  - Additional patient (X) variables
    - Socioeconomic status (more details in future mailing)
    - Insurance status
    - Type of hospital
- 

## Calculations To Be Performed by Each Country

- Descriptive Analysis
  - Summary statistics on sample size and sample characteristics (average age, % female, etc.)
  - Average values of dependent variables for particular demographic groups (for example, catheterization rate for females aged 45-64)
- Multivariate Analysis
  - Multivariate regression analysis for your country only
  - Sub-matrix calculations for your country, to contribute to formal multivariate regression analysis including all countries

# Sub-Matrix Method

- This method allows us to estimate statistical models that include data from all countries, without pooling confidential patient information from each country
- Example using multivariate regression for two countries:

$$Y_{ict} = \alpha_c + \delta_t + D_{ict} \beta + Z_{ct} \gamma + \varepsilon_{ict} = X_{ict} \phi + Z_{ct} \gamma + \varepsilon_{ict}$$

$\alpha_c$  indicator variable for country c

$\delta_t$  indicator variable for year t

$D_{ict}$  variables describing characteristics of patient i (age, gender, comorbidities, etc.)

$Z_{ct}$  variables describing characteristics of country's health care system in year t (economic and regulatory incentives)

$Y_{ict}$  dependent variable for patient: treatment, outcome, cost

## Sub-Matrix Method 2

- To simplify example further, suppose that we do not include any variables  $Z_{ct}$  so that the model is:

$$Y_{ict} = X_{ict} \phi + \varepsilon_{ict} = X \phi + \varepsilon$$

- The least-squares regression estimate for  $\phi$  is  $(X'X)^{-1} X'Y$ , which can be rewritten as

$$(X_A'X_A + X_B'X_B)^{-1} (X_A'Y_A + X_B'Y_B)$$

where A and B denote different countries; additional sub-matrices can be added to include additional countries

- Thus, the needed regression results can be calculated using sub-matrices that are based on data from individual countries; these sub-matrices include no confidential patient information

# Sub-Matrix Method 3

- We also need consistent estimates of the standard errors for the model coefficients
- To account for heteroskedasticity (errors for individual observations are not independently, identically distributed), we will use a “robust” two-step method (White, 1980, *Econometrica*)
  - Obtain preliminary coefficient estimates, as described previously
  - Use these coefficient estimates to compute residuals for each observation  $e_{ict}$  (consistent estimate of observation error,  $\varepsilon_{ict}$ )
  - Estimate weighted-least squares models, based on  $X'e$  rather than  $X$
- Can also derive estimates, standard errors for non-linear results (e.g., odds ratios) using delta method

---

## Data Validation Analysis

- Idea: Use supplementary, clinically detailed data to determine true clinical similarity of your country’s population of hospitalized AMI patients to the AMI populations of other countries
- Strongly encouraged if feasible in your country
- Key variables that ideally would be included: clinical symptoms; time to hospitalization; EKG results (ST elevation, progression); CPK results; troponin results (new); measures of AMI severity (symptoms of heart failure or shock, blood pressure, heart rate); measures of comorbidity
- Method I: Use clinically detailed data that is linked to your main hospital discharge dataset or AMI registry
  - MONICA: examples - Western Australia Linked Data Study, Denmark AMI Registry Study
  - Other clinically detailed, chart-based data: examples - US Cooperative Cardiovascular Project and National Registry of Myocardial Infarctions
- Method II: Review clinical studies of hospitalized AMI patients published in your country, and/or consult with clinical experts to obtain this information (face validity of your results)

# Role of TECH Coordinating Center

- Circulate proposed standard definitions for cases and variables, based on ICD-8,-9,and -10 codes
- Circulate specific SAS and Stata code for variable construction and construction of descriptive statistics
- Produce descriptive statistics including all countries, based on results provided from each country
- Circulate specific SAS and Stata programming code for estimating sub-matrix results in each country, and small “test” dataset so that each country can verify their ability to estimate sub-matrix models
- Combine sub-matrix from each country to obtain preliminary coefficient estimates; circulate preliminary coefficient estimates to each country with specific programming code so that each country can produce final sub-matrix results
- Integrate final submatrices to obtain model results
- Communicate with specific groups, and provide interim results for all groups, to allow quality control, review of preliminary results for possible anomalies, and validation analysis
- Maintain repository of summary data, and provide similar coordinating services for supplemental/additional studies proposed by research teams

