

EXECUTIVE SUMMARY

平成 30 年度 調査事業報告サマリー

子ども・子育て支援推進調査研究事業（調査研究課題番号 26）

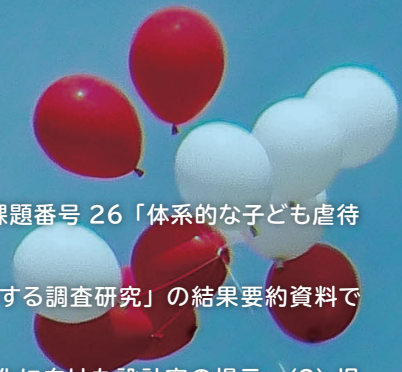
体系的な子ども虐待データベースの構築及び  
データに基づくリスクアセスメントの効果に関する調査研究



# CONTENTS 目次

はじめに（背景と目的）	P2
総合要約	P2
<b>第1部</b> 共通データベース化に向けた設計案	P3
背景と目的	P3
全国共通データベース構築に向けたプロセス	P3
必要なデータプラットフォームに向けた提案	P4
データベース構築と運用に向けた前提条件	P5
標準的データベース構築に向けて	P5
<b>第2部</b> 児童相談所データベースの項目例と活用可能性の検討	P6
目的	P6
統計解析の方法	P6
結果	P6
1. 基礎統計	P6
2. 再相談関連要因の検討	P7
3. 対応長期化との関連要因	P8
4. 機械学習の結果	P9
考察	P10

# はじめに



本稿は、平成 30 年度子ども・子育て支援推進調査研究事業 調査研究課題番号 26「体系的な子ども虐待データベースの構築およびデータに基づくリスクアセスメントの効果に関する調査研究」の結果要約資料である。本事業では、(1) 全国自治体関係機関における共通データベース化に向けた設計案の提示、(2) 児童相談所データベースの項目例とその活用可能性を検討した。

## 総合要約 IMPORTANT DISCOVERY

本事業の主たる目的は次の二点である。

1. 情報集約システムを活用した情報共有や関係機関の連携についての好事例を把握し、当該事例で用いられている情報集約システムが児童虐待の端緒となる要因の把握と、児童虐待の発生予防・早期発見に資すること示すこと。

2. 虐待統計の国際比較等を調査した平成 29 年度報告等を踏まえ、国際比較が可能な調査項目を用いた虐待統計のモデル実施を行うこと。

これらの目的を充足させるための成果物として、本稿では、①該当する自治体への実態調査を行った上で好事例の取りまとめを行い、内容を整理・分析する、②国際比較可能な虐待統計のモデル実施の結果を踏まえた分析と、改善点等をまとめた提言を行うことの 2 点が掲げられた。

情報連携と関係機関の連携に関する項目①については、リスクアセスメントが関係市町と共有されている三重県児童相談所を取り上げ、その項目情報について整理した。そして、詳細情報がデータベースに蓄積されたリスクアセスメントの情報を活用し、対応日数や虐待再相談率の関連要因を示した。事例の対応日数や虐待再相談は、児童が安全ではない状態に長期的に晒される可能性を示す一つの指標である。それらの関連要因が抽出されたことは、長期的・継続的な虐待が疑われる児童の安全確保に関して、予防・早期発見の観点から大きな貢献が得られたものと考えられる。また、三重県だけでなく、情報集約システムの活用可能性を検討する上で、神奈川県および千葉県のデータベースからも有益な知見が得られた。解析結果を踏まえ、データベースの利活用、ならびにリスクアセスメント項目の活用に関して、(a) 事例の予後予測が可能になること、(b) 実例に基づくリスクアセスメント項目の定量的な評価と改善、(c) 虐待事例に関する基礎知識の創出、(d) 共通言語として事例の重篤性が関係機関で共有可能になること、これら 4 つの観点から有用性を整理した。

国際比較が可能な情報項目を用いた分析と改善点に関する項目②については、国際比較に耐える研究用データベースに必要な機能・体制について 5 つの観点から整理した。(A) リスクアセスメント項目を含めたデータベース項目の統一・標準化、(B) データ入力画面等のデザイン / 入力の容易性を支える配慮、(C) 自治体単位ではなく全国で情報の連結を想定したデータベース設置環境の検討、(D) データ利活用を前提としたデータ“出力”機能の搭載、(E) 国際比較を前提とした海外のデータベース設計方法への対応である。現場での利便性を基礎とし、データに基づく各種検証と情報連携を実現するためのデータベース導入が期待される。

