

## 子どもについての行政データベースの構築<sup>\*1</sup>

別所 俊一郎<sup>\*2</sup>

野口 晴子<sup>\*3</sup>

田中 隆一<sup>\*4</sup>

牛島 光一<sup>\*5</sup>

川村 顕<sup>\*6</sup>

### 要 約

近年の経済学の実証分析では、行政部門が業務の一環として蓄積したデータを用いた研究も進んでいる。行政データは、行政施策の対象となる主体の悉皆データを定期的に収集したものである。業務遂行のために蓄積されたものであるから、数値の正確さ・脱落や無回答の少なさ・観測値数の多さ等の点で、従来の標本調査よりも優れている。他方で、機密性等の問題に対処するための法的規制や、煩雑な行政手続き等により、行政データの学術利用は容易ではない。本稿では、まず、行政データの学術利用の長所短所を整理し、続いて、国内外の事例を紹介する。そののちに、行政データベース構築の一例として、本特集号内の論文でも用いられている東京都足立区の児童生徒データベースの構築の経緯と構造を解説する。足立区の児童生徒データベース構築には、筆者ら研究者が大きく関与している。このような事例は国内ではそれほど多くないと思われるから、その経緯を詳らかにしておくことには意義があろう。

キーワード：行政データ，足立区

JEL classification：C81, I20, H20

- 
- \* 1 本研究は文部科学省科学研究費補助金基盤研究（B）（一般）「子どもの人的資本の蓄積メカニズムに関する実証研究—足立区の挑戦から学ぶこと—」（課題番号：16H03636）の一環として行われたものである。文中においては敬称を省略している。本稿の作成、本稿で解説しているデータの作成にあたっては、足立区役所・足立区教育委員会の皆様からさまざまな貴重な助言・支援を得た。通常の留意を持って深く感謝したい。なお、本稿の内容や意見は全て筆者らの個人的な見解であり、足立区役所、足立区教育委員会、財務省、財務総合政策研究所の見解を示すものではない。
  - \* 2 東京大学大学院経済学研究科准教授／前財務省財務総合政策研究所総務研究部総括主任研究官
  - \* 3 早稲田大学政治経済学術院教授
  - \* 4 東京大学社会科学研究所教授
  - \* 5 筑波大学システム情報系社会工学域助教
  - \* 6 早稲田大学政治経済学術院准教授

## I. はじめに

国内外において、効果的・効率的な政策の実施、国民や有権者に対する説明責任の強化の観点から、客観的な根拠に基づいて政策形成を行う必要性が指摘されており、教育政策もその例外ではない。客観的な根拠に基づくためには、その基礎となるデータの整備が不可欠である。経済学分野においては、かつては分析の対象となるデータは公刊された官庁統計の報告書に掲載された集計値であった。その後、研究者自身が、あるいは企業等が行う標本調査(アンケート調査)の個票が分析されるようになり、統計法改正を機として利用できる官庁統計の調査票情報も拡大してきた。さらに近年では、行政部門が業務の一環として収集あるいは蓄積したデータ、すなわち行政データ(administrative data)<sup>1)</sup>を用いた研究の蓄積も進んでいる。所得税・法人税・関税等の税務データ、公的医療保険におけるレセプトデータ、学校教育における学力試験データ等はこの例である。このような研究の進展の背景には、もちろん、コンピュータの計算能力向上に代表される情報処理技術の進化、政策評価のための計量経済学の方法論の革新がある。

行政データは、中央政府の所有するデータであれば一国内、地方政府の所有するデータであれば所管地域内について、行政施策の対象となる主体の全数(悉皆)データを定期的に収集したものである。業務遂行のために蓄積されたものであるから、数値の正確さ・脱落や無回答の少なさ・観測値数の多さ等の点で、官庁統計を含めて従来の標本調査よりも優れている特徴を有している。他方で、機密性やプライバシーの問題に対処するための法的規制や、煩雑な行政

手続き、あるいはデータの散在や形式の不統一による接合の困難さ等の理由により、行政データの学術利用は容易ではない。とはいえ、国内外において、このような障壁を乗り越えて行政データを基にしたデータベースを整備し、分析・研究を進める事例が着々と増加している。

本稿の目的は、データ整備の事例として、本特集号内の論文でも用いられている東京都足立区の児童生徒データベースの構築の経緯と構造について解説を加えることである。また同時に、行政データの学術利用についての国内外の事例を紹介し、今後のデータの利活用・データベース整備のための参考資料を供することも目的としている。

足立区の児童生徒データベース構築には、筆者ら研究者が大きく関与している。研究者・学識経験者として足立区役所の幹部や担当者と共に共通理解を深めて信頼関係を徐々に構築し、データベース整備や、それを利用した分析の必要性を説き、データにアクセスする機会を得て、分析を行っている。このような事例は国内ではそれほど多くないと思われるから、その経緯を詳らかにしておくことには意義があろう。

本稿の構成は以下のとおりである。続く第Ⅱ節では、研究の際に行政データを用いることの長短をまとめるとともに、海外・国内の事例のいくつかを紹介する。第Ⅲ節では足立区でのデータベース作成の経緯について、第Ⅳ節では足立区データベースの構造について説明する。最終第Ⅴ節はまとめに充てられる。

1) 津田・岡崎(2018)では「行政記録情報」という用語を当てている。なお、administrative dataという単語は民間企業が事業を通じて収集したデータを含めて用いられることがある。Einav and Levin(2014)は民間企業の業務データの活用についても述べている。

## II. 行政データの学術利用

行政データ (administrative data) の学術・研究目的の利用は社会科学分野においても進展している。たとえば, Einav and Levin (2014) は経済学の最有力誌のひとつ American Economic Review における非公開データを用いた論文の割合の推移を示しており, 公的部門の非公開データを用いた論文の割合は 2006 年には 4% だったものが, 2014 年には 26% となっている。本節では, 一般には非公開とされている行政データの研究利用の事例について概観する。なお, 本節の目的は事例の紹介にあり, すべての事例を網羅する意図はなく, また筆者たちにはそのような能力もないことを注記しておく。