## Authorship Policy Proposal

- TECH authorship principles:
  - (1) All members of all research teams who contribute to a paper are coauthors, and should receive credit for their efforts
  - (2) Authorship policy should provide opportunities for research team members who wish to write TECH papers to do so: enthusiasm and new ideas should be encouraged and rewarded
  - (3) Authorship policy should not create barriers to TECH productivity
- Proposed Policy for “Required” Studies
  - All active members of all research teams (as determined by each team) are coauthors for the required study
  - If number of authors listed on author line is limited by journal, all contributing members of all research teams will be listed as authors in footnote. Priority for inclusion on author line is:
    - (1) All authors who contributed to writing manuscript, and
    - (2) One “responsible” author from each country, designated by their research team.
    - (3) Additional “responsible” authors from each country, designated by their research team

# Authorship Policy Proposal 2

- Supplemental studies
  - No restrictions on individual teams using own data in supplemental studies (please include TECH as keyword, and let us know); authors are individual team members
  - Individual research teams, or groups of teams, may propose specific supplemental studies involving >1 country
  - Participation in supplemental studies is optional, and participants must provide required data on schedule; in general, except for regional studies (e.g., North America, Scandinavia), all countries should be given opportunity to participate
  - Reasonable time limit (2 years?) to complete study; and reasonable limit on number of studies being led by a particular group at the same time (to provide opportunity for all teams to lead supplemental studies if they wish)
  - If data from a country are used in a supplemental study, the country's research team must be asked to coauthor
  - Investigator(s) that coordinate the analysis and write the paper can be lead authors
  - TECH Coordinating Center provides usual support (circulation of memos on research plan, integration of sub-matrix results)
- Do we need an Authorship Committee to coordinate this policy?

---

## Proposed Supplemental Studies

- Effects of Socioeconomic Status
  - Will use either individual data on SES (work status, education, income, etc.) or postal/zip code data on SES linked to individual data (absolute income levels, relative income levels of patient's neighborhood)
  - These variables will be added to individual patient (X) variables to allow analysis of their effects on treatments and outcomes
  - Working group will review data available for each country
- National Procedure Rate Trends
  - Will use either nationally-representative hospital discharge data or data from procedure registries in participating countries
  - Will include repeat angioplasty procedures if possible
  - Working group will review whether all countries can provide discharge-based data (which allows direct calculation of changes in population treatment rates by age groups, and calculation of repeat procedures)
  - Will investigate whether regulatory and economic (Z) variables influence trends in procedure rates

## Proposed Supplemental Studies 2

- Cost Analysis
    - Will develop comparable measures of costs, resource use, and/or reimbursement
    - Possible sources of cost and reimbursement data: (1) data linked to individual patient data, or (2) unlinked data matched to averages for AMI patient groups (e.g., average nursing cost per regular hospital day would be multiplied by average length of stay to obtain estimate of nursing costs for AMI patients)
    - These will be additional dependent (Y) variables for analysis
    - Working group will review cost data available for participating countries, prioritizing most important areas for comparison, and specific methods for defining cost variables
  - Validation Analysis
    - Described on previous slide
    - Results from all participating countries would be compared, to assess potential differences in AMI patient populations
- 

## Proposed Supplemental Studies 3

- Additional disease studies using TECH methods and network
  - Stroke, and population rates of carotid endarterectomy
  - Breast cancer, possibly other cancers
  - Other forms of coronary heart disease (e.g., unstable angina and other acute coronary syndromes)