なお、情報共有を目的とするデータベースを実装する際の着眼点として、(1) 最新のフィードバックを常に他機関から受けられる都道府県間を越えた情報共有データベースの構想を持つこと、(2) 全国区での情報共有を可能とするデータベースによって今後の拡張性を含むコストの削減や効率性の向上が見込まれることが指摘された。

## 共通データベース化に向けた 設計案

### 背景と目的

児童虐待に関するデータベースには、「記録」と「活用」という大きく目的の異なる2つの機能が並存している。児童への対応を明確に「記録」しておくための項目には、例えば通告日時や通告経路などの情報がある。こういった項目は、児童相談所や自治体によって定められた業務手続きに沿うことで必然的に定立される。すでに自治体ごとに「記録」を管理するためにデータベース利用している場合も多い。一方、従来のような「形式的な情報 (e.g. 通告経路)」

だけでなく、事例の内容的側面を端的かつ効果的に共有する、「活用」を想定したデータベース項目を導入している自治体は少ない。そして、各自治体の所有するデータベース同士のシステム構造が全く異なることも「活用」を阻む一つの要因と考えられる。

協力の得られた自治体のデータベース項目を援用し、「自治体の共通データベースを設計する場合に、どのような項目が必要か」について整理する。具体的には、海外データベースの参照、リスクアセスメントに有用可能性のある項目などを調査し、その結果を整理する。児童相談対応に係る標準データベースの設計により、将来的には各関係機関(市区町村や保健センター、医療機関、警察、学校、園など)のデータベース同士が接続されやすくなり、児童虐待事例の内容を円滑に共有可能となることが期待される。

### 全国共通データベース構築に向けたプロセス

全国で統一した標準児相データベースを作るのであれば、まず国として標準データベース構造を一つ設定し、それに各自治体データベースのデータ構造を定義することが一番低コスト、かつ自治体毎の調整がしやすい。もしも各自治体に、データベース項目の調整を任せれば、局所的に整合性をとるようなデータベース項目の変更があちこちで起こり、結果的に多大なコストがかかることになるだろう。

また、全国で統一した標準児相データベースを設計する上で、3点重要な要素が存在する。1つ目は、先に述べたとおり、標準児相データベースが関係機関のデータベースと連携可能なように拡張性のある設計を行うことである。2つ目は、業務フローを抑え、現場職員の方々のニーズに対応した設計にすることである。これは、現場職員の方々がどのようなデータをいつ、どんなタイミングで、どのように入れるのかなどを知った上で、ほしい情報をすぐに見つけられるようにしたり、複数のデータを結合したり(例：単独で統一されている児童を、一家族として結合するなど)できるようにすることである。3つ目は、高い秘匿性を保つデータベース設計を行うことである。児童虐待に関するデータベースを作成するうえで、高い秘匿性が求められる個人情報をどのようなネットワークで接続するかといった課題が伴う。くわえて、必要な情報を効率的に記録・蓄積するためのインターフェースについての工夫も必要となる。

共通データベースを最終目標とした2019年現在、紙を使った運用が残っている自治体も少なくなく、経過記録を紙で管理している児童相談所もある。電子化していても、リスクアセスメントは全て紙、あるいは合計得点のみ電子化されており、各項目は紙しか残っていないという場合も多い。この段階から、構想して実装し、そして実証実験まで考えると1都道府県あたり2~3年程度の期間がかかることが予想される。

### データベース (DB) 整備に向けたプロセス



## 必要なデータプラットフォームに向けた提案

虐待対応用のデータベースは基本的に2つ存在し、一つは現場における情報共有用のデータベースであり、もう一つは研究用のデータベースである。前者を構築することで、現場職員は経過記録を書くだけで、自動的に福祉行政報告例のアウトプットが可能になる。そして、関係機関同士で開示可能な情報を適切に制御しつつ、適切に関係機関情報にアクセスすることができるようになる。しかしながら、政策や運用上の調整を考えると、児童相談所と関係機関までのデータベース共有も兼ねた標準的データベースを作成することよりも、最優先すべきは、ハイリスクケースを抱える全国の児童相談所同士の情報共有データベースを構築することが望ましいと考える。後者の研究用データベースを構築することで、研究者は自動整形された研究用データベースにアクセスし、新たな知見を発見することができるようになるかもしれない。研究者が研究用データベースにアクセス可能になることによって、現場へのアンケート調査を最小しつつ、現場に活用可能な知見を発見することにつながるかもしれない。そして、この研究用データベースは、現場職員の経過記録をもとに作成された情報共有データベースをもとに自動的に生成されるため、現場職員も研究者も大きな負担がなくなるのが利点である。共通データベースの構築によって、現場職員の負担軽減、研究者の自由度向上、省庁がリアルタイムに統計値を集計できるなど、三方良しの仕組みとなる。