### II-1. 行政データ利用の長所と短所

事例の紹介に先立ち, 行政データを学術目的に用いることの長所と短所についてまとめる (Figlio et al. 2015, 2017, Friedman 2015, Einav and Levin 2014a, b, Jarmin and O'Hara 2016)。ここでは, これまで用いられてきた標本調査の個票利用との対比を念頭に置く。

統計を作成するために政府や民間部門が行っている標本調査やサーベイ調査・アンケート調査と比べたときの行政データの特徴は, 基本的には全数 (悉皆) データであり, 行政の必要性から定期的に収集されたデータであるということである。このような特徴は, 学術利用においては (例外もあるが) 以下のような長所となる。

第 1 に, 観測値数が大きいことが多く, 計測の質が高いことから, 統計的推測における正確性が高まる点である。たとえば日本の大規模な調査である「全国消費実態調査」においては, 基準抽出率は 1/791.4 (2014 年調査) である。観測値数が大きいことから, より正確なノンパラメトリックな推定が可能となる。また, 近年

の標本調査で問題になりやすい脱落や無回答, 過大回答・過少回答の問題も, 行政データにおいてはより小さいと考えられる。特に, 社会調査ではシステマティックに脱落する傾向の高い「問題の多い」標本こそが分析上重要な場合には, 行政データの活用が必要不可欠となる。

第 2 に, 全数調査で観測値数が大きいことから, 一部分に着目した分析や, 効果の異質性に注目した統計的分析が可能である。Piketty and Saez (2003) などで行われた高所得者の所得シェアの分析はその典型であるし, 双子や三つ子, 障がい者, 犯罪者等についての分析もこの例であろう。また, 細かい単位での集計値の差の分析も可能となる。学力・医療費・生産性等の平均値や散らばりの地域差の分析が例として挙げられる。

第 3 に, 定期的に収集されていることから, 自然実験・準実験となりうる事象が発生したときにも前向きデータの構築が可能であり, 処置群が少ない時にも統計的解析に耐える標本規模を確保することが期待できる。逆に言えば, 自然実験となりうる事象が起きることは稀だから, 標本調査では「事前に」処置群を調査しておくことは容易ではない。

ここまで述べてきた特長は, 何らかの行政データをそれ自体のみで用いたときにも利用できる点だが, 他のデータとの接合が可能であればさらに豊かな研究が期待できる。全数であることから, 個人識別情報があればすべてのデータを基本的には接合できるし, 一般の標本調査には含まれない変数を接合できる可能性もある。フィールド実験のデータと接続し, 行政データによる追跡ができれば長期的な効果の計測ができる。親族関係を特定できれば, 世代間を超えた分析も行える。アメリカ内国歳入庁 (IRS: Internal Revenue Service) と共同して Raj

Chetty率いる研究グループが行ってきた Equality of Opportunity project, Moving to Opportunity experiment, Opportunity Insightsはその好例と言えよう。

このような行政データの長所の裏側にはもちろん欠点もある。第1に、行政データは政策を実施し、管理するために収集されたものだから、必ずしも研究目的にそぐわないことがあるし、研究目的に対応した形で計測され、構築されているとは限らない。例えば教育関連の行政データでは、学力試験の点数のデータはあっても、性格に関するビッグファイブ(「外向性」「調和性」「誠実性」「神経症的傾向」「経験への開放性」)に関するようなデータはないことが多い。あるいは例えば、「所得」の変数はなんらかの控除後の値に加工されてしまっているかもしれない。また、行政施策の対象とならない人々は行政データには記録されない。例えば、海外移住をした人などは追跡されないだろう。

第2に、データへのアクセス可能性である。デンマーク・スウェーデン・ノルウェー・フィンランド等の北欧各国は、1960年代から社会保障制度を実施するにあたり、年収等を捕捉するために全国的な個人識別番号を導入しており、この番号を用いれば行政機関内でデータの接合が可能である。しかし、全国的な個人識別番号がない、あるいは機密性やプライバシーの問題からそのような番号の利用、あるいはデータへのアクセスに法的制限がかかっている国も多い。学術目的の利用では行政データは匿名化されるが、その利用にさまざまな手続きや費用が掛かることも往々にしてある。データの形式が統一されていなかったり電子化されていなかったりすれば、データを管理する官僚機構における人的あるいは技術的資源の制約や、煩雑な行政手続きによって利用が困難となるケースもあろう。

第3に、データの詳細についての説明文書

(メタデータ)がなかったり不十分だったり公開されていないといったように、管理状態が不十分でデータインフラが未整備なことがある。行政データは基本的にはデータ作成時点での行政施策の実施のためのものだから、変数の定義が制度改正等によって変化しても、それが文書として整理されているとは限らない。二次利用の体制が整っている標本調査では、変数の定義について詳細な説明書がついていることが多いから、この点では行政データのほうが質に問題があるといえるかもしれない。

このように、行政データには長所もあれば短所もある。したがって、特定の研究目的のためにデザインされた標本調査等とは補完的に用いられるべきだろう。次に、行政データを用いた研究事例をみてみよう。

## II-2. 海外の事例

前述したように、経済学においても行政データを用いた研究の蓄積が進んでいる(Figlio et al. 2015, 2017, Friedman 2015, Einav and Levin 2014a, b, Jarmin and O'Hara 2016, Almunia et al. 2019)。Figlio et al. (2015)は教育経済学を中心に、有力学術誌に掲載されたそのような研究の一覧表を提供している。教育関連の行政データの学術利用制度については、三菱総合研究所(2016)がイングランド・オーストラリア・アメリカ・スウェーデンについて、田中(2019)がイングランドとフィンランドについて解説している<sup>2)</sup>。これらの国々は、全国的な学力調査の個別児童の結果をデータベース化し、学校の情報等と接続したうえで、学術利用のために提供している。ここでは、最も先進的な事例の一つであるイングランドについて、これら2個の文献等に基づいて整理する。