# Supplemental Regional Research Projects

- **Scandinavian Supplemental Studies**
  - Analysis of trends in AMI diagnosis, treatment, and outcomes, focusing on Scandinavian countries (Denmark, Sweden, Finland)
- **US-Canadian Supplemental Studies**
  - Data Validation: Comparison of patient and hospital characteristics and their effects on outcomes in the US and Canadian provinces
  - Methods Validation: Comparison of model results using linear methods (sub-matrix approach) versus non-linear methods (logistic analysis)
  - Relationship of AMI mortality trends in hospitalized patients to “macro” mortality trends in US and Canadian provinces
  - Comparison of incremental effects of intensive treatment in the US and Canada using instrumental variables methods
  - Comparison of trends in pharmaceutical use in post-acute AMI treatment in the US and Canadian provinces
  - Preliminary studies of applying TECH methods to other illnesses (stroke, breast cancer)

---

## Funding Opportunities

- **Validation Studies:** “audits” of diagnosis reporting in countries using diagnosis-related groups (DRGs) and other payment (government provides ongoing funding for this in US, other countries)
- **Country sources:** relate required or supplemental TECH research to your ongoing research interests and/or national funding priorities
- **Regional sources:** e.g., European Union, Framework 5 (formerly BIOMED), NATO
- **Other sources:** US funding of small grants for comparative studies, other international foundations
- **TECH Coordinating Center** can provide background materials for grant proposals

## Planning for Next TECH Global Meeting

- Aiming for conference in early September 2000
- Conference would include presentation of preliminary results from “required” analysis based on sub-matrix method, and some supplemental studies

---

## Proposed Timeline for “Required” Analysis

- October: Coordinating Center circulates details and programming language on case definition, sample description, variable construction, descriptive statistics, and preliminary sub-matrix calculations
- December 15 or earlier: Preliminary statistics on AMI patient samples (sample sizes by year, any anomalies), description of planned validation studies (optional but strongly encouraged)
- February 28: Descriptive statistics on required variables
- March 30: Each country reports multivariate regression results based on its own data

## Proposed Timeline for “Required” Analysis 2

- April 30: Each country reports initial sub-matrix results
  - May 31: Coordinating center circulates coefficient estimates for use in calculating final sub-matrices
  - June 30: Final sub-matrices reported
  - July 15: Coordinating center circulates preliminary results
  - Early September: Second TECH Global Conference at Stanford -- review and discussion of preliminary results and publication strategy, discussion of data validity studies, discussion of other supplemental studies, plans for further comparative analyses
  - Additional dates for conference calls, regional meetings, etc. will be added as needed
-

## 共同研究 3

「所得分配に関する国際比較研究」

(平成 11～13 年度)

## 所得分配に関する国際比較研究

平成11年度中間報告

阿部 彩 大石亜希子

本研究は、厚生科学研究・政策科学推進研究（指定研究）「社会保障の改革動向に関する国際共同研究」の一環である「所得分配に関する国際比較研究」として行われ、国民生活基礎調査と所得再分配調査を使用した研究を行う。同プロジェクトの目的は (1)先進諸国の社会保障改革の動向をフォローする(2)日本の所得分布の現状と動向を国際比較を交えて分析する(3)現役低所得層内の所得再分配を分析し、低所得者層を推計する、ことである。

平成11年度は、関連文献をレビューし、研究の分析手法を検討した。その結果、以下の6つのテーマと手法を研究の方向性として確定した。うち、テーマ5については、予備的分析を行い、その結果は「季刊社会保障研究」(148号)に掲載される予定である（別添1）。

### 分析テーマ1：生涯所得不平等度指標の計測、税や社会保障が生涯所得に及ぼす再分配効果の計測

#### <主旨>

日本のジニ係数は1980年代以降、上昇傾向にある。これについては、日本はすでに欧米諸国を上回る不平等社会になったという主張がある一方で、ジニ係数の上昇は主に高齢化や世帯構成の変化によるものだという指摘もでており、経済格差論争が続いている。