また、標準データベースの構築を考えるにあたっては、データベースそのものだけでなく、業務効率化に向けた包括的なプラットフォーム開発を合わせて行うことが重要である（下図）。現場に必要なのは最新情報が入ってきたら、自動でリアルタイムに同期される仕組みである。これらを可能にするには、ただ自治体内でデータを溜めて共有するという発想ではなく、自治体内外であっても、同様の情報共有ができるようなデータベースの構築をイメージする必要がある。

そのための基盤は LGWAN（Local Government Wide Area Network）として既に地方自治体向けにサービスが提供されている。LGWAN では、インターネットから遮断され、地方公共団体のみがアクセスでき、また LGWAN 内でのセキュリティが担保されていることで、個人情報などの機微情報も載せて良い運用ルールとポリシーとなっている。

### 必要なデータプラットフォームへ

データベースはシステムの一部。さらに現場を支えるテクノロジーの開発へ



#### タブレットアプリ

使いやすいUI/UXを整えたアプリ



#### API等による柔軟な制御

自治体間を越えた  
他機関情報共有とアクセス権管理



#### クラウドデータベース

セキュリティの高い監視と柔軟で使いやすいDB構造  
研究に必要なデータは、個人情報は自動でマスクされ  
研究者が自分で取得可能にすれば、現場アンケート調査は  
最小限にできる



#### 将来的にAIやRPAなどの自動化

標準的なデータプラットフォームが溜まることで、  
リスクの自動分類、リアルタイムシミュレーション、  
経過記録の自動化、データに基づく人材育成へ



#### 子ども安全にのみフォーカスできる職場環境

事務や記録などをとことん効率化し、  
児童相談所職員の方々にとって  
ほぼ100%子どもの安全に勤務時間を割けるようにする

## データベース構築と運用に向けた前提条件

データベース構築と運用に際し、前提条件が4つある。1つ目は、運用する上で、データ入力の件数コントロール（全国共通計上方法）を決定し、データベース入力定着に向けた研修といった人材育成を継続的に行う必要がある。現状、各自治体のデータ入力ルールが全国的に明示されていないため、統計の取り方やケースのカウントの仕方自体も自治体によって異なっている。たとえ標準データベースを導入しても、データ入力ルールが統一されていないと結果の歪みにつながる可能性がある。

2つ目は、現場実務の負担を総合的に軽減できるようなデータベース構築を行うことである。具体的には、使いやすいフロントデザイン（外観）、福祉行政報告例の自動エクスポート（排出）機能、リスクアセスメントの各項目得点の保存によるリスク評価遷移の可視化などを満たすことが特に重要である。

3つ目は、研究用データベースの機能として、国際比較や年度をまたぐ検討のために、データを一括エクスポート可能な機能やAPIを利用した自動抽出及び整形機能を持つことである。ただ、これは技術的な問題であり、基本的にはデータベース設計時に容易に解決可能であると考えられる。

4つ目は、情報共有データベースと研究用データベースを用いて、政策・人員予算請求・業務管理に活用できるようなデータベース設計を行うことである。具体的には、翌年度における虐待通告数、一時保護数、措置数の予測を立て、その数に基づく予算請求ができるような資料や、政策や人員配置予算、及び業務量管理に参照すべきは、例えば通告数の平均値だけでなく、最も忙しい日時はいつか、また最大瞬間風速だと何件の通告がくるかといった予測資料を作成するための基礎材料となる情報を現場職員が容易に準備可能になることをさす。

また、児童相談所の運用効率化は、その地域の特性を吟味することが重要である。本来、人数の設置基準について、国として都市部・地方部に対する公平性から、子ども当たりの配置人数を一律にすることは現状致し方ないところがある。ただし、吟味すべきは、その計算基準は児童人口だけで良いのかという問題である。児童人口だけでなく、本来は地域における犯罪率や貧困率など、ケースの困難度（リスク）などについても調整した形で、配置人数を議論することが望ましいと考えられる。即ち、データに基づいた判断が可能になると、全て一律の人員配置基準は公平でない可能性もある。いずれにせよ、配置基準なども運用面において、データに基づく判断が重要である。