イングランドでは、教育省(Department for Education)が児童・生徒に関する各種データを接合した全国生徒データベース(NPD:

2) その他、行政データの学術利用の仕組みについては、内山ほか(2018)、津田・岡崎(2018)を参照せよ。ChettyとSaezがアメリカ内国歳入庁のもつ税務データにアクセスした経緯についてはMervis(2014)が興味深い。また、集計等を行ってオープンデータ化する事例については豊福(2017, 2018)を参照。

National Pupil Database) を整備している。研修の受講等の必要な申請手続きを経て、一定の基準を満たせば、学校・自治体・他の政府機関等に属する外部の研究者も NPD の個票データを二次利用・分析することが可能である。

NPD は、イングランド国内の学校教育の質的改善、学校教育のアカウントビリティの確保を目的としており、1999 年から運用されている。NPD の活用は、新たなデータ収集費用・実験費用の節約となる。このデータベースには、全国テスト (National Test) と一般中等教育修了試験 (General Certification of Secondary Education : GCSE) の結果、性別・年齢・民族的属性等の児童生徒の基本属性、給食費免除・特別支援教育の必要性・居住地郵便番号等の児童生徒のその他の属性、地方当局が所管の学校から収集した規模や生徒数等の学校属性等が保管されている。家計情報は郵便番号単位で地域ごとの集計値と接合されている。

教育省は NPD に公表されるデータをもとに自治体や学校に対する財政支援の予算を決めているから、自治体や学校のデータ提供は必須であるし、データを提供する動機がある。データ収集に際しては、データを入力する学校・自治体側の負担軽減のために、政府デジタルサービス (Government Digital Service) の協力のもとでインターフェイスの統一等の再設計が進められている。

NPD の学術利用は 2000 年代前半から進められてきたが、当初は 10 名ほどの研究者グループが二次利用へ向けてのインフラ整備を行っていた。この初期メンバーのうちには、当時の教育省の統計担当者と個人的なつながりを強く持つ研究者がおり、このネットワークがインフラ整備に活用された。現在では研究者グループは 100 名以上に拡大している。NPD の研究目的での利用には申請が必要であるが、利用申請を行う前に 1~2 日の講習を受講して情報セキュリティに関するテストに合格する必要がある。また、NPD の二次利用へ向けてのインフラ整備を行った研究者グループによる NPD の

利用促進のための集中講義を大学で行うなど、NPD の研究利用促進活動も行なっている。

イギリスには Fischer Family Trust という慈善団体が運営する FFT Datalab が収集している学校や地方政府の教育データベースもある。このデータベース構築の初期においても、地方政府からのデータ提供には、地方行政関係者との個人的なつながりを持つ研究者がきっかけとなったとされている。

### II-3. 国内の事例

国内においては、本特集号の前半の論文が活用しているように、「全国学力・学習状況調査」の児童生徒個票を学術目的で利用可能な仕組みが存在している。これに加え、いくつかの地方政府が独自に教育関係の行政データを用いたデータベース整備を進めている。

とくに小中学校の教育政策に関する客観的な根拠を求めることを考えるとき、日本においては地方行政データの活用は大きな可能性を秘めている。理由は以下のとおりである。第 1 に、全国データではないために代表性に欠ける可能性があるとはいえ、行政データは基本的には特定地域の悉皆データであり、アンケートなどの標本調査に比べて回収率・回答率は極めて高く、標本選択の問題が少ない。

第 2 に、日本の地方自治体、とくに市町村は幅広い政策分野の執行機関となっており、年金を除くほとんど全ての再分配的施策を実施している (林 2008, p. 64)。後述の足立区が重視しているように、小中学生に対する教育施策は、他の政策分野、とりわけ医療や福祉政策と深く関連している。地方自治体は関連施策についてのデータを保有していることから、これらのデータを接合して多面的な分析が可能になりうる。たとえば、学力調査のデータに、就学援助等の経済状況に関連する情報や健診等の健康関連情報、就学前教育の情報等を接合できるかもしれない。

第 3 に、日本の地方自治体は首長と議会という二元代表制を採っており、制度的な権限配置

を見る限りでは首長の影響力は議会よりも強く、官僚制に対する議会の制御はかなり限定的である(曾我・待鳥2007, p. 45-48)。このもとでは、首長のリーダーシップによりデータの接合・整備が進む可能性がある。このような整備は、ともすれば縦割りにになりがちな中央省庁間ではより困難であろう。

第4に、教育・福祉政策に関連する中央政府からの規制・誘導はさまざまにあるものの、地方自治体には地方単独事業として独自施策を実施する余地がある。独自事業は全国展開していない政策の実験場とみることでもできる。また、施策の設計によっては、実験・準実験的な状況を作ることもできるだろうから、データの整備をあらかじめ組み込んでおけば制度の高い政策評価が可能となる。

三菱総合研究所(2018)では、先進的な取り組みの事例や課題について、文献調査・インタビュー調査・アンケート調査・討論会を通じた検討を行っている。ここではこの報告書やその後の進展を踏まえて、本稿で用いた足立区以外の事例について簡単にまとめておく。

### II-3-1. 埼玉県

埼玉県は、児童生徒の学力や学習に関する情報を収集して教育施策や指導を改善するため、2015年度から、さいたま市を除く県内公立小中学校に通う小学4年生から中学3年生を対象とした「埼玉県学力・学習状況調査」を実施している。対象は、2015年度で小学校708校・中学校360校であり、受検者数は小中ともに約15万人、合わせて約30万人である。この調査の特徴は、項目反応理論(IRT: Item Response Theory)に基づく調査手法を採用し、学力の経年変化を継続して把握できる調査としている点である。埼玉県の資料では、IRTを用いた自治体初の調査である。また、教科に関する調査(学力テスト)に加え、学習に対する意識や生活の様子を調査する児童生徒質問紙調査、学校・市町村教育委員会に対する質問紙調査を実施している。