しかしながら、ジニ係数をはじめとする従来型の不平等度指標は横断面での所得分配状況を示しているにすぎず、実際に人々の間で不平等化が進んでいるかどうかを経済学的に検証するには、人々の「生涯所得（あるいはその近似としての消費水準）」の分散を生まれ年（コホート）別に比較するほうが望ましいといわれている。すなわち、生まれ年が同じ世代について、高齢になるほど格差が拡大していくのかどうか、また、若い世代ほどライフサイクルでみて不平等化が進んでいるかどうかを検証する必要がある。さらに、不平等度の変化のうち、人口構成の変化や世代内格差、世代間格差がどの程度寄与しているのか、個々の社会保障制度は世代内格差の縮小に貢献しているのか、あるいは世代間格差の縮小に貢献しているのかといった分析も、政策効果を計測する上で必要である。こうした研究は米国のDeaton教授、Paxson教授らの論文が先駆けとなって現在、米国や英国、台湾など各国で進行中である。

本プロジェクトでは、豊富な世帯情報・個人情報を含むマイクロデータを使用することにより、属性の差や所得概念の差を調整した上で、日本の所得不平等化が実際に進んでいるのか、いないのか、また、国際的にみてどの程度の不平等度に位置づけられるのか明らかにする。

主な作業内容としては、ライフサイクルでみた所得分散・消費分散を計測し、税や社会保障

の生涯所得に及ぼす再分配効果を国際比較を交えて検討する。

◆ 所得分散・消費分散の計測

複数年次のマイクロデータを使用して生まれ年（コーホート）別に生涯所得・生涯可処分所得・生涯消費を計算し、その分散をライフサイクルおよび世代間で比較する。また、個々の税や社会保障が世代内および世代間での再分配にどのように寄与しているかを検証する。

対象とする所得の範囲：世帯別および個人別の以下の項目（拠出金に関しては所得から控除するために使用）

- ・雇用者所得（基礎・所得票 9）
- ・事業所得（基礎・所得票 10）
- ・農耕・畜産所得（基礎・所得票 11）
- ・家内労働所得（基礎・所得票 12）
- ・家賃・地代（基礎・所得票 14）
- ・利子・配当金（基礎・所得票 15）
- ・仕送り（基礎・所得票 17）
- ・その他の所得（基礎・所得票 18）
- ・内容別のその他の所得（再分配・調査票(1)）
- ・公的年金・恩給（基礎・所得票 13）
- ・制度別の公的年金・恩給（再分配・調査票(1)）
- ・その他の社会保障給付金（基礎・所得票 16）
- ・制度別のその他の社会保障給付金（再分配・調査票(1)）
- ・現物給付・措置費（再分配・調査票(1)）
- ・所得税課税の有無（基礎・所得票 19）
- ・所得税額（基礎・所得票 20）（再分配・調査票(1)）
- ・住民税課税の有無（基礎・所得票 21）
- ・住民税額（基礎・所得票 22）（再分配・調査票(1)）
- ・固定資産税課税の有無（基礎・所得票 25）
- ・固定資産税額（基礎・所得票 26）（再分配・調査票(1)）
- ・自動車税など（再分配・調査票(1)）
- ・社会保険料支払いの有無（基礎・所得票 23）
- ・社会保険料支払額（基礎・所得票 24）
- ・制度別の社会保険料拠出金（再分配・調査票(1)）
- ・内容別のその他の拠出金（再分配・調査票(1)）

対象とする消費：世帯別の以下の項目

- ・ 5 月中の家計支出額（基礎・世帯票 7）
- ・ 5 月中の家計支出額のうち、耐久消費財支出・冠婚葬祭費の有無（基礎・世帯票 8）
- ・ 5 月中の家計支出額のうち、耐久消費財購入額（基礎・世帯票 9）
- ・ 5 月中の家計支出額のうち、冠婚葬祭費（基礎・世帯票 10）