## 標準的データベース構築に向けて

児童相談所の標準的データベースを作る際には、現場の情報共有データベースと研究用データベースが想定される。まずは現場で使いやすい業務用（連携機能含む）データベースを作り、そこから自動的に研究用データベースとして抽出できる形式が望まれる。これら構築に際し、現場のデータをまとめるという点において、一定の費用が必要不可欠であり、国から各自治体への予算補助は必須といえる。そのため、標準的なデータ構造を全国統一で示すという国主導のスタンスが必要となる。そのため、都道府県をまたぐ情報共有を前提に、標準データベースとそれに関する情報セキュリティ基準について、厚労省をはじめ、情報セキュリティやネットワークを管轄する総務省、情報技術を構築する経産省との調整が重要と考えられる。

## 目的

第二部におけるデータ解析の目的は、蓄積されたデータベースの活用方法を実践的に示すとともに、現状のデータが有する特徴や情報量の多さについて検討することである。そこで、協力の得られた3つの自治体（三重県・神奈川県・千葉県）のデータを用いたデータ解析を実施する。具体的には、データの活用により、(1) 事例の再相談要因の検討、(2) 事例の長期化関連要因の検討、(3) 予後予測のためにデータベースにどの程度の情報量が含まれるかの検討を行う。三重県については、5年間に及ぶリスクアセスメント項目データが含まれており、その活用可能性についても検討する。

## 統計解析の方法

再相談予測の関連要因を検討する際にはロジスティック回帰モデルを利用した。ロジスティック回帰モデルは、目的変数（予測対象）が「あり・なし」などの二値の場合に用いられる回帰分析の一種であり、説明変数（予測のための変数）ごとに「目的変数とどの程度の関連性が認められるか」についての係数値とオッズ比を得ることができる。

事例の対応日数に関連する要因の検討では、負の二項分布を用いた回帰モデルによる解析を行なった。解析で得られる結果は「○があるほど、対応期間が長い」「▲▲に該当するほど、対応期間が短い」といった関連性を示す係数値である。

機械学習とは、過去のデータに潜在するパターンを数量的に調べ上げ、見つかったパターンに従って将来の結果を予測する解析手法の総称を指す。その性能を示す指標として、本稿では精度（全予測対象のうち、何パーセントの事例について正しく予測できたか：Accuracy）を主に用いる。本稿で機械学習を利用する目的は、「データベースが、事例の予後を予測するためにどの程度の情報量を有しているか」について検討することである。情報を引き出すための手段として、具体的には、勾配ブースティング木（Extreme Gradient Boosting Tree, XGBoost）、ランダムフォレスト、全結合多層ニューラルネットワーク、バイズ多層ニューラルネットワーク、それらの結果を統合したアンサンブルモデルを利用する。

## 結果

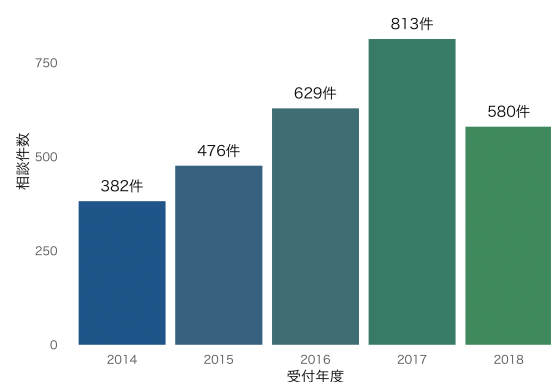
本要約資料では、協力を得られた自治体のうち、リスクアセスメントに関する複数の項目をデータベースに蓄積している三重県のデータを主として報告することで、リスクアセスメントデータの活用例を示す。なお、以下で利用する「再相談」は、「対応期間中か否かに拘らず、同一児童IDが将来的に反復出現しているレコード」と定義した。自治体によってデータベースへのレコード登録方式が異なるため、単純に自治体間で比較できないことに留意されたい。

### 1 基礎統計

三重県データベースから抽出したデータセットを解析用に結合した結果、解析対象となったのは述べ7395件の虐待相談レコードであった。年度別の虐待通告件数は、2017年度で222件となっている。2018年度については、11月31日時点までのデータであるため、参考値となる（右図）。