この調査の個票データは、学力向上のための指導改善を進めるためとして、専門的知見を有する研究機関へ提供されている。たとえば2016年度には慶應義塾大学SFC研究所を委託先とし、中室牧子・慶應義塾大学准教授と益川弘如・静岡大学准教授が主な研究担当者として分析に当たっている。

さらに埼玉県は、この調査に関心のある自治体・企業から構成されるコンソーシアムを立ち上げている。これまでのコンソーシアムには、自治体からは、埼玉県戸田市・埼玉県羽生市・山梨県・福島県・福島県郡山市・福島県西会津町・京都府・広島県・広島県福山市・鳥根県雲南市・香川県・高知県・佐賀県の担当者が出席している。

### II-3-2. 埼玉県戸田市

埼玉県戸田市は埼玉県南部に位置し、北は政令指定都市・県庁所在地であるさいたま市に、南は東京都板橋区・北区に接している。戸ヶ崎勤教育長は2015年に着任して以来、さまざまな改革を進めているが、そのなかに、客観的な根拠を重視した教育政策の試みがある。また、AIでは代替できない能力・AIを活用できる能力の育成を目指した産業界・大学等との連携も推進している。戸田市では、上述の埼玉県学力・学習状況調査とともに、全教員を対象とした質問紙調査も実施されている。

外部との共同研究は客観的な根拠を重視した教育政策立案のための一つとして進められ、戸田市は、学校や教室におけるデータを外部の企業や研究者に実証研究の材料として提供している。共同研究においては、複数の企業・研究者と連携することにより、外部関係者間の競争を促している。たとえば2019年度では、慶應義塾大学・筑波大学・国立情報学研究所・ベネッセ・LITALICO・IGSとの共同研究が進められている。また、LoiLo・Google・ベネッセ・富士電機ITソリューションと連携してEdTechも推進されている。これまでに約70の大学、NPO、企業等の連携を構築してきている。

### II-3-3. 大阪府箕面市

大阪府箕面市は大阪府北部に位置している。阪急電鉄箕面線が走っており、大阪圏のベッドタウンとして発展してきた（北大阪急行電鉄南北線の延伸が決まっている）。2008年、倉田哲郎が当時の最年少市長として就任（34歳）し、「子育てしやすさ日本一」を重点プランの一つに掲げた。

箕面市では「子ども成長見守りシステム」というデータベースを2014年から整備している。このシステムには、箕面市内に在住する0～18歳の子どもに関する、箕面市役所内の各部局が保有する（が散在していた）データが統合されている。そのデータには、箕面市が独自に実施する「箕面学力・体力・生活状況総合調査（箕面子どもステップアップ調査）」の結果のほか、生活保護・児童扶養手当・学童保育・医療費助成・虐待相談・教育相談・学習支援等の受給・利用状況が含まれ、それらが定期的に更新されている。箕面子どもステップアップ調査は小学1年生から中学3年生までを対象に、学力と体力の調査、学習習慣・運動習慣・生活習慣・自己認識等のアンケート調査を行うものである。ただし、公立小中学校が対象であるため、私学に通っている子どもについては学力等のデータはない。

箕面市は条例を改正し、心身の保護・生活の維持の目的に限定して、このシステムの目的外使用と外部利用提供を認めることとした。これを利用した成果の一つが、日本財団（2018）である。この論文では2014～16年にかけての3年分のデータが利用されている。

### II-3-4. 兵庫県尼崎市

尼崎市は兵庫県の南東に位置する中核市であり、東は大阪市に接している。2010年、稲村和美が女性としては当時の最年少市長として就任（38歳）した。稲村市政は2期目に公約「尼崎ネクスト30」のなかで「子育て支援センター機能の充実」を掲げ、データ収集を進めることとした。2017年には尼崎市学びと育ち研究所を設置、大阪大学教授の大竹文雄を所長として迎え、中室牧子・慶應義塾大学教授、北野幸子・神戸大学准教授、岡田知雄・神奈川工科大学教授、中尾繁樹・関西国際大学教授、西山将広・神戸大学助教の5名が首席研究員となっている。彼ら首席研究員は、尼崎市から子どもに関するさまざまな個票データを使用した研究を進め、2018年には第1回の報告会を開催している。

尼崎市は、2017年には市立小学校5年生とその保護者、市立中学校2年生とその保護者の全数を対象とする「尼崎市子どもの生活に関する実態調査」を実施し、児童生徒の生活や意識の実態を調査している。この調査の個票については希望する研究者等へ提供される制度が整っている。また、2018年からは小学1年生から中学2年生までを対象とした市独自の学力テストを開始している。

このように、いくつかの自治体では子どもに関する個票を整備し、データベース化して研究者に提供している。自治体から見れば分析を依頼・委託した形になっており、研究者から見ればデータの提供を受けた形となっている。ただし、データ提供の条件はさまざまである。そこで次に、本研究で用いる足立区のデータ構造について説明する。

## III. データ構築までの経緯

本特集号の野口ほか論文と別所ほか論文は、本プロジェクトの成果である。まず、これら論

文の問題意識・背景を解説するためにも、足立区の近年の区政運営について述べておく。

### Ⅲ-1. 区政運営と政策評価

足立区は東京都の北東に位置し、江戸四宿のひとつとして栄えた千住宿のあった地域を含む、23特別区の一つである。中心地のひとつは、JR東日本・東京メトロ・東武鉄道・つくばエクスプレスが乗り入れる北千住駅の周囲である。この千住地域は荒川の南側に位置しているが、面積の大半は荒川以北の地域が占めている。人口は2018年初では68.5万人である。