コーホート作成およびグループ分けに必要な属性項目：

- ・ 生活保護の状況（基礎・世帯票 5）
- ・ 単独世帯の区分（基礎・世帯票 6）
- ・ 作付可能な耕地面積（基礎・世帯票 11）
- ・ 最多所得者の世帯員番号（基礎・世帯票 12）
- ・ 世帯主との続柄（基礎・世帯票 14）
- ・ 性別（基礎・世帯票 15）
- ・ 生年月（基礎・世帯票 16）
- ・ 配偶者の有無（基礎・世帯票 17）
- ・ 医療保険の加入状況（基礎・世帯票 18）
- ・ 傷病の状況（基礎・世帯票 19）
- ・ 所得を伴う仕事の有無（基礎・世帯票 20）
- ・ 勤めか自営かの別（基礎・世帯票 21）
- ・ 年金の加入状況（基礎・世帯票 22）
- ・ 公的年金・恩給の受給状況（基礎・世帯票 23）
- ・ 生活意識の状況（基礎・所得票 28）
- ・ 世帯主の税法上の扶養親族数（基礎・所得票 27）
- ・ 世帯主情報（年齢、性別、配偶関係、配偶者の年齢、勤めか自営か、医療保険の加入状況、傷病状況、年金受給の有無）
- ・ 高齢者情報（60 歳以上の有無、人数、構成、60 歳以上のみ、65 歳以上の有無、人数、65 歳以上のみ）
- ・ 児童等情報（18 歳未満未婚者数、20 歳未満未婚者数、6 歳未満の児童の有無）
- ・ 年金の状況（加入者数、受給者数）

## 分析テーマ 2：世代間の受益と負担の分析（世代会計）

### <主旨>

世代会計の目的は長期的な財政収支を世代毎の負担と受益に焦点を当てて定量的に分析し、財政の持続可能性と世代間均衡を回復するための政策の選択肢を検討することである。世代会計では政府の諸制度（税制、社会保障制度、教育など）ごとに、税金や保険料のかたちでそれぞれの年齢の平均的男女がどれだけの金額を負担し、また現金給付やサービスのかたちでどれ

だけの額を受益しているかを計測する。このようにして求めたコーホート別の対政府収支をもとに、世代間の受益と負担の不均衡度合を検討することができる。

これまでのわが国の世代会計は、総務庁の「家計調査」などに基づいて作成されていたため、個人別の正確な税金・社会保険料負担が明らかでなく、また、年齢区分も粗いなど、諸外国の世代会計と比較して不十分なものであった。本研究では「所得再分配調査」のクロス再集計から得られる正確な世帯および個人情報をもとに、生まれ年（コーホート）別に世代会計を作成する。

#### ◆クロス再集計

世代会計の準備作業として、分析テーマ1の作業における「属性項目」を表頭項目とし、各種所得や拠出金を表側項目とするクロス再集計を行い、それらをコーホート別に再集計して対政府収支表（世代会計）を作成する。集計様式は別添4に示した。

### 分析テーマ3：年金資産と引退行動の分析

#### <主旨>

諸外国での研究でも明らかなように、高齢期の引退・就業の意思決定に際しては、引退を1年先延ばしすることによって、生涯年金資産がどれだけ変化するかが重要な要因となっている。しかしながら、日本においては個々人の年金資産の推計に必要な情報が少なかったため、これまでそうした実証分析は進んでいなかった。ここでは「所得再分配調査」から得られる個人および世帯ベースでの社会保険料や年金受給額に関する豊富な情報をもとに個々人および夫婦の年金資産額を計算し、その分配の現状をみることに加え、在職高齢年金制度が高齢者の労働供給に及ぼす影響を複数時点で比較したり、高齢期の所得分配に及ぼす影響を比較する。具体的な作業手順は以下の通りである。

- ・ 個人属性、世帯属性別、年齢別の雇用者所得データをもとに各個人について過去および将来の賃金プロファイルを作成する。
- ・ 作成された賃金プロファイルと現実の社会保険料拠出額などから、受給可能な年金額を計算。簡易生命表の死亡率データなどから生涯年金資産額を計算する。
- ・ 同様にして配偶者についても生涯年金資産額を計算し、夫婦年金資産額を得る。
- ・ 引退を1年先延ばしした場合の年金資産額の変化分を計算し、引退・就業行動との関係を以下のような回帰式で分析する。

#### 従属変数

- ・ 勤めか自営かの別
- ・ 雇用者所得の記入の有無（記入があるのに無職の場合は、離職したとする）

#### 説明変数