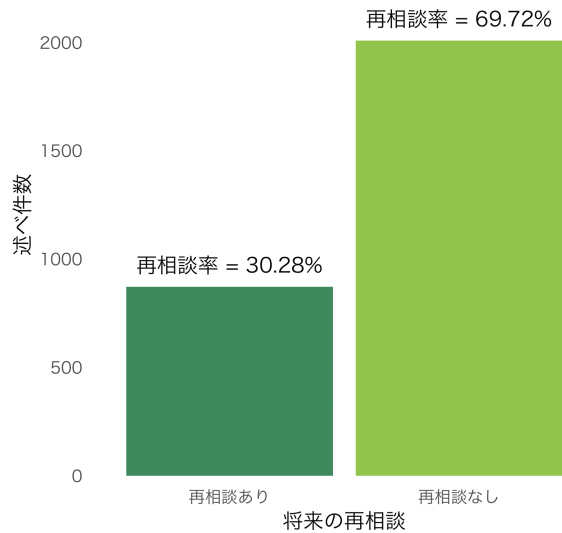
対応日数が算出可能なデータは全件のうち1026件であったため、日数別の割合を報告する。三重県においては、身体的虐待と心理的虐待の事例が占める割合が多く、その3割から4割程度が一年以上の対応事例であった（次ページ右パネル）。また、述べ件数で計上した再相談の割合は、30.28%であり、全体の4分の1程度を占めた（次ページ左パネル）。

三重県の年度別虐待通告述べ件数  
2018年度は2018年11月31日までのデータ



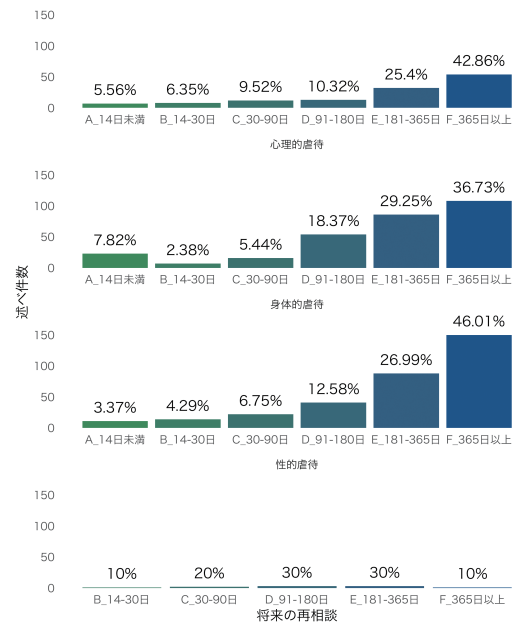
## 三重県における再相談発生率

将来的な再相談事例を1とカウント(述べ件数を利用)



## 三重県：虐待種別対応日数の割合

将来的な再相談が観測された事例を1とカウント(述べ件数を利用)



## 2 再相談関連要因の検討

三重県におけるリスクアセスメントデータを用いたロジスティック回帰を実施した。一部内容の重複するアセスメント項目(e.g. 種別性虐待と性虐待(疑い含む)や、児童の年齢と0-5歳未満の項目)などは、多重共線性のリスクを考慮し片方を除外した。推定の結果得られた調整オッズ比(Adjusted Odds Ratio, 以下 OR)とその95%信頼区間(95%Confidence Interval, 以下95%CI)を算出したところ、最も後の再相談と関連したのは過去の係属歴(OR = 2.55, 95%CI[2.31, 2.83])であった。さらに、保護者の年齢が若い(OR = 1.49, 95%CI[1.32, 1.72])、首から上の傷あざ・腹部の打撲痕(OR = 1.23, 95%CI[1.08, 1.40])、児童が帰宅することに恐怖を感じている場合(OR = 1.23, 95%CI[1.04, 1.41])、保護者が攻撃的な態度を有する場合(OR = 1.17, 95%CI[1.05, 1.31])、児童に理由不明の傷あざがある場合(OR = 1.15, 95%CI[1.03, 1.28])であった(下図)。なお、これらは全て「再相談と関連する要因」「該当により再相談が予測される要因」であって、因果関係ではないことに留意されたい。