2002年11月、足立区は公募による「足立区新基本構想策定区民委員会」を設置し、その結果は翌2003年3月に「検討結果」としてまとめられた。この結果を受けて2003年7月に「足立区基本構想審議会」が設けられ、この審議会は2004年3月に答申を提出した。この答申に基づき、「協働で築く力強い足立区の実現」を基本理念とする足立区基本構想が策定され、2004年10月に足立区議会で可決された。およそ10年を経て基本構想は見直されることとなり、2015年7月に足立区基本構想審議会が設置された。この審議会の答申に基づき、2016年には「協創力でつくる 活力にあふれ 進化し続ける ひと・まち 足立」を謳うあらたな基本構想が策定された。

2004年の基本構想策定後、2007年6月に近藤やよい区長が当選した。近藤区長のもとで2009年度には、重要かつ喫緊の課題を解決するための「足立区重点プロジェクト推進戦略」が策定された。この「戦略」では、優先的に取り組むべき課題が「子ども」「暮らし」「まちづくり」「経営改革」の4分野にまとめられた。重点プロジェクトは、後述する区民評価委員会でもとくに評価の対象とされている。これら重点プロジェクトを推進するなかで、足立区は「治安・学力・健康・貧困の連鎖」の4つをボトルネック的課題、すなわち、克服しない限り区内外から正当な評価が得られない根本的課題と位置付けた(秋生2019)。治安については、刑法犯認知件数を減少させるべく、「美しいまちは安全なまち」を合言葉に「ビューティフル・ウィンドウズ運動」を展開した。学力については、

小中学校の学力テストの結果に課題があることから、基礎学力の定着を目指した試みをさまざまに進めている。本特集号の別所ほか論文はこの試みの一つを評価しようとするものである。健康については、健康寿命が都平均より2歳短いことから、「糖尿病対策アクションプラン」など糖尿病対策に特化した施策を進めている。貧困の連鎖は、治安・学力・健康の3問題の根底にある共通の原因と位置付けられ、足立区は子ども支援に政策資源を投入してきた。「東京・足立区の給食室～毎日食べたい12栄養素バランスごはん～」(足立区「おいしい給食」製作委員会2011)が発行されるなど、知名度の高まった「おいしい給食」による子どもの食生活習慣の改善もその一環に位置づけることができる。このようななか、2014年1月に「子どもの貧困対策の推進に関する法律」が施行されたことを契機に、2015年度に「未来へつなぐあだちプロジェクト(足立区子どもの貧困対策実施計画)」が策定された。このもとで、出産前からライフステージごとに早め早めの対策が試みられている。

さて、2004年の基本構想では「協働」が強調され、その実現のために、「情報公開や行政評価などの様々なしくみを通じて、協働を制度的に保障することが必要」とされた。そのため、2005年度には要綱に基づき、2004年度に実施された事業を対象として、区民評価が行われた。翌2006年度からは区民評価委員会は条例設置となった。すなわち、基本構想に従って2004年度に制定された足立区自治基本条例第15条では行政評価が義務付けられており、これに従って2006年3月には足立区区民評価委員会条例が成立したからである。この条例に基づき、足立区では引き続き区民評価を行っている。区民評価は、事業を実施する区役所・行政や、事業の立案にもかかわる区議会以外の、一般区民や学識経験者の視点から評価を行い、評価の客観性を高めるための仕組みである。

区民評価委員会は、いくつかの分科会から構成されているが、その構成は変化している。条

例設置として初の委員会である2006年度委員会では、基本計画に定める114施策全てを評価対象とし、「まちづくりと安全の分科会」「健康と福祉の分科会」「教育と産業の分科会」「環境と区民生活としくみづくり分科会」の4分科会が設置されていた。「足立区重点プロジェクト推進戦略」の策定後の2010年度委員会では、重点プロジェクトが「子ども」「くらし」「まちづくり」「経営改革」の4分野にまとめられたことを受け、「子ども」「くらし」「まちづくりと経営改革」の3分科会が設置された。2012年度には、重点プロジェクトに含まれない事務事業を評価の対象とする「一般事務事業見直し分科会」が設置され、4分科会となった。その後、2017年には新たな基本計画が始まったことを受け、重点プロジェクトを評価対象とする分科会は「ひと」「くらしと行財政」「まちと行財政」の3つへ再編され、4分科会体制となっている。

### Ⅲ-2. 研究者の関与

足立区に限らず、中央省庁や地方自治体に設置される審議会・調査会等の諮問機関や懇談会・研究会・委員会等の行政運営上の会合においては、学識経験者が委員の一部として招かれることが多い。足立区区民評価委員会は区長が委嘱する17名以内の委員から構成されるが、ここにおいても同様である。要綱設置であった第1回の会長は川崎一泰・東海大学助教授（当時）であり、引き続き、会長と分科会長は学識経験者が就くのが通例となっている。特筆すべき点として、区民評価委員会の使命はあくまでも区民目線から施策の評価を行うことであるため、学識委員もより客観的・中立的な立場から区政の評価に当たることになる。

本稿の著者のうち、野口・田中・別所は区民評価委員会の学識経験者委員として足立区との接点を持った。すなわち、野口は第1回の2005年度から委員を務め、2009年から副会長、2014～2015年度には前任の出口保行・東京未来大学教授のあとを襲って評価委員長であっ

た。田中は2012～2015年度の子ども分科会長、2016～2018年度の評価委員長を、別所は2012～2016年度の一般事務事業見直し分科会長を務めた。

さて、区民評価委員としての活動を通じて、われわれは区の持つ各種データの存在と、それらを活用した評価方法の改善の可能性に気が付いた。この点を説明するために、まず区民評価委員会での評価の方法について、野口・田中・別所の3名がともに委員であった2012～2016年度の状況を述べておく。