### 要因別による、虐待再発確率(オッズ比)

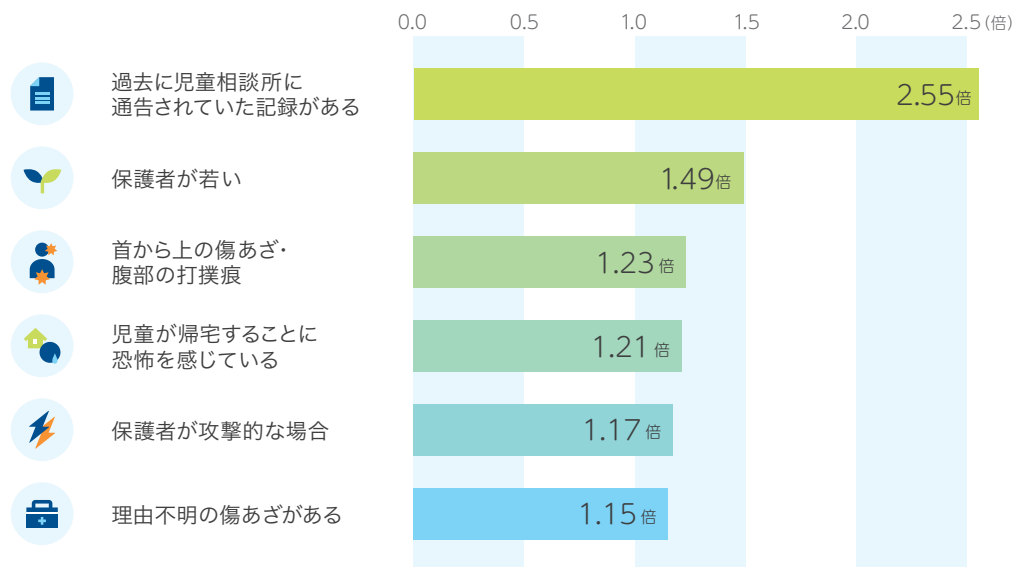


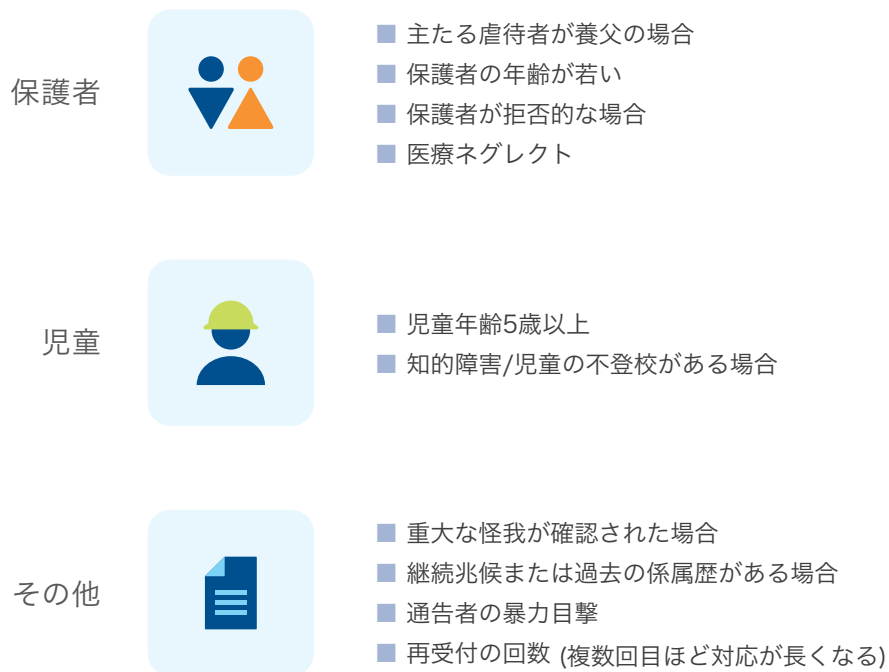
図 三重県：将来の再相談に関わる関連要因(調整オッズ比)



### 3 対応長期化との関連要因

リスクアセスメント項目を含めた三重県データを活用し、負の二項分布を用いた事例の長期化要因について検討を行なった。その結果、主たる虐待者が養父の場合 ( $\beta = 0.15$ )、保護者の年齢が若い場合 ( $\beta = 0.02$ )、保護者が拒否的な場合 ( $\beta = 1.15$ )、医療ネグレクトが該当する場合 ( $\beta = 1.37$ )、児童の年齢が5歳以上の場合 ( $\beta = 0.27$ )、知的障害を伴う場合 ( $\beta = 0.54$ )、不登校がある場合 ( $\beta = 0.95$ )、重大な怪我が確認された場合 ( $\beta = 0.16$ )、虐待の継続兆候が認められる場合 ( $\beta = 0.52$ )、過去の係属歴がある場合 ( $\beta = 0.29$ )、通告者が暴力を目撃していた場合 ( $\beta = 1.46$ )、再相談の回数が多い場合 ( $\beta = 0.52$ ) に、事例対応が相対的に長期化する傾向が示唆された(下図)。