分科会による差異はあるものの、評価は、事前の調書等の読み込み、ヒアリング、ヒアリング後の検討、の順に進行する。まず、事務局によって評価の対象となる事業が選定され、委員には庁内評価を経た事業評価調書が渡される。この事業評価調書には、事業の目的や概要、根拠法令等と並び、活動指標や成果指標、投入資源、成果分析・課題分析が共通の様式に従ってまとめられている。活動指標とは、区の活動の度合いを示す指標であり、たとえばイベントや講座の開催回数が用いられる。成果指標は活動指標よりも政策目的に直結した指標であり、アンケートで測った満足度や、給食残菜率等が設定される。活動指標・成果指標については事前に目標値が設定され、実績値と比較した達成率も記入されている。投入資源はおもに事業費と人件費であり、区が支出した事業費とともに、国と都からの補助金や、参加費等の受益者負担金も計上される。成果分析・課題分析は、所管部署が活動指標・成果指標の達成率等をもとに行った結果であり、達成状況やその原因、指標で計上されなかったその他の実績、今後の課題や環境の変化等がまとめられている。

事業評価調書とともに事前資料を受け取った評価委員は、ヒアリングに先立って、これらの資料を基に所管部署に対する質問項目を検討する。この質問はおもに、事業に関する用語や指標の定義の確認、追加資料の請求である。検討された質問項目は所管部署に送られる。ヒアリングでは、所管部署からの質問に対する回答の

のち、更問と回答が繰り返される。

ヒアリング終了後、評価委員は評価シートを作成し、庁内評価の結果とともに公表される。区民評価は、5段階の評点とコメントから構成される。5段階の評点は、重点プロジェクトの場合、「反映結果に対する評価」「目標・成果に対する評価」「今後の方向性への評価」「全体評価」の4項目について与えられる。

このような区民評価の手続きにおいては、区役所に保管されているデータが集計されて公表される。たとえば、本特集号の別所ほか論文で検討の対象となっている小中学生に対する補習（小学生基礎学習教室・中学生補習講座）は、重点プロジェクト事業の一つとして区民評価の対象であった。この事業の活動指標は補習実施延時間数であり、成果指標として、区が実施する学力テスト（足立区基礎学力に関する総合調査）において平均正答率30%未満の児童生徒の割合や、補習の直前直後に対象児童に対して行われた学力テストの結果が用いられている。

区民評価では、公募による一般区民も評価委員として活動することや、年度ごとの速報性の必要から、単純に計算された指標が用いられることはやむを得ない。しかし、近年の計量経済学が強調するように、単純な指標では政策の効果を適切に数量的に評価することはできないことがしばしばある。このような観点から、野口を中心として、公表される集計データの背後にある個票を活用した定量的な行政評価の可能性を担当者に提案し続けることとなった。野口・田中がともに子ども分科会長であったこと、また前述したように足立区で「治安・学力・健康・貧困の連鎖」の4つをボトルネック的課題と認識しており、足立区の子どものための「育ち」に

関する現状把握が必要との知見が得られたことから、子どもに関する行政データの利活用がとくに重点的に訴えられた。

度重なる研究者からの提案・説得を受けた足立区との交渉は、2016年度から本格化し、2016年末からデータ整備が始まった。データの構造については次節で説明するが、その前に、データ利用の枠組みについて次小節で述べる。

### Ⅲ-3. データ利用の枠組み

本特集号の野口ほか論文と別所ほか論文は、われわれが実施する「子どものそだちの決定要因に関する実証的研究」に対して足立区が提供したデータを用いた分析の結果である。データ提供は、足立区と、各研究者が所属する部局がそれぞれに締結した基本協定書に従って行われた。基本協定書では、協定期間や経費負担のほか、損害賠償・守秘義務等について規定されている。データの提供にあたっては、事前に足立区情報公開・個人情報保護審議会の承認が必要とされ、提供されるデータからは個人情報削除され匿名化されている。

提供されたデータは、足立区役所内で匿名化された。ただし、この研究が始まるまでは、これらデータはしばしば異なる様式やファイル形式で、多くの場合は年度ごとに別々に、また別々の担当課・担当係によって保管されていた。また、児童生徒に共通の個人識別番号は用いられていなかった。したがって、提供前に、これらのデータを接合する、あるいは接合可能な状態に整備する必要があった。

そこで、本研究においては、上記データ整備及び匿名化処理の作業を、研究者自身が研究機関からの職員派遣による形で実施した。

## IV. データの構造

本研究で用いられているデータは、足立区内の区立小中学校に通学する児童・生徒の悉皆データである。区教育委員会は区内に在住する児童・生徒の学齢簿を保管しているから、これに子どもに関する各種調査の個票、区独自のさまざまな取り組みや学校・教員に関する情報を接合してデータベースを作成している。現在の標本期間は2009年度から2017年度であるが、いくつかの情報については接合が完了していない部分がある。

子どもに関する調査とは、ここでは「足立区基礎学力定着に関する総合調査」「足立区で実施した児童・生徒の体力の調査」「全国学力・学習状況調査」「長期欠席者リストの調査」「QUアンケート調査」の5調査である。このうち、「足立区基礎学力定着に関する総合調査」と「足立区で実施した児童・生徒の体力の調査」がわれわれの研究で最もしばしば用いられる調査となっているが、これらは区立小中学校に通う児童・生徒を対象とした調査であるため、われわれのデータベースには、足立区内に在住していても、私立学校に通っている児童生徒のデータは実質的に含まれていない。また、「足立区基礎学力定着に関する総合調査」は小学2年生から中学3年生までを対象としているため、小学1年生のデータは含まれていない。この調査は毎年4月に実施されているから、前学年の成果を計測していると解釈すると、中学3年生のデータが含まれていないとも言える。この意味では、中学から私学に進学した生徒については小学6年生のデータが欠落している。小学校については各年度1学年あたり5,000人強の児童がおり、データベースに含まれる児童生徒の延べ数は412,462人である。

区のさまざまな取り組みや教員・学校のデータとしては、9つの情報の接合を試みている。

すなわち、(1)就学援助の申請と受給状況、(2)就学前の幼稚園・保育園等の在園児名簿、(3)学校選択の状況、(4)多層指導モデル(MIM)対象児童名簿、(5)小学校基礎学習教室参加児童・中学校補習教室参加生徒名簿、(6)中学校夏季勉強合宿参加生徒名簿、(7)教員情報、(8)学校情報、(9)学級閉鎖状況、である。