#### 虐待事例対応が長期化する要因



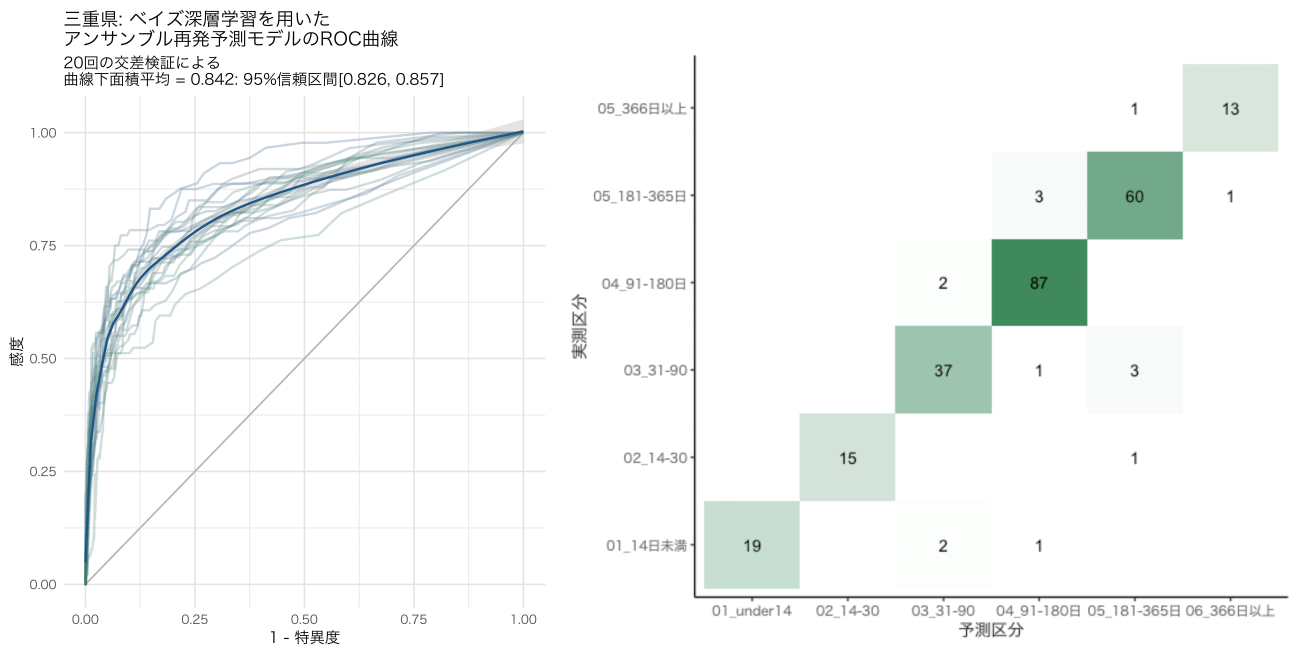
#### 神奈川県・千葉県における長期化関連要因について

神奈川県データを用いて事例の対応長期化関連要因について検討した結果、長期化を積極的に予測する要因として児童福祉施設の入所対応 ( $\beta = 1.50$ )、市からの相談・通告 ( $\beta = 1.07$ )、児童福祉司指導対応 ( $\beta = 0.95$ )、障害程度 = 2 ( $\beta = 0.47$ )、教育委員会からの相談 ( $\beta = 0.3$ )、障害程度 = 1 ( $\beta = 0.26$ )、ネグレクト事例 ( $\beta = 0.2$ )、主たる虐待者が実母の場合 ( $\beta = 0.11$ ) となった。千葉県のデータでは、一時保護の実施 ( $\beta = 0.68$ )、教育委員会・学校からの通告 ( $\beta = 0.49$ ;  $\beta = 0.48$ )、措置の実施 (施設・里親等全て) ( $\beta = 0.45$ )、安全確認の回数が多いほど ( $\beta = 0.29$ )、市区町村や指定都市などの自治体からの通告 ( $\beta = 0.25$ ;  $\beta = 0.19$ )、母親からの相談 ( $\beta = 0.18$ )、医療機関からの通告 ( $\beta = 0.16$ ) において、相対的に対応日数が長期化する傾向が認められた。