これらの情報を接合し、氏名等の個人情報を削除して匿名化したものが、研究者へと提供されている。なお、これらの情報を得るにあたっては区役所内に保管されているデータのみを用い、学校現場へのデータ収集に関する協力依頼は一切行っていない。

現在われわれが利用しているデータは、これまでに記述してきたようなデータの接合が完璧に行われたものではない。さまざまな要因で欠落している部分がある。ここではそのいくつかについて指摘する。

第1に、データベースは学齢簿を基本としているが、学齢簿自体がもともとは繰り返しクロスセクションデータであるので、同一の児童・生徒を追跡できていないことがある。公立小中学校において留年はほぼ存在しないが、転校・名字の変更・氏名の表記変更等により、年度をまたいで児童・生徒を紐づけることができなかったケースがいくつかあった。

第2に、2013年度から2017年度までについては電子媒体の学齢簿を用いることができたが、2009年度から2012年度までについては利用可能ではなかったため、「足立区基礎学力定着に関する総合調査」のデータを学齢簿の代わりに用いている。区立小中学校に通う児童生徒のかなりの部分がこの調査に回答しているものの、本来の学齢簿に比べると欠落がある。その原因はいくつかある。一つには、特別支援学級の児童生徒や、あるいは病欠等により、調査に

回答していない児童生徒が存在することである。また、この「総合調査」の氏名のデータは、児童生徒が回答用紙に記入した氏名であるため、小学校低学年で漢字表記ができていなかったり、異体字（「渡辺」と「渡邊」など）を用いていた<sup>3)</sup>、名字の変更が反映されていない<sup>3)</sup>、名字の変更が反映されていない<sup>3)</sup>、本来の学齡簿の表記と異なる表記になっていることがある。このような氏名の表記ゆれや転校については目視によって紐づけをしたものの、当然のことながら完全には解決していない。

第3に、上記の「児童生徒が自ら記入した氏名を鍵とした接合」に起因する紐づけの不完全

性が、他のデータとの接合においてもしばしば発生している。また、氏名情報が利用可能でないデータもあった。

第4に、教員データについてである。現行制度では、区立小中学校の教員の人事権は東京都が持っている。そのため、足立区では区立小中学校に勤務する教員の情報は把握しているものの区外での履歴の情報は無い。すなわち、たとえば区外からの異動が発生したときに、前任校の情報や、都に採用されてからの年数等の情報は、区は保有していない。

また、もちろん、データの接合作業における人為的な誤りも存在する可能性は否定できない。

## V. 結びにかえて：今後の課題

効果的・効率的な政策の実施、国民や有権者に対する説明責任の強化の観点から、教育政策においても客観的な根拠に基づいた政策形成の推進が必要であろう。そのためには根拠を提供するデータの整備が不可欠であり、本稿ではそのようなデータの源として行政データを利用した事例を紹介するとともに、本特集号内の論文でも用いられている足立区の児童生徒データベースの構築の経緯と構造について解説を加えた。行政データは、中央政府の所有するデータであれば一国内、地方政府の所有するデータであれば所管地域内について、全数（悉皆）データを定期的に収集したものであり、数値の正確性・脱落や無回答の少なさ・観測値数の多さ等、標本調査よりも優れている特徴を有している。他方で、さまざまな理由により、行政データの学術利用は容易ではない。そのようななか、国内外で行政データの利用事例が蓄積されつつある。東京都足立区においても、研究者がデータの接合を行った児童生徒データベースが整備さ

れつつある。

行政データの学術利用のさらなる推進へ向け、日本における今後の課題をいくつか提示したい。アメリカでは2016年3月にEvidence-Based Policymaking Commission Act of 2016が成立し、これに基づいて超党派でCommission on Evidence-based Policymakingが設置された。津田・岡崎（2018）は、この委員会の最終報告書“The Promise of Evidence-Based Policymaking”の内容を紹介し、データへの適切なアクセスを阻害している要因として、一貫性のない法令・データの管理状態・煩雑な行政手続き、が指摘されているとしている。これらの問題は、日本においてもあてはまるように思われる。

われわれがデータベース構築に携わった足立区においては、ICT・データ利活用・情報システム管理推進体制を構築している（秋生2019）。この体制では、副区長をCDO（最高データ責任者）兼CIO（最高情報責任者）とし、

3) 著者の一人である別所の「別」の字も、戸籍では少し異なる。

政策経営部のなかに ICT 戦略推進担当課を設置、この課にデータ連携担当係を置くとともに、非常勤の専門家をエビデンスに基づく政策形成を推進する CDO 補佐として置いている。また、政策経営部のなかにシステム運用管理を担当する情報システム課があり、こちらに非常勤の専門家をシステム運用管理を担当する CIO 補佐として配している。このようなデータインフラの制度的整備が行政側にも必要かもしれない。

このようなデータインフラ整備のひとつの到達点は、イングランド等の事例で見られたように、匿名化された個票データを外部の研究者が二次利用・分析できるような仕組みであろう。行政側が研究課題を公募し、審査を通った研究者がデータをオンサイト施設やリモートアクセスによって利用する制度は、税務データについてすでにイギリスやフィンランドで運用されている (Almunia et al. 2019)。

三菱総合研究所 (2016) は、自治体に対してエビデンスに基づく教育政策の現状や課題等についてのアンケートを行った結果を示している。それによれば、個人情報を含むデータを自

治体外部の研究者等へ提供できる自治体は都道府県で 6.4%、指定都市で 5.0%、中核市で 7.4% に過ぎない。他方で、エビデンスに基づく教育施策を進めていくうえでの課題として、都道府県・指定都市・中核市のいずれにおいても過半数の自治体が「エビデンス収集や分析の知見を有する人材が不足している」を挙げている。本稿で取り上げた国内事例ではいずれも外部人材との連携を進めていたし、国際的に最も先進的な事例の一つであるイギリスの NPD においても、形成過程の初期においては研究者と行政担当者との個人的なつながりが大きな役割を果たしていた (田中 2019)。日本においては、研究者の側が行政の論理を理解していないとの批判もある (鈴木 2018)。短期的には行政と研究者の目的にずれ違いはあるかもしれないが、長期的には、両者はより好ましい政策・よりよい社会の形成という点で、その目的はより似通ったものであるはずである (Almunia et al. 2019)。客観的な根拠に基づいた政策形成の基盤となるデータ整備のための官学連携は、今後の大きな課題の一つであろう。

## 参 考 文 献

- 秋生修一郎 (2019) 「未来へつなぐあだちプロジェクト：計画策定とデータ活用」神奈川県政策研究フォーラム発表資料 (2019 年 1 月 29 日)
- 足立区「おいしい給食」製作委員会 (2011) 『日本一おいしい給食を目指している 東京・足立区の給食室 毎日食べたい 12 栄養素バランスごはん』泰文堂。
- 内山融・小林庸平・田口壮輔・小池孝英 (2018) 「英国におけるエビデンスに基づく政策形成と日本への示唆—エビデンスの「需要」と「供給」に着目した分析—」RIETI Policy Discussion Paper Series 18-P-018
- 鈴木亘 (2018) 「EBPM に対する温度差の意味するところ」『医療経済研究』30 巻 1 号, pp. 1-4.
- 曾我謙悟・待鳥聡史 (2007) 『日本の地方自治：二元代表制政府の政策選択』名古屋大学出版会。
- 田中隆一 (2019) 「根拠に基づく政策形成と統計的会計検査：イギリスおよびフィンランドの教育・統計データベース調査を基に」会計検査院平成 30 年度海外行政実態調査報告書。
- 津田広和・岡崎康平 (2018) 「米国における Evidence-based Policymaking (EBPM) の動向」RIETI Policy Discussion Paper Series 18-P-016.
- 豊福晋平 (2017) 「教育分野におけるオープンデータ活用の実態」『情報処理学会研究報告 IPSJ SIG Technical Report』2017-CLE-23, pp. 1-4.

- 豊福晋平 (2018) 「教育のオープンデータは普及するか」『情報処理』59巻5号, pp. 467-470.
- 日本財団 (2018) 「家庭の経済格差と子どもの認知能力・非認知能力格差の関係分析. 25万人のビッグデータから見えてきたもの」
- 林正義 (2008) 「地方分権の経済理論」貝塚啓明・財務総合政策研究所編著『分権化時代の地方財政』中央経済社, 第2章, pp. 43-69.
- 三菱総合研究所 (2016) 「学力調査を活用した専門的な課題分析に関する調査研究報告書」平成27年度文部科学省委託調査「全国学力・学習状況調査の結果データベースの構築・運用に関する調査研究」
- 三菱総合研究所 (2018) 「エビデンスに基づく教育政策の在り方に関する調査研究報告書」平成29年度文部科学省委託調査「教育改革の総合的推進に関する調査研究」
- Almunia, M., Harju, J., Kotakorpi, K., Tukiainen, J., & Verho, J. (2019), "Expanding access to administrative data: the case of tax authorities in Finland and the UK", *International Tax and Public Finance*, vol. 26, no. 3, pp. 661-676.
- Card, David E., Raj Chetty, Martin S. Feldstein, and Emmanuel Saez. (2010), "Expanding Access to Administrative Data for Research in the United States", *National Science Foundation White Paper*.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. (2014), "Where is the land of opportunity? The geography of intergenerational mobility in the United States", *Quarterly Journal of Economics* vol. 129, no. 4, pp. 1553-1623.
- Crato, Nuno, and Paolo Paruolo. (2019), "The Power of Microdata: An Introduction", In: Crato, Nuno, and Paolo Paruolo (eds), *Data-Driven Policy Impact Evaluation*, Springer Open, pp. 1-14.
- Einav, Liran, and Jonathan Levin. (2014a), "Economics in the Age of Big Data", *Science*, vol. 346, no. 6210, 1243089.
- Einav, Liran, and Jonathan Levin. (2014b), "The Data Revolution and Economic Analysis", *Innovation Policy and the Economy* vol. 14, pp. 1-24.
- Figlio, D., Karbownik, K., & Salvanes, K. G. (2016), "Education Research and Administrative Data", *Handbook of the Economics of Education*, vol. 5, pp. 75-138.
- Figlio, D., Karbownik, K., & Salvanes, K. (2017), "The Promise of Administrative Data in Education Research", *Education Finance and Policy*, vol. 12, no. 2, pp. 129-136.
- Jarmin, Ron S., and Amy B. O'Hara. (2016), "Big data and the transformation of public policy analysis", *Journal of Policy Analysis and Management* vol. 35, no. 3, pp. 715-21.
- Kum, Hye-Chung, Stanley Ahalt, and Thomas M Carsey. (2011), "Dealing with Data: Governments Records", *Science*, vol. 332, no. 6035, pp. 1263.
- Larrimore, J., Mortenson, J., & Splinter, D. (2019), "Household Incomes in Tax Data: Using Addresses to Move from Tax Unit to Household Income Distributions", *Journal of Human Resources*, forthcoming.
- Mervis, Jeffrey. (2014), "How two economists got direct access to IRS tax records", <https://www.sciencemag.org/news/2014/05/how-two-economists-got-direct-access-irs-tax-records>.
- Piketty, T., & Saez, E. (2003), "Income Inequality in the United States, 1913-1998", *Quarterly Journal of Economics*, vol. 118, no. 1, pp. 1-41.
- Røed, K., & Raaum, O. (2003), "Administrative registers-Unexplored reservoirs of Scientific Knowledge?", *Economic Journal*, vol. 113, no. 488, pp. F258-F281.