## 4 機械学習の結果

三重県データでは、リスクアセスメントデータを含めた虐待対応レコードのうち、60% を学習データ、20% をアンサンブル学習用データ、残りの20% を検証用データに無作為に分割し、機械学習を行なった。20回の交差検証法による性能評価の結果、リスクアセスメント時点の情報から将来の再相談に対する予測精度は Accuracy = 0.896 となった（平均精度の95%CI [0.89, 0.90]）。20回の交差検証法で得られた結果を利用し、ROC 曲線を図示した（下図左パネル）。その結果、平均曲線下面積は 0.842 となり、その95%信頼区間は下限が 0.826、上限が 0.857 となった。

対応日数を予測するため、リスクアセスメント項目を含めたデータセットの90% を学習データとし（ランダムに選抜）、残りの10% を検証データとして機械学習モデルの構築を行なった。日数自体を目的変数として学習させ、予測日数とともにカテゴリ化し、その精度を検証対象とした。その結果、6カテゴリ（マルチクラス）での精度は 0.934 となり、対応日数を高い精度で予測することができた（下図右パネル）。



千葉県および神奈川県データを活用した機械学習では、XGBoost を用いた再相談予測・対応日数予測を行なった。結果を以下に示す。

	三重県	神奈川県	千葉県
将来再相談の予測精度	89.6%	82.4%	74.6%
対応日数の予測精度 (6 カテゴリ)	93.4%	37.5%	42.2%

## 考察

リスクアセスメントの元来の目的は、事例が有する様々なリスク（死亡リスクを代表とし、児童安全が脅かされる諸々のリスク等）を取りこぼしなく把握し、適切な対応につなげることである。それゆえ、リスクアセスメントは事例の内容的色相を強く帯びる情報となる。言い換えれば、該当項目の内容によって、事例の特徴が理解・イメージできる情報となる。そのようなアセスメント項目を運用し、データベースに蓄積することにより、複数の活用可能性が展開される。第一に、本項で示した通り、事例の予後予測が可能となる。後の再相談や対応期間の長期化が項目への該当から具体的イメージを持って見通せるようになり、したがって、児童が継続的に危険状態に晒される可能性をあらかじめ見積もることができるようになる。

第二に、実例に基づくアセスメント項目の検証が可能になる。項目がもつリスク検知能力を評価し更新することで、より高い精度で危険性を見落としを防止する基準の策定が可能になる。データを元にした評価項目を用いることにより、より客観的な評価方法が定立可能となる。

第三に、リスクアセスメント項目を基本とした知識の集積が可能となる。例えば、リスクアセスメント項目の蓄積によって、どのような危険性を抱える事例が存在するのか、そのパターンを把握することが可能となる。多種多様な事例の様相をもつ虐待事案の中に、共通性やパターンが共通認識的知識として得られれば、より詳細を検討する研究へとつながるだけでなく、効果的な対応を考える研修材料としても活用することができる。

そして第四に、“共通言語”として事例の重篤性を情報共有可能になる。児童相談所を中心とする関連機関で、基本情報とともにアセスメント項目情報が共有されることで、その予後予測知見や重篤どの認識の共通化を統一的な枠組みで図ることができる。リスクに関する基盤的な情報が共通化されることによって、円滑な連携につながることを期待される。



国立研究開発法人産業技術総合研究所 人工知能研究センター  
高岡 昂太 (受託代表者)・坂本 次郎・北條 大樹・橋本 笑穂・  
坂上佐知子・村川 尚子・先光 毅士・北村 光司・本村 陽一

社会福祉法人 恩賜財団母子愛育会 愛育研究所  
客員研究員 山本 恒雄

福岡市こども総合相談センター  
所長 藤林 武史

千葉県市川児童相談所  
所長 渡邊 直

三重県児童相談センター  
センター所長 鈴木 聡

神奈川県中央児童相談所  
課長 鈴木 浩之

愛育クリニック小児精神保健科  
部長 小平 雅基

大阪大学大学院 医学系研究科  
特任助教 安藤 絵美子

東京医科歯科大学  
プロジェクト助教 伊角 彩

University of Oklahoma Health Science Center  
Research Fellow 山岡 佑衣

---

〒135-0064 東京都江東区青海 2-4-7

産業技術総合研究所 臨海副 都心センター 別館 (バイオ・IT 棟)

03-3599-8108 (研究班 代表電話番号)

2019/3/31